BGRS/SB-2020

The 12th International Multiconference

Bioinformatics of Genome Regulation and Structure/ Systems Biology

BGRS/SB-2020
NOVOSIBIRSK, RUSSIA
06–10 JULY, 2020

https://bgrssb.icgbio.ru/2020/

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences

Novosibirsk State University

BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE/SYSTEMS BIOLOGY (BGRS/SB-2020)

The Twelfth International Multiconference

Abstracts

06–10 July, 2020
Novosibirsk, Russia

Novosibirsk
ICG SB RAS
2020

УДК 575
В60

Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2020) :
The Twelfth International Multiconference (06–10 July 2020, Novosibirsk, Russia); Abstracts /
Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences;
Novosibirsk State University. – Novosibirsk: ICG SB RAS, 2020. – 683 pp. – ISBN 978-5-91291-051-7.

DOI 10.18699/BGRS/SB-2020-000

Organizers

Institute of Cytology and Genetics, SB RAS (ICG SB RAS)

Novosibirsk State University (NSU)

Ministry of Science and Higher Education of the Russian Federation (Minobrnauki of Russia)

The Vavilov Society of Geneticists and Breeders

State Scientific-Research Institute of Physiology and Basic Medicine (PhBMRI)

Vavilov Journal of Genetics and Breeding

Contacts
All questions, please contact the official address
of the organizing committee of the conference: bgrs2020@bionet.nsc.ru

Organizing Committee Chairman: Zubova Svetlana
Address: Lavrentieva Avenue, 10, Novosibirsk, 630090
Phone: +7 (383) 363 4977, e-mail: svetazubova@gmail.com

ISBN 978-5-91291-051-7

Program committee

Nikolai Kolchanov, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia (Conference Chair)
Ralf Hofestädt, University of Bielefeld, Germany
Mikhail Fedoruk, Rector of Novosibirsk State University, Novosibirsk, Russia
Alexey Kochetov, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Dmitry Afonnikov, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Lubomir Aftanas, State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk, Russia
Konstantin Anokhin, SIC "Kurchatov Institute", Moscow, Russia
Yurii Aulchenko, Novosibirsk State University, Institute of Cytology and Genetics of SB RAS, PolyOmica, Novosibirsk, Russia
Matteo Barberis, University of Surrey, Guildford, United Kingdom
Tatyana Chernigovskaya, St. Petersburg State University, St. Petersburg, Russia
Valery Danilenko, The Vavilov Institute of General Genetics of RAS, Moscow, Russia
Sergey Deyev, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of RAS, Moscow, Russia
Sergey Goncharov, Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia
Vladimir Ivanisenko, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Irina Ivshina, Institute of Ecology and Genetics of Microorganisms of UB RAS, Perm, Russia
Elza Khusnutdinova, Institute of Biochemistry and Genetics, Ufa Scientific Center of RAS, Ufa, Russia
Natalya Kolosova, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Olga Krebs, Heidelberg Institute for Theoretical Studies, Germany
Dagmar Kulms, University Hospital Carl Gustav Carus Technical University, Dresden, Germany
Sergey Lashin, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Inna Lavrik, Otto von Guericke University, Magdeburg, Germany
Olga Lavrik, Institute of Chemical Biology and Fundamental Medicine of SB RAS, Novosibirsk, Russia
Andrey Letyagin, Research Institute of Clinical and Experimental Lymphology – Branch of Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Vsevolod Makeev, Vavilov Institute of General Genetics of RAS, Moscow Institute of Physics and Technology, Moscow, Russia
Mikhail Marchenko, Institute of Computational Mathematics and Mathematical Geophysics of SB RAS, Novosibirsk, Russia
Yuri Matushkin, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Mikhail Moshkin, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Anton Nizhnikov, Saint Petersburg State University, St. Petersburg, Russia
Galina Orlova, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Andrey Palyanov, A.P. Ershov Institute of Informatics Systems of SB RAS, Novosibirsk, Russia
Nikolay Pimenov, The Federal Research Centre "Fundamentals of Biotechnology" of RAS, Moscow, Russia
Vladimir Popov, The Federal Research Centre "Fundamentals of Biotechnology" of RAS, Moscow, Russia
Vladimir Poroikov, Institute of Biomedical Chemistry, Moscow, Russia
Nikolai Ravin, The Federal Research Centre "Fundamentals of Biotechnology" of RAS, Moscow, Russia
Igor Rogozin, National Institutes of Health, USA
Elena Salina, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Maria Samsonova, St. Petersburg State Polytechnic University, St. Petersburg, Russia
Anna Trubacheva, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia
Dmitry Ushakov, Institute of Psychology of RAS, Moscow, Russia
Andrey Vasiliev, Koltzov Institute of Developmental Biology of RAS, Moscow, Russia

Organizing committee

Svetlana Zubova, ICG SB RAS (Chairperson)
Olga Petrovskaya, ICG SB RAS (Secretary scientist)
Anton Bogomolov, ICG SB RAS
Tatyana Chalkova, ICG SB RAS
Larisa Eliseeva, ICG SB RAS
Tatyana Karamysheva, ICG SB RAS
Andrey Kharkevich, ICG SB RAS
Grigoriy Khazankin, Chair of the EMBS Chapter of the IEEE Russia Siberia Section, NSU
Tatyana Lakhova, ICG SB RAS
Pavel Linkevich, ICG SB RAS
Alexey Mukhin, ICG SB RAS
Zakhar Mustafin, ICG SB RAS
Artem Pronosin, ICG SB RAS
Dmitry Rasskazov, ICG SB RAS
Anna Smirnova, ICG SB RAS
Erlan Tokpanov, ICG SB RAS
Vladimir Zamyatin, ICG SB RAS

The conference is held with the support of the Kurchatov Genomics Center
of the Institute of Cytology and Genetics, SB RAS (Novosibirsk), project No. 075-15-2019-1662

The conference was supported by the Russian Foundation for Basic Research, project No. 20-04-22002

Partners

BGRS/SB-2020 conference is organized with the support of PMI Science (PMI)

Vavilov Journal of Genetics and Breeding

Конференции.ru

Frontiers in Genetics

MDPI IJMS

Sponsors

GOLD SPONSOR

HELICON COMPANY

SILVER SPONSORS

ROCHE_DIAGNOSTICS_RUS

CHIMMED COMPANY

BIOLINE

DIA-M

BIOCAD

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

Director: Corr. Member of the RAS *Alexey V. Kochetov*
Academic Advisor*:* Professor, Full Member of the RAS *Nikolay A. Kolchanov*
Academic Secretary: Candidate of Science (Biology) *Galina V. Orlova*
Phone: +7 (383) 363 4985 ext. 1336, e-mail: gorlova@bionet.nsc.ru

The Institute was founded in 1957, among the first institutions of the Siberian Branch of the Russian Academy of Sciences. It is situated in the Novosibirsk Akademgorodok. Presently, ICG SB RAS is an interdisciplinary biological center, which ranks among the leading biological institutions in Russia. The second step of the restructuring of the Federal Research Center Institute of Cytology and Genetics was completed in May 2017. Presently, ICG includes three affiliated branches: 1) Siberian Research Institute of Plant Production and Breeding (SibRIPPB). The institute is located in Krasnoobsk Village and the Novosibirsk rural area. It conducts academic, prospective, and applied studies including the collection, examination, preservation, and utilization of plant genetic resources for obtaining new biological knowledge; expansion and improvement of crop gene pools; 2) Research Institute of Clinical and Experimental Lymphology (RICEL); 3) Research Institute of Internal and Preventive Medicine (RIIPM). RICEL and RIIPM are situated in the Sovetskiy and Oktyabr'skiy districts of Novosibirsk. They conduct academic, prospective, and applied studies in molecular medicine and human genetics. They also provide medical care.

Tasks of ICG SB RAS: Solution of top-priority problems in the development of the Russian science and technology sector in plant genetics and breeding, animal genetics and breeding, human genetics, biotechnology, and fundamental medicine by applying methods of molecular genetics, cell biology, and computational biology.

Strategic goal: Integrated studies in plant genetics and breeding, animal genetics and breeding, human genetics, fundamental medicine, and biotechnology by applying methods of molecular genetics, cell biology, and computational biology from the generation of academic knowledge to the solution of top priority problems set by Russian agricultural, biotechnological, biomedical, and pharmaceutical industries.

Staff: As on June 1, 2020, ICG included 120 scientific units, which employed 1436 members; of them 548 researchers, 2 RAS Advisors, 7 Full Members of the RAS, 5 Corresponding Members of the RAS, 94 Doctors of Science, and 316 Candidates of Science.

Postgraduate education and residency training: As on June 1, 2020, ICG trains 60 postgraduates.

Publications: The Institute ranks among acknowledged leaders in Russian biology. Numerous works of its researchers are published in Russian and foreign academic journals. In 2019 the overall number of publications in peer-reviewed periodicals is 530. The overall number of ICG publications indexed in the Web of Science Core Collection is 484. In 2015–2019 papers of the researchers of the Institute received 12 362 citations accounted by Web of Science.

Auxiliaries: Core facility “Center for Genetic Resources of Laboratory Animals”, which includes the unique research unit “SPF vivarium”, and seven shared access centers (www.bionet.nsc.ru/uslugi/).

Institute of Cytology and Genetics is the founder of the following journals (mass media): *Vavilov Journal of Genetics and Breeding*, *Letters to Vavilov Journal of Genetics and Breeding*, *Atherosclerosis*, *Siberian Scientific Medical Journal*, and *Live Science*.

The Federal Research Center Institute of Cytology and Genetics is looking to cooperate with scientific and commercial enterprises.

Address: Prospekt Lavrentyeva 10, Novosibirsk, 630090 Russia; *phone*: +7(383) 363-49-80;
fax: +7(383) 333-12-78; *URL*: www.bionet.nsc.ru; *e-mail*: icg-adm@bionet.nsc.ru

PMI SCIENCE

PHILIP MORRIS INTERNATIONAL

PMI Science – научная платформа компании Philip Morris International (PMI)[1], предназначенная для публикации и распространения научной информации об усилиях компании по разработке и оценке продуктов, которые могут снизить индивидуальный риск и вред для населения, связанный с употреблением табака

Несколько лет назад компания **Philip Morris International (PMI)** вступила в эпоху трансформации и взяла на себя обязательства по созданию будущего без сигаретного дыма. Курение табака вызывает ряд серьезных заболеваний и увеличивает риск преждевременной смерти. Стратегии борьбы с табакокурением в большинстве стран сосредоточены на мерах регулирования спроса и предложения, их цель — не допустить появления новых курильщиков, снизить потребление табака и стимулировать отказ от курения. Эти меры привели к снижению распространенности курения за последние три десятилетия, но вряд ли они ликвидируют курение полностью.

Основной источник проблемы известен: десятилетия научных исследований демонстрируют, что главной причиной болезней, связанных с курением, являются вредные и потенциально вредные компоненты, образующиеся в результате горения табака. Именно горение вызывает появление большинства вредных химических соединений, обнаруживаемых в сигаретном дыме. Поэтому Компания сосредоточена на разработке ассортимента продуктов с пониженным риском для совершеннолетних курильщиков. Полный переход на такую продукцию может быть значительно менее вредным, чем продолжение курения традиционных сигарет.

[1] Аффилированной компанией Philip Morris International в России является ООО «Филип Моррис Сэйлз энд Маркетинг».

Мы начали разрабатывать продукцию с пониженным риском почти 10 лет назад. В области разработки бездымных продуктов у нас уже более 4600 зарегистрированных патентов и более 6300 на рассмотрении. Более 400 ученых, инженеров и технических специалистов мирового уровня разрабатывают и проводят клиническую оценку влияния на здоровье человека бездымных продуктов нового поколения.

Многоуровневый подход ФМИ к научным исследованиям:

- основан на практиках, уже много лет применяемых в фармацевтической отрасли, а также на методических рекомендациях Управления по контролю качества продуктов питания и лекарственных препаратов США (U.S. FDA), включая Проект методических рекомендаций FDA по табачной продукции с модифицированным риском;
- осуществляется с соблюдением таких международных стандартов, как Правила надлежащей лабораторной практики (GLP) и Правила надлежащей клинической практики (GCP).

Наша программа научной оценки основывается на сотрудничестве и экспертизе в области химии, токсикологии, биологии, информатики, медицины, восприятия и поведения и включает в себя пять ступеней научной оценки наших бездымных продуктов:

1. Разработка продукта. Подтверждение пониженного уровня образования вредных и потенциально вредных веществ.
2. Токсикологическая оценка. Подтверждение пониженного уровня токсичности и риска в лабораторных условиях.
3. Клиническая оценка. Подтверждение снижения вредного воздействия на организм и риска развития заболеваний у курильщика после переключения на альтернативные бездымные продукты пониженного риска.
4. Восприятие и поведение. Оценка восприятия и поведения в отношении бездымных продуктов совершеннолетними курильщиками, некурящими и бывшими курильщиками.
5. Долгосрочная оценка. Мониторинг безопасности и использования нашей продукции в реальных условиях.

Мы руководствуемся высокими научными стандартами и следуем принципам прозрачности и открытости в научно-исследовательской деятельности. Вся информация о методиках и результатах исследований публикуется на нашем научном портале PMIScience.com, который также доступен и на русском языке.

ООО «КОМПАНИЯ ХЕЛИКОН» — с 1997 года один из ведущих российских поставщиков оборудования, реагентики и расходных материалов для клеточной и молекулярной биологии, ориентированный на внедрение новейших разработок в российскую лабораторную практику.

Направления деятельности:

- молекулярная и клеточная биология,
- биоиндустрия,
- криминалистика,
- ветеринария,
- пищевая безопасность,
- клиническая диагностика.

Продуктовое портфолио включает более 70 мировых брендов, среди которых Bio-Rad, Beckman Coulter, QIAGEN, MGI, Eppendorf, Fluidigm, BMGLabtech, Merck, SSI и др.

Компания имеет R&D и производственную базу и выпускает лабораторную продукцию под собственной маркой «Хеликон»: оборудование, расходные материалы и комплектующие для вертикального и горизонтального электрофореза, системы гель-документирования, штативы для пробирок и дозаторов и др.

Особенностью работы с компанией является возможность бесплатного тестового использования некоторых видов продукции до принятия решения о ее покупке. Доставка и инсталляция в лаборатории Клиента осуществляется за счет Компании Хеликон.

Наличие развитой логистической и складской сети позволяет доставлять заказы в кратчайшие сроки. Региональные представительства компании находятся в Санкт-Петербурге, Новосибирске, Казани, Ростове-на-Дону и Екатеринбурге.

Контакты:

ООО «Компания Хеликон»

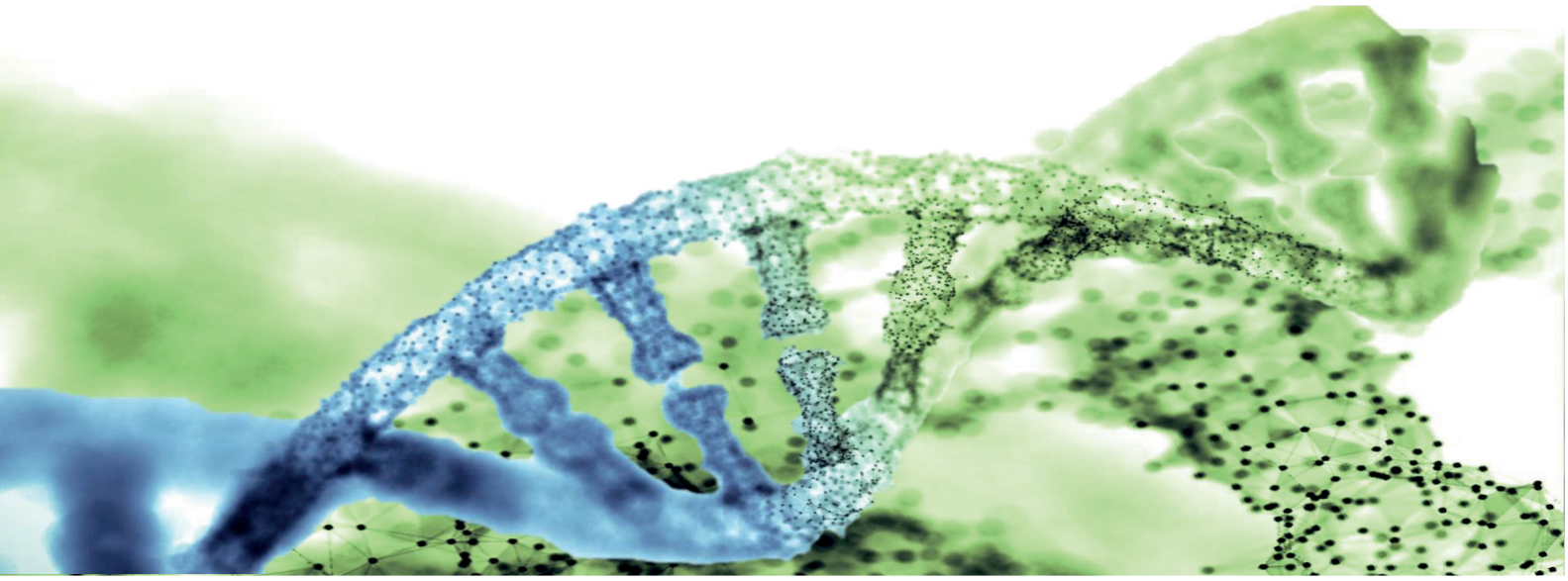
121374, Москва, Кутузовский проспект, д. 88

8 800 770 71 21 (бесплатно для всей России)

+7 499 705 50 50 (в Москве)

mail@helicon.ru

www.helicon.ru



ООО «Рош Диагностика Рус» – официальный импортер продукции Roche в России и лицензиат компании F.Hoffmann–La Roche Ltd.

Подразделение Roche Sequencing Solutions (RSS) разрабатывает и предлагает комплексные системы реагентов NGS пробоподготовки для широкого применения, как в научных, так и в клинических исследованиях.

Системы реагентов KAPA для приготовления библиотек в NGS

В портфолио RSS входит полный комплект реагентов KAPA для приготовления ДНК и РНК библиотек в NGS: ферменты и буферы для пробоподготовки; индексированные адаптеры, совместимые с известными NGS секвенаторами; магнитные частицы для очистки НК; а также наборы для количественной оценки готовых библиотек методом ПЦР в реальном времени.

Панели для таргетного секвенирования KAPA TE Hyper

Таргетные панели KAPA TE Hyper основаны на гибридизации ДНК зондов с готовыми NGS библиотеками. С помощью панелей KAPA TE Hyper Вы можете получить все интересующие регионы в одной пробирке и исключить все остальные участки генома. Рош предлагает Вам самостоятельно создавать свои собственные кастомные панели для обогащения геномов любых биологических организмов, а также воспользоваться готовыми панелями, как для клинических, так и для фундаментальных исследований.

moscow.reception1_dia@roche.com
https://sequencing.roche.com
Тел.:+7 (495) 229-69-99 Факс: +7 (495) 229-62-64 115114, Россия, Москва, ул. Летниковская, д. 2, стр. 2, Бизнес-центр «Вивальди Плаза

ThermoFisher
SCIENTIFIC

Современная высокотехнологичная генетическая лаборатория – это возможность использования различных методов для проведения исследований: ПЦР, секвенирование CE, NGS. Поможем подобрать оборудование, расходные материалы и реактивы для ваших исследований.

Генетические анализаторы – это высокое качество анализа и высокая степень точности, возможность прочтения длинных фрагментов.

- **SeqStudio**
- **3500**
- **3500 xl**
- **3730 NEW**
- **Ion GeneStudio S5**
- **Ion Torrent Genexus**

Цифровая ПЦР -
QuantStudio 3D

Амплификаторы RT – семейство приборов QuantStudio: от QS1 до QS 12K

ХИММЕД

115230, Москва, Каширское шоссе, дом 9, корп. 3
Тел.: +7 (495) 640 4192, (499) 682 6555
E-mail: biomol@chimmed.ru
www.chimmed.ru

BioTek's complex for long-term cellular studies

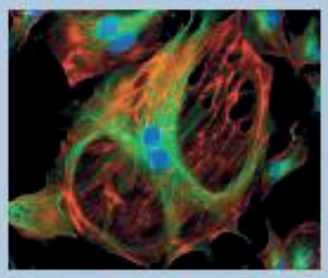
Dispenser or washer MultiFlo™, 405™, EL406

Automated Incubator BioSpa™ BioSpa™8

Multi-Mode Reader or Cell-Imager Epoch™, Synergy™, Cytation™

Compact and affordable solution for cellular research using an automated incubator:

- Cells Cultivation for up to 2 weeks without intervention of the operator
- 8 compartments dedicated for cell culture
- Parallel use of several independent protocols
- CO_2, O_2, temperature and humidity control
- Working with multi-well microplates, Petri dishes, culture flasks
- Continuous monitoring and recording of steps and conditions for the working experiment



Analysis of the results:

- Visualization of cells in bright field, phase contrast and fluorescence mode
- Multiple (EX and EM) spectrum analysis
- OD, fluorescence and luminescence measurements

Sample preparation:

- Replacement of the medium during cultivation
- Adding reagents
- Automated processing or staining procedures

It may include various combinations of washing / dispensing instruments and multi-mode readers or imagers for data logging.

BioLine LLC is the official distributor for BioTek Instruments (USA) in Russia

BioLine группа компаний

Bioline LLC
Russia, 197046,
St. Petersburg,
Pinsky alley, 3, lit. A

tel : +7 (812) 320 49 49
fax: +7 (812) 320 49 40
e-mail: main@bioline.ru
www.bioline.ru

Technical assistance number for all regions of Russia:
8 800 333 00 49

Фрагментный анализ ДНК, РНК и белков

Системы капиллярного электрофореза Qsep

ДИА•М
современная лаборатория

BiOptic

Скрининг ПЦР-продуктов

Генотипирование

Количественный и качественный анализ библиотек для NGS и нанопорового секвенирования

Анализ фрагментированности геномной ДНК

Анализ внеклеточной ДНК

Анализ олигонуклеотидов

Анализ продуктов рестриктазного расщепления

Анализ концентрации и целостности РНК (RNA Quality Number)

- Автоматическая подача образцов (от 1 до 96);
- несколько типов картриджей для разных молекул (фрагментированная ДНК, геномная ДНК, РНК, белки), рассчитанных на анализ 100–300 образцов;
- время анализа одного образца — от 2 до 7 минут;
- разрешение — 1–4 нуклеотида (в диапазоне длин от 100 до 500 нуклеотидов);
- чувствительность — от 100 пг/мкл;
- объём образца — 1 или 20 мкл, из которых в капилляр вводится 1 пл;
- визуализация, количественный и качественный фрагментный анализ с помощью ПО;
- низкая стоимость анализа — **от 15 рублей за образец**.

www.dia-m.ru

Symposium
Genomics, transcriptomics, bioinformatics

DOI 10.18699/BGRS/SB-2020-001

AD ASTRA: the database of Allelic Dosage-corrected Allele-Specific TRAnscription factor binding suggests causal regulatory sequence variants of pathologies

Sergey Abramov
Vavilov Institute of General Genetics
Russian Academy of Sciences
Moscow, Russia
Abramov.SA@phystech.edu

Alexandr Boytsov
Vavilov Institute of General Genetics
Russian Academy of Sciences
Moscow, Russia
boytsovs.av@phystech.edu

Bykova Dariia
Faculty of Bioengineering and Bioinformatics
Lomonosov Moscow State University
Moscow, Russia
db.bykova@gmail.com

Eugene Baulin
Institute of Mathematical Problems of Biology RAS - the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences
Pushchino, Russia
baulin@lpm.org.ru

Ivan Yevshin
BIOSOFT.RU LLC
Novosibirsk, Russia
ivan@developmentontheedge.com

Fedor Kolpakov
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
fkolpakov@gmail.com

Vsevolod J Makeev
Vavilov Institute of General Genetics
Russian Academy of Sciences
Moscow, Russia
vsevolod.makeev@vigg.ru

Ivan V Kulakovskiy
Engelhardt Institute of Molecular Biology
Moscow, Russia
ivan.kulakovskiy@gmail.com

***Abstract* — We present a comprehensive meta-analysis of more than 7 thousand ChIP-Seq experiments that allowed the identification of more than 100 thousand allele-specific binding sites across human transcription factors and cell types. Particularly, the allele-specific binding was found at many regulatory variants having known associations with various pathologies.**

Keywords* — *Transcription Factors, ChIP-Seq, Allele-Specific Binding, Allelic Dosage, Phenotypic Traits, Single-Nucleotide Variants, SNV, Single-Nucleotide Polymorphisms, SNP

Introduction

Sequence variants in genomic regulatory regions act as a source of many hereditary phenotypic traits including predisposition to various pathologies. Variants in promoters and enhancers may affect gene expression regulation at the transcription level by altering transcription factor (TF) binding. Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) with enough sequencing depth allows detecting events of allele-specific binding (ASB) i.e. differential TF affinity to alternative alleles of a particular single-nucleotide variant, existing in a heterozygous state in a given cell type. However, most of the published ChIP-Seq data is obtained in cell types with complex aneuploidy and multiple copy-number variants (CNVs), complicating statistical assessment of ASB events due to biased read counts from alternative alleles arising from non-equal background allelic dosage (BAD, the combination of chromosome ploidy and CNV effects).

Results

We present a new computational approach to ASB calling by accounting for BAD through de novo identification of chromosome duplications and CNVs directly from the read coverage at alternative alleles in aggregated data for each particular cell type. We successfully validated the algorithm using data on copy numbers in immortalized cell lines from the COSMIC database. We also introduced a new statistical approach to estimating ASB significance accounting for the read mapping bias by separately assessing read frequencies at the reference and alternative alleles with the negative binomial test.

We have applied the pipeline for more than 7 thousand ChIP-Seq read alignments available in the GTRD database, identified nearly 35 million candidate single-nucleotide variants with coverage of no less than 10, systematically assessed BAD in more than five hundred of cell types, and found more than 100K ASB events with coverage no less than 30 and 5% FDR.

We have annotated the ASBs by using existing public genotype-phenotype association databases, such as EBI GWAS Catalog, and ClinVar, and identified enriched association of ASB SNPs with different pathologies. Particularly, higher enrichment was detected for ASBs of multiple TFs, suggesting that rewiring of a regulatory network has greater potential to be deleterious compared to disruption or creation of individual regulatory sites.

Acknowledgment

This study was supported by the Russian Foundation for Basic Research grant 18-34-20024.

DOI 10.18699/BGRS/SB-2020-002

Bioinformatic screening for subtilisin-like peptidases in dikaryotic fungi

Nikita Alkin
MSU, Moscow, Russia
nikita9801@mail.ru

Yakov Dunaevsky
MSU Belozersky IPCB, Moscow, Russia
dun@belozersky.msu.ru

Mikhail Belozersky
MSU Belozersky IPCB, Moscow, Russia
mbeloz@belozersky.msu.ru

Galina Beliakova
MSU, Moscow, Russia
adm-odo@yandex.ru

Valeriia Tereshchenkova
MSU, Moscow, Russia
v.tereshchenkova@gmail.com

Elena Elpidina
MSU Belozersky IPCB, Moscow, Russia
elp2@yandex.ru

***Abstract* — Subtilisin-like peptidases (SLPs) play an important role in digestion and host colonization of pathogenic fungi. Analysis of 42 fungal genomes from all major taxa of Ascomycota and Basidiomycota was performed. Homologs of enzymes from all 4 major families of SLPs were found is the examined genomes. The most abundant types of the SLP homologs found were those of kexin and proteinase K families; pyrolysin homologs were detected in the half of the observed species; in addition, several homologs of OSP family proteins were found. Alignments of homologous amino acid sequences with type SLPs indicated presence of intact conserved domains and active sites. Number of SLP homologs differ distinctly between species from different environmental niches with the highest variety of SLPs found in pathogenic and parasitic fungi.**

Keywords — fungal peptidases, subtilisin family

Introduction

Subtilisin-like peptidases (SLPs) form one of the two largest superfamilies of serine peptidases along with chymotrypsins. Up to date SLPs were discovered and characterized in both prokaryotes and eukaryotes including all major lineages of the latter. Fungal SLPs are a subject of a high interest due to their distinguished role in an infection process of entomopathogens and in colonization process of endophytic fungi. Though there are works which trace evolutionary history of fungal SLPs within particular fungal taxa, the information about the variety and distribution of SLPs throughout Ascomycota and Basidiomycota remains rather incomplete. The objective of this research is a search for genes of SLPs in genomes of different species of dikaryotic fungi (Ascomycota and Basidiomycota) and subsequent analysis of predicted SLP sequences.

Methods

The search for sequenced and annotated fungal genomes was accomplished in NCBI public database (https://www.ncbi.nlm.nih.gov/). Amino acid sequences of endophytic ascomycete *Epichloë festucae* SLP homologs were obtained by Protein BLAST service (https://blast.ncbi.nlm.nih.gov/Blast.cgi). *E. festucae* was chosen as a sample species for homolog search due to its well-described 15 putative SLP gene system. The query used included SLPs belonging to four families: proteinase K-type (prtA-prtF, prtI, prtJ), kexins (kexA, kexB), oxidatively stable proteases or OSPs (prtL) and pyrolysins (prtG, prtH, prtK, prtM). Amino acid sequences of biochemically characterized SLPs (proteinase K from *Parengyodontium album*, kex2 from *Saccharomyces cerevisiae*, KP-43 from *Bacillus sp*. strain KSM-KP43 and pyrolysin from *Pyrococcus furiosus*) and information about their active sites was achieved in SWISS-PROT database (http://www.uniprot.org/). The homologs with E-value lower than 1*10-20 and cover higher than 75% were used in subsequent analysis. Isoelectric point, number of amino acid residues and molecular weight were calculated via Isoelectric point calculator website (http://isoelectric.org/calculate.php) for each homolog found. Multiple alignment of amino acid sequences and phylogenetic tree building were performed via Clustal Omega service (http://ebi.ac.uk/Tools/msa/clustalo/). Mutations in the active sites of inspected SLP homologs were identified in comparison to the amino acid sequences of type SLPs after the alignment. Potential transmembrane domains of the detected homologs were analyzed using TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) and potential signal peptides were predicted through SignalP-5.0 server (http://www.cbs.dtu.dk/services/SignalP). Batch CD-search service (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) was used to identify conserved domains (CDs) in analyzed homologs.

Results

We have analyzed 42 fungal species with sequenced genomes from all subphyla of Ascomycota and Basidiomycota (2 species of Taphrinomycotina, 4 species of Saccharomycotina, 27 species of Pezizomycotina, 1 species of Pucciniomycotina, 3 species of Ustilaginomycotina and 6 species of Agaricomycotina). The chosen species represent different life forms (mycelial, yeast and dimorphic) and ecological niches (biotrophic and necrotrophic plant pathogens, xylotrophic and saprotrophic free-living fungi, animal parasites and nematophagous fungi, alkaliphilic fungi, etc.).

Analyzed fungal genomes contained homologs of *E. festucae* SLPs from all four major families, some of the homologs were presented in several copies. Homologs of prtD belonging to proteinase K family, were found in genomes of all the species studied except *Hypocrella siamensis*; kexin homologs were found in all the species excluding *Taphrina deformans*. All examined species except *T. deformans* contained 1-2 kexin homologs in their genomes; on the contrary, the number of prtD homologs per genome varied from 0 (*H. siamensis*) up to 19 (enthomopathogenic *Metarhizium anisopliae*) with prevalence of rich gene clusters in genomes of phytopathogenic and enthomopathogenic fungi, e.g. *Magnaporthe grisea* (6 homologs) and *Moelleriella libera* (12 homologs) respectively. Homologs of prtL, a

member of OSP family, were found only in genomes of *M. anisopliae* (2 homologs), *Fusarium graminearum* and *Mycosphaerella graminicola* (1 homolog). Pyrolysin homologs were found in genomes of 22 out of 42 species, the most common number of homologs was 2 genes per genome, the richest pyrolysin gene cluster consisting of 16 homologs was found in genome of plant pathogen *Claviceps purpurea*.

The obtained sequences of SLP homologs have the same active site composition as type SLPs of corresponding families. The conservative amino acid triad of serine peptidases (serine, histidine, and aspartic acid) is present in each discovered kexin and prtL homolog and in the majority of prtD and prtG homologs, which is an indirect evidence that these peptidases retain enzyme activity. Search for CDs shows that every investigated kexin homolog contains the same set of CDs as the type yeast kexin, namely peptidase S8 catalytic domain and proprotein convertase domain; this may be another argument for functionality of the homologs. All investigated homologs of pyrolysin are distinguished by an insertion in peptidase S8 catalytic domain, homologs of proteinase K demonstrate N-terminal inhibitory domain prior to the catalytic domain. Some SLP homologs have long insertions close to C-terminus, for example, kexin from Saitoella complicata contains about 200 extra amino acid residues in comparison to yeast kexin.

Isoelectric point (pI) of the main part of detected SLPs ranges within pH 4,5-6. The vast majority of SLP homologs contain N-terminal signal peptide of 16-20 amino acid residues in length, which is consistent with secretory pathway origin of SLPs. Analysis of the fungal SLPs on the presence of potential transmembrane domains demonstrates that all obtained kexin homologs contain 1 such domain in the C-terminus third of a sequence. On the contrary, proteinase K, pyrolysin and OSP families have no transmembrane domains and are more likely secreted to the extracellular space, where these enzymes might serve for nutrition. It is worth to notice, that pyrolysins and OSPs predominantly cleave peptide bonds after arginine or lysine residues while SLPs from proteinase K family demonstrate broad substrate specificity and are restricted only by carboxyl group of proline residues. Thus, these SLPs might form universal cleavage tandem with proline-specific peptidases in order to hydrolyse glutamine-rich and proline-rich gluten proteins to non-inflammatory oligopeptides, which is crucial for gluten-free food industry.

Cladistic analysis of SLPs is particularly complicated due to multiple events of gene loss and gene cluster formation in some lineages. The obtained phylogenetic tree of kexin homologs resembles the data collected in the similar research from another set of fungal species [1]. There is a distinguished level of concordance between the both cladograms and modern consensus taxonomy, it is possible to outline well-resolved monophyletic clades at least at order scale; the main observed disaccordance with modern taxonomy on both trees is clusterization of Basidiomycota fungi with members of one of the basal subphyla of Ascomycota, with Saccharomycotina fungi in this research.

Conclusions

1. Analysis of 42 fungal genomes revealed homologs of enzymes from all 4 major families of SLPs.
2. Kexin and proteinase K homologs were present in the vast majority of studied species; pyrolysin homologs are found in the half of observed species; homologs of OSP family proteins are relatively scarce throughout species of higher fungi.
3. The obtained homologous amino acid sequences retain intact conserved domains and active sites.
4. Number of SLP homologs of a particular species is mostly determined by ecology, not by taxonomy; the reachest variety of SLPs is found in pathogenic and parasitic fungi.

Acknowledgment

The work was supported by the RFBR grants No. 19-04-00852 A.

References

[1] Li. J. et al. (2017) Phylogenomic evolutionary surveys of subtilase superfamily genes in fungi. Scientific Reports. 7: 45456.

DOI 10.18699/BGRS/SB-2020-003

Potential of whole genome sequencing in the assessment of sensitivity of clinical isolate *M. tuberculosis* to antibiotics

Olga Berdyugina
IIP RAS, Ekaterinburg, Russia
berolga73@rambler.ru

***Abstract* — To study the potential of whole genome sequencing in assessing the sensitivity of the clinical isolate of *M.tuberculosis* is an important research task. The data obtained can be used when choosing antibiotics in the treatment of a patient.**

Keywords* — *whole genome sequencing, M.tuberculosis, antibiotic

Motivation and aim

Motivation

Genomic research in phthisiology allows you to get closer to solving complex scientific and practical problems. The study of molecular genetic factors of multi drug-resistant (MDR) and extensively drug resistant (XDR) of *M.tuberculosis* is an important area of research.

Aim

The aim of this work was to study the potential of whole genome sequencing in assessing the sensitivity of the clinical isolate of *M.tuberculosis* to antibiotics.

Methods

Clinical and radiological evaluation of the patient was performed upon admission to a medical institution, as well as at the stages of treatment, in accordance with the requirements of providing medical care to patients with pulmonary tuberculosis. The nature of tissue reactions was established by pathomorphological studies. A study was carried out on the lung resect with staining according to Van Gieson and Weigert. The nucleotide sequence of the clinical isolate of *M.tuberculosis* was studied by the method of genome sequencing, which is based on the principle of research by synthesis (pyrosequencing) using the GS Junior instrument (Roche, Switzerland). In addition, the following were used: thermal cycler for amplification of nucleic acids C1000 Touch with CFX96 module, Flash Gel Dock electrophoresis system, handheld fluorimeter Quanti-Fluor-ST, DynaMag ™ magnetic tripod, GS Junior Seq Reagents and Enzymes reagents, GS Junior Sequencing Buffers, GS Junior Packing Beads & Supplement CB, GS Junior Control Bead Kit, GS Junior Maintenance Wash Kit, GS Junior PicoTiterPlate Kit. Work with the data obtained as a result of the study included the use of comparative analysis involving information from international databases, and the use of nonparametric methods.

Results

Sequencing of the clinical isolate of *M.tuberculosis* from a patient with fibro-cavernous pulmonary tuberculosis was performed. The diagnosis is confirmed by culture growth during bacteriological examination. Concomitant pathology was diabetes. When a patient was admitted to the hospital, drug resistance to rifadin, tubazide, ethambutol, streptomycin was revealed, which developed three years after the diagnosis and irregular intake of prescribed anti-TB drugs. The length of hospital stay was 147 days. Treatment included valvular bronchial blockage, pneumoperitoneum, a course of capreomycin, pyrazinamide, cycloserine, ofloxacin, and paraaminosalicylic acid. The result of treatment is improvement.

Sequencing of the clinical isolate of *M.tuberculosis* from this patient was carried out in two stages. This was due to the fact that the molecular genetic structure of *M.tuberculosis* is characterized by the presence of a large number of triple G-C bonds, which leads to difficulties in their cleavage during the reaction.

As a result of the work carried out at one of the intermediate stages, namely, at the stage of evaluating the result of emulsion PCR, after the saturation of particles for sequencing was carried out, it was found that the obtained number of particles is insufficient to identify the result and there is a high risk of ineffective research.

The start of the re-study with a change in the ratio of the starting reagents led to the proper analysis result.

The result of the work was the identified nucleotide sequence of the clinical isolate of *M.tuberculosis* belonging to an infected patient.

The genotype of the studied *M.tuberculosis* isolate consisted of 4.25 million bp. There were 1060 large contigs, 4006 medium contigs. The largest contig was 39 thousand bp in length. The average reading length of the fragments was 502 nucleotides. It was aligned 95,21% of all fragments obtained. The minimum coverage of the fragments according to the results of the work was to be at least 4, in our study it was 15, which confirms the high quality of the study, as well as the high reliability of the data.

To study the information obtained, we used international databases (ncbi, rast, blast) containing previously sequenced sequences of *M.tuberculosis*, including data on at least 15 groups of *M.tuberculosis*, containing genomes from 4.20 to 4.54 nucleotide base pairs, that encode from 3739 to 4258 genes.

A comparison of our own data with the results of international databases of *M.tuberculosis* strains made it possible to establish that the studied clinical isolate had genes, encoding penicillin-binding proteins. The isolate was sensitive to penicillin.

Thus, the results of whole genome sequencing in assessing the sensitivity of clinical isolates of *M.tuberculosis* to antibiotics can be used in the selection of patient therapy.

Acknowledgment

This work was carried out as part of the state assignment of the Institute of Immunology and Physiology, Ural Branch of the Russian Academy of Sciences (topic No. AAAA-A18-118020590108-7).

DOI 10.18699/BGRS/SB-2020-004

Analysis tandem repeats and retrotransposons of *Shepherdia argentea* (Pursh) Nutt.

Karina Bone
Laboratory of plant genetic engineering
All-Russia Research Institute of Agricultural Biotechnology
Russian State Agrarian University - Moscow Timiryazev Agricultural Academy
Moscow, Russia
karinabone@mail.ru

Olga Razumova
Laboratory of applied genomics and crop breeding
Kurchatov Genomic Center
All-Russia Research Institute of Agricultural Biotechnology
razumovao@gmail.com

Ilya Kirov
Laboratory of marker-assisted and genomic selection of plants
All-Russia Research Institute of Agricultural Biotechnology
Moscow, Russia
kirovez@gmail.com

Gennady Karlov
All-Russia Research Institute of Agricultural Biotechnology
Moscow, Russia
karlovg@gmail.com

***Abstract* — Buffalo berry (*Shepherdia arg.)* is a dioecious plant of the *Elaeagnaceae* family, being widely used in tradicional medicine. In our work, we carried out sequencing of two Shepherdia plants followed by repeatome annotation. Tandem repeats and mobile elements including both DNA transposons and retrotransposons have been identified, classified and their genome abundancy were predicted. Comparative analysis of the repeatomes of male and additional Shepherdia plant having unknown sex revealed differences in several repeat families.**

Keywords* — *shepherdia arentea, satellite DNA, mobile elements,

Introduction

Buffalo berry is a dioecious plant of the *Elaeagnaceae* family, it is a close relative of sea buckthorn, whose homeland is North America. As well as sea buckthorn, buffalo berry is widely used in tradicional medicine, as it is a source of useful elements - in the berries of buffalo berry there is a high content of vitamin C, carotene, catechins. Among the economically significant signs of buffalo berry, one can note such signs as the ability to enrich the soil with nitrogen due to the presence of nitrogen-fixing bacteria on the roots and resistance to drought. As well as her close relative, sea buckthorn, buffalo berry begins to bear fruit only by the age of 5, which, together with dioeciousness, complicates its selection. In this regard, it seems relevant to study the genetics of this species, to create the basis for marker selection.

Methods

In our work, we performed Illumina sequencing of two *shepherdia* plants. At the first stage, we checked the quality of the obtained sequences using the Fastqc program [1].Adapters were trimmed using the Trimmomatic program [2], and then the trimmed sequences were checked again in Fastqc. After that, using the RepeatExplorer program [3], we performed a graph-based clustering and tandem repeat analysis of the sequences of two plants of *Shepherdia arg*. — male and unknown sex. The total number of input read pairs was: 11,134,794 for male plant and 6,407,362 for the plant with unknown sex. In both plants, we analyzed 3 million randomly selected reads. The resulting consensus sequences of satellite repeats were compared between two plants using BLASTN program [4]. Primers were designed to detect for chromosome mapping using fluorescence in situ hybridization (FISH).

Phylogenetic tree

The reverse transcriptase domain of retrotransposon elements were extracted from the contigs assembled by RepeatExplorer using DANTE (Domain based Annotation of Transposable Elements, Neumann et al. (in press)). Sequences of retrotransposons Ty3 / Gypsy and Ty1 / Copia based on reverse transcriptase sequences were aligned by MUSCLE [5] and Neighbor-Joining tree was constructed for Ty3/Gypsy and Ty1/Copia.

Results

Using RepeatExplorer clustering of Illumina genomic reads for two Shepherdia plants tandem repeats and retrotransposons were found. A deeper bioinformatics analysis and annotation of DNA repeats of male and unknown sex shepherdia plants revealed 8 satellite repeats and 7 satellite repeats, respectively.. Among Ty1/Copia retrotransposons we found fifteen clusters corresponding to Ale subfamilies, twelve clusters corresponding to Angela/ Tork subfamilies, three clusters corresponding TAR subfamilies, seven clusters corresponding Ivana subfamilies and two clusters corresponding SIRE subfamilies. Among Ty3/Gypsy retronsposons we found eleven Athila subfamilies, three Tekay subfamilies, one Galadriel subfamily, four CRM subfamilies and sixty Reina subfamilies.

Acknowledgment

This work was supported by a grant from the Russian Foundation for Basic Research, agreement No. 20-016-00145 A.

References

[1] Fastqc Version 0.11.8: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[2] Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.
[3] Novak, P., Neumann, P., Pech, J., Steinhaisl, J., Macas, J. (2013) - RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. Bioinformatics 29:792-793.
[4] Zhang, J. & Madden, T.L. (1997) "PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation." Genome Res. 7:649-656. PubMed
[5] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput.Nucleic Acids Res. 32(5):1792-1797.

DOI 10.18699/BGRS/SB-2020-005

The first insights into regulation of cell transdifferentiation during gut regeneration in *Eupentacta fraudatrix*

Alexey Boyko
NSCMB FEB RAS, Vladivostok, Russia
Alteroldis@gmail.com

Igor Dolmatov
NSCMB FEB RAS, Vladivostok, Russia
idolmatov@mail.ru

Abstract — The holothurian Eupentacta fraudatrix is a unique organism for studying regeneration mechanisms. It has been repeatedly confirmed that regeneration is only due to the transformation of terminally differentiated cells. In this study, we examined changes in gene expression during gut regeneration of the holothurian E. fraudatrix and found 11 TFs, which are candidates for the role of transdifferentiation regulators

Keywords — echinoderms, transdifferentiation, gene ontology, gene expression, RNA-seq

Motivation and aim

Motivation

It is well-known that regeneration in *E. fraudatrix* is only due to the transformation of terminally differentiated cells. In addition, at the histological level, this process has been well studied previously. However, the molecular basis of cell transdifferentiation mechanisms in *E. fraudatrix* remains unknown.

Aim

The morphological features of regeneration of internal organs and cell transdifferentiation in the holothurian *E. fraudatrix* have been well studied [1]. In order to understand, to a first approximation, the possible molecular processes that occur during transdifferentiation in holothurians, we decided to use RNA-seq. In this work, we tried to obtain the complete transcriptome and determine the differential expressed genes (DEGs) at the stage of active cell transdifferentiation and transcription factors (TFs), which are candidates for the role of transdifferentiation regulator.

Methods

Tissue of gut was sampled from the intact gut and on the third (first stage of regeneration), fifth–seventh (second stage), and tenth (third stage) days post-evisceration. Each sample was sequenced and represented in two biological replicates with 5 individuals, pooled together, per replicate. Then the clean reads were corrected and assembled using SPAdes 3.13 software and our own Python script as described in the article [2]. Annotation was carried out by NR NCBI, Echinobase and Human Ensemble protein databases. Orthologs of human proteins were identified using a custom Python script that implements modified reciprocal method for finding the best hit. To identify the most likely candidates for the role of transdifferentiation regulators, only the TFs satisfying the three conditions were taken into consideration: the values of TPMs (Transcripts Per Kilobase Million) of the second stage higher than unity; the average value of LogFC (logarithm of fold change with base 2) of the second stage relative to the first and third stages more than half; the value of LogFC (logarithm of fold change) of the second stage relative to the first or third stages more than unity. The differential expression was evaluated for the sequences using Bowtie v2.3.4, RSEM v1.3.1 and DESeq2 v1.18 software. The enrichment analysis of biological processes and pathways was performed with GSEA software in accordance with the EnrichmentMap protocol for RNA-seq data [3]. Only processes or pathways containing at least one of the 11 most likely candidates for the role of transdifferentiation regulators were used. Then, results of the enrichment analysis were visualized using EnrichmentMap plug-ins in Cytoscape software

Results

Transcriptomes of intestinal anlage of the three stages of regeneration, as well as the normal gut, were sequenced with an Illumina sequencer (San Diego, CA, USA). We identified 14,617 sea urchin protein homologs, of which 308 were transcription factors. After analysing the dynamics of gene expression during regeneration and the map of biological processes in which they participate, we identified 11 factors: Ef-EGR1, Ef-ELF, Ef-GATA3, Ef-ID2, Ef-KLF1/2/4, Ef-MSC, Ef-PCGF2, Ef-PRDM9, Ef-SNAI2, Ef-TBX20, and Ef-TCF24. With the exception of TCF24, they are all involved in the regeneration, development, epithelial-mesenchymal transition, and immune response in other animals. We suggest that these transcription factors may also be involved in the transdifferentiation of coelomic epithelial cells into enterocytes in holothurians.

Acknowledgment

The results were obtained using the equipment of CKP «Primorsky aquarium», National Scientific Center of Marine Biology FEB RAS (Vladivostok, Russia). This study was supported by the Russian Foundation for Basic Research (grant No. 19-34-90015).

References

[1] Mashanov V.S. et al. (2005) Transdifferentiation in holothurian gut regeneration. Biol. Bull. 209: 184–193.

[2] Boyko A.V. et al. (2019) Reference assembly and gene expression analysis of Apostichopus japonicus larval development. Sci. Rep. 9: 1–11.

[3] Merico D. et al. (2010) Enrichment Map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 5: e13984.

DOI 10.18699/BGRS/SB-2020-006

Metavirome analysis of Baikal sponges

Tatyana Vladimirovna Butina
Laboratory of Analytical and Bioorganic Chemistry
LIN SB RAS
Irkutsk, Russia
tvbutina@mail.ru

Yurij Sergeevich Bukin
Laboratory of Genosystematics
LIN SB RAS
Irkutsk, Russia
bukinyura@mail.ru

Igor Veniaminovich Khanaev
Laboratory of Ichthyology
LIN SB RAS
Irkutsk, Russia
igkhan@lin.irk.ru

***Abstract* — Sponges are the oldest multicellular invertebrates (phylum Porifera); they are ecologically important components of marine and freshwater benthic environments. Associated communities of sponges include a variety of microorganisms: fungi, algae, archaea, bacteria and viruses. The aim of this study was to elucidate the genetic diversity of viruses in the community of Baikal endemic sponges *Baikalospongia bacillifera* using metagenomic approach. We have shown for the first time a high genetic, potential taxonomic and functional diversity of dsDNA viruses in these Baikal sponges. Identified viral sequences belonged to 28 viral families that infect a wide range of organisms. The bacteriophages of the *Myoviridae*, *Siphoviridae* and *Podoviridae* families dominated in the samples. The viruses of the *Phycodnaviridae*, *Poxviridae* *Mimiviridae*, *Herpesviridae*, *Baculoviridae*, *Polydnaviridae* and *Iridoviridae* families were also the most numerous. Viral communities of visually healthy and diseased Baikal sponges were significantly different. Analysis of viral sequences has indicated 22 functional categories of proteins and revealed a wide variety of structural viral proteins and enzymes. Among those the genes of proteins involved in the metabolism of host cells were also identified. Thus, the role of viruses in sponges may be both in the regulation of the number and diversity and in the maintenance of the vital activity of their hosts and the associated community as a whole.**

Keywords* — *viruses, viral diversity, metagenomics, virome, sponge holobiont, freshwater sponges, Lake Baikal

Introduction

Sponges are the oldest multicellular invertebrats (phylum Porifera) that represent complex symbioses (holobionts) in marine and freshwater ecosystems. These unusual animals are active biofilters and, due to their mass distribution, they play an important role in the ecosystem of Lake Baikal.

The community of sponges includes various microorganisms – fungi, algae, archaea, bacteria and viruses; they also serve as food and habitat for many invertebrates. The number, diversity and the role of viruses in the sponge holobionts may be very significant; however, viruses in the symbiotic associations of sponges remain poorly understood compared to other microorganisms. The first studies of the genetic diversity of sponge holobiont viruses have recently been published and concern some species of sponges inhabiting reefs [1, 2] and Lake Baikal (branched sponge *Lubomirskia baikalensis*) [3]. The results of these studies demonstrate the high diversity and important role of viruses in the composition of marine and freshwater sponge communities.

Against the backdrop of global climate change and increasing anthropogenic pressure in Lake Baikal, the mass destruction and death of sponges is registered, starting in 2011. The causes of the disease are still unknown.
The aim of this work was to study the genetic diversity of viruses in the associated community of the Baikal endemic sponges *Baikalospongia bacillifera* (diseased and visually healthy) using the metagenomic approach. Sponges *B. bacillifera* – one of the most abundant in Lake Baikal, they have a massive globular shape.

I. Materials, and Methods

The sponges *B. bacillifera* were sampled in the southern basin of Lake Baikal (near Bolshiye Koty, 51°54'07.5"N, 105°06'12.0"E) at depths of 16-20 m in May 2018 using lightweight diving equipment. Two individuals of *B. bacillifera* of 5-7 cm^3 in volume were collected: one looked healthy (Sv2478), and another had necrosis lesions (Sv2475). The sponge samples were twice washed in sterile Baikal water and thoroughly homogenized. Then homogenates were frozen in nitrogen and transported to the laboratory. The samples were gently thawed, twice diluted with SM buffer (0.2 M NaCl; 10 mM $MgSO_4$; 50 mM Tris HCl, pH 7.5), shaken with a Heidolph Multi Reax Vortex Mixer (10,000 rpm, 30 min), and were centrifuged 400 g for 15 min followed by 16,000 g for 30 min. The aqueous fraction was passed through a syringe filter with a pore size of 0.2 μm (Sartorius) and treated with DNase I and RNase A enzymes (Thermo Fisher Scientific) to remove contaminating nucleic acids. Viral DNA was extracted by ZR Viral DNA kit (Zymo Research).

The preparation and sequencing of DNA libraries were performed in The Center of Shared Scientific Equipment "Persistence of microorganisms" of Institute for Cellular and Intracellular Symbiosis UB RAS, Russia. The paired-end libraries were prepared using a NEBNext Ultra II FS DNA Library Prep Kit for Illumina (NEB) according to the manufacturer's protocol. The validation of DNA libraries was verified by Agilent 2100 Bioanalyzer (Agilent Technologies). Sequencing of the libraries was conducted on a MiSeq genome sequencer using MiSeq Reagent Kit v3 (2x300cycles, Illumina).

The primary processing (quality control and trimming) of the metavirome datasets (paired reads of 2x300 bp) was performed using the R package "ShortReads" [4]. The first (up to 15) and last (up to 30) nucleotides with low quality were removed. The sequences of less than 80 nucleotides were excluded from datasets.

Bioinformatic analysis of data was performed as described previously [3]. Briefly, taxonomic identification of viral sequences was performed using the BLASTn algorithm [5] against NCBI RefSeq viral complete genomes database (September 2018 release). The sequence reads were considered 'identified' if they had a relative in the reference database with an e-value of $\leq 10^{-5}$ and bit score ≥ 50. For the functional annotation of viral sequences, we used the local Blastx application [5] and COG database [6]. The reliability of the difference between two *B. bacillifera* viral communities

was estimated using the chi-square test. Statistical calculations were performed using the R packages “vegan” and “pvclust”.

Results and discussion

The raw data contained 3 842 088 and 5 035 528 pair sequence reads for the samples (Sv2475) and (Sv2478), respectively. After quality processing of data, we have obtained 3 375 063 and 4 063 311 reads, ranging from 80 to 256 bp. Of them, 97 557 and 88 517 sequences were identified as viral using the NCBI RefSeq viral genomes database, accounting for ca. 2.9 and 2.2% of datasets.

The families *Myoviridae*, *Phycodnaviridae*, *Siphoviridae*, *Poxviridae*, *Podoviridae*, *Mimiviridae*, *Herpesviridae*, *Baculoviridae,* and *Iridoviridae* were the most numerous, represented more than 1% of the sequences and in total accounted for more than 70% of the identified virome sequences (Fig. 1). We did not classify the significant parts of viral reads (21.4 and 23.9% in the samples Sv2475 and Sv2478, respectively) at the family rank. The diversity, richness and difference of two viral communities were estimated using Shannon, Simpson, ACE and Chao1 indices [7] (Table 1), rarefaction technique and chi-square test. The rarefaction curves for the both samples reached a plateau (data not shown).

TABLE 1 - BIODIVERSITY AND RICHNESS INDICES FOR THE VIROME DATASETS

Samples	Chao1/ACE	Shannon index	Simpson index
*B. bacillifera*_Sv2475	986/986	5.26	0.98
*B. bacillifera*_Sv2478	973/973	5.36	0.98

A comparative analysis of the diseased and visually healthy sponges did not reveal significant differences at the family level, the composition of the families was the same, but their percentages varied. In general, the viral communities of visually healthy and diseased Baikal sponges were significantly different (p-value < 2.2e-16). This indicates the changes in the structure of microbial communities in the affected sponges and the propagation of microorganisms and viruses that are not characteristic for the normal state of the sponges.

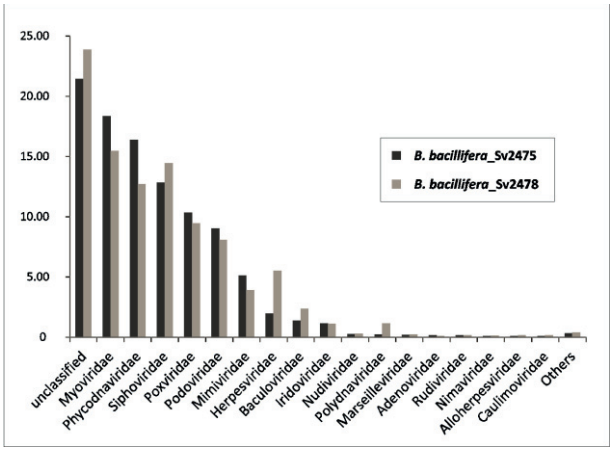


Fig. 1. The proportion of identified DNA viral families and viruses that were unclassified at the family rank

The comparison of revealed viral reads with the COG database has indicated the 22 functional categories of proteins and enzymes. The most representative (more than 5%) were proteins of replication, recombination and repair, nucleotide transport and metabolism, and mobile genomic elements (prophages and transposons).

Structural viral proteins, as well as the enzymes involved in the synthesis, modification, and packaging of DNA/RNA (methylases, reductases, terminates, helicases, polymerases, etc.) were the most numerous in *B. bacillifera*. The proteins of the host metabolism (auxiliary metabolic genes, AMGs) were also identified. The viruses, in particular bacteriophages, may synthesize the proteins necessary to maintain the vital activity of bacteria under unfavorable conditions. This, in turn, contributes to greater production of viral particles. Thus, viruses can play an important role in sponges by regulating the number and variety of micro- and macroorganisms, as well as participating in the metabolism and maintenance of the vital activity of their hosts.

References

[1] P.W. Laffy, E.M. Wood-Charlson, D. Turaev, K.D. Weynberg, E.S. Botté et al., “HoloVir: A workflow for investigating the diversity and function of viruses in invertebrate holobionts”, Front. Microbiol., vol. 7, 822, 2016.

[2] P.W. Laffy, E.M. Wood-Charlson, D. Turaev, S. Jutz, C. Pascelli et al., “Reef invertebrate viromics: diversity, host specificity and functional capacity”, Environ. Microbiol., vol. 20, pp 2125-2141, 2018.

[3] T.V Butina, Yu.S. Bukin, I.V. Khanaev et al., “Metagenomic analysis of viral communities in diseased Baikal sponge *Lubomirskia baikalensis*”, Limnology and Freshwater Biology, vol. 1. pp. 155-162, 2019.

[4] M. Morgan, S. Anders, M. Lawrence, P. Aboyoun, H. Pages, R. Gentleman, “ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data”, Bioinformatics, vol. 25, pp. 2607–2608, 2009.

[5] S.F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, “Basic local alignment search tool”, J. Mol. Biol., vol. 215, pp. 403–410, 1990.

[6] R.L. Tatusov, M.Y. Galperin, D.A. Natale, E.V. Koonin, “The COG database: a tool for genome-scale analysis of protein functions and evolution”, Nucleic Acids Res., vol.28, pp. 33-36, 2000.

[7] M.O. Hill, “Diversity and evenness: a unifying notation and its consequences”, Ecology, vol. 54, pp. 427–432, 1973.

DOI 10.18699/BGRS/SB-2020-008

Transcriptome-based analysis of lectins and carotenoid metabolism enzymes in amphipods of the Lake Baikal region

Polina Drozdova
Institute of Biology
Irkutsk State University;
Baikal Research Centre
Irkutsk, Russia
0000-0003-3955-6105

Alexandra Dolgikh
Dept. of Genetics and Biotechnology
St. Petersburg State University;
Bioinformatics Institute
St. Petersburg, Russia
0000-0003-1845-9701

Anna Nazarova
Institute of Biology
Irkutsk State University
Irkutsk, Russia
0000-0002-5388-2559

Yulia Lubyaga
Institute of Biology
Irkutsk State University;
Baikal Research Centre
Irkutsk, Russia
0000-0002-2494-8723

Maxim Timofeyev
Institute of Biology
Irkutsk State University;
Baikal Research Centre
Irkutsk, Russia
0000-0002-5250-6818

Abstract — **In this work, we explored two groups of proteins, the carbohydrate-binding lectins and enzymes of carotenoid metabolism, in amphipods of the Lake Baikal region based on the available transcriptome data. We found that different species possess quite similar repertoires of these proteins. The obtained sequence data are instrumental for further molecular-level exploration of amphipod physiology.**

Keywords — Carotenoid metabolism, immunity, lectins, Crustacea, Amphipoda, Baikal

Introduction

The ancient Lake Baikal, located in Eastern Siberia, hosts a spectacular diversity of largely endemic fauna [1]. Among the most diverse groups of its inhabitants are amphipods (Crustacea, Malacostraca, Amphipoda), comprising over 350 species and subspecies in the lake. At the same time, shallow gulfs of the lake and smaller water bodies in the region are inhabited by just one species, which is a potential invader but so far not found in the open part of Lake Baikal [2]. A number of amphipod species of the Baikal region have been used in ecophysiology- and ecotoxicology-, and biomonitoring-related studies [3-6, *etc.*]. An important contributor to the changes in the physiological state of an organism is its immune system. Recently, next-generation sequencing data for Baikal amphipods became available [7, 8], and these data can help in the exploration of the immune system of these emerging model species. In this work, we focused on two factors important for the immune system of crustaceans, lectins [9] and carotenoid metabolism [10].

Materials and methods

To check the presence of transcripts encoding putative lectins and carotenoid metabolism enzymes, we searched the available transcriptome assemblies of amphipods of the Baikal region [7, 8] with known sequences of these enzymes found in other species [9, 11, *etc.*] using the blastx and blastp algorithms of the ncbi-blast [12] and diamond [13] packages. Protein sequences were predicted with TransDecoder [14]. Estimation of the expression level was performed with Salmon [15] using a wrapper script provided by Trinity [14].

Results and Discussion

Lectins

Lectins are a very ancient and versatile group of proteins containing carbohydrate-binding domains. Members of different lectin subfamilies bind to various carbohydrates (*e.g.*, mannosides, galactosides, sialic acids), even though the ranges of substrates are overlapping, have characteristic folds and act both inside the cells and in the extracellular matrix [16]. The results of the search for lectins were similar for two endemic amphipod species of Lake Baikal (*Eulimnogammarus cyaneus* and *E. verrucosus*) and a potential invader species, *Gammarus lacustris* (Table 1). Further exploration of the expression levels of these transcripts revealed a relatively high organism-level expression (>10 transcripts per million) only in type C lectins in all three species.

This result may mean that some of the type C lectins are expressed constitutively and/or in multiple tissue types, while the other types are expressed only in particular types of cells or only in response to infection.

TABLE I. PRESENCE OF DIFFERENT FAMILIES OF LECTINS AND ENZYMES OF CAROTENOID METABOLISM IN AMPHIPOD TRANSCRIPTOMES

Functional group	Protein (sub)family	Number of identified transcripts[a]
Lectins	C	136 / 111 / 159
	S (galectin)	14 / 9 / 13
	M	22 / 12 / 14
	L	10 / 8 / 9
	P	8 / 2 / 5
Carotenoid metabolism enzymes	CrtW (ketolase)	5 / 6 / 4
	CrtZ (hydroxylase)	48 / 35 / 21
	CrtS	117 / 84 / 138

[a] Shown are the counts of transcripts encoding proteins similar to the reference sequences of the corresponding family in *E. cyaneus*, *E. verrucosus* and *G. lacustris*, respectively (e-value < 10^{-5}).

Carotenoid metabolism enzymes

It is well-known that the absolute majority of animals do not synthesize carotenoids but need to rely on their presence in the food source, and crustaceans are no exception [17]. However, they are capable of bioconversion between carotenoids, mainly by oxidizing orange beta carotene present in their diets into dark-red astaxanthin via a cascade of four reactions. These reactions require two enzyme activities, a 3,3'-hydroxylase (CrtZ) and a ketolase (CrtW). These enzymes have been explored in marine bacteria [18]; moreover, some fungi seem to possess a bifunctional enzyme (CrtS) [19]. Thus, we looked for the known enzymes of these

three groups from bacteria, fungi, and crustaceans and found that putative CrtZ and CrtS proteins were found in all transcriptomes, while CrtW proteins are much less widespread and diverse. It is important to note that the analysis included transcriptome assemblies derived from whole animals, *i.e.*, including all symbiotic microorganisms, which can serve as the source of the identified enzymes (examples in Table 1).

Interestingly, no obvious difference between brightly colored deep red species and white-colored ones was found with this analysis. There are obviously other factors that contribute to the color formation, such as the level and composition of carotenoids, which can vary extensively between Baikal amphipod species [20], and carotenoid-biding proteins, which can form complexes of various colors [17].

Conclusion

The analysis of amphipod transcriptomes helped us discover many putative lectin and carotenoid metabolism enzyme sequences, which can be used in further molecular-level exploration in studies of amphipods of the Baikal region.

Acknowledgment

The study of carotenoid metabolism enzymes was supported by the Russian Science Foundation, project number 19-74-0045. The study of lectins was funded by the RFBR and the Government of the Irkutsk Region, project number 19-42-383007.

References

[1] M. E. Cristescu, S. J. Adamowicz, J. J. Vaillant, and D. G. Haffner, "Ancient lakes revisited: from the ecology to the genetics of speciation," Molecular Ecology, vol. 19, no. 22, 2010, pp. 4837-4851.

[2] V. V. Takhteev, "On the current state of taxonomy of the Baikal Lake amphipods (Crustacea: Amphipoda) and the typological ways of constructing their system," Arthropoda Selecta, vol. 28, no. 3, 2018, pp. 374-402.

[3] M. A. Timofeyev, C. Wiegand, B. K. Burnison, Z. M. Shatilina, S. Pflugmacher, C. E. W. Steinberg, "Impact of natural organic matter (NOM) on freshwater amphipods," Science of the Total Environment, vol. 319, no. 1-3, 2004, pp. 115-121.

[4] D. Axenov-Gribanov, D. Bedulina, Z. Shatilina, L. Jakob, K. Vereshchagina, Y. Lubyaga *et al*., "Thermal preference ranges correlate with stable signals of universal stress markers in Lake Baikal endemic and Holarctic amphipods," PloS ONE, vol. 11, no. 10, 2016, 0164226.

[5] L. Jakob, D. S. Bedulina, D. V. Axenov-Gribanov, M. Ginzburg, Z. M. Shatilina, Y. A. Lubyaga *et al*., "Uptake kinetics and subcellular compartmentalization explain lethal but not sublethal effects of cadmium in two closely related amphipod species," Environmental science & technology, vol. 51, no. 12, 2017, pp. 7208-7218.

[6] A. Gurkov, E. Shchapova, D. Bedulina, B. Baduev, E. Borvinskaya, I. Meglinski, M. Timofeyev, "Remote in vivo stress assessment of aquatic animals with microencapsulated biomarkers for environmental monitoring," Scientific reports, vol. 6, no. 1, 2016, 36427.

[7] S. A. Naumenko, M. D. Logacheva, N. V. Popova, A. V. Klepikova, A. A. Penin, G. A. Bazykin *et al*., "Transcriptome-based phylogeny of endemic Lake Baikal amphipod species flock: fast speciation accompanied by frequent episodes of positive selection," Molecular ecology, vol. 26, no. 2, 2017, pp. 536-553.

[8] P. Drozdova, L. Rivarola-Duarte, D. Bedulina, D. Axenov-Gribanov, S. Schreiber, A. Gurkov *et al.,* "Comparison between transcriptomic responses to short-term stress exposures of a common Holarctic and endemic Lake Baikal amphipods," BMC genomics, vol. 20, no. 1, 2019, 712.

[9] A. G. Lai, A. A. Aboobaker, "Comparative genomic analysis of innate immunity reveals novel and conserved components in crustacean food crop species," BMC genomics, vol. 18, no. 1, 2017, 389.

[10] K. Tan, H. Zhang, L.-S. Lim, H. Ma, S. Li, H. Zheng, "Roles of carotenoids in invertebrate immunology," Frontiers in Immunology, vol. 10, 2020, 3041.

[11] N. Mojib, M. Amad, M. Thimma, N. Aldanondo, M. Kumaran, X. Irigoien, "Carotenoid metabolic profiling and transcriptome-genome mining reveal functional equivalence among blue-pigmented copepods and appendicularia," Molecular ecology, vol. 23, no. 11, 2014, pp. 2740-2756.

[12] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden. "BLAST+: architecture and applications," BMC bioinformatics, vol. 10, no. 1, 2009, 421.

[13] B. Buchfink, C. Xie, and D. H. Huson, "Fast and sensitive protein alignment using DIAMOND," Nature methods, vol. 12, no. 1, 2015, pp. 59-60.

[14] B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, *et al.,* "De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis," Nature protocols, vol. 8, no. 8, 2013, pp. 1494-1512.

[15] R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford, "Salmon provides fast and bias-aware quantification of transcript expression," Nature methods, vol. 14, no. 4, 2017, pp. 417-419.

[16] G. S. Gupta, "Lectins: An Overview", in Animal lectins: form, function and clinical applications, Springer Science & Business Media: Vienna, 2012, pp. 3-25.

[17] T. Maoka. "Carotenoids as natural functional pigments," Journal of natural medicines, vol. 74, 2019, pp. 1-16.

[18] D. P. L. Toews, N. R. Hofmeister, and S. A. Taylor, "The evolution and genetics of carotenoid processing in animals," Trends in Genetics, vol. 33, no. 3, 2017, pp. 171-182.

[19] J. F. Martín, E. Gudiña, and J. L. Barredo, "Conversion of β-carotene into astaxanthin: Two separate enzymes or a bifunctional hydroxylase-ketolase protein?," Microbial Cell Factories, vol. 7, no. 1, 2008, p. 3.

[20] B. Czeczuga, "Carotenoids in thirteen species of gammaridae from Lake Bajkał," Comparative Biochemistry and Physiology Part B: Comparative Biochemistry, vol. 50, no. 2, 1975, pp. 259-268.

DOI 10.18699/BGRS/SB-2020-010

Statistical relations between N-glycome of circulating immunoglobuline G and total plasma N-Glycome

Sofya G. Feoktistova
Laboratory of Glycogenomics
Institute of Cytology and Genetics
Novosibirsk, Russia
sayfutdinovas@gmail.com

Sodbo Sharapov
Laboratory of Glycogenomics
Institute of Cytology and Genetics
Novosibirsk, Russia
ORCID: 0000-0003-0279-4900

Yakov A. Tsepilov
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia
ORCID: 0000-0002-4931-6052

Tim Spector
Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences
King's College London
London, United Kingdom
ORCID:

Gordan Lauc
Genos Glycoscience Research Laboratory
Zagreb, Croatia
ORCID: 0000-0003-1840-9560

Frano Vuckovic
Genos Glycoscience Research Laboratory
Zagreb, Croatia
ORCID: 0000-0002-4845-0882

Yurii S. Aulchenko
Laboratory of Glycogenomics
Institute of Cytology and Genetics
Novosibirsk, Russia
ORCID: 0000-0002-7899-1575

***Abstract* — Glycosylation is the most common co-translational and post-translational modification of proteins. Glycans influence the physical properties of proteins as well as their biological functions. Alteration in glycosylation is observed in many human diseases.Defining genetic factors, altering glycosylation, can provide a basis for novel approaches to diagnostic and pharmaceutical applications. Predictionof IgG N-glycome from total plasma N-Glycome (TPNG) will allow to increase the power of genetic analysisof tissue specific glycosylation processes.**

Keywords* — *N-glycome, glycomics, IgG, prediction models

Introduction

The majority of human proteins are post-translationally modified by covalent addition of one or more complex oligosaccharides (glycans). Alterations in glycosylation processing are associated with numerous diseases, and glycans are attracting increasing attention both as disease biomarkers and as targets for novel therapeutic approaches[1]. Immunoglobulin G(IgG) is most abundant glycoprotein in human blood plasma: while total concentration of glycoproteins in human plasma is about 30 mg/mL, the concentration of IgG is about12 mg/mL [1]. The IgG is secreted to bloodstream by plasma cells. IgG is one of the most studied glycoproteins in terms of structural and functional aspects of glycosylation. The N-glycosylation patterns of IgG have been shown to be altered under various physiological and pathological conditions [3,4].

Ultra PerformanceLiquid Chromatography (UPLC) is the most broadly used method for measuring N-Glycome of both total plasma proteins and IgG. Recent advances in discovery of the genetic control of proteins N-Glycomeusing genome-wide association studies (GWAS) showed substantial overlap between genetic control of N-glycosylation of IgG and total plasma glycoproteome [5-7], which, perhaps, is not so surprising given IgG makes about 1/3 of total plasma glycoproteome.

These observationscould motivate development ofmethods for prediction ofN-Glycome profile of IgG from the TPNG, as well as prediction of TPNG ex. contribution from IgG.The latter could increase the power of discovery of genetic loci, that influence glycosylation of non-IgG proteins, thus increasing the chances of finding regulators, working in tissues and cell types other than plasma cells.

Here, we aimed to characterize multivariate statistical relationships between N-glycome of circulating IgG and TPNG.

Materials and methods

To studyempirical relationships between TPNG data and IgG N-glycome we used TwinsUKcohort, where both TPNG and IgG N-Glycome profiles were measured using UPLC technology about 6 000 samples.TheIgG and TPNG N-glycomeprofiles contain 24 and 36chromatographic peaks, correspondently. Approximately 1800 samples were set off as a test data set.

We applied different linear multiple regression models to predict each IgG glyco-peak (IGP) from plasma glyco-peaks (PGPs), so we had 24 outcomes and 36 predictors:

$$IGP_i \sim PGP1 + \ldots + PGP36 \quad (1)$$

We used regression models without penalization, as well as with the L1 (ridge) and L2 (LASSO) penalization.

We have also applied canonical correlation analysis.

In total, we trained 24 * 4 = 96 models.For penalized models, estimation of meta-parameters was performed using cross validation.

The test data set was used to evaluate the prediction quality.The proportion of variance explained was used as a metric of prediction quality for specific glycans, and root mean square error was used as a metric for overall prediction accuracy.

Results

We proposed set of approaches that allows prediction of IgG N-glycome from TPNG, and evaluate. All approaches have shown moderate prediction power.

ACKNOWLEDGMENT

The work of S.Sh., S.F., and Y.A., who coordinated and supervised this study, performed data analysis, interpretation of the results and writing of the initial text,was funded by the Russian Science Foundation grant number 19-15-00115. and supported by the Russian Ministry of Education and Science under the 5-100 Excellence Programme.

The generation of the data used in this study wassupported by the European Structural and Investment Funds IRI grant (#KK.01.2.1.01.0003) and Croatian National Centre of Research Excellence in Personalized Healthcare grant (#KK.01.1.1.01.0010) to GL. TwinsUK(TS) is funded by the Wellcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

REFERENCES

[1] Clerc, Florent, et al. "Human plasma protein N glycosylation." *Glycoconjugate journal* vol. 33, no.5, pp. 309-343, 01-June-2016

[2] K. Ohtsubo and J. D. Marth, "Glycosylation in Cellular Mechanisms of

[3] Health and Disease," *Cell*, vol. 126, no. 5. pp. 855–867, 08-Sep-2006

[4] I. Gudelj *et al.*, "Low galactosylation of IgG associates with higher risk for future diagnosis of rheumatoid arthritis during 10 years of follow-up," *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1864, no. 6, pp. 2034–2039, 2018.

[5] W. Peng *et al.*, "Clinical application of quantitative glycomics," *Expert Rev. Proteomics*, vol. 15, no. 12, pp. 1007–1031, Dec. 2018.

[6] J. E. Huffman *et al.*, "Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults," *Hum. Mol. Genet.*, vol. 20, no. 24, pp. 5000–5011, Dec. 2011.

[7] S. Z. Sharapov *et al.*, "Defining the genetic control of human blood plasma N-glycome using genome-wide association study," *Hum. Mol. Genet.*, vol. 28, no. 12, pp. 2062–2077, Mar. 2019.

[8] L. Klarić *et al.*, "Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases," *Sci. Adv.*, vol. 6, no. 8, p. eaax0301, Feb. 2020.

DOI 10.18699/BGRS/SB-2020-011

Competition and collaboration in the miRNA science field

Artemiy Firsov
Computer Science and Computer Engineering
Institute of Informarics Systems
Novosibirsk, Russia
0000-0002-7681-1032

Igor Titov
Laboratory of Molecular Genetic Systems
Institute of Cytology and Genetics
Novosibirsk, Russia
titov@bionet.nsc.ru

Abstract — **In this work we present the analysis of the characteristics of institutions interactions in the miRNA science field using the data from PubMed digital library. We identified the leaders of the field - China, USA -, characterized the interactions and described the country level features of co-authorship. We observed that the USA were leading in the publication activity until China took the lead 4 years ago. However, the USA are the main co-authorship driver in this field. We have also identified the pioneers and show, that they are the leaders of co-authorship activity. We compare the publications activity patterns on the organization level, identifying leaders. We compare the organization interaction graph with the authors interaction graph, and provide additional insights of miRNA science field evolution.**

Keywords — k-mer, n-gram, dbscan, identification, mirna, timsort, kofer, digital library, co-authorship

Introduction

Many digital libraries appeared with the growth of the Internet, thus, we got an opportunity to query articles metadata, gather some statistics, etc. This includes understanding the authors/institutions activity, their interactions, and other characteristics. One can also prove that the Paretto rule for the institutions' publication activity holds true [1], or that the idea spreads from one author to another like the virus spreads from one person to the others. Having this information, we can further use it to predict the new science field creation, popularity of the particular science field. In general, it can be used in social informatics.

Now the new science field is progressing – "science of science" [2]. It is a trans-disciplinary field of science that aims to understand the evolution of ideas, choice of a research problem of particular scientist, etc. Without analyzing interactions between authors, institutions, and other, such field just cannot exist.

However, to do that one should know to which real author/institution the authors name/affiliation from the paper corresponds to. This disambiguation issue gets harder considering big datasets, such as PubMed with $3 * 10^7$ articles. It becomes more complicated when you consider the presence of non-affiliation data in the affiliation field, errors in the author name/affiliation or the mixed institution names for different authors. E.g. if the *Author1* has "Institute of Cytology and Genetics, Novosibirsk, Russia" institution and the *Author2* has "Institute of Mathematics, Novosibirsk, Russia" institution, their resulting affiliation for collaborative paper might be "Institute of Cytology and Genetics, Institute of Mathematics, Novosibirsk, Russia".

In this paper, we aim to provide the solution for Affiliation Disambiguation problem and provide the analysis of the miRNA science field using the solution.

Methods and Materials

Prerequisites

The basic idea of the work is to get the groups of organizations mentions, which contain only mentions of one institution. After that we may use that information to build the co-authorship graph of organizations/countries, get static and dynamic characteristics of the science field, etc.

So on the first step we solved the clustering problem of institutions names writings. In our case, the closer the grouping is to the ground-truth grouping (i.e. one group contains all and only affiliations that refer to the same institution), the better the clustering is. For that sake, we have gathered the labeled the Novosibirsk miRNA dataset mentioned below.

Dataset

To conduct experiments, we have gathered two datasets from PubMed digital library using MEDLINE file format. First one is the Novosibirsk dataset, that consists of the preprocessed affiliations of the Novosibirsk institutions. We labeled this dataset to have the ground-truth affiliation clustering to further use it for clustering algorithm hyperparameters fine-tuning. You can find the additional information on the Novosibirsk dataset, and how we used it to fine-tune our algorithm, in our previous paper dedicated to the inter-country statistics of the miRNA science field [3].

Second one is the miRNA dataset gathered with the *miRNA* search query and its forms (*mi-RNA, miRNAS*, etc) on the PubMed website over *Title* and *Abstract* fields. The data we use was gathered on 18th of November, 2019.

Clustering and similarity

After the pre-processing stage, where we remove non-affiliation data and split the mixed affiliations, the clustering stage is performed. We tried different techniques for the clustering – K-Means and DBSCAN [4] – and different popular similarity functions applied to the preprocessed affiliations strings – Levenshtein, Jaccard, Smith-Waterman. We also tried the KOFER algorithm [3] to find similarity between affiliations.

We found out that KOFER algorithm along with Dice [5] distance metric gives better results on the labeled Novosibirsk dataset; thus, we used that approach on the full miRNA dataset.

Results

We present the results for country level co-authorship in the miRNA field, as well as the organization level statistics, i.e. pioneer analysis, leaders, etc. Using the KOFER algorithm we were able to cluster the miRNA affiliations data. From the 530,982 affiliations we got 34,081 clusters. i.e. institutions.

PubMed statistics

We got the cumulative per year number of publications on the PubMed digital library. We found out that the total number of publications added yearly to PubMed grows exponentially:

$$n(t) = e^{0.0492*x + 13.1902} \quad (1),$$

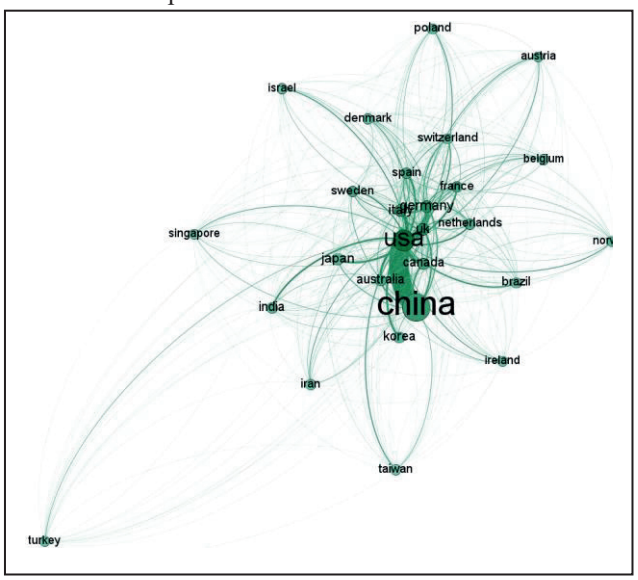
where t – 0, 1, 2 … and 0 corresponds to the starting year of publication activity within the miRNA field, i.e. 2001

miRNA statisics

The growth of the publications per year within the miRNA field, however, is reaching its peak, and can further be estimated using the models with saturation, such as log-normal and log-logistic functions.

Countries interactions

146 countries have publications in the miRNA science field. To see how countries interact, we have built the graph, leaving only the major countries in this filed, i.e. having more than 500 publications.



From this graph, it is clear that the USA and China are the leaders in this field, actively publishing with each other. This graph is also almost full, however, some of the countries have never published together, such as Iran-Israel, Iran-Brazil, Turkey-Belgium.

Additionally, analyzing the plots of publications per year, we noticed that the USA had the rapid start in this field, reaching 100 publications in 2004. However, the growth started to reduce over time, whereas China had higher growth, which led to China becoming the new leader in 2013. The USA on the other side remained the major driver of the international publications, and has more than others publications with China, Italy, UK, Canada, Germany, Korea, Japan.

Organization interactions

During the 2001-2019 period, there were 88,715 publications, 12,857 of which were international. The field until some time did not have much joint publications, however things changed in 2013. That year USA and China started actively publish together.

The interesting thing that the pioneers (the organizations, which published during the period 2001 – 2003) within this field have higher international publications rate than the overall rate – 6128 publications were published with pioneers, 2640 of which were international. Moreover, first four most active international publishers are the pioneers - University of California, Harvard Medical School, University of Texas MD Anderson Cancer Centre, Chinese Academy of Sciences.

We observer that out of 34,081 institutions, there are only 1,655 organizations, for which the number of their mentions is higher than 50 (1% of maximum mentions of the leader organization - Central South University, China). All 46 pioneers are in this main group.

Conclusion

In this work we have managed to cluster the miRNA science field affiliations data using the KOFER algorithm.

Using the clustering results, we were able to get properties of country level interactions, see that China is currently the leading country in this field, however the USA is the biggest driver of international publications.

The pioneers, such as University of California, Harvard Medical School, University of Texas MD Anderson Cancer Centre, Chinese Academy of Sciences, play big role within this field, leading the international activities. The current organization leader of the miRNA science field is Central South University, China.

REFERENCES

[1] I. I. Titov, A. A. Blinov, “Research of the structure and evolution of networks of scientific cooperation on the basis of analysis of Novosibirsk publications on the field of biology and medicine,” Vavilov Journal of Genetics and Breeding, vol. 18, 2014.

[2] S. Fortunato, C. Bergstrom, K. Borner, J. Evans, D. Helbing, S. Milojevic, A. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespignani, L. Waltman, D. Wang, A.-L. Barabasi, “Science of science”, Science, vol. 359, 2018

[3] A Firsov, I. Titov, “Inter-country Competition and Collaboration in the miRNA Science Field”, Perspectives of System Informatics, pp. 26-40, 2019

[4] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, “A Density-Based Algorithm for Discovering Clusters a Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, pp. 226-231, 1996

[5] L. R. Dice, “Measures of the Amount of Ecologic Association Between Species”, vol. 26, pp. 297-302, 1945

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

DOI 10.18699/BGRS/SB-2020-012

Functional roles of the E3 ubiquitin ligase HYD in *Drosophila* tissues

Iuliia Galimova
IMCB SB RAS, Novosibirsk, Russia
galimova@mcb.nsc.ru

Natalia Dorogova
ICG SB RAS, Novosibirsk, Russia
dorogova@bionet.nsc.ru

Svetlana Fedorova
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
fsveta@bionet.nsc.ru

Abstract — **Drosophila tumor suppressor HYD is required for the regulation of cell proliferation, growth and differentiation. HYD is involved in cell processes at different levels: it can regulate gene expression by binding to its promoters, it binds to some RNA for gene silencing and it participates in protein degradation due to its ubiquitin-ligase activity. Analysis of *Drosophila hyd* mutants revealed that HYD functions depend on tissue and stage of development: misexpression in larval somatic tissues causes over-proliferation, defects of meiosis and spermatid differentiation in spermatogenesis, and massive cell death with occasional germline overproliferation during oogenesis.**

Keywords — ubiquitin ligase, tumor suppressor, overproliferation, cell death, germline and somatic tissues

Motivation and aim

The Ubiquitin-Proteasome System (UPS) is an important regulator of cell signaling and proteostasis, which are essential to a variety of cellular processes. Components of UPS are implicated in regulation of cell cycle progression, signal transduction, DNA damage response, metabolism and transcriptional control. E3 ubiquitin ligases play key role in UPS functioning. Ubiquitin ligases (E3 enzymes) transfer ubiquitin from ubiquitin-conjugating (E2) enzymes to target proteins. By determing the selection of target proteins, modification sites on those target proteins, and the types of ubiquitin modifications that are formed, E3 enzymes are key specificity factors in ubiquitin signaling.

One of the most interesting E3 ligases is drosophila tumor suppressor HYD (hyperplastic disc). It is required for the regulation of cell proliferation during development. Mutations in the *hyd* gene result in developmental abnormalities that include adult sterility caused by germ cell defects [1]. In eye imaginal disc HYD function differs from what is expected for conventional tumor suppressor behaviour [2]. In eye disc homozygous *hyd* mutant clones induce non-autonomous overproliferation of nearby tissue [2].

UPS is highly conserved through evolution. Human orthologue of drosophila HYD – named EDD1 or UBR5 - was identified in 1998 [3]. UBR5 expression is deregulated in many cancer types, amplification of *UBR5* has been reported in human breast and ovarian cancer [4]. However, the mechanism by which UBR5 may contribute to tumor initiation and progression remains poorly defined.

Methods

To define the role of HYD in *Drosophila* germ line tissues and reveal the cause of sterility we studied the processes of spermatogenesis and oogenesis in drosophila *hyd* mutants. We used Bloomington *Drosophila* stocks: kni hyd15 e1 /TM3, Sb1 (3718) and yw; PBac{3HPy+}hydC017/TM3, Sb1 Ser1 (16256) as the source of mutant *hyd* alleles. Detailed cytological analysis of oogenesis and spermatogenesis was performed using fluorescent and electronic microscopy.

Results and discussion

Analysis of spermatogenesis showed that mutations of this gene cause multiple cell division abnormalities in spermatogenesis, but do not lead to hyperplasia of both generative and somatic tissue [5]. In contrast, main effect of *hyd* mutations in oogenesis was massive cell death. However, about 5% of ovaries contained egg chambers with abnormal amount of germ cells – 2-4 times higher than in control (32- or 64-cell cysts in comparing to normal 16-cells). It indicates 1 or 2 extra rounds of cell division of mutant cystoblasts. It should be noted that *hyd* mutations did not affect somatic follicular cells.

Thus, in different tissues mutations of *hyd* lead to various consequences: overproliferation in somatic tissues, defects of meiosis and spermatid differentiation in drosophila spermatogenesis, and massive cell death with occasional overproliferation during oogenesis. So, HYD functions may differ depending on cell context. It was shown that HYD can regulate gene expression at transcriptional level [6] and translational level, owing to its unique feature – presence of PABC domain, commonly found in pABP proteins, whose function is related to mRNA translation. Also it is known that mammalian EDD plays a critical role in miRNA silencing [7] and participates in DNA damage response [8]. Interesting, that E3 ubiquitin ligase activity is dispensable for EDD function in miRNA silencing. The PABC domain of EDD is essential for its silencing function.

It is clear that HYD possess many apparently divergent roles in multiple pathways in different tissues. Earlier studies have shown that tumor suppressor genes can function in many tissues, but only in some of them they cause hyperplasia [9]. Futher researches are required to better define the role of HYD in different cell processes.

Acknowledgment

Supported by the ICG SB RAS budget Project (0324-2019-0042-C-01).

References

[1] Mansfield E.A. et al. (1994). Genetic and molecular analysis of hyperplastic discs, a gene whose product is required for regulation of cell proliferation in Drosophila melanogaster imaginal discs and germ cells. Dev. Biol. 165(2): 507-526.

[2] Lee J.D. et al. (2002). The ubiquitin ligase Hyperplastic discs negatively regulates hedgehog and decapentaplegic expression by independent mechanisms. Development. 129(24): 5697-5706.

[3] Callaghan M.J. et al. (1998). Identification of a human HECT family protein with homology to the Drosophila tumor suppressor gene hyperplastic discs. Oncogene. 17: 3479-3491.

[4] Clancy J.L. et al. (2003). EDD, the human orthologue of the *hyperplastic discs* tumour suppressor gene, is amplified and overexpressed in cancer. Oncogene. 22: 5070-5081.

[5] Pertceva J.A. et al. (2010) The role of *Drosophila hyperplastic discs* gene in spermatogenesis. Cell Biol. Int. 34(10): 991-996.

[6] Wang G. et al. (2014). Hyperplastic discs differentially regulates the transcriptional outputs of hedgehog signaling. Mech. Dev. 133: 117-125.

[7] Su H. et al. (2011). Mammalian hyperplastic discs homolog EDD regulates microRNA-mediated gene silencing. Mol Cell. 43(1): 97–109.

[8] Henderson M.J. et al. (2002). EDD, the human hyperplastic discs protein, has a role in progesterone receptor coactivation and potential involvement in DNA damage response. J. Biol. Chem.19;277(29):26468-78.

[9] Gateff E. (1994). Tumor suppressor and overgrowth suppressor genes of Drosophila melanogaster: developmental aspects. Int. J. Dev. Biol. 38(4):565-90.

DOI 10.18699/BGRS/SB-2020-013

Promoter expression landscape in skeletal muscle in hindlimb suspension and recovery model in rat

Guzel Gazizova
Kazan Federal University, Kazan, Russia
grgazizova@gmail.com

Islam Nigmetzyanov
KFU, Kazan, Russia
islamka32@gmail.com

Ilia Akberdin
BIOSOFT.RU, LLC; Novosibirsk State University, Novosibirsk, Russia
akberdinir@biosoft.ru

Sergei Pintus
BIOSOFT.RU, LLC; Institute of Computational Technologies, Novosibirsk, Russia
sspintus@biosoft.ru

Oksana Tyapkina
FRC KSC RAS, Kazan, Russia
antitoxin@icloud.com

Fedor Kolpakov
Institute of Computational Technologies, Novosibirsk, Russia
fkolpakov@gmail.com

Ruslan Deviatiiarov
KFU, Kazan, Russia
ruselusalbus@gmail.com

Leniz Nurullin
FRC KSC RAS, Kazan, Russia
leniz2009@gmail.com

Oleg Gusev
KFU, Kazan, Russia
RIKEN, Yokohama, Japan
oleg.gusev@riken.jp

***Abstract* — Loss of muscle mass and function during long period of physical inactivity still remains a clinical problem for humanity because the significant reducing of life quality and increasing mortality. In recent studies global gene expression in atrophied and recovered skeletal muscles on different animal models was analyzed. However, previously whole-genome regulation of atrophy and following recovery at promoter and enhancer level has not been studied. To identify transcription initiation sites (TSS), and evaluate full-genome RNA expression at the level of individual promoters and enhancers during unloading and subsequent recovery in rats two types of muscles, "slow" (m. Soleus) and "fast" (m. EDL), were examined in rats in normal conditions, after 1, 3 and 7 days of hindlimb suspension and following 1, 3 and 7 days of recovery using CAGE (Cap Analysis of Gene Expression) method followed by Illumina HiSeq 2500 sequencing. We obtained 9971 unique CAGE clusters, among which 6377 were promoters associated with genes. Differential expression of genes and their promoter activity were strongly varied in m. EDL and m. soleus within suspension-recovery time course. This study provides the first systematic annotation of promoters landscape and genes activated in "fast" and "slow" muscle types under induced atrophy and following recovery in rats.**

Keywords* — *muscle atrophy, muscle recovery, gene expression, promoter, CAGE

Motivation and aim

Skeletal muscle represents adaptive system characterizing by plasticity. It is capable to remodeling in response to different external and internal stimuli. Mechanical unloading causes muscle atrophy but at the same time physical activity induces hypertrophy. Loss of muscle mass and function during long period of physical inactivity still remains a clinical problem for humanity because the significant reducing of life quality and increasing mortality. For resolving this task understanding molecular regulatory mechanisms of disuse muscle atrophy and recovery are required. In recent studies global gene expression in atrophied and recovered skeletal muscles on different animal models was analyzed. However, previously whole-genome regulation of atrophy and following recovery at promoter and enhancer level has not been studied. Here, for the first time we perform analysis of regulatory genome elements in skeletal muscles in hindlimb suspension-recovery rat model. One of the most sensitive approaches for such analysis is Cap Analysis of Gene Expression method (CAGE), which allows determine transcription initiation sites (TSS) with up to one nucleotide precise, thus, evaluate the transcriptional activity of genes at promoter and enhancer level. In current study we aimed to identify transcription initiation sites (TSS), and evaluate full-genome RNA expression at the level of individual promoters and enhancers during unloading and subsequent recovery in rats.

Methods

Two types of muscles, "slow" (m. Soleus) and "fast" (m. EDL), were examined in rats in normal conditions, after 1, 3 and 7 days of hindlimb suspension and following 1, 3 and 7 days of recovery using CAGE (Cap Analysis of Gene Expression) method followed by Illumina HiSeq 2500 sequencing. After quality check and filtration, CAGE reads were mapped to the current rat genome assembly rn6 (2014) by using bwa and then clustered by python script. Further annotation of CAGE peaks, analysis of differential expression, and functional terms enrichment were proceeded through R environment.

Results

We obtained 9971 unique CAGE clusters, among which 6377 were promoters associated with genes. The remaining promoters were suggested to be unannotated transcripts, remotely located promoters of known genes and non-coding RNAs.

Differential expression of genes and their promoter activity were strongly varied in m. EDL and m. soleus within suspension-recovery time course: "slow" m. soleus has no significant changes of transcriptional activity up to 7 days of suspension, but drastically shifted during recovery, while "fast" m. EDL shows quick and stable response to the stress and fast recovery after placing in normal conditions. First atrophic changes were marked after 3 days of hindlimb suspension in soleus muscle and associated with upregulation of proteasome and peptidase complex genes. After the first day of recovery genes involved to transition between fast and slow fibers were activated in m. EDL. In addition, key transcription factors regulating the activity of specific muscle promoters were identified. Also enrichment of signaling pathways by differentially expressed genes was conducted.

This study provides the first systematic annotation of promoters landscape and genes activated in "fast" and "slow" muscle types under induced atrophy and following recovery in rats.

Acknowledgment

The reported study was funded by RFBR according to the research project No. 17-00-00243.

DOI 10.18699/BGRS/SB-2020-015

Comparison of brain transcriptome profiles of short-lived and long-lived species of *Nothobranchius*

Zulfiia Guvatova
Center for Precision Genome Editing and Genetic Technologies for Biomedicine
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences
Moscow, Russia
guvatova.zulfiya@mail.ru

George Krasnov
Center for Precision Genome Editing and Genetic Technologies for Biomedicine
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences
Moscow, Russia
gskrasnov@mail.ru

Sergey Simanovsky
Laboratory for Ecological Monitoring of Nuclear Power Plant (NPP) Regions and Bioindication
Russia A.N. Severtsov Institute of Ecology and Evolution
Moscow, Russia
sergey.a.simanovsky@gmail.com

Alexander Frolov
Laboratory for Ecological Monitoring of Nuclear Power Plant (NPP) Regions and Bioindication
Russia A.N. Severtsov Institute of Ecology and Evolution
Moscow, Russia

Nataliya Gladysh
Veterinary biology faculty
MSAVM&B - MVA named after K.I. Skryabin
Moscow, Russia
natalyagladish@gmail.com

Anna Kudryavtseva
Center for Precision Genome Editing and Genetic Technologies for Biomedicine
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences
Moscow, Russia
rhizamoeba@mail.ru

***Abstract* — Annual fishes of the genus *Nothobranchius* from East Africa is a promising vertebrate model in comparative and ageing studies. *Nothobranchius* show accelerated growth and age-related changes at all levels of organization. The present work aims to compare brain transcriptome profiles of short-lived *N. rachovii* with longer-lived *N. korthausae* and *N. guntheri*.**

Keywords* — *Nothobranchius, RNA-seq, brain aging

Introduction

In recent years, annual fishes of the genus *Nothobranchius* have become powerful vertebrate animal models for the study of aging. Due to the short lifespan, *Nothobranchius* manifest changes associated with age such as skin color fade, spine curvature, weight loss, decline in cognitive abilities and an increased incidence of tumours in a few months after hatching. Moreover, killifish are sensitive to variations in water temperature, which allows easy manipulating their rate of aging under laboratory conditions.

Another important aspect is that different species of the genus *Nothobranchius* show differences in life expectancy, making them useful for comparative ageing research [1]. It has been suggested that the lifespan determination in the genus *Nothobranchius* depends on the rainfall pattern in the biotope of origin and differences in rearing and maintenance conditions. Maximum lifespan for these fish have been observed that range from 3 for *N. furzeri* to 28 months for *N. guentheri*.

It is known that there are conserved features linked to brain aging of mammals, such reduced adult neurogenesis and down-regulation of synaptic and axonal proteins expression. For well-known short-lived teleost fish *N. furzeri* brain aging has also been shown [2]. On the other hand, significant brain regenerative properties and adult neurogenesis were observed for other teleost fish as medaka and zebrafish [3]. The aim of the present study was to compare brain transcriptome profiles of several short- and longer-lived species of the genus *Nothobranchius.*

Materials and Methods

Fish strains and maintenance

N. rachovii Biera-98, N. korthausae MI TZN 08-4, and N. guntheri AS were obtained from commercial supplier (Peter Covar, Brno) and acclimated to the Center for Precision Genome Editing and Genetic Technologies for Biomedicine at the Engelhardt Institute of Molecular Biology. To avoid the effects of maintenance conditions, animals were bred under identical conditions. Fish were kept in 2 litre tanks at 27°C under a 14 h: 10 h day:night regime and fed twice a day with *Chironomus* larvae

RNA isolation, library preparation and transcriptome sequencing

Total RNA was isolated from brain of male and female specimens aged 3 months (adult) using MagNA Pure Compact RNA Isolation Kit on MagNA Pure Compact Instrument (Roche, Switzerland) according to manufacturer's protocol. RNA quantity and quality were evaluated using a Qubit 2.0 fluorometer (Life Technologies, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Samples passed quality (RIN>8) were used for subsequent experiments. Double stranded cDNA library was prepared by using TruSeq RNA Library Prep Kit v2 following manufacturer's protocol from 1 mg of total RNA. Quality of prepared cDNA libraries were checked with bioanalyzer Agilent 2100 using High Sensitivity DNA chip (Agilent Technologies, USA). Then, libraries were normalized to 4 nM, pooled together in equal volumes, and sequenced on the NextSeq™500 system (Illumina, USA).

NGS Data Processing

About 100 million paired-end Illumina reads (2*150 bp) were derived for each of 3 *Nothobranchius* species (*N. rachovii, N. korthausae, N. guntheri*). Reads were trimmed with trimmomatic 0.39, and then reads were processed using cutadapt 2.8 to ensure the complete removal of adapters. Next, we performed transcriptome assembly using Trinity 2.9.1. For this purpose, we pooled libraries. The assembled transcripts were analyzed for bacterial contamination using Kraken 2 (MiniKraken DB_8GB), and bacterial contigs have

been removed. Overall statistics have been calculated using QUAST. The completeness of the assembled transcriptomes was evaluated using BUSCO 4.0.5 (Vertebrata and Actinopterygii OrthoDB v10 datasets). The assembled transcripts were annotated using Trinotate 3.2.0 pipeline. Next, we evaluated gene/transcript expression profiles and calculated ExN50 statistics. For this purpose, reads were mapped to the assembled transcripts using bowtie2 2.4.1 and then transcript/gene expression levels were inferred by RSEM 1.3.3. The derived read counts were transformed into FPKM values.

Next, we tried to identify orthologous gene groups for all 3 organisms using OrthoFinder. For this purpose, we translated the derived transcripts using TransDecoder and selected the longest ORF for each gene. The derived amino acid sequences and selected transcript were supplied to OrthoFinder. To reduce CPU time, we limited this analysis only to annotated genes. Finally, for each orthologous gene group, we compared gene expression levels (in terms of FPKM) between 3 organisms. Because of inevitable presence of the experimental biases between organisms (sequence-dependent efficacy of PCR amplification, sequencing), only 3-fold or greater expression level changes were considered as significant.

Results

The overall lengths of the assemblies varied from 393 to 458 Mb; the total number of genes varied from 56 to 71 thousand. From 19 to 26 thousand genes were successfully annotated with Trinotate. About 4% transcripts were marked by Kraken as contaminants. According to the results of BUSCO analysis, the completeness of the assemblies varied from 91% to 94% (Vertebrata dataset) and from 82% to 86% (Actinopterygii dataset; this suggests that Vertebrata dataset is more appropriate for *Nothobranchius sp*.). Using OrthoFinder, we successfully found orthologous groups for 54% annotated genes (all 3 species), 84% (at least 2 species).

Differential expression (DE) analysis showed that genes associated with chromatin remodeling were the most differentially expressed. According to the results, the genes CBX1A, CBX7A were up-regulated (counts per million (CPM)> 3, false discovery rate (FDR) <0.05) in short-lived N. rachovii by 8 times compared with long-lived N. guntheri and 5 times compared to N. korthausae. We also observed the down-regulation of enhancer of zeste 1(*EZH1)* expression levels *in N. rachovii (CPM>3, FDR<0.05) when compared to N. korthausae*, but not to *N. guntheri.* The change of expression of these genes with age was also observed in human brain [4]. Pathway enrichment analysis revealed a difference in expression of gene associated with «Ribosome», «RNA processing» and «Lysosome» GO terms. We also found that short-lived fish had reduced expression of genes related to such age-associated pathways as Wnt and Noth signaling pathways. Thus, the results obtained in the work might contribute to better understanding of differences in lifespan between species of this genus.

ACKNOWLEDGMENT

This work was supported by grant 075-15-2019-1660 from the Ministry of Science and Higher Education of the Russian Federation.

REFERENCES

[1] T. Genade M. Benedetti, E. Terzibasi, P. Roncaglia, D.R. Valenzano, A. Cattaneo, A. Cellerino. “Annual fishes of the genus Nothobranchius as a model system for aging research.,” Aging Cell, vol. 4, no. 5, pp. 223–33, Oct. 2005, doi: 10.1111/j.1474-9726.2005.00165.x.

[2] M. Baumgart M. Groth, S. Priebe, A. Savino, G. Testa, A. Dix, R. Ripa, F. Spallotta, C. Gaetano, M. Ori, E. T. Tozzini, R. Guthke, M. Platzer, and A. Cellerino. “RNA-seq of the aging brain in the short-lived fish N. furzeri - conserved pathways and novel genes associated with neurogenesis.,” Aging Cell, vol. 13, no. 6, pp. 965–74, Dec. 2014, doi: 10.1111/acel.12257.

[3] P. M. Loerch, T. Lu, K.A. Dakin, J.M. Vann, A. Isaacs, C. Geula, J. Wang, Y. Pan, D.H. Gabuzda, C. Li, T.A. Prolla, B.A. Yankner. “Evolution of the aging brain transcriptome and synaptic regulation.,” PLoS One, vol. 3, no. 10, p. e3329, Oct. 2008, doi: 10.1371/journal.pone.0003329.

[4] S. Horvath Y. Zhang, P. Langfelder, R.S. Kahn, M.P. Boks,K. van Eijk, L.H. van den Berg, R.A. Ophoff. “Aging effects on DNA methylation modules in human brain and blood tissue.,” Genome Biol., vol. 13, no. 10, p. R97, Oct. 2012, doi: 10.1186/gb-2012-13-10-r97.

DOI 10.18699/BGRS/SB-2020-016

Genome assembly and annotation of *Nothobranchius rachovii* killifish

Zulfiia Guvatova
Center for Precision Genome Editing and Genetic Technologies for Biomedicine
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences
Moscow, Russia
guvatova.zulfiya@mail.ru

George Krasnov
Center for Precision Genome Editing and Genetic Technologies for Biomedicine
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences
Moscow, Russia
gskrasnov@mail.ru

Anastasiya Snezhkina
Center for Precision Genome Editing and Genetic Technologies for Biomedicine
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences
Moscow, Russia
leftger@rambler.ru

Artemy Tokarev
Veterinary biology faculty
MSAVM&B - MVA named after K.I. Skryabin
Moscow, Russia
artemiy.tokarev@mail.ru

Maria Fedorova
Laboratory of Postgenomic Research
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences
Moscow, Russia
fedorowams@yandex.ru

Anna Kudryavtseva
Center for Precision Genome Editing and Genetic Technologies for Biomedicine
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences
Moscow, Russia
rhizamoeba@mail.ru

***Nothobranchius rachovii* is a species of freshwater annual killifish native to Africa. This organism is an attractive model in aging studies because of short lifetime and rapid maturation. In this study, we present draft genome assembly of this organism based on Nanopore reads and polished with Illumina reads. We tested four popular Nanopore reads assemblers (Flye, MaSuRCA, Shasta, wtdbg2) and two polishers (Pilon and MaSuRCA-polish tool). We selected the best combination of these tools in terms of contig length and assembly completeness.**

Keywords: Nothobranchius rachovii, genome assembly, Nanopore,

Introduction

Nothobranchius rachovii is a species of killifish native to Africa (southern Mozambique). Due to very short duration of the rain season, the natural lifespan of these animals is limited to a few months and their captive lifespan is likewise short. Along with *N. furzerii,* another species from this genus, *N. rachovii* is an attractive model system for ageing and disease research [1, 2]. With respect to the molecular, cellular, and integrative traits of aging, these species show significant similarities to mammals, including humans [1, 2]. The complete genome assembly of *N. rachovii* still lacks.

The genome of *N. furzerii*, a close phylogenetic relative of *N. rachovii* [3], is abnormally enriched with tandem repeats (21% of its length), which has been suggested as a factor in its fast ageing [4]. On other hand, this complicates the genome assembly and make challenging to get high accuracy, attractive contig lengths stats and low misassembly rate.

In recent years, Oxford Nanopore Technology (ONT) sequencers are widely used to produce high-quality and inexpensive genome assemblies of various organisms, including those with complex genomic structure (low complexity, abundance of tandem repeats, pronounced heterozygosity, polyploidy). The field of the development of software for the genome assembly based on ONT reads is one of the fastest growing at the moment.

In this study, we aimed to derive high accuracy genome assembly of *N. rachovii* based on ONT and Illumina reads. For this purpose, we tested several most popular and recent assemblers and polishers, compared the results in terms of assembly completeness, contig lengths and similarity to the assembled genome of *N. furzerii.*

Materials and Methods

Sample preparation, sequencing

Genomic DNA was extracted from muscle tissue of N. rachovii using DNeasy Blood and Tissue Kit (QIAGEN, USA) according to the manufacturers instructions. To generate the long reads DNA library from 1 μg of high molecular weight DNA was performed following the SQK‑LSK108 protocol (Oxford Nanopore Technologies, UK). Sequencing was performed on five R9.4.1 flow cells on a MinION (Oxford Nanopore Technologies, UK). The short-reads was generated using HiSeq 2000 Sequencing System (Illumina, USA) according to manufacturer's instructions. The libraries for sequencing on HiSeq 2000 System (Illumina, USA) were constructed using 500 ng of genomic DNA using the TruSeq DNA LT kit(Illumina, USA) according to the manufacturer's instructions. RNA extracted from muscle tissue of fish was used for library construction and RNA sequencing (RNA-seq) on NextSeq500 System (Illumina, USA).

Genome assembly and annotation

First, we derived nucleotide sequences (fastq) from the obtained raw electrical signal intensity data (fast5) using Guppy and fast "flip-flop" high accuracy calling algorithm. Using trimmomatic we filtered out low-quality reads with average phred-scaled Q < 4 (no trimming has been applied). Next, we performed primary genome assemblies of *N. rachovii* genome (based on ONT reads) using four different assemblers: Flye (2.7), MaSuRCA (3.3.4), Shasta (0.4.0), and wtdbg2 (2.5). The derived contigs were polished by Racon (1.4.3) with ONT reads (2 iterations) and then polished with Illumina reads parallelly by Pilon (1.23) and by MaSuRCA-polish tool (3.3.4).

At this step, we calculated assembly metrics using QUAST and evaluated assembly completeness using BUSCO (Vertebrata and Actinopterygii datasets). For the further

analysis, we selected only the best assembly. Additionally, we measured the efficacy of polishers, also by BUSCO metrics.

Next, we excluded redundant contigs (coming from diploidy or mis-assembled contigs) using PurgeHaplotigs tool (Feb 2020). At this step, we adjusted PurgeHaplotigs parameters in order to maximally shorten assembly length, but not to impair its completeness (in terms of BUSCO metrics). Finally, the derived pool of contigs was mapped to the database of bacterial genomes (MiniKraken DB) in order to eliminate contaminants.

The derived contigs were mapped to the reference genome of *Nothobranchius furzerii* (CCSH00000000.1) and indel/SNP rates were evaluated. Dotplots were generated by LAST and mummer. Additionally, we tried to find orthologous between *N. furzerii* and *N. rachovii* genes using OrthoFinder.

The annotation of the assembled contigs was performed by funannotate (1.7.0) pipeline. For this purpose, we performed transcriptome assembly of *N. rachovii* RNA-Seq reads using Trinity (genome-guided), and then the derived transcripts were mapped to the genomic contigs using minimap2. These BAM files were transferred to the PASA pipeline (incorporated into funannotate) in order to train Augustus, snap, GlimmerHMM and pass the derived results to the EvidenceModeler. After filtering gene models, gene annotation was performed (using UniProt, BUSCO, eggNOG, InterProScan databases).

Results

In the present work, we performed assembly of N. rachovii genome using ONT reads and four assemblers. Among them, Flye demonstrated the best results in terms of contig lengths and assembly completeness. After eliminating redundant contigs, the genome length comprised 1183 Mb with N50 = 7.5 Mb. The completeness of the assembly was evaluated as 98.6% (completed/duplicated BUSCOs; Vertebrata dataset), 95.4% (Actinopterygii dataset). Shasta gave worse results in terms of assembly completeness, whereas MaSuRCA lower N50 with comparable assembly completeness. wtdbg2 demonstrated the worse result.

26.897 protein-coding genes have been identified in the assembly. There were about 1.8 alternatively spliced transcript variants per gene. Besides, we identified 231 tRNA genes. The length of mitochondrial genome comprised 21 Kb, contained 23 tRNAs and 13 protein-coding genes. For 19.844 genes we did find orthologs in *N. furzerii* reference genome. The overall sequence similarity between *N. furzerii* and *N. rachovii* comprised 82.5% for the entire genome, 93.5% for coding nucleotide sequences, and 96% for the encoded protein sequences.

Consclusions

In this study, we present a draft genome assembly of *Nothobranchius rachovii*, a species of killifish native to Africa. This organism is considered at attractive model system for ageing and disease research. Its genome shares high homology to the *Nothobranchius furzerii* and is highly enriched with tandem repeats, which is thought to be one of the reasons for its rapid aging.

Acknowledgment

This work was supported by grant 075-15-2019-1660 from the Ministry of Science and Higher Education of the Russian Federation.

References

[1] C.Y. Hsu, Y.C. Chiu, W.L. Hsu, Y.P. Chan. “Age-related markers assayed at different developmental stages of the annual fish Nothobranchius rachovii,” J. Gerontol. A. Biol. Sci. Med. Sci., vol. 63(12), pp.1267-76, Dec 2008. doi: 10.1093/gerona/63.12.1267

[2] M. Platzer, C. Englert. “Nothobranchius furzeri: A Model for Aging Research and More,” Trends. Genet., vol. 32 (9), pp. 543–552, Sept 2016. doi: 10.1016/j.tig.2016.06.006.

[3] A. Dorn, Musilová Z, Platzer M, Reichwald K, Cellerino A. “The strange case of East African annual fishes: aridification correlates with diversification for a savannah aquatic group?” BMC Evol Biol, vol. 14, p.210, 2014 Oct. doi: 10.1186/s12862-014-0210-3.

[4] K. Reichwald, C. Lauber, et al. “High tandem repeat content in the genome of the short-lived annual fish Nothobranchius furzeri: a new vertebrate model for aging research,” Genome Biol, vol. 10(2), R16, Feb 2009. doi: 10.1186/gb-2009-10-2-r16.

DOI 10.18699/BGRS/SB-2020-017

Transcription factor Kaiso regulates cell division in the developing mouse brain

Nina Illarionova
ICG SB RAS, Novosibirsk, Russia
nina.illarionova@gmail.com

Maria Borisova
ICG SB RAS, Novosibirsk, Russia
zolotykh@bionet.nsc.ru

Ekaterina Bazhenova
ICG SB RAS, Novosibirsk, Russia
ekaterina.yu.bazhenova@gmail.com

Daria Fursenko
Institute of Gene Biology RAS,
Moscow, Russia

Daria Zabelina
NSU, Novosibirsk, Russia
dazabelina@gmail.com

Nikita Khotskin
ICG SB RAS, Novosibirsk, Russia
khotskin@bionet.nsc.ru

Alexander Kulikov
ICG SB RAS, Novosibirsk, Russia
v_kulikov@bionet.nsc.ru

***Abstract* — Kaiso is a transcription factor that is known to regulate cell division in different cell types, including cancer cells. Kaiso binds kaiso specific binding site and/or methylated CpG islands on the promoter region of different target genes. Kaiso silencing may lead to an increase or a decrease in cell proliferation, which is cell specific and is probably dependent on the methylation status of Kaiso binding sites at the target promoter [1]. In various mammalian cell types Kaiso was shown to regulate cell proliferation via its target genes of cyclins E1 and D1 guarding the G1 to S phase transition in cell division [1].**

Keywords* — *kaiso; cell division; cyclins; hippocampus

Motivation and aim

Motivation

Kaiso is widely expressed throughout the brain regions, however its regulation of cell division in brain development was not investigated.

Aim

Our aim was to study the role of the transcription factor Kaiso in cell division in the developing mouse brain.

Methods

The study was performed on littermate C57BL/6j mice of WT and Kaiso knockout (KO) genotype. Using the method of immunohistochemistry with a cell division marker bromodeoxyuridine we have calculated newborn cells within the 24 h period in the hippocampus, cortex, striatum and subventricular zone of the lateral ventricles on the second day after birth. Using real-time PCR we have analyzed expression profiles of associated with cell division genes (*c-myc*, *CCNE1* (cyclin E1) and *CCND1* (cyclin D1)) in hippocampus, cortex and striatum at three different ages: embryonic day 16 (E16), postnatal day 2 (P2) and postnatal day 40 (P40).

Results

Our preliminary data showed that in the developing hippocampus (dentate gyrus region) there was a significant increase in the number of newborn cells in KO mice (17 ± 1) compared with the WT mice (12 ± 1 per 100 μm2; $p < 0{,}001$). The number of newborn cells co-stained with the neuron marker (NeuN) was also significantly greater in KO mice (6 ± 1 per 100 μm2) compared with the WT mice (4 ± 1 per 100 μm2; $p < 0{,}05$).

In the cortex, striatum, and subventricular zone of the lateral ventricles, there were no significant differences between genotypes in the number of newborn cells.

We have performed expression profiling of genes associated with cell division regulation (*c-myc, CCNE1 and CCND1*) that were previously noted to be regulated by Kaiso in different cell types [1,2]. A significant decrease in the expression of the *c-myc* gene was shown in the hippocampus of KO mice at the age P2 compared to WT (KO: 2,7 ± 0,3; WT: 4,5 ± 0,7; $p < 0{,}05$). On the mRNA level neither *CCNE1* nor *CCND1* expression was different between the genotypes in all selected brain regions and corresponding age groups.

Acknowledgment

Supported by the RFBR (18-04-00869 A), budget project (AAAA-A17-17072710029-7) and using the equipment of the Center for Laboratory Animal Genetic Resources Center, Federal Research Center of Cytology and Genetics SB RAS, supported by the Russian Ministry of Education and Science (RFMEFI62119X0023).

References

[1] Pozner A. et al. (2016) Cell-specific Kaiso (ZBTB33) Regulation of Cell Cycle through Cyclin D1 and Cyclin E1. THE JOURNAL OF BIOLOGICAL CHEMISTRY. 291(47): 24538–24550.

[2] Koh D. et al. (2013) Kaiso is a key regulator of spleen germinal center formation by repressing Bcl6 expression in splenocytes. Biochemical and Biophysical Research Communications. 442: 177–182.

DOI 10.18699/BGRS/SB-2020-018

PCR dependent biases could significantly affect quantitative estimation of plant mix composition

Valeriia Kaptelova
Central Research Institute of Epidemiology, Moscow, Russia
valeriia.kaptelova@gmail.com

Denis Omelchenko
Institute of Information Transmission Problems, Moscow, Russia
omdeno@gmail.com

Andrey Ayginin
Center for Strategic Planning and Management of Biomedical Health Risks Moscow, Russia
ayginin75@gmail.com

Andrei Samoilov
Central Research Institute of Epidemiology, Moscow, Russia
andrei.samoilov@gmail.com

Maria Logacheva
Skolkovo Institute of Science and Technology, Skolkovo, Russia
Lomonosov Moscow State University, Moscow, Russia
maria.log@gmail.com

Anna Fedotova
Skolkovo Institute of Science and Technology, Skolkovo, Russia
Lomonosov Moscow State University, Moscow, Russia

Kamil Khafizov
Center for Strategic Planning and Management of Biomedical Health Risks Moscow, Russia
kkhafizov@gmail.com

Anna Speranskaya
Central Research Institute of Epidemiology, Moscow, Russia
Lomonosov Moscow State University, Moscow, Russia
hanna.s.939@gmail.com

Anastasia Krinitsina
Lomonosov Moscow State University, Moscow, Russia
I.M. Sechenov First Moscow State Medical University, Pharmaceutical Natural Science Department, Moscow, Russia
ankrina@gmail.com

Elena Korneenko
Central Research Institute of Epidemiology, Moscow, Russia
lennatta@yandex.ru

***Abstract* — Metagenomic analysis using high-throughput sequencing is an intensively developing approach nowadays. One of its problems is adequate quantification of components in metagenomic samples.**

Keywords — food; nrITS1; metagenomics; polymerase; NGS

Motivation and aim

Metagenomic analysis using high-throughput sequencing is an intensively developing approach nowadays. In contrast to the qualitative analysis, the quantitative one still remains very challenging. The main factors that hamper quantitative analysis are differences in GC-content and length of marker regions. Our aim was to investigate and improve the potential of metabarcoding as analytic systems for quantitative identification of complex plant mixes used for food. In plants nrITS1 sequence length can vary almost twice. The percentage of the GC-composition also varies significantly: min10%/median/max10% were found 48.1/58.0/67.5.

Methods

We prepared and analyzed the artificial plant mixes from leaves of seven cultivated plants used for food, as well as model species *Arabidopsis thaliana*. Plant species were chosen according to their different GC-composition (from 53% to 72,9%) and length of *nrITS1* (from 192 to 348 bp) in order to be able to compare the results of the evaluation. The following species were used for the experiments: *Oryza sativa* L. (rice), *Arabidopsis thaliana* (L.) Heynh, *Helianthus annuus* L. (sunflower), *Fagopyrum esculentum* Moench (buckwheat), *Cucurbita pepo* L. (pumpkin). Equal weights of leaf powder were mixed, then DNA was extracted using sorbent purification-based Diamond DNA Plant kit (ABT, Russia). nrITS1 barcodes were amplified (in three replicates for each mix) using polymerases from different vendors: Encyclo proofreading polymerase («Encyclo Plus PCR kit», Evrogen, Russia) and Q5 High-Fidelity polymerase («Q5® Hot Start High-Fidelity 2X Master Mix», NEB, England). Libraries were constructed using Ion Xpress™ Plus Fragment Library Kit (Thermo Fisher Scientific) according to manufacturer's instructions, including amplification by Platinum HiFi Polymerase. Another approach was library preparation using the same kit but without amplification (PCR-free). Further all libraries were sequenced on Ion Torrent S5.

Results

For the libraries prepared from DNA samples with the standard protocol Ion Xpress Plus the congruence with expected results was extremely poor for the mixes containing *O. sativa* (Table 1). *O. sativa ITS1* has high GC content - 72,9%. On the other hand, the *nrITS1* of *F. esculentum* consists of 68,5% GC (i.e. the value is also high) but analysis of samples containing this component showed correct results, perhaps due to shorter length of barcode (in comparing with *nrITS1* of the other components in the same samples). We made a series of additional experiments using fragment length analysis of amplified libraries and showed that at least one of the reasons for bias is inefficient amplification of extremely high GC-rich fragments by Platinum HiFi polymerase during library preparation. When PCR-free amplification protocol was used the quantification, results were improved significantly for *O. sativa* containing mixes. The aim of this investigation was to check the hypothesis that GC-composition and length of DNA-barcodes could significantly affect quantitative identification of multicomponent plant foods. We found that this dependence exists indeed and can affect PCR efficiency leading to yield inaccurate results of quantification. It may be possible to improve the interpretation of sequencing data using optimization of experimental protocols, additional controls [1] and through the implementation of special algorithms that correct sample preparation biases through correction coefficients.

ACKNOWLEDGMENT

Supported by the Ministry of education and science of Russia, project # 14.609.21.0101, unique grant identifier RFMEFI60917X0101.

REFERENCES

[1] D.O. Omelchenko, A.S. Speranskaya, A.A. Ayginin, K. Khafizov, A.A. Krinitsina, A.A. Fedotova, D.V. Pozdyshev, V.Y. Shtratnikova, E.V. Kupriyanova, G.A. Shipulin, M.D. Logacheva. (2019) Improved Protocols of ITS1-Based Metabarcoding and Their Application in the Analysis of Plant-Containing Products. Genes. 10(2): 122.

TABLE 1

Components of mix	**GC-composition (%)**	**Length (b.p.)**	**Q5 High-Fidelity Hot Start polymerase, % of reads**	**Encyclo proofreading polymerase, % of reads**	**Q5 High-Fidelity Hot Start polymerase, % of reads (PCR-free)**	**Encyclo proofreading polymerase, % of reads (PCR-free)**
			Average (SD)	***Average (SD)***	***Average (SD)***	***Average (SD)***
A. thaliana	55,9	348	25,8(4,61)	24,87(10,9)	28,1(3,12)	33,5(6,41)
C. pepo	53	192	74,14(4,61)	68,27(1,15)	71,73(2,93)	66,5(6,41)
A. thaliana	55,9	348	43,76(6,28)	N/A	45,27(2,02)	57,63(0,81)
F. esculentum	68,5	259	56,09(6,32)	N/A	54,57(2,08)	42,4(0,87)
A. thaliana	55,9	348	98,3(0,45)	99,53(0,15)	58,33(8,58)	78,43(13,21)
O. sativa	72,9	275	1,14(0,59)	0,47(0,15)	41,5(8,42)	21,5(13,17)
H. annuus	50	343	62,61(6,86)	55,17(6,78)	58,67(5,13)	63,43(15,72)
F. esculentum	68,5	259	37,09(7,28)	44,77(6,9)	40,45(4,85)	36,57(15,72)
	59,5	300	8,85(11,73)	42,07(2,5)	20,07(17,44)	N/A

DOI 10.18699/BGRS/SB-2020-019

Glycosylation of immunoglobulin G is a complex trait regulated by a large network of genes

Lucija Klarić
MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh Edinburgh, United Kingdom Genos Glycoscience Research Laboratory Zagreb, Croatia

Yakov A. Tsepilov
Laboratory of Recombination and Segregation Analysis Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia Novosibirsk State University Novosibirsk, Russia

Chloe M. Stanton
MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh Edinburgh, United Kingdom

Yurii S. Aulchenko
Laboratory of Glycogenomics Institute of Cytology and Genetics of the SBRAS Novosibirsk, Russia PolyOmica 's-Hertogenbosch, The Netherlands

Gordan Lauc
Genos Glycoscience Research Laboratory, Zagreb, Croatia Faculty of Pharmacy and Biochemistry, University of Zagreb Zagreb, Croatia

Caroline Hayward
MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh Edinburgh, United Kingdom Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh Edinburgh, United Kingdom

Introduction

The majority of proteins undergo post-translational glycosylation, in which complex carbohydrates are attached to the surface of proteins. Effector functions of Immunoglobulin G are regulated by the composition of a glycan moiety, thus affecting activity of the immune system. Aberrant glycosylation of IgG has been observed in many diseases, but little is understood about the mechanisms behind these changes. Here we show that the glycan fraction of IgG is a complex trait under control of an interconnected set of genes and an intermediate link between genotype and disease. In the largest genome-wide association study of IgG N-glycosylation to date (N=8,090) we found 27 associated loci (15 novel) and developed a data-driven network approach to propose how these genes form a functional network regulating glycosylation of IgG. From this network we confirmed in-vitro that the transcription factor IKZF1 regulates the expression of glycosyltransferase FUT8, resulting in increased levels of fucosylated glycans. We also found strong in-silico evidence that RUNX1 and RUNX3 transcription factors, together with SMARCB1 chromatin remodelling protein, regulate expression of glycosyltransferase MGAT3. We showed that glycosylation variants supporting this network are pleiotropic with inflammatory and autoimmune diseases and expression of genes involved in regulation of glycosylation enzymes. This study showed new evidence that variation in key transcription factors coupled with regulatory variation in glycosylation enzyme genes drives changes in IgG glycosylation, modifying IgG function, and has a significant influence on inflammatory and autoimmune diseases.

Aknowledgements

We thank Massimo Mangino, Timo Tõnis Sikka, Tõnu Esko, Eugene Pakhomov, Perttu Salo, Joris Deelen, Stuart J. McGurnaghan, Toma Keser, Frano Vučković, Ivo Ugrina, Jasminka Krištić, Ivan Gudelj, Jerko Štambuk, Rosina Plomp, Maja Pučić-Baković, Tamara Pavić, Marija Vilaj, Irena Trbojević-Akmačić, Camilla Drake, Paula Dobrinić, Jelena Mlinarec, Barbara Jelušić, Anne Richmond, Maria Timofeeva, Alexander K. Grishchenko, Julia Dmitrieva, Mairead L. Bermingham, Sodbo Zh. Sharapov, Susan M. Farrington, Evropi Theodoratou, Hae-Won Uh, Marian Beekman, Eline P. Slagboom, Edouard Louis, Michel Georges, Manfred Wuhrer, Helen M. Colhoun, Malcolm G. Dunlop, Markus Perola, Krista Fischer, Ozren Polasek, Harry Campbell, Igor Rudan, James F. Wilson, Vlatka Zoldoš, Veronique Vitart and Tim Spector for their contribution to this work. ORCADES DNA extractions and genotyping were performed at the Genetics Core of the Clinical Research Facility, University of Edinburgh. We acknowledge the invaluable contributions of the research nurses in Orkney, the administrative team in Edinburgh, and the people of Orkney. COGS thanks the participants in all the studies who contributed to this piece of work and all the recruitment teams and collaborators who made these studies possible. We acknowledge the excellent technical support from M. Walker, and we are grateful to R. Wilson, D. Markie, and all those who continue to contribute to recruitment, data collection, and data curation for the Study of Colorectal Cancer in Scotland studies. The work was supported by funding for the infrastructure and staffing of the Edinburgh CRUK Cancer Research Centre. We acknowledge the expert support on sample preparation from the Genetics Core of the Clinical Research Facility, University of Edinburgh. We also acknowledge all the staff of several institutions in Croatia that supported the CROATIA_Vis and CROATIA_Korcula, fieldwork, including, but not limited to, the University of Split and Zagreb Medical Schools, Institute for Anthropological Research in Zagreb, and the Croatian Institute for Public Health in Split. Genotyping was performed in the Genetics Core of the Clinical Research Facility, University of Edinburgh. We are grateful to the Russian Federal Science and Technology Program of Development of Genetic Technologies.

References

[1] Klaric L, Tsepilov YA, Stanton CM *et al*. Glycosylation of Immunoglobulin G Is Regulated by a Large Network of Genes Pleiotropic With Inflammatory Diseases. Science Advances 2020 Feb 19;6(8):eaax0301. doi: 10.1126/sciadv.aax0301

DOI 10.18699/BGRS/SB-2020-020

Predicting elongation efficiency of gene translation for annotation of bacterial genomes: a case study for biosynthetic gene clusters of nonribosomal peptides

A.I. Klimenko
Kurchatov Genomics Center,
Institute of Cytology and Genetics,
ICG SB RAS
Novosibirsk, Russia
klimenko@bionet.nsc.ru

Yu.G. Matushkin
Institute of Cytology and Genetics,
ICG SB RAS
Novosibirsk, Russia

D.A. Afonnikov
Kurchatov Genomics Center,
Institute of Cytology and Genetics,
ICG SB RAS, Novosibirsk, Russia
Novosibirsk State University, NSU
Novosibirsk, Russia

Abstract — The gene expression levels for bacteria are largely determined by the efficiency of translation elongation. We have performed bioinformatic elongation efficiency analysis of NRP biosynthetic gene clusters (BGCs) obtained from ANTISMASH-DB using whole-genome sequences of bacterial genomes that are available at NCBI Genbank. The analysis has provided the information about distribution of nonribosomal peptide biosynthetic gene clusters in bacteria and their putative translation elongation efficiency.

Key words — nonribosomal peptides; translation elongation efficiency; bacteria; genome annotation

Motivation and Aim

It is hard to underestimate the crucial role that efficiency of gene expression plays in synthetic biology and genome engineering. Though being affected at several levels such as transcription, translation, posttranslational modification, etc., the gene expression levels for bacteria are largely determined by the efficiency of translation elongation [1]. Nonribosomal peptides (NRPs) constitute an important fraction of bacterial peptidomes acting as antibiotics, toxins, surfactants, siderophores, anti-tumor agents and immune response modifiers [2]. Biosynthesis of NRPs is dependent on particular enzymes - nonribosomal peptide synthetases (NRPSs), which are encoded by clusters of genes in bacterial genomes [3].

Methods and Algorithms

The EloE software [4] is a tool for gene ranking based on their putative translation elongation efficiency inferred from their nucleotide sequences taking into account such factors as codon composition, presence and stability of secondary structures in mRNA [1]. The obtained predicted values correlate with available experimental data on gene expression in different microorganisms [4]. Thus, EloE is useful as a bioinformatic tool for genome annotation that enables a researcher with a capacity to infer *a priori* estimates of gene expression efficiency based on whole-genome nucleotide sequences only. We have performed bioinformatic analysis of NRP biosynthetic gene clusters (BGCs) obtained from ANTISMASH-DB [5] using whole-genome sequences of bacterial genomes that are available at NCBI Genbank. The analysis is based on the method predicting gene translation elongation efficiency that is implemented in the EloE software [4]. Statistical and bioinformatic analysis scripts have been developed on Python using software library Biopython.

Results

The statistical analysis has revealed the taxa possessing genomes enriched by BGCs. The analysis of predicted efficiency of translation elongation has shown that while only 6.6% NRPS genes fall into the category of high predicted efficiency of translation, there is a number of BGCs distinguished by both large absolute number of genes with high predicted efficiency of translation and their percentage. These BGCs belong to such genera as *Pseudomonas*, *Staphylococcus*, *Corynebacterium*, *Streptomyces*, *Amycolatopsis*, *Paenibacillus*, *Rhodococcus* and *Burkholderia*.

Conclusion

The performed bioinformatic analysis has provided the information about distribution of nonribosomal peptide biosynthetic gene clusters in bacteria and their putative translation elongation efficiency. Thus, the analysis of translation elongation efficiency is useful as a high-throughput technique for gene ranking in bacterial genomes, which can be regarded as a rough estimate of expression level for various groups of genes of interest.

Acknowledgment

This study was funded by the Russian Foundation for Basic Research (Grant Nos. 17-00-00470 (K), 17-00-00462). The computational resources of the Joint HPC Facility 'Bioinformatics' was used with the support of the budget project No. 0324–2019-0040-C-01.

References

[1] V. A. Likhoshvai and Y. G. Matushkin, “Nucleotide composition-based prediction of gene expression efficacy,” *Mol. Biol.*, vol. 34, no. 3, pp. 406–412, 2000.

[2] S. Caboche, M. Pupin, V. Leclère, A. Fontaine, P. Jacques, and G. Kucherov, “NORINE: A database of nonribosomal peptides,” *Nucleic Acids Res.*, vol. 36, no. SUPPL. 1, pp. 326–331, 2008.

[3] R. D. Süssmuth and A. Mainz, “Nonribosomal Peptide Synthesis—Principles and Prospects,” *Angew. Chemie - Int. Ed.*, vol. 56, no. 14, pp. 3770–3821, 2017.

[4] V. Sokolov, B. Zuraev, S. Lashin, and Y. Matushkin, “Web application for automatic prediction of gene translation elongation efficiency.,” *J. Integr. Bioinform.*, vol. 12, no. 1, p. 256, 2015.

[5] K. Blin, M. H. Medema, R. Kottmann, S. Y. Lee, and T. Weber, “The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D555–D559, 2017.

DOI 10.18699/BGRS/SB-2020-021

Differentially expressed genes associated with TMPRSS2-ERG molecular subtype of prostate cancer

Anastasiya Andreevna Kobelyatskaya
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
kaa.chel@mail.ru

Elena Anatolevna Pudova
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
pudova_elena@inbox.ru

George Sergeevich Krasnov
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
gskrasnov@mail.ru

Anna Victorovna Kudryavtseva
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
rhizamoeba@mail.ru

Kirill Mikhailovich Nyushko
Urological department
FSBI NMRRC
Moscow, Russia
kirandja@yandex.ru

Boris Yakovlevich Alekseev
Urological department
FSBI NMRRC
Moscow, Russia
mnioi@mail.ru

Abstract — **Prostate cancer (PC) is one of the most common and socially significant oncological diseases in men. This study examined the transcriptome profile of the most common molecular genetic subtype of prostate cancer, TMPRSS2-ERG. As a result of bioinformatics analysis conducted on the basis of The Cancer Genome Atlas project data, 115 differentially expressed genes were identified for this study group, and in particular, the most over-expressing genes were identified: *ALOX15*, *CACNA1D*, *EML6*, *HLA-DMB*, *NKAIN1*, *OGDHL*, *PLA1A*, *SYT13*. Enrichment pathways analysis showed that these genes are participants in important oncologically significant pathways, which emphasizes the association of this molecular subtype with an unfavorable prognosis for prostate cancer.**

Keywords — TMPRSS2-ERG, TCGA, prostate cancer, sequencing

Introduction

Prostate cancer (PC) is one of the most common and socially significant oncological diseases in men. At the molecular genetic level, prostate cancer can be divided into seven subtypes, the most common among them being the subtype characterized by the presence of the chimeric transcript TMPRSS2-ERG (up to 45% of cases) [1, 2]. This transcript is the result of translocation between exon 1 of *TMPRSS2* gene and exon 4 of *ERG* gene [2, 3]. According to the results of numerous studies, the presence of the TMPRSS2-ERG transcript is most often considered an important predictor of unfavorable prognosis [1, 2, 4]. The aim of our study is to investigate the characteristics of the transcriptome profile of this molecular subtype, which can serve as a basis for understanding the mechanisms of progression in prostate cancer and help in the search for informative prognostic markers.

Methods

The study included PC samples RNA-Seq data of The Cancer Genome Atlas project. The cohort was divided into two groups: tumors with TMPRSS2-ERG fusion transcript (88 cases) and TMPRSS2-ERG-free tumors (117 cases). Samples have appropriate *ERG* expression signature. Differential expression analysis was performed in statistical environment R using the EdgeR package was used. The Mann-Whitney test, Exact test and quasi-likelihood method (QLF) were used for statistical analysis. To exclude false positive results FDR method was used. Enrichment pathways analysis was performed using clusterProfiler and ReactomePA packages. The results were considered statistically significant with a QLF p-value <0.05.

Results

We found out at list 115 differentially expressed genes between the studied groups (fig.1 – Differentially expressed genes TMPRSS2-ERG molecular subtype PC). Among those next genes are of most interest with fold-expression level more than 4 for TMPRSS2-ERG-positive group: *ALOX15*, *CACNA1D*, *EML6*, *HLA-DMB*, *NKAIN1*, *OGDHL*, *PLA1A*, *SYT13*. The following genes showing overexpression in TMPRSS2-ERG-positive group of samples are marked: *ALOX15*, *CACNA1D*, *EML6*, *HLA-DMB*, *NKAIN1*, *OGDHL*, *PLA1A*, *SYT13*. According to the results of enrichment pathways analysis, these genes are participants in the following cancer-significant pathways in the KEGG database associated with the progression: Arachidonic acid metabolism (hsa00590), Focal adhesion (hsa04510), Mucin type O-glycan biosynthesis (hsa00512), Notch signaling pathway (hsa04330), PI3K-Akt signaling pathway (hsa04151), Prostate cancer (hsa05215), Sphingolipid metabolism (hsa00600). Thus, the results underscore the potential association with an unfavorable prognosis in the group of samples belonging to the molecular subtype TMPRSS2-ERG.

Acknowledgment

This work was funded by the Russian Science Foundation grant no. 18-75-10127. This work was performed using the equipment of EIMB RAS “Genome” center (http://www.eimb.ru/ru1/ckp/ccu_genome_c.php).

References

[1] Arora K., Barbieri C.E. (2018) Molecular Subtypes of Prostate Cancer. Curr Oncol Rep. 20(8):58. doi: 10.1007/s11912-018-0707-9.

[2] Cancer Genome Atlas Research Network. (2015) The molecular taxonomy of primary prostate cancer. Cell. 163(4):1011-25. doi: 10.1016/j.cell.2015.10.02.

[3] Krumbholz M., Agaimy A., Stoehr R., Burger M., Wach S., Taubert H., Wullich B., Hartmann A., Metzler M. (2019) Molecular Composition of Genomic TMPRSS2-ERG Rearrangements in Prostate Cancer. Dis Markers. 2019:5085373. doi: 10.1155/2019/5085373.

[4] Wang Z., Wang Y., Zhang J., Hu Q., Zhi F., Zhang S., Mao D., Zhang Y., Liang H. (2017) Significance of the TMPRSS2:ERG gene fusion in prostate cancer. Mol Med Rep. 16(4):5450-5458. doi: 10.3892/mmr.2017.7281.

DOI 10.18699/BGRS/SB-2020-022

Constructing a pipeline for genome variant / gene functioning hybrid prioritization: a case study of type II diabetes

Irina Kolesnikova
LLC NCGI, Novosibirsk, Russia
i.kolesnikova@mygenetics.ru

Valery Polunovsky
LLC NCGI, Novosibirsk, Russia
valeriy.polunovskiy@mygenetics.ru

Konstantin Gunbin
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
genkvg@bionet.nsc.ru

Abstract — **According to recently proposed "omnigenic model" the genes implicated in complex diseases determination can be divided into core genes and peripheral genes. Mutations in core genes directly affect disease development, while mutations in peripheral genes can only indirectly modulate disease risk. In this study in order to discriminate core genes and their major regulators, we hierarchically combine genome variant prioritization with the prioritization of genes and target tissues.**

Keywords — hierarchical prioritization, genome variant, gene functioning, type II diabetes

Motivation and aim

One of the major challenge in the interpretation of GWAS results is the only modest or even very weak contribution of important loci to the total heritability of complex diseases. Earlier an "omnigenic model" proposed in which genes implicated in complex diseases determination divided into core genes and peripheral ones according to the gene network of trait [1]. It is of big importance that modern theoretical investigations uncover that biological reason of modest or weak implication of variations in core genes to the total heritability of complex traits is primary rooted in strict negative selection on core genes [2]. Therefore, it is tempting to analyze GWAS data taking into consideration modern views on the gene network basis of complex traits and on the evolutionary reasons of effect sizes distribution across variants uncovered by GWAS investigations.

Methods

In order to do so, we construct pipeline composed of hierarchically combination of prioritizations procedures starting from genome variant prioritization and ending by target tissues prioritization.

First, we focused on the most stable and most studied single nucleotide variants (SNVs) that are uniquely annotated starting from the NCBI dbSNP v. 150, stable SNVs (SSNVs). We performed a frequency and a functional filtering of the selected SSNVs. Frequency filtering of SSNVs was made using the ANNOVAR platform.

Functional filtering of SSNVs was the following: the effect of each SSNVs located in the coding regions or splicing sites must have been thorough tested on protein structure and function; the effect of each regulatory SSNVs must have the experimental and consistent confirmation of biological significance. All SSNVs should be characterized by an evolutionary and functional conservation, located inside (and not between) linkage disequilibrium blocks for the Asian and European populations, and have well-characterized allele frequencies in the public databases. In order to prioritize selected SSNVs in each trait under analysis, we have developed a consistency function that takes into account the above described characteristics types of genome variants and based on nonparametric statistics and randomization procedures.

Second, we prioritize genes containing SSNVs (or regulated by upstream regions containing SSNVs) according to (1) the characteristics of gene haploinsufficiency, (2) the gene networks related to the standardized vocabularies of phenotypic abnormalities from Human Phenotype Ontology (HPO) resource and vocabularies of medical diagnoses from International Classification of Diseases (ICD) v. 10, (3) the position of the gene in the topology of these gene networks, (4) the relation of the gene to the known thorough tested network modules of complex diseases, and (5) the pleiotropy of gene across various complex traits. We developed normalization function that reconcile the above-mentioned gene characteristics.

Third, if the determination of complex disease dispersed through various tissues and organs, we assigned a tissue score based on the position of genes in the network topologies and on the level of their tissue-specific expression.

In our work, we used GeneAtlas [3] as the main GWAS data source. We choose GeneAtlas because 1) it is the only resource that has largest number of SNVs ever (more than 34M of all SNVs including imputed ones); 2) it deposited GWAS data for more than 650 clinical and more than 110 normal human traits; 3) it is based on an unprecedented number (> 450,000) of experimentally genotyped, unrelated individuals. These unique GeneAtlas features allowed [3]: 1) to obtain the highest heritability of the studied traits (significant reduction of the "missing heritability" artifact), 2) to get rid of the Winner's Curse artifact overestimating the strength of SNV effect size on small samples.

Results

A hierarchical prioritization pipeline for searching and prioritizing genes and SNVs associated with a risk to diseases was developed. We check our pipeline on wellinvestigated type II diabetes. Significant SNPs were selected in more than 100 genes / regulatory regions strongly associated with type II diabetes, its complications and comorbid diseases (Table 1).

NOS3	6.752	42.753
MC4R	7.392	40.975
TERF2IP	3.069	39.417
NEUROG3	9.068	38.841
TUFM	10.113	37.669

The use of information on the gene networks structure and tissue-/organ- specificity of gene expression clarifies substantially our initial SSNVs prioritization, however, analysis of all GeneAtlas traits is significantly complicated by poor compliance between ICD and HPO vocabularies.

TABLE 1. THE TOP-30 LIST OF PRIORITIZED GENES ASSOCIATED WITH TYPE II DIABETES AND RELATED TO PANCREAS FUNCTIONING

Gene name	Normalised functioning score in pancreas	Composite score
MICB	13.457	195.911
EHMT2	5.167	180.766
KCNK3	7.713	146.456
AGT	6.367	119.92
APOE	8.236	101.094
ZFP36L2	12.935	90.303
CTRB1	14.31	75.996
MAP2K7	7.863	74.064
OPRL1	3.209	62.166
SPRY2	5.893	58.241
REL	7.016	50.483
HLA-DPB1	8.152	47.214
LDLR	10.696	45.945
HLA-C	13.271	45.123
TAP1	3.554	33.596
HLA-DPA1	9.605	32.237
LIPC	4.019	23.976
ATP2A1	5.899	23.589
HNF1A	7.84	23.197
DYNC2LI1	2.162	22.966
MFF	5.807	22.501
ANK1	6.117	21.774
SH3YL1	5.179	21.561
TCEA2	3.132	21.461
ACP5	3.637	20.822

References

[1] Liu X, Li YI, Pritchard JK. (2019) Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell. 177(4):1022–1034.e6.

[2] O'Connor LJ, et al. (2019) Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. Am J Hum Genet. 105(3):456–476.

[3] Canela-Xandri O, Rawlik K, Tenesa A. (2018) An atlas of genetic associations in UK Biobank. Nat Genet. 50(11):1593–1599.

DOI 10.18699/BGRS/SB-2020-024

Functional annotation of the transcription factors from *Methylotuvimicrobium alcaliphilum* 20Z^R

Semyon K. Kolmykov
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS
FRC Institute of Cytology and Genetics SB RAS Novosibirsk, Russia
kolmykovsk@gmail.com

Nikita V. Ivanisenko
FRC Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
n.ivanisenko@gmail.com

Ivan S. Evshin
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
ivan@developmentontheedge.com

Mikhail Kulyashov
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS Novosibirsk State University
Novosibirsk, Russia
m.kulyashov@mail.ru

Tamara M. Khlebodarova
FRC Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
tamara@bionet.nsc.ru

Ilya R. Akberdin
BIOSOFT.RU, LLC
FRC Institute of Cytology and Genetics SB RAS
Novosibirsk State University
Novosibirsk, Russia
akberdinir@gmail.com

***Abstract* — Methane is a promising carbon source for biosynthesis of biotechnologically useful compounds using aerobic methanotrophic bacteria as biocatalysts. Despite more than a century-long history of discovering and studying of methanotrophic microorganisms, knowledge of the molecular mechanisms of gene expression regulation by transcription factors in these bacteria is very limited with only a few isolated cases being published. Therefore, the identification of potential transcription factors for methanotrophic organisms and their target genes is not only a foreground fundamental problem in the research field of methanotrophy, but it is also especially relevant for the active development of biotechnological application of methane-oxidizing microorganisms. In this study a comparative genomics approach together with the structural modeling techniques were applied to reveal the TFs in the 20ZR genome and predict their target regulatory genes.**

Keywords — Transcription factors, haloalkaliphilic, aerobic, methanotrophs, bioinformatics

Introduction

Haloalkaliphilic aerobic methanotrophs, including *Methylotuvimicrobium alcaliphilum* 20Z^R strain (20ZR) are standing out as the promising microbial "factories" in industrial research as new sources of enzymes and protein based materials that are resistant to high salt and pH levels, and natural producers of amino acids, sugars, and osmoprotectants [1-3]. Moreover, the use of modern genetic engineering approaches, methods of genomics and system biology allows to significantly expand the scope of their application in the production of biofuels, biodegradable polymers and other valuable commercial products.

Functional annotation and search for new transcription factors (TFs) of 20ZR present enormous fundamental and applicable interest in understanding the life-cycle and biochemical pathways regulation. In the current work a comparative genomics approach together with the structural modeling techniques were applied to reveal the TFs in the 20ZR genome and predict their target regulatory genes.

Methods

Identification of potential TFs encoding in 20ZR genome

Prediction of TFs and DNA-binding proteins was carried out using tools that analyze the level of homology and similarity with known DNA-binding proteins and TFs. InterProScan package was applied to identify TFs and DNA-binding domains in the set of protein sequences [4]. To improve the quality of putative TFs identification, a list of HMM models of DNA-binding protein and TF domains for subsequent filtering of the InterProScan results was made based on information from the P2TF [5] and DBD [6] databases. Additionally to identify putative transcription factors in the set of proteins with unknown function DBD-Threader [7] and TFPredict [8] were used.

Identification of transcription factor target genes

To solve the problem of searching for TF target genes, an analysis of the scientific publications on transcriptional regulation of the genes expression of 20ZR and closely related groups of bacteria were carried out. The obtained set of regulated genes was expanded to include genes located in the same operons as the identified target genes. Information on the structure of 20ZR operons was obtained from the database of prokaryotic operons (Prokaryotic Operon DataBase; ProOpDB) [9].

In order to identify potential TF binding sites (TFBSs), promoter regions of 20ZR genes and their orthologs were analyzed using the RSAT software package [10] and frequency matrices of TF binding sites (PFM) from the PRODORIC2 [11] and RegulonDB [12] databases. Additionally a comparative analysis of putative 20ZR TFs and TFs of prokaryotes with known PFMs of TFs (DBs PRODORIC2 and RegulonDB) was conducted. Based on the obtained data on the level of homology, a consensus PFMs for the 20ZR TFBSs were constructed. The resulting consensus PFMs were used to search for potential TFBSs. The set of target genes under consideration will be expanded based on the obtained list of TFBSs.

To reconstruct 20ZR regulons based on the available genome-wide gene expression profiles (RNA-seq), correlation dependencies were analyzed and the corresponding clusters of coexpressed putative target genes and TFs were identified.

A structural annotation of TFs was carried out using homology modeling techniques, including MODELLER software [13]. Prediction and analysis of TFs interaction with target DNA fragments were analyzed using FoldX [14] and Rosetta molecular modeling software [15].

Results And Discussion

Functional and structural annotations of TFs of 20ZR were carried out for the first time. The comparative genomics approach together with molecular modeling techniques allowed to reveal the majority of TFs and assign its potential regulatory function.

The screening for DNA binding domains in the 20ZR proteome allowed us to predict the new putative TFs with unknown function. For these proteins the potential regulons in the 20ZR were predicted and ranked using *in silico* protein-dna binding energy estimations.

Obtained results provide essential data to reconstruct gene regulatory networks operating 20ZR metabolism and reveal their unique features for haloalkaliphilic aerobic methanotrophs.

REFERENCES

[1] Y.A., Trotsenko and V.N. Khmelenina, "The biology and osmoadaptation of haloalkaliphilic methanotrophs," Microbiology, vol. 71, № 2, pp.123-132, 2002.

[2] C.A. Henard and M.T. Guarnieri, "Metabolic Engineering of Methanotrophic Bacteria for Industrial Biomanufacturing," in Methane Biocatalysis: Paving the Way to Sustainability, Springer, Cham, 2018, pp. 117-132.

[3] S. Nariya and M.G. Kalyuzhnaya, "Diversity, Physiology, and Biotechnological Potential of Halo (alkali) philic Methane-Consuming Bacteria," in Methanotrophs, Springer, Cham, 2019, pp. 139-161.

[4] P. Jones, D. Binns, H.Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, and S. Pesseat, "InterProScan 5: genome-scale protein function classification," Bioinformatics, vol. 30, №9, pp.1236-1240, 2014.

[5] P. Ortet, G. De Luca, D.E. Whitworth, and M. Barakat, "P2TF: a comprehensive resource for analysis of prokaryotic transcription factors," BMC genomics, vol. 13, № 1, p. 628, 2012.

[6] D. Wilson, V. Charoensawan, S.K. Kummerfeld, and S.A. Teichmann, "DBD—taxonomically broad transcription factor predictions: new content and functionality," Nucleic acids research, vol. 36, suppl_1, pp.D88-D92, 2007.

[7] M. Gao, and J. Skolnick, "A threading-based method for the prediction of DNA-binding proteins with application to the human genome," PLoS computational biology, vol. 5, № 11, p.e1000567, 2009.

[8] J. Eichner, F. Topf, A. Dräger, C. Wrzodek, D. Wanke, and A. Zell, "TFpredict and SABINE: sequence-based prediction of structural and functional characteristics of transcription factors," PloS one, vol. 8, № 12, p.e82238, 2013.

[9] B. Taboada, R. Ciria, C.E. Martinez-Guerrero, and E. Merino, "ProOpDB: Prokaryotic Operon DataBase," Nucleic acids research, vol. 40, № D1, pp.D627-D631, 2011.

[10] N.T.T. Nguyen, B. Contreras-Moreira, J.A. Castro-Mondragon, W. Santana-Garcia, R. Ossio, C.D. Robles-Espinoza, M. Bahin, S. Collombet, P. Vincens, D. Thieffry, and J. van Helden, "RSAT 2018: regulatory sequence analysis tools 20th anniversary," Nucleic acids research, vol. 46, № W1, pp.W209-W214, 2018.

[11] D. Eckweiler, C.A. Dudek, J. Hartlich, D. Brötje, and D. Jahn, "PRODORIC2: the bacterial gene regulation database in 2018," Nucleic acids research, vol. 46, № D1, pp.D320-D326б 2017.

[12] A. Santos-Zavaleta, H. Salgado, S. Gama-Castro, M. Sánchez-Pérez, L. Gómez-Romero, D. Ledezma-Tejeida, J.S. García-Sotelo, K. Alquicira-Hernández, L.J. Muñiz-Rascado, P. Peña-Loredo, and C. Ishida-Gutiérrez, "RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12," Nucleic acids research, vol. 47, № D1, pp.D212-D220, 2018.

[13] N. Eswar, D. Eramian, B. Webb, M.Y. Shen and A. Sali, "Protein structure modeling with MODELLER," in Structural proteomicsm, Humana Press, 2008, pp. 145-159.

[14] A.D. Nadra, L. Serrano and A. Alibés, "DNA-binding specificity prediction with FoldX," in Methods in enzymology, vol. 498, Academic Press, 2011, pp. 3-18.

[15] S.J. Fleishman, A. Leaver-Fay, J.E. Corn, E.M. Strauch, S.D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon and J. Meiler, "RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite," PloS One, vol. 6, № 6, p.e20161, 2011

DOI 10.18699/BGRS/SB-2020-025

GTRD – an integrated view on transcription regulation

Fedor A. Kolpakov
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
fkolpakov@gmail.com

Semyon K. Kolmykov
FRC Institute of Cytology and Genetics SB RAS
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
kolmykovsk@gmail.com

Mikhail A. Kulyashov
Novosibirsk State University
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
m.kulyashov@mail.ru

Ivan S. Evshin
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
ivan@biosoft.ru

Yury V. Kondrakhin
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
yvkondrat@mail.ru

Ruslan N.Sharipov
Novosibirsk State University Institute of Computational BIOSOFT.RU, LLC
Novosibirsk, Russia
shrus79@biosoft.ru

***Abstract* — GTRD - Gene Transcription Regulation Database (http://gtrd.biouml.org/) contains uniformly annotated and processed NGS data (ChIP-seq, ChIP-exo, DNase-seq, MNase-seq, ATAC-seq and RNA-seq) related with transcription regulation. All cell types (cell lines and tissues) presented in the GTRD were arranged into a dictionary, linked with specialised ontologies (BRENDA, Cell ontology, Uberon, Celosaurus, Experimental factor ontology) and further linked with experiments in specialised databases on transcription regulation: FANTOM 5, ENCODE and GTEx. For integrated view in transcription regulation we have developed web-interface that provides browsing, advanced search possibilities, an integrated genome browser and integrated reports by given cell type, transcription factor or gene.**

Keywords* — *GTRD, transcription regulation, NGS, uniformed analysis, data integration

Introduction

Transcription regulation is a complex process that involves a lot of participants. There is a lot of experimental data of different types obtained for different cell types and different experimental conditions which are stored in different databases.

An important task is integration of such data for a given cell type and experimental condition to get an integrated view on main events related with transcription regulation. This is one of the main goals of GTRD database [1].

Results

The suggested approach is shown on figure 1:

1) we are collecting and uniformly process[1] following NGS data:

- ChIP, Chip-exo - to identify binding regions for transcription factors (TFs) and their coactivators, as well as to identify regions with corresponding histone marks.

- DNase-seq, MNase-seq and ATAC-seq - to identify genome regions with open chromatin and footprints for TF binding sites;

- RNA-seq data where gene expression was changed by TFs activation or knockout to identify genes regulated by these TFs.

2) all cell types (cell lines and tissues) presented in the GTRD were arranged into a dictionary, linked with specialised ontologies (BRENDA, Cell ontology, Uberon, Celosaurus, Experimental factor ontology) and further linked with experiments in specialised databases on transcription regulation: FANTOM 5, ENCODE and GTEx [2].

3) for integrated view in transcription regulation we have developed web-interface:

- based on BioUML platform[3] - it provides browsing and displaying information, advanced search possibilities and an integrated genome browser;

- based on BeanExplorer technology[4] - it provides integrated reports by given cell type, TF or gene.

ACKNOWLEDGMENT

This study was supported by the Russian Science Foundation, grant No. 19-14-00295.

REFERENCES

[1] Yevshin I, Sharipov R, Kolmykov S, Kondrakhin Y, Kolpakov F. GTRD: a database on gene transcription regulation - 2019 update. Nucleic acids research. 2019 Jan 8;47(D1):D100-5.

[2] Kulyashov M., Kolmykov S., Yevshin I., Kolpakov F. “Description, Characteristic And Algorithm For Creation Of A Dictionary Of Cell Types And Tissues In The Gtrd Database”, CEUR-Workshop proceedings, 2020, Vol-2569, p.13-18.

[3] Kolpakov F, Akberdin I, Kashapov T, Kiselev L, Kolmykov S, Kondrakhin Y, Kutumova E, Mandrik N, Pintus S, Ryabova A, Sharipov R. BioUML: an integrated environment for systems biology and collaborative analysis of biomedical data. Nucleic acids research. 2019 Jul 2;47(W1):W225-33.

[4] Bean Explorer https://github.com/DevelopmentOnTheEdge/be5

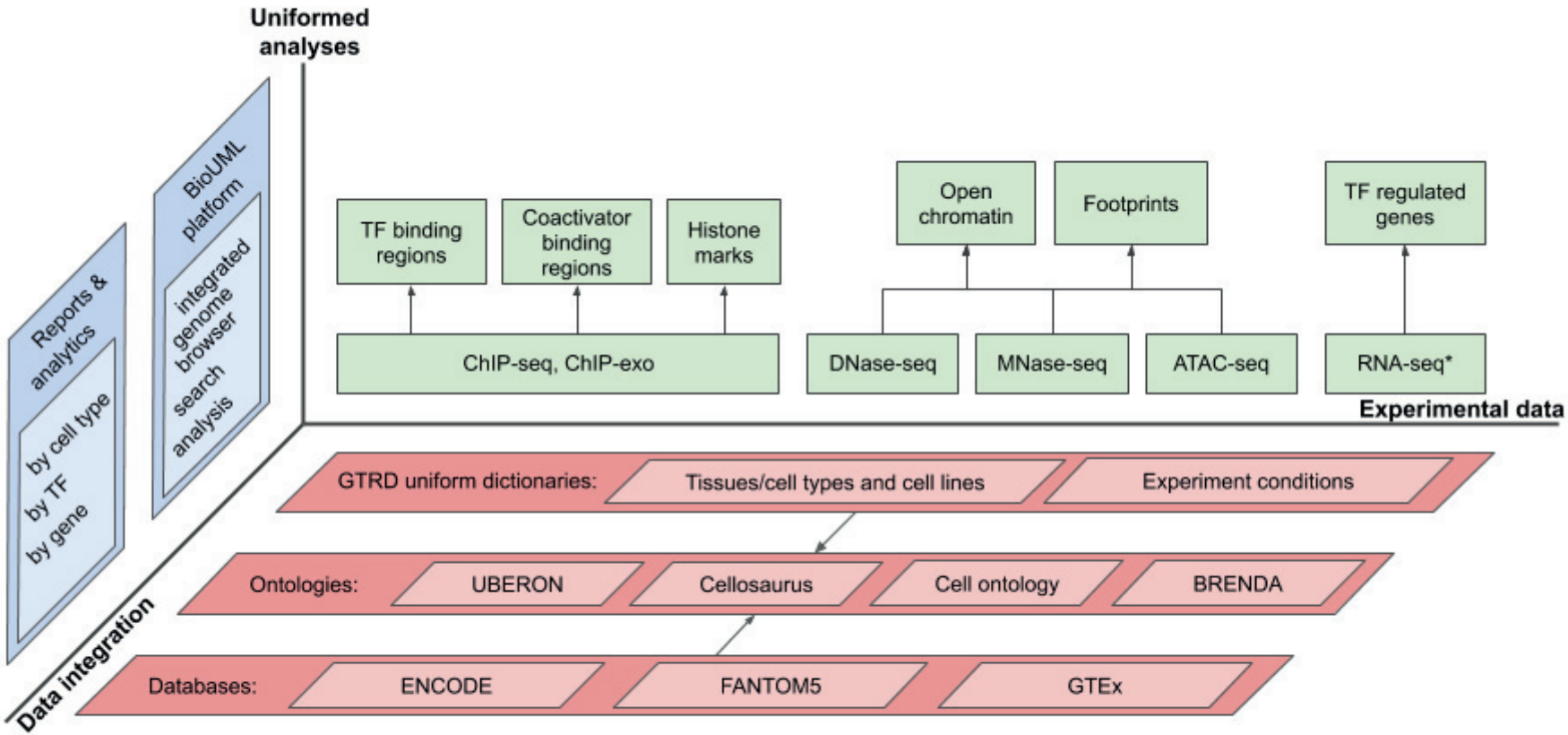


Fig. 1. GTRD - an integrated view on transcription regulation. * RNA-seq data for different TFs activation or knockout.

DOI 10.18699/BGRS/SB-2020-026

Detection of alphacoronavirus in bat fecal samples from Volgograd region

E.V. Korneenko
Central Research Institute of Epidemiology, Moscow, Russia, lennatta@yandex.ru

A.E. Samoilov
Central Research Institute of Epidemiology, Moscow, Russia
andrei.samoilov@gmail.com

I.V.Artyushin
Lomonosov Moscow State University, Moscow, Russia, sometyx@gmail.com

A.V. Dudorova
Lomonosov Moscow State University, Moscow, Russia

V.G. Dedkov
Pasteur Institute, Saint-Petersburg, Saint-Petersburg, Russia

A.D. Matsvay
FSBI "Center of Strategic Planning" of the Ministry of Health, Moscow, Russia

E.V. Pimkina
Central Research Institute of Epidemiology, Moscow, Russia

M.V. Safonova
Plague Control Center, Federal Service on Consumers' Rights Protection and Human Well-Being Surveillance, Moscow, Russia

A.S. Speranskaya
Central Research Institute of Epidemiology, Moscow, Russia, Lomonosov Moscow State University, Moscow Russia, hanna.s.939@gmail.com

Abstract — Bats are natural reservoirs of many emerging viruses and their virome can differ according to the regions. In this study we have found a genome fragment of Coronaviridae which was 99% percent identical to Alphacoronavirus N.las/C/Spain/2007 in bat fecal sample from Volgograd region

Keywords — bats, viruses, coronavirus, high throughput sequencing

Motivation and aim

Bats are natural reservoirs of many emerging viruses and their virome can differ according to the regions. The aim of this study was to analyse virome of bat fecal samples which were collected in Volgograd region (Russian Federation).

Materials and methods

Fecal samples from bat species (*Nyctalus noctula, Vespertilio murinus, Nyctalus lesleri, Myotis daubentonii, Eptesicus serotinus, Pipistrellus kuhlii*) were collected in 2016 in Volgograd regions (Yaminsky farm, Zimnyatskiy farm) and transported to laboratory of Moscow State University (MSU) in RNAlater solution. Notably, that transportation took some days and no special conditions were applied. Then, extraction using RNeasy Lipid Tissue Mini Kit (Qiaqen, Germany), reverse transcription using REVERTA-L (AmpliSens, Russia), PCR amplification on primers described in previous study [1] and library preparation was performed using Illumina TruSeq (in house) protocol. Libraries were sequenced using Illumina MiSeq platform. High throughput sequencing data were assessed with FastQC, filtered by trimming using trimmomatic 0.38, assembled by SPAdes 3.13, taxonomy of resulting contigs was identified by the BLASTn.

Results and discussion

For 26 samples we performed library preparation and sequencing on Illumina Miseq Platform. In one sample we have found a genome fragment of Coronaviridae which was 99% percent identical to Alphacoronavirus N.las/C/Spain/2007 accession number -HQ184051.1. Interestingly, this virus was found in a greater noctule bat (Nyctalus lasiopterus) in Spain in 2007 [2]. According to taxonomy Alphacoronavirus N.las/C/Spain/2007 does not relates to the aetiological agents of severe acute respiratory syndrome (SARS), which was identified as a 2b β-CoV, Middle East respiratory syndrome-CoV (MERS-CoV) is as 2c β-CoV or coronavirus disease 2019 (COVID-19) SARS-CoV-2. This finding can present interest for the further study of virus migration in bats in different regions of Russia Federation.

References

[1] Virome analysis for identification of viruses in bat species from Moscow region, Speranskaya A.S., Pimkina E.V., Artyushin I.V., Shipulin G.A., Deviatkin A.A., Kuleshov K.V., Dedkov V.G., Safonova M.V., Bioinformatics of Genome Regulation and Structure/BGRS-2016

[2] Mol. Biol. Evol., 28 (2011), pp. 2731-2739

[3] Lelli, D.; Papetti, A.; Sabelli, C.; Rosti, E.; Moreno, A.; Boniotti, M.B. Detection of Coronaviruses in Bats of Various Species in Italy.Viruses 2013, 5, 2679-2689

DOI 10.18699/BGRS/SB-2020-027

EPhIMM: computational workflow for fast phylogenetic inference based on multiple alignment of prokaryotic single-copy marker genes

Aleksei Korzhenkov
NRC Kurchatov Institute, Moscow, Russia
oscypek@ya.ru

***Abstract* — EPhIMM — a computational workflow for fast and accurate phylogenetic inference for prokaryotic genomes was developed and tested on bacterial and archaeal genomic datasets. Source code is freely available at https://github.com/laxeye/EPhIMM.**

Keywords — genomics, phylogenetics, workflow

Motivation and aim

Motivation

Multiple sequence alignment (MSA) and phylogenetic inference are one of the most time- and resource-consuming operations in bioinformatical analysis. The growing number of complete and draft prokaryotic genomes enforces search of an effective computational workflow for phylogenetic analysis.

Aim

The aim of the study was to select and test a combination of bioinformatic tools for sequence search, alignment and phylogenetic inference, each of them could be used on a common PC or laptop and create a workflow for user-friendly phylogenetic analysis.

Methods

41 protein coding prokaryotic single copy marker genes that have HMM profiles were used from CheckM software [1]. CDS were predicted using prodigal [2] with default settings whether genomic sequences were used. CDS for marker genes were found using hmmsearch [3] with default setting except e-value = 1e-10 against corresponding HMM profiles from pfam [4]. Sequences were extracted using esl-sfetch tool from easel software package (https://github.com/EddyRivasLab/easel). When several protein coding sequences from one genome matched same HMM profile a sequence with lowest mean distance to all other sequences was selected using BioKotlin (https://github.com/laxeye/BioKotlin).

Three MSA tools were tested with default settings: Clustal Omega [5], Muscle [6] and MAFFT [7]. Columns, containing more than 50% of gaps were removed from the resulting alignments using BioKotlin.

Three tools for phylogenetic tree inference were tested: RAxML [8], PhyML [9] and FastTree claimed to be several magnitudes faster than RAxML and PhyML [10]. LG substitution matrix was used for all tools. CAT approximation with 25 and 20 rate categories for RaxML and FastTree respectively was used. Phylogenetic trees were visualized and topologies of the trees were compared using ete3 [11].

Bash script was written to manage transition of user data from one tool to another.

Three sets of high quality complete genomes (one genome per species) of type strains were analyzed: *Bacillaceae,* Euryarchaeota and *Gammaproteobacteria*. Genomes of *Staphylococcus aureus subsp. aureus* DSM 20231, *Sulfolobus acidocaldarius* DSM 639 and *Campylobacter fetus subsp. fetus* NCTC 10842 were used as outgroups respectively. Totally 39, 117 and 332 genomes were in the first, second and third set respectively. Predicted protein coding sequences were downloaded from NCBI RefSeq Release 95.

Results

A PC with 8 GB DDR3 and Intel Core i7-4790 3.60GHz with Ubuntu 18.04 operating system was used to test the worflow. RAM usage appeared to be the main restriction: PhyMl required more than 8 GB on *Gammaproteobacteria* dataset, while Clustal Omega was unable to perform MSA on all datasets, for data on time consumption see Table 1.

Phylogenetic analysis of 117 genomes on reference PC took roughly 10 minutes resulting in phylogenetic tree with mean support value of nodes boo.

Result of the work is EPhiMM – a workflow including tools for protein prediction (prodigal), marker gene search and extraction (hmmer and easel), multiple sequence alignment (MAFFT) and phylogenetic tree inference (FastTree) coupled with helper tool (BioKotlin). EPhIMM may use nucleotide genomic sequences or sets of predicted proteins. The workflow depends on mentioned tools and Java virtual machine 8 which could be easily installed using Bioconda. Source code and documentation are freely available at https://github.com/laxeye/EPhIMM under MIT license.

Table 1. Phylogenetic inference based on aminoacid sequences

Genomic set	*Bacillaceae*	Euryarchaeota	*Gammapro-teobacteria*
Genome count	39	117	332
Extraction of marker genes, m:s	0:31.2	1:27.9	4:56.5
Sequence alignment (MAFFT FFT-NS-2), m:s	0:53.8	5:55.1	15:47.2
Sequence alignment (MUSCLE), m:s	4:51.4	20:05.9	70:49.3

Alignment length, aa	30452	39007	63616
Proportion of gaps in alignment, %	49.32	66.56	73.36
Phylogenetic tree inference (FastTree), m:s	0:53.1	2:58.6	10:07.6
Phylogenetic tree inference (RAxML), h:m:s	0:26:12	3:8:16	19:21:38
Phylogenetic tree inference (PhyML), m:s	14:51.68	123:28.4	N/A

ACKNOWLEDGMENT

Supported by a grant of Ministry of Science and Higher Education of Russian Federation allocated to the Kurchatov Center for Genome Research (075-15-2019-1659).

REFERENCES

[1] Parks, D. H. et al. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome research, 25(7), 1043-1055.

[2] Hyatt, D. et al. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics, 11(1), 119.

[3] Eddy, S. R. (2011). Accelerated profile HMM searches. PLoS computational biology, 7(10), e1002195.

[4] Finn, R. D. et al. (2015). The Pfam protein families database: towards a more sustainable future. Nucleic acids research, 44(D1), D279-D285.

[5] Sievers, F. et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology, 7(1).

[6] Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 32(5), 1792-1797.

[7] Katoh, K., Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772-780.

[8] Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312-1313.

[9] Guindon, S. et al. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology, 59(3), 307-321.

[10] Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 - Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE, 5(3):e9490.

[11] Huerta-Cepas, J., Serra, F., Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Molecular biology and evolution, 33(6), 1635-1638.

DOI 10.18699/BGRS/SB-2020-028

RTrans: a pipeline for multi-way analysis of differential gene expression profiles

George Sergeevich Krasnov
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
gskrasnov@mail.ru

Anastasiya Andreevna Kobelyatskaya
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
kaa.chel@mail.ru

Anastasiya Vladimirovna Snezhkina
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
leftger@rambler.ru

Vladislav Sergeevich Pavlov
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
vladislav1pavlov@gmail.com

Elena Anatolevna Pudova
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
pudova_elena@inbox.ru

Anna Victorovna Kudryavtseva
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
rhizamoeba@mail.ru

Abstract — **RNA-Seq is a widespread technique routinely used to reveal differentially expressed (DE) genes, splicing events and affected signaling pathways. Usually, RNA-Seq data analysis needs skills in programming and bioinformatics. We aimed at developing easy-to-use pipeline covering multiple aspects of DE analysis. Here we present RTrans pipeline aimed at evaluating differential gene expression, pathway enrichment, WGCNA co-expression analysis and visualization. It provides rapid analysis of read counts data including MANOVA, paired tests, non-parametric tests. RTrans represents a versatile pipeline facilitating DE analysis based on RNA-Seq and other quantitative data.**

Keywords — RNA-Seq, pipeline, differential expression analysis

Introduction

RNA-Seq is a widespread technique routinely used to reveal differentially expressed (DE) genes, splicing events and affected signaling pathways. Usually, RNA-Seq data analysis needs skills in programming and bioinformatics. Last years, several pipelines facilitating transcriptome analysis have been developed (DEBrowser, UTAP, iDEP, Chipster, RNACocktail, DEapp) but there is still a need for tools that would be convenient in daily practice [1-6].

Hence, we aimed to develop RNA-Seq data processing pipeline that would satisfy several requirements:

- include various statistical test, both parametric and non-parametric, and multifactor ANOVA;
- be convenient to use in everyday practice, require minimal user attention, not require filling up numerous forms at each step of the analysis;
- quickly make several intergroup comparisons or ANOVA analyses in one run;
- include pathway enrichment analyses across several databases, and WGCNA;
- create interactive and flexible reports, including heatmaps, MDS plots, enrichment maps, etc.

Methods

We present RTrans, a pipeline for analyzing gene expression data (read counts) that aims to automate the analysis, minimize the needs for user attention, and parallelize the analyses. Before RTrans launch, users only need to fill out Excel workbook describing experimental design and provide read counts data. RTrans is written in R and Python. It is based on several Bioconductor packages: edgeR and DESeq (differential expression); clusterProfiler, topGO, ReactomePA (enrichment analyses); pathview, ggplot2, pheatmap, etc. (visualization). It includes several tests for differential expression analysis: quasi-likelihood F-test, likelihood ratio test, exact Fisher's test for GLM (edgeR); non-parametric Mann-Whitney, Wilcoxon, Kruskal-Wallis test; Student's test; Spearman and Pearson correlations; comparison of distribution densities. RTrans also considers samples pairing (e.g tumor-normal). Additionally, RTrans analyzes dependencies between gene expression and transcript lengths; it stratifies genes into several bins depending on transcript lengths and then normalize read count within each bin. This procedure partially eliminates the effect of 3'-bias coming from RIN imbalance. RTrans performs enrichment tests on GO, KEGG, Reactome, WikiPathways, Disease Ontology, DisGeNET, Network of Cancer Genes databases. It creates GO/KEGG-centric DE expression plots, heatmaps, MDS plots, Venn diagrams, enrichment maps, and other reports. Most of the reports generated by RTrans are formatted Excel workbooks.

Results

The developed pipeline aims to automate the analysis, minimize the needs for user attention, and parallelize the analyses. Before RTrans launch, users only need to fill out Excel workbook describing experimental design and provide read counts data. RTrans produces Excel-formatted differential gene expression reports with embedded sparklines/heatmaps illustrating gene expression distribution, pathway enrichment reports (Gene Ontology, KEGG, Reactome, WikiPathways, Disease Ontology, DisGeNET, Network of Cancer Genes databases), heatmaps, MDS plots, Venn diagrams, enrichment maps, GO/KEGG-centric DE expression plots, and other reports. RTrans represents a versatile pipeline facilitating DE analysis based on RNA-Seq and other quantitative data. The pipeline is available at https://github.com/gskrasnov/RTrans

ACKNOWLEDGMENT

This work was funded by the Russian Science Foundation grant no. 18-75-10127. This work was performed using the equipment of EIMB RAS "Genome" center (http://www.eimb.ru/ru1/ckp/ccu_genome_c.php).

REFERENCES

[1] Ge SX, Son EW, Yao R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. BMC Bioinformatics. 2018 Dec 19;19(1):534. doi: 10.1186/s12859-018-2486-6.

[2] Kallio MA, Tuimala JT, Hupponen T, Klemelä P, Gentile M, Scheinin I, Koski M, Käki J, Korpelainen EI. Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics. 2011 Oct 14;12:507. doi: 10.1186/1471-2164-12-507.

[3] Kohen R, Barlev J, Hornung G, Stelzer G, Feldmesser E, Kogan K, Safran M, Leshkowitz D. UTAP: User-friendly Transcriptome Analysis Pipeline. BMC Bioinformatics. 2019 Mar 25;20(1):154. doi: 10.1186/s12859-019-2728-2.

[4] Kucukural A, Yukselen O, Ozata DM, Moore MJ, Garber M. DEBrowser: interactive differential expression analysis and visualization tool for count data. BMC Genomics. 2019 Jan 5;20(1):6. doi: 10.1186/s12864-018-5362-x.

[5] Li Y, Andrade J. DEApp: an interactive web interface for differential expression analysis of next generation sequence data. Source Code Biol Med. 2017 Feb 3;12:2. doi: 10.1186/s13029-017-0063-4. eCollection 2017.

[6] Sahraeian SME, Mohiyuddin M, Sebra R, Tilgner H, Afshar PT, Au KF, Bani Asadi N, Gerstein MB, Wong WH, Snyder MP, Schadt E, Lam HYK. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. Nat Commun. 2017 Jul 5;8(1):59. doi: 10.1038/s41467-017-00050-4.

DOI 10.18699/BGRS/SB-2020-029

Advanced data curation in GTRD database: hierarchical dictionaries of cell types and experimental factors

Mikhail A. Kulyashov
BIOSOFT.RU, LLC
State University
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
m.kulyashov@mail.ru

Semyon K. Kolmykov
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
kolmykovsk@gmail.com

Fedor A. Kolpakov
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
fkolpakov@gmail.com

Ivan S. Yevshin
BIOSOFT.RU, LLC

***Abstract* — GEO database contains a lot of different experiments about transcription regulation. Most part of such experiments (ChIP-seq, DNase-seq and Histone-ChIP-seq) for 9 species (*Homo sapiens, Mus musculus, Rattus norvegicus, Danio rerio, Caenorhabditis elegans, Drosophila melanogaster, Saccharomyces cerevisiae, Schizosaccharomyces pombe and Arabidopsis thaliana*) were annotated and uniformly processed in GTRD database. Here we are describing the latest advances in these data annotations for GTRD database to formalise experimental data: hierarchical dictionaries of cell types and experimental factors. This approach helps us to integrate experimental data by cell lines (cell types and tissues) and experimental conditions as well as simplifies searching for relevant information for a user.**

Keywords* — *GTRD, database structure, experiments annotation, experiments comparison

Introduction

GTRD database contains a large number of ChIP-seq, DNase-seq and Histone-ChIP-seq experiments which were manually annotated [1].

Chip-seq experiment shows binding regions for transcription factors. DNase-seq experiment gives an opportunity to identify regions of open chromatin. Histone ChIP-seq shows a presence of different types of histones modifications, such as methylation, acetylation and e.t.c. It should be noted that results of such experiments are specific for cell type and experimental conditions (different treatments, conditions of experiment, characteristics of donors such as sex or age and many other).

An important task is integration of such data for a given cell type and experimental condition to get an integrated view on main events related with transcription regulation. This is one of the main goals of GTRD database.

To solve this task we are building a hierarchical dictionary which contains all cell lines, tissues and cell types which were described in GTRD as well as a hierarchical dictionary of different experimental conditions. Using this approach we also can compare experiments on the same cell lines, cell types or tissues in different databases. This is further development of our previous work [2].

Materials and Methods

Dictionary of cell types and tissues:

All cell lines, tissues and cell types which were used in experiments, described in GTRD, contain 2 fields that were used for building the dictionary [2]:

Source - tissues or cell types from which cells were isolated in experiments or system, which includes this cell, if isolation region was not indicated;

Cell type - cells on which experiment was made.

First step in process of creation a hierarchical dictionary was to separate sources to *main clusters* - a global groups which unify different sources, souch organs, embryonic parts, cell lines and some other in union groups:

Cell line - this group includes cell lines, which (permanently established cultures) like MCF-7 cell line, or cell types, which were derived from cell lines, for example WA09 derived fibroblasts.

Embryo - are all cell types, which were described as whole embryos and this group includes a lower cluster embryonic cells in which presented all cell types and tissues which were derived from embryos, like mesendoderm, embryonic organs e.t.c.

Stem cell - stem cell of different stages (totipotent, pluripotent, multipotent, e.t.c.), and all cell types which were derived from stem cells, for example embryonic stem cells derived epithelial cells.

Multicellular organism - this cluster includes all tissues and cell types in which were used whole multicellular

organism and all smaller clusters which describe different tissues of the organism, for example brain, spleen, bone-marrow and many others.

Whole organism - whole unicellular organisms, which are included in GTRD database, in different strains and developmental stages.

Other - this cluster includes all cell types which now are not classified to any cluster.

Dictionary of experiments description:

Dictionary of *experimental factors (ExpFactors)* includes all factors which can influence an experiment. This dictionary is creating simultaneously with the process of manual annotation and all these factors are adding by annotator.

Similar with cell types we also can group experimental factors into clusters like treatment, conditions, age e.t.c., main of them were described in our previous work [2]. They are the starting point of corresponding classification trees. Then we can manually specify child-parent relationships, for example, "doxycycline" will be parent for "doxycycline for 45 mins".

Database structure:

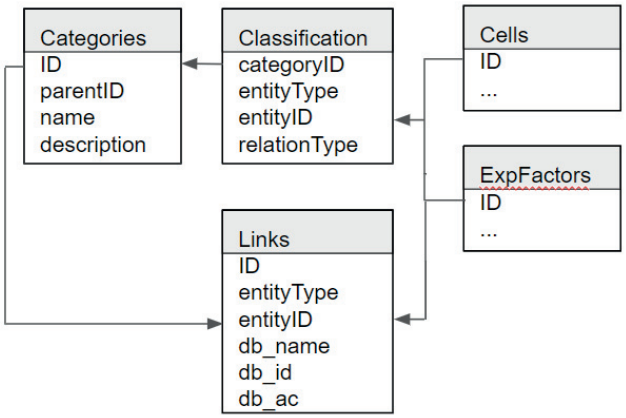


Fig. 1. Fragment of GTRD database structure for hierarchical classification of cell types and experiment conditions

We extended GTRD database scheme for hierarchical classification of data presented in tables Cells (cell types) and ExpFactors (experimental factors):

Categories - describes classification tree where the clusters described above are the roots (parentID is null).

Classification - links flat dictionary of cell types and experimental factors with categories tree.

Links - links Categories, Cells and ExpFactors with specialized ontologies: and linked it with specialized ontologies: BRENDA [3], Cell ontology [4], Uberon [5], Celosaurus [6], Experimental factor ontology [7].

Tables described above were filled semi-automatically. Part of information was extracted from mentioned above specialised ontologies which were downloaded in OBO format and translated into relational database. Then this information was checked and supplemented by annotator using specialized GTRD tools.

Results

Hierarchical dictionaries for all cell types and experimental conditions for annotated in GTRD database were developed. This approach helps us to integrate experimental data by cell types and experimental conditions as well as simplifies searching for relevant information for a user.

ACKNOWLEDGMENT

This work was supported by the Russian Science Foundation. (grant No. 19-14-00295).

REFERENCES

[1] Yevshin I., Sharipov. R, Kolmykov S., Kondrakhin Y., Kolpakov F. “GTRD: a database on gene transcription regulation—2019 update”, Nucleic Acids Research. 2019, Volume 47, Issue D1, Pages D100–D105.

[2] Kulyashov M., Kolmykov S., Yevshin I., Kolpakov F. “Description, Characteristic And Algorithm For Creation Of A Dictionary Of Cell Types And Tissues In The Gtrd Database”, CEUR-Workshop proceedings, 2020, Vol-2569, p.13-18.

[3] Gremse M, Chang A, Schomburg I, et al. “The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources”, Nucleic Acids Res. 2011, D507–D513.

[4] Bard J., Rhee S.Y., Ashburner M. “An ontology for cell types”, Genome Biol. 2005;6(2):R21.

[5] Mungall C.J., Torniai C., Gkoutos G.V. et al. “Uberon, an integrative multi-species anatomy ontology”, Genome Biol. 2012, 13, R5.

[6] Bairoch A. “The Cellosaurus, a Cell-Line Knowledge Resource”, J Biomol Tech. 2018, 29(2):page 25–38.

[7] Malone J., Holloway E., Adamusiak T., et al. “Modeling Sample Variables with an Experimental Factor Ontology”, Bioinformatics. 2010, 26(8):1112-1118.

[8] Kolpakov F., Akberdin I., Kashapov T., Kiselev I., Kolmykov S., Kondrakhin Y. et al. “BioUML: an integrated environment for systems biology and collaborative analysis of biomedical data”, Nucleic Acids Research, Volume 47, Issue W1, 02 July 2019, Pages W225–W233

DOI 10.18699/BGRS/SB-2020-030

The limits of the additive model for adult height

Ivan A. Kuznetsov
Center of Life Sciences
Skolkovo Institute of Science and Technology
Moscow, Russia
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia
Ivan.Kuznetsov@skoltech.ru

Georgii A. Bazykin
Center of Life Sciences
Skolkovo Institute of Science and Technology
Moscow, Russia

Sergei A. Slavskii
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Tatiana I. Axenovich
Laboratory of recombination and segregation analysis
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Tatiana I. Shashkova
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Fyodor A. Kondrashov
Institute of Science and Technology
Vienna, Austria

Yurii S. Aulchenko
Laboratory of recombination and segregation analysis
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Abstract — **The classical approach to analysis of polygenic quantitative traits assumes use of normal approximation and the additivity of effects. For more than a century, adult height served as an exemplary trait justifying this type of analysis and serving as a test case. We demonstrate that the classical approach to analysis of height has met its limits in contemporary large populations. In particular, we demonstrate the existence of weak, but highly significant non-additive interactions of genetic and environmental factors. In the conventional model, the better fit to the data that was achieved by accounting for these interactions came at the expense of the mean-model's increased complexity. The complexity of the model could be reduced if log-normal approximation was used.**

Keywords — height, quantitative trait, distribution, scale effects

Introduction

Adult height inspired the first biometrical and quantitative genetic studies [1,2] and is a test case trait for understanding the genetic component of heritability [3–5]. In statistical textbooks and in human genetics height serves as an example of a trait that is shaped by a process whereby the independent contributions of many genetic and environmental factors are summed up, resulting in a normal distribution. At the same time, results from anthropological and socioeconomic studies of distinct populations suggest that human height may be distributed log-normally, which could imply that genetic and environmental factors affecting human height multiply. The aim of this study was to compare two height distribution approximations: normal and log-normal, using a large sample size.

Methods

In the first stage we used data on average male and female height in various populations in the world. Then we analysed adult height in 369,153 white British participants from the UK Biobank study [6]. The participants belonged to groups defined by place of birth and self-reported ethnicity. We considered sex, genotype, and residual (combining socioeconomic and study covariates) effects. The genotype was included in analysis as a polygenic height score, defined as the weighted prevalence of height-increasing alleles in the genotype according to study of Wood et al [5]. The three predictors were strongly associated with height in a linear model with sex explaining 50.8%, polygenic score explaining 5.8% and residual predictor explaining 4.5% of height variance (all p-values $< 10^{-100}$).

Results

We found that the standard deviation, but not the coefficient of variation, of mean height in populations increases with mean height. Also, we revealed that the linear coefficient between mean female and male height significantly distinguishes from 1, which is consistent with the multiplicative effect of sex. On individual-level data, we showed an increase of the effect of sex, polygenic score and environmental predictor with an increase of mean group height, the standard deviation in these groups also increased. Besides this, significant statistical interactions between sex, polygenic score, and environmental predictor were revealed in the linear model for height. Similar analyses for the logarithm of height prediction showed no evidence for any effect of studied predictors increases with mean height or interactions.

Conclusion

Thus, our analysis suggests that the log-normal approximation may be a useful alternative to the normal approximation in analysis of big (hundreds of thousands of individuals) height data, analysis of heterogeneous populations, and analysis of the extremes of height.

Acknowledgment

The work of IAK, SAS and TIS were supported by Russian Ministry of Science and Education under the 5-100 Excellence Programme. The work of YSA and TIA was supported by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genetics SB RAS (project #0324-2019-0040-C-01). FAK is supported by the ERC Consolidator Grant (ChrFL: 771209). This research has been conducted using the UK Biobank Resource (project # 41601, "Non-additive effects in control of complex human traits").

References

[1] Galton, F. Regression Towards Mediocrity in HereditaryStature. The Journal of the Anthropological Institute of Great. Britain and Ireland vol. 15 246 (1886).
[2] Fisher, R.A. XV. The Correlation between Relatives on theSupposition of Mendelian Inheritance. Earth Environ. Sci.Trans. R. Soc. Edinb. 52, 399–433 (1919).
[3] Aulchenko, Y.S. et al. Predicting human height by Victorian and genomic methods. Eur. J. Hum. Genet. 17, 1070–1075 (2009).
[4] Visscher, P. M. Sizing up human height variation. Naturegen. v. 40 489–490 (2008).
[5] Wood, A. R. et al. Defining the role of common variation in thegenomic and biological architecture of adult human height. Nat. Genet. 46, 1173–1186 (2014).
[6] Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. Nature vol. 562 203–209 (2018).

DOI 10.18699/BGRS/SB-2020-031

Functioning of unique nitrile-detoxifying system in soil xenobiotic degrader Rhodococcus *rhodochrous*: a whole-genome transcriptomic approach

Konstantin V Lavrov
NRC "Kurchatov institute – GosNIIgenetika"
lavrov.ko@gmail.com

Artem S Kasianov
Vavilov Institute of General Genetics, Moscow, Russia
artem.kasianov@gmail.com

Tatyana I Kalinina
NRC "Kurchatov institute – GosNIIgenetika"
tatyana.i.kalinina@googlemail.com

Andrey D Novikov
NRC "Kurchatov institute – GosNIIgenetika"
alexm19@mail.ru

Alexander S Yanenko
NRC "Kurchatov institute – GosNIIgenetika"
yanenko@genetika.ru

Abstract — **Organic nitriles are significant participants of interactions between soil living organisms, thus their synthesis and degradation are of interest. Using transcriptomic approach, we revealed a genetic cascade, realizing nitrile degradation in powerful nitrile degrader – *Rhodococcus rhodochrous*. We also hypothesized a genetic "network" for nitrile/amide utilization, connected with cobalt homeostasis.**

Keywords* — *Rhodococcus, nitrile, amidase, cobalt, transcriptome, pathway regulation

Motivation and aim

Organic nitriles, both natural (mainly plant-derived [1]) and artificial (pesticides), are widely presented in soil. Nitriles are toxic, thus their detoxifying is an important part of interactions within soil micro- and macrobiota. Nitrilemediated interactions can be complex, including pest protection or variants of cooperation. An examples of possible complexity are that plant growth hormone indoleacetic acid can be synthesized as a result of enzymatic hydration of indole acetonitrile in bacteria, or, nitriles can be synthesized by bacteria *via* aldoxime dehydratase activity

Bacteria, which considered being main degraders of nitriles, are able to do it using two catabolic pathways – starting with a nitrile hydratase/amidase system (NH/AM) or with nitrilase [3, 4]. Known xenobiotic degraders - *Rhodococcus* bacteria – are also known hosts for both pathways. Remarkable features of cobalt-dependent NH/AM system in *Rhodococcus* are fast, nitrile-dependent switching on, very high synthesis of NH/AM enzymes (up to 50% of soluble intracellular proteins), and very high enzyme activity of NH [5]. Due to this unique combination of features, *Rhodococcus* species possessing cobalt-dependent NH/AM system are interesting representatives of soil nitriledegrading community.

A few examples of such system are known up to date, and its genetic regulation is poorly studied. It was only reported that cobalt ions and amides (products of NH activity) can activate transcription of NH/AM, and roles of three regulator genes located near NH were revealed [6]. We aim to uncover a more wide genetic "network" which participates in regulation and functioning of NH/AM system in *Rhodococcus*, and here we report a whole-genome transcriptomic approach to solve this task.

Methods

R. rhodochrous strain M8, containing NH and AM genes [7], was grown on four variants of glucose containing minimal medium, with following additions: (1) NH_4Cl, (2) NH_4Cl + $CoCl_2$, (3) urea, (4) urea + $CoCl_2$. The first was "blank" variant without activators of NH/AM, and in other cobalt/urea were probable independent (in 2 and 3) or cooperative (in 4) activators. Cells were collected at both mid-exponential and stationary phases of growth. Thus, 8 variants of cells were obtained, each as a sum of 10 independent reproducible growth experiments.

Total RNAs were extracted, quality controlled, separated on two aliquots, and resulting 16 samples were further processed independently. After clean-up from ribosomal RNA, cDNA libraries were prepared, and further sequenced on Illumina HiSeq 2500 platform (at Genotek, Moscow, Russia). Reads were mapped on *R. rhodochrous* M8 genome, and relative transcription levels (RTL) of genes were calculated with DESeq2 software. RTLs were expressed in RPKM units, and reflects, in principle, the following ratio: (number of reads mapped on ORF) / (number of all reads in the sample * length of ORF). Differences in RTLs between growth conditions were calculated as fold change (RTL_1/RTL_2).

Results general observations

In all growth conditions, minimal RTLs within the genome were approx. 10, and maximal – 16 000 - 100 000 units (max. 10 000-fold difference). RTLs of approx. 70% of genes were below 100, and of approx. 5% - higher than 1000 units. Between growth conditions, fold change of RTLs of 90-99% genes were lower than 10, and maximal fold change was 250.

Growth conditions 1 (NH_4^+) resulted in lowest RTLs of majority of genes, and in 2-4 conditions (with NH/AM activators) RTLs of many genes increased up to 250-fold. RTLs of very few genes decreased in 2-4, no more than 10-fold (except cluster of 5 ribosomal proteins genes, which decreased 40-fold). Significant differences in RTLs between 2-4 and 1 variant were observed only at exponential growth phase (log-phase). Data from log-phase was described below, if other not indicated. Variant 1 of growth conditions was chosen as blank not only for NH/AM RTLs comparison, but for comparison of RTLs throughout the genome.

Response of NH and AM to activators

Of two NHs presented in M8 genome, Co-type and Fe-type, only first transcribed at significant levels (500 – 100 000) and varied between growth conditions. RTLs of the last were below 80 and practically constant. Cobalt itself did not activate Co-NH, and urea itself activated it by 40fold. Co + urea conditions were the best for NH activation in log-phase (125-fold). In these conditions at stationary phase, RTL of Co-NH remained high (1,3-fold change comparing to log-phase). Also, RTLs of NH were greatly higher than RTLs of all other genes (approx. 90 000, 10-fold higher than highest other genes (11 000 for cold-shock protein)).

Of 15 available AMs, two AMs (formamidase *fmd5* and aliphatic amidase *amiE*) activated too in these conditions (11 and 25-fold). RTLs of *fmd5* and *amiE* further increased at stationary phase, comparing to log-phase (5- and 1.5-fold change). We suggest that these AMs are involved in biochemical cascade with Co-NH. In it, an amide starts to appear at log-phase as a result of NH activity, and then AMs continued to degrade it to acid.

RTLs of these AMs were relatively high (11 000 and 3700), thus felling into 0.4% portion of highly transcribing genes. A portion of NH/AM transcripts in total mRNA was as high as 23%, indicating a high significance of this pathway to bacterium. Interestingly, both AMs were much more induced by urea, than by urea + cobalt (100300-fold *versus* 11-25-fold, correspondingly). This indicates that the AMs can be involved in other yet unknown biochemical cascade (s). Taking in account this observation, and also a presence of Fe-NH and several other AMs, which are silent in given conditions, we can suggest a multiple, intersecting and branched nitrile detoxifying pathway in *Rhodococcus* bacteria.

Response of other genes to activators

Cobalt itself activates to some extent many genes throughout the genome (5-15-fold change). A few of it seemed to be related to cobalt (uptake, efflux, chelation, and biosynthesis of cobalamin), but majority are not, including clusters of probable prophages. When urea is present too, the majority of these not related to cobalt genes are again switched off. We suggest that cobalt causes massive stress response, which is significantly decreased when NH is synthesized and cobalt is incorporated into its active center.

Urea itself activates much smaller number of genes, among which are mainly probable genes of urea/ammonia transport and utilization, and, surprisingly, a cobalt-related genes - cobalamin biosynthesis cluster of 4 genes, including cobalt transporters (100-fold change), and separate cobalamin biosynthesis genes *cbiX, cobW, cbiM* (10-fold change). This observation can indicate possible existence of unknown regulation of cobalamin biosynthesis, and, also possible interconnection between pathways of cobalt and urea in the cell.

Acknowledgment

This work was financially supported by the Ministry of Education and Science of Russian Federation (Contract No. 075-15-2019-1658, creation and development of the "Kurchatov Genomic Center").

References

[1] Fraser F. Fleming. Nitrile-containing natural products. Natural Product Reports, 1999,16, 597-606, https://doi.org/10.1039/A804370A

[2] Ken-Ichi Oinuma, Yoshiteru Hashimoto, Kazunobu Konishi, Masahiko Goda, Takumi Noguchi, Hiroki Higashibata, Michihiko

[3] Kobayashi. Novel Aldoxime Dehydratase Involved in CarbonNitrogen Triple Bond Synthesis of *Pseudomonas chlororaphis* B23. J Biol Chem, 278(32), 29600-8, 2003 Aug 8, DOI: 10.1074/jbc.M211832200

[4] Yanenko, Alexander, Osswald, Steffen. Hydrolysis of Nitriles to Amides. in Enzyme Catalysis in Organic Synthesis, 2012. pp. 531-544 https://doi.org/10.1002/9783527639861.ch13

[5] Oßwald, Steffen, Yanenko, Alexander. Hydrolysis of Nitriles to Carboxylic Acids. in Enzyme Catalysis in Organic Synthesis, 2012. pp 545-559, https://doi.org/10.1002/9783527639861.ch14

[6] Konstantin V Lavrov, Anna O Shemyakina, Elena G Grechishnikova, Andrey D Novikov, Tatyana I Kalinina and Alexander S Yanenko. In vivo metal selectivity of metal-dependent biosynthesis of Co-type nitrile hydratase in *Rhodococcus* bacteria: a new look at the nitrile hydratase maturation mechanism? Metallomics, 2019, 11, 1162. https://doi.org/10.1039/C8MT00129D

[7] K. V. Lavrov, A. O. Shemyakina, E. G. Grechishnikova, A. D. Novikov, D. D. Derbikov, T. I. Kalinina, A. S. Yanenko. New cblA gene participates in regulation of cobalt-dependent transcription of nitrile hydratase genes in *Rhodococcus rhodochrous*. Research in Microbiology, 2018, May Jun;169(4-5):227-236, https://doi.org/10.1016/j.resmic.2018.03.006

[8] Novikov AD, Lavrov KV, Kasianov AS, Gerasimova TV, Yanenko AS. Draft genome sequence of *Rhodococcus* sp. Strain M8, which can degrade a broad range of nitriles. Microbiology Resource Announcements 2018;6. https://doi.org/10.1128/genomeA.0152617.135.

DOI 10.18699/BGRS/SB-2020-032

Asymmetry of motifs conservation within composite elements

Victor Levitsky
ICG SB RAS, Novosibirsk, Russia
levitsky@bionet.nsc.ru

Elena Zemlyanskaya
ICG SB RAS, Novosibirsk, Russia
ezemlyanskaya@bionet.nsc.ru

Dmitry Oshchepkov
ICG SB RAS, Novosibirsk, Russia
diman@bionet.nsc.ru

Tatyana Merkulova
ICG SB RAS, Novosibirsk, Russia
merkulova@bionet.nsc.ru

***Abstract* — The cooperative binding to genomic DNA of at least two transcription factors is the widespread mechanism of transcription regulation. Cooperating TFs are recognized through the analysis of co-occurrence of their binding sites. Recently we proposed Motifs Co-Occurrence Tool (MCOT) package that allowed predicting for a single ChIP-seq dataset composite elements (CEs), i.e. overrepresented pairs of spaced or overlapped motifs. Now we updated MCOT package and specified prediction of CEs with one of participating motifs possessing higher conservation than another one. Additionally, for each predicted CE we tested the hypothesis on the significance of asymmetry toward one or other motif. Hence, MCOT package now provides deeper insight in mechanism of cooperative action of TFs in gene regulation.**

Keywords* — *chromatin immunoprecipitation, transcription factor binding site recognition, conservation of motifs, motifs co-occurrence

Motivation and aim

Motivation

Tissue-, cell- and stage-specific regulation of gene expression is produced through multiple interactions of transcription factors (TFs) with respective regulatory elements called binding sites (BSs) or motifs. Typically, each TF functions in tight cooperation with other TFs [1]. One of the major mechanisms of TF-TF pair binding to DNA consists in preliminary DNA binding of one TF and subsequent binding to TF-DNA complex of other TF. Thus, we may expect that behaviour of two TF is asymmetric. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) analysis becomes the gold standard for protein/DNA-binding annotation at the whole genome level. In particular, ChIP-seq approach has been widely applied for annotation of TFs binding sites (motifs). The standard analysis pipeline at the final stage proposed application of *de novo* motif search tool, e.g. [2]. We recently proposed the Motifs Co-Occurrence Tool (MCOT) package that allowed predicting spaced as well as overlapped pairs of co-occurred motifs (composite elements, CEs) for a single ChIP-seq dataset [3]. A specific feature of MCOT is the analysis of the range of conservation for each motif in a CE.

Aim

In the current study, we demonstrated an example of the updated MCOT package application. The analysis of motifs conservation within pairs of co-occurred motifs allowed recognizing a specific type of CEs with asymmetric conservation toward either an Anchor or Partner motif.

Methods

We applied MCOT package as described earlier [3]. This tool annotates pairs of overrepresented motifs, i.e. composite elements (CE). Input data of MCOT comprise the peak dataset in FASTA format, the nucleotide frequency matrix for Anchor motif (that respected to potential BSs of immunoprecipitated TF), and either a Partner motif or the list of potential Partner motifs extracted from Hocomoco human or mouse collections [4].

MCOT classifies CEs according to mutual orientation of motifs, e.g. for heterotypic CEs there are four distinct orientations (Fig. 1A). There are three distinct cases of mutual locations: Full/Partial overlaps and Spacer, consequently MCOT uses five computation flows (Full, Partial, Overlap, Spacer and Any, Fig. 1B). MCOT applies the recognition model of Position Weight Matrix (PWM) for mapping. For each hit of PWM the false positive rates (FPR) for a whole-genome dataset of promoters, so that $-\mathrm{Log}_{10}[\mathrm{FPR}]$ value reflects the conservation of motif. For each computation flow, MCOT calculates p-values of Fisher's exact tests that estimates the CE significances and the significance of asymmetry (Fig. 1, C-G).

We modified the original MCOT algorithm [3] to estimate the asymmetry within predicted CEs as follow. We classified all them into two classes with more conservative Anchor or Partner motifs (see Fig. 1, C).

$\{-\mathrm{Log}_{10}[\mathrm{FPR}_{\mathrm{ANCHOR}}] > -\mathrm{Log}_{10}[\mathrm{FPR}_{\mathrm{PARTNER}}]\}$ (1a)

$\{-\mathrm{Log}_{10}[\mathrm{FPR}_{\mathrm{ANCHOR}}] \leq -\mathrm{Log}_{10}[\mathrm{FPR}_{\mathrm{PARTNER}}]\}$ (1b)

We classified the conservation of each motif within the ranges of 12 conservation levels, [<3.5], [3.5..3.7], [3.7..3.9] etc. up to [5.3..5.5] and [>5.5]. We computed the counts of CEs from the foreground and background datasets $\mathrm{Obs}_{i,j}$ and $\mathrm{Exp}_{i,j}$ that had distinct combinations of conservation levels. Here indices i and j denote Anchor and Partner motifs, resectively.Finally, the per mille measure transforms the absolute CE counts to relative ones as follow: $\{1000*\mathrm{Obs}_{i,j}/\mathrm{Obs}\}$ and $\{1000*\mathrm{Exp}_{i,j}/\mathrm{Exp}\}$.

Results

We illustrate calculation of CE asymmetry with the ChIP-seq data for FoxA2 in mouse liver [5]. In this study, besides the enrichment of the Anchor FoxA2 motif authors revealed its co-occurrence with Partner HNF1 motif. MCOT analysis confirmed this co-occurrence, but also found the significant asymmetry in predicted CEs toward Partner HNF1β motif ($p < 2E\text{-}28$ and $p < 4E\text{-}17$ for CEs with an overlap of motifs and with a spacer). Fig. 1H shows the difference between relative frequencies of observed and expected CEs with overlap of FoxA2 and HNF1β motifs (see Methods).

The recent version of MCOT package is available at https://gitlab.sysbio.cytogen.ru/academiq/mcot-kernel.

Acknowledgment

The work was supported by Russian Foundation for Basic Research (project # 18-29-13040) and State Budget Project 0324-2019-0040.

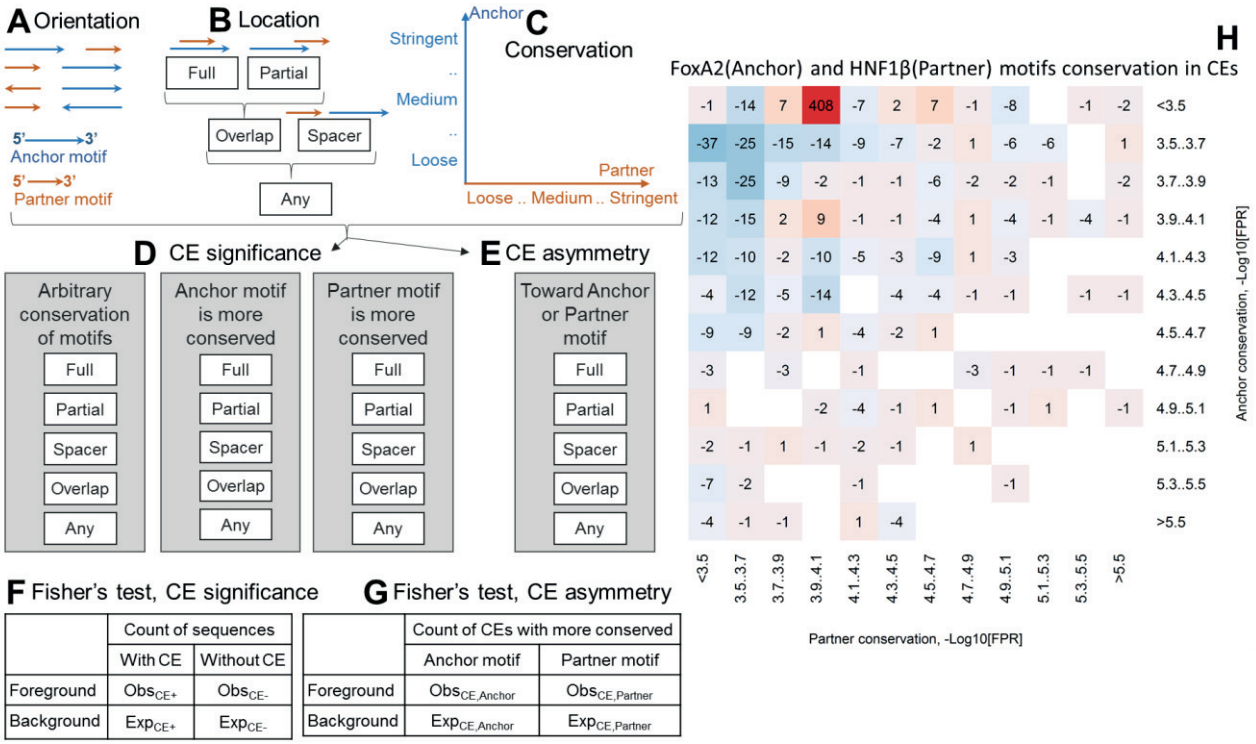


Fig. 1. MCOT package basic concept (**A-G**) and example analysis of asymmetry within predicted CEs FoxA2/HNF1β. (**H**). **A, B and C.** MCOT classifies all CEs according to mutual orientation, location and conservation of two motifs; the next analysis (**D** and **E**) is subdivided on five computation flows (Full, Partial, Overlap, Spacer and Any), that were defined according mutual location of motifs (**B**). For each potential CE we compute (1) it significance with no respect to conservation of motifs or (2,3) two significances for co-occurred pairs with one of motifs (Anchor or Partner) with more conserved hits than for other one (**D**); The asymmetry between Anchor and Partner motifs can be significant toward one of these motifs. MCOT constructs 2x2 contingency tables for calculation of significance of CEs (**F**) and CE asymmetry (**G**). To compute CE significance (**F**), we take peaks with hits of both motifs and compare fractions of peaks/permuted sequences that contained CEs. To compute the CE asymmetry (**G**) we compile two lists of all CEs for peaks and permuted sequences, and compare the fractions of CEs with more conserved Anchor and Partner motifs. **H.** The color shows the difference between observed (peaks) and expected (permuted sequences) relative abundances of CEs with an overlap of Anchor FoxA2 (axis Y) and Partner HNF1β (axis X) motifs for ChIP-seq data from liver [5], in per mille. FoxA2 and HNF1β motifs we derived from Homer de novo motif search [2] and Hocomoco mouse core collection [4]

REFERENCES

[1] Reiter F. et al. (2017) Combinatorial function of transcription factors and cofactors, Curr. Opin. Genet. Dev., 43:73–81.

[2] Heinz S., et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell., 38(4), 576-589.

[3] Levitsky V. et al. (2019) A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package, Nucleic Acids Res., 47(21): e139.

[4] Kulakovskiy I.V. et al. (2018) HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Res., 46(D1), D252-D259.

[5] Wederell E.D., et al. (2008) Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing, Nucleic Acids Res., 36(14):4549–4564.

DOI 10.18699/BGRS/SB-2020-033

Patterns of maternal and paternal inheritance in Russian populations

Anton Logachev
Dobzhansky Center for Genome Bioinformatics
St. Petersburg State University
St. Petersburg, Russia
a.logachev@spbu.ru

Daisuke Hirata
Dobzhansky Center for Genome Bioinformatics
St. Petersburg State University
St. Petersburg, Russia
d.khirata@spbu.ru

Gaik Tamazian
Dobzhansky Center for Genome Bioinformatics
St. Petersburg State University
St. Petersburg, Russia
g.tamazian@spbu.ru

***Abstract* — Genome Russia (GR) study is a national collaboration to expand our understanding of demographic history of peoples of Russia, to get deep insight into population genetics and to create a detailed genetic map of a country. Here we report analyses of matrilineal and patrilineal inheritance derived from whole genome sequencing data collected in the framework of the project. We examined sequences of 171 Y-chromosomes and 322 mtDNA ascertained from 14 ethnicities. We determined haplogroups for both Y-chromosomes and mtDNA sequences and estimated their distribution across the country. We performed phylogenetic and population genetic analyses for mtDNA and Y-chromosomes separately and then compared the results to reveal genetic history of indigenous groups living in Russia.**

Keywords — Genome Russia Project, Y-chromosome, mtDNA, haplogroups

Introduction

To decipher the population and demographic history of indigenous Russian ethnicities, migratory routes, and settlings of humankind in Russia, we analyzed mitochondrial genomes and Y-chromosome sequences from WGS datasets generated in the course of GR study. Phylogenetic and population genetic analyses were performed to reveal the maternal and paternal genetic history of the indigenous ethnic groups in Russia. By comparing both mitochondrial and Y-chromosomal DNA results, we inferred differences in the genetic history of both maternal and paternal population divergence and migration histories.

Materials and methods

We performed population genetic analyses for unrelated individuals (243 mtDNAs and 171 Y-chromosomes) from 14 populations (Central Russians, Southern Russians, Western Russians, Northern Russians, Adygeys, Seto, Komi, Udmurts, Khanty, Tatars, Bashkirs, Yakuts, Kalmyks, and Udege).

To assemble and genotype mitochondrial genomes we applied the MToolBox v.1.1 pipeline. In subsequent analyses we only used biallelic sites. To obtain mtDNA haplogroup information from each individual, the mtDNA haplogroup assignment of 322 individuals was searched by Haplogrep2 [2] using variant information generated with MToolBox as an input and PhyloTree.org - mtDNA tree Build 17 [3,4] as a reference. To identify Y-chromosomal haplogroups in males we performed SNP calling, and then tested and used different publicly available software tools e.g., AMY-tree [5], YFitter [6], Yleaf [7], and assembled them into our own pipeline. We then used available results from previous studies and 1K Genome project to construct phylogenetic trees for newly identified haplogroups in studied ethnicities.

To elucidate the population genetic status of each population, molecular diversity indices (Haplotype diversity, mean number of pairwise differences, number of segregating sites), Neutrality test (Tajima's *D*), and population pairwise F_{ST} for the whole mtDNA sequences and for male-specific regions in Y-chromosome were calculated with Arlequin v.3.5.2.2 [8]. Multi-dimensional Scaling (MDS) based on population pairwise F_{ST} genetic distances was conducted using 'cmdscale' function as implemented in R.

Results

We identified mtDNA haplogroups (Macro-haplogroup, Major haplogroup, and Haplogroup) distribution of the 243 individuals. Our analyses demonstrate that haplogroup composition of three populations, Yakuts, Kalmyks, and Udege was distinct from those of the remaining populations. Thus, macro-haplogroup M, which was mainly found in Asia and Siberia, is dominant in Yakuts, Kalmyks and Udege, while macro-haplogroup R, which was mainly found in Europe, is dominant in the remaining populations examined in the study, including ethnic Russians. Interestingly, Udege can be identified by a unique major haplogroup combination (M7, M8, and N9), which is only found in this ethnicity among the 14 populations analysed.

Ethnic Russians, belonging to the East Slavs ethnic group are clustered together with Komi and Seto, the Finno-Ugric peoples, as revealed with the MDS plot. Surprisingly, according to our results, Tatars, being a Turkic nation, are closely related to their geographic neighbors, Khanty and Udmurts (the Finno-Ugric peoples), rather than to the members of the same Turkic language family Yakuts and Bashkirs. Adyghe, Kalmyk, and Udege peoples attributed to Circassians, Mongolic, and Tungusic language families, respectively, show similarity between genetic and linguistic distinctness. Overall, mtDNA from non-ethnic Russians is heterogeneous compared to the homogenous ethnic Russians. The Y-chromosome analyses shows that while subclades of R1a1 and R1b1 haplogroups are common for all ethnic Russian populations, the central and northern Russians also have a significant proportion of N1a haplogroup. Other Y-chromosomal haplogroups in Russians are less-frequent. Komi and Khants both have a typical Finno-Ugric N1a haplogroup. Adyghe people have different subclade of G2a haplogroup commonly encountered in the Caucasus. Haplogroup frequency patterns revealed in Bashkirs show equal distribution of J1, J2a, E1b and C. We also noticed that Y-chromosomal haplogroups diversity is higher in populations of ethnic Russians and Bashkirs as compare to Adyghe and Komi people.

In this study we examined recent genetic history of mtDNA and Y-chromosome haplotypes in the gene pool of 14 ethnicities from present day Russia and documented divergence across various populations.

AKNOWLEDGMENTS

We are grateful to all collaborators from Genome Russia Consortium for their valuable comments and input. We further acknowledge all the donors contributing to the project.

REFERENCES

[1] Calabrese, C. *et al.* MToolBox: A highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. *Bioinformatics* (2014).

[2] Weissensteiner, H. *et al.* HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* (2016).

[3] van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* 30, 386–394 (2009).

[4] van Oven, M. PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci. Int. Genet. Suppl. Ser.* (2015).

[5] Van Geystelen et al. AMY-tree: an algorithm to use whole genome SNP calling for Y chromosomal phylogenetic applications. *BMC Genomics* 14:101 (2013).

[6] Karafet et al. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18: 830-838 (2008).

[7] Ralf et al. Yleaf: Software for human Y-chromosomal haplogroup inference from Next-Generation Sequencing data. *Mol. Biol. Evol.* 35(5): 1291-1294 (2018).

[8] Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567 (2010).

DOI 10.18699/BGRS/SB-2020-034

Molecular basis of phosphoryl guanidine oligonucleotides elongation by Taq DNA polymerase

Alexander Lomzov
ICBFM SB RAS, Novosibirsk, Russia
lomzov@niboch.nsc.ru

Dmitrii Pyshnyi
ICBFM SB RAS, Novosibirsk, Russia
pyshnyi@niboch.nsc.ru

***Abstract* — Phosphoryl guanidine oligonucleotides (PGO) become essential tools for basic research in molecular biology, molecular diagnostics and perspective for the therapeutics. To determine molecular basis of the observed effects of PG modification on elongation efficiency we performed molecular dynamics simulation of Taq polymerase / PGO complexes. The data obtained shows most important points for the analysis of protein / PGO interaction and gives useful information for development of improved PCR analysis.**

Keywords — Phosphoryl guanidine oligonucleotides, Taq polymerase, molecular dynamics

Motivation and aim

Motivation

Phosphoryl guanidine oligonucleotides (PGO) become essential tools for basic research in molecular biology, molecular diagnostics and perspective for the therapeutics [1-4]. A number of problems in PCR diagnostics could be solved by the using of PGOs. The influence of phosphoryl guanidine (PG) group(s) on elongation of primers by Taq DNA polymerase were studied experimentally. To determine molecular basis of the observed effects we analyzes of Taq polymerase / DNA complexes using computational approaches.

Aim

To determine molecular basis of the observed effects of PG modification on elongation efficiency we performed molecular dynamics simulation of Taq polymerase / PGO complexes.

Methods

The complex was build based on the structures PDBIDs: 1TAQ and 5E41. Lacking fragments of protein were added in disordered form.

Classical full atomic molecular dynamics simulations in implicit solvent water model were performed using Amber software, ff99bsc1 force field for DNA and ff14SB - for protein. The data obtained was analyzed using ccptraj and UCSF Chimera software.

Results

A number of complexes Taq polymerase and native and PG-containing oligonucleotides were analyzed. All complexes were stable along MD trajectory (50 – 100 ns). Most important PG positions of different diastereomers which are critical for PGO elongation were determined. The data obtained in good agreement with the experimental results. The results of the study shows most important points for the analysis of protein / PGO interaction and gives useful information for development of improved PCR analysis.

Acknowledgment

Supported by the RSF (18-14-0035)7 and by State project No. A-0309-2016-0004.

References

[1] Kupryushkin, M. S., Pyshnyi, D. V., Stetsenko, D. A. (2014). Phosphoryl guanidines: a new type of nucleic acid analogues. Acta Naturae, 6(4): 116-118.

[2] Garafutdinov, R. R., Sakhabutdinova, A. R., Kupryushkin, M. S., & Pyshnyi, D. V. (2020). Prevention of DNA multimerization using phosphoryl guanidine primers during isothermal amplification with Bst exo-DNA polymerase. Biochimie. 168: 259-267.

[3] Lomzov, A. A., Kupryushkin, M. S., Shernyukov, A. V., Nekrasov, M. D., Dovydenko, I. S., Stetsenko, D. A., Pyshnyi, D. V. (2019). Diastereomers of a mono-substituted phosphoryl guanidine trideoxyribonucleotide: isolation and properties. Biochemical and biophysical research communications. 513(4): 807-811.

[4] Kuznetsov, N. A., Lebedeva, N. A., Kuznetsova, A. A., Rechkunova, N. I., Dyrkheeva, N. S., Kupryushkin, M. S.,... & Lavrik, O. I. (2017). Pre-steady state kinetics of DNA binding and abasic site hydrolysis by tyrosyl-DNA phosphodiesterase 1. Journal of Biomolecular Structure and Dynamics. 35(11): 2314-2327.

DOI 10.18699/BGRS/SB-2020-035

Comparative genomics and quantitative proteomics reveal differentially produced proteins underlying virulence and host specificity in *Bacillus thuringiensis*

Yury Malovichko
Laboratory for Proteomics of Supra-Organismal Systems
All-Russian Research Institute of Agricultural Microbiology
Saint Petersburg, Russia
yu.malovichko@arriam.ru

Daria Gorbach
Department of Biochemistry
St. Petersburg State University
Saint Peterburg, Russia
daria.gorba4@yandex.ru

Maria Belousova
Laboratory for Proteomics of Supra-Organismal Systems
All-Russian Research Institute of Agricultural Microbiology
Saint Petersburg, Russia
m.belousova@arriam.ru

Ekaterina Romanovskaya
Department of Biochemistry
St. Petersburg State University
Saint Peterburg, Russia
e.romanovskaya@spbu.ru

Elena Lukasheva
Department of Biochemistry
St. Petersburg State University
Saint Peterburg, Russia
elena_lukasheva@mail.ru

Christian Ihling
Department of Pharmaceutical Chemistry and Bioanalytics
Institute of Pharmacy, Martin-Luther Universität Halle-Wittenberg
Halle, Germany
christian.ihling@pharmazie.uni-halle.de

Andrej Frolov
Department of Biochemistry
St. Petersburg State University
Saint Petersburg, Russia
Department of Pharmaceutical Chemistry and Bioanalytics
Institute of Pharmacy, Martin-Luther Universität Halle-Wittenberg
Halle, Germany
Andrej.Frolov@ipb-halle.de

Anton Nizhnikov
Laboratory for Proteomics of Supra-Organismal Systems
All-Russian Research Institute of Agricultural Microbiology
Saint Petersburg, Russia
Department of Genetics and Biotechnology
St. Petersburg State University
Saint Peterburg, Russia
a.nizhnikov@arriam.ru

Kirill Antonets
Laboratory for Proteomics of Supra-Organismal Systems
All-Russian Research Institute of Agricultural Microbiology
Saint Petersburg, Russia
Department of Genetics and Biotechnology
St. Petersburg State University
Saint Peterburg, Russia
k.antonets@arriam.ru

***Abstract* — *Bacillus thuringiensis* is a Gram-positive spore-forming bacterium known for its insecticidal activities. Although its features of high virulence and exceptional host specificity are wiely known and have conditioned its use as a source of novel biopesticides, molecular mechanisms underlying these traits remain elusive and are usually attributed to its repertoire of proteinaceous toxins. In this work we used combined proteogenomic approach to dissect <>. We used three different strains of B. thuringiensis belonging to the var. *thuringiensis*, *darmstadtiensis* and *israeliensis* and one derivate of B. thuringiensis var israeliensis, which lost the ability to produce Cry-toxins. By using hybrid Oxford Nanopore and Illumina sequencing we achieved replicon-level genome assemblies of the studied strains which, upon annotation, facilitated comparison of virulence factor repertoires and putative reasons of virulence loss in derivate strain, Further application of quantitative HPLC-Orbitrap-MS and proteome level-comparison between vegetative and sporulating cultures allowed us to identify the proteins, which were differentially produced in the strains at each stage. While proteins differentially produced at vegetative stage related mostly to cell metabolism with few virulence factors captured, those identified in spores included different spore coat proteins, flotillin-like proteins involved in in cell differentiation, exosporium proteins and a wide set of virulence factors including proteinaceous Cry toxins. Taking together the data obtained in this study indicate the impact of the accessory genes produced at various stages of colony development on pathogenicity-associated phenotypic traits in strains with different host ranges. To our knowledge, this is the first proteogenomic study of *B. thuringiensis* aimed to compare both genomic and proteomic profiles between serovars.**

Keywords* — *bacillus thuringiensis, biopesticides, toxins, genomics, proteomics, pathogenicity, virulence, specificity, nanopore, illumina, orbitrap

Introduction

Bacillus thuringiensis, commonly referred to as *Bt*, is a spore-forming bacterium of Firmicutes phylum mostly known as a pathogen of various invertebrates, especially insects [1]. A peculiar feature of this bacterium is that, despite a vast range of potential hosts, specificity of particular strains is usually confined to a usually narrow set of invertebrate species. Traditinally such specificity is attributed to proteinaceous toxins produced either by vegetative cells (e.g. Vip and Sip toxins) or at the onset of sporulation, of which three-domain toxins are the most well-known [2]. This paradigm, however, does not take into account numerous auxiliary virulence factors, such as chitinases, metalloproteases and phospholipases, which emerge to gain growing attention of microbiologists [3]. The abundance of this factors and distribution of respective genes among existent *Bt* strains confounds the exploration of molecular mechanisms modulatin *Bt* host specificity. Although genomics and other -omics techniques are being widely used in *Bt* studies, none of them have incorporated these approaches to compare virulence factor repertoires across strains or serovars. In this work we perform whole genome sequencing (WGS) of four *Bt* strains belonging to three different serovars followed by quantitative mass-spectrometry (MS) of respective vegetative and sporulating cultures. Our results indicate a set of proteins which may be related to specific virulence features known for the tested strains and patterns of their production,

Materials And Methods

Three virulent *Bt* strains belonging to *darmstadtiensis* (strain 109/25), *israeliensis* (strain 800/3) and *thuringiensis* (strain 800/15) serovars and an avirulent derivate of var. *israeliensis* strain were used in this work. Genomic DNA was extracted from vegetative bacterial cultures and sequenced with Illumina HiSeq X (Illumina, San Diego, CA, USA) and MinION (Oxford Nanopore Technologies, Oxford, UK). Genome assembly was performed using Flye with ON-based error correction was performed by four-wise launching of Minimap2 and Racon followed by Medaka and llumina-based assembly correction performed with Pilon. Quality control was performed using QUAST, and assembly graphs were visualized using Bondage. Obtained assemblies were annotated using Prokka with *Bacillus*-related protein seqeunces from NCBI IPG as a database. Genome-level contig alignment was performed using MUMmer.

For proteomic assays, total protein fraction was extracted from both vegetative cells grown on LB medium and sporulating culture grown on T3 medium to induce spore formation, Qualitative proteome analysis was performed using electrospray quadruple mass-spectrometry coupled with fast performance liquid chromatography (FPLC-ESI-MS) of spots selected by two-dimensional difference gel electrophoresis (2D-DIGE). For quantitative analysis, a high-performance liquid chromatography/Orbitrap electrospray mass-spectrometry (HPLC-Orbitrap-MS) was used. Resulting spectra were annotated using Progenesis QI and Perseussoftware. Acquired data were preprocessed, normalized and imputed at protein level using Msnbase R package. Differential expression was assessed by independent use of two R packages, limma and MsqRob. Proteome-wide functional annotation was performed using eggNOG v2.0 with subsequent over-representaion assessment in differentially expressed gene subsets using topGO and BOG R packages.

Results

Genomes of all four strains used in this work were assembled up to replicon level. Subsequent annotation revealed all of these strains possessed genes encoding chitinases, phospholipases C and CalY metalloproteases, while distribution of several other virulence factors such as ColB, NlpC and NprB were confined to particular strains. Also, the Cry gene repertoire for each strain was evaluated. It appeared that strain 800/15 bear genes for Cry1Ba1 and Cry1Ab12, strain 109/25 possess genes encoding Cry1Ea10 toxin and strain 800/3 genome encodes Cry4Ba1, Cry4Aa2 and Cry10Aa3 according to Bt Nomenclature []. Pathogenicity loss in the avirulent strain was attributed to the absence of the Cry-encoding plasmid present in the progenitor strain.

Qualitative proteomic assay revealed 21 differentially produced protein among four strains in question. Further quantitative HPLC-Orbitrap-MS showed that all four strains show similar patterns of differences between vegetative and sporulating cells which reflect processes underlying cll compaction and spore formation with a median of 20 proteins upregulated in vegetative cells and 30 proteins upregulated in spores. At the same time, interstrain comparisons at the respective stages pointed out differences potentially involved in pathogenicity phenotypes. Proteins differentially produced at the vegetative stage mostly involve cellular metabolism enzymes with only few virulence factors captured, including flagellins and virulence-related metalloproteases. Those present in the spores, however, are enriched with virulence determinants including Cry toxins as well as different proteins composing spore coat, exosporium and cellular membranes. Taken together, obtained data not only revealed changes in the B. thuringiensis proteomes associated with vegetative cells-spore transition, but also shed light on the differences between the strains at both stages, which might help to better understand mechanisms of host specificity.

ACKNOWLEDGMENT

This work was supported by the Russian Foundation for Basic Research, Project 20-316-70020.

REFERENCE

[1] B Raymond, PR Johnston, C Nielsen-LeRoux, D Lereclus, N Crickmore. *Bacillus thuringiensis*: an impotent pathogen?. Trends Microbiol. 2010;18(5):189–194. doi:10.1016/j.tim.2010.02.006

[2] L Palma, D Muñoz, C Berry, J Murillo, P Caballero. *Bacillus thuringiensis* toxins: an overview of their biocidal activity. Toxins (Basel). 2014;6(12):3296–3325. Published 2014 Dec 11. doi:10.3390/toxins6123296

[3] YV Malovichko, AA Nizhnikov, KS Antonets. Repertoire of the Bacillus thuringiensis Virulence Factors Unrelated to Major Classes of Protein Toxins and Its Role in Specificity of Host-Pathogen Interactions. Toxins (Basel). 2019;11(6):347. Published 2019 Jun 17. doi:10.3390/toxins11060347

DOI 10.18699/BGRS/SB-2020-036

Software pipeline for the analysis of the functional role of nucleotide substitutions in regulatory regions of genes and its testing on polymorphisms associated with obesity

Ekaterina Alekseevna Matrosova
Department of Natural Sciences
Novosibirsk State University
Novosibirsk, Russia
MatrosovaEA@bionet.nsc.ru

Vadim Mikhailovich Efimov
Laboratory of Molecular Genetic Systems, ICG SB RAS
Novosibirsk, Russia
Novosibirsk State University
Novosibirsk, Russia
efimov@bionet.nsc.ru

Elena Vasilevna Ignatieva
Laboratory of Evolutionary Bioinformatics and Theoretical Genetics, ICG SB RAS
Novosibirsk, Russia
eignat@bionet.nsc.ru

Abstract — **Genome-wide association studies have shown that approximately 90% of the nucleotide substitutions associated with diseases are located outside the coding regions of the genes, with about 40% occurring in regulatory regions. However, the molecular mechanisms by which such SNPs influence incidence of diseases are still poorly understood. In this work, we have developed a software pipeline that allows us to predict the potential effects of nucleotide substitutions on potential transcription factor binding sites (TFBSs). This pipeline integrates and analyzes information obtained from the UCSC Variant Annotation Integrator and PERFECTOS-APE programs, and the dbSNP, Ensembl and HOCOMOCO databases. The operation of the pipeline was tested on two sets of SNPs. All of them were located in the *BDNF* gene regulatory regions in the vicinity of the SNP rs11030104 associated with elevated body weight. SNPs from the first set were in linkage disequilibrium with rs11030104 and the second set contained random SNPs. We found some differences in the functional characteristics of TFBSs and corresponding transcription factors identified on the basis of the analysis of these two sets of SNPs.**

***Keywords — polymorphism (SNP), transcription factor binding sites (TFBSs), software pipeline, obesity*.**

Introduction

Genome-wide association studies have shown that approximately 90% of the nucleotide substitutions associated with diseases are located outside the coding regions of the genes, with about 40% occurring in regulatory regions [1]. However, the molecular genetic mechanisms of the influence of the revealed polymorphisms on the development of pathologies have been studied insufficiently. Such genetic variants can be either causative, that is, change the functional activity of regulatory DNA sites (transcription factor binding sites), or they may mark specific haplotypes containing causal SNPs [1–3].

The aim of this work was to create a software pipeline to evaluate the potential effects of nucleotide substitutions on TFBSs. Functionality of the software pipeline was tested on nucleotide variants in regulatory regions of the *BDNF* gene associated with the development of monogenic forms of obesity [4]. The relevance of the chosen research object is due to the fact that today obesity is an extremely common disease in both developed and developing countries. In addition, obesity may increase the risk of many health problems [5].

Materials and methods

To write scripts of the software pipeline, the Python was used. Data on polymorphisms was obtained from dbSNP, Ensembl, UCSC table browser, and UCSC Variant Annotation Integrator program output. Weight matrix data and the types of DNA-binding domains of transcription factors (TFs) were obtained from HOCOMOCO [6]. Prediction of the effects of nucleotide substitutions on TFBSs was carried out using the PERFECTOS-APE program using the default settings (P-value cutoff = $5*10^{-4}$, Fold change cutoff = 4).

The set of SNPs № 1 (case) included 87 SNPs that were in linkage disequilibrium in the European population with the *BDNF* gene polymorphism rs11030104 associated with increased body weight [7]. The identifiers of these SNPs were extracted from the LDproxy database using the criterion $r^2>0.6$. The set of SNPs № 2 (control) included 86 SNPs localized in the vicinity of rs11030104 polymorphism but not linked to it and having a minor allele frequency over 0.01. To compare the results of the analysis of case and control, the statistical chi-squared test and Kullback criterion were used [8]. The p-value threshold for a statistical significance was 0.05.

Results

We have developed a software pipeline that includes two original Python scripts that integrate and analyze information obtained from the UCSC Variant Annotation Integrator and PERFECTOS-APE programs, as well as from the dbSNP, Ensembl, and HOCOMOCO databases (Fig. 1). At the output, the pipeline provides information on the effects of nucleotide substitutions on the similarity of a DNA regions with the TFBS weight matrices, on the positions of SNPs in the genome and relative to gene transcription start sites, on the TFs, as well as the p-values characterizing the similarity of pairs of sites with their matrices, its fold change, and some other data.

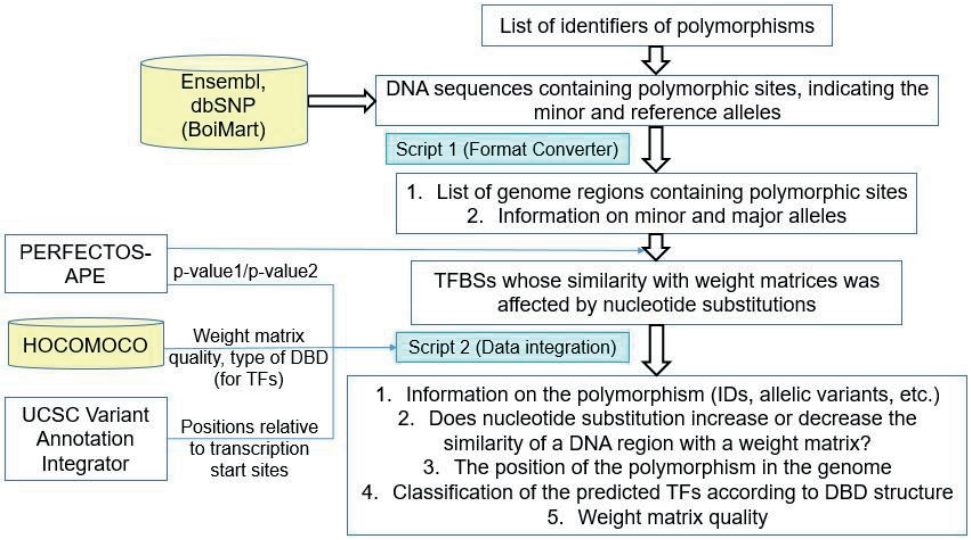


Fig. 1. The scheme of the software pipeline, which allows to analyze the effects of nucleotide substitutions on TFBSs

The software pipeline was tested on the case and the control sets of polymorphisms (see Materials and methods). As a result, 423 and 505 TFBSs were predicted, whose similarity with weight matrices changed by nucleotide substitutions (Table 1). According to the 1000 Genomes Project report [9], for SNPs located in conserved fragments of regulatory regions of the human genome the ratio between the number of nucleotide substitutions damaging the TFBSs and the number of substitutions leading to their occurrence is 4:1. Both SNPs sets showed a significant difference in the proportions between positive and negative effects of nucleotide substitutions on TFBSs in comparison with this ratio. In both cases, an approximately equal numbers of both positive and negative effects of nucleotide substitutions on the similarity of TFBSs with the corresponding weight matrices were predicted (Table 1).

TABLE 1 – RESULTS OF ANALYSIS OF SNPS EFFECTS ON TFBSS, OBTAINED BY THE SOFTWARE PIPELINE

SNPs set	**Number of SNPs in the set**	**Number of identified TFBSs**	**Number of TFBSs for which the similarity with the weight matrix has**		**Number of TFs corresponding to identified TFBSs**
			increased	***decreased***	
№ 1 (case)	87	423	215	208	246
№ 2 (control)	86	506	231	275	266

It was revealed that the predicted sites corresponded to 246 and 266 transcription factors (Table 1). When comparing the distributions of these TFs sets by the classes of their DNA-binding domains (DBDs), significant differences were identified using the chi-squared test, as well as using Kullback criterion. When comparing the proportion of TFs of a particular DBD class, significant differences were found between sets of TFs for the classes of *Basic* and *Other all-alpha-helical* types of DBDs. When comparing the proportions of binding sites for specific TFs using Kullback criterion, significant differences were revealed between sets of TFs predicted for the case and the control sets of SNPs.

Similar results were obtained when comparing the sets of predicted TFBSs and TFs for the control SNPs set (on the one hand) and the sets of predicted TFBSs and TFs for SNPs from the case set that have a higher degree of correlation with rs11030104: $r^2>0.8$ and $r^2>0.9$ (on the other hand) (data not shown).

Conclusion

We have developed a software pipeline that takes a list of SNP identifiers as an input and allows us to predict the potential effects of nucleotide substitutions on potential transcription factor binding sites. The efficiency of the pipeline was tested on two sets of polymorphisms in the regulatory regions of the *BDNF* gene associated with obesity.

ACKNOWLEDGMENT

Supported by the State Budgeted Project 0324-2019-0040-C-01.

REFERENCES

[1] M. T. Maurano *et al.*, 'Systematic Localization of Common Disease-Associated Variation in Regulatory DNA', *Science (80-.).*, vol. 337, no. 6099, pp. 1190–1195, Sep. 2012.

[2] J. M. Seddon, K. T. Berggren, and L. M. Fleeman, 'Evolutionary history of DLA class II haplotypes in canine diabetes mellitus through single nucleotide polymorphism genotyping', *Tissue Antigens*, vol. 75, no. 3, pp. 218–226, Mar. 2010.

[3] Y. Zhang, M. Manjunath, S. Zhang, D. Chasman, S. Roy, and J. S. Song, 'Integrative Genomic Analysis Predicts Causative Cis - Regulatory Mechanisms of the Breast Cancer–Associated Genetic Variant rs4415084', *Cancer Res.*, vol. 78, no. 7, pp. 1579–1591, Apr. 2018.

[4] J. Gray *et al.*, 'Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF) gene.', *Diabetes*, vol. 55, no. 12, pp. 3366–71, Dec. 2006.

[5] K. R. Fontaine, D. T. Redden, C. Wang, A. O. Westfall, and D. B. Allison, 'Years of life lost due to obesity', *JAMA*, vol. 289, pp. 187–193, 2003.

[6] I. V Kulakovskiy *et al.*, 'HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis', *Nucleic Acids Res.*, vol. 46, no. D1, pp. D252–D259, Jan. 2018.

[7] A. E. Locke *et al.*, 'Genetic studies of body mass index yield new insights for obesity biology', *Nature*, vol. 518, no. 7538, pp. 197–206, Feb. 2015.

[8] S. Kullback, *Information theory and statistics*. Peter Smith, 1978.

[9] 1000 Genomes Project Consortium *et al.*, 'An integrated map of genetic variation from 1,092 human genomes', *Nature*, vol. 491, no. 7422, pp. 56–65, Nov. 2012.

DOI 10.18699/BGRS/SB-2020-037

Whole genome sequencing and assembly of *Saccharomyces cerevisiae* genomes using Oxford Nanopore data

Andrew G. Matveenko
Dpt. of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia
a.matveenko@spbu.ru

Anton B. Matiiv
Dpt. of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia
antonmatiiv@yandex.ru

Yury A. Barbitoff
Dpt. of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia; Bioinformatics Institute, St. Petersburg, Russia
barbitoff@bk.ru

Evgenia M. Maksiutenko
Vavilov Institute of General Genetics, St. Petersburg Branch; Dpt. of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia
jmrose@yandex.ru

Svetlana E. Moskalenko
Vavilov Institute of General Genetics, St. Petersburg Branch; Dpt. of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia
s.moskalenko@spbu.ru

Alexandra V. Beliavskaia
University of Liverpool, Liverpool, UK
alex.beliavskaia@gmail.com

Alexander V. Predeus
University of Liverpool, Liverpool, UK; Bioinformatics Institute, St. Petersburg, Russia
predeus@bioinf.me

Galina A. Zhouravleva
Dpt. of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia
g.zhuravleva@spbu.ru

***Abstract* — The Peterhof genetic collection (PGC) of yeast *Saccharomyces cerevisiae* is one of the rare examples of a large genetic collection established independently of reference S288C strain. We assembled genomes of two widely used PGC strains, 1A-D1628 and 74-D694, using Oxford Nanopore MinION sequencing data. Subsequent analysis of structural variations showed a number of differences between PGC strains and S288C.**

Keywords* — Saccharomyces cerevisiae, *NGS, Oxford Nanopore, structural variations

Motivation and aim

Saccharomyces cerevisiae is one of the most thoroughly studied eukaryotic model organism. Hundreds of yeast strains have been described in wild environment, industrial lineages, and laboratories throughout the world. Most of the common laboratory yeast strains, and, specifically, the reference strain S288C ascend to the so-called "Berkeley yeast". The Peterhof genetic collection (PGC) of S. cerevisiae was founded 50 years ago and it is one of the rare examples of a large genetic collection established independently. First PGC strains ascend to a distillery lineage, which is unrelated to the Berkeley yeast. For many years, studies of translation, prion biology, and other fields have benefited from PGC, and several laboratory strains are now widely used throughout the world.

Recently, next-generation sequencing (NGS) was applied extensively to study the diversity of yeast strains. To place the PGC into known interrelations of yeast lineages we previously attempted whole genome sequencing of several strains using Ion Torrent NGS technology [1]. This allowed the identification of genomic variations that caused several phenotypic traits in these strains, e.g., clumping phenotype, phenylalanine auxotrophy, nonsense suppression caused by defective *SUP35* transcription [2], etc. The closest strain to the PGC progenitor, 15V-P4, was shown to differ greatly from other laboratory stocks. Despite the progress achieved, the obtained assemblies of the PGC genomes were incomplete and required substantial improvement.

Methods

We performed genome sequencing of two strains, 1A-D1628, and 74-D694, using Oxford Nanopore MinION. Using the obtained reads we made a draft assembly that comprises all yeast chromosomes and mtDNA. We compared several long-read assemblers, with Canu showing best results. The obtained assemblies were polished by raw Nanopore signal and, subsequently, by short reads produced by Illumina HiSeq.

Results

The final assemblies were higly contiguous: 15 and 14 chromosomes (out of 16) were fully assembled for 1A-D1628 and 74-D694, respectively. 97.3% 97.8% complete BUSCO were discovered in 1A-D1628 and 74-D694, respectively, (compared to 98.0% for reference assembly) demonstrating high reliability of the obtained sequences. We next performed structural variants analysis and discovered several gross rearrangements, as well as multiple intergenic deletions, insertions and duplications in PGC genomes. All structural variations tested have been confirmed by PCR or Sanger sequencing. We also compared the number of *CUP1* gene clusters and copper resistance of the strains. PGC strains were less resistant to copper, and contained less copies of the *CUP1* gene, as shown by qPCR. Comparative analysis and annotation of the whole genome sequences are continuing.

Acknowledgments

The work is supported by RCs "MCT" and "Biobank" of SPbSU, by the State research program 0112-2016-0015, by the RFBR grant 20-34-70156, and the RSF grant 18-14-00050.

References

[1] Drozdova, P.B. et al/ (2016). Genome sequencing and comparative analysis of Saccharomyces cerevisiae strains of the Peterhof genetic collection. PloS One, 11(5).

[2] Matveenko, A.G., Drozdova, P.B., Moskalenko, S.E., Tarasov, O.V., & Zhouravleva, G.A. (2019). Whole genome sequencing data and analyses of the underlying SUP35 transcriptional regulation for a Saccharomyces cerevisiae nonsense suppressor mutant. Data in brief, 23, 103694.

DOI 10.18699/BGRS/SB-2020-040

Whole genome analysis of clinical *Staphylococcus aureus* multi-drug resistant isolates from Moscow medical center

Yulia Mikhaylova
Central Research Institute
of Epidemiology
Moscow, Russia
mihailova@cmd.su

Andrey Shelenkov
Central Research Institute
of Epidemiology
Moscow, Russia
0000-0002-7409-077X

Yurii Yanushevich
Central Research Institute
of Epidemiology
Moscow, Russia
yanushevich@cmd.su

Valeria Fomina
National Medical and Surgical Center
named after N.I. Pirogov
Moscow, Russia
med_2006@mail.ru

Mikhail Zamyatin
National Medical and Surgical Center
named after N.I. Pirogov
Moscow, Russia
mnz1@yandex.ru

Dmitry Shagin
Central Research Institute
of Epidemiology
Moscow, Russia
shagin@cmd.su

Vasiliy Akimkin
Central Research Institute
of Epidemiology
Moscow, Russia
vgakimkin@yandex.ru

***Abstract* — *Staphylococcus aureus* is a Gram-positive bacterium that is a usual member of human body microbiota and can be frequently found in the upper respiratory tract and on the skin. It can become an opportunistic pathogen causing skin infections including abscesses, respiratory tract infections and food poisoning. Methicillin-resistant strains of *S. aureus* represent one of the major sources of healthcare-associated infections in clinical settings. Although more than 10,000 *S. aureus* genomes are currently available in public databases, the lack of associated metadata usually does not allow conducting the studies on antibiotic resistance acquisition, epidemiological surveillance and the presence of virulent factors in this important pathogen. We have performed whole-genome sequencing (WGS) of 22 clinical *S. aureus* isolates on Illumina and Oxford Nanopore platforms. Isolate typing was performed using Multilocus sequence typing (MLST), antimicrobial resistance factors were revealed using both phenotypical and genomic studies, virulence factors and plasmids were annotated in genome sequences. These data will be useful for future epidemiological and antimicrobial resistance studies of *S. aureus*.**

Keywords* — *NGS, pathogenic bacteria, whole genome sequencing, antibiotic resistance

Introduction

A number of factors determine the role of *Staphylococcus aureus* as both commensal and pathogenic bacteria. These include many determinants of virulence, as well as the ability to acquire multiple resistance to antimicrobial drugs. Infections caused by methicillin-resistant *S. aureus* (MRSA) have become a global problem.

The aim of this work is to perform whole genome sequencing (WGS)-based genotyping and genomic epidemiology analysis of multi-drug resistant *S. aureus* isolates including MRSA. In addition, several isolates obtained from the same patient were sequenced to investigate the optimal SNP threshold value for distinguishing close and distantly related isolates by their core genome comparison.

Materials and methods

WGS was performed on Illumina HiSeq platform for 22 isolates of *S. aureus* including two environmental isolates collected from Moscow medical center, and two isolates were also sequenced on MinION platform (Oxford Nanopore). Genomic DNA was isolated with DNeasy Blood and Tissue kit (Qiagen) and used for paired-end library preparation with Nextera™ DNA Sample Prep Kit (Illumina®). For MinION platform, DNA was used to prepare the MinION library with the Rapid Barcoding Sequencing kit SQK-RBK004 and was sequenced on R9 SpotON flow cell. The susceptibility was determined by the boundary concentration method on VITEK2Compact30 analyzer (bioMerieux, France).

Genome assemblies were made using SPAdes version 3.13 [1] and Unicycler [2] (hybrid long- and short read assemblies). Genome comparisons were performed using roary [3], dnadiff and custom software. Custom annotation pipeline was used for MLST and spa-typing, antibiotic resistance gene prediction, virulence factor and plasmid search using corresponding versions of public databases.

results

The population of *S. aureus* studied was represented by two dominating lines (ST8 and ST22) and 5 minor lines including one isolate possessing a novel sequence type (ST5555) and having previously unknown MLST allele (pta_647). The isolates studied were divided into 11 Spa variants with prevailing Spa-type t008. Seventeen MRSA isolates had SCCmec IV type, and one - SCCmec V cassette.

Other common determinants of antibiotic resistance were genes of resistance to penicillin, macrolides and amikacin. The set of virulence factors in *S. aureus* isolates was quite extensive (up to 20 genes), and included both structural components of the cell and secreted products. Pairwise comparisons of core-genome composition and the number of single nucleotide polymorphisms (SNP) of isolates from the same and different patients, revealed discrepancies in number of SNPs and thresholds previously proposed for strain discrimination in literature. For example, the threshold of 15 SNPs for distinguishing S. aureus isolates was proposed

earlier [4], while we have revealed up to 51 nucleotide mismatches between the core genomes of our isolates from the same patient.

Conclusions

Thus, applying WGS technology enabled us to reveal unique genome features of *S. aureus* clinical isolates. Such features are important both from the standpoint of an increased level of virulence of MRSA and evolutionary changes. The data obtained will allow us to identify new target molecules for differentiation of MRSA isolates belonging to different genetic variants. Our results suggest that although the number of core-genome SNPs proves to be useful measure for distinguishing distant bacterial strains, more research is needed to establish appropriate thresholds for it.

References

[1] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77

[2] Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595

[3] Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3

[4] Schurch et al. (2018) Clinical Microbiology and Infection, 24:350-354

DOI 10.18699/BGRS/SB-2020-041

Repetitive elements in the genome of Siberian larch (*Larix sibirica* Ledeb.)

K. A. Miroshnikova
Laboratory of Radioecology
Institute of Biophysics SB RAS
Krasnoyarsk, Russia

V. V. Biriukov
Laboratory of Genomic Research and Biotechnology
FRC KSC SB RAS
Krasnoyarsk, Russia

D. A. Kuzmin
Laboratory of Forest Genomics
SibFU, Krasnoyarsk, Russia

M. G. Sadovsky
Laboratory of Forest Genomics
SibFU, Krasnoyarsk, Russia

E. I. Bondar
Laboratory of Genomic Research and Biotechnology
FRC KSC SB RAS
Krasnoyarsk, Russia

Y. A. Putintseva
SFB&BT, SibFU, Krasnoyarsk, Russia

K. V. Krutovsky
Department of Forest Genetics and Forest Tree Breeding
Georg-August University of Göttingen
Göttingen, Germany
Laboratory of Forest Genomics
SibFU, Krasnoyarsk, Russia
Laboratory of Population Genetics
Vavilov Institute of General Genetics
Moscow, Russia
konstantin.krutovsky@forst.uni-goettingen.de

V. S. Akulova
Laboratory of Genomic Research and Biotechnology
FRC KSC SB RAS
Krasnoyarsk, Russia

V. V. Sharov
Laboratory of Genomic Research and Biotechnology
FRC KSC SB RAS
Krasnoyarsk, Russia

N. V. Oreshkova
Laboratory of Genomic Research and Biotechnology
FRC KSC SB RAS
Krasnoyarsk, Russia

Abstract — **Repetitive elements (REs) are both the problem for assembling whole genome nucleotide sequence reads and the enormous uncharted research field [1]. Besides, REs can be part of proteins that play important role in plant immunity [2]. Conifers are among ancient representatives of plants. They are keystone species in boreal ecosystems that greatly affect global climate. There are a great number of REs in the conifer genomes [3], and most of them are still not characterized. Their role in species evolution and gene regulation is still largely unknown. The main goals of this research were de novo discovery and classification of REs in the genome of Siberian larch (Larix sibirica Ledeb.), and analysis of amino acid repeats in proteins expressed in different tissues and a role that they play in response to pathogens, especially, leucine-rich repeats (LRRs).**

Keywords — Larix sibirica; Repetitive elements; Leucine-rich repeats

Introduction

Repetitive elements (REs) or repeats are sequences that occur multiple times in genome. They can be divided into two broad distinct groups — tandem and interspersed repeats, respectively. Tandem repeats are relatively short nucleotide motifs repeated consecutively one after another in a head to tail orientation. Interspersed repeats are mainly results of transposable element or transposon activity, and they are usually nonadjacent and interspersed throughout the host genome [1].

Active transposons and other mobile elements can move and change their location in genome, and some of them, such as class I transposable elements or retrotransposons can create a new copy of itself before insertion [1], [4]. The activity of transposable elements is usually suppressed by a host defense mechanism, but they can be activated by stress factors. The coding genome regions are better protected from insertion of new repeat copies. Most of the new copies can be inserted into already existing genome repeats. Moreover, repeats themselves are very prone to mutation accumulation. These two factors lead to the formation of the so-called 'genomic dark matter'. It is difficult to identify these genome regions, but they contribute significantly to the content of plant genomes [5]–[7].

Some repeats are used by organisms for their particular needs. For example, many plant proteins associated with disease resistance, and which function is to recognize pathogen molecules, have domain containing leucine rich-repeats (LRRs). The number of leucine in the LRR regions and number of the LRR regions can vary both within and between species, which provides the specificity of pathogen molecules recognition. Thus, LRRs play a very important role in plant immunity [2].

Conifer genomes have gigantic lengths, which is not due to genome-wide duplication events. It was suggested that the main cause of the increased genome sizes is the propagation of repetitive DNA sequences [3]. Thus, it is very difficult to study conifer genomes. Nevertheless, a technological breakthrough in sequencing and assembly of genomes made it possible to analyze the genomes of some conifers (see [8] for review). Siberian larch has also a huge genome of ~12 Gbp, although it is smaller than in other conifers [9]. About 80% of this genome consists of REs. Here we provide data on the preliminary analysis of the larch repeatome, including LRRs.

Methods and Algorithms

RepeatModeler was used for *de novo* discovery of REs [10], [11]. This program is widely used in the analysis of repeats in the large genomes. TEclass was used for classification of found repeats [12]. LRRs were found using

HMMER package with Hidden Markov Models (HMMs) for some LRR families from Pfam database [13], [14].

Results

In total, 1908 repeat families with a total number of 75945 iterations were discovered and the majority of them (1424 or 75%) were not classified by RepeatModeler based on the blast results. TEclass managed to classify 94% of 1908 found families, but 41% of 484 families already classified by RepeatModeler were redefined entirely differently. RepeatModeler is based on *de novo* repetition detection programs — RepeatScout and RECON. Since RepeatScout does not use all scaffolds or contigs for the analysis, but only a randomly selected part of them, we decided to analyse 2869 scaffolds from a larch assembly longer than 100 Kbp. Their total length was 360 Mbp, which corresponds to 3% of the total length of the entire assembly. LRRs also were detected in the transcriptomes of larch tissues (autumn buds, shoots, sprouts, needles, and cambium). The presence of these LRRs was approximately the same in all tissues and no more than 2%. LRRs in larch bud were compared with LRRs in the transcriptome of the Sitka spruce bud, and the number of LRRs in the spruce transcriptome was less than in the larch transcriptome (0.5% *vs* 0.9%, respectively).

Conclusion

REs were identified in the Siberian larch genome, and their preliminary analysis was performed. This study can help us better understand organization of large genomes, such as conifer genomes.

LRRs and resistance genes in transcriptomes of different tissues of Siberian larch were identified and analyzed. These sequences are associated with immune response of plants to biotic stress. Their study will also help us better understand genetic mechanisms of disease resistance in larch and other plants.

Acknowledgments

This study was funded by a research grant No. 14.Y26.31.0004 from the Government of the Russian Federation.

References

[1] J. Padeken, P. Zeller, S. M. Gasser, “Repeat DNA in genome organization and stability,” Current opinion in genetics & development, vol. 31, pp. 12-19, 2015.

[2] E. Schaper, M. Anisimova, “The evolution and function of protein tandem repeats in plants,” New Phytologist, vol. 206, no. 1, pp. 397-410, 2015.

[3] B. Nystedt et al. “The Norway spruce genome sequence and conifer genome evolution,” Nature, vol. 497, no. 7451, pp. 579-584, 2013.

[4] T. Wicker, et al. “A unified classification system for eukaryotic transposable elements,” Nature Reviews Genetics, vol. 8, no. 12, pp. 973-982, 2007.

[5] F. Maumus, H. Quesneville, “Impact and insights from ancient repetitive elements in plant genomes,” Current opinion in plant biology, vol. 30, pp. 41-46, 2016.

[6] B. Piégu, S. Bire, P. Arensburger, Y. Bigot, “A survey of transposable element classification systems–a call for a fundamental update to meet the challenge of their diversity and complexity,” Molecular phylogenetics and evolution, vol. 86, pp. 90-109, 2015.

[7] S. Bire, F. Rouleux-Bonnin, “Transposable elements as tools for reshaping the genome: it is a huge world after all!” Mobile Genetic Elements. Humana Press, pp. 1-28, 2012.

[8] E. Mosca et all. “A reference genome sequence for the European silver fir (*Abies alba* Mill.): a community-generated genomic resource,” Genes Genomes Genetics G3, vol. 9, no. 7, pp. 2039-2049, 2019.

[9] D. A. Kuzmin, et al. “Stepwise large genome assembly approach: a case of Siberian larch (*Larix sibirica* Ledeb),” BMC bioinformatics, vol.20, no. 1, pp. 37, 2019.

[10] Z. Bao, S. R. Eddy, “Automated de novo identification of repeat sequence families in sequenced genomes,” Genome research, vol. 12 no.8, pp. 1269-1276, 2002.

[11] A. L. Price, N. C. Jones, P. A. Pevzner. “De novo identification of repeat families in large genomes,” Bioinformatics, vol. 21, no. suppl_1, pp. i351-i358, 2005.

[12] G. Abrusán, N. Grundmann, L. DeMester, W. Makalowski, “TEclass — a tool for automated classification of unknown eukaryotic transposable elements,” Bioinformatics, vol. 25, no. 10, pp. 1329-1330, 2009.

[13] S. El-Gebali, et al. “The Pfam protein families database in 2019,” Nucleic acids research, vol. 47 no. D1 pp. D427-D432, 2018.

[14] J. Mistry, R. D. Finn, S. R. Eddy, A. Bateman, M. Punta, “Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic acids research,” vol. 41, no. 12, pp. e121-e121, 2013.

DOI 10.18699/BGRS/SB-2020-043

WebMCOT web-service for prediction of co-occurred DNA motifs in ChIP-seq data

Aleksey M. Mukhin
Shared Access Center Bioinformatics
Institute Cytology and Genetics
SB RAS
Novosibirsk, Russia
mukhin@bionet.nsc.ru

Victor G. Levitsky
Laboratory of Evolutionary
Bioinformatics and Theoretical
Genetics
Institute Cytology and Genetics
SB RAS
Novosibirsk, Russia
levitsky@bionet.nsc.ru

Dmitriy Y. Oshchepkov
Laboratory of Evolutionary
Bioinformatics and Theoretical
Genetics
Institute Cytology and Genetics
SB RAS
Novosibirsk, Russia
diman@bionet.nsc.ru

Sergey A. Lashin
Sector for computer analysis and
simulations of biological systems
Institute Cytology and Genetics
SB RAS
Novosibirsk, Russia
lashin@bionet.nsc.ru

***Abstract* — Regulation of eukaryotic gene expression is controlled by specific regulatory proteins transcription factors. Binding sites of transcription factors are called motifs. Conventionally, genome-wide annotation of motifs performed with chromatin immunoprecipitation followed by massive sequencing (ChIP-seq) approach. The term composite element (CE) implied two closely located and frequently occurred in genomic DNA motifs. CEs contain two overlapping or spacing motifs. Earlier we proposed Motif Co-Occurrence Tool (MCOT) package that is capable of (a) predicting CEs with both overlapping and spacing of motifs in a single ChIP-seq dataset; (b) all predicted CEs were classified by conservation of both participant motifs. This work presents a web interface WebMCOT for MCOT package.**

Keywords* — *web-service, transcription factors, site binding, motifs

Motivation and aim

Chromatin immunoprecipitation followed by massive sequencing (ChIP-seq) is the most popular approach for experimental analysis of gene expression regulation by transcription factors (TFs) through mapping of their potential binding sites (motifs). Since TF binding conventionally is cooperative, the term composite element (CE) denotes the specific combination of two motifs that exhibit qualitatively new properties, thus resulting in a new pattern of gene regulation; the motifs participating in a CE may be overlapping, or separated by a relatively short spacer [1].

We recently developed the Motif Co-Occurrence Tool (MCOT) package [2] that is capable of (a) predicting CEs with both overlapping and spacing of motifs in a single ChIP-seq dataset; (b) all predicted CEs were classified to conservation of both participant motifs, i.e. CEs with substantial asymmetry between two participating motifs were also predicted. Hence, MCOT approach have additional advantages as compared with the previous tools [3, 4].

MCOT is console program working on Windows and Linux operation systems without graphical user interface (GUI). However, SpaMo tool for prediction of CEs with a spacer [3] have web-interface and service for anyone to apply this approach remotely, so users can use only browser. WebMCOT provides web-service for application of MCOT package.

Materials and methods

For web-service development, we use different computer technologies. Firstly, we make client-server architecture through REST API interface between client and server. HTTP requests and responses are compiled with JSON schema. In addition, the “backend” consists of two parts: web-server and worker. They can be on one or more host machines connected by Redis broker messenger.

The JavaScript language, the Vue.JS library for creating the user interface and the Material Design interface style form Google, make frontend site. Backend is made by Python language with Flask library for creating REST API interface and Celery library for running background tasks.

The calculation kernel is written on the C++ language. Also R and Python scripts are developed for visualization.

Results

In this work, we developed web-service WebMCOT for prediction of co-occurred DNA motifs in ChIP-seq data. The web-service is available on the site: https://webmcot.sysbio.cytogen.ru. The help page was implemented (https://webmcot.sysbio.cytogen.ru/help). The screenshot of WebMCOT interface is provided in Fig. 1. Panel A shows the page that allows to enter input data (ChIP-seq peaks and motifs). The rest panels represent the main output data of WebMCOT. In particular, panel B shows the significance of various structural types of CEs (see the help page for details); panel C shows the distribution of mutual locations and orientations for predicted CEs; panel D shows the relative abundances of various CE types with different ratios of conservation of two participating motifs.

ACKNOWLEDGMENT

The work is partially supported by RFBR, project no. 18-29-13040 “Meta-analysis of genome-wide data based on discovery and annotation of various structural-functional transcription factor binding sites participating in composite elements” and by Budget Project No. 0259-2019-0008.

REFERENCES

[1] Kel O.V. et al. (1995) A compilation of composite regulatory elements affecting gene transcription in vertebrates, Nucleic Acids Res., 23(20):4097-4103.

[2] Levitsky V. et al. (2019) A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package, Nucleic Acids Res., 47(21):e139.

[3] Whitington T. et al. Inferring transcription factor complexes from ChIP-seq data. // 2011, Nucleic Acids Res., 39(15):e98..

[4] Jankowski A. et al. (2014) TACO: a general-purpose tool for predicting cell-type-specific transcription factor dimers, BMC Genomics, 15:208.

[5] Heinz S. et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell., 38(4):576-589.

[6] Kulakovskiy I.V. et al. (2018) HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Res., 46(D1), D252-D259.

[7] Ridinger-Saison M. et al. (2012) Spi-1/PU.1 activates transcription through clustered DNA occupancy in erythroleukemia. Nucleic Acids Res., 40(18):8927–8941.

[8] Pang S.H. et al. (2016) PU.1 cooperates with IRF4 and IRF8 to suppress pre-B-cell leukemia. Leukemia. 30(6):1375–1387.

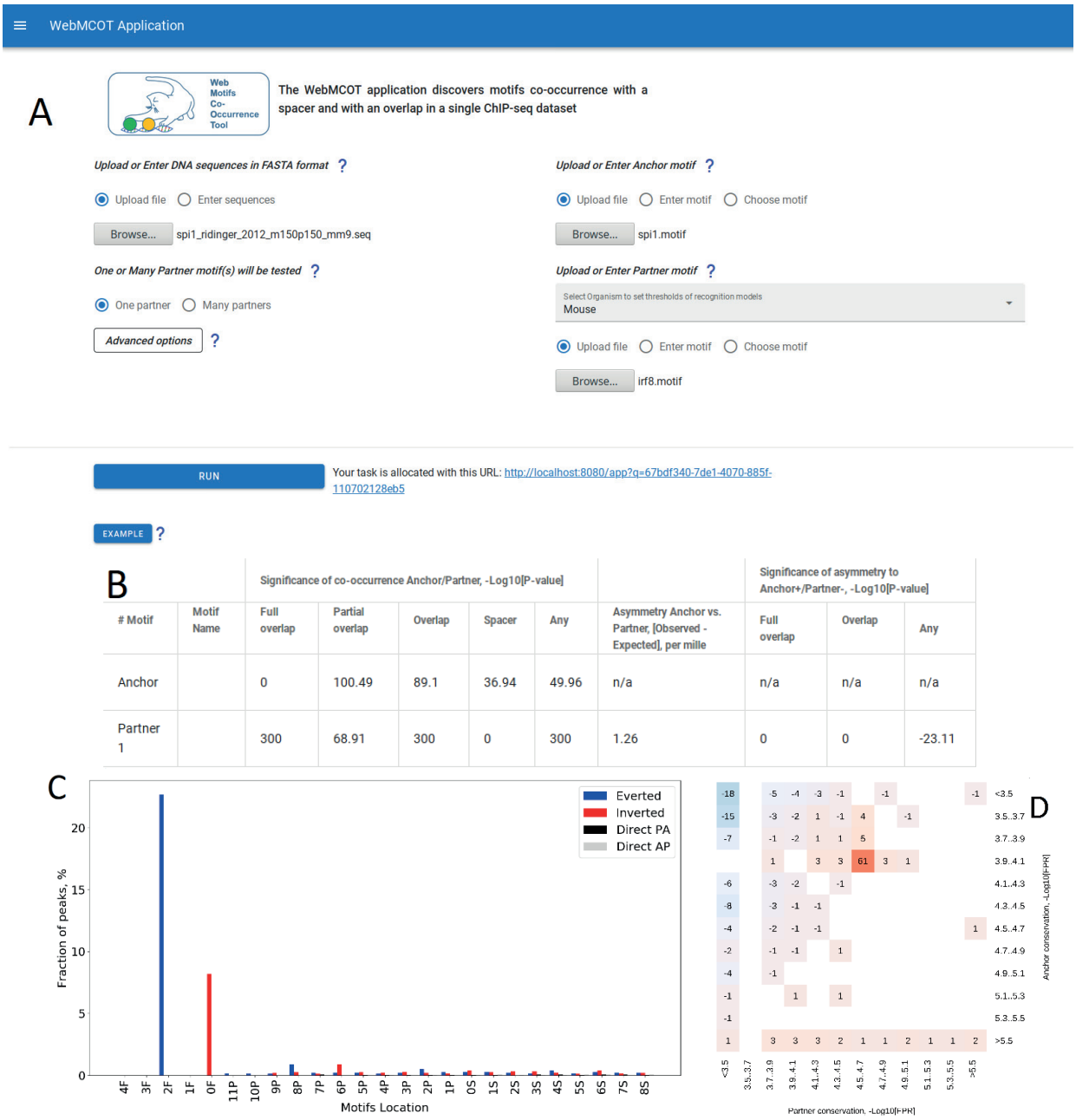


Fig. 1. WebMCOT interface for applying MCOT pipeline. A. The application page for the input data specification (nucleotide sequences of peaks and motifs). B. The summary for the significances of various structural types of predicted CRs. C. Histogram of frequencies of various structural types of CEs according to mutual locations and orientation of motifs. D. The relative abundances of various types of CEs with different ratios of conservation of two motifs. Axes X/Y show the conservation of two motifs, i.e. $-\mathrm{Log}_{10}$(FPR) value (here FPR denotes the false positive rate that respects to predicted hits). The color shows the difference between observed (peaks) and expected (permuted sequences) relative abundance of CEs with specific conservation levels. The Anchor SPI1 and Partner IRF8 motifs were derived from Homer de novo motif search [5] and Hocomoco database [6]. ChiP-Seq peaks were taken from [7]. The previous study [8] have shown that SPI1 and IRF8 cooperate to regulate early B cell development

DOI 10.18699/BGRS/SB-2020-044

Hydroxymethylation changes during early embryonic development in zebrafish

Artem Nedoluzhko
Faculty of Biosciences and Aquaculture
Nord University
Bodø, Norway
artem.nedoluzhko@nord.no

Paula Berrutti
Faculty of Biosciences and Aquaculture
Nord University
Bodø, Norway
paula.d.berrutti@nord.no

Igo Guimarães
Departamento de Produção Animal
Universidade Federal de Goiás
Goias, Brazil
igoguimaraes@ufg.br

Ioannis Konstantinidis
Faculty of Biosciences and Aquaculture
Nord University
Bodø, Norway
ioannis.konstantinidis@nord.no

Igor Babiak
Faculty of Biosciences and Aquaculture
Nord University
Bodø, Norway
igor.s.babiak@nord.no

Jorge M.O. Fernandes
Faculty of Biosciences and Aquaculture
Nord University
Bodø, Norway
jorge.m.fernandes@nord.no

***Abstract* — A large number of studies in organisms across the eukaryotic phyla have shown that DNA 5-methylcytosine (5mC) modification is one of many mechanisms that suppress gene expression.**

Keywords* — *zebrafish, Danio rerio, embryo, development, hydroxymethylation

Cytosine hydroxymethylation has been described during the past years; this DNA modification is increasingly recognized as an important component of epigenetic regulation in eukaryotes.

In the present study, we investigated if hydroxymethylation may be involved in fish embryonic development and demonstrated for the first time at a genome-wide level and single nucleotide resolution the hydroxymethylome changes during zebrafish (*Danio rerio*) embryogenesis from one-cell stage to hatching. DNA hydroxymethylation was profiled by reduced representation hydroxymethylation profiling (RRHP), as shortly described in Fig. 1.

Taken together with recently published data on 5-methylcytosine (5mC) modification events in *D. rerio*, our data unveil a new role for DNA hydroxymethylation in epigenetic regulation of fish embryonic development.

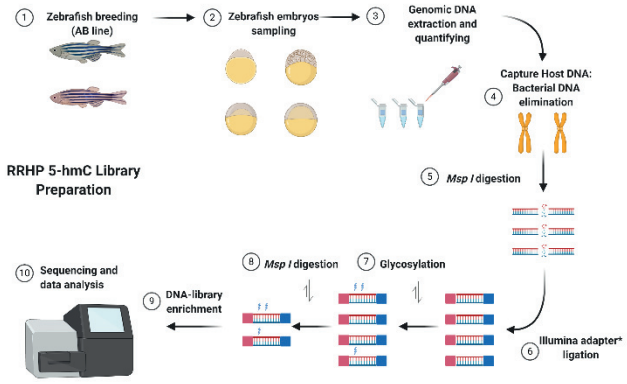


Fig. 1. An overview of the reduced representation hydroxymethylation profiling (RRHP) pipeline that has been used

ACKNOWLEDGMENT

This study has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [grant agreement no 683210] and from the Research Council of Norway under the Toppforsk programme [grant agreement no 250548/F20]. Fedor Sharko was supported by the RFBR (Russian Foundation for Basic Research) Grant no 19-04-00033.

DOI 10.18699/BGRS/SB-2020-045

Expression of DNA reparation genes in anhydrobiotic insect *Polypedilum vanderplanki*

Alexander Nesmelov
IFMB KFU, Kazan, Russia
nesmelov@gmail.com

Sabina Kondratyeva
IFMB KFU, Kazan, Russia
sabinakondr@gmail.com

Taisiya Voronina
IFMB KFU, Kazan, Russia
vorotaisiya@gmail.com

***Abstract* — Anhydrobiosis is survival of complete body desiccation, tightly linked to the ability to cope with a massive DNA damage. However, in anhydrobiotic midge *Polypedilum vanderplanki* this ability is not accompanied by a huge induction of genes of DNA reparation system which is typical for many other anhydrobiosis-related genes in this insect.**

Keywords* — *anhydrobiosis, Polypedilum vanderplanki, DNA reparation, gene expression

Motivation and aim

Motivation

P. vanderplanki is the most complex animal able to survive body desiccation. Its ability to cope with extraordinary DNA damage is an interesting phenomenon for basic science and may serve as a model for the development of dry preservation technologies.

Aim

Earlier we have shown a presence of massive DNA damage in *P. vanderplanki*, performed genome assembly and gene annotation in this insect [1, 2]. Here we aimed to identify genes of different DNA reparation systems known for Drosophila melanogaster and reveal what are the DNA reparation pathways most activated in response to desiccation.

Methods

We extended existing gene annotation in *P. vanderplanki* by the additional search for genes of DNA reparation system. We identified orthologs of the genes of different DNA reparation pathways known for D. melanogaster using OrthoFinder 2.2.7. Gene models for identified *P. vanderplanki* orthologs were estimated using Nanopore reads and EST database. Using RNA-seq data, we studied the expression of identified DNA reparation genes in *P. vanderplanki* in response to desiccation and across different life stages. Employing WGCNA analysis, we also analyzed the membership of genes of DNA reparation system in different modules of the gene coexpression network.

Results

In *P. vanderplanki* DNA reparation genes are typically expressed at a lower level and are less induced in response to desiccation, in comparison to previously identified genes related to anhydrobiosis. Most of these genes are not included in the "key" module of gene network associated with anhydrobiosis. Moreover, most of them are not assigned to any existing module, suggesting that their regulation is perturbed during anhydrobiosis induction. Among different DNA reparation subsystems, nucleotide excision repair seems to be most activated since it contains highest number of genes, significantly upregulated on a course of anhydrobiosis.

Acknowledgment

Supported by the Russian Science Foundation grant No. 19-74-00133.

References

[1] Gusev O.A. et al. (2010) Anhydrobiosis-associated nuclear DNA damage and repair in the sleeping chironomid: linkage with radioresistance. PLoS One. 5(11):e14008

[2] Gusev O.A. et al. (2014) Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. Nature Communications 5:4784

DOI 10.18699/BGRS/SB-2020-046

New approach to genome-wide automated inference of bacterial transcription factor binding sites

Yevgeny Nikolaichik
Department of Molecular Biology
Belarusian State University
Minsk, Belarus
nikolaichik@bsu.by

Pavel Vychik
Department of Molecular Biology
Belarusian State University
Minsk, Belarus
p.vychik@gmail.com

***Abstract* — We describe a method for identification of transcription factor binding sites (TFBS) in bacterial genome sequences. The method is based on two bits of information: (i) specific contacts between amino acid residues of DNA-binding domains and DNA nitrogen bases deduced from 3D structures of DNA-protein complexes and (ii) transcriptional units' organisation inferred from genome annotation. The method can be useful for correct application of experimentally verified TFBS motifs to new genomes and also for discovering TFBS motifs *de novo*. Method was verified with known *Escherichia coli* TFBSs and evaluated during genome-wide TFBS inference in *Pectobacterium spp*. which showed its high efficiency.**

Keywords* — *transcription factor binding sites; operator; regulon; genome annotation; motif discovery

Motivation and aim

Motivation

Current approaches to bacterial genome annotation are mostly restricted to coding sequences, rRNA, tRNA and in some cases ncRNA genes. None of the popular pipelines can identify regulatory elements (transcription factor binding sites, promoters and terminators), which makes functional annotation incomplete and reduces the usability of genome sequences.

Aim

To develop a reliable method, based on experimental data, for automated inference of transcription factor binding sites in bacterial genomes.

Methods

Transcription factors (TFs) were classified according to models of DNA-binding domains present in PFAM [1]. Protein and DNA motif searches were performed with hmmsearch and nhmmer from the HMMER package [2]. Transcription factor binding site (TFBS, also called operator) motif identification and comparison were done with programs from the MEME suite package [3]. Clustering of the regulatory regions to reduce redundancy was done with MeShClust [4]. Transcriptional terminators were identified with TransTermHP [5]. Pipeline glue code, regulatory region analysis, operon inference logic and graphical interface were written in Xojo and integrated into a previously developed SigmoID program [6].

Genome sequences of *Escerichia coli* K-12 substr. MG1655 (accession code NC_000913), *Pectobacterium atrosepticum* 21A (CP009125) and *P. versatile* 3-2 (CP024842) were used for testing TFBS inference algorithms.

Results

Sahota and Stormo have suggested using the information on 3D structures of TF-operator complexes for *de novo* inference of TFBS motifs [7]. The idea is based on (i) conservation of positions of amino acid residues within DNA binding domain (critical residues) that directly contact DNA bases and (ii) regulation by most TFs of their own or neighbouring genes.

We extend this idea by attaching critical residue information (CR-tag) to experimentally verified TFBS profiles. As CR tag specifies the amino acid residues making specific contacts with DNA bases within an operator, TFs with the same CR-tag can recognise the same operators. This approach allows to correctly apply experimentally determined TFBS profiles available from such databases as RegulonDB and CollecTF to genomes with no experimental TFBS data. If a genome being analysed encodes a TF with a CR-tag identical to the one in an experimentally characterised TF than a TFBS profile for the characterised TF can be safely applied to finding TFBSs in the target genome.

The success of the CR-tag based approach described above depends on the representativeness of the motif library. Its current version contains over 2000 profiles and is based on RegulonDB and CollecTF data supplemented with a selection of profiles from RegPrecise and few profiles collected from the literature. This library appears to have profiles with CR-tags matching about half of TFs of an average enterobacterium. CR-tag matches are less frequent outside *Enterobacteriales* and this frequency drops to just 1-2 % of total TF number for Gram-positive bacteria. However, a modified *de novo* approach to TFBS motif identification described below could be applied in such cases.

We have significantly improved the original *de novo* TFBS motif inference algorithm of Sahota and Stormo (Fig. 1). Slow protein database searches were replaced with fast CR-tag lookup tables, regulatory region extraction was improved (first 50 bp of ORFs and divergons are now taken into the account), redundancy elimination was improved by two clustering stages (TFs and their regulatory regions are both clustered and representatives of each cluster proceed to the next stage), motif finding with meme is run in two modes (any number and zero or one per sequence). Finally, a comparison step was added to aid in choosing the correct motif among several ones usually found. The last comparison step of the pipeline is especially important since it provides solid criteria for choosing the correct motif out of several possible ones. When testing the pipeline with known TFs, we have usually found that correct TFBS motif not only matched one or few other motifs for TFs from the same family with sufficiently low p-value, but these TFs also had similar CR-tags. Running the improved pipeline with *E. coli* genome demonstrated that it finds known motifs for over 90% of TFs that respect method assumptions.

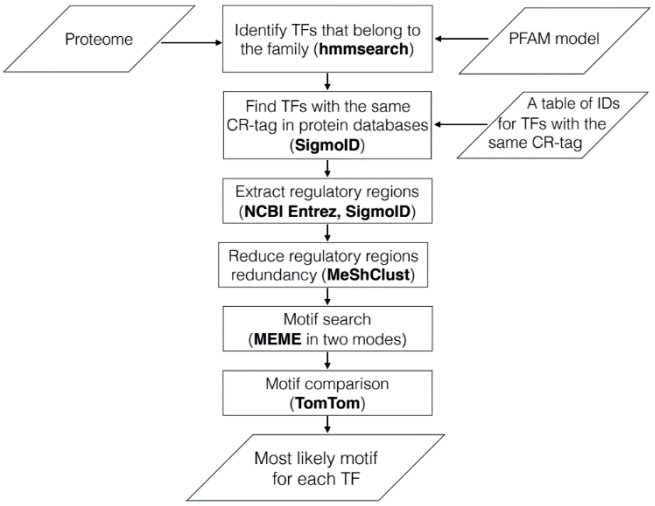


Fig. 1. *De novo* TFBS inference pipeline

Complications with the generally quite efficient *de novo* TFBS inference protocol may arise due to several factors: the absence of 3D structures of TF-operator complexes, presence of two DNA-binding domains in a TF (as in most TFs from the AraC family), indels within CR-tag region, absence of autoregulation (quite common in the LuxR family), the formation of heterodimers with other TFs and so on. To estimate the efficiency of the *de novo* protocol with a non-model organism, we have performed a full scan of *Pectobacterium atrosepticum* genome. Out of 258 TFs found by hmmscan with PFAM models, 164 TFs could be used for CR-tag based *de novo* TFBS motif inference procedure. Based on CR-tag matches to known motifs, motif comparison and correspondence to motif features, typical for TF families (like T-N11-A pattern for LysR family, direct repeats for LysR family and so on) and few *P. atrosepticum* motifs characterised experimentally, correct motifs were found for 124 (76% of the analysed, 48% of total) TFs.

We also tried to evaluate the predictive power of TFBS inference while looking for new regulatory circuits during genome-wide TFBS analysis. We scanned *P. atrosepticum* and *P. versatile* genomes with the *de novo* inferred TFBS profiles to find regulons controlled by each TF. Since pectobacteria are plant pathogens, special attention was paid to TFs that change their expression levels in planta according to available RNA-seq data [8, 9]. Few interesting observations from this analysis are listed below. LacI family TF binding site analysis combined with expression levels changes reported by RNA-seq allowed to functionally link multiple beta-glucoside utilisation loci scattered across the genome; LuxR and OmpR family analysis revealed unexpected new details of quorum sensing regulation, LysR family analysis found new global regulator of secondary metabolism, etc. We are currently verifying these *in silico* predictions experimentally.

Conclusions

The method of TFBS identification described here is universal (can be applied to any genome sequence of any bacterium) and can be expected to find up to half of TFBSs present in a genome with minimal user intervention. We expect method efficiency to improve if pipeline parameters are fine-tuned for a particular genome and TF family and with a further extension of CR-tag based motif library. Our genome-wide TFBS inference approach is capable of discovering unknown regulatory circuits. We advocate its wide use for generating testable hypotheses for further experimental research of bacterial regulatory networks. Source code and executables of the user-friendly GUI application implementing the method described here are available from github.com/nikolaichik/SigmoID.

Acknowledgment

This work was supported by a grant from Belarusian Republican Foundation for Basic Research and the State program "Biotechnology".

References

[1] R. D. Finn et al., “The Pfam protein families database: towards a more sustainable future,” Nucleic Acids Res, vol. 44, no. D1, pp. D279–D285, Jan. 2016, doi: 10.1093/nar/gkv1344.

[2] S. R. Eddy, “Accelerated Profile HMM Searches,” PLoS Comput Biol, vol. 7, no. 10, p. e1002195, Oct. 2011, doi: 10.1371/journal.pcbi.1002195.

[3] T. L. Bailey, J. Johnson, C. E. Grant, and W. S. Noble, “The MEME Suite,” Nucl. Acids Res., vol. 43, no. W1, pp. W39–W49, Jan. 2015, doi: 10.1093/nar/gkv416.

[4] B. T. James, B. B. Luczak, and H. Z. Girgis, “MeShClust: an intelligent tool for clustering DNA sequences,” Nucleic Acids Res, vol. 46, no. 14, pp. e83–e83, Aug. 2018, doi: 10.1093/nar/gky315.

[5] C. Kingsford, K. Ayanbule, and S. Salzberg, “Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake,” Genome biology, vol. 8, no. 2, p. R22, 2007.

[6] Y. Nikolaichik and A. U. Damienikan, “SigmoID: a user-friendly tool for improving bacterial genome annotation through analysis of transcription control signals,” PeerJ, vol. 4, p. e2056, May 2016, doi: 10.7717/peerj.2056.

[7] G. Sahota and G. D. Stormo, “Novel sequence-based method for identifying transcription factor binding sites in prokaryotic genomes,” Bioinformatics, vol. 26, no. 21, pp. 2672–2677, Nov. 2010, doi: 10.1093/bioinformatics/btq501.

[8] V. Gorshkov et al., “Transcriptome profiling helps to identify potential and true molecular switches of stealth to brute force behavior in Pectobacterium atrosepticum during systemic colonization of tobacco plants,” European Journal of Plant Pathology, May 2018, doi: 10.1007/s10658-018-1496-6.

[9] D. Bellieny-Rabelo, C. K. Tanui, N. Miguel, S. Kwenda, D. Y. Shyntum, and L. N. Moleleki, “Transcriptome and Comparative Genomics Analyses Reveal New Functional Insights on Key Determinants of Pathogenesis and Interbacterial Competition in Pectobacterium and Dickeya spp.,” Appl. Environ. Microbiol., vol. 85, no. 2, pp. e02050-18, Jan. 2019, doi: 10.1128/AEM.02050-18.

DOI 10.18699/BGRS/SB-2020-047

High performance pipeline for the calculation of polygenic risk scores

Arina Nostaeva
Laboratory of Recombination and Segregation Analysis
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
avnostaeva@gmail.com

Tatiana Shashkova
Laboratory of Recombination and Segregation Analysis
Institute of Cytology and Genetics
Novosibirsk, Russia

Sodbo Sharapov
Laboratory of Recombination and Segregation Analysis
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Yakov Tsepilov
Laboratory of Recombination and Segregation Analysis
Institute of Cytology and Genetics
Novosibirsk, Russia

Yurii Aulchenko
Laboratory of Recombination and Segregation Analysis
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
PolyKnomics
's-Hertogenbosch, The Netherlands

Lennart C. Karssen
PolyKnomics
's-Hertogenbosch, The Netherlands

***Abstract* — A polygenic risk score (PRS) is a value that reflects a person's predisposition to a disease or any other trait which can (partly) be explained by genetic inheritance. PRSs are often used in reports provided by genetic testing companies like 23andMe, Genotek, etc. Another way of using PRSs is to look at the distribution of PRS values for a group of people and compare them, for example, in a case-control study to find case-dependent traits. PRS models are usually based on summary statistics data from genome-wide association studies (GWAS) and take into account the linkage disequilibrium (LD) structure. We have created a pipeline for high performance PRS calculations across many traits present in the GWAS-MAP platform. The pipeline only requires individual-level data and provides the ability to select a list of traits. This pipeline will be helpful for scientific groups, working with large amounts of individual genotype data, as well as for individuals with their own personal genotype data.**

Keywords* — *genome-wide association study, summary statistics, GWAS, polygenic risk score, linkage disequilibrium, singlenucleotide polymorphisms, GWAS-MAP, database

Introduction

In the standard approach, a PRS is a linear combination of simple linear regression effect size estimates and allele counts at single-nucleotide polymorphisms (SNPs) that are selected via marker pruning coupled with a *p*-value threshold. Thus, a PRS is the attempt to obtain a single value – the score – for each person and evaluate this person's location – the risk – in the distribution of these values from a sample for a given polygenic trait. For example, for a given disease a PRS shows, for a given person, how large the risk of the disease is relation to other people [1]. Analysis of PRSs requires both genotype data of target individuals as well as GWAS results. GWAS data is needed to obtain effects sizes for further PRS calculation. GWAS give the marginal effect size for each SNP estimated by simple regression models and does not take into account the LD structure. Therefore, in order to construct a PRS model SNP effects should be re-estimated based on the genetic architecture. In case GWAS data is not available in the correct format or there are problems with evaluating the LD structure, the calculation PRS becomes a difficult procedure in terms of resources and time. We created a user-friendly pipeline on top of the GWAS-MAP platform [2] that combines all necessary pre-processing stages with the final PRS calculation and the ability of using data already available in GWASMAP. It contains basic steps such as the synchronization of individual-level data with summary statistics, the calculation of re-weighted effect estimates from the LD structure and the generation of the PRS. Our high performance pipeline also builds the distribution of polygenic risk scores for a sample of interest and conducts statistical analyses in the case of several target data groups.

Materials and methods

The GWAS-MAP platform provides access to summary statistics data of more than 7000 traits from different domains including complex traits, metabolomics, proteomics and lipidomics. The complete list of traits can be found through its web-interface called PheLiGe (https://phelige.com/). For reweighting the effect size we chose the SBayesR method [3]. This program uses a shrunk LD matrix and Markov Chain Monte Carlo approach for the estimation of the effects sizes. Two LD matrices are available in our pipeline, one obtained from individual data from the 1000G project European populations [4] and one from UK biobank data provided by the SbayesR tool. In addition to the standard approach for calculating PRS, this pipeline also allows other methods to be used such as the average approach in which the PRS of individual i is defined as:

$$PRS_i = \frac{\sum_j \beta_j G_{ij}}{N}, \qquad (1)$$

with β_j the effect size of the j^{th} SNP, G_{ij} the genotype of the j^{th} SNP for the i^{th} individual, and N the number of SNPs included in the PRS of the i^{th} individual; or the standardization approach:

$$PRS_i = \frac{\frac{1}{N}\sum_j \beta_j G_{ij} - \text{Mean}(PRS)}{\text{SD}(PRS)}, \qquad (2)$$

where SD indicates the standard deviation. Thus, a user is given the choice of model used for computing the PRS.

Results

Our PRS calculation pipeline implemented on top of the GWAS-MAP platform provides a quick and clear assessment

of target data without the need for additional processing steps. The program not only provides the user with the necessary data, but also makes it possible to easily use the large number of traits with which the GWAS-MAP platform is constantly updated. Moreover, the option for selecting different models to calculate the PRS makes analyses more flexible. The ability to conduct basic statistical analysis directly in the pipeline allows the user to immediately get results for new data or models. Thus, our pipeline is a tool that a user can access with a minimal set of data and nevertheless get a complete PRS analysis.

ACKNOWLEDGMENT

This work was partly funded by PolyKnomics BV.

REFERENCES

[1] Wray NR, Goddard ME, Visscher PM, "Prediction of individual genetic risk to disease from genome-wide association studies," Genome Res., 17(10):1520-8, October 2007.

[2] Gorev D.D., Shashkova T.I., Pakhomov E., "GWAS-MAP: a platform for storage and analysis of the result of thousands of genome-wide association scans," Bioinformatics of Genome Regulation and Structure/Systems Biology, The Eleventh International Conference, C. 43. 2018.

[3] Luke R. Lloyd-Jones, Jian Zeng, "Improved polygenic prediction by Bayesian multiple regression on summary statistics," Nature Communications, vol. 10, 5086, 2019.

[4] Adam Auton, Gonc¸alo R. Abecasis, "A global reference for human genetic variation," Nature, vol. 526, p. 68–74, 2015.

DOI 10.18699/BGRS/SB-2020-048

Short sequence repeats (SSR) under selection pressure: Cyprinidae fish case study

Mikhail Orlov
ICB RAS, Pushchino, Russia
orlovmikhailanat@gmail.com

Andrey Tykhonov
"Aqua Logo" company group, Moscow, Russia
andrew693@mail.ru

***Abstract* — Short sequence repeats (SSR) were earlier shown to have length well correlated with various genomic hotspots including intense selection pressure, chromosome rearrangement, etc. We approach this by considering SSR sets for numerous Cyprinidae fish representatives. Analysis of length of SSR as well as their oligonucleotide composition revealed difference between domesticated vs free-living fish.**

Keywords* — *SSR, short tandem repeats, Cyprinidae, genetics, selection

Motivation and aim

Motivation

Genes encoding traits under strict selection pressure were shown to have SSR (short tandem repeats) with length well correlated with the quantity of the traits [2]. SSR are capable of rapid mutational changes known as expansions (i.e. trinucleotide expansions) which are connected to the rate of various evolutionary processes [3].

Aim

Cyprinidae representatives are characterized by complex polyploidy, interspecies and interphylum hybridization as well as massive chromosome rearrangement events [1]. Here we describe SSR for 28 Cyprinidae fish (including particularly interesting and important goldfish *Carassius auratus*).

Methods

SSR data were taken from FishMicrosat database [4]. For each set corresponding to a species we have considered lengths and nucleotide composition. Namely, di-, tri- and tetranucleotide frequency was derived. The result was furtherly clusterized using hierarchical Ward's method.

Results

Accordly to the hypothesis we put forward, domesticated fish clustered mostly together. So did the free-living species. The same oligo frequency dataset was analysed using principal component analysis (pcA). This reinforced the result above while also showing significant difference between the two clusters. Indeed, TG/CA dinucleotide presented more in domesticated fish with smaller amount of AG/TC dinucleotide. This mifht suggest the presence of specific SSR region with some relation to genes under evolutionary pressure.

References

[1] Fondon J. W., Garner H. R. // 2004, V. 101(52), P. 18058–18063

[2] K. T. Xie, G. Wang, A. C. Thompson et al. // Science. 2019. V. 363. P. 81–84.

[3] Katsnelson L.B. et al. Boron A., Spoz A., Porycka K. et al. Comparative Cytogenetics. 2014, V. 8(3), P.233–248.

[4] Nagpure N.S., Rashid I., Pati R. et al. BMC Genomics. 2013, V. 14, P. 630.

DOI 10.18699/BGRS/SB-2020-049

Diversity of Cis-elements in response to dioxin in human

Evgenia Oshchepkova
ICG SB RAS, Novosibirsk, Russia
nzhenia@bionet.nsc.ru

Yana Sizentsova
ICG SB RAS, Novosibirsk, Russia
sizentsova.yans@gmail.com

Viktoriya Mironova
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
victoria.v.mironova@gmail.com

***Abstract* — In this work we studied the potential transcriptional regulators which may play role under the action of one of the most dangerous anthropogenic poisons - dioxins. Using limma, edgeR, MetaR, HOMER, Tomtom and MCOT bioinformatics tools for microarray and RNA-seq data analyzing identification of new possible cis-elements involved in the response of cells to dioxins was carried out.**

Keywords* — *dioxin, AhR, transcription regulation, cis-elements, meta-analysis.

Motivation and aim

Motivation

Dioxins, widespread environmental contaminants, cause disturbances in the functioning of the immune, endocrine, and reproductive systems; their exposure leads to developmental pathologies in embrios and newborns (in particular cleft palate), cause a decrease in IQ [1-3]. The cellular receptor to which dioxins are bind (also being the main transcriptional regulator) is known. However, the effect of dioxins has a wide range of changes in the organism, the expression of hundreds of genes are changes. The mechanisms underlying this are not clear.

Aim

Our goal was to identify potential transcription factors (TF) that may be involved in the response of cells to TCDD (2,3,7,8-tetrachlorodibenzo-p-dioxin), in addition to those whose functioning has already been shown in numerous studies. Using microarray and RNA-seq data, as well as bioinformatics tools, we searched for potential cis-elements in genes that systematically changing expression in response to dioxin.

Methods

Package limma [4] for microarray and edgeR [5] for RNA-seq datasets were used to distinguish differentially expressed genes (DEGs), which are changing their expression in response to dioxin. Two types of samples were used: TCDD-treated cell cultures and control. False discovery rate (FDR) $\leqslant 0.2$ and fold-change > 1.3. Microarray and RNA-seq datasets were taken from GEO database. To identify overreprsented potential cis-elements in promoters of DEGs the R package MetaRE [6] and HOMER v4.10 [7] were used. For this the upstream regions [-1500:+1] of 19815 Homo sapiens genes were taken from GENCODE Release v27. Tomtom motif comparison tool [8] was used to comparing overrepresented sequences founded by MetaRE with known transcription factor (TF) cis-elements (HOCOMOCO v11 core database with using the euclidean distance). The hits with an E-value <0.05 were considered as significant matches. MCOT Toolbox [9] was used to identify composite elements, functioning together with aryl hydrocarbon receptor (AhR).

Results

As a result of a meta-analysis of the upstream regions of DEGs that change expression in response to TCDD, hundreds of frequently occurring short (from 5 to 8 nucleotides) sequences were found. Using the Tomtom software package, an analysis of the correspondence of these sequences to the binding sites of known transcription factors was carried out. According to the generally accepted concept of the TCDD action on cells, the main mediator is AhR, which regulates genes by binding to the dioxin-responsive elements (DRE) in DNA. However, AhR binding sites were far from first among the other overrepresented potential cis-elements. The most frequently occurred transcription factors in results of analysis are member of KLF ad SP families. MCOT expectedly shown the presence of AhR-ARNT composite elements. Besides this composite elements with other TFs were found, in particular with KLF, ZNF and E2F families members.

Acknowledgment

Supported by RFBR (18-04-01130) and by the Complex Program for Fundamental Research of SB RAS.

References

[1] Yoshioka W., Tohyama C. (2019) Mechanisms of Developmental Toxicity of Dioxins and Related Compounds. Int J Mol Sci. 20(3). pii: E617.

[2] Lundqvist C., Zuurbier M., Leijs M., Johansson C., Ceccatelli S., Saunders M., Schoeters G., ten Tusscher G., Koppe J.G. (2006) The effects of PCBs and dioxins on child health. Acta Paediatr Suppl. 95:55-64.

[3] Wesselink A., Warner M., Samuels S., Parigi A., Brambilla P., Mocarelli P., Eskenazi B. (2014) Maternal dioxin exposure and pregnancy outcomes over 30 years of follow-up in Seveso. Environ Int. 63:143-8.

[4] Ritchie M.E., Phipson B., Wu D., Hu Y., Law C.W., Shi W., Smyth G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43:e47.

[5] Robinson M.D., McCarthy D.J., Smyth G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26:139-40.

[6] Cherenkov P., Novikova D., Omelyanchuk N., Levitsky V., Grosse I., Weijers D., Mironova V. (2018) Diversity of cis-regulatory elements associated with auxin response in Arabidopsis thaliana. J Exp Bot. 69(2):329-339.

[7] Heinz S., Benner C., Spann N., Bertolino E., Lin Y.C., Laslo P., Cheng J.X., Murre C., Singh H., Glass C.K. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576-89.

[8] Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J., Li W.W., Noble W.S. (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37:W202-8.

[9] Levitsky V., Zemlyanskaya E., Oshchepkov D., Podkolodnaya O., Ignatieva E., Grosse I., Mironova V., Merkulova T. (2019) A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package. Nucleic Acids Res. 47(21):e13

DOI 10.18699/BGRS/SB-2020-050

Analysis of the complete genome sequence of strain Concept-8, the new representative of the genus *Methylococcus*

I.Y. Oshkin
Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences
ig.owkin@gmail.com

K.K. Miroshnikov
Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences
infon18@gmail.com

D.V. Chernushkin
BIOSINTEZ, LLC rs1@gkconcept.biz

N.V. Ravin
Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia
nravin@biengi.ac.ru

V.O. Popov
Research Center of Biotechnology of the Russian Academy of Sciences
vpopov@inbi.ras.ru

V.N. Khmelenina
Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences
khmelenina@ibpm.pushchino.ru

S.E. Belova
Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences
svet-bel@mail.ru

A.V. Beletsky
Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia
mortu@yandex.ru

S.N. Dedysh
Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences
dedysh@mail.ru

S. Y. But
Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences
sergeybut20063@gmail.com

N.S. Khokhlachev
Gazprom VNIIGAZ gagarin-88@yandex.ru

A.V. Mardanov
Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia
mardanov@biengi.ac.ru

N.V. Pimenov
Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences
npimenov@mail.ru

***Abstract* — The complete genome sequence of the thermotolerant obligate methanotroph and a promising single-cell protein producer *Methylococcus* sp. Concept-8 has been. determined and analyzed. The final genome assembly represents a circular chromosome of 3.46 Mbp. The closest phylogenetic relative of strain Concept-8 is the well-studied methanotroph *Methylococcus capsulatus* Bath. ANI value estimated for the genomes of these methanotrophs constituted 88.39% which suggests that strain Concept-8 belongs to a new species of the genus *Methylococcus*. Both genomes contain two identical *rrn* operon copies, two operons *pmoBAC* encoding particulate methane monooxygenase and a single operon *mmoXYBZDC* encoding soluble methane monooxygenase. The sets of genes encoding enzymes of the C1 metabolism are also highly similar. In addition, we identified two prophage regions associated with prophages of the family *Siphoviridae* in the genome of strain Concept-8.**

Keywords* — *methanotrophs, genome analysis, complete genome sequence, single-cell protein

Motivation and aim

Motivation

Aerobic methanotrophs are a subset of methylotrophic bacteria that can utilize methane as a sole source of energy [1,2]. This unique metabolic feature is widely used for production of the single-cell protein (SCP). Large-scale industrial production of SCP from natural gas using *Methylococcus capsulatus* BSV-874 was previously implemented in USSR. Strain Concept-8 was obtained by long-term storage, selection, and purification of an industrial culture of *Methylococcus capsulatus* BSV-874. Concept-8 is considered as a promising producer of SCP as it demonstrates high growth rates and high biomass yield during growth on methane.

Aim

The present study aimed to determine the complete genome sequence of strain Concept-8. We focused on comparison of the genome sequence of strain Concept-8 with that of *Methylococcus capsulatus* Bath in order to define precise phylogenetic position of these methanotrophs and identify shared and distinct metabolic features. In modern biotechnology, comprehensive genome analysis is necessary for defining the metabolic potential and identification of possible ways for the optimization of the metabolic capabilities

Methods

We sequenced the genome of strain Concept-8 using both MinION nanopore (Oxford Nanopore Technologies, Oxford, UK) and Illumina MiSeq technologies (Illumina, Inc., USA). Genome annotation was performed using Prokka [3] and KEGG [4]. Genome-based phylogeny was determined based on GTDB-Tk pipeline [5].

Results

The final hybrid assembly was a circularized chromosome sequence of 3.46 Mbp, with an average G+C content of 63.0%. The closest phylogenetic relative of the strain Concept-8 is *Methylococcus capsulatus* Bath (98.7% similarity of the 16S rRNA gene sequences). Average nucleotide identity value estimated for these methanotrophs constituted 88.39%. This indicates that strain Concept-8 belongs to new species within the genus *Methylococcus*. In addition, we performed phylogenomic analysis using the

Genome Taxonomy Database and GTDB-Tk (https://github.com/Ecogenomics/GtdbTk). It demonstrated that strain Concept-8 and *Mc. capsulatus* Bath clustered together within *Methylococcus* group. Other closest relatives were represented by *Methylocaldum*, *Methylomagnum*, *Methyloterricola* и *Methylogaea*. The compared genomes possess two copies of operons encoding rRNA and membrane methane monooxygenase, as well as one copy of operon *mmoXYBZDC* encoding soluble methane monooxygenase. The ability to grow on methanol is explained by the presence of two methanol dehydrogenases - *MxaFJGIRACKLD* and *XoxF*. The set of genes encoding the enzymes of C1 metabolic pathways were also highly similar. Genes encoding pyrroloquinoline quinone cofactor biosynthesis proteins, tetrahydromethanopterin-linked and tetrahydrofolate-mediated pathways, and NAD-linked formate dehydrogenase were identified. The strain Concept-8 genome encodes four isoforms of glucose-6-phosphate dehydrogenase and two 6-phosphogluconate dehydrogenases/ decarboxylases that participate in oxidative ribulose monophosphate (RuMP) pathway. *Mc. capsulatus* Bath possesses only two isoforms of glucose-6-phosphate dehydrogenase and one 6-phosphogluconate dehydrogenase. A complete set of genes involved in the function of the RuMP pathway for formaldehyde assimilation was present. Both genomes encode indicative serine pathway enzymes, serine-glyoxylate aminotransferase, hydroxypyruvate reductase and 3-glycerate kinase, as well as malyl-CoA lyase and malattiokinase involved in the synthesis of glyoxylate. However, Concept-8 and Bath strains lack PEP carboxylase, one of the key enzymes of the serine pathway. *Methylococcus* sp. Concept-8 genome contains genes encoding the large (*cbbL*) and small *(cbbS*) subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), as well as the *cbbQ* gene encoding a polypeptide with the putative function of Rubisco's post-translational activator. Strain Concept-8 possesses the enzymes necessary for the synthesis of glycogen: 4-α-glucanotransferase, glucose-1- phosphate adenylyltransferase, 1,4-α-glucan-branching enzyme, glycogen synthase and ADP-ribose pyrophosphatase. Similar cluster of genes is identified in the genome of *Mc. capsulatus* Bath. The genome of *Methylococcus* sp. Concept-8 contains 2 regions associated with prophages of the family *Siphoviridae*. The prophage regions found in the genome of *Mc. capsulatus* Bath are also homologous to those from *Siphoviridae* viruses. However, they differ from those found in the genome of strain Concept-8.

ACKNOWLEDGMENT

The research was supported by the Ministry of Science and High Education of Russia, grant No 075-15-2019-1830 of the Federal Research Program, unique project identifier RFMEFI60719X0297

REFERENCES

[1] Hanson R., Hanson T. (1996) Methanotrophic bacteria. Microbiol. Rev. 60: 439–471.

[2] Trotsenko Y.A., Murrell J.C. (2008) Metabolic aspects of aerobic obligate methanotrophy. Adv. Appl. Microbiol. 63: 183–229.

[3] Seemann T. (2014) Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 30: 2068–2069.

[4] Kanehisa M. et al. (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J. Mol. Biol. 428: 726–731.

[5] Parks, D.H. et al. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36: 996

DOI 10.18699/BGRS/SB-2020-051

Genome-wide association study of Parkinson's disease using MAX3 test

Georgii Ozhegov
Kazan Federal University, Kazan, Russia
Novel Software Systems, Ltd., Novosibirsk, Russia
georgii_provisor@mail.ru

Dmitry Poverin
Novosibirsk State Technical University, Novosibirsk, Russia

Sergey Medvedev
Federal Research Center Institute of Cytology and Genetics, Novosibirsk, Russia

Suren Zakian
Federal Research Center Institute of Cytology and Genetics, Novosibirsk, Russia

Yuri Vyatkin
Novosibirsk State University, Novosibirsk, Russia
Novel Software Systems, Ltd., Novosibirsk, Russia

Sergey Postovalov
Novosibirsk State Technical University, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia

Abstract — **Whole exomes for a set of patients with Parkinson's disease (PD) were sequenced to conduct a genome-wide association study (GWAS) using MAX3 test to find novel genomic variants associated with the disease. As a result, several new variants were identified.**

Key words – Parkinson's disease; GWAS; SNP

Introduction

One of the priority task of medical genetics is to discover the genetic basis of the etiology and pathogenesis of Parkinson's disease (PD). In this work, we search for known disease markers among patients with confirmed diagnosis of PD and control group. We also conduct a genome-wide association study (GWAS) to identify further common variants that contribute to disease.

Sample preparation

A sample of 50 patients with diagnosed PD and 10 control individuals was acquired for the study. As a control group, we took examined elderly individuals with no signs of PD and for whom the development of this pathology in the future is considered as unlikely. Whole blood was withdrawn from the patients with further preparation for the sequencing by performing quality check of the libraries and sequencing them on Illumina HiSeq 2500 sequencer in Rapid Run mode. SureSelect XT Target Enrichment System kit was used resulting pair-end reads of 100 b.p. in length.

Data acquisition

The list of germline single nucleotide polymorphisms (SNPs) was obtained with Genomenal NGSWizard software [1]. Original fastq files were quality checked and mapped to the reference human genome GRCh38/hg38. For SNPs discovery GATK v. 4.1.3.0 software package was used with Best Practice guidelines applied [2]. After SNPs and sample quality control, we acquired genotype data from more than 900 thousands SNPs and short indels. All of them were annotated by SnpEff and Annovar software systems. Among all SNPs we selected around 3000 SNPs appeared in PD-associated genes. PD-associated genes were taken from OMIM, ClinVar and MDSGene databases. Selected SNPs were also analyzed with ClinVar and dbSNP databases. We observed two pathogenic SNPs: one appeared in LRRK2 gene and was found in two patients, another appeared in GLUD2 gene and was found in three patients. It should be emphasized that not a single sample from the cohort with an excluded diagnosis had a pathogenic mutation or variant with an uncertain significance (VUS). Thus, the pathogenic genetic variants for Parkinson's disease were found for 5 out of 50 patients.

Genome-Wide Association Studies

For further data analyses, we performed a GWAS. We used the MAX3 test to check the hypothesis of an association between Parkinson's disease and biallelic single nucleotide polymorphisms [3]. The MAX3 test is based on the Cochran-Armitage trend test (CATT). The MAX3 test statistic is calculated as a maximum of the absolute values of CATT statistics for the dominant, recessive, and additive inheritance mode. As a result, the MAX3 test is not much inferior in power to the trend tests in the case of the known inheritance mode, and is robust if the inheritance mode is unknown [4]. To calculate the p-value the MAX3 test the approach proposed in [5] was used. To account for multiple testing, the Bonferroni correction for significance level was performed [6]. As a result, we obtained 83 SNPs in which the p-value is less than 1e-9, while all 10 people from the control group and at least 43 people with Parkinson's disease are present in the sample.

Data analysis

A fraction of mutations is located on chromosome 17 in the locus where gene KCNJ12 is located. According to published data, the change of KCNJ12 protein could be involved in PD pathogenesis [7]. In this regard, we additionally analyzed all SNPs from this locus. 264 mutations we observed, but not a single of them was reported in ClinVar database. Most of them were previously analyzed by some projects like 1000genomes, Topmed, GnomAD, ExAC and PAGE. This helped us to determine global alternative allele frequencies for SNPs. Based on this data we isolated suspicious SNPs with high p-values and low global alternative allele frequencies. We state that this combination of signs may indicate pathogenic and associated with PD SNPs. Eventually we identify 10 SNPs with p-value lower than 7e-6 and global alternative allele frequency low than 9%.

Results and discussion

Around 10 new SNPs associated with PD and not reported previously were identified during the study. Since the sample size was small, we plan to make meta-analysis with results of another GWASes of PD in the future [8].

ACKNOWLEDGEMENTS

The work was supported by RSCF grant No. 19-75-20063.

REFERENCES

[1] Available at https://ru.genomenal.com

[2] Van der Auwera, Geraldine A., et al. "From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline." *Current protocols in bioinformatics* 43.1 (2013): 11-10.

[3] Freidlin, Boris, et al. "Trend tests for case-control studies of genetic markers: power, sample size and robustness." *Human heredity* 53.3 (2002): 146-152.

[4] Postovalov, Sergey, and R. Wyler Metge. "A power comparison of the association tests for genome-wide association studies." 2016 11th International Forum on Strategic Technology (IFOST). IEEE, 2016.

[5] Zang, Yong, Wing Kam Fung, and Gang Zheng. "Simple algorithms to calculate the asymptotic null distributions of robust tests in case-control genetic association studies in R." *Journal of Statistical software* 33.8 (2010).

[6] Bonferroni, Carlo E., C. Bonferroni, and C. E. Bonferroni. "Teoria statistica delle classi e calcolo delle probabilita'." (1936).

[7] Tian, Chuan, et al. "Potassium channels: structures, diseases, and modulators." *Chemical biology & drug design* 83.1 (2014): 1-26.

[8] Chang, Diana, et al. "A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci" *Nat Genet.* 2017. – Vol. 49(10). P. 1511-1516.

DOI 10.18699/BGRS/SB-2020-052

Identification and description of genes with a high mutation frequency in vagal paragangliomas

Vladislav Pavlov
EIMB RAS
Moscow, Russia
vladislav1pavlov@gmail.com

Anastasiya Snezhkina
EIMB RAS
Moscow, Russia
leftger@rambler.ru

George Krasnov
EIMB RAS
Moscow, Russia
gskrasnov@mail.ru

Dmitry Kalinin
A.V. Vishnevsky National Medical Research Center of Surgery
Moscow, Russia
dmitry.v.kalinin@gmail.com

Alexander Golovyuk
A.V. Vishnevsky National Medical Research Center of Surgery
Moscow, Russia
algolovyuk@inbox.ru

Anna Kudryavtseva
EIMB RAS
Moscow, Russia
rhizamoeba@mail.ru

***Abstract* — Vagal paragangliomas (VPGLs) are rare neuroendocrine tumors of the head and neck. Germline and somatic mutations in a number of genes are associated with the development of VPGLs. The MutSigCV algorithm was implemented in a search for genes characterized by a high mutation frequency in VPGLs. For this purpose, we used the previously obtained results of exome sequencing of 8 VPGL samples. Ten genes have been identified (*ZNF717, MUC16, PKD1L2, TTN, MUC3A, MUC5AC, HYDIN, SSPO, FLG, OBSCN*), that can potentially be involved in the development and progression of VPGLs. The involvement of these genes in the VPGL pathogenesis is shown for the first time.**

Keywords* — *vagal paragangliomas, tumor-associated genes, exome, high-throughput sequencing

Motivation and aim

Motivation

Vagal paragangliomas (VPGL) that arise from the vagus nerve paraganglia account for about 13% of all head and neck paragangliomas [1]. VPGLs occur as both sporadic and hereditary tumors; hereditary VPGLs are often associated with mutations in a list of driver genes.

Aim

Earlier, we performed high-performance exome sequencing of 8 VPGLs [2]. Two of them didn't possess a mutation in the known driver genes. To identify new genes that may be involved in the development of VPGL, exome sequencing data was analyzed using the MutSigCV algorithm [3].

Methods

In this work, we used previously obtained data on high-performance sequencing of the exome of 8 samples of VPGL [2]. Exome libraries were prepared with Nextera Rapid Capture Exome Kit (Illumina, USA) [4] and sequenced on a NextSeq 500 System (Illumina) in paired-end mode, 76x2 cycles. Each sample had at least 60 million reads (300× coverage). Sequencing data is available on the NCBI Sequence Read Archive (BioProject PRJNA561073).

After running the bioinformatic analysis pipeline described in our previous paper [5] we analyzed the output using the MutSigCV algorithm.

Results

Using the MutSigCV algorithm, we identified 10 genes with a high mutation frequency for VPGL cases ($q < 0.05$) (*ZNF717, MUC16, PKD1L2, TTN, MUC3A, MUC5AC, HYDIN, SSPO, FLG, OBSCN*). Some of them (*MUC16, MUC3A, MUC5AC, HYDIN)* are associated with a wide range of oncological diseases. The involvement of *ZNF717, PKD1L2, TTN, SSPO, FLG, OBSCN* genes in oncogenesis was shown for the first time. High mutation frequency of *HYDIN* and *MUC* family genes has been previously shown for carotid paragangliomas [5]. At certain loci of the 10 identified genes, mainly single non-synonymous mutations arise that lead to the impairment of the structure and function of the protein.

Acknowledgment

This work was performed using the equipment of the EIMB RAS "Genome" center (http://www.eimb.ru/ru1/ckp/ccu_genome_c.php).

References

[1] Thompson, L. D. R. (2006). World Health Organization classification of tumours: pathology and genetics of head and neck tumours. Ear, Nose & Throat Journal, 85(2), 74–74.

[2] Kudryavtseva A.V., Kalinin D.V., Pavlov V.S., Fedorova M.S., Pudova E.A., Snezhkina A. V et al.. Mutation profiling in eight cases of vagal paragangliomas. BMC Medical Genomics, in press.

[3] Lawrence M.S., Stojanov P., Polak P., Kryukov G.V., Cibulskis K., Sivachenko A., et. al. (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 499, 214--218.

[4] Snezhkina A.V., Lukyanova E.N., Kalinin D.V., Pokrovsky A.V., Dmitriev A.A., Kudryavtseva A.V. et al. (2018) Exome analysis of carotid body tumor. BMC Med. Genomics. 11, 17.

[5] Snezhkina, A.V, Lukyanova, E.N., Fedorova, M.S., Kalinin, D.V., Pudova, E.A., Kudryavtseva, A.V. et al. (2019) Novel genes associated with the development of carotid paragangliomas. Mol Biol(Mosk). Jul-Aug;53(4):613-626.

This work was financially supported by the Russian Science Foundation, grant No. 19-15-00419.

DOI 10.18699/BGRS/SB-2020-053

Disruptive natural selection by male reproductive potential prevents underexpression of the genes encoding proteins on the human Y chromosome as a self-domestication syndrome

Mikhail Ponomarenko
Systems Biology Department
Institute of Cytology and Genetics, ICG SB RAS
Novosibirsk, Russia
pon@bionen.nsc.ru

Dmitry Oshchepkov
Systems Biology Department
Institute of Cytology and Genetics, ICG SB RAS
Novosibirsk, Russia
diman@bionet.nsc.ru

Alexander Osadchuk
Animal Genetics Department
Institute of Cytology and Genetics, ICG SB RAS
Novosibirsk, Russia
osadchuk@bionet.nsc.ru

Irina Chadaeva
Systems Biology Department
Institute of Cytology and Genetics, ICG SB RAS
Novosibirsk, Russia
ichadaeva@bionet.nsc.ru

Dmitry Rasskazov
Systems Biology Department
Institute of Cytology and Genetics, ICG SB RAS
Novosibirsk, Russia
rassk@bionen.nsc.ru

Ludmila Osadchuk
Animal Genetics Department
Institute of Cytology and Genetics, ICG SB RAS
Novosibirsk, Russia
losadch@bionet.nsc.ru

Abstract — **We performed an in silico genome-wide analysis of all SNPs located within 70 bp proximal promoters in front of the all experimentally knowns starts of protein-coding transcripts from human Y chromosome within the framework of the current release #151 of the dbSNP database and GRCh38/hg38 assembly of the human reference genome, which are publicly available using the UCSC Genome Browser. As a result, we first found disruptive natural selection by male reproductive potential preventing underexpression of the Y-linked proteins under this study as if self-domestication would have happened during the human origing and evolution that could cause male fertility disorders as self-domestication syndrome.**

Keywords — reproductive potential, human, Y chromosome, gene, promoter, TATA box, TATA-binding protein, single-nucleotide polymorphism, candidate SNP marker, verification

Background

In population ecobiology, the conception of reproductive potential define the most common vital indicator of a probability estimate to produce and regrow a healthy descendant until his/her reproductive maturity at the best conditions. This term associates lifespan and living conditions of an individual with one another, which correspond to his/her disease susceptibilities encoded by his/her genome and his/her both environment and lifestyle. Female reproductive potential seems to be most investigated in depth, width, and complexity now, whereas the male one has not subjected by equal amount of attention as yet. That is why, in this work we focused our attention on the only human Y chromosome as the male specific part of the human genome and, thus, predicted *in silico* candidate single-nucleotide polymorphism (SNP) markers of male reproductive potential within the framework of the current status of both reference human genome and variome.

Materials and methods

Using the current release #151 of the dbSNP database [1] and GRCh38/hg38 assembly of the human reference genome [2] via publicly available Web-service UCSC Genome Browser [3] together with our earlier created Web-service SNP_TATA_Comparator [4], we examined *in silico* all 1206 SNPs, none of which has been characterized so far regarding any link to human diseases (for brevity, hereinafter: unannotated SNPs) within 70 bp proximal promoters of all 63 Y-linked genes.

Results and Discussions

As a result, we found 261 candidate SNP markers for male reproductive potential, which can statistically significantly change the binding affinity of TATA-binding protein (TBP) for these promoters. Among them, there are candidate SNP markers of spermatogenesis disorders (e.g., rs1402972626), pathoembryogenesis (e.g., rs28378830), adenovirus infection of spermatozoa leading to male infertility via spontaneous abortion (e.g., rs1317376848), azoospermia via testicular degeneration (e.g., rs1420856028), reduced sperm quality (e.g., rs1225019830), increased verbal IQ reducing male reproductive potential (e.g., rs1462000578), pediatric cancer (e.g., rs1483581212), as well as male anxiety damaging family relationships and mother's and children's health (e.g., rs187456378).

We first selectively verified in vitro both absolute and relative values of the analyzed TBP–promoter affinity using electrophoretic mobility shift assay (EMSA), where Pearson's simple linear correlation coefficients between our predictions and measurements were $r = 0.84$ (significance $p < 0.025$) and $r = 0.98$ ($p < 0.025$), respectively. Then, we found a twofold excess of the amount of candidate SNP markers increasing the TBP-promoter affinity relatively those for its decrease, whereas the genome-wide norm [5, 6] consists in four-fold excess of SNP-induced damages of TBP-promoter complexes in comparison with SNP-caused improvements of these complexes [6, 7] ($p < 0.05$ according to binomial distribution). This observed difference in the SNP occurrences in the entire genome as a whole and on chromosome Y in particular points to the natural selection against deficient male-specific expression of proteins that are encoded by this chromosome.

Simultaneously, the occurrences of candidate SNP markers, which can increase and decrease male reproductive potential, seem to be indistinguishable ($p < 0.05$) as if male self-domestication could have happened, with its experimentally known disruptive natural selection, according

to Belyaev's observations made during foxes domestication [9].

Finally, because there is still not enough scientific evidence that this could have happened, we discuss the human diseases associated with candidate SNP markers of male reproductive potential in comparison with the known experimental data and natural observation on differences between wolfs and dogs [10], boars and pigs [11] as well as wild and domestic ducks [12], horses [13], mice [14]. Summing up, we found commonly accepted prototype traits of anthropogenic selection associated with animal domestication for the majority of candidate SNP markers of male reproductive potential except for the only candidate SNP markers for suicide (e.g., rs772325955) and verbal IQ (e.g., rs1393008234), which could be human-specific.

Conclusions

In general, our results obtained within the framework of this work indicate that a decrease in the reproductive potential in men can be a self-domestication syndrome as a manifestation of destructive natural selection against a deficiency of proteins encoded by the human Y chromosome.

ACKNOWLEDGMENTS

The study supervision were supported by project #19-15-00075 from the Russian Science Foundation (for MP and LO, AO). The software development was financed by project #0324-2019-0040-C01 from the Russian Government Budget (for DR). The data compilation was supported by project and #18-34-00496 from Russian Foundation for Basic Research (for IC). The data analysis was supported by was supported by the Russian Federal Science & Technology Program for the Development of Genetic Technologies (for DO). The *in silico* result interpretation was supported by by project #0324-2019-0041-C01 from the Russian Government Budget (for AO).

REFERENCES

[1] S. Sherry, M. Ward., M. Kholodov, J. Baker, L. Phan, E. Smigielski, K. Sirotkin. "dbSNP: the NCBI database of genetic variation". Nucleic Acids Res. vol. 29, pp. 308-311, 2001.

[2] D. Zerbino, S. Wilder, N. Johnson, T. Juettemann, P. Flicek. "The Ensembl regulatory build". Genome Biol. vol. 16: pp. 56. 2015.

[3] M. Haeussler, B. Raney, A. Hinrichs, H. Clawson, A. Zweig, D. Karolchik, J. Casper, M. Speir, D. Haussler, W. Kent. "Navigating protected genomics data with UCSC Genome Browser in a box". Bioinformatics. vol. 31. pp. 764-766. 2015.

[4] M. Ponomarenko, D. Rasskazov, O. Arkova, P. Ponomarenko, V. Suslov, L. Savinkova, N. Kolchanov. "How to use SNP_TATA_Comparator to find a significant change in gene expression caused by the regulatory SNP of this gene's promoter via a change in affinity of the TATA-binding protein for this promoter". Biomed. Res. Int. vol. 2015. pp. 35983004625. 2015.

[5] J. Haldane. "The cost of natural selection". J Genet. vol. 55. pp. 511–524. 1957.

[6] M. Kimura. "Evolutionary rate at the molecular level". Nature. vol. 217. pp. 624-626. 1968.

[7] M. Kasowski, F. Grubert, C. Heffelfinger, M. Hariharan, A. Asabere, S. Waszak, L. Habegger, J. Rozowsky, M. Shi, A. Urban, M. Hong, K. Karczewski, W. Huber, S. Weissman, M. Gerstein, J. Korbel, M. Snyder M. 2010. "Variation in transcription factor binding among humans". Science. vol. 328. pp.: 32-235. 2010.

[8] 1000 Genomes Project Consortium, G. Abecasis, A. Auton, L. Brooks, M. DePristo, R. Durbin, R. Handsaker, H. Kang, G. Marth, G. McVean. "An integrated map of genetic variation from 1.092 human genomes". Nature. vol. 491. pp. 56-65. 2012.

[9] D. Belyaev. "The Wilhelmine E. Key 1978 invitational lecture. Destabilizing selection as a factor in domestication". J Hered. vol. 70. pp. 301-308. 1979.

[10] C. Theofanopoulou, S. Gastaldon, T. O'Rourke, B. Samuels, P. Martins, F. Delogu, S. Alamri, C. Boeckx. "Self-domestication in Homo sapiens: insights from comparative genomics". PLoS One. vol. 12. pp. e0185306. 2017.

[11] F. Almeida, M. Leal, L. Franca. "Testis morphometry, duration of spermatogenesis, and spermatogenic efficiency in the wild boar (*Sus scrofa scrofa*). Biol Reprod. vol. 75. pp. 792-799. 2006.

[12] A. Charuta, H. Markowska-Pliszka, B. Bartyzel, J.Wysocki. Size of heart of the domestic Pekin duck (*Anas platyrhynchos f. domestica*) and wild duck (*Anas platyrhynchos, Linnaeus, 1758*). Acta Sci Pol Med Veterinaria. vol. 4. pp. 11-19. 2005.

[13] P. Librado, C. Gamba, C. Gaunitz, C. Der Sarkissian, M. Pruvos, A Albrechtsen, A. Fages, N. Khan, M. Schubert, V. Jagannathan, A. Serres-Armero, L. Kuderna, I. Povolotskaya, A. Seguin-Orlando, S. Lepetz, M. Neuditschko, C. Theves, S. Alquraishi, A. Alfarhan, K. Al-Rasheid, S. Rieder, Z. Samashev, H. Francfort, N. Benecke, M. Hofreiter, A. Ludwig, C. Keyser, T. Marques-Bonet, B. Ludes, E. Crubezy, T. Leeb, E. Willerslev, L. Orlando. "Ancient genomic changes associated with domestication of the horse". Science. vol. 356. pp. 442-445. 2017.

[14] D. Blanchard, G. Griebel, R. Blanchard. "Mouse defensive behaviors: pharmacological and behavioral assays for anxiety and panic". Neurosci Biobehav Rev. vol. 25. pp. 205-218. 2001.

DOI 10.18699/BGRS/SB-2020-054

Intermediate and high-risk prostate cancer methylation analysis

Anastasiya Andreevna Kobelyatskaya
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
kaa.chel@mail.ru

Elena Anatolevna Pudova
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
pudova_elena@inbox.ru

George Sergeevich Krasnov
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
gskrasnov@mail.ru

Kirill Mikhailovich Nyushko
Urological department
FSBI NMRRC
Moscow, Russia
kirandja@yandex.ru

Boris Yakovlevich Alekseev
Urological department
FSBI NMRRC
Moscow, Russia
mnioi@mail.ru

Anna Victorovna Kudryavtseva
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
rhizamoeba@mail.ru

***Abstract* — Prostate cancer is one of the most important socially significant oncological diseases in men. To select an effective approach to therapy, prostate cancer is stratified into appropriate risk groups based on criteria such as TNM status, Gleason score and the level of prostate-specific antigen (PSA). However, to optimize therapy, additional informative markers are necessary, and the aim of this study is to search for these markers at the level of genome methylation. This work included methylation data of prostate cancer from The Cancer Genome Atlas project. The cohort involved patients with high (23 cases) and intermediate (103 cases) risk. As a result, 1,056 differentially methylated CpG sites were found between high and medium risk groups. CpG most interested sites: cg17687367 (chr13: 79936801), cg26874611 (chr5: 168147884), cg06989693 (chr5: 41409252), cg02226810 (chr6: 1605117), cg07736716 (regulation regions: 85): 85 to 91 (85): 85 858: 858: 858: 85: 86: 85: 86: 85: 86: 85: 86: 85: 86: 85: 85.**

Keywords — risk group, prostate cancer, methylation, TCGA

Introduction

Prostate cancer (PC) is one of the most important socially significant oncological diseases in men.. When disease is detected the question arises as to the choice of treatment tactics [1]. Treatment may consist of surgery, hormone therapy, or both. As a rule, determination of risk group is based on criteria such as TNM parameters, Gleason score, PSA level [2, 3, 4]. Considering above criteria and risk group a therapeutic concept is chosen. However, in practice these criteria are not enough to choose a therapeutic concept [2]. And patients receive either excessive or insufficient therapy. Additional markers of diagnosis are needed to optimize the choice of therapy. The aim of research is to identification differentially methylated CpG sites as possible diagnosis markers PC.

Methods

This work included PC samples methylation data of The Cancer Genome Atlas (TCGA) project. Either high (23 cases) or intermediate (103 cases) risk PC cases are involved into the cohort. Risk groups were determinate based on TNM parameters, Gleason score and PSA level. Differential methylation analysis was performed in statistical environment R using EdgeR and BiSeq packages. The Mann-Whitney test, beta-regression and logistic regression models were used for statistical analysis.

Results

In this study we found out 1056 differentially methylated CpG sites between high and intermediate risk groups. From those we selected the most interested CpG sites that are in regulatory regions: cg17687367 (chr13: 79936801), cg26874611 (chr5: 168147884), cg06989693 (chr5: 41409252), cg02226810 (chr6: 1605117), cg07736716 (chr8: 85379411). Above CpG sites are hypermethylated in intermediate risk group cases. These hypermethylated CpG are covered by regulatory regions such as ENSR00000064172 (promoter flank), ENSR00000776896 (enhancer), ENSR00000753523 (promoter flank), ENSR00000191969 (promoter), ENSR00000332061 (transcription factor binding site), ENSR00000860893 (transcription factor binding site). The results will be used in further validation study.

Acknowledgment

This work was financially supported by the RFBR, grant no. 17-29-06083. This work was performed using the equipment of EIMB RAS “Genome” center (http://www.eimb.ru/ru1/ckp/ccu_genome_c.php).

References

[1] Chang A.J., Autio K.A., Roach M., Scher H.I. (2014) High-risk prostate cancer-classification and therapy. Nat Rev Clin Oncol. 11(6):308-23. doi: 10.1038/nrclinonc.2014.68.

[2] Gerhauser C., Favero F., Risch T., Simon R., Feuerbach L., Assenov Y., et al. (2018) Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. Cancer Cell. 10;34(6):996-1011.e8. doi: 10.1016/j.ccell.2018.10.016.

[3] Hurwitz L.M., Cullen J., Kim D.J., Elsamanoudi S., Hudak J., Colston M., et al. (2017) Longitudinal regret after treatment for low- and intermediate-risk prostate cancer. Cancer. 1;123(21):4252-4258. doi: 10.1002/cncr.30841.

[4] Serrano N.A., Anscher M.S. (2016) Favorable vs Unfavorable Intermediate-Risk Prostate Cancer: A Review of the New Classification System and Its Impact on Treatment Recommendations. Oncology (Williston Park). 30(3):229-36.

DOI 10.18699/BGRS/SB-2020-055

Justification of measures for optimization and prevention with dysplasia of stratified squamous epithelium of the cervix in women of reproductive age

Ra'nokhon Bakhodir qizi Solieva
Department of Obstetrics and Gynecology No. 1
of the ASMI
Andijan, Uzbekistan
solieva_r@mail.ru

Dilfuza Abdullaevna Alieva
Republican specialized scientific and practical medical center
of obstetrics and gynecology
Tashkent, Uzbekistan

***Abstract* — The protection and strengthening of the reproductive health of mothers, children and adolescents is an important and priority area of health care in the Republic of Uzbekistan and is fully supported by state programs, the basis of which is the development of preventive medicine.**

Keywords — folate cycle, cervical dysplasia

Motivation and aim

Motivation

Recent decades have been marked by the rapid development of molecular biology and biochemical sciences, which contributed to the revision of previously existing ideas about various pathologies in the reproductive system as a hormone-dependent and / or iatrogenic state.

Aim

To study the role of certain folate metabolism genes in predicting dysplasia of the stratified squamous epithelium of the cervix.

Increasingly important in the formation of pathological conditions and diseases is given to violations of the processes of remodeling of connective tissue (dysplasia), which can be caused by mutation of genes under the influence of exogenous factors: adverse environmental conditions, inadequate nutrition and stress [1, 4].

The course of pregnancy, childbirth and the postpartum period in 87.5% of cases with dysplasia is complicated by the generalization of connective tissue damage involving the reproductive system in the pathological process [2].

The above facts determine the scientific novelty and theoretical significance of the research work, including the study of the relationship between cytological, morphological, genetic indicators in women with intraepithelial neoplasia of the cervix uteri (CM). Based on a study of the polymorphism of folate metabolism genes in women with dysplasia of varying severity, their role in neoplastic processes will be shown.

Methods

In order to achieve the objectives, it is planned to examine women aged 25-49 years, divided into groups: patients with cervical dysplasia (CIN) of various degrees and a group that includes practically healthy women.

All examinations will be carried out on the basis of the Andijan Regional Perinatal Center in the Department of Gynecology, and on the basis of the scientific advisory clinic of Republican specialized scientific and practical medical center of obstetrics and gynecology and on the basis of the Department of Molecular Medicine and Cellular Technologies of the Research Institute of Hematology and Blood Transfusion of the MzRUz.

Results

Discussion. Despite the large number of studies aimed at developing methods for the early diagnosis of CMM diseases, treatment methods that ensure the absence or minimum number of complications, the carcinogenic potency of background CMM diseases remains unclear. This is where the reserve is located, which allows improving the diagnosis of early manifestations of precancer [3].

Acknowledgment

With the support of the Ministry of Health of the Republic of Uzbekistan (24-02-2020, No. 8n-d / 89), Andijan State Medical Institute (No. 11-11-2019)

References

[1] Mayorov M.V. Folic acid: when, why, how much? [Electronic resource]. - Access mode: http://lib.komarovskiy.net/folievaya-kislota-kogda-zachem-skolko.html.

[2] Shikh E.V., Makhova A.A. Vitamin and mineral complex during pregnancy. M.: GEOTAR-Media; 2016.349 s.

[3] Shikh E.V., Makhova A.A. Questions of choosing the form of folate for the correction of folate status. Obstetrics and gynecology. 2018, No. 8, pp. 33-40.

[4] Botto L.D. et al. International retrospective cohort study of neural tube defects in relation to folic acid recommendations: are the recommendations working? / L.D. Botto, A. Lisi, E. Robert-Gnansia // BMJ. - 2005. - Vol. 330, N 7491. - P. 5

DOI 10.18699/BGRS/SB-2020-056

Genetic mapping of QTLs controlling the ISIAH hypertensive rat behavior in an open field tests

Olga Redina
Laboratory of Evolutional Genetics,
Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
oredina@bionet.nsc.ru

Svetlana Smolenskaya
Laboratory of Evolutional Genetics,
Wheat Genetics Laboratory,
Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
svsmol@ngs.ru

Arcady Markel
Laboratory of Evolutional Genetics,
Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
markel@bionet.nsc.ru

Abstract — **Genetic mapping of QTLs controlling the ISIAH hypertensive rat behavior in an open field tests evaluating research behavior, the level of anxiety and motor activity of animals was performed. The work was carried out using two groups of male hybrid rats F2(ISIAHxWAG) of different ages - 3-4 months (n = 103) and 6 months (n = 126) obtained by crossing hypertensive ISIAH and normotensive WAG rats whose behaviors are significantly different in all six tests in an open field taken in the analysis: the latency, the motor activity at 1-st minute of the test, motor activity on the periphery, grooming on the periphery, rearing on the periphery, defecation score. In both groups of rats, genetic loci associated with the studied traits of behavior were found and the common loci associated with both the behavioral traits and traits related to the hypertension development were identified. The results of our work suggest that genes that control behavioral patterns of ISIAH rats can either be closely linked to genes involved in the control of traits associated with the hypertensive status of these rats, or, in some loci, may have a pleiotropic effect on both behavioral traits and on the traits associated with the hypertensive status of ISIAH rats.**

Keywords — QTL analysis, behavior, open field tests, hypertensive ISIAH rats

Introduction

The increased stress sensitivity of the mammalian organism is associated with its special neuroendocrine status, which affects behavior, including motor activity, anxiety, changes in cognitive ability, and many other physiological functions, in particular, the impaired control of the blood pressure and development of hypertension [1-3]. The ISIAH rats simulating a stress-sensitive form of arterial hypertension are characterized by increased activity of the hypothalamic-pituitary-adrenal and sympathetic adrenal systems, changes in the functions and morphology of a number of peripheral target organs (brain, kidneys, adrenal glands, heart), as well as specific behavior in unfamiliar environments [4-6]. The literature suggests that changes in behavior may be a condition concomitant to hypertension. However, the genetic basis of the relationship between behavior and the body's hypertensive status have been poorly understood to date, and such studies have not been conducted on ISIAH rats before.

The aims of current work were 1) to perform genetic mapping (QTL analysis) of ISIAH rat behavior in six open-field tests evaluating research behavior, the level of anxiety and motor activity of animals, and 2) to identify the common loci associated with both the behavioral traits and traits related to the development of hypertension.

Methods

Animals

Two groups of male hybrid rats F_2(ISIAHxWAG) of different ages - 3-4 months (n = 103) and 6 months (n = 126) obtained by crossing rats of the hypertensive ISIAH strain and the normotensive WAG strain were used. The behavior of rats from these two strains is significantly different in all six tests in an open field taken in the analysis: the latency, the motor activity at 1-st minute of the test, motor activity on the periphery, grooming on the periphery, rearing on the periphery, defecation score.

QTL analysis

QTL analysis was performed using 145 polymorphic microsatellite markers in the group of younger rats, and using 149 polymorphic microsatellite markers in the group of older rats. The list of DNA markers and the sequence of primers used are given on the website of the Institute of Cytology and Genetics SB RAS (http://icg.nsc.ru/isiah/en/category/qtl/). Processing the results was carried out using specialized programs MAPMAKER/EXP3.0 and MAPMAKER/QTL1.1. Identification of common loci associated both with the behavior of rats and with other traits associated with the hypertensive phenotype of the ISIAH rats, such as blood pressure levels at rest and under stress, body weight, absolute and relative weights of target organs (kidneys, adrenal glands, heart), the concentration of corticosterone in the plasma at rest and under stress, and the concentration of norepinephrine in the hypothalamus was performed using a bivariate (multivariate) analysis in the QTL Cartographer Version 1.17 software package, JZmapqtl (statgen.ncsu.edu). Empirical significant threshold values for QTLs were determined by a permutation test using 1000 permutations of the original data.

Results

In both groups of rats, genetic loci associated with the studied traits of behavior were found. In the group of rats aged 3-4 months, statistically significant associations were found for such traits as motor activity at the 1st minute of the test at the locus on chromosome 16 (marker D16Rat58), motor activity at the periphery at the loci on chromosomes 2 and 16 (markers D2Rat157 and D16Rat32*), grooming at the periphery on chromosome 20 (marker D20Rat30) and latent period on chromosome 11 (marker D11Rat50*). In the group of rats at the age of 6 months, the most highly reliable associations were found for defecation on chromosomes 2 and 14 (markers D2Rat194* and D14Rat5*). Asterisks mark loci whose association with the trait is described for the first time.

A search for common loci showed that some loci associated with behavior overlap with loci associated with one or more of the analyzed traits characterizing the hypertensive status of ISIAH rats: blood pressure under stress, body weight, relative weight of the adrenal glands, absolute and relative kidney weight, plasma corticosterone concentration at rest and under stress, and the concentration of norepinephrine in the hypothalamus.

Conclusion

The results of our work suggest that genes that control behavioral patterns of ISIAH rats can either be closely linked to genes involved in the control of traits associated with the hypertensive status of these rats, or, in some loci, may have a pleiotropic effect on both behavior and on the manifestation of traits associated with the hypertensive status of ISIAH rats. The results obtained in this study can be used to determine candidate genes at genetic loci associated with behavioral patterns of hypertensive ISIAH rats.

ACKNOWLEDGMENT

The work was supported by Russian Foundation for Basic Research grant No. 20-04-00119a and by Budget project No. 0324-2019-0041-C-01.

REFERENCES

[1] B. Bohus, R. F. Benus, D. S. Fokkema, J. M. Koolhaas, C. Nyakas, G. A. van Oortmerssen, A. J. Prins, A. J. de Ruiter, A. J. Scheurink, and A. B. Steffens, "Neuroendocrine states and behavioral and physiological stress responses," Prog. Brain Res., vol. 72, pp. 57-70, 1987.

[2] F. Kiefer, and K. Wiedemann, "Neuroendocrine pathways of addictive behavior," Addict. Biol., vol. 9, no. 3-4, pp. 205-212, 2004.

[3] R. S. Fries, P. Mahboubi, N. R. Mahapatra, S. K. Mahata, N. J. Schork, G. W. Schmid-Schoenbein, and D. T. O'Connor, "Neuroendocrine transcriptome in genetic hypertension: multiple changes in diverse adrenal physiological systems," Hypertension, vol. 43, no. 6, pp. 1301-1311, 2004.

[4] I. O. Meshkov, T. A. Alekhina, T. A. Moreva, and A. L. Markel', "[Behavioral characterictics of ISIAH rat strain]," Zh. Vyssh. Nerv. Deiat. Im. I. P. Pavlova, vol. 62, no. 2, pp. 233-242, 2012. Russian.

[5] A. L. Markel, O. E. Redina, M. A. Gilinsky, G. M. Dymshits, E. V. Kalashnikova, Y. V. Khvorostova, L. A. Fedoseeva, and G. S. Jacobson, "Neuroendocrine profiling in inherited stress-induced arterial hypertension rat strain with stress-sensitive arterial hypertension," J. Endocrinol., vol. 195, no. 3, pp. 439-450, 2007.

[6] O.E. Redina, N.A. Machanova, V.M. Efimov, and A.L. Markel. Rats with Inherited Stress-Induced Arterial Hypertension (ISIAH strain) display Specific Quantitative Trait Loci for Blood Pressure and for Body and Kidney Weights on Chromosome 1. Clin. Exp. Pharm. Phys., vol. 33, pp. 456-464, 2006.

DOI 10.18699/BGRS/SB-2020-057

Transcriptional profiling of ventral tegmental area of male mice with alternative patterns of social behaviors

Olga Redina
Laboratory of Neuropathology Modeling, Laboratory of Evolutional Genetics,
Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
oredina@bionet.nsc.ru

Vladimir Babenko
Laboratory of Neuropathology Modeling,
Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
bob@bionet.nsc.ru

Vadim Efimov
Laboratory of Molecular Genetic Systems Modeling,
Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
efimov@bionet.nsc.ru

Dmitry Smagin
Laboratory of Neuropathology Modeling, Neurogenetics of Social Behavior Sector,
Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
smagin@bionet.nsc.ru

Irina Kovalenko
Laboratory of Neuropathology Modeling, Neurogenetics of Social Behavior Sector,
Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
koir@bionet.nsc.ru

Anna Galyamina
Laboratory of Neuropathology Modeling, Neurogenetics of Social Behavior Sector,
Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
galyamina@bionet.nsc.ru

Natalia Kudryavtseva
Laboratory of Neuropathology Modeling, Neurogenetics of Social Behavior Sector, Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
n.n.kudryavtseva@gmail.com

Abstract — The activity of ventral tegmental area (VTA) neurons plays a crucial role in the reward circuit, emotional and addictive behaviors in animals and humans. ***To elucidate the molecular mechanisms underlying the functional changes in VTA neurons under stress arising from social interactions, RNA-Seq analysis was used to compare the transcriptome profiles in the VTA of three groups of male mice: chronically winning mice with positive social experience in daily agonistic interactions, chronically defeated mice with negative social experience in daily agonistic interactions, and control mice having no experience of agonistic interactions.*** **The data obtained showed that both winning and defeated mice experience stress, however, in defeated animals, the repeated agonistic interactions have a stronger effect and cause more significant changes in the levels of gene transcription.** ***Several genes have been identified that may be involved in the determination of alternative behavioral phenotypes in groups of male mice with alternative social experience.***

Keywords — ventral tegmental area (VTA), repeated agonistic interactions, RNA-Seq, winners and defeated mice

Introduction

The mesocortical and mesolimbic dopaminergic pathways originating from the ventral tegmental area (VTA) neurons play a crucial role in reward circuit, emotional and addictive behaviors. Induction of VTA dopaminergic activity is associated with the development of a number of psychiatric disorders. Repeated positive or negative social experience in daily agonistic interactions can be considered as an effective experimental approach to elucidate the molecular mechanisms underlying the activity of VTA neurons [1]. The repeated winning or defeated experience of mice leads to the formation of distinctive patterns of behavior which might be characterized as the winning mice having experience of social victories and developing enhanced aggressiveness, and defeated mice having a chronic experience of defeats and developing a mixed anxiety/depression-like state [2-4]. The current study was aimed at searching for the key VTA genes that may be responsible for the formation of alternative behaviors in winning and defeated mice exposed to the repeated agonistic interactions.

Methods

Animals

C57BL/6J male mice aged 10-12 weeks were used. Mice were kept under standard conditions with food and water given *ad libitum*. The study was approved by Scientific Council N 9 of the Institute of Cytology and Genetics SB RAS of March, 24, 2010, N 613 (Novosibirsk). Alternative forms of social behavior were generated as a result of daily agonistic interactions of male mice as described previously [3]. Three groups of animals were used: (1) Controls – mice without the experience of agonistic interactions; (2) chronically winning mice; (3) chronically defeated mice. VTA samples from 3 animals of each group were collected-and stored in RNAlater (Life Technologies, USA) at -70°C until sequencing.

RNA-Seq analysis

The frozen VTA samples were sent to JSC Genoanalytica (Moscow, Russia, http://genoanalytica.ru) where the RNA-Seq analysis was performed. All samples were analyzed as biological replicates. The sequencing data were mapped to the mouse reference genome sequence GRCm38.p3. Cufflinks/Cuffdiff programs were run to identify the differentially expressed genes (DEGs) in the experimental and control groups. Genes were considered differentially expressed at a false discovery rate (FDR) <5%. Only annotated gene sequences were used in the analysis.

Databases used

The Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/) Pathway Database was used to identify the significantly ($p<0.05$) enriched metabolic pathways. To define the association of DEGs with behavior/neurological phenotype the Neurological Disease Portal (Phenotypes, Mouse) in Rat Genome Database (https://rgd.mcw.edu/rgdweb/portal/home.jsp?p=7) was employed. The functional annotation of DEGs was performed by the Database for Annotation, Visualization and Integrated Discovery gene annotation tool (http://david.abcc.ncifcrf.gov/). Atlas of combinatorial transcriptional regulation in mouse and man [5] was used to reveal the DEGs encoding the transcription factor genes.

Statistical methods

The obtained RNA-Seq data (in FPKM values) were analyzed using partial-least squares discriminant analysis (PLS-DA) [6] following by the Pearson correlation between gene expression levels and the coordinates of animals along the first functionally meaningful synthetic PLS-DA Axis 1. A set of expressed genes, which were characterized by the highest correlation coefficients, was considered as genes contributing the most to inter-group differences.

Results

Analysis of gene expression in VTA from winning mice compared to the control revealed 44 differentially expressed genes (DEGs), 14 of which are known as associated with the behavior and neurological phenotype. Nine DEGs encode transcription factors. The results obtained in this comparison suggest an increase in the synthesis and transport of serotonin in VTA of aggressive mice compared to the control.

Analysis of gene expression in VTA of defeated mice compared to the control revealed 188 differentially expressed genes, 64 of which are known as associated with the behavior/neurological phenotype, including six DEGs (*Camk2a, Gad2, Slc6a2, Slc6a3, Slc6a4, Tph2*) known as associated with abnormal depression-related behavior. 28 DEGs encode transcription factors. Based on changes in the level of transcription of a number of genes in VTA of defeated mice as compared to control ones, activation of calcium signaling processes, increased synthesis of GABA (gamma-aminobutyric acid), increased synthesis and transport of serotonin and dopamine, and decreased transport of norepinephrine and glycine can be assumed.

A comparison of the DEG lists revealed 23 common genes that significantly changed the level of transcription in both winning and defeated mice compared to the control. 22 of these genes either increased or decreased the level of transcription unidirectionally in both experimental groups, and only one gene (*Nrgn*, neurogranin) changed the level of transcription in winning and defeated animals in different directions with respect to the control. Accordingly, we can conclude that, regardless of whether the animals are winning or defeated, the level of transcription of a large number of genes in their VTA changes unidirectionally, which is probably the result of the reaction of the animal organism to social stress conditions. It can be assumed that genes whose expression level varies in different directions can make the contribution to the formation of alternative behavioral phenotypes in mice with alternative social experiences in daily agonistic interactions. One of these genes is probably the *Nrgn* gene. In order to identify additional genes that changed the transcription level in winners and defeated mice in different directions, a comparative analysis of gene expression in winners compared to defeated animals was performed.

When comparing the level of gene expression in VTA of winning mice versus the defeated, 207 DEGs were determined. Among them, 6 genes bidirectionally changing the level of transcription in experimental groups of mice relative to the control were found. According to technology based on PLS-DA analysis, four of them (*Six3, Ercc2, Otx2* and *Nrgn*) can make the maximum contribution to intergroup differences when comparing winning and defeated mice. Three of these genes *Ercc2, Otx2*, and *Nrgn* are associated with behavior/neurological phenotype. The most highly expressed of these genes are the *Ercc2* and *Nrgn*. Gene Ontology analysis showed that the *Ercc2* gene is associated with biological processes such as neurogenesis (gliogenesis), regulation of transcription from RNA polymerase II promoter (regulation of gene expression), programmed cell death, stem cell differentiation, response to stimulus, response to hypoxia, immune system development, aging, and the *Nrgn* gene is associated with biological processes such as nervous system development, cell-cell signaling, trans-synaptic signaling, behavior, cognition, regulation of signaling, learning or memory, learning, response to stimulus, signal transduction, regulation of synapse structure or activity, modulation of synaptic transmission, regulation of synaptic plasticity, associative learning, intracellular signal transduction. We suggest that the expression patterns of these genes may be involved in the determination of behavioral differences in mice with alternative social experiences in daily agonistic interactions.

Acknowledgment

This work has been supported by the Russian Science Foundation, Grant No. 19-15-00026.

References

[1] E. N. Holly and K. A. Miczek, "Ventral tegmental area dopamine revisited: effects of acute and repeated stress," Psychopharmacology (Berl), vol. 233, no. 2, pp. 163-186, 2016.

[2] N. N. Kudryavtseva, I. V. Bakshtanovskaya, L. A. Koryakina, "Social model of depression in mice of C57BL/6J strain," Pharmacol. Biochem. Behav., vol. 38, no. 2, pp. 315-320, 1991.

[3] N. N. Kudryavtseva, D. A. Smagin, I. L. Kovalenko and G.B. Vishnivetskaya, "Repeated positive fighting experience in male inbred mice," Nat. Protoc., vol. 9, no. 11, pp. 2705-2717, 2014.

[4] A. Shimamoto, "Social defeat stress, sex, and addiction-like behaviors," Int. Rev. Neurobiol., vol. 140, pp. 271-313, 2018.

[5] T. Ravasi, H. Suzuki, C.V. Cannistraci, et al., "An atlas of combinatorial transcriptional regulation in mouse and man," Cell, vol. 140, no. 5, pp. 744-752, 2010.

[6] M. Barker and W. Rayens, "Partial least squares for discrimination," J. Chemometrics, vol. 17, no. 3, pp. 166–173, 2003.

DOI 10.18699/BGRS/SB-2020-059

Genome distance between regulatory elements of growth-related genes may determine morpho-physiological traits in mammals

Dmitriy Romanov
Academy of biology and biotechnologies
Southern federal university
Rostov-on-Don, Russia
rdme@yandex.ru

Tatiana Shkurat
Academy of biology and biotechnologies
Southern federal university
Rostov-on-Don, Russia
tshkurat@yandex.ru

***Abstract* — The problem of growth regulation in mammals is one of the long-standing mysteries in biology.**

We found significant correlation between morpho-physiological traits and genome distances between conserved elements in neighborhoods of growth-related genes *Mycn, Plagl1* and *Ezh2* in mammals. Given these conserved elements may be regulatory as well, we propose that genome distance between them may *evolutionary* modulate gene expression of these genes and eventually affects phenotype. It opens intriguing prospect to control morpho-physiological traits of mammals (i.e. adult body mass or longevity) just simply editing genome distance between these elements.

Index Terms* — *growth regulation, conserved elements, genome distance, morpho-physiological traits, mammals

MOTIVATION AND AIM

Motivation

The problem of growth regulation in mammals is one of the long-standing mysteries in biology [1, 2].

We have previously revealed strong correlation between morpho-physiological traits and some inter-conserved element distances in the neighborhoods of *Mycn*, *Plagl1* and *Ezh2* genes [3]. We propose that genome distance between regulatory elements of the growth-related genes may modulate gene expression of this genes and eventually affects phenotype.

Aim

Many *cis*-regulatory elements being conserved elements as well, the aim of this research was to study the relationships between morpho-physiological traits of mammals and genome distances between the conserved elements in the neighborhoods of the *Mycn*, *Plagl1* and *Ezh2* genes.

METHODS

Correlation analysis

All available genome sequences of ±50000 b.p. neighborhoods of *Mycn*, *Plagl1* and *Ezh2* genes of mammals (about 130 species) were obtained.

Genome elements, preserved in each neighborhood of the gene, were considered to be conserved. Each conserved element was assigned a name, i.e. *MYCN(-6893)*, where *MYCN* denotes gene and −6893 — genome distance in b.p. between gene start and the middle of the element.

Correlation analysis of the relationships between interconserved element distances in the neighborhoods of the *Mycn*, *Plagl1* and *Ezh2* genes and morpho-physiological traits of mammals was carried out.

Adult body mass, adult body length, age of maturity and lifespan were assessed. The trait values were obtained from the PanTHERIA and AnAGE databases [4, 5].

The relationships with the Spearman's correlation coefficient greater than 0.4 and the *p*-value less than 0.05 were considered to be statistically significant. The problem of multiple comparisons was addressed by multiplying the resulting *p*-value by the number of comparisons according to the Bonferroni correction.

Gene Ontology enrichment analysis

The *BLAST* search for homologs of the conserved elements in the human genome was carried out. The set of genes in the ±50000 b.p. neighborhood of the homologs was obtained. Gene Ontology enrichment analysis of this set of genes was made with the help of *Panther classification system* [6].

RESULTS

Correlation

The correlations revealed in this study (Table I) are generally conform to those obtained previously on the set of 36 mammals [3]. Nevertheless, correlations are weak in spite of *p*-values being very small. It may be due to a fuzzy nature of morpho-physiological trait data. Strong correlation was revealed between *EZH2(-8314)– EZH2(196)* genome distance and age of maturity. Semilog plots of the inter-element distances and morphophysiological traits for the pairs of the topmost correlated elements are given in Fig. 1.

TABLE I – CORRELATION BETWEEN THE INTER-CONSERVED ELEMENT DISTANCES AND MORPHO-PHYSIOLOGICAL TRAITS OF MAMMALS

Corr. coeff.	P-value	Bonferroni corrected p-value	Element 1	Element 2	Trait
0.53	4.79E-10	4.79E-08	PLAGL1(79389)∗	PLAGL1(112171)	Adult body
-0.41	1.38E-05	1.38E-03	MYCN(-6893)	MYCN(74)	mass
-0.40	2.08E-05	2.08E-03	MYCN(-6893)	MYCN(1273)	
0.54	5.73E-09	5.73E-07	PLAGL1(79389)	PLAGL1(112171)	Adult body
-0.44	1.90E-05	1.90E-03	MYCN(-25368)	MYCN(1273)	length
-0.44	2.03E-05	2.03E-03	MYCN(-25368)	MYCN(74)	
-0.41	7.06E-05	7.06E-03	MYCN(-6893)	MYCN(74)	
0.40	1.95E-05	1.95E-03	PLAGL1(79389)	PLAGL1(112171)	Max
-0.51	1.24E-04	1.24E-02	EZH2(-8314)	EZH2(194)	longevity
-0.75	1.14E-09	1.14E-07	EZH2(-8314)	EZH2(194)	Sexual
-0.54	1.08E-06	1.08E-04	EZH2(-8314)	EZH2(1627)	maturity age

∗Elements, which overlap gene promoters, are highlighted in bold

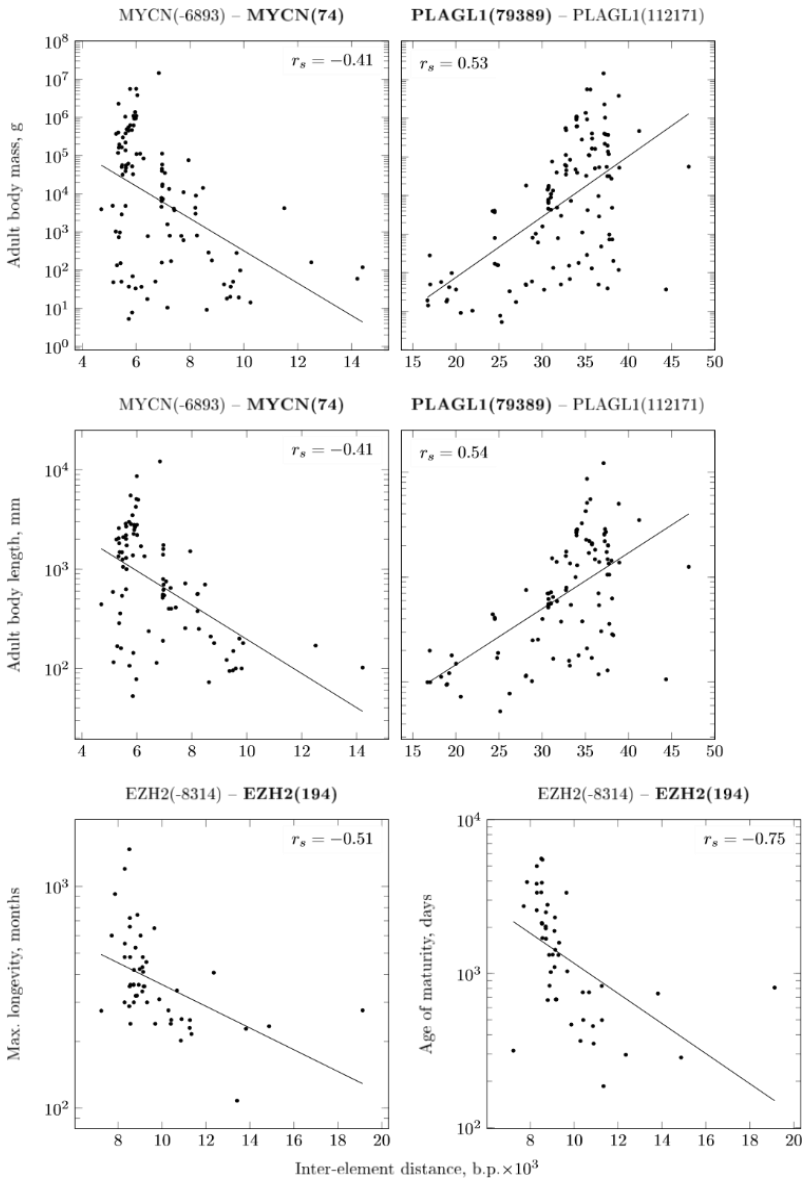


Fig. 1. Semilog plots of the inter-conserved element distances and morphophysiological trait for the pairs of the topmost correlated conserved elements in the neighborhoods of the *Mycn*, *Plagl1* and *Ezh2* genes. Elements, which overlap gene promoters, are highlighted in bold. Spearman's rank correlation coefficent is provided for every plot with the line of best fit

In each plot dots appear to line up along the straight line suggesting that morpho-physiological traits depend exponentially on the inter-element distance.

Gene Ontology enrichment analysis

The BLAST search for homologs of the topmost correlated conserved elements was made.

Surprisingly, the promoter overlapping elements *MYCN(74)*, *PLAGL1(112171)* and *EZH2(194)* were presented in a single copy in the genome, whereas the corresponding elements of the topmost correlated pair — *MYCN(-6893)*, *PLAGL1(79389)* and *EZH2(-8314)* — were scattered across the genome.

It suggests that elements *MYCN(74)*, *PLAGL1(112171)* and *EZH2(194)* may be regarded as *gene-specific* regulatory elements, whereas elements *MYCN(-6893)*, *PLAGL1(79389)* and *EZH2(-8314)* may represent *universal* regulatory elements, capable to control many neighboring genes.

Gene Ontology enrichment analysis of the set of genes in the ±50000 b.p. neighborhood of homologs of elements *MYCN(6893)*, *PLAGL1(79389)* and *EZH2(-8314)* was made.

It was found that *ACVR1B* and *ACVRL1* genes in the vicinity of homolog of element *EZH2(-8314)* are significantly overrepresent category *response to growth factor*, GO:0070848, p =0.046.

Thus, it suggests that element *EZH2(-8314)* may do be a regulatory sequence and control growth-related genes *EZH2*, *ACVR1B* and *ACVRL1* along with other genes.

ACKNOWLEDGMENT

The study was performed on the equipment of the Center of collective use “High technologies” (Southern federal university, Rostov-on-Don, Russia) the framework of the state task of Russian Ministry of education and science “Biochemical and molecular genetic studies of pathological processes mechanisms associated with socially significant diseases”, № BAZ0110/20-5-14AB.

REFERENCES

[1] J. C. Lui, J. Baron, Mechanisms limiting body growth in mammals, Endocr. Rev. 32 (3) (2011) 422–440.

[2] A. Delaney et al. Evolutionary conservation and modulation of a juvenile growth-regulating genetic program, J. Mol. Endocrinol. 52 (3) (2014) 269–277.

[3] Genome distance between conserved elements in neighborhoods of growth-regulating genes is correlated with morpho-physiological traits in mammals / D. E. Romanov, E. V. Butenko, G. B. Bakhtadze, T. P. Shkurat // Gene Reports. — 2019. — Vol. 17.

[4] Jones, K.E. et al. (2009) PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology 90, 2648.

[5] Tacutu, R et al. (2018) Human Ageing Genomic Resources: new and updated databases. Nucleic Acids Res., 46, D1:D1083-D1090.

[6] Mi, H, Muruganujan, A, Thomas, PD (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res., 41, Database issue:D377-86.

DOI 10.18699/BGRS/SB-2020-060

JetGene – an internet resource for analysis of regulatory regions or nucleotide contexts at differently translated transcripts

Sadovskaya N.S.
Tmiryazev Institute of Plant Physiology, RAS, Moscow, Russia
nataliya.sadovskaya@gmail.com

Mustafaev O.N.
Genetic Resources Institute, ANAS, Baku, Azerbaijan
orkhan@bioset.org

Tyurin A.A.
Tmiryazev Institute of Plant Physiology, RAS, Moscow, Russia

Goldenkova-Pavlova I.V.
Tmiryazev Institute of Plant Physiology, RAS, Moscow, Russia

***Abstract* — We create web database JetGene that allows to estimate the variation of nucleotide composition, codon usage frequency, to study nucleotides surrounding of the start codon and much more. In addition beta-JetGene allows user to compare two samples of mRNA.**

Keywords* — *in silico analysis, regulatory codes, motives, translation effectivity, comparison of two mRNA samples

Motivation and aim

The paradox of misfit between the levels of mRNAs and translation effectivity in the eukaryotic cells, encountered by researchers at every phase of organism development and under influence of various stress conditions at whole organism. Accordingly, in recent times researchers direct their close attention into fine mechanisms of translation.

Results

It's known that mRNA has some regulatory codes that define the individual mRNA fate in translation. For such regulatory codes and motives discovery researchers apply bioinformatics analysis of different mRNA regions. With the purpose of such regulatory codes detection in mRNA and their correlation with translational efficiency we develop an internet resource JetGene (https://jetgene.bioset.org/). This web resource contains cDNA, CDS, 5'-UTR, 3'-UTR sequences of six kingdoms of living organisms. This should be mentioned that information about CDS and cDNA is downloaded and updated regularly from Ensembl [1]. JetGene has a wide toolkit and very friendly for users. JetGene allows make a comparative analysis of sequences, namely: (i) to estimate the variation of length, nucleotide composition, frequency of codon usage, to analyze GC-content, CpG-islands, to study nucleotides surrounding of the start codon and much more; (ii) to identify and define statistically significant representation of potential regulatory contexts at mRNA with different translation efficiency. It should be noted that the analysis could be performed both by full-length transcript, by interval of transcript and by coding/non-coding regions. Sequences that selected by researchers for analysis can be extract in fasta-format from JetGene. The graphical representation of results accompanies every phase of the study. Such cute detail is greatly facilitated the work of any user.

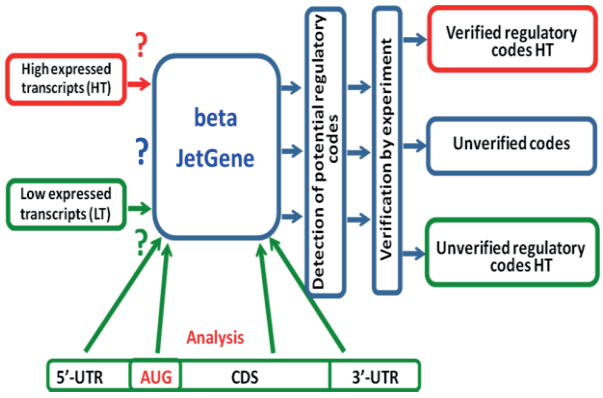


Fig. 1. Comparison of two mRNA samples by beta-JetGene

In addition, beta-version of JetGene (https://beta.bioset.org) allows user to compare two mRNA samples (Fig. 1) and to apply omics data for searching and prediction regulatory determinants of translation.

Acknowledgment

Supported by the Russian Science Foundation (project no. 18-14-00026; IVG-P).

References

[1] Yates A.D. et al. (2020) Ensembl 2020. Nucleic Acids Res. 2020 Jan 8;48(D1):D682-D688.

DOI 10.18699/BGRS/SB-2020-061

Cryptic plasmids of alfalfa root nodule bacteria – structural and functional diversity

Alla Saksaganskaia
ARRIAM, Saint-Petersburg, Pushkin, Russia
allasaksaganskaya@mail.ru

Viktoria Muntyan
ARRIAM, Saint-Petersburg, Pushkin, Russia
vucovar@yandex.ru

Alexey Afonin
ARRIAM, Saint-Petersburg, Pushkin, Russia
afoninalexeym@gmail.com

Marina Roumiantseva
ARRIAM, Saint-Petersburg, Pushkin, Russia
mroumiantseva@yandex.ru

***Abstract* — Abundance of plasmids in genomes of root nodule nitrogen fixing bacteria (rhizobia) is their common feature. Plasmids harboring genes related to symbiosis are usually called symbiotic plasmids, while rhizobia can contain differing number of cryptic plasmids, which sizes are varied from 7 kbp to 600 kbp. Cryptic plasmids were predicted to be important both for the existence of rhizobia in soil and for formation of effective nitrogen-fixing symbiosis with particular host plant. Nevertheless, assemble and annotation of such replicons became possible only in the postgenomic era. In this work a structural and functional analysis of cryptic plasmids of *Sinorhizobium meliloti* which forming symbiosis with alfalfa was done. Native isolates of *S. meliloti* were recovered from nodules of alfalfa plants adapted to salinized soils at the Aral Sea region. The sequence analysis was done for four plasmids ranging in size from 31.2 kbp to 453.7 kbp. A large number of ORFs of cryptic plasmids of *S. meliloti* are associated with metabolic processes as it was established. For the first time prophage sequences were identified in cryptic plasmids. Data proving that cryptic plasmids of native strains are related to horizontal gene transfer and could be essential for rhizobia fitness.**

Keywords* — *S. meliloti, cryptic plasmids, ORFs, COGs, prophages, horizontal gene transfer

Motivation and aim

Motivation

Cryptic plasmids (CPs) are accessory elements of the multicomponent genome of α-proteobacteria forming nitrogen-fixing symbiosis with legumes [1]. The study of the functional role of nonsymbiotic CPs has got a significant prospect in postgenomic era.

Aim

The aim of this work was the comparative analysis of CPs which were obtained as a result of genome-wide sequencing of symbiotically active strains *Sinorhizobium meliloti* AK555 and AK83 adapted to salinized area of the Aral Sea region.

Methods

Strains *Sinorhizobium meliloti* AK555 and AK83 were obtained in the frame of the INCO-COPERNICUS project (ICA2-CT-2001-10001). Cryptic plasmids assembly of AK555 was done on the basis of full-genome sequencing [2]. The annotation of plasmid sequences was done by using bioinformatics approaches (Flye, Racon, Medaka, Pilon, Prokka).

Data about pSINME02 and pSINME01 of the AK83 are from GenBank (NC_015592.1; NC_015597.1).

Results

The two nonsymbiotic plasmids as a result of genome-wide sequencing of the AK555 were assembled and annotated by using bioinformatics approaches (Flye, Racon, Medaka, Pilon, Prokka). The SMd is a replicon of 31.2 kbp with 34 ORFs and SMe is a replicon of 453.7 kbp containing 519 ORFs. Plasmids of the strain AK83 are: pSINME02 of 70.5 kbp and pSINME01 of 256.3 kbp. The first replicon contained 85 ORFs and the second one 287 ORFs.

The analysis of SMd of the AK555 revealed that 86% of its length showed 92% homology with nearly a half of the sequence pSINME02 of the AK83. But no homology was detected between above mentioned plasmids and other two plasmids SMe and pSINME01 of the both strains.

More than a half of ORFs of each examined replicons were characterized by COGs. In the result, all ORFs were allocated into 4 functional groups: cellular processes and signaling (D, M, N, O, T, U, V, W), information storage and processing (J, K, L), metabolism (C, E, J, H, I, P, Q) and poorly characterized (R, S, X). An average portions of ORFs belonged to cellular processes/signaling or metabolism groups were 0.2 and 0.34, correspondingly, per replicon. ORFs related to metabolism were detected on SMe, pSINME02, pSINME01, but that was not a case for the SMd. On the SMe was detected ORF encoding Kup2 protein involving in K+ transport, which could relate to osmoadaptation process. The Kup2 is homologous to Kup proteins encoded by other two sequences localized on the chromosome and megaplasmid SMa of the same strain. While the homology was 71 and 84%, respectively.

On plasmids SMe and SINME01 prophage sequences of different sizes were detected by using PHASTER (https://phaster.ca/). The 2 sequences of the SMe were: 12.9 kbp and 5.9 kbp, which were characterized as intact and incomplete prophages, correspondingly. On plasmid pSINME01 the 2 prophage sequences of a sizes 43.9 kbp and 9.3 kbp were identified. These sequences were characterized as intact prophage and incomplete prophages. The analysis of protein-coding genes of phage related sequences revealed homology with sequences of phages from *Siphoviridae*, *Myoviridae* and *Podoviridae* families.

Summarizing, cryptic plasmids of *S. meliloti* are enriched with ORFs related to vital metabolic processes, harboured phage related sequences. The fact of the homology between two plasmids of genetically unrelated strains evident that these plasmids are participating in horizontal gene transfer. Thus nonsymbiotic plasmids of *S. meliloti* are essential elements of rhizobia fitness.

ACKNOWLEDGMENT

Supported by the RFBR 18-04-01278A (functional role of ORFs) and by RSF 20-16-00105.

REFERENCES

[1] Lagares A., Sanjuán J., Pistorio M. (2014) The plasmid mobilome of the model plant-symbiont *Sinorhizobium meliloti*: coming up with new questions and answers. Microbiology Spectrum. 2(5):PLAS-0005-2013. doi:10.1128/microbiolspec.PLAS-0005-2013.

[2] Muntyan V.S., Baturina O.A., Afonin A.M., Cherkasova M.E., Laktionov Y.V., Saksaganskaya A.S., Kabilov M.R., Roumiantseva M.L. (2019) Draft genome sequence of *Sinorhizobium meliloti* AK555. Microbiology Resource Announcements. 8(2): e01567-18. doi: 10.1128/MRA.01567-18.

DOI 10.18699/BGRS/SB-2020-062

MicroRNA content of horse and human milk exosomes

Sergey Sedykh
ICBFM SB RAS, Novosibirsk, Russia
sedyh@niboch.nsc.ru

Anna Kuleshova
ICBFM SB RAS, Novosibirsk, Russia
aekuleshova25@gmail.com

Georgy Nevinsky
ICBFM SB RAS, Novosibirsk, Russia
nevinsky@niboch.nsc.ru

Abstract — **Exosomes are extracellular vesicles secreted by almost all cell types. Exosomes were isolated from different biological liquids: cell cultures, blood, urine, tears, saliva and milk. Milk is unique source of exosomes. Here we describe microRNA content of horse and human milk exosomes.**

Keywords — exosomes, milk, microRNA

Motivation and aim

Exosomes are 40–100 nm diameter natural vesicles, containing CD9, CD63 and CD81 tetraspanins on the surface. It was shown that milk obtained from different sources contain exosomes: human, horse, bovine, rat. Horse milk is a unique source of exosomes since is less allergenic than bovine and not prion prone. Also, horse milk can be obtained in larger amounts than human milk.

Methods

According to the results, published by the other research groups, milk exosomes contain hundreds and thousands of proteins, mRNA and microRNA molecules. Our recent results show that these numbers may be significantly overestimated. Isolation of exosomes from sediments obtained after ultracentrifugation with an additional step of gel-filtration allows decreasing the number of proteins, that co-isolate with vesicles. Many microRNAs are described in exosomes, obtained from human and bovine milk.

Results

Here we show the content of more than two dozens major microRNA molecules, isolated from human and horse milk exosomes on different stages of centrifugations, ultracentrifugations, before and after gel-filtration. According to these data, several microRNAs may be used as indicators of milk exosomes' purity.

ACKNOWLEDGMENT

The study was funded by the Russian Scientific Foundation (research project 187410055 to S. Sedykh) and by the Program of Fundamental Research of Government Academia (AAAA-A17-117020210023-1 to Prof. G. Nevinsky).

DOI 10.18699/BGRS/SB-2020-063

Novel loci associated with plasma immunoglobulin G *N*-glycosylation identified by a multivariate analysis

Alexandra S. Shadrina
Laboratory of Glycogenomics
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
ORCID: 0000-0003-1384-3413

Alexander S. Zlobin
Laboratory of Recombination and Segregation Analysis
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia

Olga O. Zaytseva
Genos Glycoscience Research Laboratory
Zagreb, Croatia
ORCID: 0000-0003-3960-5157

Gordan Lauc
Genos Glycoscience Research Laboratory
Zagreb, Croatia
ORCID: 0000-0003-1840-9560

Lucija Klaric
MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine,
University of Edinburgh
Edinburgh, United Kingdom
ORCID: 0000-0003-3105-8929

Sodbo Z. Sharapov
Laboratory of Glycogenomics
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
ORCID: 0000-0003-0279-4900

Yurii S. Aulchenko
Laboratory of Glycogenomics
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
ORCID: 0000-0002-7899-1575

Yakov A. Tsepilov
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia
ORCID: 0000-0002-4931-6052

***Abstract* — Immunoglobulin G (IgG) is the most prevalent human plasma N-glycosylated protein, which makes a significant impact into the total plasma protein glycosylation profile. Glycosylation of IgG is known to affect its biological properties; for example, sialylation of glycans attached to IgG is known to have an anti-inflammatory effect, while the absence of a core fucose in the IgG glycan structure can increase antigen-dependent cell cytotoxicity. The biochemical processes underlying protein glycosylation are well-studied; at the same time, little is known about biological network regulating these reactions. In the present study, we performed a multivariate analysis based on the summary statistics obtained in the previously published IgG N-glycome GWAS in order to discover new loci influencing IgG N-glycosylation patterns. We revealed thirty-four loci associated with the levels of plasma IgG N-glycosylation. Of these loci, eight loci have not been reported in previous works. Our results significantly expand the number of identified IgG N-glycome-associated loci and contribute to understanding the mechanisms of the genetic control of glycosylation.**

Keywords* — *human N-glycosylation; immunoglobulin G; multivariate analysis; genome-wide association study; locus

Introduction

Immunoglobulins are the most abundant glycoproteins in human blood plasma: while the total concentration of glycoproteins in plasma is about 30 mg/mL, the concentration of immunoglobulins is approximately 16 mg/mL [1]. Of these, immunoglobulin G (IgG) constitutes approximately 75%. IgG heavy chains carry bi-antennary glycan moieties covalently linked to an asparagine 297 residue. The spectrum of glycans attached to IgG molecules is called the IgG *N*-glycome. Glycosylation of IgG influences its effector functions and plays an important regulatory role in the immune system [2], [3]. Changes in the IgG *N*-glycome composition are associated with a number of diseases, such as inflammatory, infectious, autoimmune, and alloimmune diseases, and different types of cancer [3]. Biological mechanisms regulating IgG *N*-glycosylation are still not fully understood. Elucidating the genetic control of IgG *N*-glycome helps to uncover molecules and pathways involved in establishing IgG glycosylation patterns and facilitates research linking IgG *N*-glycome and human pathologies. To date, four genome-wide association studies of the total plasma IgG *N*-glycome have been published [4]–[7]. The largest study was conducted by Klarić et al. [7], which has increased the number of known loci robustly associated with IgG *N*-glycome to 23.

The aim of our study was to further increase the statistical power of the analysis and reveal novel loci associated with the plasma IgG *N*-glycome traits. Our aim was achieved by performing a multivariate analysis using the genome-wide association study (GWAS) summary statistics obtained in the previous work [7].

Materials and methods

The discovery GWAS providing summary statistics for our analysis was performed in four European-ancestry cohorts with a total sample size of 8,090 individuals [7]. We used the GWAS results for 23 IgG *N*-glycan traits directly measured by ultra-performance liquid chromatography. We grouped these 23 glycan traits into nine groups based on their general structural and chemical properties as described previously [5]. The following groups were constructed: galactosylation, monogalactosylation, digalactosylation, sialylation, monosialylation, disialylation, fucosylation, bisecting *N*-Acetylglucosamine, and one group that combined all 23 traits (*N*-glycosylation). Multivariate analysis was performed using the GWAS summary statistics- based method described by Ning et al. [8]. The statistical significance threshold was set at $P < 1.67e\text{-}09$ (5.0e-08/(21+9), where 9 is the number of trait groups and 21 is the number of principal components explaining most of the variance of IgG *N*-glycosylation traits [4]). Calculations were performed using the MultiABEL package for R software environment.

Results

Multivariate analysis revealed 34 loci associated with at least one group of IgG *N*-glycosylation traits. Of these, 31 loci were associated with 2 or more traits groups. Median number of trait groups associated with the identified loci was 3 (range 1-8). Eight loci have not been reported in previous IgG *N*-

glycome studies (**Ошибка! Источник ссылки не найден.**). The number of newly identified loci constitutes one third of the number of loci that are already known to affect IgG *N*-glycosylation. This demonstrates the efficacy of the multivariate analysis methods in investigating the genetic architecture of complex traits. The revealed loci are likely to be involved in the genetic control of complex processes regulating general IgG *N*-glycosylation pathways.

ОШИБКА! ИСТОЧНИК ССЫЛКИ НЕ НАЙДЕН. – NOVEL LOCI ASSOCIATED WITH IGG *N*-GLYCOSYLATION TRAITS REVEALED BY THE MULTIVARIATE ANALYSIS.

SNP rs number	Chr:pos[a]	*P*-value	Top trait group[b]	Nearest gene
rs11895615	2:26113120	5.48e-10	galactosylation	*ASXL2*
rs1372288	3:142901537	3.74e-11	galactosylation	*CHST2, SLC9A9*
rs9844068	3:196224038	3.23e-12	*N*-glycosylation	*RNF168*
rs3130380	6:30279130	2.62e-12	galactosylation	*HCG18*
rs1269852	6:32080191	1.98e-19	*N*-glycosylation	*ATF6B*, *TNXB*
rs1794282	6:32666526	3.85e-15	galactosylation	*HLA-DQB1 - HLA-DQA2*
rs479844	11:65551957	7.25e-15	digalactosylation	*OVOL1*
rs4561508	17:16848750	4.51e-11	fucosylation	*TNFRSF 13B*

[a.] Chromosome: position on chromosome according to GRCh37.p13 assembly

[b.] Trait group most significantly associated with the locus

ACKNOWLEDGMENT

The work of ASSh, SZSh, and YSA was funded by the Russian Science Foundation grant number 19-15-00115. The work of ASZ was supported by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genetics (project0324-2019-0040-C-01 / AAAA-A17-117092070032-4). The work of YAT was supported by the Russian Ministry of Science and Education under the 5-100 Excellence Programme. The work of LK was supported by the RCUK Innovation Fellowship from the National Productivity Investment Fund (MR/R026408/1). The work of OOZ and GL was supported by the Croatian National Centre of Research Excellence in Personalized Healthcare grant (#KK.01.1.1.01.0010). ASSh, SZSh, and YSA coordinated and supervised this study, performed data analysis and interpreted the results. YAT conceived and oversaw the study, contributed to the design and interpretation of the results. ASZ, LK, OOZ, and GL contributed to data analysis and interpretation of the results.

REFERENCES

[1] F. Clerc, K. R. Reiding, B. C. Jansen, G. S. M. Kammeijer, A. Bondt, and M. Wuhrer, “Human plasma protein N-glycosylation,” Glycoconj. J., vol. 33, no. 3, pp. 309–343, Jun. 2016.

[2] E. Maverakis et al., “Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review,” Journal of Autoimmunity, vol. 57. Academic Press, pp. 1–13, 01-Feb-2015.

[3] I. Gudelj, G. Lauc, and M. Pezer, “Immunoglobulin G glycosylation in aging and diseases,” Cell. Immunol., vol. 333, pp. 65–79, Nov. 2018.

[4] G. Lauc et al., “Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers.,” PLoS Genet., vol. 9, no. 1, p. e1003225, Jan. 2013.

[5] X. Shen et al., “Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation,” Nat. Commun., vol. 8, no. 1, p. 447, Dec. 2017.

[6] A. Wahl et al., “Genome-wide association study on immunoglobulin G glycosylation patterns,” Front. Immunol., vol. 9, no. FEB, p. 277, Feb. 2018.

[7] L. Klarić et al., “Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases,” Sci. Adv., vol. 6, no. 8, p. eaax0301, Feb. 2020.

[8] [Z. Ning et al., “Beyond power: Multivariate discovery, replication, and interpretation of pleiotropic loci using summary association statistics,” bioRxiv, p. 022269, Apr. 2019.

DOI 10.18699/BGRS/SB-2020-064

Results of genome-wide association study of plasma proteome N-glycosylation in 10,000 samples

Sodbo Sharapov
Laboratory of Glycogenomics
Institute of Cytology and Genetics
Novosibirsk, Russia
ORCID: 0000-0003-0279-4900

Sofya Feoktistova
Laboratory of Glycogenomics
Institute of Cytology and Genetics
Novosibirsk, Russia
ORCID: 0000-0002-7608-439X

Lucija Klaric
MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine,University of Edinburgh
Edinburgh, United Kingdom
ORCID: 0000-0003-3105-8929

Harry Campbell
Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh
Edinburgh, United Kingdom
ORCID: 0000-0002-6169-6262

Matthias Schulze
Department of Molecular Epidemiology, Ger- man Institute of Human Nutrition Potsdam- Rehbruecke
Nuthetal, Germany
ORCID: 0000-0002-0830-5277

Yurii Aulchenko
Laboratory of Glycogenomics
Institute of Cytology and Genetics
Novosibirsk, Russia
ORCID: 0000-0002-7899-1575

Yakov A. Tsepilov
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia
ORCID: 0000-0002-4931-6052

Eugene Tiys
Laboratory of Glycogenomics
Institute of Cytology and Genetics
Novosibirsk, Russia
ORCID: 0000-0001-5820-8646

Karsten Suhre
Department of Physiology and Biophysics, Weill Cornell Medicine-Qatar, Doha, Qatar
ORCID: 0000-0001-9638-3912

Malcolm Dunlop
Colon Cancer Genetics Group, MRC Human Genetics Unit, MRC Institute of Genetics & Molecular Medicine, Western General Hospital, The University of Edinburgh
Edinburgh, United Kingdom
ORCID: 0000-0002-3033-5851

Tim Spector
Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences
King's College London
London, United Kingdom
ORCID:

Elizaveta E. Elgaeva
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia
ORCID: 0000-0002-2484-553X

Frano Vuckovic
Genos Glycoscience Research Laboratory
Zagreb, Croatia
ORCID: 0000-0002-4845-0882

Nishi Chaturvedi
MRC Unit for Lifelong Hlth & Ageing University College London
London, United Kingdom
ORCID:

Frances Williams
Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences
King's College London
London, United Kingdom
ORCID: 0000-0002-2998-2744

Gordan Lauc
Genos Glycoscience Research Laboratory, Zagreb, Croatia
ORCID: 0000-0003-1840-9560

Abstract — **One of the common co-translational and post-translational modifications of human proteins is glycosylation. Glycosylation influences the physical properties of proteins as well as their biological function. Alteration in glycosylation is observed in many human diseases. Defining genetic factors altering glycosylation may provide a basis for novel approaches to diagnostic and pharmaceutical applications. Here, we conducted a genome-wide association study (GWAS) of human total blood plasma N-glycome measured by ultra-performance liquid chromatography technology in almost 10 000 samples.**

Keywords — GWAS, association, N-glycosylation, glycomics, genetic, glycobiology, glycan

Introduction

Glycosylation – the covalent attachment of carbohydrate (glycan) to a substrate - is a common and structurally diverse post-translational modification of protein. Addition of glycans influences the physical properties of protein (solubility, conformation, folding, stability, trafficking, etc.) [1]–[3] as well as the biological function including protein-protein, cell-cell, cell-matrix, and host-pathogen interactions [1], [2], [4]. Alteration in glycosylation is observed in many human diseases such as type 1 and 2 diabetes mellitus [5], rheumatoid arthritis [6], Parkinson's disease [7] and cancer [8], [9]. Glycans are considered potential therapeutic targets [10] and biomarkers for early diagnosis and disease prognosis [11]–[14], making glycobiology a promising field for future clinical applications. Genome-wide association studies (GWAS) allow a hypothesis-free search of genetic variants involved in the regulation of glycosylation [15]. In our recent study [16] we analyzed the genetic control of total plasma N-Glycome (TPNG) measured by ultra performance liquid chromatography (UPLC) by conducting a GWAS using up to 3 800 samples. We found and confirmed the association of twelve loci with TPNG. In the present study we present results of the next stage, where we tripled the sample size and recruited up to 10 000 samples for further discovery of genetic control of the human blood plasma proteins N-glycosylation.

Materials and methods

Quality control of the TPNG data for 10 000 samples from seven cohorts: TwinsUK, EPIC, PainOmics, SOCCS, SABRE, QMDiab, and CEDAR was performed. As the number of quantified glycan traits measured by UPLC varied

across cohorts from 36 to 42, we applied our recently developed protocol that resulted in the harmonized set of 36 glycan traits. Based on the biochemical structure of 36 original glycans, we computed an additional 81 derived traits leading to a total of 117 glycan traits. We conducted an inverse-variance weighted meta-analysis of GWAS for 117 traits from seven the cohorts. The discovery and replication datasets had sample size of 7 500 and 3 100 samples respectively. For the replicated loci, we performed an *in silico* functional annotation to prioritize potentially causal genes. Prioritization used multiple lines of evidence such as presence of predicted damaging variants in a gene, pleiotropic effects of glycan-associated SNPs on gene expression and the results of DEPICT (Data-driven Expression Prioritized Integration for Complex Traits). Meta-analysis of 117 traits and seven cohorts as well as *in silico* follow up was performed using the GWAS-MAP tool [17], that provides researchers a "swiss army knife" tool for GWAS and post-GWAS analyses. An association was considered statistically significant at the genome- wide level if the P-value for an individual SNP was less than $5\times10-8/(27+1) = 1.78\times10-9$, where 27 is an effective number of tests (traits) that was estimated as the number of principal components that jointly explained 99% of the total plasma glycome variance in the TwinsUK sample

Results

We found 38 loci that were significantly associated ($p < 1.79 \times 10^{-9}$) with at least one glycan trait. Of the 38 loci, 12 were previously reported to be associated with plasma glycome composition [15], [16], 14 loci were associated with N-glycome composition of IgG [18]– one of the most prevalent N-glycoprotein in plasma, while six loci were known to be associated with both plasma and IgG N-Glycome. For the 14 loci, the significant association with plasma glycome was demonstrated for the first time. The majority of loci were replicated in a combined set of replication cohorts. For the replicated loci we conducted an extensive *in silico* functional follow up to priorities the functional genes and variants.

ACKNOWLEDGMENT

The work of S.Sh., E.E., S.F., E.T., YA, who coordinated and supervised this study, performed meta-analysis, *in-silico* functional follow-up, interpretation of the results and writing of the initial text was funded by the Russian Science Foundation grant number 19-15-00115. The work of YT, who contributed in interpretation of the results, was supported by the Russian Ministry of Science and Education under the 5-100 Excellence Programme and by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genetics (project 0324-2019-0040-C-01/AAAA-A17-117092070032-4). The work of LK, who contributed in interpretation of the results, was supported by the RCUK Innovation Fellowship from the National Productivity Investment Fund (MR/R026408/1).The generation of the data used in this study was supported by the European Structural and Investment Funds IRI grant (#KK.01.2.1.01.0003) and Croatian National Centre of Research Excellence in Personalized Healthcare grant (#KK.01.1.1.01.0010) to GL; by 'Biomedical Research Program' funds at Weill Cornell Medicine - Qatar, a program funded by the Qatar Foundation to KS; by grants from Cancer Research UK (C348/A3758, C348/A8896, C348/ A18927) to MD and HC; Scottish Government Chief Scientist Office (K/OPR/2/2/D333, CZB/4/94); Medical Research Council (G0000657-53203, MR/K018647/1); Centre Grant from CORE as part of the Digestive Cancer Campaign (http://www.corecharity.org.uk). TwinsUK (TS) is funded by the Wellcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. EPIC-Potsdam has been funded by the German Federal Ministry of Science, the German Federal Ministry of Education and Research, and the European Union.

REFERENCES

[1] K. Ohtsubo and J. D. Marth, "Glycosylation in Cellular Mechanisms of Health and Disease," *Cell*, vol. 126, no. 5. pp. 855–867, 08-Sep-2006.

[2] D. Skropeta, "The effect of individual N-glycans on enzyme activity," *Bioorg. Med. Chem.*, vol. 17, no. 7, pp. 2645–2653, Apr. 2009.

[3] H. Takeuchi *et al.*, "O-Glycosylation modulates the stability of epidermal growth factor-like repeats and thereby regulates Notch trafficking.," *J. Biol. Chem.*, vol. 292, no. 38, pp. 15964–15973, Jul. 2017.

[4] G. Lauc, M. Pezer, I. Rudan, and H. Campbell, "Mechanisms of disease: The human N-glycome.," *Biochim. Biophys. Acta*, vol. 1860, no. 8, pp. 1574–1582, Oct. 2015.

[5] R. F. H. Lemmers *et al.*, "IgG glycan patterns are associated with type 2 diabetes in independent European populations," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1861, no. 9, pp. 2240–2249, Sep. 2017.

[6] I. Gudelj *et al.*, "Low galactosylation of IgG associates with higher risk for future diagnosis of rheumatoid arthritis during 10 years of follow-up," *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1864, no. 6, pp. 2034–2039, 2018.

[7] A. C. Russell *et al.*, "The N-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease.," *Glycobiology*, vol. 27, no. 5, pp. 501–510, May 2017.

[8] B. N. Vajaria and P. S. Patel, "Glycosylation: a hallmark of cancer?," *Glycoconj. J.*, vol. 34, no. 2, pp. 147–156, Apr. 2017.

[9] N. Taniguchi and Y. Kizuka, "Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics.," *Adv. Cancer Res.*, vol. 126, pp. 11–51, 2015.

[10] E. RodrÍguez, S. T. T. Schetters, and Y. van Kooyk, "The tumour glyco-code as a novel immune checkpoint for immunotherapy.," *Nat. Rev. Immunol.*, vol. 18, no. 3, pp. 204–211, Mar. 2018.

[11] G. Thanabalasingham *et al.*, "Mutations in HNF1A result in marked alterations of plasma glycan profile," *Diabetes*, vol. 62, no. 4, pp. 1329–1337, Apr. 2013.

[12] B. Adamczyk, T. Tharmalingam, and P. M. Rudd, "Glycans as cancer biomarkers," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1820, no. 9, pp. 1347–1353, Sep. 2012.

[13] Y. Shinohara, J. I. Furukawa, and Y. Miura, "Glycome as biomarkers," in *General Methods in Biomarker Research and their Applications*, vol. 1–2, V. R. Preedy and V. B. Patel, Eds. Dordrecht: Springer Netherlands, 2015, pp. 111–140.

[14] W. Peng *et al.*, "Clinical application of quantitative glycomics," *Expert Rev. Proteomics*, vol. 15, no. 12, pp. 1007–1031, Dec. 2018.

[15] J. E. Huffman *et al.*, "Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults," *Hum. Mol. Genet.*, vol. 20, no. 24, pp. 5000–5011, Dec. 2011.

[16] S. Z. Sharapov *et al.*, "Defining the genetic control of human blood plasma N-glycome using genome-wide association study.," *Hum. Mol. Genet.*, vol. 28, no. 12, pp. 2062–2077, Mar. 2019.

[17] D. D. Gorev *et al.*, "GWAS-MAP: a platform for storage and analysis of the results of thousands of genome-wide association scans," in *Bioinformatics of Genome Regulation and Structure$\$Systems Biology (BGRS$\$SB-2018)*, 2018, p. 43.

[18] L. Klarić *et al.*, "Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases," *Sci. Adv.*, vol. 6, no. 8, p. eaax0301, Feb. 2020.

DOI 10.18699/BGRS/SB-2020-065

Peak caller comparison through quality control of ChIP-Seq datasets

Ruslan N. Sharipov
BIOSOFT.RU, LLC;
Novosibirsk State University
Novosibirsk, Russia
shrus79@biosoft.ru

Yury V. Kondrakhin
Institute of Computational Technologies SB RAS;
BIOSOFT.RU, LLC
Novosibirsk, Russia
yvkondrat@mail.ru

Semyon K. Kolmykov
FRC Institute of Cytology and Genetics SB RAS;
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
kolmykovsk@gmail.com

Ivan S. Yevshin
Institute of Computational Technologies SB RAS;
BIOSOFT.RU, LLC
Novosibirsk, Russia
ivan@biosoft.ru

Anna S. Ryabova
Institute of Computational Technologies SB RAS;
BIOSOFT.RU, LLC
Novosibirsk, Russia
anna@biosoft.ru

Fedor A. Kolpakov
Institute of Computational Technologies SB RAS;
BIOSOFT.RU, LLC
Novosibirsk, Russia
fedor@biosoft.ru

Abstract — **Chromatin immunoprecipitation followed by high throughput sequencing, i.e. ChIP-Seq, is a widely used experimental technology for the identification of functional protein-DNA interactions. Nowadays, such databases as GTRD, ChIP-Atlas and ReMap systematically collect and annotate a large number of ChIP-Seq datasets generated by distinct peak callers, including MACS2. The quality control of such datasets is currently indispensable, since the peak callers may produce different results for the same ChIP-seq experiment. We have performed a comparative analysis of intensively used peak callers with the help of two metrics that control false positive/negative rates. We have found that MACS2 outperformed its competitors.**

Keywords — quality control, ChIP-Seq datasets, peak caller, false positives, false negatives, GTRD database

Introduction

Understanding the basic mechanisms of transcription regulation is a big problem in modern biology. Regulation of transcription is a complex process in which transcription factors play a key role. Nowadays, ChIP-Seq experiments are widely used to detect protein-DNA binding in whole genomes. To date, several databases, such as GTRD [1], ChIP-Atlas and ReMap, accumulate ChIP-Seq datasets obtained by applying various peak callers to the primary ChIP-Seq data. To control the quality of accumulated datasets distinct control metrics are used. For example, such well-known metrics as Non-Redundant Fraction (NRF), PCR Bottlenecking Coefficient 1 and 2 (PBC1 and PBC2), Normalized Strand Cross-correlation coefficient (NSC), and Relative Strand Cross-correlation coefficient (RSC) evaluate the quality of read alignments for individual genomes. These metrics were developed as part of the ENCODE project [2]. However, these metrics do not control false positive and false negative rates. Recently, two quality control metrics, namely, the False Positive Control Metric (FPCM) and the False Negative Control Metric (FNCM), were developed on the base of the population size estimation approach [3].

Several tens peak callers have been developed to generate transcription factor binding regions (TFBRs) from aligned ChIP-Seq data. [4] However, a comparative analyses [5, 6] of peak callers did not reveal so far the best among them. We performed comparative analysis of the most popular peak callers – GEM [7], MACS, MACS2 [8], SISSRs [9] and PICS [10]. For this purpose, we used the FPCM and FNCM metrics, as well as 8982 TFBR datasets from the GTRD database. The conducted comparative analysis showed that MACS2 outperformed its competitors.

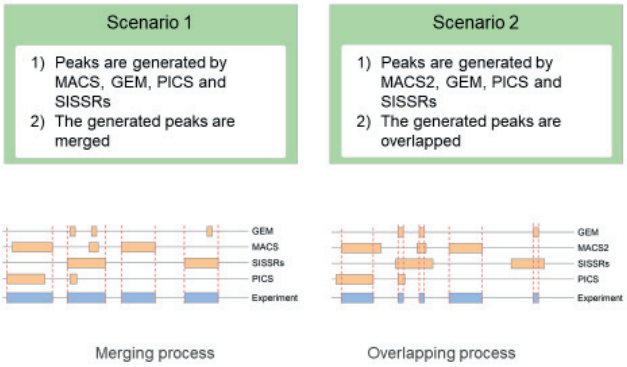
In this study we applied some rank aggregation (RA) methods for a meta-analysis of transcription factor binding sites (TFBSs) obtained from ChIP-seq data publically available in the GTRD database. Additionally, we are introducing a new RA method based on the Borda method utilizing values of FPCM and FNCM quality metrics.

Materials And Methods

To generate TFBR datasets, we used two distinct scenarios. According to scenario 1, see Fig. 1(A), the four peak callers - GEM, MACS, PICS, and SISSRs – were applied independently to the same ChIP-Seq set of reads aligned to the reference genome. Then the obtained four sets of peaks were merged into a final dataset. According to scenario 2, MACS was replaced by MACS2, and the final TFBR dataset was obtained by overlapping the peaks instead of merging them. The processes of merging and overlapping peaks are demonstrated in Fig. 1.

To compare peak callers, we used FPCM and FNCM metrics [3]. FNCM for each peak caller was defined as the ratio of the observed number of its peaks to the estimated number of genuine peaks. FNCM varies in the range [0.0; 1.0]. The closer the FNCM value to 1.0, the lower the false-negative rate, and the values closer to 0.0 indicate that a large number of genuine peaks have been missed. FPCM was defined as the ratio of the observed number of orphans in the TFBR dataset to the estimated number of true orphans. In turn, orphans were defined as separate peaks that did not overlap with other initial peaks. If the difference between the observed and estimated number of orphans is insignificant, then the FPCM should be close to 1. Such FPCM values indicate that erroneously formed peaks are practically absent. However, if the FPCM considerably exceeds 1 (e.g., FPCM > 2.0 or FPCM > 3.0), then at least half or more orphans are classified as false positives.

(A)



(B)

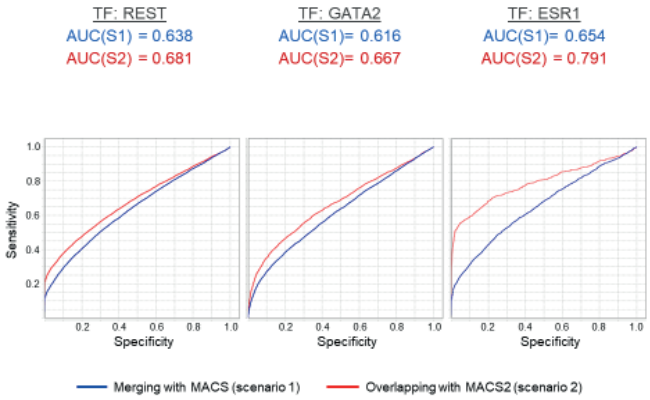


Fig. 1. Two scenarios for peak caller comparison. (A) scenario definition; (B) ROC curves and AUC values in two scenarios

Results And Discussion

To obtain reliable conclusions, we performed a comparative analysis of 8982 human ChIP-Seq experiments stored in GTRD. For each experiment, we calculated two FPCM values, say FPCM1 and FPCM2, which corresponded to scenario 1 and scenario 2, respectively. In general, in 70.9% of experiments, replacing MACS with MACS2 resulted in improved quality by lowering FPCM. In particular, for experiments PEAKS034562 (REST), PEAKS033231 (GATA2) and PEAKS034509 (ESR1), the pair of values (FPCM1, FPCM2) is equal to (14.431, 1.167), (10.778, 1.132) and (11.514, 1.204).

We performed a direct comparison of peak callers by comparing their FNCM. Thus, for experiment PEAKS034562 in scenario 1 the following FNCM values were achieved: MACS – 0.967, SISSRs – 0.909, GEM – 0.807 and PICS – 0.04. Hence, MACS has outperformed its competitors because its FNCM is maximal. Overall, in 56.1% of the experiments MACS showed better results, while GEM, SISSRs, and PICS outperformed in 18.5%, 16.2% and 9.2% of experiments, respectively. In scenario 2, MACS2 showed better results in 69.5%, while GEM, SISSRs and PICS outperformed in 8.4%, 13.9% and 8.2% of experiments, respectively.

Finally, the usefulness of transition from scenario 1 to scenario 2 can be confirmed by increasing the accuracy of identifying site motifs in the three ChIP-Seq datasets mentioned above. To identify the motif, we used the following models of position weight matrix from the HOCOMOCO database [11]:

REST_HUMAN.H11MO.0.A,
GATA2_HUMAN.H11MO.0.A,
ESR1_HUMAN.H11MO.0.A.

Fig. 1(B) demonstrates that the transition from scenario 1 to scenario 2 increased the accuracy of site identification. This increase in accuracy is in good agreement with the decrease in FPCM values.

Conclusions

Comparative analysis of GEM, MACS, MACS2, SISSRs and PICS revealed that MACS2 outperformed its competitors in terms of the FPCM and FNCM metrics.

Acknowledgment

This work was supported by the Russian Science Foundation, grant agreement 19-14-00295.

References

[1] I. Yevshin, R. Sharipov, S. Kolmykov, Y. Kondrakhin, and F. Kolpakov F, “GTRD: a database on gene transcription regulation-2019 update”. Nucleic Acids Res., vol. 47(D1), pp. D100–D105, January 2019.

[2] ENCODE Project Consortium, “An integrated encyclopedia of DNA elements in the human genome”, Nature, vol. 489, pp. 57–74, September 2012.

[3] S. K. Kolmykov, Y. V. Kondrakhin, I. S. Yevshin, R. N. Sharipov, A. S. Ryabova, and F. A. Kolpakov, “Population size estimation for quality control of ChIP-Seq datasets”, PLoS One, vol. 14(8), e0221760, August 2019.

[4] R., Thomas, S. Thomas, A. K. Holloway, and K. S. Pollard, “Features that define the best ChIP-Seq peak calling algorithms”, Brief Bioinform, vol. 18(3), pp. 441–450, May 2017.

[5] T. D. Laajala, S. Raghav, S. Tuomela, R. Lahesmaa, T. Aittokallio, and L. L. Elo, “A practical comparison of methods for detecting transcription factor binding sites in ChIP-seq experiments”, BMC Genomics, vol. 10(1), 618, December 2009.

[6] H., Koohy, T. A. Down, M. Spivakov, and T. Hubbard, “A comparison of peak callers used for DNase-Seq data”, PLoS ONE, vol. 9(5), e96303, May 2014.

[7] Y. Guo, S. Mahony, and D.K. Gifford, “High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints”, PLoS Comput. Biol., vol. 8(8), e1002638, August 2012.

[8] Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, and et al., “Model-based analysis of ChIP-Seq (MACS)”, Genome Biol., vol. 9(9):R137, September 2008.

[9] L. Narlikar and R. Jothi, “ChIP-Seq data analysis: identification of protein-DNA binding sites with SISSRs peak-finder”, Methods Mol. Biol., vol. 802, pp. 305–322, November 2011.

[10] X. Zhang, G. Robertson, M. Krzywinski, K. Ning, A. Droit, S. Jones, and et al., “PICS: probabilistic inference for ChIP-seq”, Biometrics, vol. 67(1), pp.151–163, March 2011.

[11] I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, A. V. Soboleva, A. S. Kasianov, H. Ashoor, and et al., “HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models”, Nucleic Acids Res., vol. 44(D1), pp. D116–D125, January 2016.

DOI 10.18699/BGRS/SB-2020-066

GWAS-MAP: the platform for analysis of results of genome-wide association studies

Tatiana Shashkova
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia
t.shashkova@nsu.ru

Sodbo Sharapov
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Denis Gorev
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Yakov Tsepilov
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Yurii Aulchenko
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Eugene Pakhomov
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Lennart Karssen
PolyKnomics
's-Hertogenbosch, Netherlands

***Abstract* — We present the database and the high-performance GWAS-MAP platform for aggregating, storing, processing and visualizing big data, namely the results of genome-wide and regional association studies (GWAS and RWAS). At the moment, we have compiled a database of human GWAS and RWAS, which contains more than 70 billion associations between genomic polymorphisms and different diseases, quantitative and "omics" traits. The platform and the database can be used in studies of the etiology of human diseases, in the development of predictive risk models, as well as in the search for candidate biomarkers, and candidate therapeutic targets.**

Keywords* — *GWAS-MAP, database, genome-wide association study, summary statistics, GWAS, regional-wide association study, RWAS, quantitative genetics

Introduction

The genome-wide association study (GWAS) is one of the main approaches for identifying the relationships between allelic variants of the genome and human complex traits. One of the most important advantages of GWAS is the agnostic nature of the method. The GWAS methodology has gained great popularity over the past decade [1]. Since 2007, the number of GWAS has been growing exponentially; At present, hundreds of original genome-wide association studies are published each year.

The results of GWAS are published in the form of summary statistics - for each single nucleotide polymorphism (SNP) information is provided on the frequencies of alleles and characteristics of the effect of the allele on the phenotype. For the analysis of human genomes, a GWAS assumes a study of associations between the character of interest and tens of millions of SNPs. In this regard, the results of GWAS (R-GWAS) form big data.

R-GWAS can be used to solve many problems - from research in the field of fundamental biology and genetics to the search for biomarkers and therapeutic effects.

While the number of R-GWAS collected by the scientific community and their analysis methods is increasing, the use of this data is limited by the lack of generally accepted standards. In particular, researchers are faced with the following problems. Firstly, this is a big amount of data that requires the development of specialized infrastructure. Secondly, in addition to storing primary data, it is necessary to ensure their pre-processing. Thirdly, the same type of data is generated in different laboratories using different protocols, which requires an integration procedure to ensure universal storage. Finally, it is necessary to develop user interfaces that would allow non-bioinformaticians access and analyze the data.

To help addressing these issues, we developed the GWAS-MAP platform for storing, processing, and analysis of R-GWAS. The system provides an opportunity to carry out research that will contribute to the search for new candidate biomarkers and therapeutic approaches.

Results

The GWAS-MAP consists of two data processing modules (integration and analysis of GWAS/RWAS results), an interface to the database (DB) and a web-interface (Fig. 1). Integration involves converting R-GWAS, collected from different sources, into a universal format. After integration, we perform QC and, if the R-GWAS passes the tests, we keep it in the database.

The DB consists of two components, each controlled by its own DB management system (DBMS). The first component is managed by the Clickhouse DBMS, which provides us with rapid-access storage of RGWAS accessible via powerful and flexible interface. Each record is a vector of parameters characterising an association between a SNP and a trait in a GWAS. The second component is managed by the PostgreSQL DBMS and contains information about the study design, results of analyses and other supporting information. To ensure convenient usage of GWAS data, the system also contains key information about each study, for example, date of publication, reference, and details of the experiment.

Currently, the GWAS DB contains in total data from 7,245 genome-wide and 1,25 millions regional (e.g. cis-eQTL) association scans describing 70 billions of genotype-phenotype associations. Tirty-six percents of associations are

related to complex traits, 12% are related to transcriptomics, 4% - to metabolomics, 47% - proteomics and 1% - glycomics. To ensure convenient usage of GWAS data, the system also contains key information about each study, for example, date of publication, reference, and details of the experiment.

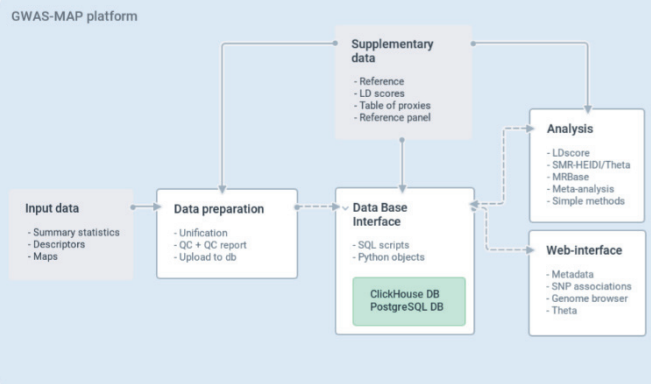


Fig. 1. GWAS-MAP platform program architecture

The system incorporates widely used methods of RGWAS analysis. These methods are focused on the identification of functional variants, genes, molecules, and traits that may be potential targets of (therapeutic) manipulation. In particular, we implemented methods of

- Linkage Disequilibrium score regression - a method for assessing the SNP-heritability of a trait and calculating genetic correlations between two traits [2].
- Mendelian randomization methods - a set of tests that allow hypothesising a causal relationship between a pair of traits [3].
- Summary-level mendelian randomization and heterogeneity in dependent instruments (SMR-HEIDI) which helps determining whether two different traits are controlled by the same functional variant in the locus or are controlled by two different functional variants in linkage disequilibrium. When used with gene expression GWAS, this can permit the establishment of the causal gene in a locus, which is not readily attained using a trait GWAS alone [4].
- Dimitrieva-Georges Theta (DGT) - is an alternative to HEIDI. This method uses weighted correlation analysis to evaluate the similarity of association profiles based on summary statistics only. The method may be more robust than SMR/HEIDI because it does not rely on availability and precise knowledge of the structure of linkage disequilibrium [5].

We also implemented meta-analysis tools that allow combining results of GWAS performed for the same trait. This module has additional QC procedure, to check that studies could be included in analysis together [6].

Finally, we developed a web-service "PheLiGe" (https://phelige.com), in order for external users to have access to the database. PheLiGe allows not only for scanning the phenome for effects of a genetic variant, but also to assess whether co-association of a variant with multiple traits may occur due to pleiotropy or the consequence of linkage. In more detail, in the "GWAS/cisQTL" tab, a user can search for specific GWAS and RWAS on multiple criteria, access metadata and generate interactive Manhattan plots. In the "Associations" tab the user can search for associations observed for a SNP of interest directly or via a proxy variant in linkage disequilibrium. The user can download the results and also graphically explore regional pattern of association around the selected SNP. Next, the user can select a "primary" and "secondary" regional association, that will be passed to the "Analysis" tab. In this tab, regional patterns of association are compared using the Dmitrieva-Georges weighted correlation method.

Conclusion

In this work, we introduced the GWAS-MAP platform, intended for storage and processing of GWAS and RWAS results. Currently, the database system contains data on more than 70 billion genotype-phenotype associations. The system interface provides a universal way of working with data, which allows quickly embed new analysis methods in the system. Users can work with the system through command line utilities, with the help of which they have the ability to download data to the system, as well as run analyzes and get their results. Approbation of GWAS-MAP platform is presented in an article on varicose veins study [7].

Acknowledgment

This work was funded by the TOP100 program of the Novosibirsk State University and by PolyKnomics BV.

References

[1] Visscher P. M. et al. "10 years of GWAS discovery: biology, function, and translation" The American Journal of Human Genetics, 2017, pp. 5-22.

[2] Bulik-Sullivan B. K. et al. "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies" Nature genetics, 2015, pp. 291.

[3] Hemani G. et al. "MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations" *bioRxiv*, 2016, pp. 078972.

[4] Zhu Z. et al. "Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets" Nature genetics, 2016, pp. 481.

[5] Momozawa Y. et al. "IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes" Nature communications, 2018, pp.2427.

[6] Winkler TW et al. "Quality control and conduct of genome-wide association meta-analyses" Nature Protocols, 2014, pp. 1192–1212

[7] Shadrina AS., et al. "Varicose veins of lower extremities: Insights from the first large-scale genetic study" PLoS genetics, 2019, pp. e1008110.

DOI 10.18699/BGRS/SB-2020-067

Hemolymph metagenome of endemic amphipod *Eulimnogammarus verrucosus* from Lake Baikal

Ekaterina Shchapova
Irkutsk State University, Irkutsk, Russia
shchapova.katerina@gmail.com

Natalia Belkova
Scientific Centre for Family Health and Human Reproduction Problems, Irkutsk, 664033 Russia
nlbelkova@gmail.com

Renat Adelshin
Irkutsk Anti-Plague Research Institute of Siberia and Far East, Irkutsk, Russia;
Irkutsk State University, Irkutsk, Russia
adelshin@gmail.com

Anton Gurkov
Irkutsk State University, Irkutsk, Russia
a.n.gurkov@gmail.com

Maxim Timofeyev
Irkutsk State University, Irkutsk, Russia
m.a.timofeyev@gmail.com

***Abstract* — A diverse number of studies indicates that the hemolymph of aquatic invertebrates, apparently healthy invertebrates, is non-sterile. In this study we for the first time characterized the bacterial diversity in hemolymph of endemic amphipods from Lake Baikal on the example of abundant littoral species *Eulimnogammarus verrucosus*.**

Keywords — hemolymph, amphipod, bacteria, Baikal

Motivation and aim

Motivation

The aquatic environment contains a diverse and rich bacterial flora and, because of this fact, the presence of microbiota in the hemolymph is mostly reported for aquatic invertebrates [1]. Moreover, some microorganisms not only survive in the animal hemolymph, but also can be beneficial for the host [2]. Certain local microorganisms can compete with incoming ones and stimulate the host immune system, but in some cases, they can also become potential pathogens [3, 4]. The hemolymph microbiota, therefore, is a factor that must be considered when predicting the impact of negative environmental conditions on invertebrate populations. Crustaceans, in particular diverse endemic amphipods (Amphipoda), are one of the main ecosystem components in pristine Lake Baikal, but their hemolymph microbiomes have never been investigated using modern sequencing techniques.

Aim

The aim of the study was to characterize the hemolymph microbiome of one of the most abundant littoral amphipods in Lake Baikal, *Eulimnogammarus verrucosus* (Gerstfeldt, 1858), by 16S rDNA amplicon meta-sequencing.

Methods

DNA was isolated from pooled hemolymph of adult amphipods caught on the Baikal shoreline from two locations separated by the Angara River outflow. The first group of animals was collected in the port Baikal in June of 2016 (2 prepared hemolymph samples, 3 individuals per pool) and the second group was caught in Listvyanka village in September of 2016 (2 samples, 4 animals per pool). Amphipods were acclimated to laboratory conditions in aquaria for about a week. The surface of amphipod chitin was sterilized with ethanol before hemolymph extraction. The DNA samples were analyzed using a 454 GS Junior sequencer (Roche) after prior amplification of V4-V6 variable regions of 16S rDNA (500F primer: CCAGCAGCYGCGGTAAN; 1064R primer CGACRRCCATGCANCACCT). The IDTAXA program with the GTDB database was used for taxonomic classification of the obtained amplicon libraries [5].

Results

Microbiomes of the analyzed hemolymph samples of *E. verrucosus* were generally similar within each of the locations but varied in composition between the locations. In particular, we identified in total 29 bacterial genera, but only five of them were present in all samples: *Acinetobacter, Flavobacterium, Arcicella, Polaromonas,* and *Pseudomonas*. As the core components of *E. verrucosus* hemolymph metagenome these genera may deserve the main attention during further research on host-symbiont relationships for amphipods of Lake Baikal.

Acknowledgment

The study was funded by the RFBR grant # 19-34-90137.

References

[1] Wang, X. W., & Wang, J. X. (2015). Crustacean hemolymph microbiota: Endemic, tightly controlled, and utilization expectable. Molecular immunology, 68(2), 404-411.

[2] Desriac, F., Le Chevalier, P., Brillet, B., Leguerinel, I., Thuillier, B., Paillard, C., & Fleury, Y. (2014). Exploring the hologenome concept in marine bivalvia: haemolymph microbiota as a pertinent source of probiotics for aquaculture. FEMS microbiology letters, 350(1), 107116.

[3] Schmitt, P., Rosa, R. D., Duperthuy, M., de Lorgeril, J., Bachère, E., & Destoumieux-Garzón, D. (2012). The antimicrobial defense of the Pacific oyster, Crassostrea gigas. How diversity may compensate for scarcity in the regulation of resident/pathogenic microflora. Frontiers in microbiology, 3, 160.

[4] Zhang, X., Sun, Z., Zhang, X., Zhang, M., & Li, S. (2018). Hemolymph microbiomes of three aquatic invertebrates as revealed by a new cell extraction method. Appl. Environ. Microbiol., 84(8), e02824-17.

[5] Murali, A., Bhargava, A., & Wright, E. S. (2018). IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome, 6(1), 1-14.

DOI 10.18699/BGRS/SB-2020-068

Computational pipeline for genomic epidemiology surveillance of pathogenic bacteria

Andrey Shelenkov
Central Research Institute of Epidemiology
Moscow, Russia
0000-0002-7409-077X

Yurii Yanushevich
Central Research Institute of Epidemiology
Moscow, Russia
0000-0001-9061-752X

Yulia Mikhaylova
Central Research Institute of Epidemiology
Moscow, Russia
mihailova@cmd.su

Vasiliy Akimkin
Central Research Institute of Epidemiology
Moscow, Russia
0000-0003-4228-9044

***Abstract* — Due to rapid development of the second and third generations of high-throughput sequencing systems, the number of available pathogenic bacteria genomes has increased dramatically in last 10 years and reached 60,000 in 2019. Novel software and computational pipelines are required to process such a huge amount of data in order to derive useful information from it. Here we present a pipeline combining various existing bacterial typing and annotation tools to create an easy-to-use and flexible environment for investigating the whole genomes of pathogenic bacteria obtained by high-throughput sequencing. The pipeline includes multi-locus sequence typing, antibiotic resistance and virulence genes' annotation, plasmid detection and typing, and various other useful features.**

Keywords* — *annotation pipeline, bacterial genome analysis, NGS, antibiotic resistance, genomic epidemiology

Introduction

The development of second- and third-generation sequencing technologies (NGS) has dramatically simplified and greatly lowered the cost of obtaining the whole genomes, especially, bacterial ones. In the field of clinical bacteriology, rapid and reproducible identification, typing and annotation of bacterial pathogens is vitally important, and, thus, NGS slightly becomes a commonly used practice for outbreak and infection transmission investigation. Although many computational pipelines for bacterial genome annotation are available, currently the need still exists for flexible tools that can be routinely used in clinical and epidemiological investigations. We have developed such a pipeline for typing and annotation of clinical and food poisoning bacterial pathogens. It can be used for scientific purposes in the emerging field of bacterial genomic epidemiology that combines the use of molecular biology, epidemiology and bioinformatics techniques and procedures for infectious disease surveillance and outbreak investigations.

Matherials and Methods

To achieve the goals set above, we have developed a set of custom scripts for seamless integration of various software tools within Linux ecosystem. The input data should include genome sequences (complete chromosomes or assembled contigs) in FASTA format. Batch processing of multiple genomes of the same or various bacterial species is possible; however, species name should be set for each group of organisms separately. First step of the pipeline includes running BLAST on 'nt' or other similar database locally to verify that genomic sequence really belongs to the organism of interest. If it does not, all downstream species-specific tests are skipped. Mixtures of different organisms are also reported.

Then MLST-typing [1] is performed if the MLST scheme is available for the species under study. After that, ResFinder [2] is called to check for the presence of known acquired antibiotic-resistance genes in genomic sequence. This search is not species-specific, so it can be performed for any bacteria. Next steps include searching for known virulence genes with VirulenceFinder [3] and detecting plasmids in input sequence using PlasmidFinder [4]. These programs rely on the databases that store the data only for some bacterial families. If the data for the current species is not available, these steps will be excluded from the pipeline.

Then additional typing is performed when it is available for the given species. For example, the typing schemes based on capsular (K, or KL) loci and oligosaccharide (O) loci are available for *Klebsiella pneumoniae* and *Acinetobacter baumannii*, SPA-typing scheme based on the sequence of a polymorphic variable number tandem repeat is used for *Staphylococcus aureus* etc.

Finally, the results obtained by the programs mentioned above are summarized and presented in concise tabular form for each defined group of bacteria.

Some additional processing steps including phylogenetic tree building are also available as optional procedures.

It is worth mentioning that the pipeline can be easily updated by adding new antibiotic resistance genes, virulence genes or plasmid sequences for various organisms to the existing local databases. These updates are as simple as adding or editing text files.

Results

The pipeline developed is routinely used in our group for rapid annotation of newly sequenced clinical bacterial isolates, mainly belonging to ESKAPE pathogen group and food poisoning pathogens. The combination of various typing schemes (MLST, cgMLST, KL, O) allows narrowing isolate classification and facilitates the investigations of antimicrobial resistance spreading within the given healthcare unit, hospital or even at the country or international level.

The example of main output file in tabular format is presented in the Table 1.

TABLE 1 – EXEMPLARY PIPELINE OUTPUT

Sample ID	Species	MLST	AR genes[a]	AR[a]	Virulence genes	KL-type
N1_49	Klebsiella pneumoniae	ST25	aac(3)-IIa, aadA1, strA; bla-OXA-48	Aminoglycoside; Beta-lactam	mrkA2, mrkB2	KL2
N2_56	Klebsiella pneumoniae	ST395	fosA	Fosfomycin	fyuA_2, irp1285	KL61

[c.] In 'AR genes' and 'AR' columns ';' is used as a separator between antibiotic groups

Conclusions

We have developed a user-friendly computational pipeline for typing and annotation of pathogenic bacteria genomes including antibiotic resistance prediction, MLST-based sequence type assignment, typing based on capsular and oligosaccharide loci, virulence gene finding and plasmid revealing, and other useful features including phylogenetic tree building. It can become a useful tool for epidemiological and genomic studies of newly sequenced bacterial pathogens. The pipeline can be updated or expanded without substantial efforts.

REFERENCES

[1] M.V. Larsen, S. Cosentino, S. Rasmussen, "Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria", J. Clin. Microbiol., vol. 50, pp. 1355-1361, 2012.

[2] E. Zankari, H. Hasman, S. Cosentino et al., "Identification of acquired antimicrobial resistance genes", J Antimicrob. Chemother., vol. 67, pp. 2640-2644, 2012.

[3] K.G. Joensen, F. Scheutz, O. Lund et al, "Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli", J. Clin. Microbiol., vol. 52, pp.1501-1510, 2014.

[4] A. Carattoli, E. Zankari, A. Garcia-Fernandez et al. "In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing", Antimicrob. Agents Chemother., vol. 58, pp. 3895-3903, 2014.

DOI 10.18699/BGRS/SB-2020-070

Allelic drop-out is a common phenomenon reducing the diagnostic yield of PCR-based target sequencing

Shestak A.G.
Petrovsky National Research Center of Surgery, Moscow, Russia
Cardiac Electrophysiology Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical

Bukaeva A.A.
Petrovsky National Research Center of Surgery, Moscow, Russia
Cardiac Electrophysiology Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical

Zaklyazminskaya E.V.
Petrovsky National Research Center of Surgery, Moscow, Russia
Cardiac Electrophysiology Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical

Saber S.
Petrovsky National Research Center of Surgery, Moscow, Russia
Cardiac Electrophysiology Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical

Abstract — **Allelic drop-out (ADO) is a known phenomenon of selective allele amplification representing the potential problem of correct DNA diagnostics. The results demonstrate the incidence of ADO reducing the diagnostic yield of PCR-based target sequencing**

Motivation and aim

Allelic drop-out (ADO) is a known phenomenon of selective allele amplification representing the potential problem of correct DNA diagnostics. Both NGS and Sanger sequencing are PCR-based methods, Sanger sequencinq is used to verify NGS results. The aim of this study is to demonstrate the incidence of ADO reducing the diagnostic yield in primary cardiomyopathy genetic testing via semiconductor NGS and Sanger sequencing of target gene panels.

Methods

We have developed 3 AmpliSeq custom gene panels for mutational screening: "K^+/Na^+ ion channels", "Desmosomal proteins", "Sarcomeric proteins", contains 1049 primer pairs (37 genes) totally, 152 kb. About 140 probands were screened with at least one of these gene panels. AmpliSeq sequences were analyzed in silico and visually compared with Sanger control sequences, noting the facts of heterozygosity loss.

Results

We have detected 12 ADO cases both in Sanger (5 cases) and AmpliSeq (7 cases) sequencing data. All ADO events happened due to frequent or rare SNVs in the oligoprimer annealing sites and were detected due to mismatch in frequent SNPs zygosity nearby. Three pathogenic variants would be missed if were not revealed by re-sequencing with alternative method and alternative oligos.

Conclusion

All PCR-based methods have a risk of ADO leading to a decrease of diagnostic yield of genetic testing. ADO can theoretically affect 1% amplicons. It seems that real scope of ADO might be much higher and depends on numbers of primer pairs. The software for ADO detection is needed.

ACKNOWLEDGMENT

This work was supported by Russian Science Foundation grant No. 16-15-10421.

DOI 10.18699/BGRS/SB-2020-071

Genome-wide association study reveals novel genetic variants associated with HIV-1C infection in Botswana population

Andrey K. Shevchenko
Theodosius Dobzhansky Center for Genome Bioinformatics
St.-Petersburg State University
St-Petersburg, Russia
ORCID: 0000-0002-7332-5661

Sergey V. Malov
Theodosius Dobzhansky Center for Genome Bioinformatics
St.-Petersburg State University
St-Petersburg, Russia
ORCID: 0000-0003-0093-6506

Alexey Antonik
Theodosius Dobzhansky Center for Genome Bioinformatics
St.-Petersburg State University
St-Petersburg, Russia
ORCID: 0000-0003-4341-1291

***Abstract* — Genome wide association studies (GWAS) allow to identify common variants associated with the trait in question. In order to efficiently search for the genetic associations we have previously developed Genome-Wide Association Tracks Chromosome Highway (GWATCH). The broad goal of the Botswana GWAS project is to identify genetic determinants of susceptibility and resistance to infection by HIV-1 subtype C among people severely affected by HIV/AIDS in Botswana. By conducting GWAS analysis on HIV1C case/control dataset consisting of 762 Tswana people (combined from two partly overlapping datasets of 809 microarray and 362 WGS samples), we found several gene regions slightly below significance level.**

Keywords — GWAS, GWATCH, HIV-1, AIDS, Botswana

Introduction

HIV-infection is a deadly chronic disease spreading across all continents. There are 37.9 million [32.7 million–44.0 million] people in the world living with HIV; 35.4 million [25.0 million–49.9 million] people who died from AIDS-related conditions since the onset of the epidemic several decades ago with the estimated 770 000 [570 000–1.1 million] who died in 2018 (UNAIDS, 2019). Despite the efforts and resources dedicated to the fight against the HIV epidemics for more than three decades, currently no effective vaccine or cure exists.

According to UNAIDS, Botswana falls into the top 3 countries with the highest HIV prevalence in the world with an adult aged 15 to 49 HIV prevalence rate of 20.3 [17.3 - 21.8] (UNAIDS, 2019). Moreover, HIV-1 subtype C (widespread in Botswana), which accounts for over half of HIV-1 infections worldwide, is understudied, because most current studies focus on other HIV subtypes more prevalent in Europe and the USA. We wanted to fill this gap and to search for genetic factors associated with HIV-1 subtype C in Botswana population.

Methods

The individuals used in this study originated from two primary populations that have been described in detail. One was a population of pregnant women (436) involved in a maternal transmission study; the other (373) was a population enrolled to identify acute or early infections. This data were obtained in collaboration with Harvard T.H. Chan School of Public Health AIDS Initiative (Boston, USA) and Botswana Harvard AIDS Institute (Gaborone, Botswana). These samples were genotyped using two technologies: 809 samples have microarray-based genotypes (this dataset is called BH-IM), and 364 samples (335 of which are overlapping with these 809) have WGS-based genotypes. The resulting genotype data was used for the association tests, performed using PLINK logistic regression. These results were loaded into the GWATCH tool [Svitin et al, 2014; http://gen-watch.org/] for further analysis and visualization.

Results

First, we were able to replicate known HIV-associated loci from [O'Brien 2013] (NCOR2, TRIM5, CXCL12 and several HLA genes). Next, we investigated the newly discovered loci associated with HIV-1 subtype C. Many of the top associated genes are known to play a role in pathways that may be important in HIV infection process, such as L3MBTL3, NEO1, CD79B, AP3B1, PTPRA and others. One of the identified loci (close to HCG22 gene) was also replicated in another dataset [Xie et al. 2016].

Conclusions

This work presents the first step towards identification of genetic variants associated with infection by HIV-1 subtype C in Botswana population. We have replicated several loci that have shown association with HIV in previous studies. In addition, we have identified several novel putative associated loci in such genes as NEO1, CD79B, L3MBTL3, AP3B1, PTPRA.

Acknowledgment

This research was supported in part by St.-Petersburg State University (grant No. 1.52.1647.2016)

References

[1] Svitin A. et al. 2014. GWATCH: a web platform for automated gene association discovery analysis. GigaScience.

[2] Xie W. et al. 2016. Genome-Wide Analyses Reveal Gene Influence on HIV Disease progression and HIV-1C Acquisition in Southern Africa. AIDS Research and Human Retroviruses.

[3] O'Brien SJ, Hendrickson SL. Host genomic influences on HIV/AIDS. Genome Biol. 2013 Jan 31;14(1):201. doi:10.1186/gb-2013-14-1-201

DOI 10.18699/BGRS/SB-2020-072

Genome-centered integrated instrumental information system modeling and interpretation of human and virus omics

Anatoliy Shlikht
dept. life safety
Far Eastern Federal University
Vladivostok, Russia
schliht@mail.ru

Natalia Kramorenko
dept. business informatics and economic-mathematical methods
Far Eastern Federal University
Vladivostok, Russia
kramnat@mail.ru

***Abstract* — Creating an integrated genome-centered intellectual system based on highly structured databases and knowledge bases for modeling and interpreting human and virus omics. The system allows you to automatically find functionally significant structures (templates, motifs, epitopes) for genomic and proteomic data in the formation of nucleo-protein complexes and antigen-antibody complexes, determine protein functions, and model spatio-temporal biochemical and physiological processes.**

Keywords* — *genomics, proteomics, databases, knowledge bases, modeling, interpretation

Introduction

Genome-centered analysis and interpretation faces the challenges of storing, searching, and modeling big omics data (genomics, transcriptomics, proteomics, metabolomics, and others). This fully applies to the tasks of viral infections. When solving problems of virology (molecular diagnostics, vaccination, search for therapeutic drugs), it is necessary to move from big omics data to functionally significant molecular structures: mutations, motives, conservative domains, enzymes, metabolic reactions and pathways, physiological systems, diseases based on bioinformatics technologies. This approach allows us to build a complete picture of the biochemical and physiological processes in a living organism, in particular, when the human organism interacts with viruses.

Methods and Algorithms

The goal of this work is to create an integrated genome-centered intellectual system based on highly structured databases and knowledge bases for processing omics data in the tasks of integrating genomic virology with the human genome.

Currently existing big weakly structured omics data from international portals (NCBI, KEGG, Ensembl, HGNC, etc.) with interactive data access technologies do not allow to effectively solve the problems of integration of genomic virology and the human genome. An effective solution to these problems can be obtained and associated with the creation of highly structured indexed omics databases and knowledge bases, followed by automatic analysis and interpretation of primary data based on expert systems technology. It is necessary to restructure primary data as much as possible, automate interactive scenarios, shifting them to intelligent expert systems.

Results

Highly structured indexed models of omics data of numerous viruses and the reference human genome have been developed [1,2], on the basis of which databases and knowledge bases have been created, which form the basis of expert systems for automatic analysis, interpretation and diagnosis of viral diseases.

The expert system (ES) has been created that automatically searches for functionally significant structures (patterns, motifs, epitopes) for both genomic data and amino acid sequences of protein in the formation of nucleo-protein complexes and antigen-antibody complexes.

Special filters have been developed that work both on the basis of Prosite patterns (https://prosite.expasy.org/) and arbitrary templates, allowing you to find functionally significant domains and define protein functions. For automatic operation of these filters, a special compiler has been developed that allows modeling various search structures and templates.

The ES automatically searches for mutations for the studied genes and proteins. ES allow efficient automatic analysis and interpretation of genomic and proteomic data, and subsequent diagnosis of diseases.

A distinctive feature of the developed ES is their Autonomous nature, which allows you to work autonomously with both the reference human genome and virus genomes without constantly accessing data from world portals. At the same time, the mode of periodic updating of omics data from world portals for data updating is provided.

The created ES allow us to trace deep connections, starting from a mutation in the gene along the chain: (DNA, RNA) – gene – information RNA – encoded protein – enzyme – metabolic reaction – metabolic pathway – physiological system – disease. This view allows the medic, biologist, geneticist, and researcher to trace the entire depth of biological and physiological processes. Created databases and knowledge bases allow you to determine the genotype of humans and a virus.

For the convenience of working with the developed systems (databases and knowledge bases), specialists who do not know information technologies have developed a special visual dialog interface. The interface is based on simple graphical constructs of visual programming.

The system can also be used effectively to analyze the results of genomic sequencing with subsequent interpretation of mutations. Based on the developed filters, the most significant mutations and their manifestation in proteins and metabolic processes are identified.

Conclusion

The developed information tool systems (databases and knowledge bases, ES) make it possible to effectively analyze the primary sequences of DNA, RNA, proteins for humans

and viruses. The created systems proved their effectiveness in diagnostics, as well as in other tasks of molecular biology: search for motives, sites, epitopes; modeling of transcripts; modeling of proteins; modeling of virus strategies; obtaining large omics quantitative data with subsequent transformation into small qualitative semantic data.

REFERENCES

[1] A.G.Shlikht, N.V.Kramorenko, "Deep bioinformatics expert system of analysis, modeling and interpretation of omics BigData of the human genome", The 3rd International Symposium "Mathematical Modeling and High-Performance Computing in Bioinformatics, Biomedicine and Biotechnology (MM-HPC-BBB-2018)". Novosibirsk, 2018. P. 66.

[2] A.G.Shlikht, N.V.Kramorenko, "Artificial intelligence in the problems of analysis and interpretation of omics human data", The 11th International Conference "Bioinformatics of Genome Regulation and Structure\Systems Biology (BGRS\SB-2018)". Novosibirsk, 2018. P. 78.

DOI 10.18699/BGRS/SB-2020-073

lncRNAs – their potential in regulation of hypertension and behavior of ISIAH rats

Ivan Sidorenko
Laboratory of Molecular Plant Pathology,
Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
vanyasidorenko22@gmail.com

Vladimir Babenko
Laboratory of Human Molecular Genetics, Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
bob@bionet.nsc.ru

Arcady Markel
Laboratory. of Evolutional Genetics,
Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
markel@bionet.nsc.ru

Olga Redina
Laboratory. of Evolutional Genetics,
Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
oredina@bionet.nsc.ru

Abstract — **Long non-coding RNAs (lncRNAs) play an important role in the control of many biological processes in the body, including the development of cardiovascular diseases and hypertension. It is believed that lncRNAs play a central role in the epigenetic control of gene expression, however, the understanding of lncRNA biological functions and interactions is still far from being complete. In this work, we identified the lncRNAs differentially expressed in the hypothalami of hypertensive ISIAH and normotensive WAG rats, and revealed lncRNA-associated differentially expressed genes (DEGs) related to hypertension and behavioral characteristics of ISIAH rats (grooming, vertical activity, hyperactivity, abnormal emotion/affect behavior (including abnormal response to novelty). The work was carried out using transcriptome sequencing (RNA-Seq method). Three lncRNAs (Bc1, RGD1562890, and Snhg4) were found, the expression of which differed in the hypothalami of hypertensive ISIAH and normotensive WAG rats. The largest number of co-regulated genes, both associated with hypertension and behavior, was found for Snhg4. These findings may be useful for further understanding the role of lncRNAs in regulating the protein coding genes and modulating processes associated with both hypertension and behavior.**

Keywords — RNA-Seq, lncRNA, hypertension, behavior, hypertensive ISIAH rats

Introduction

Long non-coding RNAs (lncRNAs) play an important role in the control of many biological processes in the body, including the development of cardiovascular diseases [1] and hypertension [2-5]. Although they have a central role in the epigenetic control of gene expression [4], the understanding of lncRNA biological functions and interactions is still far from being complete [5]. The development of NGS technologies, the results of which allow to identify the lncRNA-associated genes by calculating the correlation coefficient between lncRNA and mRNA expression levels [6-7], leads to a rapid increase in the number of studies in this field.

ISIAH rats model the stress-sensitive form of arterial hypertension. They are characterized by elevated blood pressure, characteristic structural and functional changes in target organs, increased stress sensitivity under the influence of mild emotional stress, and specific behavior in an unfamiliar environment. Increased activity of the hypothalamic-pituitary-adrenal and sympathetic adrenal systems suggest a key role for the hypothalamus in regulating the manifestation of the above phenotypic characteristics of hypertensive ISIAH rats [8]. Earlier, using transcriptome sequencing (RNA-Seq method), we determined genes differentially expressed in the hypothalami of hypertensive ISIAH and control normotensive WAG rats [9]. The current work presents the results of identification of differentially expressed lncRNAs in the hypothalami of the same rats and reveals the lncRNA-associated protein coding DEGs related to hypertension and behavioral phenotypes of ISIAH rats (grooming, vertical activity, hyperactivity, abnormal emotion/affect behavior and abnormal response to novelty).

Methods

Animals

The work was carried out using hypertensive ISIAH/Icgn (Inherited Stress Induced Arterial Hypertension) rats and normotensive WAG/GSto-Icgn (Wistar Albino Glaxo) rats at the age of 3 months. Rats were kept under standard conditions with free access to food and water. All animal experiments were approved by the Institute's Animal Care and Use Committee. Hypothalami from 3 rats in each group were collected and stored in RNA Later (Qiagen, Chatsworth, CA) at -70oC until sequencing.

RNA-Seq analysis

The collected samples were sent to JSC Genoanalytica (Moscow, Russia), where transcriptome sequencing (RNA-Seq) was performed. The details of the protocol used were described earlier [9 4]. The data obtained were mapped to the RGSC Rnor_5.0 \ rn5 reference genome. The Cufflinks / Cuffdiff programs were then used to reveal the differentially expressed transcripts. The p value cutoff for lncRNA was 0.01. The RNA-Seq data were deposited in the NCBI Short Read Archive database with Accession number PRJNA299102.

Identification of the lncRNA-associated genes

In order to obtain the lncRNA-mRNA co-expression pairs, the data (RPKM values) were log transformed, centered and normalized. Pearson correlation analysis was conducted between the lncRNAs expression and the expression profile of the protein-coding DEGs detected in the same samples. A

correlation coefficient >0.7 was recognized as significant for the identification of the lncRNA-associated genes.

Databases used

Rat Genome Database (https://rgd.mcw.edu/) was used to identify the genes associated with the hypertension and behavioral phenotypes (behavior/neurological phenotype, grooming, vertical activity, hyperactivity, abnormal emotion/affect behavior (including abnormal response to novelty). Atlas of combinatorial transcriptional regulation in mouse and man [10] was used to reveal the DEGs encoding the transcription factor genes.

Results

Three lncRNAs (Bc1, RGD1562890, and Snhg4) were found, the expression of which differed in the hypothalami of hypertensive ISIAH and normotensive WAG rats (Table 1). The relationship between these lncRNAs and the expression level of protein coding DEGs associated with hypertension and behavioral phenotypes was calculated using the Pearson correlation coefficient. Among the hypothalamic protein coding DEGs that were associated with hypertension, one gene (*Hyal1*) associated with RGD1562890 (r = -0.73) and 12 DEGs associated with Snhg4 were found. Among them, the most significant association was found with the *Igfbp2* gene (r = 0.97), which is associated with insulin resistance in the development of hypertension.

Among the hypothalamic protein coding DEGs associated with behavioral phenotypes, one gene (*Ercc2*) associated with Bc1 (r = 0.88), two genes (*Evc* and *Foxg1*) associated with RGD1562890 (r = -0.78 and r = 0.83), and 9 genes associated with Snhg4 were found. Among them, the most significant association was identified with the *Igf2* gene (r = 0.88). It should be noted that the *Ercc2* and *Foxg1* genes encode transcription factors, and the *Igf2* gene is associated with cerebrovascular disorders.

TABLE 1 – The lncRNAs differentially expressed in the hypothalami of ISIAH and WAG rats

lncRNA	Expression level in		log2 (fold_change) ISIAH/WAG	p_value
	ISIAH, RPKM	*WAG, RPKM*		
Bc1	137705.0	98177.6	0.49	1.85E-02
RGD1562890	247.4	51.4	2.27	7.00E-04
Snhg4	11.2	34.2	-1.61	4.50E-03

Conclusion

In hypothalami of hypertensive ISIAH and control WAG rats, 3 differentially expressed lncRNAs and several lncRNA-associated protein-coding DEGs were found. The most highly expressed lncRNA (Bc1) is likely to influence the behavior of ISIAH rats by regulating the expression of the *Ercc2* gene, which encodes the transcription factor and is associated with abnormal emotion/affect behavior and abnormal response to novelty, but does not affect the regulation of genes directly associated with hypertension. According to our results, we can expect that lncRNA RGD1562890 is involved both in the regulation of processes associated with hypertension and in the regulation of behavior, in particular with hyperactivity, through the regulation of the *Foxg1* gene encoding the transcription factor. The largest number of co-regulated genes, both associated with hypertension and behavior, was found for Snhg4. These findings may be useful for further understanding the role of lncRNAs in regulating the protein coding genes and modulating processes associated with both hypertension and behavior.

ACKNOWLEDGMENT

The work was supported by Russian Foundation for Basic Research grant No. 20-04-00119a and by Budget project No. 0324-2019-0041-C-01.

REFERENCES

[1] S. Shen, H. Jiang, Y. Bei, et al., "Long Non-Coding RNAs in Cardiac Remodeling," Cell. Physiol. Biochem., vol. 41, no. 5, pp. 1830-1837, 2017.

[2] K. Gopalakrishnan, S. Kumarasamy, B. Mell, and B. Joe, "Genome-wide identification of long noncoding RNAs in rat models of cardiovascular and renal disease," Hypertension, vol. 65, no. 1, pp. 200-210, 2015.

[3] L. Hou, Z. Lin, Y. Ni, et al., "Microarray expression profiling and gene ontology analysis of long non-coding RNAs in spontaneously hypertensive rats and their potential roles in the pathogenesis of hypertension," Mol. Med. Rep., vol. 13, no. 1, pp. 295-300, 2016.

[4] G. Wu, P.A. Jose, and C. Zeng, "Noncoding RNAs in the Regulatory Network of Hypertension," Hypertension, vol. 72, no. 5, pp. 1047-1059, 2018.

[5] C. Leimena and H. Qiu, "Non-Coding RNA in the Pathogenesis, Progression and Treatment of Hypertension," Int. J. Mol. Sci., vol. 19, no. 4, pii. E927, 2018.

[6] Y. Ma, T. Luo, D. Dong, et al., "Characterization of long non-coding RNAs to reveal potential prognostic biomarkers in hepatocellular carcinoma," Gene, vol. 663, pp. 148-156, 2018.

[7] L. Ma and C. Deng, "Identification of a novel four-lncRNA signature as a prognostic indicator in cirrhotic hepatocellular carcinoma," PeerJ, vol. 7, pp. e7413, 2019.

[8] A. L. Markel, O. E. Redina, M. A. Gilinsky, G. M. Dymshits, E. V. Kalashnikova, Y. V. Khvorostova, L. A. Fedoseeva, and G. S. Jacobson, “Neuroendocrine profiling in inherited stress-induced arterial hypertension rat strain with stress-sensitive arterial hypertension,” J. Endocrinol., vol. 195, no. 3, pp. 439-450, 2007.

[9] L.O. Klimov, N.I. Ershov, V.M. Efimov, et al., "Genome-wide transcriptome analysis of hypothalamus in rats with inherited stress-induced arterial hypertension," BMC Genet., vol. 17, Suppl. 1., pp. 13, 2016.

[10] T. Ravasi, H. Suzuki, C.V. Cannistraci, et al., "An atlas of combinatorial transcriptional regulation in mouse and man," Cell, vol. 140, no. 5, pp. 744-752, 2010.

DOI 10.18699/BGRS/SB-2020-074

Transcriptome analysis of the Baikal whitefish head kidney (*Coregonus* sp., Coregonidae)

Tuyana Sidorova
Laboratory of Ichtiology
Limnological Institute, Siberian Branch
of the Russian Academy of Sciences
Irkutsk, Russia
tuyana_be@mail.ru

Lubov Sukhanova
Laboratory of Ichtiology
Limnological Institute, Siberian Branch
of the Russian Academy of Sciences
Irkutsk, Russia
lsukhanova@yandex.ru

Ivan Kutyrev
Laboratory of Parasitology and
Ecology of Aquatic Organisms
Institute of General and Experimental
Biology, Siberian Branch of the
Russian Academy of Sciences
Ulan-Ude; Institute of Biology, Irkutsk
State University, Irkutsk Russia
ikutyrev@yandex.ru

Olga Mazur
Laboratory of Parasitology and
Ecology of Aquatic Organisms
Institute of General and Experimental
Biology, Siberian Branch of the
Russian Academy of Sciences
Ulan-Ude, Country
olmaz33@yandex.ru

Kirill Khabudaev
Department of Cell Ultrastructure
Limnological Institute, Siberian Branch
of the Russian Academy of Sciences
Irkutsk, Russia
onexkirill@gmail.com

***Abstract* — Tapeworms of the genus *Dibothriocephalus* are widely distributed all around the world, with some being agents of human diphyllobothriasis, one of the most important fishborne zoonosis caused by a cestode parasite. The study of immunological relations between plerocercoids of *Dibothriocephalus* dendriticus and its salmonid fish hosts is of great importance. Hereby, we carried out complete transcriptome analyses of whitefish *Coregonus baicalensis* head kidney. Using Blast2Go, we revealed transcripts distribution among categories: biological processes, molecular functions.**

Keywords* — *whitefish, head kidney, transcriptome

Introduction

Coevolution between hosts and their parasites is predicted to be dynamic and rapid, mainly because fitness costs caused by parasites are high, parasites are ubiquitous, and they often evolve rapidly. The majority of parasites have evolved strategies to evade the immune response of their hosts. Tapeworms of the genus *Dibothriocephalus* are widely distributed all around the world, with some being agents of human diphyllobothriasis, one of the most important fishborne zoonosis caused by a cestode parasite. Humans can contract this parasite by eating raw or partially cooked fish containing *Dibothriocephalus* plerocercoids [1]. We began the study of immunological relations between plerocercoids of *Dibothriocephalus dendriticus* and its salmonid fish hosts. The aim of this study was to conduct transcriptome analysis of the whitefish *Coregonus baicalensis* head kidney.

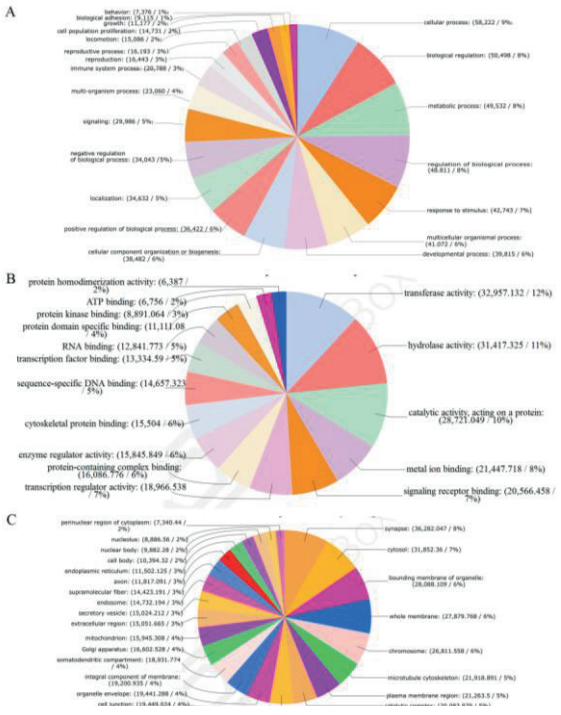
Material and methods

Samples of whitefish head kidneys were homogenized in liquid nitrogen and RNA was extracted using the RNEasy Mini (Qiagen) kit, according to manufacturer's instructions. Complementary DNA (cDNA) libraries were then prepared using 250-500 ng of total RNA. Illumina sequencing adapters were then ligated to the cDNA, which were then PCR amplified for 15 cycles of 98 ºC 10 second denaturing, 60 ºC 30 second annealing and 72 ºC 30 second extension and using a final extension step of 72 ºC for 5 minutes. The libraries were then used for clusters generation Illumina NextSeq550 and NextSeq 550 High Output v2 Kit (300 cycles) chemistry. High quality reads were merged and assembled using Trinity (with default parameters). Open reading frames were predicted using Transdecoder (http://transdecoder.sf.net) with default parameters. Conserved domains and associated Gene Ontology (GO) annotation [2] were predicted using Blast2Go.

Results and discussion

For the first time, complete transcriptome analyses of Baikal whitefish *C. baicalensis* head kidney was done. In total, 90253643 Illumina paired-end reads were generated from the cDNA libraries. Using Blast2Go, we revealed transcripts distribution among categories: biological

Fig. 1. Gene Ontology annotations. Classification of the *C. lavaretus* head kidney proteins in terms of molecular functions (A), biological processes (B), and cellular components (C).



processes, molecular functions (Fig. 1). For biological processes predominant transcripts were: cellular processes (9%), biological regulation (8%), metabolic processes (8 %), and regulation of biological processes (8%). Among molecular functions the next transcripts prevailed: transferase activity (12 %), hydrolase activity (11%), catalytic activity, acting on protein (10%). Synapse (8%) cytosol (7%), bounding membrane of organelle (6%), whole membrane (6%), and chromosome (6%) were predominant transcripts in cellular components category.

Obtained transcriptome of *C. baicalensis* head kidney will be used further to search immunocompetent genes involved in immune response against helminth invasion. Obtained information on the genes will be necessary for study of cellular and molecular relations between cestode plerocercoids and fish host immune system using real-time polymerase chain reaction, flow cytometry or confocal microscopy.

ACKNOWLEDGMENT

This study was performed under the theme of Russian government task (government registration No. AAAA-A17-117011810039-4) and funded by grant of the Russian Foundation of Fundamental Researches (grant No. 18-34-20015) and the Ministry of Science and Higher Education of the Russian Federation (project No. FZZE-2020-0026; I.A.K.).

REFERENCES

[1] J. Dupouy-Camet, H. Year, Diphyllobothrium, in: D. Liu (Ed.), Molecular detection of foodborne pathogens, Taylor and Francis, London and New York, 2010, pp. 781–788.

[2] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Geneontology: tool for the unification of biology. The Gene Ontology Consortium. NatGenet 2000;25:25–9. doi:10.1038/75556.

DOI 10.18699/BGRS/SB-2020-075

Differentially expressed microRNAs in carotid paraganglioma

Anastasiya Snezhkina
EIMB RAS, Moscow, Russia
leftger@rambler.ru

Elena Pudova
EIMB RAS, Moscow, Russia
pudova_elena@inbox.ru

Vladislav Pavlov
EIMB RAS, Moscow, Russia
vladislav1pavlov@gmail.com

Maria Fedorova
EIMB RAS, Moscow, Russia
fedorowams@yandex.ru

George Krasnov
EIMB RAS, Moscow, Russia
gskrasnov@mail.ru

Anna Kudryavtseva
EIMB RAS, Moscow, Russia
rhizamoeba@mail.ru

Abstract — Molecular mechanisms underlying the formation of rare tumors are insufficiently known. It the work, we performed the analysis of microRNA profiles in a rare neuroendocrine tumor of the head and neck, carotid paraganglioma (CPGL). We found four microRNAs (miR-1185-1/2-3p, miR-431-5p, miR-504-5p, and miR-200c-3p) that are differentially expressed among CPGL samples studied. This is the first study of microRNA expression in CPGL. Obtained results allow a better understanding of molecular changes associated with the development of the tumor.

Keywords — carotid paragangliomas, microRNA expression, tumorigenesis

Motivation and aim

Motivation

Carotid paraganglioma (CPGL) is a rare neuroendocrine tumor of the head and neck that arises from the microscopic paraganglion at the carotid artery bifurcation. CPGL amounts at more than half of all head and neck paragangliomas (65%). According to the WHO classification 2017, CPGL is characterized by the variable potential to metastasis; biomarkers of the tumor aggressiveness have not been found. Paragangliomas can occur as both hereditary and sporadic forms. Up to 40% of paragangliomas are associated with germline mutations in several known genes, predominantly in SDHx [1]. However, due to the rarity of the disease, there are little data on its genetics and molecular mechanisms behind tumor development.

Aim

We perform a complex molecular genetic study of CPGL, including whole exome, transcriptome, microRNA, and methylome analysis [2-4]. The integration of these data allows a better understanding of molecular changes associated with the tumor development and progression, as well as identifying both specific hallmarks of CPGL and common features attributed to all paragangliomas.

Methods

Total RNA was isolated from 48 tumor samples using the High Pure FFPET RNA Isolation Kit (Roche, Switzerland). MicroRNA libraries were prepared with the NEBNext Multiplex Small RNA Library Prep Kit for Illumina (NEB, USA) according to the manufacturer's protocol. High-throughput sequencing was performed on the Illumina NextSeq 500 System in the single-end mode with a length of 36 bp. Bioinformatics analysis of microRNA data was carried out using the miRge 2.0 pipeline [5].

Results

We found a list of microRNAs that are differently expressed in CPGL. Four microRNAs (miR-1185-1/2-3p, miR-431-5p, miR-504-5p, and miR-200c-3p) were characterized by the highest variability in the expression among tumor samples studied. All these microRNA were previously shown to be involved in carcinogenesis.

Acknowledgment

This work was supported by the Russian Science Foundation, grant no. 17-75-20105. The study was performed using the equipment of the EIMB RAS "Genome" center (http://www.eimb.ru/ru1/ckp/ccu_genome_c.php).

References

[1] Zhikrivetskaya S.O. et al. (2017) Molecular markers of paragangliomas/pheochromocytomas. Oncotarget. 8(15): 25756-25782.

[2] Snezhkina A.V. et al. (2018) Exome analysis of carotid body tumor. BMC Med Genomics. 11(Suppl 1): 17.

[3] Kudryavtseva A.V. et al. (2019) Mutational load in carotid body tumor. BMC Med Genomics. 12(Suppl 2): 39.

[4] Snezhkina A.V. et al. (2019) Novel potential causative genes in carotid paragangliomas. BMC Med Genetics. 20(Suppl 1): 48.

[5] Lu Y. et al. (2018) miRge 2.0 for comprehensive analysis of microRNA sequencing data. BMC Bioinformatics. 19(1): 275.

DOI 10.18699/BGRS/SB-2020-076

The rich inner world of colorado potato beetles – a metagenomic survey of viral diversity in public data

Maria Starchevskaya
SRC VB "Vector" Rospotrebnadzor,
Koltsovo, Russia
starchevskayamaria@mail.ru

Yuri Vyatkin
Novosibirsk State University,
Novosibirsk, Russia
Novel Software Systems LLC,
Novosibirsk, Russia
vyatkin@gmail.com

Denis Antonets
SRC VB "Vector" Rospotrebnadzor,
Koltsovo, Russia
Novel Software Systems LLC,
Novosibirsk, Russia
antonec@yandex.ru

***Abstract* — Recent metagenomic studies of various environments revealed an enormous number of viruses, with hundreds of previously unknown species and even new virus families, and thus greatly expanded our understanding of the virosphere. Investigations of insect virome will provide new insights into ecology and evolution of viruses and their hosts. Colorado potato beetle is one of the most serious insect pests feeding on *Solanaceae* plants. The Colorado potato beetle virome has not been studied yet and, as far as we know, our work is the first attempt to discover viral material in DNA- and RNA-Seq data of *L. decemlineata*. We also hope that identification of new viruses will help to extend the arsenal of biopesticides.**

Keywords* — *Colorado potato beetle, metagenomic analysis, virome, insect viruses

Motivation and aim

Colorado potato beetle (*Leptinotarsa decemlineata* from *Chrysomelidae* family) is one of the most serious insect pests feeding on *Solanaceae* plants. Originated from North America it became widespread worldwide and poses a significant threat to potato crops [1]. Besides, leaf feeding beetles transmit plant viruses [2,3]. Currently, new biological strategies to control the leaf feeding beetles are actively being developed [4–6], but all of them are based on either entomopathogenic fungi or bacteria, not viruses. However, entomopathogenic viruses are successfully used against Lepidoptera pests for decades [7,8]. The Colorado potato beetle virome has not been studied yet and, as far as we know, our work is the first attempt to discover the viral material in DNA- and RNA-Seq data of *L. decemlineata*. We also hope that identification of new viruses will help to extend the arsenal of biopesticides.

The main goal of this work was the identification of viral sequences from publicly available *L. decemlineata* transcriptomic and genomic samples.

Methods

L. decemlineata transcriptomic and genomic samples were taken from NCBI SRA. To minimize contamination the reads were filtered against UniVec database, containing the sequences of commonly used vectors, adapters, linkers, and primers, and also against human genome (GRCh38). The reads that were not aligned to *L. decemlineata* genome (GCA_000500325.2, Ldec_2.0, 640 Mbp) were assembled *de novo* with MetaSPAdes software [9]. Contigs with sufficiently high coverage were analyzed with BLAST and MetaGeneMark software [10]. The reads not aligned to reference genome were aligned to viral genomes extracted from NCBI RefSeq. Non-aligned reads and contigs were also analyzed with Kraken2 software [11].

Results

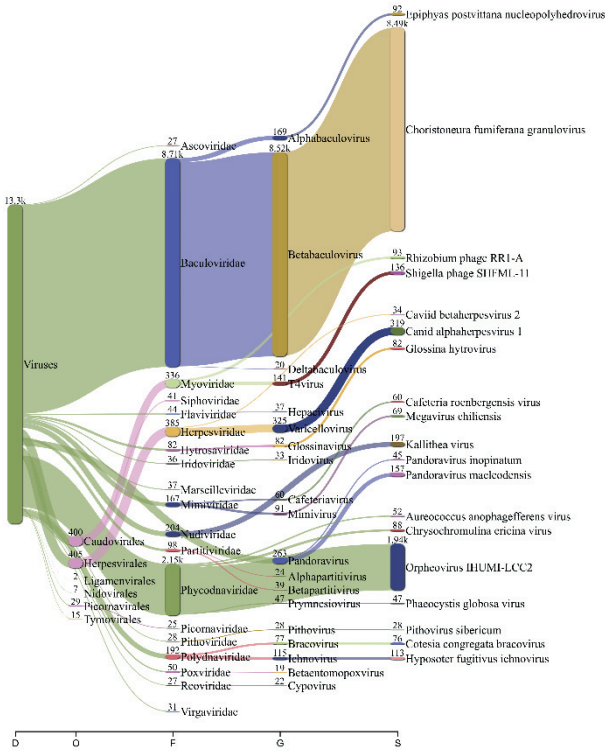


Fig. 1. Classification of reads from pooled RNA-Seq samples from *L. decemlineata* heads. The analysis was performed with Kraken2 [11]. The classification results were visualized using Pavian [12]. Only viral reads are shown

As it was expected, the most frequently found sequences belonged to common insect viruses of Iridoviridae, Baculoviridae and Iflaviridae families (Fig. 1). We were able to find whole genome sequences of several plant viruses: Potato virus S, Potato virus Y, Grapevine virus, Ligustrum virus. It was found that some genes predicted with MetaGeneMark probably belonged to Bracovirus genus. Further examination of Colorado potato beetle reference genome revealed numerous long fragments of Bracovirus genomic sequences (over 25,000 bp in total). Bracoviruses belong to Polydnaviridae virus family associated with parasitoid wasps. This is a very interesting family of viruses as they help to suppress the immune responses and alter the growth, the metabolism and endocrine system balance of wasp hosts [13]. Bracoviruses are also involved into horizontal gene transfer and they were shown to integrate into genomes of wasp hosts [14]. We analyzed 18 more Coleoptera genomes and found that Colorado potato beetle genome was the fourth most enriched with bracoviral sequences. As far as we know,

this is the first report of a polydnavirus found in *Chrysomelidae* species. However, this fact is not surprising as there is a wasp *Edovium puttleri* known to parasitize on Colorado potato beetle eggs [15] and it was even proposed as a potential biocontrol agent. We have also found a number of long contigs with relatively high reads coverage that we couldn't identify yet.

REFERENCES

[1] Krawczyk K. et al. (2015) Prevalence of Endosymbionts in Polish Populations of Leptinotarsa decemlineata (Coleoptera: Chrysomelidae), J. Insect Sci., 15:106.

[2] Wickizer S.L., Gergerich R.C. (2007) First Report of Japanese Beetle (Popillia japonica) as a Vector of Southern bean mosaic virus and Bean pod mottle virus, Plant Dis., 91:637–637.

[3] Kersch-Becker M.F., Thaler J.S. (2014) Virus strains differentially induce plant susceptibility to aphid vectors and chewing herbivores, Oecologia, 174:883–892.

[4] Charpentier G. et al. (2002) Characterization of cell lines developed from the Colorado potato beetle, Leptinotarsa decemlineata Say (Coleoptera: Chrysomelidae), In Vitro Cell. Dev. Biol. Anim., 38:73–78.

[5] Martin P.A.W. et al. (2007) Toxicity of Chromobacterium subtsugae to southern green stink bug (Heteroptera: Pentatomidae) and corn rootworm (Coleoptera: Chrysomelidae), J. Econ. Entomol., 100:680–684.

[6] Yu Y. et al. (2016) Construction of an environmental safe Bacillus thuringiensis engineered strain against Coleoptera, Appl. Microbiol. Biotechnol., 100:4027–4034.

[7] Cuartas P.E. et al. (2019) Novel biopesticide based on Erinnyis ello betabaculovirus: characterization and preliminary field evaluation to control Erinnyis ello in rubber plantations, Pest Manag. Sci., 75:1391–1399.

[8] Moore S. et al. (2015) The Cryptophlebia Leucotreta Granulovirus—10 Years of Commercial Field Use, Viruses, 7:1284–1312.

[9] Bankevich A. et al. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing, J. Comput. Biol., 19:455–477.

[10] Zhu W. et al. (2010) Ab initio gene identification in metagenomic sequences, Nucleic Acids Res., 38:e132.

[11] Wood D.E., Salzberg S.L. (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments, Genome Biol., 15:R46.

[12] Ye X. et al. (2018) Parasitoid polydnaviruses and immune interaction with secondary hosts, Dev. Comp. Immunol., 83:124–129.

[13] Strand M.R., Burke G.R. (2015) Polydnaviruses: From discovery to current insights, Virology, 479–480:393–402.

[14] Lashomb J. et al. (1987) Edovum puttleri (Hymenoptera: Eulophidae), an Egg Parasitoid of Colorado Potato Beetle (Coleoptera: Chrysomelidae): Development and Parasitism on Eggplant, J. Econ. Entomol., 80:65–68.

DOI 10.18699/BGRS/SB-2020-077

A new method for combining of genetically correlated traits by maximizing of their shared heritability

Gulnara R. Svishcheva
Laboratory of recombination and segregation analysis
Institute of cytology and genetics
Novosibirsk, Russia

Evgeny S. Tiys
Laboratory of glycogenomics
Institute of cytology and genetics
Novosibirsk, Russia

Sofya G. Feoktistova
Kurchatov genomic center
Institute of cytology and genetics
Novosibirsk, Russia

Elizaveta E. Elgaeva
Laboratory of recombination and segregation analysis
Institute of cytology and genetics
Novosibirsk, Russia

Sodbo Sharapov
Laboratory of recombination and segregation analysis
Institute of cytology and genetics
Novosibirsk, Russia

Yakov A. Tsepilov
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Abstract — **Genetic correlations between phenotypic traits are widely observed phenomena. Many groups of traits have a significant level of shared genetic background. In this work, we suggested a new method for extraction of shared genetic component for genetically correlated phenotypic traits. We applied the method to GWAS results for anthropometric traits.**

Keywords — shared heredity, anthropometry, GWAS.

Introduction

In the era of large sample sizes available for genome wide association studies (GWAS), it becomes clear that a remarkable number of human traits and diseases are highly genetically correlated. In other words, many groups of traits have a significant level of shared genetic background. There is a big temptation to be able to study this background thought analysis a new aggregate trait that approximates shared heritability of the studied traits.

Materials and methods

We propose a new strategy to the SNP-based association analysis of multiple genetically correlated traits. This approach focuses on a linear combination of the traits by maximizing of heritability explained by SNPs associated with all the traits (named as 'shared' SNPs). The first step is to decompose a matrix of genetic correlations between traits into two matrices so that one of them is explained by shared SNPs, while the other represents the genetic correlations formed by the unique genetic background of each trait. We compared several matrix decomposition techniques, including eigen-, CUR-, and NMF- decompositions. Maximum quality was achieved when we used a method that minimizes absolute values of trait-trait correlations explained the unique genetic background. The second step allows us to analytically estimate the linear combination weights of traits by maximizing their shared heritability explained only the shared SNPs. Finally, in the last step, GWAS for analyzing shared heredity is performed as a linear combination of single-trait-level GWAS summary statistics.

To assess this method we used GWAS data from such widely studied anthropometry traits as weight, BMI, waist circumference and hip circumference. Data for anthropometric traits was downloaded from http://www.nealelab.is/uk-biobank. In the first step, we found coefficients of linear trait combinations that maximizes the shared heredity between them. In the next step, we used the estimated coefficients to construct GWAS summary statistics for shared heredity. The significant threshold was set as 5×10^{-8}. The SNPs were considered from one locus if they were located within 1000kb from top associated SNP. Using Fisher's exact test we found significant enrichment of top significantly associated SNPs with combined trait among top SNPs associated with all traits.

Results

We tested our method on simulated data using different scenarios. In each scenario, we randomly generated the correlation matrix component explained by the non-shared SNPs, using parameters $w \in (0,1)$ (the contribution of each trait to shared heritability) and $e \in (0,1)$ (variance of random variables that determine correlations caused by a unique genetic background). Besides, we formed the matrix explained by the shared SNPs, using w. In each scenario, 10,000 repeats were performed. For each repeat, we tested the generated matrices for compliance with the definition of a valid correlation matrix. For small and moderate e, and moderate or large w, the mean difference between the observed and expected w was very small (0.0041 ± 0.0008).

In analysis of real data, we found 729 loci significantly associated with common genetic component (shared heredity) for the anthropometric traits. Among them, 28 loci are not significantly associated with any of the original traits and may be considered as new.

We estimated the enrichment of SH loci among the loci that are significantly associated with all individual traits versus loci that were not significant with all individual traits at the same time. *The significance of enrichment was* $p=2.3\times10^{-32}$.

We plan to apply this method for such genetically correlated traits as plasma lipid levels and several psychiatric disorders.

Acknowledgment

The work of G.S. was supported by the RFBR grant (20-04-00464). The work of S.F., E.T. and SSh was funded by the Russian Science Foundation grant number 19-15-00115. The work of EEE and YAT was supported by the Russian Foundation for Basic Research (project 19-015-00151).

DOI 10.18699/BGRS/SB-2020-078

Results of the whole-genome sequencing and annotation of the phenotype of bacteria of the genus Listeria

Marina Terekhova
St. Petersburg State University of Information Technologies, Mechanics and Optics, Saint-Petersburg, Russia
n.mim@mail.ru

Rogacheva Elizaveta
Saint-Petersburg Pasteur Institute, Saint-Petersburg, Russia
elizvla@yandex.ru

Kraeva Lyudmila
Saint-Petersburg Pasteur Institute, Saint-Petersburg, Russia
lykraeva@yandex.ru

Derevyanchenko Irina
Branch of the Federal State Health Institution "Center for Hygiene and Epidemiology in the City of St. Petersburg", Saint-Petersburg, Russia
tamiirina110804@mail.ru

Abstract* — The incidence of Listeriosis in recent years is constantly increasing. So, in highly developed countries it ranges from 0.3 to 1.5, while in Russia - only 0.04%[1]. Moreover, the indicators in Moscow and St. Petersburg exceed those in the whole country by 4 times, which is explained by the different quality of clinical and laboratory diagnostics. Moreover, mortality due to complications of listeriosis infection can reach 30% [2]. Therefore, the fact of contamination of food products with listeria is of great concern. In this case, the strains of the species *L. monocytogenes, L. innocua, L. ivanovii, L. grayii, L. seeligeri, L. welshimeri* are most often distinguished. And if earlier it was believed that representatives of the species *L. monocytogenes* possess the largest set of pathogenicity factors, then the possibilities of molecular and bioinformatics research methods provide a reliable picture with respect to other representatives of the genus Listeria, in particular *L. innocua.

Keywords — Listeria, listeriosis, sequencing

Motivation and aim

Motivation

The incidence of Listeriosis in recent years is constantly increasing. Moreover, mortality due to complications of listeriosis infection can reach 30%. Therefore, the fact of contamination of food products with *Listeria* is of great concern.

Aim

To study the pathogenic potential of strains of *L. innocua.*

Methods

The research material was 11 strains of bacteria of the genus *Listeria* isolated from food (ready-to-eat and semi-finished products: 7 strains of *L. monocytogenes* and 3 strains of *L. innocua*), as well as 1 strain of *L. monocytogenes* from a patient with listeriosis. The classic bacteriological method, mass spectrometric, molecular (genome-wide sequencing), bioinformatic analysis of genetic data were used in the work.

Results

An analysis of the Listeria genomes showed that bacteria of the species *L. innocua* have in their genome a number of virulence genes characteristic of representatives of the species *L. monocytogenes*: clpC, which encodes the production of endopeptidase, prfA, which is responsible for the production of listeriolysin A, inlJ_1 and inlJ_2, which encode the production of protein internalin — an inducer of phagocytosis.

In addition, a plasmid with the ebrB gene characteristic of strains with multi-resistant properties was found in one strain of *L. innocua* isolated from a chicken semi-finished product. Moreover, the founded plasmid is derived from *L. monocytogenes*, and this species was also isolated from the same semi-finished product. All L. innocua strains had antibiotic resistance genes: from 20 to 30 for each strain. They included genes responsible for resistance to β-lactam antibiotics, fluoroquinolones, aminoglycosides, tetracyclines, macrolides, that is, to those groups of antibacterial drugs that are used in the treatment of listeriosis infection.

The results of phylogenetic studies showed a close relationship between the genomes of bacteria of the species *L. monocytogenes* and *L. innocua*. Therefore, the proximity of bacteria in food products, in places of their storage and transportation, can lead to faster transfer of genetic information from representatives of one species to another. At the same time, strains of *L. innocua* acquire pathogenicity and antibiotic resistance factors, which allows them to become hazardous to human health along with strains of the *L. monocytogenes* species. Therefore, further study of virulence markers in other representatives of the genus *Listeria* is necessary for therapeutic and anti-epidemic measures in foci of infection.

References

[1] Alexandra Moura et al. Atypical Hemolytic Listeria innocua Isolates Are Virulent, albeit Less than Listeria monocytogenes. Infect Immun. 2019 Apr; 87(4): e00758-18.

[2] Glaser P., Frangeul L., Buchrieser C. et al. Comparative genomics of Listeria species // Science. 2016. Vol. 294. P. 849–852.

DOI 10.18699/BGRS/SB-2020-079

Differentially methylation of *ANKRD53* and *GATA3* genes in human miscarriages with trisomy 16

Ekaterina Tolmacheva
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
kate.tolmacheva@medgenetics.ru

Stanislav Vasilyev
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
stanislav.vasilyev@medgenetics.ru

Oksana Vasilyeva
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
oksana.vasilyeva @medgenetics.ru

Tatjana Nikitina
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
t.nikitina@medgenetics.ru

Elena Sazhenova
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
elena.sazhenova@medgenetics.ru

Anton Markov
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
anton.markov@medgenetics.ru

Ekaterina Serdyukova
National Research Tomsk State University, Tomsk, Russia
katya.serdyukova.1997@mail.ru

Darja Zhigalina
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
dasha_150291@hotmail.com

Igor Lebedev
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
igor.lebedev@medgenetics.ru

***Abstract* — An excess dosage of the genes at the whole chromosome can lead to a disturbance of the epigenetic background of the entire genome. We analyzed the DNA methylation of several genes that are important for normal embryogenesis in miscarriages with aneuploidy and normal karyotype.**

Keywords — DNA methylation, miscarriage, trisomy 16

Motivation and aim

Motivation

Trisomy of chromosome 16 is the most common aneuploidy among human miscarriages. Potentially, a supernumerary chromosome may affect the genome-wide methylation and disrupt the expression of genes necessary for normal embryo development. Previously, we analyzed the genome-wide methylation in chorionic villi in miscarriages with trisomy 16 using the Infinium HumanMethylation27 BeadChip (Illumina). A significant increase in DNA methylation was found for the 90 genes (deltaB>0.15) in miscarriages with trisomy 16 compared to induced abortions with normal karyotype.

Aim

We suggested that abnormal methylation in certain genes may be common for miscarriages with aneuploidy of various chromosomes.

Methods

Methylation of the individual gene promoters was studied by targeted bisulfite massive parallel sequencing in chorionic villi in 48 miscarriages with aneuploidy karyotype (26 with trisomy 16 and 22 with other autosomal trisomies and monosomy of the X chromosome), 8 miscarriages with normal karyotype and 7 induced abortions.

Results

We assessed the methylation level in promoters of two top differentially methylated genes (deltaB>0.20) - ANKRD53 and GATA3. Methylation of the CpG sites of these genes did not differ in miscarriages with normal karyotype compared with induced abortions. Miscarriages with trisomy 16, but not miscarriages with other aneuploidies, had higher levels of DNA methylation in ANKRD53 (for 5 from 40 analyzed CpG-sites, $p<0.05$) and GATA3 (for 49 from 171 analyzed CpG-sites, $p<0.05$) genes. GATA3 gene encodes transcription factor, which is a key regulator of trophoblast differentiation, and imbalance of its expression can lead to aberrant trophoblast invasion. ANKRD53 is involved in spindle dynamics and nucleus integrity in mitosis. Thus, abnormal expression of these genes can dramatically affect embryonic development and lead to mosaic karyotypes that are often observed in cases of trisomy 16.

Acknowledgment

The study was supported by the Russian Science Foundation (19-74-10026).

DOI 10.18699/BGRS/SB-2020-080

Analysis of short- and long-range interactions within potential binding sites notably extends the fraction of verified peaks in ChIP-seq data

Anton Tsukanov
ICG SB RAS, Novosibirsk, Russia
tsukanov@bionet.nsc.ru

Victor Levitsky
ICG SB RAS, Novosibirsk, Russia
levitsky@bionet.nsc.ru

Tatyana Merkulova
ICG SB RAS, Novosibirsk, Russia
merkulova@bionet.nsc.ru

***Abstract* — We developed pipeline for integrative application of various de novo motif search tools to massive sequencing data. While traditional position weight matrices (PWMs) neglect dependencies between positions of motifs, the 'short-range interactions' markov models BAMM/InMode permit only local dependencies, the 'long-range interaction' model SiteGA allows dependencies between arbitrary positions. The massive analysis of ChIP-seq data have shown that the models BAMM/InMode and PWMs recognized similar and significantly overlapped peaks; but the notable fractions of peaks predicted by SiteGA model were not predicted by other models.**

Keywords* — *chromatin immunoprecipitation, transcription factor binding site recognition, short- and long-range dependencies between nucleotide positions in binding sites

Motivation and aim

Regulation of eukaryotic gene expression achieved through compound interaction of various transcription factors (TFs). Development and massive application of next generation sequencing technologies for mapping of TF binding sites (BS) in genome, in particular ChIP-seq, provides an opportunity to study gene expression regulation in detail. Widely applied approach of TFBS prediction, the position weight matrix (PWM) relied on a relatively short motifs and proposed the additivity of various positions within potential TFBS [1,2]. Recently, to supplement the verification of potential BSs in ChIP-seq data with PWMs, various approaches that taking into account intra-motifs dependencies were applied for wide-genome application, e.g. BaMM [3] and InMode [4]. These approaches use the markov models (MMs), which neglect the additivity assumption through the concept of the order of markov chain, i.e. a short distance for a given position that may contain other dependent nucleotides. Typically, the order of MM is changed from one to five, that's way MMs may be referred as to as 'short-range interaction' models.

In this study, we proposed new 'long-range interaction' de novo motif search model SiteGA that takes into account dependencies between arbitrary positions of sites. SiteGA approach combined discriminant analysis and genetic algorithm to construct a site model [5]. To compare our approach with traditional PWMs and BAMM/InMode models we developed the integrated pipeline for ChIP-seq data verification with multiple de novo motif search models. The aim of our research does not consist in a performance comparison between various complementary motifs recognition tools, rather, we tested whether alternative tools actually capable predicted structurally different motifs.

Methods

Earlier, the training of SiteGA model for the aligned sequence dataset defined an individual as a set of N locally positioned dinucleotides (LPD) frequencies [5]. SiteGA optimized the LPD set with genetic algorithm (GA), i.e. the final LPD set is the top-scoring result of the evolution of the population of individuals. The original SiteGA approach [5] required the same length W of all sequences from the training set; each LPD was defined by location [s, e] within the whole site [1,W] and dinucleotide type d, $1 \le d \le 16$, for each LPD. For each LPD $1 \le s,e < W$ and $0 \le s\text{-}e \le 5$. In the de novo SiteGA version we preserved the length of model W and the fitness function for an individual [5]. Since any de novo motif search works with peaks of various lengths {Ln} ($W \le Ln$, $1 \le n \le N$, N is the number of peaks), we added to the definition of individual the leftmost position and orientation on of the site W in a peak. Besides earlier applied mutation and recombination operators that changed location or dinucleotide type of each LPD within the site [1, W], de novo model added in GA implementation new mutation and recombination operators. New mutations can move a position or orientation of the window W in a peak. New recombination allows an exchange between two individuals that have sites in n-th and m-th peaks in certain positions/orientation. For de novo SiteGA models we chose the same site length W = 30 nt according to previous analysis [5].

We applied the standard cross-validation method to estimate the accuracy of the models. As a result ROC curve represents the dependence between True Positive and False Positive rates (TPR and FPR, respectively). The thresholds for all models was chosen uniformly with the fixed FPR value. As described earlier [6], FPRs were estimated as probabilities of hits for the respective human/mouse whole-genome dataset of upstream regions of human/mouse protein coding genes of length 2 kb. Peaks were classified according to presence/absence of predicted hits of various models. All supported methods (estimation of model accuracy, determination of the model threshold, classification of peaks) were implemented in the python programming language.

Results

The new pipeline was developed based on the PWM (ChIPMunk, [7]), InMoDe, BaMM and SiteGA models combined with supported methods to search TFBS in ChIP-seq data. The pipeline includes several main steps: 1) training models; 2) estimating models accuracy; 3) calculating models thresholds; 4) recognition of potential sites in peaks; 5) classification of peaks according the presence of sites and their overlap within a peak sequence.

Totally, we took over 100 ChIP-seq datasets from the Cistrome database (http://cistrome.org/, [8]) in analysis. We revealed that each model among all tested (PWM, InMoDe, BaMM, SiteGA) able to verify specific peaks that do not verified by any other models. The typical results of pipeline application for NR3C1 ChIP-seq dataset (GEO: GSM1122535, CistromeDB: 37598, cell type: epithelium) are shown in Fig. 1.

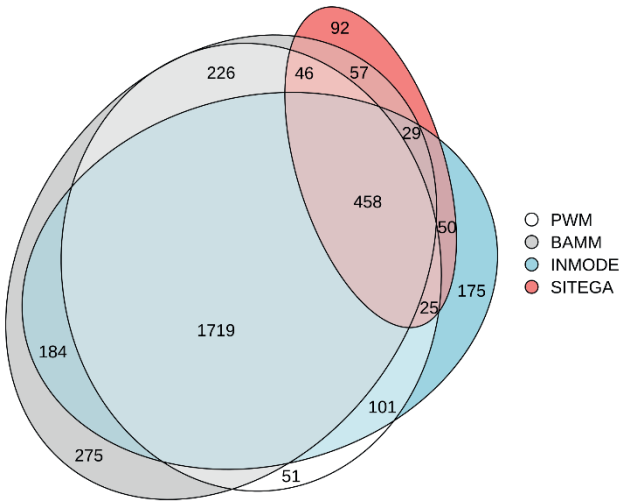


Fig. 1. The Venn diagram of the number of peaks containing predicted sites of sole models, as well as those containing predicted sites of various combinations of two, three and all four models. The diagram is drawn with eulerr [9]. ChIPseq dataset for NR3C1 we extracted from [8]. Totally, the ChIP-seq dataset consisted of 4000 peaks

ACKNOWLEDGMENT

The work was supported by Russian Foundation for Basic Research (project # 18-29-13040) and State Budget Project 0259-2019-0008-C-01.

REFERENCES

[1] Benos P.V. et al. (2002) Additivity in protein-DNA interactions: how good an approximation is it? Nucleic Acids Res., 30(20):4442-4451.

[2] Srivastava D and Mahony S. (in press) Sequence and chromatin determinants of transcription factor binding and the establishment of cell type-specific binding patterns. Biochim Biophys Acta Gene Regul Mech., 194443. doi: 10.1016/j.bbagrm.2019.194443.

[3] Siebert M. and Söding J. (2016) Bayesian Markov models consistently outperform PWMs at predicting motifs in nucleotide sequences. Nucleic Acids Res., 44(13):6055–6069.

[4] Eggeling R. et al. (2017) InMoDe: tools for learning and visualizing intra-motif dependencies of DNA binding sites. Bioinformatics, 33(4):580-582.

[5] Levitsky V.G. et al. (2007) Effective transcription factor binding site prediction using a combination of optimization, a genetic algorithm and discriminant analysis to capture distant interactions. BMC Bioinformatics, 8: 481.

[6] Levitsky V.G. et al. (2019) A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package. Nucleic Acids Res, 47(21): e139.

[7] Kulakovskiy I.V. et al. (2010) Deep and wide digging for binding motifs in ChIP-Seq data. Bioinformatics. 26(20):2622-2623.

[8] Mei S. et al. (2017) Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse, Nucleic Acids Res., 45(D1):, D658–D662,.

[9] Larsson J. (2019). eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 6.0.0, https://cran.r-project.org/package=eulerr.

DOI 10.18699/BGRS/SB-2020-081

A study of genes controlling carcinogenesis in a regenerative model flatworm *Macrostomum lignano*

Kirill Ustyantsev, Valeriya Vavilova
Sector of molecular-genetic mechanisms of regeneration
ICG SB RAS
Novosibirsk, Russia
ustyantsev@bionet.nsc.ru,
valeriya-vavilova@bionet.nsc.ru

Mikhail Biryukov
Interinstitutional laboratory of molecular paleogenetics and paleogenomics
ICG SB RAS
Novosibirsk, Russia
birykov@bionet.nsc.ru

Eugene Berezikov
Sector of molecular-genetic mechanisms of regeneration
ICG SB RAS
Novosibirsk, Russia
European Research Institute for the Biology of Ageing
Groningen, The Netherlands
ORCID 0000-0002-1145-2884
eberez@bionet.nsc.ru

***Abstract* — Free-living flatworm *Macrostomum lignano* is a novel model organism that provides a genetically tractable experimental system to specifically study the interplay between regulation of regeneration and cancer. Here, we tested two chemical carcinogens (MMS and CsA) and short wavelength ultraviolet irradiation (UVC) in *M. lignano* in order to identify genes and molecular pathways underlying carcinogenic response in this flatworms. For the first time, sensitivity of *M. lignano* to hard UV was evaluated, and it was shown that the worm can easily tolerate sterilization-level doses of higher than 100 mJ/cm2. Using differential gene expression analysis based on generated RNA-Seq data, common and individual patterns of *M. lignano* transcriptional response to the induced carcinogenesis by the tested stimuli were determined. This allowed us to select promising candidate genes for functional studies using RNAi knock-down screens and for determination of their role in stem cells regulation under regeneration and homeostasis in *M. lignano*.**

Keywords* — *Macrostomum lignano, flatworms, carcinogenesis, RNA-Seq, ultraviolet

Introduction

The restriction of regeneration is thought to evolve as a measure against cancer. However, highly regenerative flatworms do not have cancer, suggesting that they evolved some other regulatory mechanism to keep their cells under control while allowing the process of regeneration. Free-living regenerative flatworm *Macrostomum lignano* is a novel model organism that provides a genetically tractable experimental system to specifically study the interplay between regulation of regeneration and cancer [1]. Here, we studied changes in transcriptional profile in *M. lignano* which occur upon treatment by both chemical and physical carcinogenic factors to identify genes which have the most impact on stem cell regulation and proliferation control in this flatworm.

Methods and Algorithms

Worms were treated separately by two chemical carcinogenic compounds for 24 h: genotoxic methylmethansulfonate (MMS, 0.5 mM) and non-genotoxic cyclosporin A (CsA, 0.2 μM), concentrations of which were determined previously [2]. For a physical carcinogenic factorworms were exposed to short wavelength ultraviolet irradiation (UVC, 254 nm, 100 mJ/cm2) and allowed to recover for 1h or 6h. Each of the treatments, as well as intact controls, were done in four biological replicates. Total RNA was isolated from the samples and used for RNA-Seq library preparation using the *Smart-3SEQ* protocol [3]. The libraries were sequenced on the Illumina NextSeq500. Raw reads were processed by *TrimGalore* and mapped to *Mlig_3_7* genome assembly using *STAR* (v.2.5.4b). Differential gene expression (DE) analysis was performed with *edgeR* [4], setting FDR threshold to 0.001.

Results

Sensitivity of *M. lignano* to UVC was studied for the first time. It was determined that the animals can handle UVC doses higher than 100 mJ/cm2 (close to the sterilization of viral particles) without noticeable phenotype and lethality. A total of 20 RNA-Seq libraries were obtained, sequenced and analysed in this study. MDS analysis in *edgeR* supported good separation and clusterization of all the libraries according to the treatments applied with an exception of the CsA-treated samples, suggesting that CsA treatment has limited effect on transcriptional changes. Summary on the DE analysis is shown in Table 1. We assumed that *M. lignano* genes responsible for carcinogenesis regulation should be upregulated upon the treatments. To further narrow the search, we focused only on upregulated DE genes, which are specific to the worm's stem cells (neoblasts) and progenitors (Table 2). Overlapping these lists resulted in the final list of 30 candidate genes, which are being further experimentally investigated.

TABLE 1. SUMMARY OF THE DE ANALYSIS. FDR < 0.001

Treatment	Number of upregulated genes	Number of downregulated genes
MMS	1633	2560
CsA	31	6
1h post UVC	2037	1063
6h post UVC	3252	2048

TABLE 2. NUMBER OF DE GENES UPREGULATED IN *M. LIGNANO* STEM CELLS

Cell type	MMS	1h post UVC	6h post UVC
Only neoblasts	108	27	34
Neoblasts and progenitors	306	101	144

Conclusion

The obtained data lay the foundation for the development of comparative genomics among *Macrostomum* species. The list of promising candidate genes for further functional studies of molecular mechanisms underlying carcinogenesis and regeneration control in *M. lignano* was obtained.

ACKNOWLEDGMENT

The work was supported by Russian Foundation for Basic Research (RFBR) grant No. 18-04-01011.

REFERENCES

[1] S. Mouton, J. Wudarski, M. Grudniewska, and E. Berezikov, "The regenerative flatworm macrostomum lignano, a model organism with high experimental potential," Int. J. Dev. Biol., vol. 62, no. 6–8, pp. 551–558, 2018.

[2] M Biryukov, I Sukhikh, V Vavilova, K Ustyantsev, and E Berezikov, "Establishing free-living flatworm Macrostomum lignano as a model to study links between regeneration and cancer," Systems Biology and Bioinformatics, pp. 09–09, Aug. 2018.

[3] J. W. Foley et al., "Gene expression profiling of single cells from archival tissue with laser-capture microdissection and Smart-3SEQ," Genome Res., vol. 29, no. 11, pp. 1816–1825, Nov. 2019.

[4] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: A Bioconductor package for differential expression analysis of digital gene expression data," Bioinformatics, vol. 26, no. 1, pp. 139–140, Nov. 2009.

DOI 10.18699/BGRS/SB-2020-082

Web-3DPredictor: a web interface for high-resolution prediction of genome architecture

Emil Valeev
ICG SB RAS, Novosibirsk, Russia
emil@bionet.nsc.ru

Polina Belokopytova
ICG SB RAS, Novosibirsk, Russia
belokopytova@bionet.nsc.ru

Veniamin Fishman
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
minja@bionet.nsc.ru

***Abstract* — Alterations of spatial contacts of chromatin may lead to developmental disorders and cancer. We have recently shown that 3-dimensional genomic contacts could be predicted in normal and rearranged genomes using epigenetic information. Here we describe an extension of this algorithm, characterized by more accurate and higher-resolution predictions, and its web implementation which allows researchers and clinical doctors to generate patient-specific predictions of genome architecture.**

Index Terms* — *Hi-C, chromatin organization, CTCF, RNA-seq, epigenetic mechanisms

Motivation and aim

Motivation

It was recently shown that chromatin contacts could be predicted using widely-available epigenetic information and machine-learning methods [1]. These predictions could be used to dissect molecular mechanisms which connect chromosomal rearrangements with human diseases, and thus could provide useful resource for clinical genetics. However, the majority of the developed predictive algorithms are not available as a stand-alone or online software; therefore, their usage in clinics is limited. Moreover, the resolution of these algorithms is often low, thus estimation of the clinical significance of changes in 3D-genome architecture is challenging.

Aim

We aimed to extend our recently developed computational tool, 3DPredictor, to predict chromatin contacts at high resolution (1kb). Moreover, we aimed to develop a web application that allows to predict genome organization in normal and rearranged genomes, for researchers and clinical doctors who are not experienced in bioinformatics.

Methods

We used ML-based algorithm detailed in [1] to train models and predict chromatin 3D organization. Training datasets which high-resolution prediction required were taken from [2]. For web service development we used the following tools:

- Front-end: PureCSS framework, PHP
- Back-end:
- Pipeline: Bash script
- Preprocessing CTCF data: gimmemotifs [3]
- Preprocessing RNA-seq data: Python3 script
- Hi-C map: Juicer Tools [4]

Results

We showed that prediction can be significantly improved by training algorithm on higher resolution data. Also, we made a web service which performs prediction of chromatin 3Dstructure for a region of interest. Trained models for *H. sapiens* and *M. musculus* at 1- and 5-kb resolutions are provided. Web3DPredictor is hosted on the ICG server[1]. Both source code and trained models could be obtained on GitHub to deploy local server elsewhere.

AKNOWLEDGEMENT

Supported by the RFBR (18-29-13021).

REFERENCES

[1] Polina S. Belokopytova, Miroslav A. Nuriddinov, Evgeniy A. Mozheiko, Daniil Fishman and Veniamin Fishman, “Quantitative prediction of enhancer–promoter interactions”, Genome Research, 2019. doi: 10.1101/gr.249367.119

[2] Nils Krietenstein, Sameer Abraham, Sergey V. Venev, Nezar Abdennur, Johan Gibcus, Tsung-Han S. Hsieh, Krishna Mohan Parsi, Liyan Yang, Rene Maehr, Leonid A. Mirny, Job Dekker, Oliver J. Rando, “Ultra-´ structural details of mammalian chromosome architecture”, 2019. doi: 10.1101/639922

[3] Niklas Bruse, Simon J. van Heeringen, “GimmeMotifs: an analysis framework for transcription factor motif analysis”, 2018. doi: 10.1101/474403

[4] Neva C. Durand, Muhammad S. Shamim, Ido Machol, Suhas S. P. Rao, Miriam H. Huntley, Eric S. Lander, and Erez Lieberman Aiden, “Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments”, Cell Systems 3(1), 2016. doi: 10.1016/j.cels.2016.07.002

[1] Available at this link: https://genedev.bionet.nsc.ru/Web 3DPredictor

DOI 10.18699/BGRS/SB-2020-083

Genes expression related to the effects of hypoxia in the marine mussel, *Mytilus galloprovincialis*

Ekaterina Vodiasova
IBSS RAS, Sevastopol, Russia
eavodiasova@gmail.com

Aleksandra Andreyeva
IBSS RAS, Sevastopol, Russia
andreevaal@gmail.com

Anastasiya Lantushenko
SSU, Sevastopol, Russia
lantushenko@mail.ru

Yakov Meger
SSU, Sevastopol, Russia
meger_yakov@mail.ru

Irina Deghtyar
SSU, Sevastopol, Russia
skuratovskaya95@mail.ru

Dmitry Afonnikov
ICG SB RAS, Novosibirsk, Russia
ada@bionet.nsc.ru

***Abstract* — Despite that mussel is an important object of mariculture and can play the role of an ecosystem bio-indicator, the study of the effect of hypoxia on these molluscs is limited only by biochemical and RT-PCR experiments for individual genes (enzymes). There is no data on changes in the expression profile of all genes based on RNA sequencing. At the same time, an understanding of physiological processes requires comprehensive studies based on the study of the expression of all genes in target tissues. At this work, the transcriptomes of the gills of *Mytilus galloprovincialis* exposed to hypoxic stress (DO 0.3 mg/l) for 24h were sequenced using Illumina technology for the first time. The raw reads assembled into 255,302 transcripts with an N50 value of 1,273 nt. A total of 65 transcripts that differed in abundance in the hypoxia-exposed mussels were identified. Their putative function was assigned using BLAST. The data obtained will allow further large-scale studies for other targeted tissues and to study all aspects of the physiological response of *M. galloprovincialis* to the emerging oxygen deficiency in the environment.**

Keywords* — *short-time hypoxia, Mytilus galloprovincialis, transcriptomic, differential expression, RNA-seq, comparative analyses

Motivation and aim

There is a lack of works based on transcriptomics data about the expression of genes in mussels hypoxia-exposed. Most of them are devoted to the study of changes in the activity of the enzymes and their concentration in various tissues [1-4]. Very few studies analyzing gene expression changes using RT-PCR methods due to the insufficient nucleotide sequences of the target genes [2, 5, 6], transcriptomics data on the influence of hypoxia is absent. Thus, the analysis of differentially expressed genes will probably allow to identify new physiological processes that are activated in the cells of target tissues at the molecular level.

Methods

Mussels, *M. galloprovincialis* (shell length 56.3+/- 1.5 mm, weight 12.7+/-2.1 g, n=16) were placed in a tank with seawater. Hypoxic conditions were conducted by bubbling of the water with nitrogen gas reaching the concentration 0.3 mg/l. Oxygen concentration was controlled on oxygen sensor (Ohaus, USA). After 24 h incubation in hypoxia, gill tissue was obtained and RNA was extracted using Zymo Research Quick RNA mini extraction kit (R1050, Zymo Research). Total RNA was isolated from gill tissues of mussel from control (4 ind.) and experimental (3 ind.) group. The libraries were synthesized with TruSeq Stranded Total RNA Library Prep (20020596, Illumina) and single-end sequencing as 150 nt reads was done on Illumina NextSeq 500 instrument.

Quality and length trimming of the reads was conducted using Fastp v. 0.20.0. The cleaned sequencing reads were then de novo assembled using Trinity v. 2.0.6 with default parameters. Sequence similarity and annotation of transcripts was performed using local BLAST v. 2.9.0+. The predicted protein regions were mapped to Gene Ontology (GO) and Pfam. GO was used to get an overview of the proportion of enriched GO terms between normoxia and hypoxia.

Results

Differentially expressed genes in gills during oxygen depletion stress were obtained based on total RNA-seq. The raw reads assembled into 255,302 transcripts with an N50 value of 1,273 nt. Changes in gene expression in control compared to experimental groups were estimated. A total of 65 transcripts that differed in abundance in the hypoxia-exposed mussels were identified. The GO analysis was performed on the differentially expressed transcripts of mussels inside the control and experimental group. Besides, a qualitative comparative analysis was carried out on 50 differentially expressed genes that were identified in gill tissues for *M. galloprovincialis* exposed to hypoxic stress (DO 2 mg/l) for 24h.

The obtained data will allow to further study in more detail all aspects of the physiological response of *M. galloprovincialis* to the emerging oxygen deficiency in the environment.

Acknowledgment

This work is funded by State Assignment (state registration number N 0828-2018-0003), the Ministry of education and Science of the Russian Federation grant No. 14.W03.31.0015, and by RSF grant 18-14-00086 and an internal grant of SSU for 2020 No. 33/06-31.

References

[1] Anestis, A., Pörtner, H. O., & Michaelidis, B. (2010). Anaerobic metabolic patterns related to stress responses in hypoxia exposed mussels Mytilus galloprovincialis. *Journal of experimental marine biology and ecology*, *394*(1-2), 123-133.

[2] Woo, S., Denis, V., Won, H., Shin, K., Lee, G., Lee, T. K., & Yum, S. (2013). Expressions of oxidative stress-related genes and antioxidant enzyme activities in Mytilus galloprovincialis (Bivalvia, Mollusca) exposed to hypoxia. *Zoological Studies*, *52*(1), 15.

[3] Franco-Martínez, L., Tvarijonaviciute, A., Mateo, S. V., Cerón, J. J., Romero, D., Oliveira, M.,... & Martínez-Subiela, S. (2019). Evaluation of C-reactive-like protein in Mytilus galloprovincialis. *Ecological Indicators*, *106*, 105537.

[4] Haider, F., Falfushynska, H. I., Timm, S., & Sokolova, I. M. (2020). Effects of hypoxia and reoxygenation on intermediary metabolite homeostasis of marine intertidal bivalves Mytilus edulis and

Crassostrea gigas. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 110657.

[5] Woo, S., Jeon, H. Y., Kim, S. R., & Yum, S. (2011). Differentially displayed genes with oxygen depletion stress and transcriptional responses in the marine mussel, Mytilus galloprovincialis. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 6(4), 348-356.

[6] Giannetto, A., Maisano, M., Cappello, T., Oliva, S., Parrino, V., Natalotto, A.,... & Fasulo, S. (2015). Hypoxia-inducible factor α and Hif-prolyl hydroxylase characterization and gene expression in short-time air-exposed Mytilus galloprovincialis. *Marine biotechnology*, *17*(6), 768-781.

DOI 10.18699/BGRS/SB-2020-084

Search of new type of spatial organization of nucleic acids in human genome

Anastasia Zamoskovtseva
ICBFM SB RAS, Novosibirsk, Russia, NSU, Novosibirsk, Russia
a.zamoskovtseva@g.nsu.ru

Marsel Kabilov
ICBFM SB RAS, Novosibirsk, Russia, kabilov@niboch.ncs.ru

Dmitrii Pyshnyi
ICBFM SB RAS, Novosibirsk, Russia, pyshnyi@niboch.ncs.ru

Alexander Lomzov
ICBFM SB RAS, Novosibirsk, Russia
lomzov@niboch.ncs.ru

***Abstract* — It is well-known that nucleic acids perform regulatory functions in various molecular biological processes. Recently, a new type of spatial organization of nucleic acids formed by two duplex fragments and called a TW-type pseudoknot (twiplex) was found. In this work we have analyzed human genome to find these complexes. We developed an algorithm for bioinformatic search on the nucleotide sequences which potentially can form twiplex. The belonging of the obtained sequences to genes, introns or exons was analyzed.**

Keywords* — *nucleic acids complexes, algorithm, human genome

Motivation and aim

Motivation

One of the functions of nucleic acids is regulation of various biological processes. In many ways, it is determined by the secondary and tertiary structure. Recently, a new type of spatial organization of nucleic acids was discovered in the ICBFM SB RAS. This complex is formed by two duplex fragments and called a TW-type pseudoknot (twiplex). It can be formed by single or two DNA or RNA chain, or in DNA / RNA intermolecular complex.

Aim

The aim of the work is to determine the possibility of implementation of twiplex structures in living systems.

Methods

We developed an algorithm for bioinformatic search on the nucleotide sequences which potentially can form twiplex. The method was applied for the human genome analysis. We consider structures with double-stranded blocks of the length from 5 to 20 base pairs. Additional selection of sequences was carried out in order to exclude sequences forming alternative secondary structures.

Results

Genome regions potentially capable for twiplexes formation have been detected. We analyzed the belonging of the obtained sequences to genes, introns or exons. Experimental studies have begun to study the physicochemical and molecular biological properties of selected nucleotide sequences.

Acknowledgment

Supported by the RFBR (20-04-00719) and by State project No. A-0309-2016-0004.

DOI 10.18699/BGRS/SB-2020-085

A study of causal relationships between human IgG N-glycosylation traits and twelve associated diseases

Olga O. Zaytseva
Genos Glycoscience Research Laboratory
Zagreb, Croatia
ORCID: 0000-0003-3960-5157

Sodbo Z. Sharapov
Laboratory of Glycogenomics
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
ORCID: 0000-0003-0279-4900

Lucija Klarić
MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine,University of Edinburgh
Edinburgh, United Kingdom
ORCID: 0000-0003-3105-8929

Gordan Lauc
Genos Glycoscience Research Laboratory
Zagreb, Croatia
ORCID: 0000-0003-1840-9560

Yakov A. Tsepilov
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia
ORCID: 0000-0002-4931-6052

***Abstract* — N-glycosylation of IgG affects ligand binding, antigen recognition and modulates immune response. IgG N-glycome is altered in many pathological states, including cancers, autoimmune and inflammatory diseases. However, the causal relationships between diseases and IgG N-glycosylation traits remain enigmatic. In the current study we implement a Mendelian Randomization approach to study causal effect of IgG N-glycan traits on 12 diseases and vice versa. We have found limited genetic evidence that increased risk of systemic lupus erythematosus might lead to increased bisection of IgG N-glycans.**

Keywords* — *N-glycosylation, immunoglobulin G, Mendelian Randomization, inflammatory diseases, autoimmune diseases, systemic lupus erythematosus

Introduction

N-glycosylation is a post-translational modification, which impacts protein folding, stability, trafficking, ability to interact with other biomolecules, etc. The most abundant N-glycosylated antibody found in the blood plasma of healthy humans is immunoglobulin G (IgG), a major component of humoral immunity. Each IgG molecule has two conservative N-glycosylation sites in the fragment crystalizable region, and up to 20% of IgG molecules also exhibit N-glycans attached to the fragment antigen-binding domain. Majority of IgG N-glycans consist of a conserved core with two N-acetylglucosamine (GlcNAc) antennae, and may contain core-fucose, bisecting GlcNAc, and antennary galactosylation and sialylation. The structure of attached N-glycans affects the affinity of IgG to its ligands, antigen-binding and modulates immune response [1, 2]. IgG N-glycosylation profile is highly heritable [3], and a number of genetic variants showed association with IgG N-glycosylation features [4]. At the same time, N-glycosylation of IgG is altered in various physiological and pathological conditions, like ageing, cancer, inflammatory and autoimmune diseases [5].

It is still unclear, whether the lack or increase of some N-glycan structures on IgG is a risk factor for some diseases, or if it is the disease progression that leads to the changes in the IgG N-glycome. In this study we applied the two-sample Mendelian Randomization (MR) approach to investigate the causal relationships between IgG N-glycosylation traits and 12 inflammatory, autoimmune, cardiovascular and neurodegenerative diseases using summary statistics from publically available and in-house genome-wide association studies (GWAS).

Materials and Methods

Participating Cohorts

Diseases were included in the analysis based on the following criteria: 1) publically available GWAS summary statistics; 2) change of IgG N-glycosylation in the disease observed in a large-scale glycomic case-control study. Summary statistics for the IgG N-glycosylation traits were obtained from the studies of 8 cohorts of European ancestry: CROATIA-Korcula, CROATIA-Vis, ORCADES, TwinsUK, EGCUT, FINRISK, VIKINGS and CROATIA-Korcula2 [4, 6]. Following meta-analyses were performed: CROATIA-Korcula, CROATIA-Vis, ORCADES, TwinsUK (N=8090, further on referred to as the 8K cohort); 8K, EGCUT and FINRISK (N = 9190, 9K cohort); EGCUT, FINRISK, VIKINGS and CROATIA-Korcula2 (N=3147, 3K cohort).

Two-sample MR Analysis

Two-sample MR was performed using "TwoSampleMR" R package [7]. The causal effect estimate was obtained as an inverse variance weighted meta-analysis of ratios of exposure effect size and outcome effect size for each of the instruments. We performed three rounds of the analysis: in the discovery round we were testing causal effects in all possible pairs of IgG N-glycosylation traits and diseases in both directions; in the following two rounds of sensitivity analysis we were trying to replicate the causal effect of systemic lupus erythematosus (SLE) on bisection of IgG N-glycans (IgG_B) observed in the discovery round.

Discovery Round

To estimate causal effects of IgG N-glycosylation traits on diseases, we selected genetic variants that were 1) significantly associated with IgG N-glycosylation in both [4,6]; 2) the association was replicated in the 9K cohort (p-value < 5e-08). As exposure we used only summary statistics for the 56 traits that had a least 2 significant instruments. As outcomes we used publically available summary statistics for the 12 diseases. To test the causal effect of diseases on IgG N-glycosylation, as instruments we used independent genetic variants with associated p-value < 5e-08 from corresponding disease summary statistics. As outcomes we used summary statistics of 86 IgG N-glycosylation traits in the 8K cohort.

Sensitivity Analysis of the SLE Effect on IgG_B

To perform sensitivity analysis for the causal effect of SLE on IgG_B, as instruments we used a refined set of 36 genetic variants associated with SLE [8], excluding outliers identified

with the MRPRESSO package for R [9]. As outcomes we have used summary statistics for IgG_B in the 8K cohort individuals in the first round of sensitivity analysis, and in the 3K cohort for the second round.

Results

We performed MR analysis in both directions, testing causal effects of IgG N-glycosylation traits on the risk of 12 diseases and *vice versa*. On the discovery stage, a signal related to the casual link of SLE on bisection of IgG N-glycans (IgG_B) survived after the correction on multiple testing. The causal effect was estimated as 0.131 standard deviation units of IgG_B per log odds units of SLE (p value = 2.24e-09). To confirm the detected casual signal we performed two rounds of sensitivity analyses. In the first round, where a refined set of genetic variants was used as instruments, the causal effect was confirmed with p-value of 1.24e-05. However, in the second round of sensitivity analysis, where the summary statistics for IgG_B from the 3K cohort were used, the causal effect of SLE on IgG_B was completely lost. The results of all MR rounds for SLE as exposure on IgG_B as outcome are shown in Table 1. In conclusion, the observed genetic support for causal relationships between IgG N-glycosylation is limited and further studies of the regulation of IgG N-glycosylation are required.

TABLE 1 – CAUSAL EFFECT OF SLE ON BISECTION OF IGG N-GLYCANS ESTIMATED BY MENDELIAN RANDOMIZATION

Analysis round	# of SNPs	SNP selection	beta	se	p-value
Discovery	33	Clumping	0.13	0.022	2.2e-09
Sensitivity 1	36	Literature	0.12	0.028	1.2e-05
Sensitivity 2	35	Literature	-0.03	0.045	0.49

ACKNOWLEDGMENT

The work of OOZ and GL was supported by the Croatian National Centre of Research Excellence in Personalized Healthcare grant (#KK.01.1.1.01.0010). The work of SZSh was supported by the Russian Science Foundation grant number 19-15-00115. The work of LK was supported by the RCUK Innovation Fellowship from the National Productivity Investment Fund (MR/R026408/1).The work of YAT was supported by the Russian Ministry of Science and Education under the 5-100 Excellence Programme by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genetics (project 0324-2019-0040-C-01/AAAA-A17-117092070032-4).

REFERENCES

[1] C. Ferrara et al., "Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcgammaRIII and antibodies lacking core fucose", Proc. Natl. Acad. Sci. U. S. A., vol 108, no. 31, pp.12669–74, 2011.

[2] F.S. van de Bovenkamp, L. Hafkenscheid, T. Rispens, Y. Rombouts, "The Emerging Importance of IgG Fab Glycosylation in Immunity", J. Immunol., vol 196, no.4, pp. 1435–1441, 2016.

[3] C. Menni et al., "Glycosylation of Immunoglobulin G: Role of Genetic and Epigenetic Influences", PLoS One, vol. 8, no. e82558, 2013.

[4] L. Klarić et al., "Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases", Sci. Adv., vol. 6, no. 8, p. eaax0301, 2020.

[5] I. Gudelj, G. Lauc, and M. Pezer, "Immunoglobulin G glycosylation in aging and diseases", Cell. Immunol., vol. 333, pp. 65–79, 2018.

[6] G. Lauc et al., "Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers", PLoS Genet., vol. 9, no. 1, p. e1003225, 2013.

[7] G. Hemani et al., "The MR-base platform supports systematic causal inference across the human phenome", Elife, vol. 7, p. e34408, 2018.

[8] J. Bentham et al., "Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus", Nat. Genet., vol. 47, no. 12, pp. 1457–1464, 2015.

[9] M. Verbanck, C.-Y. Chen, B. Neale, R. Do, "Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases", Nat. Genet., vol. 50, no. 5, pp. 693–698, 2018.

DOI 10.18699/BGRS/SB-2020-086

Automatic annotation of operons responsible for O-antigen synthesis

Danil Zilov
Laboratory of Applied Genomics, SCAMT, ITMO University, St. Petersburg, Russia
zilov@scamt-itmo.ru

Polina Chesnokova
Laboratory of Applied Genomics, SCAMT, ITMO University, St. Petersburg, Russia
chesnokova@scamt-itmo.ru

Alexey Komissarov
Laboratory of Applied Genomics, SCAMT, ITMO University, St. Petersburg, Russia
komissarov@scamt-itmo.ru

***Abstract* — Unfortunately, the most ready-made tools for serotype determination are limited to genome *E. Coli* only. However, there is a problem due to the low accuracy of O-antigen cluster identification in the genome. Pipeline, which was created due the work, solves this problem and can be used to find and annotate a cluster of genes in genome of the most gram-negative bacteria.**

Keywords* — *O-antigen, genome assembly, annotation of genome, pipeline

Motivation and aim

Motivation

O-antigen is a polysaccharide, the major component of the outer bacterial membrane. O-antigen biosynthesis is encoded by genes, which are often clustered. Despite the simplicity of the cluster structure, its identification in microbial DNA is still a problem.

It is well known, that in the structure of O-antigen there are specific sites of binding to antibodies of the host organism, the interaction with which leads to the emergence of an immune response to infection. We are interested in the way these sites are organized. Unfortunately, the most ready-made tools for serotype determination are limited to genome *E. Coli* only. Therefore, we started to build our own tool.

Aim

Our goal is to create a pipeline that will automate the search, assembly, filtration and annotation of a cluster of genes responsible for the synthesis of O-antigen, for most gram-negative bacteria.

Methods

We used Python programming language, Snakemake workflow management system, BLAST and NCBI databases Conda package manager and few custom scripts to create pipeline.

To run the pipeline raw sequence reads or assembled genome is needed.

The main stages of data processing:

- Assessing the quality of input data, removing adapters and optical duplicates;
- Genome assembly using SPAdes and Unicycler;
- Checking for contamination and plasmids;
- Removal of contaminants and plasmids, if any;
- Annotation by Prokka and eggNOG;
- Functional annotation, extraction of genes involved in O-antigen biosynthesis;
- Abstract of operons containing genes involved in O-antigen biosynthesis taking into account possible assembly errors and contigs breaks.

We are currently annotating O-antigens of all gram-negative bacteria available in the NCBI database. The resulting information will be used to improve the existing Pipeline.

Results

The created pipeline significantly saves time for serotyping gram-negative bacteria. Additionally, it can be used in the detection of outer membrane components related to both virulence of microorganisms and symbiotic interaction between bacteria and plants.

Symposium
Systems computational biology: analysis, mathematical modeling and information technologies

DOI 10.18699/BGRS/SB-2020-087

Systems biology analysis of metabolism, signaling and gene expression regulation in human skeletal muscle

Ilya R. Akberdin
BIOSOFT.RU, LLC
FRCenter Institute of Cytology and Genetics SB RAS
Novosibirsk State University
Novosibirsk, Russia
akberdinir@gmail.com

Alexander Yu. Vertyshev
CJSC "Sites-Tsentr"
Moscow, Russia
avertyshev@mail.ru

Ilya N. Kiselev
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
axec@developmentontheedge.com

Pavel A. Makhnovskii
Institute of Biomedical Problems of the RAS, Moscow, Russia
maxpauel@gmail.com

Fedor A. Kolpakov
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
fkolpakov@gmail.com

Sergey S. Pintus
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
sspintus@developmentontheedge.com

Daniil V. Popov
Institute of Biomedical Problems of the RAS, Moscow, Russia
danil-popov@yandex.ru

Abstract — **Physical exercise training induces profound molecular genetic adaptations in skeletal muscle cells. The adaptation is ensured by instant/transient activation of signaling pathways in the muscle cells eventually resulting in an alteration of both global metabolic network fluxes and expression of various genes. Despite the experimentally based efforts to disentangle the complexity of the muscle adaptation network caused by multiple interactions and intersections on signaling, metabolic and genetic regulatory levels, the quantitative and mechanistic contribution of each component of the signaling cascades on downstream genetic regulation processes has not been fully elucidated. Data-driven mathematical models provide a rigorous way to analyze and understand such intricate biological systems. Herein a novel integrated mathematical model linking intracellular metabolism, Ca^{2+}-dependent signaling pathway and downstream transcription regulation of early and late response genes in human skeletal muscle during and after acute exercise developed in BioUML platform has been presented.**

Keywords — mathematical model, skeletal muscle, physical exercise, Ca^{2+}-dependent signaling pathway, RNA sequencing, regulation of expression, BioUML

Introduction

Skeletal muscle comprises about 40 % of total body mass in adult lean humans and plays a crucial role in the control of whole-body metabolism and exercise tolerance. Regular low-intensity exercise (aerobic or endurance training) strongly induces adaptive changes in skeletal muscle which are increased vascular and mitochondrial density, oxidative capacity and improve fat and carbohydrate metabolism. These adaptations lead to an enhancement of muscle endurance performance and reduce the risk associated with the morbidity and premature mortality of chronic cardiovascular and metabolic diseases [1; 2].

It is worth to note, although advancement in the development of high-throughput experimental techniques and generation of diverse omics data for human skeletal muscle during endurance exercise enabled to unveil key participants of the cellular response and adaptation [3-7], the system understanding of signaling-metabolic pathways relationships with downstream genetic regulation in exercising skeletal muscle is still elusive. As a complementary theoretical counterpart to the experimental investigation of molecular mechanisms underlying the skeletal muscle adaptation to the endurance training, a detailed mechanistic mathematical model provides a powerful *in silico* tool to quantitatively investigate signal transduction pathways and corresponding molecular mechanisms orchestrating gene expression dynamics during an exercise [8-11].

Methwods

Model reconstruction

We have previously developed a multi-compartmental mathematical model describing the dynamics of intracellular species concentrations and fluxes in human muscle at rest and intracellular metabolic rearrangements in exercising skeletal muscles during an aerobic exercise on a cycle ergometer [10]. As an initial model, we have used a complex model of energy metabolism in the human skeletal muscle developed by Li and coauthors [8]. We have proposed a modular representation of the complex model using BioUML platform [12]. To reconstruct the coupled signaling-gene expression regulation networks in SBGN standard [13] and build corresponding detailed mechanistic models we employed BioUML platform too [12]. The processes are distributed in two cellular compartments: signaling transduction pathway and translation, post-translation modification of the protein encoded by early response gene are represented in the cytosol, while the genetic regulation processes are located in the nucleus. The modular models of Ca^{2+}-dependent signaling pathway and AMPK isoforms activities induced by the muscle construction are represented by ordinary differential equations (ODEs).

Model validation and parameter fitting

For our model, values of the almost all kinetic parameters were extracted from previous experimental data and mathematical models describing different aspects of the metabolic network, AMPK- and Ca^{2+}-dependent signaling pathway [7-8; 14-19]. We adjusted the values of parameters that had not been reported in the published data to achieve consistency with experimental quantitative data on transcriptional activity PGC-1α mRNA and a time-scale required to reach the maximal level of PGC-1α mRNA in skeletal muscle after the endurance exercise as well as a time-scale to recover the basal level of transcriptional

activity at the rest [7]. To simulate the muscle adaptation to endurance training we used experimentally verified shift of Ca^{2+} concentration which corresponds five-fold increase of the parameter value during the transition from rest to the exercise [11]. The numerical simulations of the dynamic model developed as a system of ordinary differential equations have been conducted using VODE method implemented in BioUML platform [12]. Sensitivity analysis of the model parameters and initial values of the model variables were performed using default settings in the corresponding BioUML module.

RNAseq and proteomics data analysis

Source of RNAseq data and procedure of data selection has been presented in our previous publication [7]. Original data on diverse content of proteins in slow and fast fibres were extracted from the Supplementary Material of the paper [14]. If a protein was not presented in this data set its concentration was calculated in the basis of RNAseq data due to the fact that protein concentration correlates with RPKM and FPKM values. FPKM values from the RNAseq data were used as a basis and compared to GTEx, Average RPKM values data for corresponding proteins in the skeletal muscles extracted from The Human Protein Atlas (www.proteinatlas.org). As a control of the calculation's correctness, a protein content in the skeletal muscle cell estimated using RNAseq data was compared to published proteomics data [14]. Results of calculations for such low-presented proteins as subunits of AMPK, Calmodulin, CaMKII have demonstrated similar values of the concentrations estimated by two approaches. According to this for other low-presented proteins (i.e. transcription factors) concentrations estimated on the basis of RNAseq data were used if they were absent in proteomics study.

Results

The computational model describing a gene expression activation elicited by upstream signaling network during the transition from rest to exercise in human skeletal muscle has been developed based on modular modeling approach [12]. In turn, exercise-induced changes in gene expression levels promote the metabolic adaptation of the skeletal muscle cells by means of an activation mechanism which enhances energy metabolism via transport and reaction fluxes. In order to the integrated model represents actual changes in gene expression in exercised human skeletal muscle in more details, we replaced the general work rate parameter on the concentration of Ca^{2+}-Calmodulin complexes and incorporated PGC-1α–mediated transcription regulation of genes playing an important role in adaptation to regular exercise. *In silico* simulations demonstrate a high potential of the developed model to predict a proper cellular response to exercise based on the comparison between experimental measurements and simulated values of the model variables on all hierarchical levels (metabolic, signaling and gene expression regulation).

Acknowledgment

The study has been financially supported by RFBR grants (N_O. 17-00-00308 (K) and N_O. 17-00-00296).

References

[1] B.K. Pedersen and M.A. Febbraio, “Muscles, exercise and obesity: skeletal muscle as a secretory organ,” Nature Reviews Endocrinology, vol. 8, № 8, pp. 457–465, 2012.

[2] J.A. Hawley, M. Hargreaves, M.J. Joyne and J.R. Zierath, “Integrative biology of exercise,” Cell, vol. 159, № 4, pp. 738-749, 2014.

[3] O. Neubauer et al., “Time course-dependent changes in the transcriptome of human skeletal muscle during recovery from endurance exercise: from inflammation to adaptive remodeling,” Journal of Applied Physiology, vol. 116, № 3, pp. 274-287, 2013.

[4] K. Vissing and P. Schjerling, “Simplified data access on human skeletal muscle transcriptome responses to differentiated exercise,” Scientific data, vol. 1, pp. 140041, 2014.

[5] D.V. Popov, P.A. Makhnovskii, N.S. Kurochkina, E.A. Lysenko, T.F. Vepkhvadze and O.L. Vinogradova, “Intensity-dependent gene expression after aerobic exercise in endurance-trained skeletal muscle,” Biology of sport, vol. 35, № 3, pp. 277, 2018.

[6] J.M. Dickinson et al., “Transcriptome response of human skeletal muscle to divergent exercise stimuli.,” Journal of Applied Physiology, vol. 124, № 6, pp. 1529-1540, 2018.

[7] D.V. Popov, et al., “Contractile activity-specific transcriptome response to acute endurance exercise and training in human skeletal muscle,” American Journal of Physiology-Endocrinology and Metabolism, vol. 316, № 4, pp. E605-E614, 2019.

[8] Y. Li, R.K. Dash, J. Kim, G.M. Saidel, M.E. Cabrera, “Role of NADH/NAD+ transport activity and glycogen store on skeletal muscle energy metabolism during exercise: *in silico* studies,” American Journal of Physiology-Cell Physiology, vol. 296, № 1, pp. 25–46, 2009.

[9] Akberdin I.R., Kazantsev F.V., Ermak T.V., Timonov V.S., Khlebodarova T.M., Likhoshvai V.A. In Silico Cell: Challenges and Perspectives// Mathematical Biology and Bioinformatics. 2013. V. 8. № 1. P. 295–315.

[10] Kiselev I.N., Akberdin I.R., Vertyshev A.Y., Popov D.V. and Kolpakov F.A. A modular visual model of energy metabolism in human skeletal muscle// Mathematical Biology and Bioinformatics. 2019. V. 14. № 2. P. 373-392.

[11] I.R. Akberdin, A.Y. Vertyshev, S.S. Pintus, D.V. Popov and F.A. Kolpakov. A mathematical model linking Ca^{2+}-dependent signaling pathway and gene expression regulation in human skeletal muscle,” Mathematical Biology and Bioinformatics, vol. 14, № 2, pp. 373-392, 2020.

[12] F.A. Kolpakov et al., “BioUML: an integrated environment for systems biology and collaborative analysis of biomedical data,” Nucleic Acids Research, vol. 47, № W1, pp. W225–W233, 2019.

[13] N. Le Novere et al., “The systems biology graphical notation,” Nature Biotechnology, vol. 27, № 8, pp. 735, 2009.

[14] M. Murgia et al., “Single muscle fiber proteomics reveals fiber-type-specific features of human muscle aging,” Cell Reports, vol. 19, № 11, pp. 2396–2409, 2017.

[15] C.C. Carroll, J.A. Carrithers, T.A. Trappe, “Contractile protein concentrations in human single muscle fibers,” Journal of Muscle Research and Cell Motility, vol. 25, № 1, pp. 55–59, 2004.

[16] M. Wilhelm et al., “Mass-spectrometry-based draft of the human proteome,” Nature, vol. 509, № 7502, pp. 582, 2014.

[17] F. Edfors et al., ”Gene-specific correlation of RNA and protein levels in human cells and tissues,” Molecular systems biology, vol. 12, № 10, 2016.

[18] T.E. Jensen, J.F.P. Wojtaszewski and E.A. Richter, “AMP-activated protein kinase in contraction regulation of skeletal muscle metabolism: necessary and/or sufficient?,” Acta physiologica, vol. 196, № 1, pp.155-174, 2009.

[19] R. Kjøbsted et al., “AMPK in skeletal muscle function and metabolism”, The FASEB journal, vol. 32, №4, pp.1741-1777, 2018.

DOI 10.18699/BGRS/SB-2020-088

Phase portraits of gene networks models

Natalia Ayupova
Lab. of inverse problems of mathematical physics. Sobolev Institute of Mathematics SB RAS
Novosibirsk, Russia
ayupova@math.nsc.ru

Vladimir Golubyatnikov
Lab. of inverse problems of mathematical physics. Sobolev Institute of Mathematics SB RAS
Novosibirsk, Russia
vpgolubyatn@gmail.com

Vyacheslav Gradov
Mechanical and mathematical department Novosibirsk State University Novosibirsk, Russia
v.gradov@g.nsu.ru

Liliya Minushkina
Mechanical and mathematical department Novosibirsk State University Novosibirsk, Russia
l.minushkina@g.nsu.ru

***Abstract* — We investigate geometric and combinatorial structures of phase portraits of dynamical systems considered as models of functioning of some circular gene networks. Invariant 2-dimensional piecewise linear surfaces constructed here allow to give full description of phase portraits of these models and to control their trajectories.**

***Keywords* — gene networks models, dynamical systems, phase portraits, cycles[1]**

Introduction

We study piecewise linear dynamical systems of the type

$$dx_1/dt=f_1(x_n) - k_1x_1;\ dx_j/dt=f_j(x_{j-1}) - k_jx_j;\ (1)$$

as models of circular gene networks. Here j=1,2,…n; f_1, f_j are monotonic step functions either increasing, or decreasing, this corresponds to positive, respectively, negative feedbacks in the gene network. Some particular cases of the systems (1) are considered in [1, 2]. Let α_1, α_j be discontinuity points of the functions f_1, f_j and F_j be their maximal values. We denote by $L_j(x)$ and $\Gamma_j(x)$, decreasing and increasing step functions, respectively. Our main task in these studies of phase portraits of such systems is detection of their periodic trajectories (cycles), construction of invariant surfaces and domains etc. in order to give qualitative description of trajectories of these systems. We assume that $F_j>k_j$ for all j, otherwise the system (1) does not have cycles, as it was shown in [3]. We say that the system (1) has the Elowitz-Leibler type ([1]) if n=6, and $f_{2i-1} = L_{2i-1}$, $f_{2i} = \Gamma_{2i}$ for i = 1,2,3. Similarly, we say that the system (1) has the Glass-Pasternack type (see [2]) if $f_1= L_1$, and $f_j= \Gamma_j$ for j= 1,2,…,n.

Discretization of Phase Portraits

Let $Q_n=[0,F_1]\times\dots\times[0,F_n]$. As in [3, 4], one can show that this is positively invariant domain of the system (1), and the point $A_n=(\alpha_1, \alpha_2 ,\dots \alpha_n)$ is contained in Q_n. The hyperplanes $x_j= \alpha_j$ subdivide Q_n to 2^n smaller parallelepipeds (blocks) which we enumerate by binary multi-indices $E=(\varepsilon_1,\varepsilon_2,\dots,\varepsilon_n)$ as follows: $\varepsilon_j=0$ if $x_j\leq \alpha_j$ for all points of this block; otherwise $\varepsilon_j=1$. In each of these blocks trajectories of the system are described explicitly in a very simple way. We say that the block E has the valence ν if trajectories of its points leave it through ν faces of this block.

Results

Glass-Pasternack Systems

It was shown in [2] that the system (1) of Glass-Pasternack type has a cycle C_1 which passes through 1-valent blocks. For n=4, we construct now an invariant piecewise linear surface S_1 in the complement of union of interiors of these 1-valent blocks such that C_1 and S_1 have nontrivial link in Q_4. The surface S_1 has exactly with one vertex A_4 and eight faces. Trajectories of all points of the surface S_1 tend to A_4 as t $\to\infty$.

Elowitz-Leibler Systems

We have shown in [4] that 6-dimensional system (1) has a cycle C_2 which passes through 1-valent blocks in Q_6. Now, we construct an invariant piecewise linear surface S_2 in the union of 5-valent blocks in this invariant domain. The surface S_2 has exactly one vertex A_6. Trajectories of all points of the surface S_2 tend to A_6 as $t\to\infty$.

Gene Network with Negative Feedbacks

Consider 5-dimensional system (1) where $f_j = L$, k_j=1 for all j=1,2,…5. Let F be maximal value of the function L. This system is symmetric with respect to cyclic permutation of the variables. We show that if F>1, and F^2>5(F – 1) then this system has two cycles. One of them passes through the union of 1-valent blocks, the second cycle is contained in the union of 3-valent blocks. Both cycles generate invariant surfaces in phase portrait of the system.

Conclusions

Some of these invariant surfaces compose boundaries of attraction basins of stable cycles in phase portraits of dynamical systems described above

Acknowledgment

Supported by RFBR, grant 18-01-00057.

References

[1] M. B. Elowitz and S. Leibler, “A synthetic oscillatory network of transcriptional regulators,” Nature, vol. 403, pp. 335–338, 2000.

[2] L. Glass and J. S. Pasternack “Stable oscillations in mathematical models of biological control system,” J. Theor. Biol., vol. 74, pp. 207–223, 1978.

[3] V. P. Golubyatnikov and V. V. Ivanov, “Cycles in odd-dimensional models of circular gene networks,” Journal of Applied and Industrial Mathematics, vol. 12, pp. 648–657, 2018

[4] V. P. Golubyatnikov and L. S. Minushkina, “Monotonicity of the Poincaré mapping in some models of circular gene networks,” Journ. of Applied and Industrial Mathematics, vol. 13, pp. 472–479, 2019.

DOI 10.18699/BGRS/SB-2020-089

Cyclin/Cdk-centered regulations are pivotal for autonomous cell cycle oscillator designs in yeast

Matteo Barberis
Systems Biology, School of Biosciences and Medicine
University of Surrey
Guildford, Surrey, United Kingdom
m.barberis@surrey.ac.uk
matteo@barberislab.com

***Abstract*—Some biological networks exhibit oscillations of their components, to convert stimuli to time-dependent responses. The eukaryotic cell cycle is such a case, being governed by waves of cyclin-dependent kinase (cyclin/Cdk) activities that rise and fall with specific timing and guarantee its timely occurrence. Disruption of cyclin/Cdk oscillations could result in dysfunction through reduced cell divisions. Therefore, it is of interest to capture network designs that exhibit robust oscillations. Here I show that a minimal cell cycle network in budding yeast is able to autonomously oscillate, and that cyclin/Cdk-centered regulations sustain their oscillations. Moreover, I show that a novel and robust molecular design incorporating these regulations synchronizes Clb waves through transcription-mediated signaling.**

***Keywords*—network motifs, network dynamics, cell cycle, autonomous oscillations, timing, waves of cyclins, budding yeast**

Introduction

The eukaryotic cell cycle is robustly designed, with molecules interacting and organized within definite network topologies. A transcriptional oscillator interlocks with waves of cyclin-dependent kinases (cyclin/Cdk) to guarantee execution of a timely cell cycle progression [1]. Although details about transcription of cyclins, the regulatory subunits of these kinases, are available [2], a lack of understanding exists about network motifs responsible for the precise timing of cyclin/Cdk oscillations. Here we investigate the robustness of molecular designs interlocking the transcriptional and cyclin/Cdk1 oscillators in budding yeast. We have recently identified a transcriptional cascade that regulates the relative timing of waves of mitotic (Clb) cyclin expression. It involves the Forkhead (Fkh) transcription factors (TF), which synchronize Clb appearance [3]. Here we aim to unravel the network motif(s) responsible for timely Clb/Cdk1 oscillations that interlock Clb waves through Fkh-mediated signaling.

Methods and Algorithms

An integrated computational and experimental framework is presented. A kinetic, ODE model of a minimal Clb/Cdk1 network, capturing the sequential activation and inactivation of waves of Clb/Cdk1 complexes and their inhibitor Sic1, is simulated under a quasi-steady state assumption, which implies a fast equilibrium of Clb/Cdk1-Sic1 complex formation. The resulting model has a reduced number of parameters, allowing it to be manageable for comprehensive parameter scans. We apply the System Design Space (SDS) methodology [4,5] that scans through the parameter space to detect limit cycles. We then analyze the ability of our minimal cell cycle model to generate sustained oscillations in the form of limit cycles, on which sensitivity analysis is conducted. Subsequently, stochastic and deterministic robustness analyses of the network are performed, by testing 1024 possible (combinations of activatory and inhibitory) network motifs for their ability to generate mitotic cyclin (Clb) oscillations. Finally, the model is fitted to *in vivo* quantitative, time course data of Clb dynamics (Figure 1a). Robustness analyses are performed by testing the 1024 network motifs for their ability to fit Clb oscillations.

Results

The Clb/Cdk1 network is robust with respect to the formation of Clb waves. Autonomous oscillations, also referred to as limit cycles, are achieved by definite molecular designs. Specifically, a positive feedback loop (PFL) by Clb3/Cdk1 on *CLB3* synthesis (Clb3 PFL) is found to improve the ability of the minimal cell cycle model to produce sustained Clb/Cdk1 oscillations. In addition, a positive feedback loop by Clb2/Cdk1 on *CLB2* synthesis (Clb2 PFL) is found to take over this key role when the model takes into account the inhibition of G1/S cyclins by Clb2/Cdk1. Furthermore, two regulatory activations, i.e. Clb5 → Clb3 and Clb3 → Clb2, forming a transcription factor-mediated linear *CLB* cascade that we have recently discovered [3], are more frequently dominant in Clb/Cdk1 oscillations as compared to the feed-forward Clb5 → Clb2 regulation described earlier [6] (Figure 1b). Overall, a novel design is unraveled, that timely synchronize Clb oscillations. This design uniquely describes a molecular timer (TF) that relies on separate inputs (multiple Clb/Cdk kinase complexes) converging on a common target (TF itself).

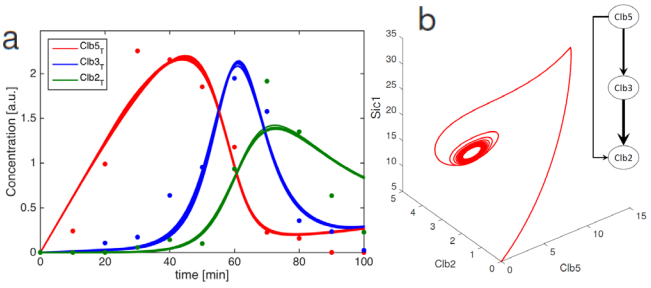


Figure 1. a) Fitting of the Clb/Cdk1 model to Clb time course data. b) Network motifs resulting in a limit cycle.

Within the design, a progressive TF activation may be realized by the sequential activation of multiple Clb/Cdk1 complexes. The ODE model predicts a definite Fkh activation pattern underlying the design motif. Experimental validation supports computational analyses, with the Clb/Cdk1-Fkh axis being pivotal for timely transcriptional dynamics. Specifically, dedicated kinase assays reveal that a progressive increase in the extent of TF phosphorylation, mediated by Clb/Cdk1 complexes, may be responsible for a timely occurrence of Clb waves.

CONCLUSIONS

Altogether, our integrative approach pinpoints how robustness of cell cycle control is realized in budding yeast, by revealing a novel and conserved principle of design that ensures a timely interlock of transcriptional and cyclin/Cdk oscillations. Given the evolutionary conservation of the cell cycle network across eukaryotes, the cyclin/Cdk network can be used as a core building block of multi-scale models that integrate regulatory modules to address cellular physiology.

ACKNOWLEDGMENT

This work was supported by the Systems Biology Grant of the University of Surrey to M.B., and by the SILS Starting Grant of the University of Amsterdam to M.B..

REFERENCES

[1] L. A. Simmons Kovacs, M. B. Mayhew, D. A. Orlando, Y. Jin , Q. Li, C. Huang, S. I. Reed, S. Mukherjee, and S. B. Haase, "Cyclin-dependent kinases are regulators and effectors of oscillations driven by a transcription factor," Mol. Cell, vol. 45, pp. 669–679, March 2012.

[2] S. B. Haase, and C. Wittenberg, "Topology and control of the cell-cycle-regulated transcriptional circuitry," Genetics, vol. 196, pp. 65–90, January 2014.

[3] C. Linke, A. Chasapi, A González-Novo, I. Al Sawad, S. Tognetti, E. Klipp, M. Loog, S. Krobitsch, F. Posas, I. Xenarios, and M. Barberis, "A Clb/Cdk1-mediated regulation of Fkh2 synchronizes *CLB* expression in the budding yeast cell cycle," NPJ Syst. Biol. Appl., vol. 3, pp. 7, March 2017.

[4] M. A. Savageau, P. M. Coelho, R. A. Fasani, D. A., Tolla, and A. Salvador, "Phenotypes and tolerances in the design space of biochemical systems," Proc. Natl. Acad. Sci. U S A, vol. 106, pp. 6435–6440, April 2009.

[5] J. G. Lomnitz, and M. A. Savageau, "Elucidating the genotype–phenotype map by automatic enumeration and analysis of the phenotypic repertoire," NPJ Syst. Biol. Appl., vol. 1, pp. 15003, September 2015.

[6] A. Pic-Taylor, Z. Darieva, B. A. Morgan, and A. D. Sharrocks, "Regulation of cell cycle-specific gene expression through cyclin-dependent kinase-mediated phosphorylation of the forkhead transcription factor Fkh2p," Mol. Cell. Biol., vol. 24, pp. 10036–10046, November 2004.

DOI 10.18699/BGRS/SB-2020-090

A Model of one central regulatory circuit

Tatyana Bukharina
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
bukharina@bionet.nsc.ru

Andrey Akinshin
Sobolev Institute of Mathematics SB RAS
Novosibirsk, Russia
Andrey.akinshin@gmail.com

Vladimir Golubyatnikov
Sobolev Institute of Mathematics SB RAS
Novosibirsk, Russia
vpgolubyatn@gmail.com

Dagmara Furman
Institute of Cytology and Genetics SB RAS
Novosibirsk State University
Novosibirsk, Russia
furman@bionet.nsc.ru

***Abstract* — Macrochaetes are sensor organs of the drosophila with a function of mechanoreceptors. An adult mechanoreceptor comprises four specialized cells. All these cells originate from the sensory organ precursor (SOP) cell. A characteristic feature of the SOP cell is the highest content of the proneural proteins ASC as compared with the surrounding cells. The accumulation of these proteins and maintenance of their amount in the SOP cell at a necessary level is provided by the gene network with the achaete-scute gene complex (*AS-C*) as its key component. The central regulatory circuit (CRC) controls the activity of this complex. The CRC activity comprises two phases differing in the time when they act. A specific feature of the second phase is the presence of PHYL protein, involved in degradation of proneural proteins. We propose a mathematical model for the CRC functioning as a regulator of the content of ASC proteins in the sensory organ precursor cell.[2]**

Keywords* — *dynamical systems, central regulatory circuit, gene networks models, achaete-scute complex, mutations

Introduction

The drosophila mechanoreceptors are the result of determined conversion of the ectodermal cells of imaginal discs into progenitor neural cells [1]. The morphogenesis of macrochaetes is a multistage process, where separation of the sensory organ precursor (SOP) cell from ectodermal cells in the wing imaginal disc is the basic event. Production and utilization of the proneural ASC proteins in SOP cell is controlled by a set of genes and the proteins they encode, which are functionally united into CRC. These genes are linked by positive and negative feedbacks mediated via the proteins and protein complexes, activating or repressing transcription of their target genes. In addition to *AS-C*, the CRC contains hairy, *senseless (sens)*, *charlatan (chn)*, *scratch (scrt)*, and *phyllopod (phyl)* genes. The general scheme of this circuit is shown on the Figure 1 [2]. We propose a mathematical model for the CRC functioning as a regulator of the content of proneural ASC proteins in the sensory organ precursor cell. –

Mathematical Model

Following the general principles of construction of gene networks models [2-4], we consider the dynamical system for non-negative variables $x_i=x_j(t)$, here and below i=1,2,…,6:

It is well known from experimental data that the PHYL protein appears in the cell in 10-12 hours later than ASC proteins [5, 6]. Thus the CRC functioning can be described by a system of delayed argument $t–\tau$ appears in the first and in the last equations on the appropriate places differential equations which is composed on the basis of the scheme depicted on the Figure 1 following general methodology of construction of such models.

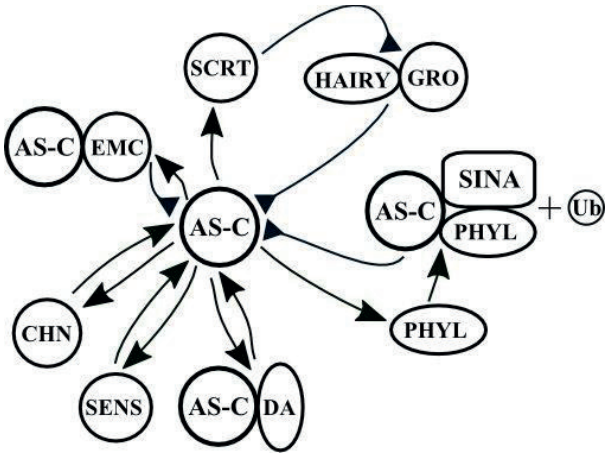


Fig. 1. Central regulatory circuit of the gene system controlling drosophila mechanoreceptors morphogenesis. Arrows denote activation of processes (positive feedbacks), stub lines, repression (negative feedbacks).

$$dx_1/dt=k_1F_1(x_1,x_2,x_3,x_5) –m_1(x_6(t–\tau))x_1 ;$$

$$dx_2/dt=k_2F_2(x_4) –m_2x_2;$$

$$dx_j/dt=k_jF_j(x_1) –m_jx_j;\ j=3,4,5;$$

$$dx_6/dt =k_6F_6(x_1,\ t–\tau) –m_6x_6.$$

CRC proteins content are denoted as follows: Achaete-Scute (ASC), $x_1(t)$; Hairy, $x_2(t)$; Senseless (SENS), $x_3(t)$; Scratch (SCRT), $x_4(t)$; Charlatan (CHN), $x_5(t)$; Phyllopod (PHYL), $x_6(t)$. Content of Ubiquitin (Ub), Seven-in-absentia (SINA), Daughterless (DA), Exstramacrochaete (EMC), and Groucho (Gro) proteins almost do not vary during the considered process, thus we assume that they are constant.

Here, the non-negative coefficients k_i reflect status of CRC genes: k_i=1 when there is no mutation in the i-th gene and synthesis rate of corresponding protein is 100%; k_i=0 when the i-th gene is mutated and there is no synthesis of corresponding protein at all. Positive functions F_i denote velocities of synthesis of CRC proteins. The positive coefficients m_i characterize velocities of CRC proteins degradation. Positive summand k_2F_2 in the second equation of the system describes negative feedback between SCRT and HAIRY (Figure 1).

Supported by RFBR, grant 18-01-00057.

Software

In order to fulfill numerical experiments with the model described above, we have elaborated specialized software on the base of the language R see https://www.r-project.org and the package Shiny (http://shiny.rstudio.com).

This software is realized as a client –server application:

All calculations are done on the cloud server, and the results are presented at https://andreyakinshin.shinyapps.io/as-c_with_delays

Numerical solutions of differential equations with delayed arguments use the package deSolve https://andreyakinshin.shinyapps.io/as-c_with_delays which was already applied to numerical modeling of other gene networks [2].

Numerical Results of Modeling

Results of numerical experiments with the system for different parameter values of this system are presented at:

https://gist.github.com/AndreyAkinshin/fb2cae36b89838ad5dcac706b1d232a1

https://gist.github.com/AndreyAkinshin/a7450a9d234da67fee8e67ab6f4364f8

https://gist.github.com/AndreyAkinshin/3f2c3c49a954a2c938c5355c1263f554

Comparison of these results (see links above) shows that:

1. For the coefficient k_1=1 the graph for content of ASC proteins is adequate to known experimental data. According to them ASC proteins content increases until the appearance of the PHYL protein in the SOP cell and then quickly decreases.

2. The shapes of the graphs for concentration of ASC proteins in the case k_1=1, k_1=0.66, and k_1=0.645 does not change considerably. However, for k_1=0.6 the function $x_1(t)$ tends to zero. This corresponds to inhibition of mechanoreceptor's formation.

3. For the values of the coefficient k_1 listed above, variations of the graphs of concentrations $x_i(t)$, $i\neq2$, of other components of CRC have similar characters.

4. The graph of the function $x_2(t)$, i.e., content of HAIRY, in general repeats the corresponding curve for k_2=1, with slightly higher numerical values. This result is biologically reasonable, since *AS-C* mutants have reduced levels of the SCRT protein which suppresses of *hair*y expression.

Conclusions

The main goal of this investigation was to create an improved model of CRC functioning. The simulation results are in good agreement with known literature data. So, the model developed reflects the pattern of CRC functioning correctly and allows to simulate not only the changes in the ASC content in the case of mutational changes in the corresponding gene, but also evaluate the contribution of these changes in content of other CRC proteins.

REFERENCES

[1] E.C. Lai, V. Orgogozo, "A hidden program in Drosophila peripheral neurogenesis revealed: fundamental principles underlying sensory organ diversity," Dev. Biol., vol. 269, pp. 1-17, 2004.

[2] T. Bukharina, D. Furman, V. Golubyatnikov, "A model study of the morphogenesis of *D.melanogaster* mechanorecetpors," JBCB, vol. 13, pp. 1540006-1–1540006-15, 2015.

[3] G.V. Demidenko, "Systems of differential equations of higher dimensions and delay equations," Siberian mathematical journal, vol, 53:6, pp. 1021 – 1028, 2012.

[4] G.V. Demidenko, N.A. Kolchanov., V.A. Likhoshvai., Yu.G.Matushkin, S.I. Fadeev, "Mathematical simulation of regulatory circuits of gene networks," Comput. Math. Math. Phys., vol. 44, pp. 2166 –2183, 2014.

[5] P.J. Chang, Y.L. Hsiao, A.C. Tien., Y.C, Li, H.Pi, "Negative-feedback regulation of proneural proteins controls the timing of neural precursor division," Development, vol. 135, plp. 3021–3030., 2008.

[6] H. Pi, S.K. Huang, C.Y. Tang, Y.H. Sun, C.T. Chien, "Phyllopod is a target gene of proneural proteins in Drosophila external sensory organ development," Proc Natl Acad Sci USA, vol.101, pp. 8378–8383, 2004.

DOI 10.18699/BGRS/SB-2020-091

Recognition of flavivirus species on the base of coding genome sequences

Maria Chaley
Institute of Mathematical Problems of Biology RAS –
Branch of Keldysh Institute of Applied Mathematics RAS
Pushchino, Russia
chaley@phystech.edu

Vladimir Kutyrkin
Department of Comp. Mathematics and Math. Physics
Moscow State Technical University n.a. N.E. Bauman
Moscow, Russia
vkutyrkin@yandex.ru

***Abstract* — Computational method for recognition the flavivirus species, including their subtypes, is proposed, that based on known sequence of viral genome. The method uses frequency characteristics of amino acid codons in the whole coding sequences of polyproteins. High reliability of this method is supported by correct recognizing of the 15 groups from different species and subtypes of flavivirus genomes.**

Keywords — flavivirus genome; latent profile triplet periodicity; flavivirus species recognition

Introduction

Flaviviruses belong to genus of arboviruses from Flaviridae family and they are transmitted through the mosquito and tick stings. Flaviviruses invade nervous system, inducing heavy diseases of human along with domestic and wild animals and birds [1–5]. Among them the most known diseases are tick-borne encephalitis, dengue fever, yellow fever, West Nile fever. To identify flavivirus species or subtype the reverse transcription and polymerase chain reaction with further cDNA sequencing of viral genome are applied along with traditional enzyme-linked immunosorbent assay (ELISA). Nucleotide sequence obtained in the result is compared with other known viral genome sequences in the GenBank.

Flavivirus genome is represented by single-stranded RNA of positive polarity, having length about 11 thousand nucleotides. Sequentially arranged with each other and translated in the same reading frame, the genes of viral proteins form a whole coding sequence (CDS) of polyprotein.

In present work the 15 groups of flavivirus polyprotein CDSs with total number of the representatives equal to 7060 were considered. Whole coding sequences of ten flavivirus species such as Bagaza virus, Japanese encephalitis virus, Murray Valley encephalitis virus, Powassan virus, Saint Louis encephalitis virus, Duck Tembusu virus, Tick-borne encephalitis virus, Usutu virus, the viruses of West Nile and Yellow fevers were analyzed. Besides, the four groups of polyproteins' CDSs were separately considered for the subtypes of Dengue fever virus (Dengue serotype 1, Dengue serotype 2, Dengue serotype 3, Dengue serotype 4), as well as Kunjin virus which is now considered as Australian subtype of West Nile fever virus was distinguished in a special group. In all analyzed CDSs of flavivirus proteins latent triplet periodicity was recognized, that is referred to earlier introduced new type of profile periodicity [6]. As more general model, explaining existence of such the latent periodicity, so called the SHOC-model has been proposed [7]. This model is unambiguously characterized by random codon which is a random value with the meanings in triplets of the genetic code. Consequently, in frame of the SHOC-model for coding DNA sequence, where the latent triplet periodicity is observed, an estimate for random codon's probability distribution may be obtained as a set (vector) of frequencies of the genetic codons occurred in this sequence. So, such the DNA sequences may be classified according to the types of the random codon distributions.

Thereby, analyzed 15 groups of the CDSs of flavivirus polyproteins induce a set of corresponding sampled probability distributions, forming a set of the points in the 64-dimentional space. Clusterization problem for such the points' set, where each distinct cluster is characterized by only one of the species or subtypes of the flavivirus considered, is addressed in the work. Algorithm proposed in the work is based on sampled probability distribution of the codons and it allows unambiguously assigning polyprotein CDS to a cluster of corresponding flavivirus species or subtype.

Materials and Methods

In the present work flavivirus genomes were selected from the GenBank release 231.0, dated April 15 2019. It should be noted that the introns in flavivirus genomes are practically absent, so flavivirus genome CDS is represented by uninterrupted coding sequence of polyprotein. Table I presents qualitative (**j**) and quantitative ($\mathbf{N_j}$) content of the 15 groups of flavivirus polyprotein CDSs analyzed.

TABLE I

j	Species and subtypes of flaviviruses	N_j
1	Bagaza virus	18
2	Dengue serotype 1 virus	1895
3	Dengue serotype 2 virus	1412
4	Dengue serotype 3 virus	922
5	Dengue serotype 4 virus	190
6	Japanese encephalitis virus	300
7	Kunjin virus	43
8	Murray Valley encephalitis virus	14
9	Powassan virus	21
10	Saint Louis encephalitis virus	38
11	Duck Tembusu virus	110
12	Tick-borne encephalitis virus	175
13	Usutu virus	147
14	West Nile virus	1652
15	Yellow fever virus	123

The mean value of frequency vector of the codons of the genetic code was calculated over all the genomes for each group. Such a vector unambiguously characterizes corresponding group, i. e. flavivirus species or subtype (serotype).

Question of identity with one or another group for flavivirus genome analyzed was solved by the following way. Original statistics was proposed in the work that calculates a distance of frequency vector for individual flavivirus genome from the mean values of frequency vectors for all other groups. Flavivirus genome analyzed was attributed to a group that showed minimum of distance between such the vectors. To prove a correctness of group choice the following additional procedure was used. For flavivirus genomes in each group a maximal distance of all corresponding codon frequency vectors from the mean value of codon frequency vector in the group was calculated. Choice of identity group was confirmed, if minimal distance found was no more such the maximal distance in the group.

Results

Method proposed in the work allows determining flavivirus species, if coding sequence of its polyprotein is known. This method was tested in recognition of 10 species and five subtypes of the flaviviruses with the genomes significantly represented in the GenBank. It was shown [8] that for 10 species and two subtypes the method provides 100% reliability. Mistakes in flavivirus species recognition amount to no more than 0.5%. In particular, 23 genomes among 7060 flavivirus genomes analyzed were not recognized and it was only two cases of wrong species determination.

Conclusion

New, effective and rapid method was proposed in the work to identify the species of flaviviruses, giving rise to heavy diseases of human, animals and birds, after viral genomes have been sequenced.

In general, application of the method proposed is not restricted by identifying RNA viruses only, particularly, flaviviruses. The method may be used for proving virus species from other genus and families, independently of RNA or DNA viruses are analyzed.

Method proposed in the work significantly simplifies a procedure of virus identification that is based on sequence of its genome, as no additional knowledge about conservative fragments of viral genome is needed, because alignment with other viral genomes from international databases is not carried out.

References

[1] J. Fernández-Pinero, I. Davidson, M. Elizalde et al., “Bagaza virus and Israel turkey meningoencephalomyelitis virus are a single virus species,” J. Gen. Virol., vol. 95, pp. 883-887, 2014.

[2] W. Zhang, S. Chen, S. Mahalingam, et al., “An updated review of avian-origin Tembusu virus: a newly emerging avian Flavivirus,”. J. Gen. Virol., vol. 98, pp. 2413-2420, 2017.

[3] E. Benzarti, A. Linden, D. Desmecht, et al., “Mosquito-borne epornitic flaviviruses: an update and review,” J. Gen. Virol., vol. 100, pp. 119-132, 2019.

[4] A. Diaz, L. Coffey, N. Burkett-Cadena, et al., “Reemergence of St. Louis encephalitis virus in the Americas. Emerging Infectious Diseases,” vol. 24., pp. 2150-2157, 2018.

[5] M. Clé, C. Beck, S. Salinas, et al., “Usutu virus: A new threat?”, Epidemiol. Infect., vol. 147. Article No. e232, 2019.

[6] M. Chaley, V. Kutyrkin, “Spectral-statistical approach for revealing latent regular structures in DNA sequences,” in Data Mining Techniques for the Life Sciences, O. Carugo, F. Eisenhaber, Eds. New York: Springer Science+Business Media, 2016, pp. 315-340.

[7] M. Chaley, V. Kutyrkin, “Stochastic model of homogeneous coding and latent periodicity in DNA sequences,” J. Theor. Biol., vol. 390. pp. 106-116, 2016.

[8] M. Chaley, Zh. Tyulko, V. Kutyrkin, “Flavivirus species recognition based on the polyprotein coding sequences,” Math. Biol. Bioinf., vol. 14, pp. 533-542, 2019. (In Russian)

DOI 10.18699/BGRS/SB-2020-092

Development of a method for recognizing biomedical entities in the texts of scientific articles

Stepan Derevyanchenko
Academy of Sciences
Novosibirsk, Russia
Novosibirsk State University
Novosibirsk, Russia sod97@yandex.ru
ORCID ID: 0000-0001-9433-8341

Abstract — In this work, the problem of recognizing named entities in the texts of scientific articles of biological orientation was solved. Using a combination of machine learning methods allowed us to achieve high recognition quality indicators.

Keywords — neural network, prediction, machine learning

Introduction

Currently, there are thousands of scientific journals in the world that publish research results in various fields of biology and medicine. Huge databases of patents are collected. The number of articles and patents increases exponentially over time. The analysis of such a huge number of experimental facts presented in text sources (scientific publications and patents) requires the use of automatic methods for extracting knowledge (text-mining). At the moment, most automatic methods of extracting knowledge are based on algorithms of computational linguistics and automatic text processing (Natural Language Processing). One of the most important tasks of modern computational linguistics is the task of recognizing named entities (Named Entity Recognition). In terms of its importance within this field, this task is comparable to the task of defining parts of speech in classical linguistics.

Results

In this paper, the problem of recognizing named entities is solved using the corpus of scientific articles from the PubMed Internet resource.

The paper considers a neural network approach for solving the problem of recognizing named entities. Specifically, a multi-layer recurrent neural network consisting of the following layers is used:

- Layer of trainable context-dependent vectors for words (Embeddings) [1]
- Bidirectional LSTM network [2]
- Fully connected layer (Linear)

In addition, classical machine learning models are used as a method for online verification of the neural network, which determines the presence of this entity in the text, which helps to significantly reduce the number of errors. the Russian

The paper compares the quality of logistic regression, naive Bayes, random Forest [4], and gradient boosting on trees (XGBoosting) [3]. As a result of the research, high quality metrics for recognizing biomedical entities (proteins, metabolites, organisms, diseases, etc.) were obtained.

Acknowledgment

The authors thank the Siberian Branch of the Russian Academy of Sciences for the budget project No. 0324-20190040-C-01.

References

[1] T. Mikolov, I. Sutskever, K. Chen, G. S Corrado, J. Dean
[2] “Distributed Representations of Words and Phrases and their
[3] Compositionality“ Advances in Neural Information Processing Systems, pp. 3111–3119, 2013
[4] S. Hochreiter, J. Schmidhuber “Long short-term memory“ Neural computation, vol. 9 (8), pp. 1735–1780, 1997.
[5] Y. B. Yann LeCun Leon Bottou, P. Haffner “Gradient-based learning applied to document recognition“ IEEE, 1998.
[6] L. Breiman “Random Forests“ Machine Learning, vol. 45 (1), pp. 5–32, 2001.
[7] K. S. Jones “A Statistical Interpretation of Term Specificity and its Application in Retrieval“ Journal of Documentation, vol. 60 (5), pp. 493–502, 2004.

DOI 10.18699/BGRS/SB-2020-093

SINE and LINE-1 competition for energy resources determines cell fate

Maria Duk
The Ioffe Institute, St. Petersburg, 194021 Russia,
Centre for Genome Bioinformatics, St. Petersburg University,
St.Petersburg, Russia,
duk@mail.ioffe.ru

Alexandra Chertkova
Peter the Great St.Petersburg Polytechnic University,
BioCAD,
St.Petersburg, Russia,
tschertkowa.alex@gmail.com

Vitaly Gursky
The Ioffe Institute, St. Petersburg, 194021 Russia,
Peter the Great St.Petersburg Polytechnic University,
St.Petersburg, Russia,
gursky@math.ioffe.ru

Maria Samsonova
Peter the Great St.Petersburg Polytechnic University,
St.Petersburg, Russia,
m.g.samsonova@gmail.com

Alexander Kanapin
St. Petersburg State University,
St.Petersburg, Russia,
a.kanapin@spbu.ru

Anastasia Samsonova
St. Petersburg State University,
St.Petersburg, Russia,
a.samsonova@spbu.ru

***Abstract*— Much of our DNA is repetitive sequence comprised of mobile genetic elements. Cell's energy expenditures associated with their life cycle and containment are negligible in normal conditions, as human cells have evolved to coexist with transposable elements and use them to their advantage. In disease, when genome defense mechanisms are compromised, energy costs associated with transcription, translation and integration of mobile elements soar. Here we demonstrate how a competition between two major types of human transposons may determine cell fate in pathological conditions.**

Keywords — retrotransposable elements, SINE, LINE-1, mathematical modelling, cell energy

Motivation and Aim

Throughout its life, the eukaryotic cell balances energetic costs of processes such as growth, division, movement, as well as genome maintenance, transcription and translation. Of these, the latter processes are the most energy consuming [1], as protein synthesis accounts for ~75% of total cell's energy budget [2]. The bulk of our genomes (~45%) is made of retrotransposable elements (rTEs), - (semi)autonomous mobile genetic elements, capable of propagation through the genome via a copy-and-paste mechanism [3]. In disease situation, when genome protection mechanisms are compromised, unrestricted TEs mobilization may introduce new gene regulatory circuits, lead to genome instability, spontaneous mutagenesis and lead to considerable expansion of translatable genome [4,5]. An intriguing question arises: Is it possible for a cell to run out of energy and cell-building materials due to unrestricted propagation of transposable elements and associated energetic burden?

Here we present a mathematical model of a cell incapable of efficient protection from novel TE insertions (i.e., with compromised genome defense). In this model, we consider competition between two widespread types of rTEs, namely, SINEs and LINEs for various cell resources (e.g., energy, ribosomes and proteins), as well as for proteins essential for transposition.

Materials and Methods

We built a mechanistic model of a cell using principles given in [6]. The model includes cell energy (such as ATP molecules), number of copies of transposable elements of both types, a corresponding number of transcripts, proteins ORF1p and ORF2p encoded by LINE-1, and, finally, the remaining cellular proteins and their mRNAs as variables. Since the processes of translation and integration of TEs are not instantaneous, to account for that and include dependence on number of available ribosomes and proteins, we also consider numbers of complexes mRNA+ribosome and mRNA+ORF2p as variables. Various parameters required by the model, corresponding to concentrations of molecules, reaction rates etc., were derived from literature sources and databases, such as BioNumbers [7]. Since the number of rTEs increases with time, this is not a steady-state model, although concentrations of other molecules may reach a plateau resembling a quasi-stationary state. The system is pretty stable in this state, however certain cell resources may be exhausted.

Results

The coefficients of the model were varied to examine its behavior under different scenarios, including those where TE competition for energy resources and compromised genome defense mechanisms determine cell fate. Our numerical experiments demonstrate that the cell is protected from fatal energy deprivation, by constraints on rTE integration and transcription rates. Furthermore, variations in ribosome numbers have no effect on cell fate whatsoever. We've shown that cell survival depends on the ratio between transcription rates of LINE and SINE elements. Interestingly, yet another essential factor contributing to cell survival is a fierce competition between LINE and SINE transcripts for the LINE-encoded reverse transcriptase. A complete shutdown of genome defense mechanisms results in strong fluctuations in concentration of LINE protein products, thus exhausting energy resources and cell-building materials.

Aberrant genome protection triggers unrestricted propagation of transposable elements, however cell death or survival is determined by the competition between different types of retrotransposable elements: increased number of LINE transcripts is a killer, while elevated SINE activity protects the cell, at the cost of, perhaps, a higher mutation rate.

AKNOWLEDGEMENTS

This work was supported by RSF fund, grant 20-14-00072 to AK, AS (study design, data analyses).

REFERENCES

[1] Lynch M, Marinov GK. The bioenergetic costs of a gene. Proceedings of the National Academy of Sciences. National Acad Sciences; 2015 Dec 22;112(51):15690–15695. PMCID: PMC4697398

[2] Lane N, Martin W. The energetics of genome complexity. Nature. Nature Publishing Group; 2010 Oct 21;467(7318):929–934.

[3] Beck CR, Garcia-Perez JL, Badge RM, Moran JV. LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet. 2011 Sep 22;12:187–215. PMID: 21801021.

[4] Belancio VP, Deininger PL, Roy-Engel AM. LINE dancing in the human genome: transposable elements and disease. Genome Med. 2009 Oct 27;1(10):97. PMCID: PMC2784310.

[5] Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. Science. American Association for the Advancement of Science; 2016 Feb 12;351(6274):aac7247–aac7247. PMCID: PMC4788378.

[6] Weiße AY, Oyarzún DA, Danos V, Swain PS. Mechanistic links between cellular trade-offs, gene expression, and growth. Proc Natl Acad Sci USA. 2015 Mar 3;112(9):E1038–E1047.

[7] Milo R, Jorgensen P, Moran U, acids GWN, 2010. BioNumbers—the database of key numbers in molecular and cell biology.

DOI 10.18699/BGRS/SB-2020-094

An architecture-independent algorithm for microRNA target prediction

Natalya Fokina
Department of biology and general genetics
Moscow State Medical University
Moscow, Russia
retrospector@ispras.ru

Alexander Grinev
Department of biology and general genetics
Moscow State Medical University
Moscow, Russia
grinev.ab@1msmu.ru

***Abstract* — Nowadays the research on RNA interference is a subject of growing interest. One of the most perspective directions of this research is the use of bioinformatics methods to predict possible interactions of microRNAs with their mRNA counterparts, which allows to increase the accuracy and efficiency of their confirmation using laboratory means. In spite of this fact, the existing tools are unable to meet all the needs of molecular biology. An example of a problem currently lacking an appropriate solution is the study of RNA interference between organisms belonging to different taxonomic groups, notably, in parasite-host systems. Parasites utilise the host cell as a source of nutrients and habitat; molecular genetic profiles of both participants of this interaction are subject to dynamic changes. In this paper we present a novel algorithm for detection of interactions involving RNA molecules in a wide range of organisms. The proposed solution exceeds the existing ones in several characteristics and, in our view, can be broadly used in molecular biology and bioinformatics.**

Keywords* — *RNA interference, microRNA, algorithm, target prediction, RNA

Introduction

Since the discovery of RNA interference in 1997 the existence of this mechanism has been demonstrated in numerous organisms belonging to different taxonomic groups. The research on this mechanism is of great significance not only for fundamental biology but also for its applied fields: medicine, veterinary science, agriculture, fisheries and forestry. Due to the availability of genome sequences of various organisms, considerably higher performance of modern computers and lower cost of equipment and machine time compared with laboratory research, experimental studies are often preceded by the identification of potential interactions *in silico*. Nevertheless, microRNA target prediction faces two converse problems. Firstly, existing tools demonstrate high level of errors of the second kind, which result in mRNAs that are in fact not valid targets for the corresponding microRNAs being classified as their targets and, in turn, impede a further *in vitro* analysis of the microRNAs in question. On the other hand, the number of errors of the first kind is also high. The existing instruments are primarily aimed at target mRNA identification in humans and several model organisms, such as *Drosophila melanogaster*, *Mus musculus*, *Caenorhabditis elegans*, *Danio rerio*, etc., for which these interactions have been investigated experimentally; high precision of these tools is achieved through the use of species-specific heuristics. Instruments implementing more generic algorithms are uncommon. The most popular of them, rna22 [1], is notably less accurate and slow in comparison with the aforementioned tools [2]. Furthermore, the vast majority of such tools, including those implementing algorithms that have got the most potential for further refinement, are closed-source. Thus, the confluence of the aforementioned factors makes it relevant to develop a novel algorithm for identification of potential interactions between microRNAs and mRNAs that does not utilise information about the architectures of the genomes of the investigated organisms and has got acceptable accuracy and performance.

The proposed solution

A characteristic feature of all non-coding RNAs participating in RNA interference is the existence of the so called seed region, i.e. a short sequence of nucleotides with almost perfect complementarity to the corresponding sequence in a target mRNA. Identification of this complementary sequence in a genome in the first and foremost problem in microRNA target prediction. Thus, the problem under consideration is to identify all potential microRNA targets by locating subsequences complementary to all of the possible microRNA seed regions within a whole genome given as a single sequence or a set of sequences. Formally, given two sets of strings, M and G, and two integers $l \in N$ and $k \in Z, k \geq 0$ ($k < l$), for all substrings r_i, $|r_i| = l$ of each string $r = rc(m)$, $m \in M$, where $rc()$ is the function of reverse complement construction, it is necessary to identify in each string $g \in G$ positions of substrings g_j such that $DL(r_i, g_j) \leq k$, where DL is the Damerau-Levenstein distance [3].

It is worth pointing out the key differences of our approach from those implemented in other tools. First of all, many tools utilise the observation that a seed region usually includes nucleotides from 2 to 8 of a microRNA, which naturally leads to the increase in performance of such tools. However, this assumption is not valid for all organisms and can not be adopted by an architecture-independent algorithm. Oppositely, several algorithms expect that all microRNA targets are located within 3'-untranslated regions (3'UTR) of mRNAs, which is also not a general truth [4]. Besides, the latter assuption requires the genome to be annotated, which makes these tools inapplicable for many organisms since their genome sequences are lacking appropriate annotations, and their creation is a separate complex and laborious task. Moreover, some tools imply that the user provides them with a list of mRNAs rather than a whole genome sequence, which significantly reduces the complexity of search but for the same reason significantly narrows the range of organisms for which the tools can be utilised. To sum it up, for an algorithm to meet the requirement of being architecture-independent, it is inevitable to reject all the aforementioned assumptions.

In the proposed algorithm, instead, we make use of the general observations. Since the length of the seed region l and the threshold Damerau-Levenstein distance k are typically small, we explicitly generate all the strings for which we search, i.e. we transform the set S to the set S^0 such that:

$$\forall s^0 \in S^0 \Leftarrow\Rightarrow \exists s_i \in s, s \in S \wedge |s_i| = l \wedge DL(s^0, s_i) \leq k \quad (1)$$

We proceed with searching these strings using the AhoCorasick algorithm for multiple string search [5]. The choice of the Aho-Corasick algorithm has been also supported by the fact that the most common mode of usage of our tool is the simultaneous detection of targets for multiple microRNAs. Thus, the workflow of the proposed algorithm is as follows:

1) transform the specified set of microRNAs M to the set R by applying the reverse complement function to all $m \in M$;
2) construct the set of strings to search R^0 by generating all the possible permutations of all the substrings of the specified length l for strings $r \in R$ according to (1);
3) run the Aho-Corasick algorithm to search all the strings from R^0 in the set of strings G representing the genome.

It is noteworthy that the proposed approach leaves plenty of room for further improvement as there are no data dependencies between different parts of the investigated genome, hence the algorithm can be easily applied in parallel to its various fragments using a given number of computational nodes. Even greater performance gains can be achieved through the use of hardware vectorisation and parallelisation, such as a SIMD instruction set (e.g. MMX, SSE and AVX on x86/x86 _64 processors or AltiVec/VMX on PowerPC), as well as GPU computing.

Conclusions

The present paper addresses the core problem lying behind the design of the string-based architecture-independent approach to microRNA target prediction, which is the detection of genome subsequence complementary to the microRNA seed region. We have developed an efficient algorithm allowing to detect these regions within a given genome. The algorithm has been implemented as a software tool; its performance has been evaluated in comparison with rna22. The comparison has demonstrated the significantly higher performance of the developed tool: on a set of protozoan genomes it outperforms rna22 by several orders of magnitude. Performance testing was conducted on a 3.40 GHz Intel Xeon E3-1240 v2 computer with 8 Gb of RAM under Ubuntu Linux 14.04 x86 64. It has shown that calculating targets for a single microRNA located within a single chromosome (constituting on average 1.5 megabases) takes about 20 minutes in case of our tool and from 7 to 8 hours in case of rna22.

Still, to be implemented as a piece of off-the shelf software the approach requires refinements dealing with the high level of errors of the second kind. Possible solutions to this problem include target site accessibility and free energy analyses or the use of machine learning. Other directions for future work include further performance improvement by means of parallelisation and vectorisation.

References

[1] K. Miranda, H. Tien, Y. Tay, Y.-S. Ang, W. L. Tam, A. Thomson, B. Lim, and I. Rigoutsos, "A pattern-based method for the identification of microrna binding sites and their corresponding heteroduplexes," Cell, vol. 126, pp. 1203–1217, October 2006.

[2] A. Oliveira, L. Bovolenta, P. Nachtigall, M. Herkenhoff, N. Lemke, and D. Pinhal, "Combining results from distinct microrna target prediction tools enhances the performance of analyses," Frontiers in Genetics, vol. 8, April 2017.

[3] D. Gusfield, Algorithms on strings, trees, and sequences: computer science and computational biology. Cambridge University Press, 2007.

[4] A. L. Riffo-Campos, I. Riquelme, and P. Brebi, "Tools for sequencebased miRNA target prediction: What to choose?" International Journal of Molecular Sciences, vol. 17, p. 1987, December 2016.

[5] A. Aho and M. Corasick, "Efficient string matching: An aid to bibliographic search," Commun. ACM, vol. 18, pp. 333–340, June 1975.

DOI 10.18699/BGRS/SB-2020-095

Computer-assisted analysis of caspases molecular evolution

Alexey Zamaraev
MSU, Moscow, Russia
a-zamaraev@yandex.ru

Gelina Kopeina
MSU, Moscow, Russia

Konstantin Gunbin
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
genkvg@gmail.com

Abstract — **The functions of any protein are driven by their chemical and physical properties, which, in turn, are determined by steric and physico-chemical folding requirements. Therefore, it is expected that replacement of the amino acid tightly interacting with a large number of other amino acids is related to changes in the context of interactions in the protein globule. Recent studies of the protein evolution revealed various signatures of substitution asymmetry and heterotachy. Here, based on reconstruction of ancestral libraries we analyze the substitution asymmetry in the molecular evolution of caspases protein family characterized by huge number of molecular functions.**

Keywords — caspases, molecular evolution, substitution asymmetry, ancestral libraries

Motivation and Aim

Up to now, the vast majority of available procedures of reconstruction of ancestral sequences are based on the symmetric single (applied to all protein sites) matrix of amino acid substitution rates. It seems obvious that new software tools for ancestral protein reconstruction, taking into account substitution limitations from the 3D protein structure and from the stability of its folding (for example, ProtASR), should be most useful. However, unfortunately, the experimentally resolved 3D protein structures are still lacking. Another way for substitution asymmetry accounting in the ancestral protein reconstruction is the construction of ancestral libraries [1]. To make the ancestor libraries accurate enough, it has recently been proposed to use the AltAll reconstruction approach. This approach combines all possible alternative states introduced into one protein, and then characterizes this protein with a set of these states [2, 3]. It has been shown that this approach significantly improves the imperfection of individual ancestral sequences reconstructed by Bayesian approach.

Methods

Multiple alignment was done using PROMALS. The best reversible and symmetric models of amino acid substitution rates were selected by IQTree v. 1.5.4 (in our case it is C20+G4). We adjusted initial phylogenetic tree topology using Metazoa species tree from TimeTree DB by the TreeFix v. 1.1.10 software. After this, the branch lengths were re-optimized using IQTree v. 1.5.4 and the best reversible model of amino acid substitution rates. Bayesian sampling of ancestral sequences in each internal tree node was carried out using PhyloBayes v. 4.1, the CAT model [4], and 6 rate categories of sites. In order to construct complete and truncated (using our modified approach, called 'AltAll * N') libraries of ancestral sequences we used AltAll * N' procedure. Our procedure 'AltAll * N' is an iterative rewriting of all probable (with a posteriori probability> 0.1) alternative states in a consensus ancestral sequence of given tree internal node. For example, if there are 3 alternative states in site A and 4 alternative states in site B of ancestral sequence (node) X, then we must rewrite this ancestral sequence 4 times to get 4 alternative ancestors: a) a sequence consisting of the best states in A- and B-sites, b) a sequence with the second most probable states of sites A and B, c) a sequence with the third probable states of sites A and B, and d) a sequence with the third probable state in sites A and with the fourth probable in site B.

Substitution asymmetry during protein evolution was detected by analyzing deviations in the protein evolution rates from the reversible and symmetric model of amino acid substitution rates on each of the branches of the protein tree (1) and by a comparative analysis of the lengths of the internal tree branches taking into account the structure of the protein (2). (1) In order to solve first task, we: a) reconstructed a protein-specific, time-reversible model of amino acid substitution rates using Modelestimator software; b) for each possible amino acid replacement on each internal node of the tree, we calculated d=PPa*PPb*2*NC, where PPa and PPb are the posterior probabilities of amino acids a and b, a $\neq$ b, NC=1/(1+e^(200*RFab)), RFab is the relative rate of ab substitution in the time-reversible model of amino acid substitution rates; c) summed d values along all sites in each internal tree node and calculated the natural logarithms of these sums ($\ln(\sum d)$); d) in order to identify branches with a maximum $\ln(\sum d)$, we conducted nonparametric comparison of these values over the entire tree. (2) In order to solve second task, we used truncated 'AltAll * N' libraries, particularly we: a) for each alternative ancestral sequence in each internal tree node, using the RaptorX_Property Fast software system, we assigned secondary structures to each amino acids, derived amino acid solvent accessibilities, and predicted disorder for each amino acid residue; b) calculated the frequency of changes of these measures between all alternative ancestral sequences of neighboring nodes residing on each internal tree branch; c) in order to identify branches with maximum structural changes, conducted nonparametric comparison of the abovementioned frequencies of changes across the entire tree.

Results

We extracted and verified 1565 Deuterostomia caspase proteins from NCBI GenBank. After that, we processed this protein set as described above.

One of the most promising observation we made is the finding of strict substitution asymmetry in the diversification of lower rodents, characterized by higher longevity comparing to higher rodents (Fig 1). The same evolutionary event is characteristic for caspase 7 in Xenopus (Fig 2).

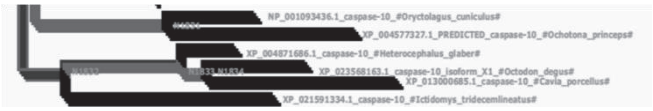


Fig. 1. Selected part of caspases phylogenetic tree, containing caspase 10 of lower rodents. Each branch composed of 5 lines, corresponding to (from down to up) changes in amino acid disorder, protein secondary structures (3 and 8 types of secondary structure), amino acid solvent accessibilities, and deviations from the reversible and symmetric model of amino acid substitution. Black lines show terminal branches, light grey indicates evolution without significant deviations, while dark grey indicates significant deviations from expectations.

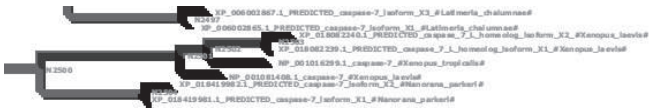


Fig. 2. Selected part of caspases phylogenetic tree, containing caspase 7 of Xenopus. Same designations as in Fig. 1.

It is of interest, that such evolutionary events are characteristic for the divergences of large protein clades tightly related with origination of large taxonomical groups, for example, avian caspase 6, mammalian caspase 13, etc.

ACKNOWLEDGMENT

This work supported by the grant from the RSF # 19-1500125. BZ is supported by the Swedish (190345) and Stockholm (181301) Cancer Societies.

REFERENCES

[1] Gumulya Y, Gillam EM. (2017) Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro'approach to protein engineering. Biochem J. 474(1):1–9.

[2] Anderson DP, et al. (2016) Evolution of an ancient protein function involved in organized multicellularity in animals. elife. 5:e10147.

[3] Eick GN, et al. (2017) Robustness of reconstructed ancestral protein functions to statistical uncertainty. Mol Biol Evol. 34(2):247–61.

[4] Quang le S, Gascuel O, Lartillot N. (2008) Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics. 24(20):2317–2323.

DOI 10.18699/BGRS/SB-2020-096

ANDDigest: a text-mining based computer system for generating digests in the field of biology

Timofey Ivanisenko
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk,
Russia
NSU, Novosibirsk, Russia
itv@bionet.nsc.ru

Pavel Demenkov
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk,
Russia
demps@bionet.nsc.ru

Vladimir Ivanisenko
Kurchatov Genomics Center,
ICG SB RAS,
Novosibirsk, Russia
NSU, Novosibirsk, Russia
salix@bionet.nsc.ru

Nikolay Kolchanov
Kurchatov Genomics Center,
ICG SB RAS,
Novosibirsk, Russia
kol@bionet.nsc.ru

***Abstract* — The vast volumes of accumulated scientific literature on the Internet have made the task of finding the information that most fully corresponds to the area considered by the researcher one of the most critical problems. The developed ANDDigest tool allows performing a search in scientific literature, using the combination of dictionaries and user-provided keywords. Also, the system allows obtaining information about trends, reflecting a dynamic of interest to biological objects, recognized in texts, based on their mentions in PubMed.**

Keywords* — *text-mining, ANDSystem, knowledge retrieval, dynamics of interest, trend analysis

Motivation and Aim

Motivation

The vast amounts of published scientific literature made the task of searching for the information relevant to the field of the researcher's study one of the most important problems. Well-known search engines, such as Google Scholar, Web of Science, PubMed, Scopus, and others, are highly effective in finding documents; however, they do not allow the use of an ontology of the subject area [1, 2, 3]. Programs based on automatic text analysis provide the extraction of knowledge and its presentation in the form of semantic networks [4]. Of particular interest among them are systems providing a full cycle of knowledge engineering, including their automatic extraction, integration, presentation in the form of semantic networks, visualization, and analysis. Examples of such software products are STRING [5], Pathway Studio [6], MetaCore [7, 8], etc.

Aim

Earlier, we developed the ANDSystem tool [9, 10, 11, 12] designed to implement a full cycle of engineering and knowledge management. However, the drawback of these systems, in contrast to universal search engines (Google and others), is the tight attachment to their ontologies that describe the subject area, which limits the possibilities of expanding search queries beyond the limits of the used ontology.

Methods

All pre-processing of the textual data, including the conversion, normalization, syntactic analysis, named entity recognition, etc. performed using the text-mining module of ANDSystem. The associations between recognized entities from the ANDSystem ontology established using the STRING co-occurrence-based text-mining scoring scheme [13]. Each recognized entity presented by the trend based on the dynamics of its mentions in the scientific literature by years. The estimation of strength for each trend calculated using a non-parametric Mann-Kendall test [14].

Results

The developed ANDDigest system designed for the complex search, analysis, and graphical representation of information retrieved from the textual sources. The system allows to perform:

(1) automatic adjustment of the subject area using the ontological description of the subject area specified in the form of dictionaries of objects. The current version of the program uses the ANDSystem ontology, which contains dictionaries for a wide range of fundamental molecular genetic entities and processes (such as proteins, genes, miRNAs, metabolites, drugs, biological processes, and diseases, etc.);

(2) an advanced search for scientific publications and international patents and other textual documents based on integrated queries, with taking into account the hematology of the subject area and user-specified keywords;

(3) analysis and graphical representation of found documents in the form of short descriptions – digests, including:

1. sorting by publication date, citing number, etc.
2. visualization of digests with a mapping of entities used by the ontology of the subject area;
3. assessment and visualization of trends, describing the dynamic of interest to the objects identified in documents, based on the frequency of their mentions in publications.

The web version of ANDDigest is free and available at http://anddigest.sysbio.ru

Acknowledgment

Supported by the Kurchatov's Genomics Center of the ICG SB RAS (agreement with the Ministry of Science and Higher Education of the Russian Federation No. 075-15-2019-1662).

References

[1] Beel J., Gipp B. (2009) Google Scholar's Ranking Algorithm: An Introductory Overview. In: In Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09). 1: 230-241.

[2] McEntyre J., Ostell J. (2002) The NCBI Handbook. Bethesda (MD): National Center for Biotechnology Information (US).

[3] Jacso P. (2005) As we may search - Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. Current Science. 89: 1537-1547.

[4] McEntyre J., Ostell J. (2013) The NCBI Handbook. Bethesda (MD): National Center for Biotechnology Information (US).

[5] Szklarczyk D. et al. (2016) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic acids research. gkw937.

[6] Nikitin A. et al. (2003) Pathway studio -- the analysis and navigation of molecular networks. Bioinformatics. 19(16): 2155-2157.

[7] Nikolsky Y. et al. (2005) Biological networks and analysis of experimental data in drug discovery. Drug discovery today. 10(9): 653-62.

[8] Ekins S. et al. (2006) Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms. Xenobiotica. 36(10-11): 877-901.

[9] Demenkov P.S. et al. (2012) ANDVisio: A new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. In Silico Biology. 11(3, 4): 149-161.

[10] Ivanisenko V.A. at al. (2015) ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. BMC Syst Biol. 9(S2): S2.

[11] Ivanisenko V.A. et al. (2019) A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. BMC Bioinformatics. 20(1): 34.

[12] Saik O.V. et al. (2019) Prioritization of genes involved in endothelial cell apoptosis by their implication in lymphedema using an analysis of associative gene networks with ANDSystem. BMC Med Genomics. 12(2): 47.

[13] Franceschini A. et al. (2012) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41(D1): D808-D815.

[14] Libiseller C., Grimvall A. (2002) Performance of partial Mann-Kendall tests for trend detection in the presence of covariates. Environmetrics. 13(1): 71-84.

DOI 10.18699/BGRS/SB-2020-097

In silico model of glioma MTS growth. Effects of compression and mechanical ECM remodeling

Vladimir Kalinin
R&D Sector of TMA, Ireland
vladimir.kalinin@tma-science.ie

***Abstract* — a continuous *in silico* model is built to simulate glioma Multicellular Tumor Spheroid (MTS) growth. Visco-elastic cell aggregation and elastic Extracellular Matrix (ECM) represent two compressible constituents of the composite system. The simulations show the important role of compressibility effects in avascular tumor, such as densification of the central tumor zone with uniform distribution of total stress over the major MTS part. The model demonstrates effects of cell- cell adhesion and mechanical deformation of ECM scaffold on cell transport, generally conditioned by hydrostatic forces in cell- ECM composite and cell motility. Transitions between three cell phenotypes: proliferative, hypoxic and necrotic is triggered by the local oxygen concentration that is being limited by oxygen diffusion into spheroid from the periphery of the calculation domain. MTS stresses undergo sharp change within a thin Transition Layer on the spheroid surface, which affects mechanically the structure of ECM scaffold. The simulations demonstrate stimulating effect of this mechanical remodeling on glioma cell invasion.**

Keywords* — *Multicellular Tumor Spheroid, Extracellular Matrix, cell- ECM stress, in silico model

Introduction

A dramatic role of cancer mechanobiology in its progression has been clearly demonstrated within the last decade [1]. The current model is built to study mechanobiological interplay between stress effects in compressible cell- ECM composite, cell- cell adhesion and cell phenotype transitions on glioma invasion.

Model basics

The basic model assumptions are as follows: (a) Spherically symmetric simulation domain contains two components: cells and ECM, and does not present any phenomenologically determined spatial subdivisions of its physical content into different parts; (b) the cellular component may include multiple voids, so being porous generally behaves as an isotropic visco- elastic compressible agglomeration; (c) ECM is isotropic, compressible, linear elastic, porous material. Cells proliferation and their motion through ECM scaffold produce elastic stresses in ECM; (d) cells exert attractive and repulsive forces on each other depending on intercellular distances. The lower limit of cell concentration in the colony is determined by the critical distance between them. Beyond this cut-off point ($C_{tot} < C_\infty$) cells cannot forcibly affect each other; the technically maximum cell packing density C_{max} determines upper cut-off poin; (e) the total stress distribution in the composite system follows iso-strain approximation. This intercellular force field together with ECM stress gradients cause motion of the cells tending to minimize the total potential energy of the system; (f) Cell motility is modeled through their diffusion. The diffusion rate is specific for different cell phenotypes. Stress in cellular component σ_c is

$$\sigma_C = \gamma K_{C0}\left(\frac{C_{tot}}{C_{eq}} - 1\right), \text{ where}$$

$$\gamma = \begin{cases} 1, C_{tot} \geq C_{eq} \\ \frac{27}{4}\frac{\sigma_{att}}{K_{C0}}\frac{C_{eq}}{(C_{eq}-C_\infty)^3}(C_{tot}-C_\infty)^2, \; C_{eq} > C_{tot} \geq C_\infty \\ 0, C_{tot} < C_\infty \end{cases}$$

$C_{tot} = C_p + C_m + C_n$; C_p – concentration of proliferating cells; C_m – concentration of hypoxic cells; C_n – concentration of necrotic cells; $C_{tot} = C_{eq}$- stress equilibrium; $C_{tot} < C_{eq}$ generates contractile stresses and $C_{tot} > C_{eq}$ - compression. K_{C0} is a bulk modulus of the cell aggregation at $C_{tot} = C_{eq}$.

The behavior of cell- ECM system is similar to thermal expansion. Quasi-static stress conditions are being assumed in the model. Strain-displacement and stress-strain relations [2] yield the governing equation for displacement *u,* resulting from cellular dynamics as:

$$\frac{\partial}{\partial r}\left(\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 u\right)\right) = \frac{(1+\nu)}{(1-\nu)}\delta_V \text{ , where}$$

ν is ECM Poisson's ratio; δ_V- the inner volumetric stain produced by expanding cell aggregation inside the matrix

$$\delta_V = \frac{1}{K_{ECM}}\frac{d\sigma_C}{dr} = \frac{\gamma K_{C0}}{K_{ECM}}\frac{d}{dr}\left(\frac{C_{tot}}{C_{eq}}\right)$$

K_{ECM} - ECM bulk modulus. Radial and tangential components of the strain tensor are as follows: $\varepsilon_{rr} = du/dr$; $\varepsilon_{\theta\theta} = \varepsilon_{\varphi\varphi} = u/r$. The components of ECM stress tensor are as:

$$\sigma_{rr} = \frac{K_{ECM}}{(1+\nu)}\left((1-\nu)\,\varepsilon_{rr} + 2\nu\varepsilon_{\varphi\varphi}\right), \quad \sigma_{\varphi\varphi} = \frac{K_{ECM}}{(1+\nu)}\left(\nu\varepsilon_{rr} + \varepsilon_{\varphi\varphi}\right).$$

And the total stress in ECM $\sigma_{ECM} = (\sigma_{rr} + 2\,\sigma_{\varphi\varphi})/3$

Iso-strain approximation gives for total stress in the composite system:

$$\sigma_{tot} = \left(\frac{C_{max} - C_{tot}}{C_{max}}\sigma_{ECM} + \frac{C_{tot}}{C_{max}}\sigma_C\right).$$

Dynamics of the composite system

Transport, proliferation and phenotype transition for the three groups of cells are described by the equation:

$$\frac{\partial C_i}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 v_D C_i\right) = \frac{1}{r^2}\frac{\partial}{\partial r}\left[D_i r^2 \frac{\partial}{\partial r}\left(C_i\right)\right] + S_i$$

where C_i is one of the three cell phenotypes C_p, C_m, C_n;

D_i are diffusion coefficients of i^{th} type of cell: D_p, D_m and obviously $D_n = 0$; v_D is drift velocity of cells for the case of stress-dependent ECM permeability, as per model [3]:

$$v_D = -M_0\left(1-\frac{\varepsilon_v}{3}\right)^2 \frac{\partial \sigma_{tot}}{\partial r},$$

where ε_v is ECM dilatation and M_0 is mobility of cells in unstressed ECM. The model assumes that the cell population growing into ECM scaffold causes its expansion, stretching pores and therefore affecting its permeability and further-mobility of invasive cells. Velocity v_D is the same for all types of cells in the model. S_i represents sources S_p, S_m and S_n for proliferative, hypoxic and necrotic cells correspondingly.

$$S_p = \frac{C_p}{\tau_p}\left(1-\frac{C_{tot}}{C_q}\right) - h_{pm}\frac{C_p}{\tau_{pm}} + h_{mp}\frac{C_m}{\tau_{mp}},$$

$$S_m = h_m\frac{C_m}{\tau_m}\left(1-\frac{C_{tot}}{C_q}\right) + h_{pm}\frac{C_p}{\tau_{pm}} - h_{mp}\frac{C_m}{\tau_{mp}} - h_{nec}\frac{C_m}{\tau_n},$$

$$S_n = h_{nec}\frac{C_m}{\tau_n},$$

where τ_p and τ_m are characteristic times of proliferation for proliferative and hypoxic cells, τ_{pm}, τ_{mp}, τ_n are phenotype transition times: proliferative to hypoxic, hypoxic to proliferative and hypoxic to necrotic cells correspondingly. The maximum carrying capacity C_q is considered here as the upper cut-off limit, assessed as $C_{max} = 8.3 \times 10^8$ cm^{-3} basing on [4]. The switch factors h_{pm}, h_{mp} and h_{nec} are Heaviside functions of oxygen concentration triggering local cell phenotype transition. Oxygen is critical metabolite and the only nutrient in this model. Diffusion of oxygen and its local consumption by the cells determine the local concentration n_{ox} as per equation

$$\frac{\partial n_{ox}}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left[D_{ox}r^2\frac{\partial}{\partial r}(n_{ox})\right] - (\alpha_p C_p + \alpha_m C_m)$$

n_{ox} is oxygen concentration, D_{ox} - diffusion coefficient of oxygen, α_p and α_m - oxygen uptake for proliferative and hypoxic cells correspondingly. Boundary Conditions for the governing equations are based on **s**pherical symmetry that dictates symmetry conditions for all profiles X_i about r=0: $dX_i/dr\ (r=0) = 0$. We assume zero displacement and strain on the distant border of the domain r= R_0: $U(r=R_0) =$, $\varepsilon_{rr}\ (r=R_0) = 0$.

Initial conditions were set in accordance with experimental setup [4]. The initial spheroid radius has been set to $R_0 = 2.5 \times 10^{-4}$ cm and cell concentration $C_{init} = 8.0 \times 10^7$ cells/cm^3, along with a uniform initial distribution of oxygen in the domain $n_{ox}(r, t=0) = 4$mg/L. Rheological and cellular process parameters have been taken from [1, 4-6].

Results and Discussion

The simulated cell density profiles are in line with experimental data [4] (fig.1). Cell densities measured in the center cannot be explained by local cell proliferation only. Densification of the spheroid center takes place through its compression due to pressure exerted by the expanded ECM. The resulting cell density profile takes elongated shape by day 3 with extended invasive zone (Fig. 1).

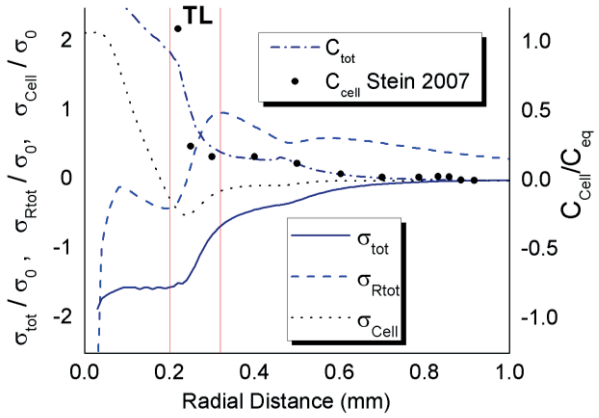


Fig. 1. Calculated distribution of C_{tot} for day 3 of MTS growth corresponds well to the experimental data of [4]; C_{eq}= 3.5 x 10^8 cells/cm^3. Distribution of total stress σ_{tot} / σ_0 is calculated within iso-strain model, also its radial component σ_{Rtot} / σ_0 and stress in cellular components C_{cell}/C_e are shown; σ_0= 100Pa

Cell- ECM interaction forms Transition Layer (TL) on the spheroid surface, where intercellular stress changes rapidly from compressive to tensile character, while the components of ECM stress exhibit also sharp, but opposite transition (fig.1). ECM reaction on MTS expansion forms compressive radial stress, along with circumferential tensile stress within (TL). This specific stress distribution within TL obtained in simulation should affect the structure of ECM scaffold. In case of randomly arrayed fibers they should expectedly be rearranging in front of TL, taking directions normal to MTS radius and align to its surface following traction forces inside TL. This scenario of mechanical collagen remodeling has been observed *in vitro* in [7], where these traction forces are assumed to facilitate tumor invasion providing tracks through mechanical alignment in matrix structure. Another effect of mechanical ECM remodeling through matrix deformation studied quantitatively in the current simulations is mechanical modulation of matrix porosity and therefore drift velocity of cells v_D, as shown in section III. The average effect of this matrix deformation resulted in increase of tumor invasion speed up to 20%.

References

[1] B. Ananthanarayanan, Y. Kim, S. Kumar, **"**Elucidating the mechanobiology of malignant brain tumors using a brain matrix-mimetic hyaluronic acid hydrogel platform," Biomaterials, 32(31), 2011, pp.7913-23.

[2] L.Landau, E.Lifshitz, Theory of Elasticity, Course of theoretical Physics. Pergamon Press, Bristol 1970

[3] F.J. O'Brien, B.A. Harley, M.A. Waller et al., " The effect of pore size on permeability and cell attachment in collagen scaffolds for tissue engineering," Technology and Health Care, 15(1) , 2007, pp. 3-17.

[4] A.Stein, T. Demuth, D.Mobley, M. Berens, L.Sander, "A mathematical model of glioblastoma tumor spheroid invasion in a three-dimensional in vitro experiment," *Biophysical Journal*, **92**(1), 2007, pp. 356–365.

[5] C.B. Raub, A.J. Putnam, B.J. Tromberg et al.,"Predicting bulk mechanical properties of cellularized collagen gels using multiphoton microscopy," Acta Biomater., 6(12), pp.4657–4665, Dec 2010.

[6] A.Martínez-González, A. Calvo, G. F., Pérez Romasanta, et al., "Hypoxic cell waves around necrotic cores in glioblastoma: a biomathematical model and its therapeutic implications." Bulletin of mathematical biology, 74(12), 2012, pp. 2875-96.

[7] K. S. Kopanska, Y. Alcheikh, R. Staneva, et al. "Tensile Forces Originating from Cancer Spheroids Facilitate Tumor Invasion," PLOS ONE, 11(6), June 7, 2016.

DOI 10.18699/BGRS/SB-2020-098

ECM stiffness effects and subtumor formation in glioma growth. *In silico* model

Vladimir Kalinin
R&D Sector of TMA, Dundalk, Ireland
vladimir.kalinin@tma-science.ie

Abstract — **This study is based on *in silico* model and demonstrates how variations of Extracellular Matrix (ECM) stiffness may affect glioma invasive scenarios. The model** describes **cell proliferation and transport through elastic and compressible cell- ECM composite, including "go/grow" transitions determined by oxygen concentrations. Simulations generate invasive scenarios conditioned only by set of mechanical characteristics of cell-ECM composite typical for gliomas. Multicellular Tumor Spheroid (MTS) growing in soft matrix exhibits gentle cell density profile with elongated low density invasive zone, while rigid ECM conditions sharp edge of MTS and higher invasion speed. Moreover, the highest speed of tumor invasion is obtained at intermediate values of ECM stiffness, when a compact cell cluster is being formed instead of invasive zone as a second proliferative layer out of the core spheroid.**

Keywords — Multicellular Tumor Spheroid, Extracellular Matrix, glioma invasiveness, cell adhesion, ECM stiffness, in silico model

Introduction

It has been found that purely mechanical characteristics of glioma microenvironment play pivotal role in the invasion process [1-4]. The importance of cell- ECM interaction and the role of matrix density, elasticity and porosity have emerged as a particularly strong regulators of migration and proliferation rate of malignant brain tumor cells [2,3]. Typically, the invasive properties of glioma grow up at increase of ECM stiffness [2-4]. Some cases however present exceptions, e.g. demonstrating biphasic function of rigidity-dependence [3] or rigidity-insensitive high invasive capacity even in soft matrigel 3D cultures [4]. Therefore, in spite of general trends every case is unique regarding the behavior of each specific glioma cell line. The model used in the current study is not intended to simulate these unique cases. It does not take into account individual glioma cell line attributes and therefore cannot reproduce the variety of invasive patterns typically exhibited by different gloma cell lines [5]. Based on average glioma cell characteristics, the model generates "first approximation" scenarios of glioma invasion invariable for all glioma cell lines. These basic scenarios do not involve signaling regulation of glioma, but present a useful starting point for further analysis of every unique case. The *in silico* model describes visco-elastic dynamics of cells-ECM composite and is capable to simulate the interplay between elastic cell- ECM interaction, cell- cell adhesion and cell phenotype transition in the frame of "go/grow" approach [6]. The simulations aimed at studying glioma MTS growth as it depends on ECM stiffness.

Model assumptions

Spherically symmetric 1D simulation domain contains two components: cells and ECM as constituents of a composite system. ECM is considered to be purely elastic compressible scaffold, while the compressible cell aggregation exhibits viscoelastic properties and also effects of cell-cell adhesion. The total stress distribution in cell-ECM composite build the driving forces for cell transport. There are three cell phenotypes considered in the model following soft version of go/grow scenario [6]: proliferative cells that can proliferate and have low, non- zero, motility; hypoxic cells that have low, non- zero, proliferation rate and exhibit high motility, and necrotic cells, which are inert- do not exhibit motility and do not proliferate. Diffusion of oxygen towards the spheroid center from the spherical domain border and its local consumption determine the distribution of oxygen concentration along with phenotype transition rates. The model is formulated in more details in [7], including initial and boundary conditions. Model validation has been carried out on the results of two *in vitro* studies of glioma MTS growth [1,8] and exhibited rather good agreement with the experimental data, see [7].

Effect of matrix stiffness

A series of MTS growth calculations have been carried out at variation of ECM (collagen) stiffness from E_{ECM} = 500Pa to 5kPa. All calculations demonstrate a presence of Transition Layer (TL) on spheroid surface, where MTS mechanical characteristics undergo a sharp transition. The sharp raise of the total stress across the transition layer builds hydrostatic pressure gradients within this narrow area directed outwards across the TL. Therefore, TL represents a potential barrier for motile cells motion out of MTS core. The resulting pressure force steepens the front of cell density distribution, pushing back some of the motile cells traveling out of the core. The stress barrier of TL depends monotonically on ECM stiffness. So, that MTS growth in relatively soft collagens of E_{ECM} ~ 500Pa [8] exhibits typically thin TL and elongated invasive zone around the core spheroid with rarefied cell population in it [7,8]. The stiff matrix of E_{ECM} = 2.5kPa generates tenfold higher stresses within TL. This difference can be easily seen on peak magnitudes of the radial stress component σ_{Rtot}/σ_0 comparing TLs on Fig.1a with the data obtained in [7] for E_{ECM} = 500Pa. The counteraction of rigid matrix against expanding cell aggregation forms a sharp front at its edge without elongation of cell density profile into peripheral zone typical for soft matrixes, see [7]. The simulations show monotonic decrease of average spheroid size at increase of ECM stiffness along with growth of its expansion speed.

Formation of subtumors

Another interesting effect of ECM stiffness variations is a discrete form of invasive zone obtained at E_{ECM} = 2.5kPa. The increase of matrix stiffness reduces concentration of motile cells coming through TL barrier and ultimately leads to contraction of invasive zone into a second narrow proliferative layer outside of the core spheroid (Fig.1).

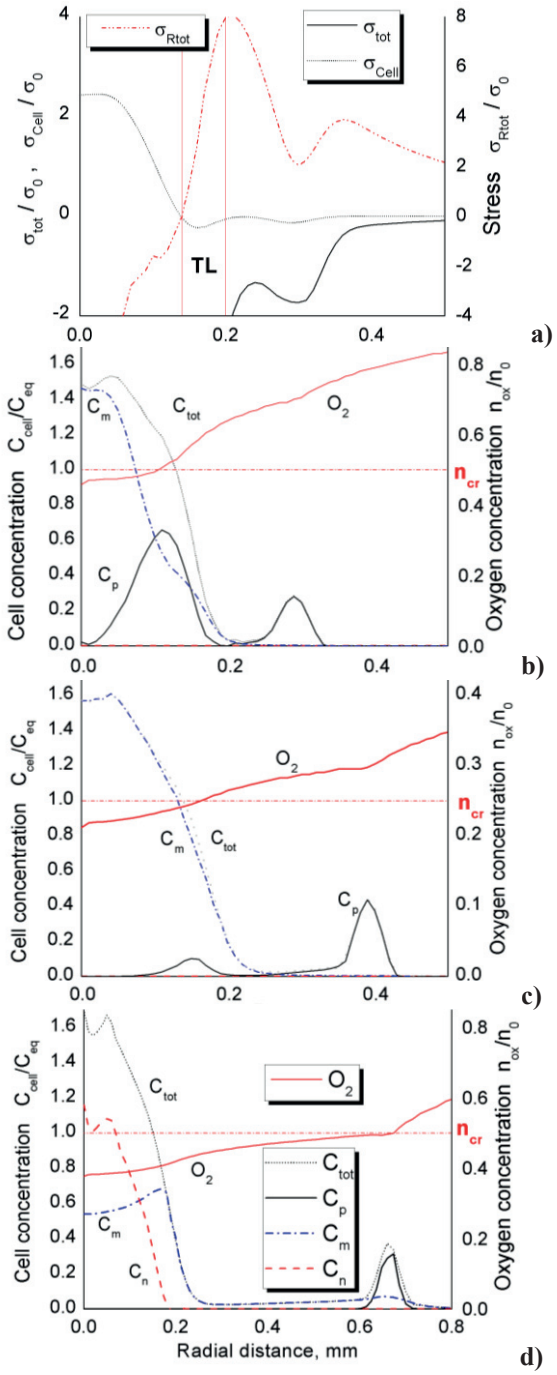


Fig. 1. Formation of Subtumor days 2, 3 and 5 of MTS growth at elevated matrix stiffness E_{ECM} = 2.5kPa: (a) stress distribution for day 2 including total stress in cells-ECM composite σ_{tot}, intercellular stress distribution σ_{cell} and radial component of the total stress $\sigma_{Rtot.}$; σ_0= 100Pa; (b,c,d) radial cell density distributions for proliferative cells C_p, hypoxic C_m, necrotic C_n and total cell density C_{tot}; C_{eq}= 3.5 x 10^8 cells/cm^3. Distribution of oxygen concentration is also shown on these graphs; the dotted line n_{cr} shows the threshold of O_2 level for proliferative–to-hypoxic cell phenotype transition, n_0 = 4 mg/L;

By day two shown on the figure, the major part of the main proliferative layer is located in the area of oxygen shortage and undergoes cell phenotype transition, while another small layer is formed out of the main one. Cell- cell adhesion and rigid matrix keep this layer compact instead of elongated invasive zone typical for soft matrixes. A radial component of the total stress in cell- ECM system exhibits strong gradient on the inner side of the newly formed layer (fig.1a) pushing it out of the core spheroid. By day three (fig.1c) over 90% of the main proliferative rim turns hypoxic, as the oxygen level inside the core part drops below the transition threshold (fig.1c). By day 5, the simulation shows that only the newly formed distant cluster of cells is out of the large hypoxic area, while the core spheroid turns hypoxic with growing necrotic zone in its center. Interestingly, the MTS does not form a distant proliferative layer at higher ECM stiffness E_{ECM} = 5kPa forming only a sharp edge of the core spheroid. The tumor invasion speed in our simulations exhibits biphasic function of E_{ECM} with its maximum around E_{ECM} ~ 2.5kPa, when the distant proliferative layers are being formed.

Conclusions

Three scenarios of glioma invasion at different ECM stiffness values, driven by mere mechanical principles of continuum and iso-strain dynamics of composite system have been shown in simulations. These scenarios may contribute into glioma variability *in vivo* being mediated within signaling feedback loop of self-consistent glioma regulation.

References

[1] V.D. Gordon, M.T. Valentine, M.L. Gardel et al., " Measuring the mechanical stress induced by an expanding multicellular tumor system: a case study", Experimental Cell Research 289, 2003, pp.58 – 66.

[2] B. Ananthanarayanan, Y. Kim, S. Kumar, "Elucidating the mechanobiology of malignant brain tumors using a brain matrix-mimetic hyaluronic acid hydrogel platform," Biomaterials, 32(31), 2011, pp.7913-23.

[3] T.A. Ulrich, E. de Juan Pardo, S. Kumar, "The mechanical rigidity of the extra-cellular matrix regulates the structure, motility, and proliferation of glioma cells", Cancer Res, 69(10), 2009, pp. 4167-74.

[4] T.J. Grundy, E. de Leon, K.R. Griffin et al., "Differential response of patient-derived primary glioblastoma cells to environmental stiffness", *Scientific Reports* 6, 23353, 2016.

[5] M.Vinci, C.Box, S.A. Eccles, "Three-Dimensional (3D) Tumor Spheroid Invasion Assay", Journal of Visualized Experiments (99), e52686, 2015.

[6] H. Hatzikirou, D. Basanta, M. Simon et al.,"'Go or grow': The key to the emergence of invasion in tumor progression?", Mathematical Medicine and Biology, 29(1), 2012, pp. 49-65

[7] V. Kalinin, "*In Silico* model of glioma MTS growth. Effects of compression and mechanical ECM remodeling", 'Mathematical methods and high-performance computing in the life sciences, biomedicine and biotechnology' at 12th Int.Multiconf. BGRS/SB-2020, 06-10 July 2020, Novosibirsk, Russia

[8] A.Stein, T. Demuth, D.Mobley et al.,, "A mathematical model of glioblastoma tumor spheroid invasion in a three-dimensional in vitro experiment," *Biophysical Journal*, 92(1), 2007, pp. 356–365.

DOI 10.18699/BGRS/SB-2020-099

MGSGenerator 1.5: software tool for reconstructing mathematical models of metabolic networks

F.V. Kazantsev
Kurchatov genomics center
ICG SB RAS
Novosibirsk, Russia
kazfdr@bionet.nsc.ru

S.A. Lashin
Kurchatov genomics center
ICG SB RAS
Novosibirsk, Russia,
Novosibirsk State University
Novosibirsk, Russia
lashin@bionet.nsc.ru

***Abstract* — There are lots of sources of information on metabolic networks but not so many sources dealing with existing mathematical models in spite of the fact that the modeling approach is an essential way for metabolic networks analysis. One of the ways to get ready-to-simulate frame models is the using of template approaches for mathematical model reconstruction, which take into account the nature of metabolic processes. However, building of more precise models typically requires reusing of existing models already adapted to experimental data. The aim of our research is to develop a tool that makes complete ready-to-simulate mathematical model for an input metabolic network. We present MGSGenerator 1.5 as the tool relied on API of existing model sources.**

Keywords* — *Mathematical model, E.coli, API, model generation, metabolic networks

Motivation and Aim

Mathematical modeling is widely used in microbiology investigations for metabolic networks analysis for the substance yield forecasting or adjustment appropriate environment conditions. *In silico* simulation allows us to check hundreds parameters for a model in a short period of time, depending on available computational environments.

Metabolic network reconstruction/analysis is the part of genome annotation steps in microbiology research. In this work we focus on the steps that go after. Metabolic network is the structural (graph) model that shows which substances interact. We are interested in the analysis of those interactions in time and in space. For the purpose, we should process structural model into dynamic model, which may be formalized in different forms. In this study we obtain Ordinary Differential Equation (ODE) model. One of the ways to get ready-to-simulate frame ODE models is the using of template approaches based on the standard equations for various types of metabolic processes. However, building of more precise ones typically requires reusing of existing models already adapted to experimental data.

The aim of our research is to develop a tool that makes complete ready-to-simulate mathematical model for an input metabolic network. The tool should be able to address API of existing model sources.

Results

We have advanced our previous tool MGSGenerator [1] to generate frame models. The novel version is 1.5. The tool processes the following steps:

- It takes structural model of metabolic network as an input. It decomposes network on subsystems and processes each of them.
- Each subsystem is compared with the set of existing models taken from various sources. The main source at present is MAMMOTh database [2] containing mathematical models of enzymatic reaction of *E. coli* adapted to experimental data.
- If appropriate structure model was not found, then the rule-based frame mathematical model generation process starts.
- To provide ready-to-simulate model, processes of substances sink and inflow are added.
- Finally, we obtain SBML model that then could be explored by Copasi tool (copasi.org) or any other simulation tools supporting SBML import.

We have tested the tool with metabolic networks of different size, from single metabolic pathway to the whole-genome metabolic network.

Acknowledgment

The study was supported by the Budget Project No. 0324-2019-0040-C-01 and Kurchatov Genomics Center of ICG SB RAS.

References

[1] F. Kazantsev, I. Akberdin, K. Bezmaternykh, V. Likhoshvai, (2009) “The tool for automatic generation of gene networks mathematical models”, VOGiS Herald, 13(1):163-169 [in Russian]

[2] F. Kazantsev, I. Akberdin, S. Lashin, N. Ree, V. Timonov, A. Ratushnyi, T. Khlebodarova, and V. Likhoshvai, (2018) “MAMMOTh: a new database for curated mathematical models of biomolecular systems”, Journal of Bioinformatics and Computational Biology, 16(01): 1740010, https://doi.org/10.1142/S02197200174001

DOI 10.18699/BGRS/SB-2020-100

Motility and fitness of microorganisms in dynamic aquatic ecosystems: a simulation study

A.I. Klimenko
Kurchatov Genomics Center,
ICG SB RAS
Novosibirsk, Russia
klimenko@bionet.nsc.ru

Yu.G. Matushkin
ICG SB RAS
Novosibirsk, Russia

S.A. Lashin
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
Novosibirsk State University,
Novosibirsk, Russia

***Abstract* — Motility is one of the key factors of adaptation in scarce marine environments inhabited by bacteria. This simulation study addresses the question of how an ability for adaptive migrations influences the evolutionary success of a population in various conditions. We examined the model of competition of motile and sessile populations of microorganisms in a confined aquatic environment supplied with a periodic batch nutrient source and assessed the fitness of both. We studied the effect of such factors as nutrient concentration in a batch, batch period, mortality term and penalty for migration and determined the conditions favouring "nomad" strategy of motile population and those favouring "settler" strategy of sessile one.**

Key words — motility; marine bacteria; agent-based modelling; ecological modelling

Motivation and Aim

In contrast to well-studied symbiotic gut microbiota living in abundant conditions, marine bacteria live in the world of scarce and ephemeral nutrition sources. Therefore, motility becomes one of the key factors of adaptation in these conditions. However, the advantage granted by motility comes with its cost and current estimates show that energy expenditure of various marine bacteria ranges from 2% to 50% of their total energy budgets [1] with different species varying in motility. While the processes of adaptation mediated by chemotaxis are relatively well-studied at the micro-scale level, there is a lack of understanding of how an ability for adaptive migrations influences the evolutionary success of a population.

Methods and Algorithms

We have used the Haploid Evolutionary Constructor 3D (HEC 3D) [2] software complex to build the model of competition of motile and sessile populations of microorganisms in a confined aquatic environment supplied with a periodic batch nutrient source. The HEC 3D allows creating multilayer ecological models of microbial populations inhabiting spatially structured environments. To describe a motile population capable of adaptive migrations via chemotaxis more adequately, the basic functionality of HEC 3D was extended to take into account the variability in migration traits such as energetic cost of migration.

Results

To estimate what conditions allow the motile population to make use of its advantage and dominate over the sessile one, we have investigated a model of a community consisting of two microbial populations in different model scenarios varying such parameters as nutrient concentration in a batch, batch period, mortality term and penalty for migration. It turned out that more dynamic and scarce environments favour dominance of motile populations whereas rich and stagnant environments tend to promote the dominance of sessile microorganisms. Moreover, dense-dependent quadratic mortality is more detrimental for motile population than linear one. We have found out that there is a turning point in the migration penalty value determining whether "nomad" strategy of motile population is adaptive or not. However, this value heavily depends on such conditions as nutrient availability. We have also shown that even without penalties for migration sessile population following the "settler" strategy can achieve dominant position in the ecosystem via decreasing local nutrient availability in the nutrient source's vicinity.

Conclusion

Thus, motile populations relying on local optimizing strategy tend to follow benign conditions and fail enduring stress associated with crossing the valleys of suboptimal nutrient availability. Though providing more adequate description, taking into account energetic cost of migration does not change the major trends in the competition between two types of microorganisms relying on different strategies in terms of motility.

Acknowledgment

This study was funded by the Budget Project No. 0324-2019-0040-C-01.

References

[1] R. Stocker and J. R. Seymour, "Ecology and physics of bacterial chemotaxis in the ocean.," *Microbiol. Mol. Biol. Rev.*, vol. 76, no. 4, pp. 792–812, 2012.

[2] A. I. Klimenko, Y. G. Matushkin, N. A. Kolchanov, and S. A. Lashin, "Modeling evolution of spatially distributed bacterial communities : a simulation with the haploid evolutionary constructor," *BMC Evol. Biol.*, vol. 15, no. Suppl 1, p. S3, 2015.

DOI 10.18699/BGRS/SB-2020-101

Candidate SNP markers of rheumatoid arthritis changing the affinity of TATA-binding protein for the human gene promoters expo disruptive selection of immunoactivative and immunosuppressive genenets that provoke and prevent this disorder, respectively, as if it could be a self-domestication syndrome

Natalya Klimova
Molecular Genetics Department
ICG SB RAS
Novosibirsk, Russia
klimova@bionet.nsc.ru

Irina Chadaeva
Systems Biology Department
ICG SB RAS
Novosibirsk, Russia
ichadaeva@bionet.nsc.ru

Evgeniya Oshchepkova
Systems Biology Department
ICG SB RAS, Novosibirsk, Russia
nzhenia@bionet.nsc.ru

Dmitry Oshchepkov
Systems Biology Department
ICG SB RAS
Novosibirsk, Russia
diman@bionet.nsc.ru

Mikhail Ponomarenko
Systems Biology Department
ICG SB RAS
Novosibirsk, Russia
pon@bionen.nsc.ru

Academician Vladimir Kozlov
RIFCI SB RAS
Novosibirsk, Russia
niiki01@online.nsk.su

Abstract — **Here we conducted a computational genome-wide study of the all known single-nucleotide polymorphism (SNP) of 70 bp proximal promoters of 67 human rheumatoid arthritis (RA)-related genes that displayed disruptive natural selections of immunoactivative or immunosuppressive genes raising or reducing risks of RA, respectively, as if it maybe a domestication syndrome. That is why, we confirmed it in vivo using the genome-wide transcriptome profiling (RNA-seq assay) of the differentially expressed genes (DEGs) within hypothalamus of adult male rats (Rattus norvegicus) of two unique outbred lines bred in aggressiveness and tameness as an animal model of human diseases (statistical significance $p_{adj} < 0.025$ at Pearson's χ^2 criterion with Bonferroni's correction).**

Keywords — rheumatoid arthritis, human, gene, promoter, TATA-binding protein (TBP) binding site (TATA box), single nucleotide polymorphism (SNP), candidate SNP marker, verification

Background

Rheumatoid polyarthritis (RA) is an autoimmune disease with autoantibodies (e.g., antibodies to citrullant antigens) and proinflammatory cytokines (e.g., TNF-α, IL-6) that are involved in the induction of chronic synovitis, bone erosion, followed by deformity [1]. Currently, it is commonly accepted that immunopathogenesis is mostly contributed by the mechanisms of the breakdown of immune tolerance to its own antigens that is characterized by an increase in the activity of T-effector cells, causing RA symptoms [2]. On the contrary, a low activity of regulatory cells (e.g., regulatory T-cells (Treg) and myeloid suppressor cells) is met in RA [3]. That is why, to say that it is the immunosuppressor cell deficit that is the central feature in RA pathogenesis [4]. It is thought that lifestyle and living conditions define a half of the RA risks and genetic susceptibility to RA do for another half [5]. Additionally, immune cells of adaptive memory in a given individual keep information on all diseases he / she survived that increase his / her resistance to them in future [6]. That is why here using our SNP_TATA_Comparator [7], we predicted RA-related candidate SNP markers, found regularities in their genome-wide frequencies, and tested them in vivo using an animal (rat) model of the human health.

Materials and Methods

We extracted 1834 SNPs of 70 bp promoters of 67 RA-related genes using UCSC Genome Browser [8], dbSNP [9] and GRCh38/hg38 assembly of the human reference genome [10] as well as treated them using our SNP_TATA_Comparator [7] as discribed elsewhere [7]. We did the experiment *in vivo* on 6 adult male rats (*Rattus norvegicus*) of 2 outbred lines bred (over 90 generations) for aggressive or tame behavior under standard conditions of the Conventional Animal Facility of ICG SB RAS (Novosibirsk, Russia), as described elsewhere [11]. All rats came from different litters and were unrelated, each weighing 250–270 g, 4 months old, marked by "-3" or "+3" on the rank scale from "-4" to "+4" (i.e., maximum aggressiveness and tameness). After their euthanasia, the biosamples of hypothalamus was prepared and stored at -70 ^{O}C until use as approved by Interinstitutional Bioethics Board at ICG SB RAS in line with Directive 2010/63/EU of the European Parliament and of the Council of September 22, 2010, on protections of animals used for scientific needs, as described in [12].With 100 mg fractions taken from each above-mentioned hypothalamic biosample from three aggressive and three manual rats, using an Illumina NextSeq 550 and NextSeq® 500/550 High Output Kit v2 (75 cycles) cassettes, we sequenced through direct reading 24 barcoded RNA-Seq libraries of 75 nt fragments that lead the total volume of 400 million reads. We processed them by our conveyor, as: FastQC [13] → Trimmomatic [14] → TopHat2 [15] → SAMtools [16] → HTSeq [17] → DeSeq2 [18]. This resulted in relative expression of genes mapped into the rat reference genome rn6 in hypothalamus of tame and aggressive rats. We verified them with our measurements of them made by quantitative polymerase chain reaction (qPCR) in our work [19]. Finally, among 67 genes studied, we found those whose expression were significantly different between aggressive and tame rats, evaluated the manifestations of them in terms of susceptibility to RA according to the clinical markers of RA

and, estimated the statistical significance of them with Bonferroni correction.

Results and Discussions

First, using the only clinical SNP marker s2276109 of RA [20], we tested adequacy of candidate SNP markers of this disease predicted by SNP_TATA_Comparator [9] Then, for 646 SNPs in question within 27 human genes clinically associated with disorders comorbid to RA, we found 83 and 73 candidate SNP markers, which can elevate (e.g., rs1185314734) or reduce (e.g., rs1190659847) expression of these genes, respectively, as well as 91 and 54 ones for the worsened (e.g., rs1189849606) or relieved (e.g., rs549858786) RA. It means the genome-wide neutral drift [21-23] towards susceptibility to RA in line with the common view that diversity of adaptive memory immune cells of a given individual grow with each inflammation he/she survived that improves his/her resistance to them, but this immune memory growth raises risks of impaired autoimmune tolerance (e.g., RA risks). Thus, by the same way we confirmed this finding in the case of nine genes, which are most GWAS-associated with RA [24], and, independently, in the case of seven immunoactivative genes. Cntrary, when we dealt with 24 immunosuppressive genes, we found 64 and 111 candidate SNP markers aggravating and relieving RA, respectively. This difference between immunoactivative and immunosuppressive genes means disruptive natural selection as if RA could be a domestication syndrome. That is why, using RNA-seq assay we compared aggressive and tame rats, both primary experimental and resulting computational data of which are available for readers, as: “ftp://ftp.RNAseq.cytogen.ru:h3&3r” using “login/password: Ht3&3Rats”. It turned out that 10 among 67 genes analyzed here are differentially expressed in hypothalamus of adult male aggressive versus domesticated rats at partial significance threshold $p_{RNAseq} < 0.05$. For four genes *Apoa1, Hbb, Hbb-b1*, and *Sod1*, it is known that over- and underexpression of them worsen RA, while six rest genes *Ccr7, Ebi3, Il1b, Il2rb, Il3ra*, and *Il25* overexpress in tame rats in line with the physiological markers of the worsen RA. Within Pearson's χ^2-criterion, it is susceptibility to RA in the tame rats ($p_{\chi 2} < 0.005$). So, $p_{adj}=N*p_{\chi 2}*p_{RNAseq}=67*0.005*0.05 <0.025$, as Bonferroni's correction.

Conclusions

We found in silico disruptive selections of immunoactivative or immunosuppressive genes corresponding high or low risks of RA, as if it could be a domestication syndrome, which we confirmed at animal model of human health..

ACKNOWLEDGMENTS

Manuscript writing was supported by project #0324-2019-0040-C-01 (for MP). The experiment *in vivo* was supported by project #0324-2019-0042-C-01 (for NK, IC). The data analysis and biological interpretation were supported by the Russian Federal Science & Technology Program for the Development of Genetic Technologies (for DO, EO).

REFERENCES

[1] J. Smolen, D. Aletaha, Nat Rev Rheumatol, vol. 11, pp. 276-289, 2015.
[2] J. Sarkander et al., Clin Transl Immunology, vol. 5, pp. e120, . 2016.
[3] O. Alsaed et al., Eur J Case Rep Intern Med, vol. 5, pp. 000895, 2018.
[4] C. Lawson et al., Rheumatology (Oxford), vol. 45, pp. 1210-1217, 2006.
[5] N. Nair et al., Pharmacogenomics, vol. 18, pp. 1323-1332, . 2017.
[6] J. Sarkander, et al., Clin Transl Immunology, vol 5, pp. e120, . 2016.
[7] M. Ponomarenko et al.. Biomed. Res. Int. vol. 2015. pp. 35983004625. 2015.
[8] M. Haeussler et al., Bioinformatics. vol. 31. pp. 764-766. 2015..
[9] S. Sherry et al., Nucleic Acids Res. vol. 29, pp. 308-311, 2001.
[10] D. Zerbino et al., Genome Biol. vol. 16: pp. 56. 2015.
[11] D. Belyaev, P. Borodin, Biol Zent Bl, vol. 100, pp. 705–714, 1982.
[12] I. Plyusnina, I. Oskina, Physiol Behav, vol. 61, pp. 381–385, 1997.
[13] K. Kroll et al., Cancer Inform, vol. 13(Suppl 3), pp. 7-14, 2014.
[14] A. Bolger et al., Bioinformatics, vol. 30, pp. 2114-2120, 2014.
[15] D. Kim et al., Genome Biol, vol. 14, pp. R36, 2013.
[16] Li H. Bioinformatics, vol. 27, pp. 1157-1158, 2011.
[17] S. Anders et al., Bioinformatics, vol. 31, pp. 166-169, 2015.
[18] M. Love et al., Genome Biol, vol. 15, pp. 550, 2014.
[19] D. Oshchepkov et al., Front Genet,vol.. 10, pp. 1267, 2019.
[20] G. Hunninghake et al., N Engl J Med, vol. 361, pp. 2599-2608, . 2009
[21] J. Haldane. J Genet. vol. 55. pp. 511–524. 1957.
[22] M. Kimura, Nature. vol. 217. pp. 624-626. 1968.
[23] 1000 Genomes Project Consortium, “Nature. vol. 491. pp. 56-65. 2012.
[24] J. Karami et al., Gene, vol. 702, pp. 8-16, 2019.

DOI 10.18699/BGRS/SB-2020-102

Searching for alternatively splicing group II introns

Nikolay Kobalo
The Laboratory of Computational Problems of Geophysics
The Institute of Computational Mathematics and Mathematical Geophysics
Novosibirsk, Russia
rerf2010rerf@yandex.ru

Denis Vorobyev
INSERM U981
Gustave Roussy Cancer Center
Villejuif, France
titov@bionet.nsc.ru

Igor Titov
The Laboratory of Molecular-Genetics Systems
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
titov@bionet.nsc.ru

***Abstract* — Group II introns are mobile elements present in bacteria, as well as in the eukaryotic organelles. This type of intron is probably the ancestor of nuclear introns. Group II introns have a conservative secondary structure, which determines their ability to self-splice. It is known that some group II introns have the ability to alternative splicing [1]. The possibility of alternative splicing of group II introns is explained by the fact that they can have several different variants of the secondary structure, each of which leads to its own splicing reaction. Based on the introns of group II with the secondary structure described in the literature, we have constructed a generalized model of the structure of introns of this type. In addition, we searched for introns corresponding to our model in the RFAM database and found introns with the possibility of alternative splicing. For the found introns, we performed an analysis of coadaptive nucleotide substitutions in the secondary structure. We also analyzed the effect of alternative splicing on proteins that are the product of the corresponding genes.**

Keywords* — *intron group II, secondary structure, alternate splicing

Materials and Methods

- To build a model of the secondary structure of group II introns, we used a database of known intron structures [2]
- We searched for introns with the possibility of alternative splicing in an open database of bacterial and eukaryotic introns of group II RFAM [3].
- To search for a given secondary structure in genomic sequences, we used the Rscan program [4].

Secondary Structure Models

The secondary structure of group II introns consists of 5 structural elements called domains. The largest of them is the first domain. At the first stage, for each of the intron domains, except 4, we built several models of the secondary structure. Each model generalized its domain for a group of introns with a similar structure. In addition to the secondary structure itself, described by the positions and lengths of its stems and loops, our model also included:

- Consensus sequences for splicing sites.
- The presence of tertiary interactions of introns. Among them are ebs1 / ibs1, ebs2 / ebs2, α-α`, β-β`, κ, ε.

In order to evaluate the quality of our models, we calculated the distribution of z-score of the energy for the secondary structure on a random sample of nucleotide sequences. Random sequences had the same dinucleotide composition as real introns. Using the obtained distribution, we set the threshold for the energy of the secondary structure: z-score <= –1.48.

We used the constructed models to restore the secondary structure introns from the RFAM database. To do this, we looked for sequences that satisfy our constraints on structures, including constraints on tertiary interactions and the z-score threshold. We also analyzed the protein sequences of the fourth domain of the detected introns for the presence of standard RT motifs in them [5]. The group II introns discovered in this way and their secondary structures were described in [6]. In total, we restored the complete secondary structure of 394 eukaryotic group II introns. Among them, 3 introns of subtype IIB, and the rest IIA. 18 found introns belong to the nuclear genome, 10 to the mitochondrial, 127 to the chloroplast genome and 196 to the plastid genome.

Search for alternative splicing

Then we used the found boundaries of all domains to detect introns containing several valid splicing variations. Such introns can potentially have the ability to alternative splicing. On the identified candidates, we carried out an analysis of coadaptive nucleotide substitutions in the secondary structure in order to establish the most probable path of their evolution. In addition, for introns located close to each other in the genome, we analyzed their interactions and mutual evolution.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of LipsBonnie A. McNeil, Dawn M. Simon and Steven Zimmerly. Alternative splicing of a group II intron in a surface layer protein gene in Clostridium tetani. Nucleic Acids Research, 2014, Vol. 42, No. 3.

[2] Dai, L., Toor, N., Olson, R., Keeping, A., and Zimmerly, S. (2003). Database for mobile group II introns. Nucleic Acids Res. 31: 424-426.

[3] Kalvari I. et al. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 46(D1):D335-D342. DOI: 10.24411/9999-017A-2019-10001 10.1093/nar/gkx1038 8–73.

[4] http://www.softberry.com/freedownloadhelp/rna/rscan/rscan.all.html#RScanDescription

[5] Zimmerly, S., Wu, L. (2014). An unexplored diversity of reverse transcriptases in bacteria. Microbiol. Spectr. 3 (2), MDNA3–0058-2014. doi: 10.1128/microbiolspec.MDNA3-0058-2014

[6] Titov I, Kobalo N, Vorobyev D and Kulikov A (2019) A Bioinformatic Method For Identifying Group II Introns In Organella Genomes. *Front. Genet.* 10:1135. doi: 10.3389/fgene.2019.01135

DOI 10.18699/BGRS/SB-2020-103

BioUML – universal platform for analyses of biomedical data

Fedor A. Kolpakov
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
fedor@biosoft.ru

Ivan S. Evshin
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
ivan@biosoft.ru

Ilya N. Kiselev
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
axec@biosoft.ru

Anna S. Ryabova
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
anna@biosoft.ru

Yury V. Kondrakhin
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
yvkondrat@mail.ru

Sergey S. Pintus
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
sspintus@biosoft.ru

Elena O. Kutumova
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
elena.kutumova@biosoft.ru

Nikita V. Mandrik
Institute of Computational Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
manikitos@biosoft.ru

Alexander E. Kel
BIOSOFT.RU, LLC
Novosibirsk, Russia
geneXplain GmbH,
Wolfebuttel, Germany
alexander.kel@biosoft.ru

***Abstract* — BioUML (homepage: http://www.biouml.org, main public server: https://ict.biouml.org) is a web-based integrated environment (platform) for systems biology and analysis of biomedical data generated by omics technologies. Here we are presenting further extensions of BioUML platform that make it universal platform for analysis of wide range of biomedical data: 1) integration with Docker technology that allows easily pack and run existing software and databases; 2) integration of noVNC web client to work with software with rich graphic interface via web; 3) new integration with Galaxy and tool shed that allows reuse wide range of existing software for data analysis; 4) integration with Jupyter hub for interactive data analysis; 5) support of Common Workflow Language (CWL).**

Keywords — BioUML, Docker, Galaxy, Jupyter notebooks, CWL, data analysis

Introduction

BioUML is a web-based integrated environment (platform) for systems biology and the analysis of biomedical data generated by omics technologies[1]. BioUML spans a comprehensive range of capabilities, including access to biological databases, powerful tools for systems biology (visual modelling, simulation, parameters fitting and analyses), a genome browser, scripting (R, JavaScript) and a workflow engine. Due to integration with the Galaxy platform and R/Bioconductor, BioUML provides powerful possibilities for the analyses of omics data. The plug-in-based architecture allows the user to add new functionalities using plug-ins. To support collaborative work on scientific projects, there is a central authentication and authorization system (https://bio-store.org). The diagram editor enables several remote users to simultaneously edit diagrams.

Results

Figure 1 demonstrates the suggested architecture of universal platform for analysis of wide range of biomedical data:

1) Uniformly packed and described computer programs for data analysis and visualization:
- Docker technology is used to easily pack and run existing software and databases;
- Docker registry is used as repository packed software images;
- Common Workflow Language (CWL) and Galaxy formats are used to describe computer programs and parameters to start them.

2) Core - BioUML platform [1]. It also provides communication with integrated tools (Galaxy, Jupyter hub, R, Docker, noVNC, etc.).

3) IT infrastructure where the platform was installed.

4) external resources can be used to store users data. They can be connected to the platform via specialised API. .

Acknowledgment

This study was supported by the Russian Science Foundation, grant No. 19-14-00295.

References

[1] Kolpakov F, Akberdin I, Kashapov T, Kiselev L, Kolmykov S, Kondrakhin Y, Kutumova E, Mandrik N, Pintus S, Ryabova A, Sharipov R. BioUML: an integrated environment for systems biology and collaborative analysis of biomedical data. Nucleic acids research. 2019 Jul 2;47(W1):W225-33.

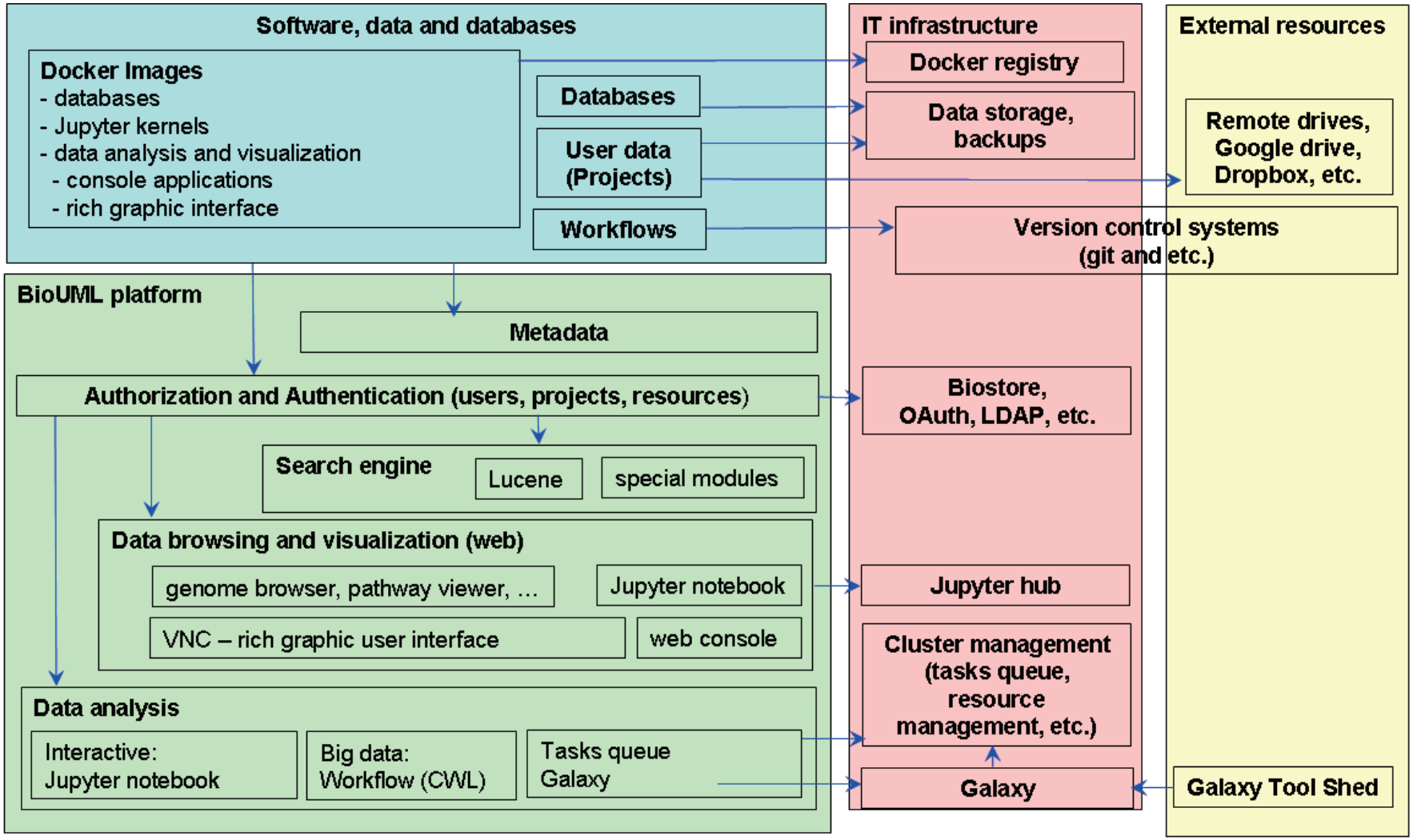


Fig. 1. Structure of universal platform for analyses of biomedical data.

DOI 10.18699/BGRS/SB-2020-104

Computer reconstruction of the ecological structure of intestinal microbiota communities based on high-throughput sequencing data

Andrew Kopochev
Kurchatov Genomics Center,
ICG SB RAS
Novosibirsk, Russia
andrew.kropochev@gmail.com

Alexandra Klimenko
Kurchatov Genomics Center,
ICG SB RAS
Novosibirsk, Russia
klimenko@bionet.nsc.ru

S.A. Lashin
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
Novosibirsk State University, NSU
Novosibirsk, Russia

Abstract — **Human gut microbiota is essential for human health. Under the recent avalanche of metagenomic high-throughput sequencing data methods for its bioinformatic analysis and ecological reconstruction gain particular relevance. In this study, we employ the trait-based ecology approach to create a method of computer reconstruction of the ecological structure of intestinal microbiota communities based on high-throughput sequencing data and apply it to the synthetic community comprised of the key representatives of the human gut microbiota. Using this structure allowed us to evaluate the abundance of respective functional groups and infer a knowledge about the processes associated with them. The developed method can be applied to analyze intestinal microbial communities of other related species.**

Keywords — human gut microbiota, trait-based ecology, bioinformatics

Motivation and Aim

The microbiota of the human gut is a complex microbial community that plays an important role in maintaining human health. Despite the fact that the microbiota of the human gut is being studied thoroughly, at present there is no comprehensive picture of the ecological interactions between its major representatives.

Since natural microbial communities consist of a large number of species, reconstruction of their ecological structures is a challenging problem. Therefore, trait-based ecology approach could be helpful as it reduces the complexity of the system putting emphasis on ecological function rather than taxonomy. Application of this approach to the study of specific ecosystems requires a detailed understanding of the basic mass and energy flows in the system and the role of each functional group in these processes.

In our study, we have applied trait-based ecology approach for computer reconstruction of the ecological structure of intestinal microbiota communities based on high-throughput sequencing data.

Methods

For our study, we used available data on a synthetic microbial community, which served as a model for a core human gut microbiota. It consists of 14 species that demostrate the basic toolbox of metabolic capacities determining major processes in the human gut [1]. The ecological structure has been reconstructed using expert analysis of the literature and bioinformatic analysis filling the gaps in knowledge obtained from the literature.

De novo transcripts were assembled using the Trinity platform [2] for working with RNA-Seq data in non-model organisms. The default settings for unpaired reads were used.

Trait-based ecology approach was applied to identify the functional groups that can be used to classify the species in this community taking into account major processes in the human gut performed by its microbiota.

Results

We have carried out an expert reconstruction of the ecological structure of the microbial community comprised of key representatives of the human gut microbiota. We distinguish such functional groups as *universal polysaccharide utilizers, butyrate producers, acetogens, sulfate reducers and mucin degraders*. For each functional group, a set of genetic markers was determined identifying representatives of this group in the analyzed data sets from the synthetic community of the human intestine [1]. We assembled a *de novo* metatranscriptome and used these markers to evaluate abundance of the functional groups in the community.

Acknowledgment

The study was supported by the Budget Project No. 0324-2019-0040-C-01 and Kurchatov Genomics Center of ICG SB RAS.

References

[1] Desai et al. (2016) A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell. 167: 1339-1353

[2] Haas B.J. et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8: 1494–1512.

DOI 10.18699/BGRS/SB-2020-105

Mathematical modeling and scenarios of COVID-19 epidemy in Moscow and Novosibirsk region based on SEIR-HCD model

Olga Krivorotko
NSU, Novosibirsk, Russia
krivorotko.olya@mail.ru

Nikolay Zyatkov
ICM&MG SB RAS, Novosibirsk, Russia
nikolay.zyatkov@gmail.com

Daria Andornaya
Baker Hughes, NSU, Novosibirsk, Russia
ermolenko.dasha@mail.ru

Sergey Kabanikhin
NSU, Novosibirsk, Russia
ksi52@mail.ru

***Abstract* — In December 2019, there was an outbreak of pneumonia of unknown cause in Wuhan, Hubei province in China. The first comfirmed case in Russian Federation was in March 2020. Mathematical modelling is one of the effective ways to prediction of the epidemiology situation in region. In paper we divide all population in seven group and apply SEIR-type model for COVID-19 epidemy. It is necessary to identify unknown parameters of mathematical model for more precise prediction of epidemic situation in local region as well as in whole country. The identification problem is formulated as a problem of minimization of misfit function that solved by differential evolution algorithm. Scenarios of COVID-19 propagation in Moscow city and Novosibirsk region are proposed considering quarantine restrictions, tested cases, confirmed and death cases, isolation rate.**

Keywords* — *COVID-19, mathematical modeling, SEIR-HCD model, inverse problem, identifiability, sensitivity analysis, prediction, optimization, differential evolution

Motivation and aim

Motivation

A new outbreak of pneumonia in Russian Federation caused by coronavirus COVID-19 is happened in March 2020. It is necessary to model scenarios of disease propagation depend on restrictive measures, isolation rate, social distancing, density of population and economy factors.

Aim

The epidemic situation is different in regions of Russian Federation (Moscow, Novosibirsk region, Far East, etc.). Earlier we have developed a mathematical model of HIV and tuberculosis co-infections [1] based on SEIR-type structure that predicted an epidemic situation of HIV and tuberculosis in endemic regions of Russian Federation. The aim of our research is to create and investigate mathematical model of propagation and prediction of COVID-19 in regions of Russian Federation based on identification of unknown parameters of model using additional measurements about confirmed, critical and death cases, isolation rate and passenger flows between cities.

Methods

The SEIR-HCD mathematical model that divide population in fixed region in seven groups with conversion factors and described by system of nonlinear ordinary differential equations [2] is applied for COVID-19 propagation. The parameter identification problem for model SEIR-HCD using time-dependent data of confirmed, critical and death cases, Yandex isolation rate (inverse problem) is reduced to a minimization problem of misfit function. Firstly, sensitivity-based identifiability analysis is performed to fix fewer sensitive parameters to measurements and avoid ill-posedness [3]. Secondly, differential evolution approach is applied to minimize misfit function and find the appropriate set of parameters with further prediction.

Results

The SEIR-HCD mathematical models was investigated for Moscow city and Novosibirsk region using additional information about confirmed, critical and death cases, Yandex isolation rate from March 23, 2020. It is shown that probability of shutdown the ventilator of critical cases, infection parameter between infected and susceptible populations as well as recovered rate should be controlled by the regularization of algorithm of inverse problem solution. The prediction of the peak of the epidemic in Moscow that was in May 7 has an error in 2 days and 174 confirmed cases using different time-data: 23.03-21.04 and 23.03-01.05. Moreover, the error in prediction till June 20, 2020, is less then 7%. The model scenario of COVID-19 epidemic in Novosibirsk region shows the growth of confirmed and hospitalized cases till the beginning of July 2020.

Acknowledgment

Supported by the Russian Science Foundation (project 18-71-10044) and by the President Scholarship of Russian Federation (agreement 075-15-2019-1078 (MK-814.2019.1)).

References

[1] Kabanikhin S., KrivorotkoO., Kashtanova V. (2018) A combined numerical algorithm for reconstructing the mathematical model for tuberculosis transmission with control programs. Journal of Inverse and Ill-Posed Problems. 26(1): 121-131.

[2] Unlu E. et al. (2020) Epidemic analysis of COVID-19 outbreak and counter-measures in France. MedRxiv 2020.04.27.20079962.

[3] Krivorotko O.I., Andornaya D.V., Kabanikhin S.I. (2020) Sensitivity analysis and practical identifiability of some mathematical models in biology. Journal of Applied and Industrial Mathematics. 14(1): 115-130.

DOI 10.18699/BGRS/SB-2020-106

Genome-scale metabolic modeling of 2,3-butanediol production by *Geobacillus icigianus*

Mikhail Kulyashov
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS,
Novosibirsk, Russia
Novosibirsk State University,
Novosibirsk, Russia
m.kulyashov@mail.ru

Ilya R. Akberdin
BIOSOFT.RU, LLC
FRC Institute of Cytology and Genetics SB RAS,
Novosibirsk, Russia
Novosibirsk State University
Novosibirsk, Russia
akberdinir@gmail.com

***Abstract* — 2,3-butanediol (2,3-BTD) is an important substrate for chemical production and at the same time is highly promising bacterial-based platform substances. *Geobacillus icigianus* is a strain of thermophilic genus *Geobacillus*, which is currently considered as the potential bacterial chassis that can be used in biotechnology. A genome-scale metabolic model of the bacteria has been built using a computational pipeline for autogeneration with consequent manual curation. The current version of the model comprises 1678 reactions, 1590 metabolites and 1316 genes and it is the largest known model for genus *Geobacillus*. In this work we demonstrate that *Geobacillus icigianus* can be potentially used for the production of 2,3-butanediol from different carbon sources, one of which is glycerol – a byproduct of chemical production. Furthermore, this model can be used as a theoretical platform to gain insight into the metabolism of the thermophilic bacteria and to predict more favor pathways for genetic modifications of *Geobacillus icigianus* strain in biotechnological goals.**

Keywords* — *Genome-scale modeling, Geobacillus icigianus, 2,3-butanediol, thermophilic organisms, biotechnology

Introduction

Due to the gradual depletion of fossil fuel sources, rising oil prices, and the aggravating environmental situation, which lead to the tight control of the chemical industry, a question of creating biological factories for the production of chemical substances is becoming crucial. One of the important substances in the chemical industry is 2,3-butanediol (butadiene glycol-2,3). The potential of bacteria in production of 2,3-butanediol was shown in the early 20th century [1]. Moreover, the production of 2,3-BTD was shown for thermophilic bacteria of genus *Geobacillus* [2]. Based on published data we decided to identify this capability by *Geobacillus icigianus* - new strain of thermophilic bacteria [3]. To solve this issue we have been used a genome scale metabolic modeling approach, which could give an opportunity to investigate a bacterial metabolism [4].

Materials and Methods

Reconstruction of the mathematical model

The complete genome of *Geobacillus icigianus* strain was extracted from NCBI Refseq database [5]. Initially we re-annotated this genome by means of RAST [6]. The procedure was conducted using standard RAST annotation scheme. At the next step we used automatic generation pipeline which is presented in web-service Kbase [7] to generate a genome-scale metabolic model. Module from Kbase for building of genome-scale models was harnessed with standard parameters including gap-filling algorithm. Biomass equation was generated automatically and stoichiometrically equivalent to biomass equation of *Bacillus subtilis*. A quality of the draft model was checked out using Memote web service [8], which demonstrated that consistency of the developed model is 92%. To improve the model consistency, we manually curated the draft model in order to modify SEED [9] reaction names and ID's on their equals from BIGG database [10]. We wrote a script on Python 3.6 using Cobrapy package and replaced all IDs which compose information about BIGG ID. Afterwards, we added boundary conditions for drain reactions which are necessary to describe wildtype growth of the strain. D-Glucose (glucose) was used as the first carbon source for the growth and lower bound was set up equal to -17 mmol/gDCWl/h^{-1} according to the published data for closely related species [11]. The modified GSM model of the strain was uploaded and analyzed via Optflux tool [12]. Flux balance analysis conducted in this tool using pFBA approach showed that growth rate on glucose, as a single carbon source, compose 0.5 mmol/gDCWl/h^{-1}.

To consider the growth of the bacteria on other carbon sources we added exchange reactions for metabolites growth ability on which was shown in the published data: *glycerol, L-arabinose and D-xylose* [3]. Metabolic analysis of the model revealed that to consider the growth on xylose and arabinose some metabolic reactions, which were not presented in the model, are required. Analysis of the metabolic pathways in closely related species of genus *Geobacillus* and metabolic pathways of *Bacillus subtilis* in SEED, KEGG [13] and BIGG databases resulted in an addition of the next set of reactions for growth on xylose: D-Xylose exchange, D-xylose reversible transport, Xylulokinase (EC: 2.7.1.17) reactions; for growth on arabinose: L-ribulokinase (EC:2.7.1.16) and L-ribulose-phosphate 4-epimerase (EC: 5.1.3.4) reactions. The presence of all proteins encoded in the *Geobacillus icigianus* genome for this set of reactions was checked out using Blastp web-service. Flux balance analysis for growth of the strain on abovementioned carbon sources was perfomed in Optflux tool using pFBA [14] approach too. Visualization of pFBA outcomes was carried out through Escher web-service [15].

Model modification for 2,3-butanediol production

To identify a list of reactions required for 2,3-butanediol production we conducted a literature analysis. It turned out that metabolic reactions to synthesize the substance in the bacteria comprise: acetolactate synthase (EC:2.2.1.6), acetolactate decarboxylase (EC: 4.1.1.5) and (R,R)-butanediol dehydrogenase (EC:1.1.1.4). Acetolactate synthase was originally presented in the model, while other metabolic reactions were added using Cobrapy. Moreover, we needed to add (R,R)-butanediol transport и (R,R)-

butanediol exchange reactions. The final version of the GSM model was uploaded into Memote web-service which demonstrated that the model consistency did not change compared to the draft model and equals to 92%.

Model analysis for 2,3-butanediol production optimization

To identify genetic modifications in order to increase 2,3-butanediol production and simultaneously do not significantly reduce biomass value we used evolutionary optimization approach via Optflux. To conduct this type of the analysis we selected 5 basic simulation algorithms: pFBA, MiMBL [16], MOMA [17], LMOMA [18] and ROOM [19]. All algorithms were started with 5000 maximum evolutionary functions and with maximum number modifications equal to 2. Optimization algorithm was chosen considering specific options of simulation methods. LMOMA, MOMA and pFBA simulation methods were run with Strengh Pareto Evolutionary Algorithm for reaction under/over expression. MiMBL and ROOM methods were initialized with Strengh Pareto Evolutionary Algorithm for the gene under/over expression.

Results

Thus, we generated the first GSM model for *G.icigianus* using Kbase web-service and final version of the model includes 1678 reactions, 1590 metabolites and 1316 genes. Flux balance analysis of the model showed that flux distribution in *G.icigianus* differs from *B.subtilis*, a model microorganism metabolic pathways and biomass equation of which were employed as a template for our model at the building stage. For instance, there are changes in electron donor/acceptor reactions in the citric acid cycle (TCA); missed reaction PGL (EC: 3.1.1.31), which is catalyzed by not thermostable enzyme, and the model describes the metabolic feature. Moreover, simulations of the GSM model demonstrated that an electron donor for oxidative phosphorylation depends on the carbon source. Growth on all substrates excluding glycerol showed a presence of lactate and succinate as excreted compounds, and it is consistent with experimental data [3]. The developed model predicts that glucose and glycerol (0.5 mmol/gDCWl/h^{-1}) are the most effective substrates for the growth, but growth on glycerol needs more oxygen. Furthermore, we matched built GSM model with published one for *Bacillus subtilis* (iYO844) [19]. As a result, the growth rate of *G.icigianus* is higher than one of *B.subtilis* for analogous substrate uptake rate which corresponds to 1% glucose concentration in the media according to the published data [20].

Model optimization for 2,3 butanediol production

Optimization analysis of the model in Optflux tool indicated that all carbon substrates can be used for 2,3-BTD production. It worth to note that MOMA and ROOM algorithms were not able to calculate the model. However, LMOMA, MiMBL, pFBA algorithms predicted glycerol as the most promising substrate for 2,3-BTD production by *G.icigianus*. All reaction modifications to improve the 2,3-BTD production predicted by the algorithm somehow affect two metabolic aspects: 1) modifications of the TCA cycle that lead the reduction of succinate production and 2) modifications that result in anaerobic or microaerobic growth conditions. It is interesting that all predicted ways of the metabolic modifications were earlier experimentally verified for another biotechnological species [1] and for closely related *Bacillus subtilis* [21].

References

[1] Celińska, E., and W. Grajek. "Biotechnological production of 2, 3-butanediol—current state and prospects." Biotechnology advances 27.6 (2009).

[2] Xiao, Z., Wang, X., Huang, Y. et al. “Thermophilic fermentation of acetoin and 2,3-butanediol by a novel Geobacillus strain”. Biotechnol Biofuels 5, 88 (2012).

[3] Bryanskaya, A. V., Rozanov, A. S., Slynko, N. M., Shekhovtsov, S. V., & Peltek, S. E. Geobacillus icigianus sp. nov., a thermophilic bacterium isolated from a hot spring. International journal of systematic and evolutionary microbiology, 65(3), 864-869, (2015).

[4] Simeonidis, E., Price, N.D. “Genome-scale modeling for metabolic engineering”. J Ind Microbiol Biotechnol 42, 327–338 (2015).

[5] Bryanskaya A. V. et al. “Draft genome sequence of *Geobacillus icigianus* strain G1w1T isolated from hot springs in the Valley of Geysers, Kamchatka (Russian Federation)”, Genome Announc.. – T. 2. – №. 5. – С. e01098-14 (2014).

[6] Aziz, R.K., Bartels, D., Best, A.A. et al. “The RAST Server: Rapid Annotations using Subsystems Technology”, BMC Genomics 9, 75 (2008).

[7] Arkin, A., Cottingham, R., Henry, C. et al. “KBase: The United States Department of Energy Systems Biology Knowledgebase”, Nat Biotechnol 36, 566–569 (2018).

[8] Lieven, C., Beber, M.E., Olivier, B.G. et al. “MEMOTE for standardized genome-scale metabolic model testing”, Nat Biotechnol 38, 272–276 (2020).

[9] Devoid S., Overbeek R., DeJongh M., Vonstein V., Best A.A., Henry C. (2013) “Automated Genome Annotation and Metabolic Model Reconstruction in the SEED and Model SEED”. In: Alper H. (eds) Systems Metabolic Engineering. Methods in Molecular Biology (Methods and Protocols), vol 985. Humana Press, Totowa, NJ.

[10] Charles J Norsigian, Neha Pusarla, John Luke McConn, James T Yurkovich, Andreas Dräger, Bernhard O Palsson, Zachary King, “BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree”, Nucleic Acids Research, Volume 48, Issue D1, 08 January 2020, Pages D402–D406.

[11] Lisowska, B., “Genomic analysis and metabolic modelling of *Geobacillus thermoglucosidasius* NCIMB 11955”, Thesis, Department of Biology & Biochemistry of University of Bath, (2016).

[12] Rocha, I., Maia, P., Evangelista, P. et al. “OptFlux: an open-source software platform for in silico metabolic engineering”, BMC Syst Biol 4, 45 (2010).

[13] Kanehisa M. et al. “The KEGG database”, Novartis Foundation Symposium. – Chichester; New York; John Wiley; 1999, 2002. – С. 91-100.

[14] Lewis NE, Hixson KK, Conrad TM, et al. “Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models”. Mol Syst Biol. (2010).

[15] Zachary A. King, Andreas Dräger, Ali Ebrahim, Nikolaus Sonnenschein, Nathan E. Lewis, and Bernhard O. Palsson. “Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways”, PLOS Computational Biology 11(8), (2015).

[16] Brochado AR, Andrejev S, Maranas CD, Patil KR. “Impact of stoichiometry representation on simulation of genotype-phenotype relationships in metabolic networks”, PLoS Comput Biol. 2012.

[17] Segrè D, Vitkup D, Church GM. “Analysis of optimality in natural and perturbed metabolic networks”, Proc Natl Acad Sci U S A. 2002.

[18] Becker, S., Feist, A., Mo, M. et al. “Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox”, Nat Protoc 2, 727–738 (2007).

[19] Shlomi, Tomer et al. “Regulatory on/off minimization of metabolic flux changes after genetic perturbations.” Proceedings of the National Academy of Sciences of the United States of America vol. 102,21 (2005).

[20] Oh, You-Kwan, et al. "Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data." Journal of Biological Chemistry 282.39 (2007).

[21] Li, L., Zhang, L., Li, K. et al. A newly isolated *Bacillus licheniformisstrain* thermophilically produces 2,3-butanediol, a platform and fuel bio-chemical. Biotechnol Biofuels 6, 123 (2013).

DOI 10.18699/BGRS/SB-2020-107

Integration of transcriptomics data into a genome-scale metabolic model of the methanotrophic bacterium *Methylotuvimicrobium alcaliphilum* 20Z^R

Mikhail A. Kulyashov
BIOSOFT.RU, LLC
Novosibirsk State University
Institute of Cytology and Genetics
SB RAS, Institute of Computational
Technologies SB RAS Novosibirsk,
Russia
m.kulyashov@mail.ru

Semyon K. Kolmykov
BIOSOFT.RU, LLC
Institute of Computational Technologies
SB RAS, Institute of Cytology and
Genetics SB RAS, Novosibirsk, Russia
semyonk@developmentontheedge.com

Ivan S. Evshin
BIOSOFT.RU, LLC
Institute of Computational Technologies
SB RAS, Novosibirsk, Russia
ivan@developmentontheedge.com

Tamara M. Khlebodarova
Kurchatov Genomics Center,
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
tamara@bionet.nsc.ru

Nikita V. Ivanisenko
Kurchatov Genomics Center,
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
n.ivanisenko@gmail.com

Ilya R. Akberdin
BIOSOFT.RU, LLC
Institute of Cytology and Genetics SB
RAS, Novosibirsk State University
Novosibirsk, Russia
akberdinir@gmail.com

***Abstract* — Aerobic methane-oxidizing bacteria or methanotrophs have the unique ability to grow on methane as their sole source of carbon and energy. The main metabolic steps of the methane utilization by microorganisms have been identified and well-studied to date. However, a detailed understanding of molecular genetic mechanisms that provide an adaptive response at the level of transcription regulation to various growth conditions, high and low pH, temperature, and salinity is still elusive. To solve the issue we have conducted a detailed theoretical study of the molecular mechanisms of gene expression regulation in the bacterium *Methylotuvimicrobium alcaliphilum* 20Z^R (hereinafter 20ZR) based on the integration of original omics data into genome-scale metabolic model of the 20ZR.**

Keywords* — *methane, methanotrophs, central metabolism, transcription, regulation, genome-scale modeling, Methylotuvimicrobium alcaliphilum 20Z^R

Introduction

Methane is a valuable source of energy, but on the other hand, it is a significant global product of recyclable waste and one of the most dangerous greenhouse gas [1]. In turn, methane is a promising carbon source for biosynthesis of biotechnologically useful compounds using aerobic methanotrophic bacteria as biocatalysts, and its oxidation by microbial communities of methanotrophic organisms is, in fact, the only biological mechanism for regulation and reduction of its content in the atmosphere [2-4].

In last decades, laboratory isolates of new pure cultures of methanotrophs were discovered and obtained, including facultative methanotrophs, extremophilic species and anaerobes, which greatly expanded our ideas and knowledge about both the taxonomic diversity of these microorganisms and their physiological capabilities [5-7]. Among them, haloalkaliphilic aerobic methanotrophs like 20ZR are standing out as the most promising microbial "factories" in industrial research as new sources of enzymes and protein based materials that are resistant to high salt and pH levels, and natural producers of amino acids, sugars, and osmoprotectants [8-11].

Over the past two decades of intensive experimental-theoretical studies in the field of methanotrophy outstanding progress has been made in understanding of mechanisms for methane capture from the environment by methane-consuming bacteria, identification of key metabolic steps of its utilization and the impact of various factors on both methane consumption rate and activity of metabolic pathways for its bioconversion. However, detailed understanding of molecular regulatory mechanisms and diversity of regulatory relationships in molecular-genetic machinery of methanotrophs (metabolic systems - transporters and donors of electrons for the first methane oxidation reaction; influence of the specific intracytoplasmic membrane structure on rate and mechanisms of enzymatic reactions; mechanisms and ability for methanotrophic fermentation; the impact of quorum sensing on metabolic modes) has still been elusive. Elucidation of these fundamental principles of the methane utilization and bioconversion will provide economically reasonable use of methanotrophs in industrial biotechnology [4; 12].

Recent advances in high throughput experimental technologies and its application to investigations of structural-functional organization of methanotrophs metabolism not only resulted in significant progress in understanding of crucial mechanisms of the methane utilization by these bacteria, but also led to the establishment of a holistic vision of all cellular functions of methanotrophs on different hierarchical level of their organization (genome, transcriptome, proteome and metabolome levels). The latter, to a large extent, is due to the possibility to use modern methods of bioinformatics and systems biology, which allow ones to integrate accumulated datasets and bulk of knowledge on molecular-genetic systems of the methane bioconversion into mathematical models [13-14]. Developed genome-scale metabolic models for methane-consuming bacteria, in turn, enable both *in silico* predictions of features of metabolic systems of the methane utilization for a certain representative of the aerobic methanotrophs and evaluation of their biotechnological capabilities as a producers of value-added target compounds [15-24].

Methods

Transcriptomics data analysis

To analyze original transcriptomic data on the growth of *M. alcaliphilum 20Z*^{R} under various cultivation conditions, obtained as part of a collaboration with Prof. Marina Kalyuzhnaya (San Diego California State University, USA), the already published protocol [25] has been used.

Development of the genome-scale metabolic model

Original transcriptomic data [25] have be incorporated into our previously developed metabolic flux model of *Methylotuvimicrobium alcaliphilum 20Z*^{R} [20] using the COBRAme approach [26], which is a part of the popular tool for the development and analysis of genome-scale metabolic models, COBRApy [27].

Results

Original published transcriptomics data generated in different growth conditions of the methanotroph's cultivation have been integrated into earlier developed genome-scale *20Z*^{R} metabolic model [20]. It is the first attempt to integrate this type of experimental data into the metabolic model of methanotrophic bacterium. Extended version of the model have enabled more precise predictions for growth of the methanotroph, its metabolic potential as a producer of target biotech compounds depending on both growth conditions (copper and/or lanthanum presence, methane or methanol as a carbon source) and corresponding levels of genes expression.

References

[1] S.E.M. Fletcher and H. Schaefer, "Rising methane: A new climate challenge," Science, vol. 364, № 6444, pp. 932-933, 2019.

[2] P.J. Strong, M. Kalyuzhnaya, J. Silverman and W.P. Clarke, "A methanotroph-based biorefinery: potential scenarios for generating multiple products from a single fermentation," Bioresource Technology, vol. 215, pp. 314-323, 2016.

[3] S. Cantera, S. Bordel, R. Lebrero, J. Gancedo, P.A. García-Encina, and R. Muñoz, "Bio-conversion of methane into high profit margin compounds: an innovative, environmentally friendly and cost-effective platform for methane abatement," World Journal of Microbiology and Biotechnology, vol. 35, № 1, p.16, 2019.

[4] M.G. Kalyuzhnaya, D. Collins and L. Chistoserdova, "Microbial Cycling of Methane," 2019.

[5] Y.A. Trotsenko and J.C. Murrell, "Metabolic aspects of aerobic obligate methanotrophy," Advances in applied microbiology, vol. 63, pp.183-229, 2008.

[6] S.N. Dedysh and C. Knief, "Diversity and phylogeny of described aerobic methanotrophs," in Methane Biocatalysis: Paving the Way to Sustainability, Springer, Cham, 2018, pp. 17-42.

[7] M.G. Kalyuzhnaya, O.A. Gomez and J.C. Murrell, "The methane-oxidizing bacteria (methanotrophs)," in Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes, 2019, pp.1-34.

[8] V.N. Khmelenina, M.G. Kalyuzhnaya, N.G. Starostina, N.E. Suzina, and Y.A. Trotsenko, "Isolation and characterization of halotolerant alkaliphilic methanotrophic bacteria from Tuva soda lakes," Current Microbiology, vol. 35, № 5, pp.257-261, 1997.

[9] Y.A. Trotsenko and V.N. Khmelenina, "The biology and osmoadaptation of haloalkaliphilic methanotrophs," Microbiology, vol. 71, № 2, pp. 123-132, 2002.

[10] C.A. Henard and M.T. Guarnieri, "Metabolic Engineering of Methanotrophic Bacteria for Industrial Biomanufacturing," in Methane Biocatalysis: Paving the Way to Sustainability, Springer, Cham., 2018, pp. 117-132.

[11] S. Nariya and M.G. Kalyuzhnaya, "Diversity, Physiology, and Biotechnological Potential of Halo (alkali) philic Methane-Consuming Bacteria," in Methanotrophs, Springer, Cham., 2019, pp. 139-161.

[12] L. Chistoserdova, "Methanotrophy: An Evolving Field," in Methane Biocatalysis: Paving the Way to Sustainability, Springer, Cham., 2018, pp. 1-15.

[13] I.R. Akberdin, M. Thompson and M.G. Kalyuzhnaya, "Systems Biology and Metabolic Modeling of C 1-Metabolism," in Methane biocatalysis: Paving the way to sustainability, Springer, Cham., 2018, pp. 99-115.

[14] Y. Zheng and L. Chistoserdova, "Multi-omics Understanding of Methanotrophs," in Methanotrophs, Springer, Cham., 2019. pp. 121-138.

[15] M.G. Kalyuzhnaya, S. Yang, O.N. Rozova, N.E. Smalley, J. Clubb, A. Lamb, G.N. Gowda, D. Raftery, Y. Fu, F. Bringel and S. Vuilleumier, "Highly efficient methane biocatalysis revealed in a methanotrophic bacterium," Nature communications, vol. 4, p.2785, 2013.

[16] A. de la Torre, A. Metivier, F. Chu, L.M. Laurens, D.A. Beck, P.T. Pienkos, M.E. Lidstrom and M.G. Kalyuzhnaya, "Genome-scale metabolic reconstructions and theoretical investigation of methane conversion in *Methylomicrobium buryatense strain* 5G (B1)," Microbial cell factories, vol. 14, № 1, p. 188, 2015.

[17] M.G. Kalyuzhnaya, A.W. Puri and M.E. Lidstrom, "Metabolic engineering in methanotrophic bacteria," Metabolic engineering, vol. 29, pp. 142-152, 2015.

[18] A. Demidenko, I.R. Akberdin, M. Allemann, E.E. Allen and M.G. Kalyuzhnaya, "Fatty acid biosynthesis pathways in *Methylomicrobium buryatense* 5G (B1)," Frontiers in microbiology, vol. 7, p. 2167, 2017.

[19] A. Gilman, Y. Fu, M. Hendershott, F. Chu, A.W. Puri, A.L. Smith, M. Pesesky, R. Lieberman, D.A. Beck and M.E. Lidstrom, "Oxygen-limited metabolism in the methanotroph *Methylomicrobium buryatense* 5GB1C," PeerJ, vol. 5, p. e3945, 2017.

[20] I.R. Akberdin, M. Thompson, R. Hamilton, N. Desai, D. Alexander, C.A. Henard, M.T. Guarnieri and M.G. Kalyuzhnaya, "Methane utilization in *Methylomicrobium alcaliphilum 20Z*^{R}: a systems approach," Scientific reports, vol. 8, № 1, p. 2512, 2018.

[21] C. Lieven, L.A. Petersen, S.B. Jørgensen, K.V. Gernaey, M.J. Herrgard and N. Sonnenschein, "A genome-scale metabolic model for *Methylococcus capsulatus* (Bath) suggests reduced efficiency electron transfer to the particulate methane monooxygenase," Frontiers in microbiology, vol. 9, p.2947, 2018.

[22] A.D. Nguyen, J.Y. Park, I.Y. Hwang, R. Hamilton, M.G. Kalyuzhnaya, D. Kim and E.Y. Lee, "Genome-scale evaluation of core one-carbon metabolism in gammaproteobacterial methanotrophs grown on methane and methanol,". Metabolic Engineering, vol. 57, pp. 1-12, 2019.

[23] S. Bordel, , Y. Rodríguez, A. Hakobyan, E. Rodríguez, R. Lebrero and R. Muñoz, "Genome scale metabolic modeling reveals the metabolic potential of three Type II methanotrophs of the genus Methylocystis," Metabolic engineering, vol. 54, pp. 191-199, 2019.

[24] C.A. Henard, I.R. Akberdin, M.G. Kalyuzhnaya and M.T. Guarnieri, "Muconic Acid Production from Methane using Rationally-engineered Methanotrophic Biocatalysts," Green Chemistry, vol. 21, № 24, pp.6731-6737 , 2019.

[25] I. Akberdin, D. Collins, R. Hamilton, D.Y. Oshchepkov, A. Shukla, C. Nicora, E. Nakayaku, J.N. Adkins and M.G. Kalyuzhanaya, "Rare Earth Elements Alter Redox Balance in *Methylomicrobium alcaliphilum 20Z*^{R}," Frontiers in microbiology, vol. 9, p. 2735, 2018.

[26] C.J. Lloyd, A. Ebrahim, L. Yang, Z.A. King, E. Catoiu, E.J. O'Brien, J.K. Liu and B.O. Palsson, "COBRAme: A computational framework for genome-scale models of metabolism and gene expression," PLoS computational biology, vol. 14, № 7, p. e1006302, 2018.

[27] A. Ebrahim, J.A. Lerman, B.O. Palsson and D.R. Hyduke, "COBRApy: constraints-based reconstruction and analysis for python," BMC systems biology, vol. 7, № 1, p. 74, 2013.

DOI 10.18699/BGRS/SB-2020-108

The algorithm for finding potentially oscillating behavior in enzymatic systems

Tatiana N. Lakhova
Sector of Bioinformatics and Information Technologies in Genetics
Kurchatov genomics center ICG SB RAS
Novosibirsk, Russia
tlakhova@bionet.nsc.ru

Fedor V. Kazantsev
Sector for computer analysis and simulations of biological systems
Kurchatov genomics center ICG SB RAS
Novosibirsk, Russia
kazfdr@bionet.nsc.ru

Yuriy G. Matushkin
Laboratory of molecular genetic systems ICG SB RAS
Novosibirsk, Russia
mat@bionet.nsc.ru

Sergey A. Lashin
Sector of Bioinformatics and Information Technologies in Genetics
Kurchatov genomics center ICG SB RAS
Novosibirsk State University
Novosibirsk, Russia
lashin@bionet.nsc.ru

***Abstract* — Enzymatic reactions regulate a lot of processes in a cell. Chemical compounds such as various inhibitors, activators, cofactors, allosteric regulators, etc. may influence reaction rate. We are interested in the oscillatory processes that play role in the functioning of biological systems. It is known that oscillatory behavior frequently emerges via influence of the positive and/or negative feedback loops. Development and analysis of mathematical and computer models may be the tools to explore such behavior. Therefore, the aim of our research is to develop the technology that can help find potentially oscillating microbial enzymatic subsystems and to explore their behavior.**

Keywords* — *mathematical model, feedback loop, oscillatory behavior, microbial enzymatic system

Motivation and Aim

Enzymatic reactions play an important role in metabolic activity of microbial cells. Series of reactions combines into the chain of substance modification, forming metabolic pathways. Often, enzymatic reaction products both may activation and slow down the reaction rate not only self but processes before or after it on the metabolic pathway. This influence is called positive or negative feedbacks. These feedbacks are responsible for complex behavior of biological systems functioning such as oscillations, switches etc.

Exploring of this complex metabolic network behavior frequently requires mathematical and computer modeling approaches.

In this study, we are focused on the processes with oscillatory behavior. It is unclear how to highlight such processes in a metabolic network; what properties identify the oscillatory behavior in the cells as well as if some non-oscillatory systems may become oscillatory under certain conditions or not.

Network reaction properties can only be fully understood by viewing experimental data from a theoretical perspective and by quantitative mathematical modelling of such processes. So, the aim of our research is to develop the technology that can highlight potentially oscillating microbial enzymatic reaction chains and to explore their behavior.

Results

We have developed the pipeline that can help find potentially oscillating enzymatic subsystems in metabolic network and to explore their behavior. The algorithm consist of the following steps:

The input of the algorithm receives structural (graph) model of a metabolic network.

Structural model decomposition (i.e. finding of subgraphs) and processing analysis have performed with Python to get network contours that are fits to certain criterion. Visual analysis of the selected contours has been carried out with Cytoscape (https://cytoscape.org/).

For each contour, we compiled complete mathematical model in terms of ordinary differential equations. We have used the Mammoth database (http://mammoth.biomodelsgroup.ru) [1] on this step as source of adapted to experimental data set of *Escherichia coli* enzymatic reaction mathematical models (we processed over 300 enzymatic reaction models).

Computer simulations were performed with Copasi (http://copasi.org). On this step, we have selected the parameters of models to find oscillatory behavior.

At the end, we have found several directed circuits in the structural model each of which has been checked for potential oscillation behavior. We have shown oscillatory behavior of the directed circuits involved into the tryptophan synthesis pathway.

There are a lot of open access databases containing information on metabolic pathways and gene networks that could be used as data source for model reconstruction process.

Acknowledgment

The study was supported by the Budget Project No. 0324-2019-0040-C-01 and Kurchatov Genomics Center of ICG SB RAS.

References

[1] Kazantsev F, Akberdin I, Lashin S, Ree N, Timonov V, Ratushny A, et al. MAMMOTh: A new database for curated mathematical models of biomolecular systems. J Bioinform Comput Biol. 2018;16:1740010 (16 pages). doi:10.1142/S0219720017400108.

DOI 10.18699/BGRS/SB-2020-109

Multi-class brain tumor segmentation via multi-sequences MRI mixture data preprocessing

Andrey Letyagin
Research Institute of Clinical and Experimental
Lymphology, Branch of IC&G SB RAS
Novosibirsk, Russia
letyaginay@bionet.nsc.ru

Bair Tuchinov
SDAML lab.
Novosibirsk State University
Novosibirsk, Russia
bairts@gmail.com

Evgeniy Pavlovskiy
SDAML lab.
Novosibirsk State University
Novosibirsk, Russia
pavlovskiy@post.nsu.ru

Sergey Golushko
Novosibirsk State University
Novosibirsk, Russia
s.k.golushko@gmail.com

Evgeniya Amelina
SDAML lab.
Novosibirsk State University
Novosibirsk, Russia
amelina.evgenia@gmail.com

Vladimir Groza
Median Technologies
Valbonne, France
vladimir.groza@gmail.com

Mikhail Amelin
Radiodiagnosis division
FSBI "Federal Neurosurgical Center"
Novosibirsk, Russia
amelin81@gmail.com

Nikolay Tolstokulakov
SDAML lab.
Novosibirsk State University
Novosibirsk, Russia
n.tolstokulakov@g.nsu.ru

***Abstract* — In this paper, we extend the previous work on the robust pre-processing technique which allows to consider all available information from MRI scans by composition of T1, T1C and FLAIR sequences in the unique input. Such approach enriches the input data for the automatic segmentation process and helps to improve the accuracy of the segmentation performance.**

Keywords — Neural Network, Deep Learning, Semantic segmentation, Medical Imaging

Proposed method also demonstrates significant improvement on the multiclass segmentation problem with respect to Dice metrics compare to similar training / evaluation procedure based on any single sequence regardless of the chosen neural network architecture.

Obtained results demonstrates significant evaluation improvement while combining three MRI sequences in the 3-channel RGB like image for considered problem of multiclass brain tumor segmentation. We also provide results of comparison of various gradient descent optimization methods and of different backbone architectures. We found that different algorithms worked best for different tumors, but no single algorithm ranked in the top for all types of tumors simultaneously. Final improvements on the test part of our dataset are in the range of 6–9% on the trained model according to the Dice metric with the best value of 0.949.

ACKNOWLEDGMENT

The reported study was funded by RFBR according to the research project No. 19-29-01103.

REFERENCES

[1] Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., et al. *Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation*. MICCAI Brainlesion Workshop. pp. 450–462. Springer, 2017.

[2] Myronenko, A. *3D MRI Brain Tumor Segmentation using Autoencoder Regularization*. MICCAI Brainlesion Workshop. pp. 311–320. Springer, 2018.

[3] Letyagin, A. Y., Golushko, S. K., Rzaev, J. A., Amelin, M. E.,

[4] Pavlovskiy, E. N., Tuchinov, B. N., ... Degtyareva, L. O. *Artificial Intelligence for Imaging Diagnostics in Neurosurgery.* In 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 336-337). IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC.

[5] Vladimir Groza, Bair Tuchinov, Evgeniy Pavlovskiy, Evgeniya Amelina, Mihail Amelin, Sergey Golushko, Andrey Letyagin. *Data Preprocessing via Multi-sequences MRI Mixture to Improve Brain Tumor Segmentation* Bioinformatics and Biomedical Engineering, 2020 (pp. 695-704). Springer, 2020.

DOI 10.18699/BGRS/SB-2020-110

Mathematical model of punctuated equilibrium evolution

Vitaly A. Likhoshvai
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Tamara M. Khlebodarova
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
tamara@bionet.nsc.ru

***Abstract* — Fossil record of Earth describing the last 500 million years is characterized by evolution discontinuity as well as recurring global extinctions of some species and their replacement by new types, the causes of which are still not clear. We developed a model of evolution of Earth's biota based on the universal laws of living systems functioning – self-reproduction, dependence of reproduction efficiency and mortality on biota density, mutational variability in the process of self-reproduction and selection of the most adapted individuals. We have shown that oscillatory dynamic of biota density is generated due to only these four factors in populations with sexual type reproduction. The same factors can explain such aspects of punctuated evolution observed in the fossil record as extinction catastrophes, rapid growth phases and stasis phases of species diversity. Therefore, both cyclic and intermittent dynamics of changes in the diversity of organisms with sexual reproduction that make up the Earth's global ecosystem observed in the last 500 million years may well be explained solely by internal laws of self-development with no impact of external factors.**

Keywords* — *fossil records, punctuated evolution, mass extinctions, dynamic systems, complex dynamics, modeling

Introduction

The theory of punctuated equilibrium proposed by Stephen Jay Gould and Niles Eldredge in 1972 does not belong to strict theories [1, 2]. It is based on some «empirical generalizations» of a number of facts long noticed by paleontologists, which indicate that, in the course of evolution, long periods of stability when the main features of species remain unchanged alternate with short intervals of rapid qualitative changes, which are characterized by «sudden» disappearance of species followed by their replacement by new types. Authors of this theory [1, 2] and other researchers [3-6] have found quite vivid examples in the fossil record confirming the existence of this pattern. At present, it can be considered proven that the «discontinuity» of evolution at the paleontological level is reflected at the molecular level [7-9]. However, there is still no understanding of the causes of discontinuity and uneven pace of evolution [10].

We suggested that «sudden» extinctions of species in different periods of the Earth's history and their replacement by new types, known as global extinctions, are a reflection of the phenomenon of punctuated evolution at a planetary level. The analysis of the causes of global extinctions in the life history on Earth showed that abiogenic factors are recognized as prevailing, and the various sets of which explain most of the extinctions described in the Earth fossil record [11]. However, they cannot explain the periodicity of extinctions observed in paleontological level [12-15] and the uneven pace of evolution observed in phylogenetic studies [7-9]. We suggested that there are internal biotic causes determining the dynamics of the evolutionary process. Earlier, using a minimal logistic model of the global ecosystem, we managed to interpret the periodicity of global extinctions seen in the Earth fossil record during the last 500 million years [16]. Discontinuity and uneven pace of evolution have not been investigated.

Model

We developed a frame model of the evolutionary self-development of a large ecosystem. In general, this model is represented by two equations:

$$\begin{cases} \dfrac{dx_i}{dt} = k_{in,i} + x_i \cdot \left(\mathrm{W}_i - \mathrm{Y}_i \right), \; i = 1,...,n, \\ \dfrac{du}{dt} = p_u u \cdot \mathrm{W}_{\sigma(u)} \cdot \dfrac{\partial}{\partial u} \left(\dfrac{d}{dt} \mathrm{F}_{\sigma(u)} \right), u \in U \end{cases} \quad (1)$$

First equation describes self-development of a large ecosystem over time, where the variable $x(t)$ is the density of biota consisting of n species, which is regulated by the by the most common laws of self-reproduction W and the «mortality» Y of individuals. Second equation describes the evolution of ecosystem parameters in general. Here, p_u is the rate constants for mutation-associated change in the values of the evolving parameter u and $\mathrm{F}_{\sigma(u)}$ is the "fitness" functional that is part of the evolution equation in the form of its total time derivative $d\mathrm{F}_{\sigma(u)}/dt$, which differs this equation from similar ones used in models of population evolution.

We describe the self-reproduction rate of the biota by the classical logistic model, which assumes a simple positive feedback between diversity and its growth rate (more parents – more descendants), and negative feedback based on the limited ecological space available for reproduction:

$$W = \left(k_{aSx} + k_{Sx} x \right) \left(1 - \frac{x}{C} \right)$$

where, k_{aSx} and k_{Sx} are the rate constants for asexual and sexual reproduction, parameter C has the meaning of the maximum possible biota density.

The specific rate of «mortality» of the biota we describe by the following function:

$$\mathrm{Y} = k_d + D \cdot H$$

where, k_d denotes the specific rate of «mortality», which is determined by environmental conditions. The second addendum ($D \cdot H$) sets the specific rate of «mortality», which is determined by the internal laws of the ecosystem's functioning. Here D is the rate constant of «mortality» and function H determines the mechanism of the influence of biota density on the «mortality» of biota individuals. Assume that H is a bounded monotonically increasing function of the biota density. This assumption is implemented through the function:

$$H = \frac{k_d * x * e^{\frac{x}{K_E}-1}}{K_D + x * e^{\frac{x}{K_E}-1}}.$$

Here $r = e^{\frac{x}{K_E}-1}$ denotes the coefficient of aggression, K_D and K_E – parameters.

Model (1) allows to design ecosystems with an arbitrary number of species and form of interactions. However, to analyze the problem of evolutionary discontinuity, we consider an ideal ecosystem consisting of only one species at any moment (n=1). Let us designate it as "transient population". It is also accepted that in the model of evolution of a "transient population" (1) two parameters, reproduction parameter C and «mortality» parameter D, are evolving. Two new parameters appear, p_C and p_D, which are the rate constants for mutation-associated change in the values of the evolving parameter C and D. The "fitness functional" $F \equiv W$, as shown earlier [16].

Results

Numerical calculations was shown that only stationary dynamics $x(t)$ is observed in model (1) at k_{Sx}=0 (no sexual type of reproduction). At the same time, at k_{aSx}=0 (no asexual type of reproduction) in model (1) for certain values of the parameters p_C and p_D undamped oscillatory dynamics of $x(t)$ variation was observed. Thus, the oscillatory dynamic is generated due to only four factors: 1) presence of sexual reproduction; 2) negative impact of population density on the rates of reproduction and mortality of biota individuals; 3) genetic variation caused by mutations that occur during replication; 4) evolutionary selection directed towards increasing the adaptability of the individuals to habitat conditions. Namely, the same factors can explain such aspects of punctuated evolution observed in the fossil record as extinction catastrophes, rapid growth phases and stasis phases of species diversity.

Consider a full cycle presents a typical scenario of the evolutionary turn of ecosystem development, in the sense that qualitative properties observed over a given time interval are reproduced in the course of evolution. Four phases of the parameter $x(t)$ evolution derived from the model (1) calculation are shown in Fig 1. f1 phase is characterized by a fast decrease in $x(t)$ to almost zero and qualitatively corresponds to the degree of mass extinctions observed on Earth at least three times: 65, 250, 370 million years ago (see Fig. 1c in [13]). f2 phase is characterized by the low $x(t)$ value and qualitatively corresponds to the stages of life development that follow mass extinctions. f3 phase is characterized by a fast increase in $x(t)$ and qualitatively corresponds to the stages of explosive growth in the terrestrial biodiversity. f4 phase is characterized by a high $x(t)$ value, which continues to increase slowly and qualitatively corresponds to those stages of life development on Earth, in which a great variety of life forms and a relatively low growth rate are observed. Usually such stages of evolution are described as stasis. In the model, losses and gains in species diversity inevitably repeat an unlimited number of times with approximately equal time intervals. Therefore, both cyclic and intermittent dynamics of changes in the diversity of organisms with sexual reproduction that make up the Earth's global ecosystem observed in the last 500 million years may well be explained solely by internal laws of self-development with no impact of external factors.

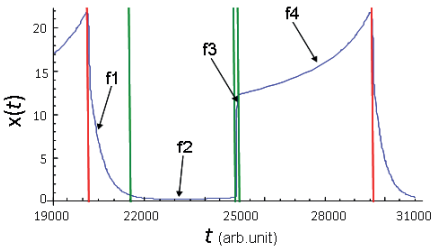


Fig. 1. Phases of $x(t)$ parameter evolution demonstrated on the example of one full cycle of oscillation of its values. Red vertical lines indicate the boundaries of the cycle, green vertical lines indicate the boundaries of the phases within the cycle.

Acknowledgment

This work was supported by the Program of Fundamental Studies of the Siberian Branch of the Russian Academy of Sciences (project No. 0324-2019-0040-C-01).

References

[1] S.J. Gould and N. Eldredge, "Punctuated equilibria: the tempo and mode of evolution reconsidered", Paleobiology, vol. 3, pp. 115-151, 1977.

[2] S.J. Gould and N. Eldredge, "Punctuated equilibria comes of age", Nature, vol. 366, pp. 223-227, 1993.

[3] R.K. Bambach, "Species richness in marine benthic habitats through the Phanerozoic", Paleobiology, vol. 3, pp. 152-167, 1977.

[4] P.O. Williamson, "Palaeontological documentation of speciation in Cenozoic molluscs from Turkana basin", Nature, vol. 293, pp. 437-443, 1981.

[5] J.J. Jr. Sepkoski, "Extinctions of life", Los Alamos Science, vol. 6, pp. 36-49, 1988.

[6] J.B. Jackson and A.H. Cheetham, "Tempo and mode of speciation in the sea", Trends in Ecology and Evolution, vol. 14, pp. 72-77, 1999.

[7] M. Pagel, C. Venditti and A. Meade, "Large punctuational contribution of speciation to evolutionary divergence at the molecular level", Science, vol. 314, pp. 119-121, 2006.

[8] S.A. Palmer, A.J. Clapham, P. Rose, F.O. Freitas, B.D. Owen, D. Beresford-Jones, J.D. Moore, J.L. Kitchen and R.G. Allaby, "Archaeogenomic evidence of punctuated genome evolution in Gossypium", Molecular Biology and Evolution, vol. 29, pp. 2031-2038, 2012.

[9] Y.I. Wolf, C. Viboud, E.C. Holme, E.V. Koonin and D.J. Lipman, "Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus", Biology Direct, vol 1, p. 34, 2006.

[10] K.L. Voje, "Tempo does not correlate with mode in the fossil record", Evolution, vol. 70, pp. 2678-2689, 2016.

[11] T.M.Khlebodarova and V.A. Likhoshvai, "Causes of global extinctions in the history of life: facts and hypotheses," Vavilovskii zhurnal genetiki i selektsii, vol. 24, pp. , 2020 .

[12] D.M. Raup and J.J. Jr. Sepkoski, "Periodic extinction of families and genera", Science, vol. 231, pp. 833-836, 1986.

[13] R.A. Rohde and R.A. Muller, "Cycles in fossil diversity", Nature, vol. 434, pp. 208-210, 2005.

[14] A.L. Melott, "Long-term cycles in the history of life: periodic biodiversity in the paleobiology database", PLoS One, vol. 3, p. e4044, 2008.

[15] G.G. Roberts and P.D. Mannion, "Timing and periodicity of Phanerozoic marine biodiversity and environmental change", Scientific Reports, vol. 9, № 1, p.:6116, 2019.

[16] V.A. Likhoshvai, V.V. Kogai, S.I. Fadeev and T.M. Khlebodarova, "Stasis and periodicity in the evolution of a global ecosystem: the minimum logistic model," Mathematical Biology and Bioinfjrmatics, vol. 12, pp. 120-136, 2017 (Russian).

DOI 10.18699/BGRS/SB-2020-111

On region based inference in genome wide association study

Sergey V. Malov
Theodosius Dobzhansky Center for Genome Bioinformatics
St.-Petersburg State University
St-Petersburg, Russia
ORCID: 0000-0003-0093-6506

Alexey Antonik
Theodosius Dobzhansky Center for Genome Bioinformatics
St.-Petersburg State University
St-Petersburg, Russia
ORCID: 0000-0003-4341-1291

Andrey K. Shevchenko
Theodosius Dobzhansky Center for Genome Bioinformatics
St.-Petersburg State University
St-Petersburg, Russia
ORCID: 0000-0002-7332-5661

***Abstract* — We consider an advanced framework for genome wide association study (GWAS) based on the signal localization. Instead of looking for single genetic markers associated with the phenotype we attempt to discover regions of genetic markers, which are associated with phenotype. We focus on the localization by moving sums of negative logarithms of p-values, which is efficient for discovery of wide regions of genetic markers associated with phenotype. Particularly, we expect the method should be efficient for searching genes containing a number of genetic markers, if any reconstruction in a gene related to a phenotype. The method is implemented in GWATCH software for visualization and interpretation results of multiple statistical tests for genome associations. We discuss some features of the GWATCH software and apply them for HIV/AIDS cohort study from Botswana.**

Keywords* — *GWAS, signal localization, chi-square test

Introduction

Genome wide association study (GWAS) is one of the most common and important methods is genetic research, set up on the statistical conclusions based on the results of multiple statistical tests. The aim aim of the GWAS is to discover genetic markers associated with the observed feature (phenotype). Increasing the number of available genetic markers leads to increasing statistical corection that makes it impossible reliable discovery of the associations. We consider an advanced method in GWAS based on signal localization [1]. Instead of looking for the single genetic markers associated with the phenotype we attempt to discover regions of genetic markers, which are associated with phenotype. The method is not efficient for associations related to single markers, but it tends to be efficient when a sufficient number of markers associated with phenotype within the same region of genome. For example our method is expected to be efficient if any reconstruction in a gene containing a number of genetic markers leads to phenotype changing. Genome Wide Tracks Chromosome Highway (GWATCH) is the advanced statistical software designed to automate the GWAS data analysis, visualization of results and meta-analysis. We discuss features of GWATCH related to region-based association discovery with some applications for real data analysis.

Signal localization method

Signal localization is quite a general methodology combining signal detection and signal indentification methods.

The signal localization framework

Let H_i, i=1,…,n be the set of null hypotheses. Introduce a family of localization sets $\mathsf{W}=\{W_j\}_{j=1}^{k}$, where each of the localization sets W_j are some subsets of indices, j=1,…,k. The localization hypothesis $\widetilde{H_j}$ holds, if all of the hypotheses $H_i, i \in W_j$ holds. Rejection of $\widetilde{H_j}$ means that there is at least one association (a signal) in the localization set W_j. Note that if H_i is the hypothesis of independence between the observed feature Y and the control variable X_i, $i=1,...,n$, the localization hypothesis is weaker than the hypothesis of full independence between Y and the vector $(X_i, i \in W_j)$, and, therefore, methods which are capable of testing the hypothesis of full independence are applicable to test the localization hypothesis as well, but may be more conservative. Choice of the localization family is rather variative. The methods of controlling an error in multiple testing remains important in the localization problem. Enriching the family of localization sets tightens the multiple testing correction, whereas selection of poor localization family is quite subjective in most cases, and user should posses a valuable arument for that choice (e.g. disjoint regions of successive markers of size k looks speculative, but selection genes as the localization setes may be natural in some cases). Further we use moving window localization, which is quite objective, and multiple testing correction does not exceed one from the classical GWAS. Let T_i be a test statistic capable of testing the null hypothesis H_i, p_i be the corresponding p-value, i=1,…,n. It looks natural to apply methods of meta-analysis based on the p-values $p_i, i \in W_j$ for testing the localization hypothesis $\widetilde{H_j}$, j=1,…,k. Fisher and Tippett approaches are applicable for testing the localization problem.

Localization using Fisher's statistic

Usage of the Fisher's combined probability method for the moving localization sets with the same number of available markers leads to the moving average process. Another choice is to use moving localization windows of the same length. In both cases the multiple testing correction is comparable with one from the classical GWAS, and it cannot be sufficiently improved in general case. The distribution of the Fisher's combined test statistic for independent markers is well known, but if the genetic markers are in linkage disequilibrium the true distribution of the statistic is not available. The naive p-values obtained under independence assumption are less then the true (adjusted) ones in most cases of dependence. Moreover, in most of practical cases the naive p-values are not srtictly related to the adjusted ones. Using the naive instead of adjusted p-values may lead to false discoveries.

The distribution of the Fisher's statistic under full independence hypotesis can be evaluated by using random permutation method. The random permutation method over all localization windows requires a lot of computations and we are using the adaptive approach as discussed in [1], but even doing that it is unattainable to obtain all the adjusted p-values. In order to discover the most significant regions we

are using the three step algorithm: (i) obtaining the naive p-values; (ii) selecting top hits of the localization windows with smallest naive p-values (e.g. top 5000), or with the naive p-values not exceeding some fixed threshold (e.g.10^{-8}), and calculating adjusted p-values for the selected localization windows; (iii) calibrating the smallest adjusted p-values in order to eliminate their bias related to the random permutation method.

Localization using χ^2

In a special case of binary phenotype and binary genetic marker (e.g. the domonant, the recessive model or the allelic approach) we developed the signal localization method based on χ^2 distribution. The chi-square test statistic for the 2×2 contingency table is actually a square of asymptotically normal statistic ξ. For each genetic marker in a fixed localization window we calculate the asymptotically normal statistic ξ_i and estimate the asymptotic correletion matrix Σ of the random vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_s)$, where s is a size of the localization window. Now, using the convergence $\boldsymbol{\xi}^T\Sigma^{-1}\boldsymbol{\xi} \Longrightarrow \chi_s^2$ we obtain the adjusted p-value. Poor estimation quality of the correlation matrix for wide localization windows is an open problem of this method. For reasonable sizes of localization windows the method is much faster then the perutation one, and it allows to compute adjusted p-values for all the localization windows.

GWATCH statistical software

Genome Wide Association Tracks Chromosome Highway (GWATCH) is the interactive software designed for efficient computation of the multiple tests on association between phenotype and any genotype available, advanced 3D visualization and automated analysis of multiple testing results as well as its meta-analysis [2]. The GWATCH as its input accepts both original genotypes and clinical data to calculate multiple statistical tests for further analysis, and pre-computed results of multiple tests supplied by the user. In the current version only the categorical tests (combined exact Fisher's test and χ^2) and the linear regression model are implemented for multiple tests computation. Other types of tests should be pre-computed by user and imported into GWATCH. We are planning to extend the list of multiple tests (first of all, for GLM and survival data analysis) available for computation in the future.

Main features

The 3-D highway vizualization concept is implemented on the top of WebGL API and renders navigable 3D view of the SNP data, which assists user to identify regions of high statistical significance and, therefore, to warrant any further investigation. Results of different statistical tests on the associations for large number of SNP's are produced and visualized in the form of colored 3D bars located according to the SNP position in genome and the type of statistical test. The height of a bar corresponds to the P-value for that SNP and test, and the color indicates the association nature (positive or negative), which determined by the sign of quantative association statistic (QAS). Genes related to SNP's are displayed along the highway. Then user can interactively retrieve the important information on the selected genetic marker (SNP), generate TRAX report which contains explanatory analysis of the genotype-phenotype association and complete list of results of the statistical tests available for the selected marker, as well as various reports based on tests results in the neghbourhood of the selected SNP in either SVG or PDF format.

Top hits and the localization

The baseline top hits tables contains the most significant p-values, top values of the QAS, top mean negative log p-values and top of the naive P-values for several types of localization windows. These tables are assembled automatically after data loading and test calculation. User can enrich the two last types of top hit tables by adding the adjusted p-values.

Applications

Botswana has the third highest HIV prevalence in the world (UNAIDS, 2019). We use GWATCH software to identify single variants and regions of genome demonstrating noticable association with HIV-1 subtype C infection in a Botswana cohort using a combination of whole-genome sequencing (WGS) and microarray genotype data. The genetic and clinical data for two cohorts were obtained in collaboration with Harvard T. H. Chan School of Public Health AIDS Initiative (Boston, USA) and Botswana Harvard AIDS Institute (Gaborone, Botswana). As a result, we have replicated several loci that have shown association with HIV in previous studies [3]. In addition, we have identified several novel putative associated loci slightly below the statistical significance threshold. Applying the localization by moving average with the window sizes of 21, 41, 71, 201 loci, 100Kbp and 200Kbp we obtain using the WGS dataset two novel putative regions associated significantly with the HIV-1C infection. The multiple testing correction is evaluated by simulations under the assumption of independent tests.

ACKNOWLEDGMENT

This research was supported in part by St.-Petersburg State University (grant N_O. 1.52.1647.2016).

REFERENCES

[1] S.V. Malov, A. Antonik, M. Tang, A. Berred, Yi Zeng and S.J. OBrien "Signal localization: a new approach in signal discovery," Biometrical Journal, vol. 59(1), pp. 126–144, 2017.

[2] A.Svitin, S.Malov, N.Cherkasov, P.Geerts, M.Rotkevich, P.Dobrynin, A. Shevchenko, Li Guan, J. Troyer, S. Hendrickson-Lambert, H.H. Dilks, T.K. Oleksyk, S. Donfield, E. Gomperts, D.A. Jabs, Efe Sezgin, M. Van Natta, P.R. Harrigan, Z.L. Brumme & S.J. OBrien, "GWATCH: a web platform for automated gene association discovery analysis," GigaScience, 3:18, 2014.

[3] W. Xie, D. Agniel, A. Shevchenko, S.V. Malov, A. Svitin, N. Cherkasov, M.K. Baum, A. Campa, S. Gaseitsiwe, H. Bussmann, J.Makhema, R. Marlink, V. Novitsky, T.-H. Lee, T. Cai, S.J. O'Brien, and M. Essex, Genome-Wide Analyses Reveal Gene Influence on HIV Disease Progression and HIV-1C Acquisition in Southern Africa. AIDS Research and Human Retroviruses, vol. 33(6), pp. 597–609, 2017.

DOI 10.18699/BGRS/SB-2020-113

Mathematical modeling of allergenic pollen propagation in atmospheric layer

Medveditsyna Olga Sergeevna
dept.physics and medical informatics
Kirov State Medical University
Kirov City, Russian Federation
ossitnikova@yandex.ru, 0000-0002-4557-6623

Rychkov Sergey Leonidovich
research and educations center of biotecnology
Vyatka State University
Kirov City, Russian Federation
rychkov@list.ru 0000-0002-7592-617X

Shatrov Anatoly Victorovich
dept.of automatics and computing Vyatka State University, dept.physics and medical informatics Kirov State Medical University
Kirov City, Russian Federation
avshatrov1@yandex.ru 0000-0002-5295-571X

***Abstract* — The quasi-two-dimensional model of impurity propagation early designed elsewhere is modified for transporting of allergenic plant pollen from spread forested areas in vicinity of a large city. The model includes consideration of mesoscale hydrothermodynamical processes in the lower atmosphere taking into account thermal nonuniformities of the underlying surface in the urban and suburban environs. The boundary conditions and the model coefficients are determined using the parametrization method. Some results of numerical calculations are presented. The calculations were performed using parallelized algorithms on the cluster supercomputer of the Vyatka State University. They show that, due to the action of an inhomogeneous horizontal temperature gradient in the lower atmosphere, vortex flows can be formed above populated areas.**

Keywords* — *allergenic pollen, mathematical modeling, impurity propagation, high-performance calculations

Introduction

Pollen grains contained in the atmosphere have the ability to cause allergic diseases. An increase in the number of pollen allergy diseases in the second half of the 20th century led to growth of interest in the atmospheric transfer of pollen around the world. 2018 is called by World Health Organization (WHO) to be the year of the allergy pandemic. The prevalence (30-60% of the population) of allergic diseases (every third inhabitant of the planet suffers from allergic rhinitis and every tenth of bronchial asthma) turned allergies into a global medical and social problem. The percentage of people suffering from allergies (mainly among the young people) is significantly higher in highly developed countries than in developing ones. The problem of allergic diseases spreading among children is very acute. Moreover, there are much fewer allergic among the villagers, while in Moscow every third one suffers from hay fever, in Berlin - every fourth, in New York - every sixth. The reason is that allergies are caused not so much by the plants themselves as their pollen, which absorbs all the harmful emissions and polluting particles that are presented in catastrophic amounts in the air of the metropolis. A sharp increase in the number of allergic diseases (AD) occurs in April-May in conditions of central Russia, when the flowering of alder, birch, willow, maple and poplar begins. During this period, pollen calendars are being formed for most large cities, patterns of pollen content of certain plant species in the atmosphere are investigated, the influence of meteorological factors is determined, a network of permanently operating stations for monitoring pollen is created [1]. The well-known information search engine Yandex publishes daily a map of distributions of pollen emissions of various origin in vicinities of large cities. Yandex.Pogoda forms a special map for those persons who are allergic to pollen.

Morphology of a pollen allergy

However, these distributions are not confirmed with numerical parameters and can be considered as illustrative material of possible occurrences of aerosol pollen components. An important factor in assessing the degree of pollution by aerosol pollen is the knowledge of the physical parameters during the pollen transfer. Data for the mass proportion of grains of various tree species and the procedure of measuring them are given in [2]. Table I shows the data for the distribution of masses of grains and the corresponding proportion of the main tree species for the European part of Russia.

TABLE I. DISTRIBUTION OF POLLEN GRAINS ON MASS

Wood species	*Mass of grains*	*Percentage in distribution*
Fir	82,4	0,06
Spruce	63,1	0,24
Pine	15,5	0,18
Linden, oak, maple, elm	10,7	0,02
Aspen	4	0,1
Birch tree	3,9	0,26
Alder	3,5	0,09
Poplar	3,5	0,03
Willow	2,5	0,02

One can see that the presented data are grouped by grain weight in three main groups: heavy (spruce, fir and pine), medium (linden, oak, maple, elm) and light (aspen, birch, alder, poplar, willow). Taking into account the presented data, attention should be paid to the fact that the mass of allergenic pollen grains has the greatest influence on the transfer processes and there is a group of light fractions. It is the last group that represents the most dangerous impact on the development of AD in large cities.

Math Model and Simulation of Problem

The statement of the problem of the transfer of pollutants in the form of aerosol components involves the use of a sequence of mathematical models built on the principle "from simple to complex." This sequence is determined by the choice of modeling scale. If we define a certain, rather small neighborhood of the impurity distribution region over a finite period of time, then it is quite reasonable to use simplified (including stationary) models allowing exact solutions. But it is necessary to take into account the complex dynamics of the velocity and temperature fields, surface inhomogeneity and boundary conditions in the case of problems with extended geometry or with sufficiently powerful pollutant sources, for example, when estimating emissions from large industrial enterprises. In [3], [4], mesoscale models of the bioaerosol components transport are considered both for individual types of plant pollen [5] and in the aggregate of inert impurities [6]. The equations of momentum, temperature, impurities and moisture transfer are used for calculations [7]. In [6] the mesoscale model of the surface layer is used to calculate transfer of fungal mold and estimate its interaction with atmospheric flow moisture. This work presents a mesoscale quasi-two-dimensional model of the transfer of pollen grains of various fractions, which are the most common types of plant aerosol allergens: birch, alder, poplar, maple, willow on a micro- and mesoscale [8]. A finite-difference explicit calculation method was used for numerical implementation. The construction of a parallel version of the calculation algorithm was based on the principle of geometric decomposition of the grid domain.

Results of Modeling

The parallel computational algorithm was implemented in the Fortran version of Intel Cluster Studio for Linux Open MP, installed on the Vyatka State University HPC Enigma X000 cluster supercomputer. The calculations were carried out on the basis of the system of equations with the initial and boundary conditions. The explicit difference scheme [3,4,7] was used on a 1000x1000 grid. In accordance with the theory of Monin and Obukhov [8], the coefficients of vertical and horizontal turbulent viscosity, thermal conductivity, and diffusion for mesoscale turbulent processes in the lower atmosphere were assumed being the same, namely, $k_m, T, S = lD^2$, $D = 400m$ and $A_m, T, S = 400m^2/s$. The wind velocity c_g was varied from 1 to 10 *m/s*. In most of calculations the velocity was 2-5 *m/s* in this case, the temperature inhomogeneity effect on the wind ow in the vicinity of a heat source is most clearly expressed. The interaction between aerosol impurity and the underlying surface was taken into account on the basis of the information on the nonuniformities of the temperature and absorption coefficient distributions taken from the map of land utilization of the computation domain. The air temperature T_S varied from 18° C outside populated areas to 23° C in the city of Kirov. A minimum temperature was observable at the north boundary of the area. The coefficient of impurity absorption by the underlying surface was taken to be $\alpha = 0.0139m^{-1}$ outside populated areas and $\alpha = 0.00139m^{-1}$ on their territories. Spread impurity sources were located on the underlying surface, on the center of the region under consideration (in vicinity of the city territory).

In the calculations it was also taken that intensity of plant pollen sources are about 10^{-2} kg/m^2. Distribution of contaminations by the wind from north-west on Fig.1 is presented.

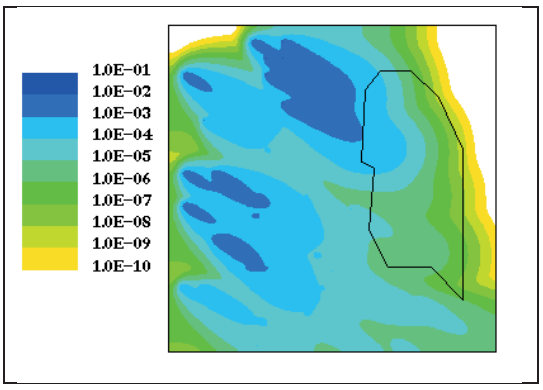


Fig. 1. Distribution of concentration of pollen for “light group” (aspen, birch, alder, poplar, willow) by the wind from north-west

References

[1] Allergophone. URL: https://allergotop.com/ (date of coll 20.10.2019)

[2] V.V. Golovko, P.K. Kutsenogiy, E.I. Kirov, V.L. Istomin, V.A. Ryzhakov “Pollen components of atmospheric aerosols the neighborhood of Novosibirsk” [In Russian], // Optic of Atmosphere and Ocean, No 6, 1998, pp. 645-649

[3] A.V. Shatrov, K.G. Shvarts “Numerical Modeling of Mesoscale Atmospheric Impurity Transport Processes in the Environs of the City of Kirov”// Fluid Dynamics. Vol. 46, No 2. 2011, pp. 333-340. DOI: 10.1134/S0015462811020165

[4] K.G. Shvarts, V.A. Shklyaev “Mathematical Modeling of Mesoscale and Largescale Processes of Transport in Baroclinic Atmosphere” [in Russian]// Moscow-Izhevsk: Institute of Computing Research, 2015.

[5] A.V. Starchenko “Modelling of transfer of aerosol impurity in the city scale” // II-th International conference "Climatology and Glatiology of Siberia", Tomsk, 20-23, october 2015.

[6] K.G. Shvarts, Yu.A. Shvarts and V.A. Shklyaev “The two-dimensional model of mesoscale processes in the lower atmosphere with allowance for inhomogeneity of temperature and air humidity.”// Vycisl. meh. splos. Sred – Computational Continuum Mechanics, vol. 8, no. 1, 2015, pp. 5-15 DOI: https://doi.org/10.7242/1999-6691/2012.5.3.32

[7] T.V. Naumovich, S.L. Rychkov, A.V. Shatrov, and K.G. Shvarts, “Software Complex for Modeling the Biotechnological Impurity Transport in the Ground Layer in the City of Kirov”// Proc. of the IV All-Russian Sci. Conf. on Math. Modeling of the Developing Economics and Ecology EKOMOD-2009 [in Russian], Vyatka Univ. Press, Kirov, 2009, pp.- 256-270.

[8] A.S. Monin, Theoretical Fundamentals of Geophysical Hydrodynamics [in Russian], Gidrometeoizdat, Leningrad 1988.

DOI 10.18699/BGRS/SB-2020-114

Named entity recognition in medical texts in Russian using deep learning models

I.V. Moskalev
ASU, Barnaul, Russia
moskalev.igor.v@gmail.com

L.A. Khvorova
ASU, Barnaul, Russia
khvorovala@gmail.com

***Abstract* — The application of contextual and domain-specific pre-trained word embeddings for recognition of medical concepts in free-text clinical notes in Russian is considered. As it is known, a large amount of medical data is stored in electronic form, a significant part - in an unstructured form (medical history, extracts, description of the results of various tests). This data contains a large amount of useful information for the diagnosis of diseases. The results of the experiments showed the effectiveness of applying contextual language models which pre-trained on medical texts in the task biomedical named entity recognition.**

Keywords — information extraction of medical data; medical text processing; biomedical named entity recognition; deep learning; word embeddings

Motivation and Aim

Motivation

In recent years, relevant scientific areas in medicine are: the transition to advanced digital, intelligent technologies, the creation of large-scale data processing systems, the development of machine learning and artificial intelligence, the transition to personalized medicine, high-tech healthcare and health-saving technologies. One of the current developing areas of artificial intelligence application is the analysis of medical data and the creation of clinical decision support systems.

There is a large amount of medical data in electronic form [1]. A significant part of this data is stored in an unstructured form (health records, discharge summaries, results of various tests and procedures). However, this data contains a large amount of information.

Medical texts are characterized with non-academic text style, the presence of a large number of acronyms and abbreviations, and incomplete sentences. One of the important tasks in the analysis of medical texts is the biomedical named entity recognition (BioNER). A large number of studies are devoted to this problem in the English language [2]–[4]. But also there are a number of works for texts in Russian [5]. This paper presents the application of advanced approaches of text processing in the task of BioNER in Russian texts.

Aim

Medical data contains important information about patients. Results of tests are stored mainly in a structural form. Data such as health histories, medical examinations reports, results of procedures have an unstructured form (natural language text). The article shows the advantage of applying contextual and domain-specific pre-trained deep learning language models in biomedical named entity recognition.

Methods

A number of approaches were considered including the applying of contextual versus non-contextual word embeddings, and pre-training on general domain texts versus domain-specific texts. We compared the performance of non-contextual word embeddings such as word2vec, GloVe, FastText and contextual word embeddings such as ELMo and BERT. Also we experimented with general domain and domain-specific pre-training. For general domain pre-training we used data from the Russuan part of Wikipedia Corpus and Corpus of Russian news articles collected from Lenta.Ru. For domain-specific pre-training we used pages of Russian part of Wikipedia Corpus which titled medical concepts and texts of the clinical protocols of the Ministry of Health. To perform medical concepts identification we used the Bi-LSTM+CRF model with non-contextual and ELMo embeddings. In case of using BERT-based models we added a simple linear layer or Bi-LSTM layer for classification. For the final optimization and evaluation of the models, an annotated dataset was prepared. It contains about 1000 free-text clinical notes manually annotated with Brat annotation tool. For the evaluation we calculated precision, recall and F1 score.

Results

The paper assesses the quality of developed methods and models of extracting information from clinical texts in Russian. The use of data mining and processing techniques will allow to automate the solution of many medical problems arising in clinical practice, thus improving the quality of primary medical care. The results of the research show the effectiveness of using contextual and domain-specific pre-trained language models in the biomedical named entity recognition task which indicates the prospect of using developed models for clinical decision support systems.

References

[1] P. B. Jensen, L. J. Jensen, and S. Brunak, “Mining electronic health records: Towards better research applications and clinical care,” Nat. Rev. Genet., vol. 13, no. 6, pp. 395–405, Jun. 2012.

[2] B. De Bruijn, C. Cherry, S. Kiritchenko, J. Martin, and X. Zhu, “Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010,” Journal of the American Medical Informatics Association, vol. 18, no. 5, pp. 557–562, 2011.

[3] S. Jonnalagadda, T. Cohen, S. Wu, and G. Gonzalez, “Enhancing clinical concept extraction with distributional semantics,” Journal of Biomedical Informatics, vol. 45, no. 1, pp. 129–140, 2012.

[4] M. Habibi, L. Weber, M. Neves, D. L. Wiegandt, and U. Leser, “Deep learning with word embeddings improves biomedical named entity recognition,” Bioinformatics, vol. 33, no. 14, pp. i37–i48, 2017.

[5] A. O. Shelmanov, I. V. Smirnov, E. A. Vishneva, “Information extraction from clinical texts in Russian,” Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”, vol. 14, no. 21, pp. 537–549, 2015.

DOI 10.18699/BGRS/SB-2020-115

A computational approach to investigation of *C. elegans* backwards crawling mechanism via simulation of involved nervous and muscular cells activity driving body movement

Andrey Yu. Palyanov
A.P. Ershov Insitute of Informatics Systems, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
palyanov@iis.nsk.su

Natalia V. Palyanova,
Institute of Molecular Biology and Biophysics,
Novosibirsk, Russia,
natasha@soramn.ru

***Abstract* — A small and well-studied invertebrate, *C. elegans*, is often considered to become the first organism reproduced in the form of a computer simulation. However, its nervous system, with hundreds of different ion channels and thousand of connections between neurons remains too sophisticated for a complete reconstruction nowadays. The subject of our research is a small subsystem including few dozens of neurons, body wall muscle cells, proprioceptors and interactions between them, which was proved to be sufficient for producing backwards crawling independently of the rest nervous system. We present the models of all key components and interactions between them, as well as the software system for performing simulations.**

Keywords* — *C. elegans, computational modeling, backwards crawling system, biomechanics, neurobiology, rhythmic activity, proprioceptive feedback

Motivation and aim

Motivation

In computational neurobiology it is highly important to know whether the model of neural activity is able to produce the same dynamics in the same situation in comparison with the real organism. For *C. elegans*, even in case when it is freely moving at natural conditions, it is possible to observe the time dependence of its individual neurons activity [1]. The system responsible for backwards crawling is an outstanding object for modeling and tuning of the model via comparison with the real experimental data.

Aim

The purpose of this investigation is to build a detailed biologically reasonable simulation of the described system, composed of the nervous and muscular cells with quite well-known properties and structure of the interconnections, but without the knowledge about how all these functions together, taking into account proprioception and feedback from the virtual physical environment. The second part to explore the cabability of simulated system to produce the expected behavior – backwards crawling – via optimization of the model's electrophysiological parameters keeping them within biologically reasonable intervals of values.

Methods

3D physical body of *C. elegans* and biomechanics of its locomotion is simulated using our software system named Sibernetic [2], which is able to calculate the dynamics of liquid, elastic matter, contractile matter (for muscle fibers simulation) and elastic films (optionally impermeable to liquid). Sibernetic is able to function together with the NEURON simulation environment, aimed at biological neurons and neural networks modeling, performing data exchange to send sensory information from physical body to nervous system and activating signals from electro-physiological representation of motorneurons to their biomechanical model.

C. elegans neurons differ from those of "more advanced" biological species which have spiking neurons interacting via action potentials – there is evidence that signals travelling through *C. elegans* nervous system propagate passively, in a gradual manner [3]. Previously we have created the basic model of the *C. elegans* neuron (within the NEURON simulator), with typical morphological properties and electrophysiological parameters responsible for passive signal propagation [4], defining its speed and length constant. However, there is an experimental evidence that some of *C. elegans* neurons can function as non-bursting oscillators – like A-class motorneurons, exhibiting intrinsic rhythmic activity at frequencies less than 1 Hz (without external stimulation) [5]. In contrast to *C. elegans* neurons, its body wall muscles are able to fire action potentials [6], even spontaneously.

The main types of interactions between the mentioned cells are the following:

1) Chemical synapses and gap junctions between neurons

2) Graded transmission at the neuromuscular junctions of *C. elegans*, encoded by post-synaptic current bursts [7] – more frequent action potentials cause larger degree of muscle contraction and vice versa

3) Spatially adjacent body wall muscles are connected via gap junctions, which, at some degree, synchronize action potentials and Ca^{2+} transients in them [8].

4) A-class motorneurons, which were active while backward locomotion, possess long undifferentiated processes, oriented anteriorly regarding to neuron's soma positions, which hypothetically function as stretch receptors [9].

Results

In addition to Sibernetic and its interaction with the NEURON, we've developed a program, which takes morphological data and information about the architecture of *C. elegans* neural network as an input, takes into account a set of parameters describing the models of muscular and nervous cells mentioned above, and automatically generates a software code written in NEURON's built-in programming language. The resulting code represents a model of electrophysiological activity and interaction between the following components, for which there is an experimental evidence that they are enough for producing backwards crawling movement [5]. It includes the majority of 95 body

wall muscles (except those which are located in the head and are driven by other mechanisms) and 21 A-class motorneurons (VA1…VA12 and DA1…DA9, innervating ventral and dorsal muscles, correspondingly), which are distributed nearly uniformly from head to tail. In more details this system is described in our recent work [10].

What is currently included into the simulation:

1) The majority of 95 body wall muscles, excluding 6 of 23 or 24 cells in each of 4 muscle quadrants, for which we reproduce their simple geometry and location within the body

2) Motorneurons DA1...DA9 and VA1…VA12 with stretch-receptive processes.

3) Muscle arms from dBWML/dBWMR body wall muscles to DA- neurons.

4) Synaptic connections between muscle arms and DA-neurons, connected via "Exp2Syn" NEURON's synapse model.

5) Synchronization between each pair of adjacent i-th and (i+1)-th dBWML muscle cells via gap junctions

6) Induced random activity of DA neurons, imitating their intrinsic rhythmic activity with typical frequencies.

ACKNOWLEDGMENT

Supported by the Russian Federation for Basic Research grant No. 18-07-00903.

REFERENCES

[1] Global brain dynamics embed the motor command sequence of Caenorhabditis elegans. Kato S, Kaplan HS, Schrödel T, Skora S, Lindsay TH, Yemini E, Lockery S, Zimmer M. Cell. 2015 Oct 22;163(3):656-69

[2] A. Palyanov, S. Khayrulin, S.D. Larson. "Three-dimensional realistic model of the Caenorhabditis elegans body and muscle cells in liquid and gel environments for behavioral analysis", *Phil Trans R Soc B: Biol Sci*, 373(1758): 20170376, 2018.

[3] W.N. Nickell, R.Y.K. Pun, C.I. Bargmann, C.I. Kleene. "Single ionic channels of two *Caenorhabditis elegans* chemosensory neurons in native membrane", *J. Membrane Biol.,* 189:55–66, 2002.

[4] A.Yu. Palyanov, A.S. Ratushnyak. "Some Details of Signal Propagation in the Nervous System of C. elegans", *Russian Journal of Genetics: Applied Research*, 5(6): 642–649, 2015.

[5] S. Gao, S.A. Guan, A.D. Fouad et al. "Excitatory motor neurons are local oscillators for backward locomotion", *eLife*, 7:e29915, 2018.

[6] S. Gao, M. Zhen. "Action potentials drive body wall muscle contractions in Caenorhabditis elegans". *Proc Natl Acad Sci USA*, 108(6): 2557–2562, 2011.

[7] P. Liu, D. Chen, Z.-W. Wang. "Postsynaptic current bursts instruct action potential firing at a graded synapse", *Nature Communications*, 4(1911), 2013.

[8] P. Liu, D. Chen, Z.-W. Wang. "Gap junctions synchronize action potentials and Ca2+ transients in Caenorhabditis elegans body wall muscle", *J. Biol. Chem.*, 286(51):44285-93, 2011.

[9] J.G. White, E. Southgate, J.N. Thomson, S. Brenner. "The structure of the nervous system of the nematode Caenorhabditis elegans", *Philos Trans R Soc Lond B Biol Sci*, 314(1165):1-340, 1986.

[10] A.Yu. Palyanov and N.V. Palyanova. "On Prerequisites for Revealing *C. elegans* Backward Crawling Mechanism through Computer Simulation of Key Involved Subsystems", *International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*: 0944-0949, 2019.

DOI 10.18699/BGRS/SB-2020-116

Adjoint ensemble methods for inverse modeling of biological processes

Alexey Penenko
Institute of Computational Mathematics and Mathematical Geophysics SB RAS
Novosibirsk, Russia
a.penenko@yandex.ru

Ulyana Zubairova
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
ulyanochka@bionet.nsc.ru

Alexander Bobrovskikh
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
alex-bobrovskih@mail.ru

Alexey Doroshkov
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russian Federation
ad@bionet.nsc.ru

***Abstract* — In the paper, we consider the inverse modeling framework based on the ensembles of the adjoint equation solutions and apply it to the model of the antioxidant system of the plant cell. We use algorithms based on sensitivity operators to identify coefficients of ordinary differential equations of production-destruction type. The analysis of the sensitivity operator allows estimating the inverse problem solution before the actual solution of the inverse problem.**

Keywords — inverse modeling, adjoint equations, sensitivity operator, production-destruction models

Introduction

In order to study complicated natural processes through mathematical modeling, it is necessary to take into account both mathematical models of the processes and the available measurement data collected about these processes. We refer to this modeling approach as inverse modeling.

The adjoint ensemble framework allows reformulating an inverse model for a system of ordinary or partial differential equations as a parametric family of quasilinear operator equations with the sensitivity operators [1, 2]. These quasilinear equations can be both solved and analyzed by the appropriate mathematical methods. To analyze the identifiability of the inverse problem with respect to the model parameters, we use the aggregates of the sensitivity operator, analogous to the generalized sensitivities. To solve the inverse problem, we use a Newton-Kantorovich type algorithm to solve the constructed quasilinear operator equations.

We consider the parameter identification problem for the basic model of the antioxidant system of a plant cell as an example of the approach application.

Acknowledgment

The work was supported by the Russian Foundation for Basic Research grant 19-07-01135 in the development of algorithms and the Russian Foundation for Basic Research grant 19-44-543021 in part of the development and research of the biological model.

References

[1] A. Penenko, “A Newton-Kantorovich Method in Inverse Source Problems for Production-Destruction Models with Time Series-Type Measurement Data,” Numerical Analysis and Applications, vol. 12, pp. 51-69, 2019.

[2] A. Penenko, U. Zubairova, Z. Mukatova, S. Nikolaev, “Numerical algorithm for morphogen synthesis region identification with indirect image-type measurement data,” Journal of Bioinformatics and Computational Biology, vol. 17, 1940002, 2019.

DOI 10.18699/BGRS/SB-2020-117

Modeling the mutual relationship between the circadian clock and inflammation response

Nikolay Podkolodnyy
ICM&MG SB RAS
Novosibirsk, Russia
pnl@bionet.nsc.ru

Natalya Tverdokheb
Systems biology dapartment
ICG SB RAS
Novosibirsk, Russia
nata@bionet.nsc.ru

Olga Podkolodnaya
Systems biology dapartment
ICG SB RAS
Novosibirsk, Russia
opodkol@bionet.nsc.ru

***Abstract* — The computer model combining the circadian oscillator (CO) with the NAMPT/NAD+/SIRT1 system and the immune and inflammatory response system through the transcription factor NF-κB was developed. Modeling showed that an age-related decrease in SIRT1 activity might be one of the causes of malfunction of CO, which may also lead to disturbances in the circadian rhythms of the body as a whole. It is likely that age-related changes in SIRT1 activity caused by NAD+ depletion can contribute to the separation of metabolic processes and circadian rhythm, which in itself can play an important role in the aging process. The simulation also showed that a decrease in the activity of Sirt1 with age leads to an increase in the activity of NF-kB, which can contribute to the aggravation of the course of inflammatory diseases.**

Keywords* — *circadian clock, SIRT1, NAD+, aging, mathematical modeling, inflamation

Motivation and Goals

The main function of the circadian clock is to synchronize the cell biological processes to adapt it to rhythmically changing external influences. Circadian disruption can cause various diseases and disorders that manifest during a person's life.

It is generally accepted that circadian clocks function due to transcriptional-translational feedbacks between genes / proteins of the core of the circadian oscillator (CO), which includes seven groups of genes - Clock (gene *Clock* and its homolog *Npas2*), Bmal (genes *Bmal1* and *Bmal2*), Per (genes *Per1*, *Per2* and *Per3*), Cry (genes *Cry1* and *Cry2*), CK1 (genes *CK1e* and *CK1d*), Rev-erb (genes *Rev-erbα* and *Rev-erbβ*) and Ror (genes *Rorα*, *Rorβ* and *Rorγ*), which form oscillations with a period close to 24 hours.

Studies of the last decade allow us to take a fresh look at the possible connections between the circadian rhythm and other functional systems of the body.

In particular, new data on the involvement of NAD+ - dependent histone deacetylase SIRT1 in the integration of circadian rhythm and metabolic regulation pathways allow us to define the new NAD+ function as a "metabolic oscillator" [1]. SIRT1 is one of the metabolic sensors that respond to various external stimuli.

Its deacetylating activity affects a wide range of substrates, among which there are both structural and regulatory proteins, so changes in its activity affect a wide variety of processes at both the cellular and systemic levels. Its effects may be associated with the survival and differentiation of cells, the development of the body and lifespan, metabolism, inflammation, the development of neurodegenerative diseases and cancer [2,3].

SIRT1 targets in the CO core can be the BMAL1 gene and protein, PER2 protein, as well as chromatin in the region of CLOCK:BMAL1 binding sites. The enzymatic activity of SIRT1 depends on the level of NAD+, which in this case acts as a co-substrate for this deacetylase. The bottleneck link in the biosynthesis and recycling of NAD+ is the enzyme NAMPT. *Nampt* gene expression is regulated by the transcription factor CLOCK:BMAL1 and exhibits circadian oscillations at the level of mRNA and protein. In accordance with this, the level of NAD+ and, therefore, the activity of SIRT1 also changes with a 24-hour period.

Thus, the functioning of the NAMPT/NAD+/SIRT1 system is closely related to the cell's CO and we can be assumed that the interaction of CO with many biological systems can be carried out through this system. In particular, now it has been shown that the transcription factor NF-κB (the main regulator of the immune and inflammatory response) is an important regulator of the circadian clock which necessary to maintain the circadian rhythm both at the molecular and behavioral levels. The CO elements interacting with NF-κB are the heterodimeric transcription factor CLOCK:BMAL1, its subunits CLOCK and BMAL1, as well as histone deacetylase SIRT1.

The aim of the work was to develop a computer model combining the CO with the NAMPT/NAD+/SIRT1 system, which ensures the interaction of the circadian clock system with the immune and inflammatory response system through the transcription factor NF-κB.

Background and Methods

We reconstructed the circadian rhythm regulation gene network of mammals based on the annotation of 280 scientific publications and information from databases which includes 55 genes, 126 proteins, 16 non-protein substances, 439 reactions and regulatory events.

The reconstructed gene network describes the interactions of a number of CO elements including heterodimeric transcription factor CLOCK:BMAL1, its subunits CLOCK and BMAL1, as well as histone deacetylase SIRT1 with transcription factor NF-κB, which is the main transcriptional regulator of the immune and inflammatory response systems.

The transcription factor NF-κB is the main transcriptional regulator of the immune and inflammatory response system. In most cell types, NF-κB is predominantly represented by the p65:p50 heterodimer [4]. The activity of the transcription factor NF-κB is regulated at the level of post-translational modification of its subunits. One of these modifications is acetylation of p65 with CLOCK acetyltransferase, a subunit of the main transcriptional regulator of the cell CO CLOCK:BMAL1 and deacylation of the same site with SIRT1 deacetylase.

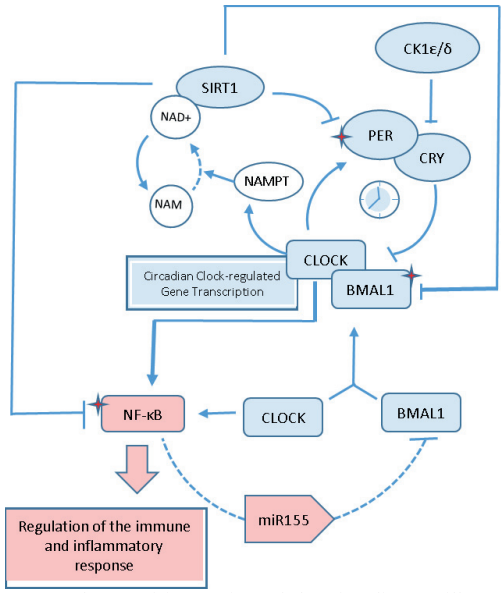


Fig. 1. Scheme of interaction of the circadian oscillator (CO) with the NAMPT/NAD+/SIRT1 system and the immune and inflammatory response system through the transcription factor NF-κB

High *Bmal1* expression suppresses the activating effect of CLOCK on the transcriptional activity of NF-κB and is manifested as a decrease in the expression of NF-κB-regulated genes [4, 5]. It is likely that BMAL1 and p65 compete for interaction with CLOCK, which can lead to disruption of the circadian oscillator mechanism of the cell under conditions of increased activity in response to inflammatory stimuli (Fig. 1).

These starting points were used to build a model for the interaction of the circadian rhythm regulation system with the immune system and the inflammatory process.

Results and Discussion

We modified and expanded the mathematical model of mammalian CO created by Kim and Forger [6], adding the following subsystems to it:

• The NAMPT / NAD+ / SIRT1 pathway.

• The pathway of enhancing the degradation of PER2 after its deacetylation with the SIRT1 enzyme.

• The pathway of the effect of SIRT1 on the transcription of the gene *Bmal1* and inhibition of the CLOCK:BMAL1 function associated with the E-box through histone deacylation.

• The pathway of interaction of CO with the transcription factor NF-κB, which is the main transcriptional regulator of the immune and inflammatory response systems.

The extended model is presented as a system of 186 ordinary differential equations (186 variables and 81 parameters).

The dynamics of changes in the main components of the extended model are consistent with the experimental data that were used in the original model of Kim and Forger [6].

Modeling while minimizing the activity of SIRT1 showed that the expression of CO genes keeps rhythm, however, the amplitude of oscillations and the general expression level of a number of genes, including *Per2*, *Bmal1* and *Rev-erb*, decreased, which is consistent with experimental data [7]. In addition, a decrease in the oscillation amplitude of the BMAL1 protein level and the CLOCK:BMAL1 complex, the main positive regulator of CO transcription, was noted.

Based on experimental data on the age-related dynamics of SIRT1 activity [7, 8] we included age as a parameter in the model. This allowed us to model age-related changes in the behavior of CO. Modeling revealed a decrease in the expression of mRNA of the *Bmal1*, *Per2, Rev-Erbα* genes and a prolongation of the oscillation period with age, which has experimental confirmation.

Our simulation also showed that the activity of NF-kB has a pronounced circadian character with a maximum at 6 o'clock in the morning. An analysis of the mortality rate of mice depending on the dose and time of endotoxin injection was performed in [4] where it was shown that the maximum mortality rate of a mouse from a pathological reaction is observed for endotoxin, administered at 6 a.m. - the time of maximum activity of NF-kB detected as a result of modeling.

Thus, our results indicate that an age-related decrease in SIRT1 activity may be one of the causes of impaired functioning of CO, which may also lead to disturbances in the circadian rhythms of the body as a whole. The simulation also showed that a decrease in the activity of Sirt1 with age leads to an increase in the activity of NF-kB, which can contribute to the aggravation of the course of inflammatory diseases.

Acknowledgment

This work was supported by project 0324–2019-0040-C-01 from the Russian State Budget.

References

[1] S. Imai, "Clocks in the NAD World: NAD as a metabolic oscillator for the regulation of metabolism and aging". Biochim Biophys Acta. 2010, Vol.1804(8), pp.1584-1590.

[2] S. Masri, "Sirtuin-dependent clock control: new advances in metabolism, aging and cancer". Curr Opin Clin Nutr Metab Care. 2015 Nov. Vol.18(6), pp.521-527.

[3] S.H. Cho, J.A. Chen, F. Sayed, M.E. Ward, F. Gao, T.A. Nguyen, G. Krabbe, P.D. Sohn, I. Lo, S. Minami, N. Devidze, Y. Zhou, G. Coppola, L. Gan, "SIRT1 deficiency in microglia contributes to cognitive decline in aging and neurodegeneration via epigenetic regulation of IL-1β". J Neurosci. 2015 Jan 14;35(2), pp.807-818.

[4] M.L. Spengler, K.K. Kuropatwinski, M. Comas, A.V. Gasparian, N. Fedtsova, A.S. Gleiberman, I.I. Gitlin, N.M. Artemicheva, K.A. Deluca, A.V. Gudkov, M.P. Antoch, "Core circadian protein CLOCK is a positive regulator of NF-κB-mediated transcription". Proc Natl Acad Sci U S A. 2012 Sep 11;109(37), pp.E2457-E2465.

[5] A.M. Curtis, C.T. Fagundes, G. Yang, E.M. Palsson-McDermott, P. Wochal, A.F. McGettrick, N.H. Foley, J.O. Early, L. Chen, H. Zhang, C. Xue, S.S. Geiger, K. Hokamp, M.P. Reilly, A.N. Coogan, E. Vigorito, G.A. FitzGerald, L.A. O'Neill, "Circadian control of innate immunity in macrophages by miR-155 targeting Bmal1". Proc Natl Acad Sci U S A. 2015 Jun 9;112(23), pp. 7231-7236.

[6] J.K. Kim, D.B. Forger, "A mechanism for robust circadian timekeeping via stoichiometric balance". Mol Syst Biol. 2012, Vol.8, p.630.

[7] H.C. Chang, L. Guarente, "SIRT1 mediates central circadian control in the SCN by a mechanism that decays with aging". Cell. 2013 Jun 20;153(7), pp.1448-1460.

[8] R.H. Wang, T. Zhao, K. Cui, G. Hu, Q. Chen, W. Chen, X-W. Wang, A. Soto-Gutierrez, K. Zhao, C-X. Deng, "Negative reciprocal regulation between Sirt1 and Per2 modulates the circadian clock and aging". Sci Rep. Vol.6, 28633 (2016).

DOI 10.18699/BGRS/SB-2020-411

SEIR-D for COVID-19 modelling and inverse problems

Aleksei Prikhodko
NSU, Novosibirsk, Russia
ICMMG SB RAS, Novosibirsk, Russia
a.prikhodko@g.nsu.ru

Maxim Shishlenin
NSU, Novosibirsk, Russia
ICMMG SB RAS, Novosibirsk, Russia
mshishlenin@ngs.ru

Sergey Kabanikhin
ICMMG SB RAS, Novosibirsk, Russia
ksi52@mail.ru

Nikita Prokhoshin
NSU, Novosibirsk, Russia
n.prokhoshin@g.nsu.ru

***Abstract* — The work provides a mathematical description of SEIR-D model and numerical method for identification of parameters of this problem by data for Moscow and Novosibirsk region. We will present a qualitative and quantitative analysis of numerical results and an analysis of the obtained forecasts for the development of the COVID-19 epidemic in Moscow and Novosibirsk region.**

Keywords — SEIR-D, COVID, coefficient inverse problem

MOTIVATION AND AIM

Motivation

In 2019, a new coronavirus (COVID-19) was discovered in Wuhan, China, which has never been detected in humans before.

Using mathematical modelling it is possible to predict the propagation of epidemy.

Aim

The aim is to calculate the scenario for the development of the disease in specific Russian cities. For example, we want to know when the incidence is expected to peak in particular city, when it will be possible to reach a plateau, and how the epidemic is expected to decline in each particular case. The mean control reproduction number was estimated in [2, 3]. In [4] the exact analytical solution of the Susceptible-Infected-Recovered (SIR) epidemic model was obtained in a parametric form.

METHODS

In SEIR-D, the population is divided into five groups: susceptible (uninfected) individuals; infected individuals without symptoms; individuals with symptoms; cured individuals; cases of the disease among the population. To the four traditional SEIR, D - deceased is added. This model described by a system of 5 nonlinear ordinary differential equations. Unknown model parameters are calculated by minimizing the cost function, using global optimization methods.

RESULTS

Mathematical model of SEIR-D based on the SEIR structure were analyzed, inverse problem of parameter refinement for Moscow and Novosibirsk region were formulated, and algorithms for solving inverse problems were developed. Scenarios for the development of the spread of the COVID-19 epidemic in Moscow and Novosibirsk region for various time periods of measurement and forecasting are constructed.

ACKNOWLEDGMENT

Supported by the RF Government 0315-2019-0005.

REFERENCES

[1] Daley D. J., Gani J. Epidemic modelling: an introduction. – Cambridge University Press, 2001. – T. 15.*)*

[2] Gumel, A.B.; Ruan, S.G.; Day, T.; Watmough, J.; Brauer, F.; van den Driessche, P.; Gabrielson, D.; Bowman, C.; Alexander, M.E.; Ardal, S.; et al. Modelling strategies for controlling SARS outbreaks. Proc. R. Soc. Lond. B.2004, 271, 2223–2232.

[3] Majumder, M.S.; Rivers, C.; Lofgren, E.; Fisman, D. Estimation of MERS-Coronavirus reproductive number and case fatality rate for the Spring 2014 Saudi Arabia outbreak: Insights from publicly available data. PLoS Curr. 2014, 18, 6.

[4] Harko, Tiberiu; Lobo, Francisco S. N.; Mak, M. K. (2014). "Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates". Applied Mathematics and Computation. 236: 184–194.

DOI 10.18699/BGRS/SB-2020-118

Parallel bias metadynamics and sketch-map dimensionality reduction as powerful tools to explore free energy landscapes of intrinsically disordered peptides

Olga Rogacheva
Laboratory of Biomolecular NMR,
SPbU, St. Petersburg, Russia
FSBRI "IEM", St. Petersburg, Russia
o.rogacheva@spbu.ru

Omar Valsson
MPIP, Mainz, Germany
valsson@mpip-mainz.mpg.de

Andrey Badanin
SPbU, St. Petersburg, Russia
a.badanin@spbu.ru

Olga Shamova
FSBRI "IEM", St. Petersburg, Russia
shamova@iemspb.ru

Abstract **— A free energy landscape reconstruction is challenging even for a small intrinsically disordered peptide because conformational ensembles of intrinsically disordered peptides are heterogeneous and span numerous degrees of freedom. We combined two recently proposed modeling approaches, parallel bias metadynamics and sketch-map, to reconstruct and analyze the free energy landscape of the small peptide containing an RGD motif (HGRGDLGRLKK). This peptide originates from the transforming growth factor beta 3 proprotein and binds to the integrin aVb6 with high efficiency and selectivity. This makes it a promising agent for cancer and fibrosis drug development since integrin aVb6 is overexpressed in these pathological conditions. The simulations indicate that the peptide is mostly disordered; however, residues 4-7 exhibit a high percentage of the helical and turns structures. These residues are known to arrange in a helix when the peptide occupies the ligand-binding site of the integrin aVb6. So, the tendency of these residues to form helical structures in solution probably explains the high affinity of the peptide for the integrin. We will use these results in further simulations of chimeric peptides incorporating this peptide as a building block.**

Keywords — intrinsically disordered peptides, parallel bias metadynamics, sketch-map, RGD-motif, free energy landscape

Motivation and Aim

A free energy landscape reconstruction is challenging even for a small intrinsically disordered peptide (IDP) because conformational ensembles of IDPs are heterogeneous and span numerous degrees of freedom. Defining stable states in this ensemble is also a non-trivial task. *A*n RMSD-based clustering is not the best choice for IDPs since separation quality is achievable only at very small cutoff values. A dihedrals-based clustering also leads to a vast amount of different conformations that do not make sense, and we cannot interpret. Bearing in mind the challenges of IDPs modeling, we combined two recently proposed approaches, parallel bias metadynamics [1] and sketch-map [2], to reconstruct and analyze the free energy landscape of the small peptide containingan RGD motif (HGRGDLGRLKK). This peptide originates from the transforming growth factor beta 3 proprotein and binds to the integrin aVβ6 with high efficiency and selectivity [3]. This makes it a promising agent for cancer and fibrosis drug development since integrin aVβ6 is overexpressed in these pathological conditions. The knowledge of the free energy landscape topology of this peptide will be helpful when introducing modifications into its sequence or using it as a building block for larger chimeric peptides. All these modifications may shift populations of the peptide conformations and result in a molecule that is incapable of receptor binding.

Methods

We used parallel bias metadynamics and the recently proposed ff99SB-disp force field [4], which is equally suitable for IDPs and folded proteins to explore the free energy landscape of the proposed peptide in solution. Biased potentials were applied along seven collective variables, including the similarity of the peptide to different regions of Ramachandran plot and order parameters of amide NH groups [5]. We furthermore performed long unbiased molecular dynamics simulations and compared the results with ones obtained in the parallel bias metadynamics simulations.

Results

On the limited time scales, parallel bias metadynamics efficiently samples a broader range of phase space than conventional molecular dynamics. Although these two methods converge in the low-lying minima well, only parallel biased metadynamics reproduce populations of high-energy states. Sketch-map is a dimensionality reduction algorithm that we used to build a 2D map of all peptide backbone dihedral angles. This allowed us to identify clusters of conformations with the elements of secondary structure and to calculate their populations. The simulations indicate that the peptide is mostly disordered with a tendency to form secondary structures of helical and beta-hairpin motifs. Perfect alpha helices and beta hairpins are poorly populated in the ensemble of conformations of this peptide as their population does not exceed 5%. However, looking at individual residues, we observe that residues 4-7 exhibit a 20-50% population of helical and turn structures. These residues are known to arrange in a helix when the peptide occupies the ligand-binding site of the integrin aVβ6. So, the tendency of these residues to form helical structures in solution probably explains the high affinity of the peptide for the integrin.

References

[1] Pfaendtner J., Bonomi M. (2015) Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics. Journal of chemical theory and computation. 11(11): 5062-5067.

[2] Ceriotti M. et al. (2011) Simplifying the representation of complex free-energy landscapes using sketch-map. PNAS, 108(32): 13023-13028.

[3] Dong X.C., Hudson N.E., Lu C., Springer T.A. (2014) Structural determinants of integrin β-subunit specificity for latent TGF-β. Nature Structural &Molecular Biology, 21(12): 1091-1096.

[4] Robustelli P., Piana S., Shaw D.E. (2018) Developing a molecular dynamics force field for both folded and disordered protein states. PNAS, 115 (21): E4758-E4766.

[5] Palazzesi F., Valsson O., Parrinello M. (2017) Conformational entropy as collective variable for proteins. 8(19): 4752-4756.

DOI 10.18699/BGRS/SB-2020-119

Development of a database on radioresistant microorganisms as a useful tool for biotechnology

Alina Ryabova
KFU, Kazan, Russia
urban-nomad@yandex.ru

Azat Kadirov
KFU, Kazan, Russia
bilinet@mail.ru

Elena Shagimardanova
KFU, Kazan, Russia
NSU, Novosibirsk, Russia
rjuka@mail.ru

***Abstract* — Mechanisms of prokaryotic radioresistance are an essential subject in radiation biology. We designed a specialized database aimed at making a comprehensive repository of identified radioresistant microorganisms with experimentally confirmed data.**

Keyword s— radioresistance, database, microorganism

Motivation and Aim

Motivation

We are yet far from detailed understanding of the extraordinary radioresistance strategies evolved by radioresistant organisms. There is an increasing need to compile the entire data about radioresistant microorganisms in a centralized and organized manner for better understanding the mechanisms of resistance to high doses of ionizing radiation.

Aim

We aim to collect and to make available for the scientific community the most up-to-date and relevant information about known radiation-resistant creatures. We designed the first version of database of radioresistant prokaryotes, as a preliminary step towards the development of a comprehensive resource for research and biotechnological purposes.

Methods

Based on the available literature, we selected studies from peer-reviewed journals supported by robust experimental data. As a result, about 80 species of radioresistant prokaryotes were selected, and their description formed the body of the database.

Results

The database schema consists of six tables, allowing to search and to retrieve any stored biological data. Among the main tables is the Taxonomy table with a detailed description of the systematic position of each species. The Irradiation table contains information about the type of ionizing radiation that the species has been exposed to, available experimental D10 values and detected lethal doses. The Mechanisms of Resistance table stores confirmed molecular mechanisms contributing to irradiation endurance. The next table refers to cross-resistance to other types of abiotic stresses. The Genomes table was obtained from NCBI databases, and contains data of sequenced strains for each species with NCBI taxonomy ID numbers, assembly level, total genome length (Mb), percentage of GC content, gene and protein count, and accession numbers for GenBank and RefSeq assemblies. The later table provides a list of references. The species name is a central link for all tables in the database. This database of radioresistant microorganisms enables quick access and easy search for pertinent information with emphasis on irradiation procedures, mechanisms of protection and repair, and genome characteristics. We plan to expand the database, adding new species of radioresistant bacteria and archaea species, and to improve by integrating new tools into the database. Such a comprehensive repository could provide opportunities in biotechnology, bioengineering, or therapeutics for multiple diseases.

ACKNOWLEDGMENT

Supported by the Ministry of Science and Higher Education of the Russian Federation (id number: RFMEFI61419X0004).

DOI 10.18699/BGRS/SB-2020-120

Stability of equilibrium points in a predator-prey model with delayed argument

Maria Skvortsova
Sobolev Institute of Mathematics SB RAS
Novosibirsk, Russia
sm-18-nsu@yandex.ru

Timur Yskak
Sobolev Institute of Mathematics SB RAS
Novosibirsk, Russia
istima92@mail.ru

***Abstract* — We consider a system of differential equations with delayed argument, which describes the interaction between predator and prey populations. We consider questions of stability of equilibrium points and study asymptotic properties of solutions. We describe sets of initial vector-functions, for which solutions stabilize at infinity. We establish estimates of solutions characterizing the rate of stabilization. The results are obtained using modified Lyapunov–Krasovskii functionals.**

Keywords* — *predator-prey model, delay differential equations, equilibrium points, asymptotic stability, estimates of solutions, attraction set, modified Lyapunov–Krasovskii functional

We consider a system of delay differential equations describing the interaction between predator and prey populations [1]:

$$x'(t) = x(t)(r - ax(t) - bz(t)),$$

$$y'(t) = kbx(t - \tau)z(t - \tau) - (D+\mu)y(t),$$

$$z'(t) = Dy(t) - \nu z(t),$$

where $x(t)$ is the density of prey at time t, $y(t)$ is the density of immature predator at time t, $z(t)$ is the density of mature predator at time t, $r > 0$ is the intrinsic growth rate of prey, $\mu > 0$ is the death rate of immature predator and $\nu > 0$ is the death rate of mature predator, constant $k > 0$ denotes the coefficient in conversing prey into new immature predator, constant $\tau > 0$ denotes the time delay due to gestation of mature predator, constant $D > 0$ denotes the rate of immature predator becoming mature predator. The system originates from predator-prey model of Lotka–Volterra type.

We consider questions of stability of equilibrium points and study asymptotic properties of solutions to this system. We describe sets of initial vector-functions, for which solutions stabilize at infinity. We establish estimates of solutions characterizing the rate of stabilization. The results are obtained using modified Lyapunov–Krasovskii functionals [2].

Acknowledgment

The work is supported by Mathematical Center in Akademgorodok, the agreement with Ministry of Science and High Education of the Russian Federation number 075-15-2019-1613.

References

[1] W. Wang and L. Chen, “A predator–prey system with stage-structure for predator,” Comput. Math. Appl., vol. 33(8), pp.83–91, 1997.

[2] G. V. Demidenko and I. I. Matveeva, “Asymptotic properties of solutions to delay differential equations,” Vestn. Novosib. Gos. Univ., Ser. Mat. Mekh. Inform., vol. 5(3), pp. 20–28, 2005.

DOI 10.18699/BGRS/SB-2020-121

Development and analysis of AIDS epidemic agent-based computer model applying an algorithm for explicit calculation of HIV replicability

Anna Smirnova
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
asmirnova@bionet.nsc.ru

Mikhail Ponomarenko
ICG SB RAS, Novosibirsk, Russia
pon@bionet.nsc.ru

Sergey Lashin
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
lashin@bionet.nsc.ru

***Abstract* — Different strains of HIV contribute differently to the course of the disease. For its evaluation, 1,336 HIV strains were analyzed. An agent model of the spread of HIV infection in the population has been developed. We analyzed 5 scenarios for the development of the HIV epidemic in Russia, depending on the initial data. Without additional measures, after 10 years, the percentage of HIV patients in Russia will increase from 1% to 2.45%. Comprehensive measures to increase the use of barrier contraception, reduce the number of joint injections among drug users and improve the situation with treatment coverage can reduce the percentage of patients from 1% to 0.3% and prevent the emergence of new patients.**

Keywords* — *agent-based model, prediction, simulation, TBP/TATA, HIV, AIDS

Motivation and aim

The AIDS epidemic is an extreme problem that is relevant not only for Russia, but also for the whole world. According to statistics, about 1% of the Russian population is HIV-infected. HIV has a large number of genetic variants that differ in both replication and virulence. The duration of the incubation period and the severity of the disease in the active phase also depend on these factors. Currently, a large number of mathematical and computer models of HIV infection have been created that consider infection at different levels of biological organization: from population-epidemiological to molecular-genetic. However, in the modern literature, we have not encountered models that combine the genetic and population-ecological levels.

Methods

The statistical method [1] of computer prediction of the affinity of the TBP/TATA complex was used to assess the replication and virulence of various genetic variants of HIV. The Python programming language was used to process the HIV strain database and visualize the data. MAFFT and BioPython were used to build a phylogenetic tree. The agent model of the AIDS epidemic is developed in the C++programming language.

Results

The first step was a genetic analysis of the HIV strains obtained from the NCBI database. Based on the analysis of 1336 HIV strains using statistical methods for predicting the affinity of the TBP/TATA complex, clustering into 3 visible groups was obtained (Fig. 1).

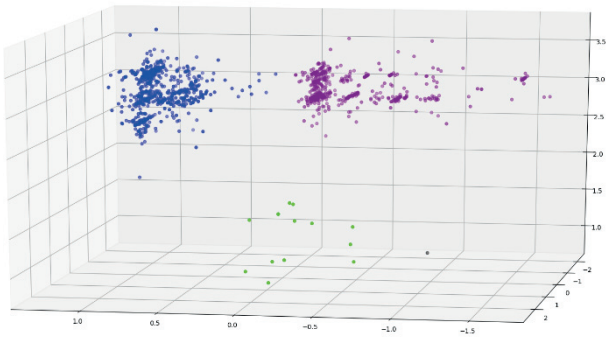


Fig. 1. Groups obtained based on the analysis of 1336 HIV strains

On the phylogenetic tree, the obtained clustering was confirmed for two groups (Fig. 2).

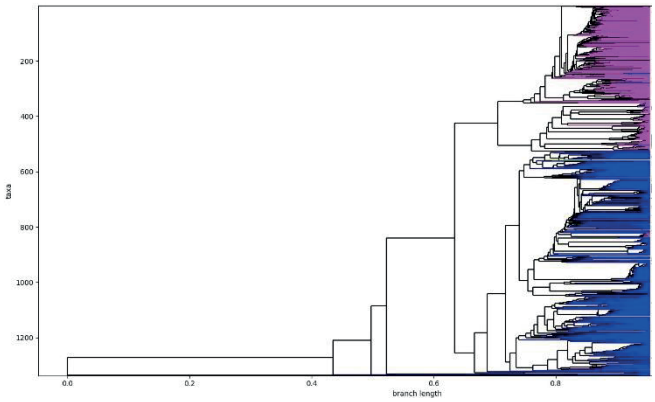


Fig. 2. The phylogenetic tree of 1336 HIV strains

In the second step, an agent computer model of the HIV epidemic was developed and implemented taking into account socio-epidemic, geographical and genetic factors.

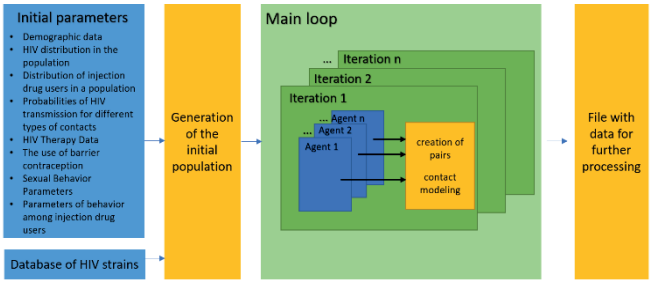


Fig. 3. HIV agent model diagram

35 strains characteristic of the territory of Russia were selected from a total of 1336 strains. Using the statistical method of computer prediction of the affinity of the TBP/TATA complex, we obtained the virulence values of these strains and integrated them into the agent model. Using

the model, forecasts were made for a 10-year period of development of the HIV epidemic in Russia for 5 scenarios of initial parameters.

The following scenarios were considered:

1) Development of the current situation without additional measures

2) Increased use of barrier contraception + decrease in the share of joint injections

3) Increased use of barrier contraception + increased treatment coverage and decreased viral load

4) Decrease in the share of joint injections + increase in treatment coverage and decrease in viral load

5) Combining all 3 options - barrier contraception, treatment and injections.

For scenario №1, the model predicts an increase in the number of HIV-infected in the population by more than two times to 2.45%. The percentage of new HIV infections per year remains stable. The most effective strategy was №5, which suggests - 1) an increase in the proportion of men and women using barrier contraception to 0.8 (instead of 0.337 for men and 0.238) 2) a decrease in the share of joint injections to 0.01 (instead of 0.4) and 3) increasing the proportion of patients using treatment to 0.9 and increasing the proportion of patients with reduced viral load among those on treatment to 0.9 (instead of 0.45 and 0.62, respectively). This scenario showed a 3-fold decrease in the percentage of HIV patients in the population from 0.97 to 0.3% over 10 years and an almost complete halt in the growth of new HIV cases.

Figure 4 shows the dynamics over 10 years of the representation of 35 strains selected for Russia in the population. The most common strain MH330341 is presented 2 times more often than the least represented MK984160.

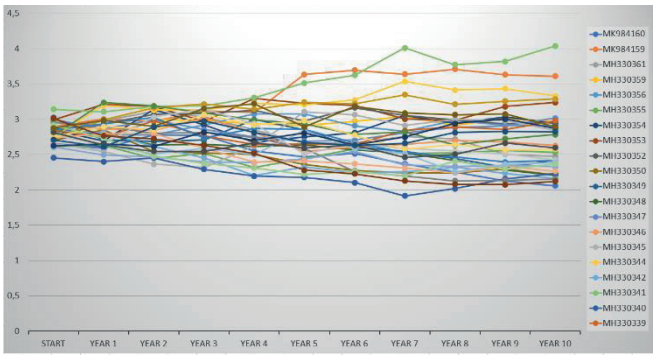


Fig. 4. The dynamics of the representation of strains in the general population over time

ACKNOWLEDGMENT

The study was supported by the Budget Project № 0259-2019-0008-C-01.

REFERENCES

[1] Suslov V.V., Ponomarenko P.M., Efimov V.M., Savinkova L.K., Ponomarenko M.P., Kolchanov N.A., SNPs in the HIV-1 TATA box and the AIDS pandemic. Journal of Bioinformatics and Computational Biology. 2010;8(3):607-625

DOI 10.18699/BGRS/SB-2020-122

Mathematical modeling of robust pattern formation in the *Drosophila* eye disc

Svetlana Surkova
SPbPU, St. Petersburg, Russia
surkova_syu@spbstu.ru

Vitaly Gursky
Ioffe Institute, St. Petersburg,Russia
gursky@math.ioffe.ru

Konstantin Kozlov
SPbPU, St. Petersburg, Russia
kozlov_kn@spbstu.ru

Sergey Nuzhdin
USC, Los Angeles, USA
snuzhdin@usc.edu

Maria Samsonova
SPbPU, St. Petersburg, Russia
m.samsonova@spbstu.ru

***Abstract* — We apply mathematical modeling approach to study mechanisms underlying the robust formation of periodic pattern during the *Drosophila* eye development. Model is fitted to the preliminary quantitative data - concentrations of the main regulators of R8 cell selection and specification. We analyze how biological noise propagates within the model to explain the molecular mechanisms responsible for noise buffering.**

Keywords* — *Drosophila eye imaginal disc, quantitative gene expression, mathematical modeling, robustness of development

Motivation and Aim

Morphogenesis of many biological structures is characterized by formation of periodic patterns from nearly homogeneous tissues. However, mechanisms underlying the robustness of such patterns are still not fully clarified. Specification of photoreceptor cells during eye development in *Drosophila* is an interesting example of rapidly forming highly ordered structure. This structure is initially determined by the precise spatial arrangement of R8 cells (presumptive photoreceptors R8) in the eye imaginal disc during larval stage of *Drosophila* development. Despite large amount of experimental data on molecular mechanisms of eye disc patterning, quantitative studies were published only recently. The expression levels of main regulators of R8 cell specification showed significant variation between *Drosophila* inbred lines, sexes and species [1]. This is unclear, how a precise structure of spatial distribution of R8 cells is formed despite variable levels of regulators, and what molecular mechanisms are involved in stabilization of this structure. To address these questions, we use mathematical modeling approach. The published modeling studies are formulated in terms of 'generalized' activators and repressors involved in R8 cell specification [2,3]. The objective of this study was to develop the approach where the mechanism of pattern formation is inferred as a consequence of the correct description of experimental data by the model.

Methods

The preliminary dataset used for model development included confocal images of *Drosophila* eye imaginal discs stained for the expression of *atonal, hedgehog, hairy, Delta* (Ali et al., 2019), *Notch* and *cubitus interruptus*. We used Hybridization Chain Reaction (HCR) technique for mRNA detection [4] and combined it with immunohistochemistry. The MrComas method was applied for quantification of gene expression from the confocal images [1]. Data on expression of other regulators of R8 cell selection and specification were obtained from literature.

Results

The model is based on the regulatory network which comprises the main factors and interactions between them relying on the available experimental information. However, the type of regulation (activation or repression) still needs to be clarified for some interactions. The model is formulated in terms of ordinary differential equations. Cell signaling is implemented in the model in two ways. First, transport of cell nonautonomous factors (e.g. Hedgehog) is described by the two-dimensional Laplacian operator, which is the diffusion operator discretized on the cell grid. Second, we include the Delta-Notch signaling mechanism in the model. We started with modification of the modeling approach implementing 'switch and template' pattern formation mechanism [2] and a generalized model with a short-range activator [3] by replacing the generalized activators and repressors with the real regulators of R8 cell specification. We fit this model to the quantitative expression data and analyze the resulted model behavior. The noise propagation in the model is estimated with the help of parameters characterizing the pattern regularity, which include the inhomogeneity index of spatial distances between the differentiated R8 cells, the inhomogeneity index of angles between photosensitive units (ommatidia), and the total number of ommatidia. We study perturbations of spatial positions of differentiating cells, as well as perturbations of parameter values, initial conditions, and disc geometry.

Acknowledgment

This study was supported by Russian Foundation for Basic Research (RFBR) grant 20-04-01047-a.

References

[1] Ali S, Signor SA, Kozlov K, Nuzhdin SV. (2019). Novel approach to quantitative spatial gene expression uncovers genetic stochasticity in the developing Drosophila eye. Evol Dev. May;21(3):157-171.

[2] Lubensky DK, Pennington MW, Shraiman BI, Baker NE. (2011) A dynamical model of ommatidial crystal formation. Proc Natl Acad Sci USA. 108: 11145–11150.

[3] .Gavish A, Shwartz A, Weizman A, Schejter E, Shilo B-Z, Barkai N. (2016) Periodic patterning of the Drosophila eye is stabilized by the diffusible activator Scabrous. Nat Commun. 7: 10461.

[4] Choi, H.M.T., Beck, V.A., Pierce, N.A. (2014) Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. ACS Nano 8: 4284-4294.

DOI 10.18699/BGRS/SB-2020-123

A feedback loop enrchment analysis in gene network of bronchial asthma and pulmonary tuberculosis interaction

Evgeny S. Tiys
Laboratory of Computer-Assisted Proteomics
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Pavel S. Demenkov
Laboratory of Computer-Assisted Proteomics
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Vladimir A. Ivanisenko
Laboratory of Computer-Assisted Proteomics
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Nikolay A. Kolchanov
Systems Biology Department
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

***Abstract* — Feedback loops play a key role in management of processes including biological processes. In gene networks, feedback loops are presented by directed cycles, that can contain positive and negative regulatory links. We suggest that presence of such cycles can characterize gene networks build for biological processes. Earlier we developed a web-tool FunGeneNet, which can estimate enrichment of biologically meaningful networks with functional links. Based on this work, here we show that the network of interactions between bronchial asthma and pulmonary tuberculosis are enriched by short cycles.**

Keywords — gene networks, random networks, enrichment analysis, feedback loops

Introduction

Feedback loops are known to play important regulatory role in biological processes [1-5]. It is know that feedback loops are significantly frequently occurs in biological networks in compare to random ones obtained by edge permutation [6]. Associative gene networks are useful tool for integration of various types of biological information. Previously, we showed that networks build for biological processes can be distinguished from random networks by number of edges [7]. The question arises as to whether the associative gene networks build for biologically meaningful sets of genes can be distinguished from the meaningless ones by the number of cycles in them? In this pilot work we show how to check it with a network of interaction between bronchial asthma and pulmonary tuberculosis [8] as an example.

Materials and methods

To build associative gene networks the ANDSystem database is used [9]. Random networks obtained as analyzed networks with swapped vertex labels were used to model biologically irrelevant networks. To faster the algorithm only vertices of the considered networks were swapped in compare to methods permuting the whole global network. A limitation of the swapping process is the following: swapped vertices have strictly the same degree in the global graph as it implemented in FunGeneNet [7]. A network of interactions between asthma and tuberculosis obtained earlier [8] and containing 19 genes was taken as a biologically meaningful network. Also we considered a subnetwork of this network, which is relevant to th1- and th2-immune responses including IL2, IL12, TNF, IFNG, IL1B, IL6, IL4, IL10 genes. We excluded undirected types of interactions from analysis before comparing the networks with random ones. The following types of interactions were considered: up- and down- regulation of expression, activity and degradation. Directed cycles of length 3 were counted in the analyzed and random networks.

Results

We obtained significant differences in the number of cycles for both whole network of bronchial asthma – pulmonary pulmonary tuberculosis interaction and th1-th2-network ($p<0.001$). These results support us to test if other biologically relevant sets of genes are enriched with the feedback loops. We plan to check whether the sets of genes associated with the Gene Ontology biological processes contain bigger number of cycles than randomly chosen sets of genes.

Acknowledgment

This work was supported by budget project AAAA-A17-117092070032-4.

References

[1] Jr. Ferrell, "Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability," *Current opinion in cell biology*, vol. 14, no. 2. pp. 140–148, 2002

[2] Yu Z *et al.* "A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation," *The Journal of cell biology*, vol. 182, no. 3, pp. 509–517, 2008.

[3] S. Sun *et al.* "miR-146a and Krüppel-like factor 4 form a feedback loop to participate in vascular smooth muscle cell proliferation," *EMBO reports*, vol. 12, no. 1, pp. 56-62, 2011.

[4] J. Locke *et al.* "Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana," *Molecular systems biology*, vol. 2, no. 1, 2006.

[5] S. Brabletz *et al.* "The ZEB1/miR-200 feedback loop controls Notch signalling in cancer cells," *EMBO reports*, vol. 30, no. 4, pp. 770-782, 2011.

[6] R. Milo *et al.* "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824-827, 2002.

[7] E. Tiys *et al.* "FunGeneNet: a web tool to estimate enrichment of functional interactions in experimental gene sets," *BMC genomics*, vol. 19, no. 3, pp. 76, 2018.

[8] E. Bragina *et al.* "Insights into pathophysiology of dystropy through the analysis of gene networks: an example of bronchial asthma and tuberculosis," *Immunogenetics*, vol. 66, no. 7-8, pp. 457-465, 2014.

[9] V. Ivanisenko *et al.* "ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology," *BMC systems biology*, vol. 9, no. S2, pp. S2, 2015.

DOI 10.18699/BGRS/SB-2020-124

Errors in miRNA recognition

Pavel Vorozheykin
NSU, Novosibirsk, Russia
Pavel.Vorozheykin@gmail.com

Igor Titov
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
Titov@bionet.nsc.ru

Abstract — **Selecting and using the key miRNA features are crucial for the identification of miRNA sequences. These criteria may help us to take out junk data from the miRNA databases and to improve the prediction of miRNA ends in order to obtain more accurate forecasting of the seed-related functions. We consider different processes of the miRNA identification to determine their role and influence on the prediction quality.**

Keywords — miRNA, pre-miRNA, miRBase, MirGeneDB, prediction method, miRNA origin, miRNA biogenesis

Motivation and Aim

MicroRNAs constitute a large class of small non-coding RNAs which clearly play an important role in the biological regulation system. Great interest of researchers led to a rapid increase of both publications and annotated miRNA sequences in the miRBase database [1]. However, only a small part of them are completely described and experimentally validated. During the process of data accumulation, the number of invalid samples is inevitably growing. To date the miRBase contains not only the false positives but lacks large number of *bona fide* miRNAs and pre-miRNAs [2]. So, the existing annotations and validation criterias are becoming obsolete and need to be revised and upgraded from time to time.

Results

Each step of the work with miRNAs introduces some distortions in processed data, so called natural, experimental and computational errors.

Natural errors occur when the miRNAs are cutting by the Dicer and Drosha complexes and appear as the heterogeneous miRNA ends. However, these defects can be effectively corrected by exosomes' trimming and tailing mechanisms. Such processes of miRNA maturation produce a natural noise, which (among other mechanisms) serves as an evolutionary driver of miRNA development and is involved in the creation of novel miRNAs.

Natural errors have significantly impact on the experimental errors. This type of errors is due to incorrect experimental validation of the miRNA functions. As a result, the miRNA database is updated with incorrect annotated miRNAs some of which belong to the transcriptional garbage (e.g. tRNAs). Other false-positive miRNAs demonstrate abnormal heterogeneous ends or cannot be observed in repeatable experiments at all (e.g. xenomiRs).

Finally, the natural and experimental errors are accumulated in training datasets which leading to inaccuracy of the computational methods. Existing computational and heuristic methods which predict the miRNA based on the criteria from different areas of knowledge about miRNAs. Ones predict the miRNAs by applying the evolutionary conservation of sequences and functions across the related species, the sequence and structure similarity and cluster analysis of the miRNA locations. Other methods implement machine learning algorithms (Hidden Markov Models, Support Vector Machines, Naive Bayesian Classifiers, etc.) using the context-structural characteristics of the miRNA sequences and their precursors (pri-/pre-miRNAs). To improve the method's efficiency ones add more intrinsic sequence features, such as oligonucleotide motifs, nucleotide frequencies, MFE of the precursor hairpin, conserved stems and loops, etc. Finally, the most costly and complex methods take the advantages of the next generation sequencing and experimental validation [3]. To increase the quality the prediction methods may include several criteria simultaneously [4].

We consider the contribution of these independent and combined criteria for the classification of putative miRNAs and on the prediction of miRNAs.

Conclusion

Results can be useful to adjust the methods either to predict new microRNAs or to verify already annotated ones. Also, the description of the significant miRNA features can shed light on the miRNA's evolutionary origin and on the specificities of the pri-/pre-/miRNA processing and context-structural characteristics of sequences.

References

[1] Kozomara A., Birgaoanu M., Griffiths-Jones S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Research. 47(D1):D155–D162.

[2] Fromm B., Billipp T., Peck L.E., et al. (2015). A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. Annual review of genetics. 49:213-242.

[3] Kozomara A., Griffiths-Jones S. (2019). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research. 42(D1):D68–D73.

[4] Fromm B., Domanska D., Høye E., et al.(2020). MirGeneDB 2.0: the metazoan microRNA complement. Nucleic Acids Research. 48(D1): D132–D141.

DOI 10.18699/BGRS/SB-2020-125

Gene network of type 2 diabetes: reconstruction and analysis

Zamyatin Vladimir
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
zamyatin@bionet.nsc.ru

Mustafin Zakhar
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
mustafin@bionet.nsc.ru

Matushkin Yury
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
mat@bionet.nsc.ru

Afonnikov Dmitry
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
ada@bionet.nsc.ru

Klimontov Vadim
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
klimontov@mail.ru

Lashin Sergey
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
lashin@bionet.nsc.ru

***Abstract* — Currently, due to the appearance of an increasing number of genomic, molecular, histological data, there is an intensive detailing of individual molecular genetic systems of the human body and phenotypic deviations caused by violations in one or more elements of these systems. At the same time, there is no, as such, a holistic understanding of the nature of the formation and course of type II diabetes, which includes an enough available experimental data. In this study, we reconstructed gene networks of transcriptional regulation, functional connectivity, and protein-protein interactions for type 2 diabetes. Information on the evolutionary age of genes was superimposed on the network; it was shown that the genes in question are predominantly "evolutionarily old". New genes have been found that were not previously associated with type 2 diabetes, such as *PER1, PER3, ARHGEF4*, genes whose homologues are associated with the onset of diabetes in mice - *CLOCK, ARNTL* (encoding transcription factors) - the goals for subsequent experimental confirmation. Validation of gene networks by analysis of tissue-specific expression has shown that most genes included in putative signal transduction pathways are expressed in the same tissues.**

Keywords — gene networks, type 2 diabetes, SNP

Motivation and aim

Motivation

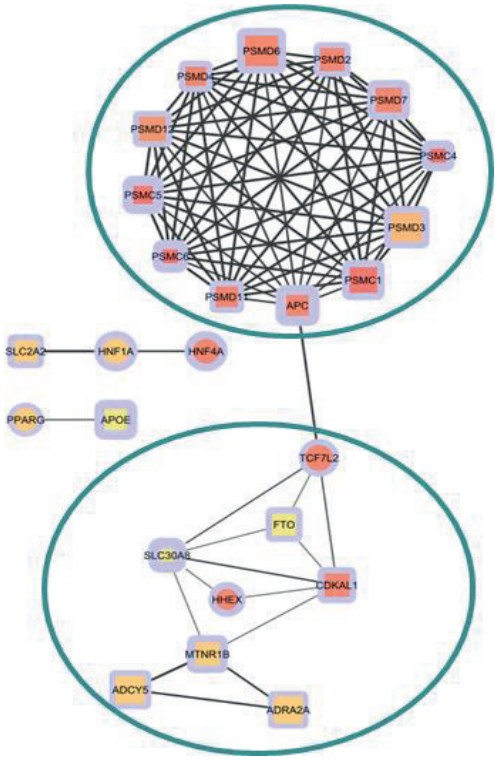
Type 2 diabetes is a multifactorial disease caused by a combination of genetic and environmental factors. Today, systems biology approaches, and in particular, network approaches, are powerful tools for studying multifactorial diseases. They are based on knowledge of the physical or functional interactions between molecules, usually presented as a network of bonds. Such networks not only report connections between individual nodes, but also implicitly organize higher-level cellular connections. In this paper, we consider networks of transcriptional regulation, protein-protein interactions, functional connectivity, bringing them to the general value of gene networks - molecular genetic systems, which ensure the development of phenotypic properties of organisms based on genetic information.

Aim – reconstruction and analysis of the gene network of type 2 diabetes mellitus.

Methods

Based on the analysis of literature data, as well as a number of databases, a sample of genes associated with type 2 diabetes was formed. Among them, transcription factors (TFs) were identified, and using the Hocomoco database [1], TF binding sites were searched in 5' non-coding gene regions from the obtained samples. Using the Orthoscape program [2], information was obtained on the philostratigraphic age and divergence index for each gene. Intrapopulation variability was estimated using SNP data from the project "1000 human genomes" [3]. Gene networks were reconstructed in the Cytoscape program [4] based on data on interactions extracted from GeneMANIA [5] and StringDB [6].

I. Results



Evolutionary network map of functional connectivity. The bulk of the network genes, based on our calculations, is rather "evolutionarily old". Two circuits are distinguished - signal transmission paths

After applying the previously obtained information, an analysis of gene networks was carried out. Regulatory circuits were identified, the violations in the elements of which can lead to disease. They contain new genes not previously associated with T2D, such as *PER1, PER3, ARHGEF4*, genes whose homologues are associated with the onset of diabetes in mice - *CLOCK, ARNTL* (encoding transcription factors) [7], as well as genes that are not in the original sample but associated with T2D, such as *PDX1*, which validates our chosen methodology.

The gene network analysis is an effective tool for a comprehensive theoretical study of human diseases and can be used not only to search for and prioritize genes associated

with diseases, but also to explain the molecular genetic mechanisms behind them.

ACKNOWLEDGMENT

Supported by the RFBR (20-04-00885).

REFERENCES

[1] Kulakovskiy IV, Vorontsov IE, Yevshin IS, Soboleva AV, Kasianov AS, Ashoor H, Ba-Alawi W, Bajic VB, Medvedeva YA, Kolpakov FA, Makeev VJ. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Research. 2015; 44 (D1): D116-25. doi: 10.1093/nar/gkv1249.

[2] Mustafin, ZS, Lashin, SA, Matushkin, YG, Gunbin, KV, Afonnikov, DA. Orthoscape: a cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles. BMC Bioinformatics. 2017; S1(18): 1–9. doi: 10.1186/s12859-016-1427-5

[3] Auton, Adam, Abecasis, Gonçalo R., Altshuler, David M. Durbin, Richard M et al. A global reference for human genetic variation // Nature. 2015. doi: 10.1038/nature15393

[4] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B; Ideker T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research. 2003; 11(13): 2498–2504. doi: 10.1101/gr.1239303

[5] Szklarczyk D, Franceschini A, Wyder S, Forslund K et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research. 2014; D1 (43). doi: 10.1093/nar/gku1003

[6] Warde-Farley D, Donaldson S, Comes O, Zuberi K et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Research. 2010; S2 (38). doi: 10.1093/nar/gkq537

[7] Marcheva B Ramsey K, Buhr E, Kobayashi Y, et al. Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. Nature (2010). doi:10.1038/nature09253

Symposium
Genomics, bioinformatics and evolution

DOI 10.18699/BGRS/SB-2020-126

CryProcessor: a novel tool for mining Cry toxins in *Bacillus thuringiensis* sequencing data

Kirill Antonets
Laboratory for Proteomics of Supra-Organismal Systems
All-Russia Research Institute for Agricultural Microbiology
Saint Petersburg State University
St. Petersburg, Russia
k.antonets@arriam.ru

Yuri Malovichko
Laboratory for Proteomics of Supra-Organismal Systems
All-Russia Research Institute for Agricultural Microbiology
Saint Petersburg State University
St. Petersburg, Russia
yu.malovichko@arriam.ru

Anton Nizhnikov
Laboratory for Proteomics of Supra-Organismal Systems
All-Russia Research Institute for Agricultural Microbiology
Saint Petersburg State University
St. Petersburg, Russia
a.nizhnikov@arriam.ru

Anton Shikov
Laboratory for Proteomics of Supra-Organismal Systems
All-Russia Research Institute for Agricultural Microbiology
Saint Petersburg State University
St. Petersburg, Russia
a.shikov@arriam.ru

Rostislav Skitchenko
ITMO University
St. Petersburg, Russia
rost20151995@gmail.com

***Abstract* — *Bacillus thuringiensis* is a gram-positive bacterium, which produces very diverse class of protein toxins, Cry. Precise annotation of these toxins in the bacterial genomes is a very challenging problem due to the sequence similarity between different toxins can be lower than 40%. Here we describe a novel tool, CryProcessor, for searching Cry toxins in protein sequences and in raw sequencing reads.**

Keywords* — *Bacillus thuringiensis, Cry toxins, data mining

Motivation and aim

Bacillus thuringiensis is a gram-positive bacterium, widely used in agriculture, because it is a natural pathogen of insects. This species of bacteria produces very diverse class of protein Cry toxins, which are highly specific against particular species of insects. To date, more than 700 different Cry toxins are described, which can share less than 40% sequence similarity. This makes the search for Cry toxins, particularly the new ones, in bacterial genomes very difficult. Here we present a novel tool for searching Cry toxin sequences.

Methods

CryProcessor tool is based on several HMM for each domain of Cry proteins. The pipeline is written in Python3.7 and is able to find Cry toxin sequences in the sets of protein sequences or in the raw Illumina reads. It also provide domain coordinates for found toxin sequences. Detailed information is available at https://github.com/lab7arriam/cry_processor. Web-version of the tool is accessible at https://lab7.arriam.ru/tools/cry_processor.

Results

We have tested our tool, CryProcessor in all modes with protein sequences of all available *B. thuringiensis* genome assemblies, and also with several SRA sequencing data. It has been shown to be more sensitive and fast comaparing with the existing analogue, BtToxin_scanner [1]. It has been also able precisely to predict the domain coordinates of the found toxins, unlike BtToxin_scanner.

Acknowledgment

This work was supported by the Russian Science Foundation (grant No. 18-76-00028).

References

[1] Ye W., Zhu L., Liu Y., Crickmore N., Peng D., Ruan L., Sun M. “Mining new crystal protein genes from *Bacillus thuringiensis* on the basis of mixed plasmid -enriched genome sequencing and a computational pipeline”. Appl. Environ. Microbiol., vol. 78, pp.4795-4801, 2012

DOI 10.18699/BGRS/SB-2020-127

Methylation and expression profiles in Apoe vicinity point to specific neighboring interaction of Apoe and TOMM40 genes: implication for the Alzheimer disease

Vladimir Babenko
Laboratory of Neuropathology Modeling, Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
bob@bionet.nsc.ru

Roman Babenko
Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
babe-roman@yandex.ru

Evgeny Rogaev
Institute of Basic Genetics, GSP-1 Gubkina 3 Moscow, 11999, Russia
rogaev@vigg.ru

***Abstract* — We assessed the dynamics of 8 genes including *TOMM40, Apoe* and six adjacent ones for overall chromatin marks landscape, including methylation profiles across ENCODE brain cell lines, and histone and *ctcf* marks. We revealed the region manifests Hi-C topology dynamics in a cell-specific manner. Additionally, based on methylation and histone marks profiles we underscore competitive manner of genes expression implying disrupted locus wide genes expression balance in Alzheimer patients due to *Apoe* extended locus methylation profile alteration.**

Keywords* — *Alzheimer disease, methylation, aging, chromatin, GWAS

Introduction

As it is follows from its definition, Late Onset Alzheimeir Disease (LOAD) features late onset disease manifestation assuming some age – related processes are implicated. One of the factors could be well known age-related methylation profile alteration. Another specific to LOAD fact is the observation of extra high significance of SNPs, corresponding to Apoe Epsilon alleles, and LOAD linkage found in the successions of Genome Wide Association Studies (GWAS). That is why several recent studies [1, 2] explored association of epigenetic factors such DNA methylation and age and genotype in disease cohorts. Recent article [3] found significant differences in hippocampus and cerebellum methylation rate in *TOMM40* promoter between LOAD patients' samples and controls. Also, evidence reported in [4] that besides *Apoe* gene, there is a range of SNPs within expanded Apoe related region significantly associated with LOAD risk, underlining *PVRL2- CLPTM1* region is implicated in LOAD genetic risk rate.

Research

We scrutinized *Apoe* neighborhood and built the profiles of methylation and co-expression for 8 genes in the vicinity including *Apoe*. We observe specific pairwise correlation variation when comparing DNA methylation rate and particular gene expression. In particular, we elucidated complex 5' and 3' interplay across 8 adjacent genes region focused on *Apoe*, implying competitive expression of the neighbor genes depending on the methylation profile. As a conclusion, we report extensive cross-talk of methylation profile in the region, affecting the chromatin environment in the region, including ctcf and histone methylation marks.

Conclusion

We found that methylation profile alterations in *Apoe* region drastically change the balance of expression rates in genes of the region.

Acknowledgment

The work was supported by Russian Foundation for Basic Research grant 18-254-1764.

References

[1] Ma Y, Smith CE, Lai CQ, Irvin MR, Parnell LD, Lee YC, Pham L, Aslibekyan S, Claas SA, Tsai MY, Borecki IB, Kabagambe EK, Berciano S, Ordovás JM, Absher DM, Arnett DK. Genetic variants modify the effect of age on APOE methylation in the genetics of lipid lowering drugs and diet network study. Aging Cell. 2015;14(1):49–59. https://doi.org/10.1111/acel.12293.

[2] Babenko VN, Afonnikov DA, Ignatieva EV, Klimov AV, Gusev FE, Rogaev EI. Haplotype analysis of APOE intragenic SNPs. BMC Neurosci. 2018; 19(Suppl 1):16. doi: 10.1186/s12868-018-0413-4

[3] Shao Y, Shaw M, Todd K, Khrestian M, D'Aleo G, Barnard PJ, Zahratka J, Pillai J, Yu CE, Keene CD, Leverenz JB, Bekris LM. DNA methylation of TOMM40-APOE-APOC2 in Alzheimer's disease. J Hum Genet. 2018 Apr;63(4):459-471. doi: 10.1038/s10038-017-0393-8. Epub 2018 Jan 25. PubMed PMID: 29371683; PubMed Central PMCID: PMC6466631.

[4] Zhou X, Chen Y, Mok KY, Kwok TCY, Mok VCT, Guo Q, Ip FC, Chen Y, Mullapudi N; Alzheimer's Disease Neuroimaging Initiative, Giusti-Rodríguez P, Sullivan PF, Hardy J, Fu AKY, Li Y, Ip NY. Non-coding variability at the APOE locus contributes to the Alzheimer's risk. Nat Commun. 2019 Jul 25;10(1):3310. doi 10.1038/s41467-019-10945-z. PubMed PMID: 31346172; PubMed Central PMCID PMC6658518.

DOI 10.18699/BGRS/SB-2020-128

Diversity and evolution of *Tat* LTR retrotransposon structures in non-flowering plants

Mikhail Biryukov
Interinstitutional laboratory of molecular paleogenetics and paleogenomics
ICG SB RAS
Novosibirsk, Russia
birykov@bionet.nsc.ru

Kirill Ustyantsev
The sector of molecular-genetic mechanisms of regeneration
ICG SB RAS
Novosibirsk, Russia
ORCID 0000-0003-4346-3868

***Abstract* — Retrotransposons (RTs) can be characterized as assemblies of protein domains each of which can have an independent evolutionary fate. This led to a modular view on RTs evolution. An illustrative example is independent acquisition of additional ribonuclease H (aRNH) domains by distant evolutionary lineages of long terminal repeat (LTR)-RTs in genomes of plants and parasitic protists oomycetes and vertebrate retroviruses. In all the cases structures of the elements converged to a single 'retroviral' variant when the aRNH domain is fixed right after the original one [1, 2]. Only LTR RTs of the *Tat* clade in plants exhibit structural diversity with various positions of aRNH in relation to the other domains. However, previous study of *Tat* LTR-RTs diversity was mostly limited by flowering plants [1]. Currently, numerous genomes of non-flowering plants assemblies are available for the analysis, opening a perspective to study general tendencies in *Tat* LTR-RTs evolution in distant plant taxa. We analyzed 78 genome assemblies of gymnosperms, ferns, clubmosses, mosses, and green algae. Our phylogenetic analysis revealed another example of convergent evolution to the 'retroviral' structure in plant LTR-RTs outside *Tat*, supporting its selective advantage.**

Keywords* — *LTR retrotransposons, convergent evolution, ribonuclease H, RNH, Tat

Introduction

Retrotransposons can be characterized as assemblies of protein domains each of which can have an independent evolutionary fate. This led to a modular view on RTs evolution. An illustrative example is independent acquisition of additional ribonuclease H (aRNH) domains by distant evolutionary lineages of long terminal repeat (LTR)-RTs in genomes of plants and parasitic protists oomycetes and vertebrate retroviruses. In all the cases structures of the elements converged to a single 'retroviral' variant when the aRNH domain is fixed right after the original one [1], [2]. Only LTR RTs of the *Tat* clade in plants exhibit structural diversity with various positions of aRNH in relation to the other domains. However, previous study of *Tat* LTR-RTs diversity was mostly limited by flowering plants [1]. Currently, numerous genomes of non-flowering plants assemblies are available for the analysis, opening a perspective to study general tendencies in *Tat* LTR-RTs evolution in distant plant taxa.

Methods and Algorithms

We implemented a standalone pipeline written in Python which uses NCBI RPS-BLAST and CDD to mine LTR-RTs sequences with aRNH in genome assemblies, automatically annotates domain structure of the identified elements, clusters them and returns non-redundant set of best-scored representative sequences for phylogenetic analysis. The phylogeny was evaluated using IQ-TREE.

Results

We analyzed 78 genome assemblies of gymnosperms, ferns, clubmosses, mosses, and green algae. Four LTR-RTs structures differing in aRNH position were identified, one of which with aRNH before GAG found in clubmosses is observed for the first time. LTR-RTs from ferns and some clubmosses formed a distinct cluster outside *Tat*, where the majority of elements have the 'retroviral' structure. The phylogenetic analysis demonstrated that acquisition of aRNH by ferns and clubmosses appears to be independent from the *Tat* clusters. Among coniferous gymnosperms all previously reported structures were confirmed, while in ancient *Ginkgo biloba* and *Gnetum montanum* LTR-RTs with the 'retroviral' and typical gymnosperm variant were found. Phylogenetic analysis of aRNH supports origin of the 'retroviral' structure in *Tat* before the split between *G. biloba*, *G. montanum* and flowering plants.

Conclusion

Our phylogenetic analysis revealed another example of convergent evolution to the 'retroviral' structure in plant LTR-RTs outside *Tat*, supporting its selective advantage.

Acknowledgment

Supported by the Russian State Budget Project № 0259-2019-0010, and № 0324-2019-0040-C-01.

References

[1] K. Ustyantsev, O. Novikova, A. Blinov, and G. Smyshlyaev, "Convergent Evolution of Ribonuclease H in LTR Retrotransposons and Retroviruses," Mol. Biol. Evol., vol. 32, no. 5, p. 1197, May 2015.

[2] K. Ustyantsev, A. Blinov, and G. Smyshlyaev, "Convergence of retrotransposons in oomycetes and plants," Mob. DNA, vol. 8, no. 1, p. 4, Dec. 2017.

DOI 10.18699/BGRS/SB-2020-129

New data on Acanthobdellida phylogeny based on complete mitochondrial genomes

Alexander Bolbat
Laboratory of Analytical and Bioorganic Chemistry
Limnological institute SB RAS
Irkutsk, Russia
bolbatav@lin.irk.ru

Gennadiy Vasiliev
Dept of Genomic Research
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
genn@bionet.nsc.ru

Vera Bogdanova
Laboratory of Fabaceae plants genetics and evolution
Institute of Cytology and Genetics
SB RAS, Nobosibirsk, Russia
Vera@bionet.nsc.ru

Irina Kaygorodova
Laboratory of Analytical and Bioorganic Chemistry
Limnological institute SB RAS
Irkutsk, Russia
irina@lin.irk.ru

Abstract — **Acanthobdellida is a unique order of leech-like piscine parasites that are believed to occupy an intermediate evolutionary position between Oligochaeta and Hirudinea due to mosaic combination of their morphological traits. Previous molecular studies yielded controversial phylogenies for this group. Using next generation sequencing technology, we reconstructed mitochondrial genomes of both Acanthobdella peledina and Paracanthobdella livanowi. The reconstruction of phylogeny based on these sequences revealed the congruence of our results with one of the existing phylogenetic schemes. Pairwise distances of 2% between geographically distinct samples in the A. peledina lineage suggest the lack of cryptic diversity within this species. P-distances between representatives of species A. peledina and P. livanowi are 14%. Assuming the rate of species divergence is approximately the same for all annelids, the values of interspecific p-distances within Acanthobdellida are lower compared to those of the most studied Amynthas. This fact casts doubt on the validity of the genus Paracanthobdella.**

Keywords — mitochondrion, complete genome, Annelida, Acanthobdellida, phylogeny.

Introduction

Acanthobdellida (Annelida, Clitellata) is a unique order of leech-like parasires that combine «archaic» traits of oligochaete worms with more «progressive» traits of true leeches (Hirudinea). Due to this combination a proposal was put forward that Acanthobdellida is a link in evolution from Oligochaeta and Hirudinea [1]. Despite the existence of an alternative hypothesis on acanthobdellids's belonging to oligochaetes, and convergent evolution of leech-like traits [2], the first hypothesis was more generally accepted. Early attempts at molecular methods application were in line with this suggestion and demonstrated A. peledina confidently clustering in between Oligochaeta and Hirudinea clades [3]. However, these studies were based on poor data (e.g., short and contaminated sequences) as was later acknowledged by their authors who, nevertheless, came to similar conclusions with newer analyses [4]. Later, the use of anchored hybrid enrichment technology yielded a completely different phylogeny, placing A. peledina deep within Hirudinea clade as a sister taxon to Piscicolidae group also parasitizing fish [5].

1) The taxonomy of Acanthobdellida remained controversial throughout the 20th century. This order was considered to be a typical member of subclass Hirudinea or an independent branch by different authors. The dispute is believed to be settled by Madill et al. who split Annelida into Oligochaeta, Acanthobdellea, Branchiobdellida and Hirudinea [6]. However, the question of interrelationship between closely related species A. peledina and P. livanowi remains still open. To date, there was no molecular data available on P. livanowi. This species, which lives in the fresh waters of Kamchatka [7], was discovered by Epstein as the only sister species to A. peledina, widely distributed in the Arctic. Later, V. Epstein singled out a separate family for the Kamchatka species.[8]. The doubts about this decision are still present.

The purpose of this study was to clarify acanthobdellids' evolutionary position and to see for the first time the level of divergence between A. peledina and P. livanowi, basing on complete mitochondrial sequences.

Materials and methods

Acanthobdellids specimens were collected in the northen regions of Eurasia (repubics of Komi and Soha, and Irkutsk region) on salmon fish and preserved in 70% alcohol. Each individual was wiped with cotton swabs and rinsed in fresh ethanol to remove mucus and dissected to remove guts with its contents to prevent contamination by host DNA. Total DNA was extracted with DiaGene Cell and Tissue kit and stored in deionized water. DNA samples were subsequently sheared to 300 bp size on Covaris M220 sonicator. Illumina libraries were prepared using NEBNext Ultra II DNA Library Prep Kit and sequenced with the use of Illumina NextSeq 550 Mid Output Kit v2.5. De-novo genome assembly was performed in MIRA v5.0 [9] and additionally confirmed with MitoZ [10]. The obtained sequences were aligned against GenBank sequences with MARS [9] and Muscle [11] aligners in succession. Bayesian phylogeny was reconstructed using BEAST2 [12] with parameters recommended by jModelTest 2 [13].

Results and discussion

The assembled Acanthobdellida mitochondrion genomes were aligned against homological data of the closest taxa (Polychaeta, Oligochaeta and Hirudinea) uploaded from the GenBank.. Due to the presence of partial sequences and large number of indels, 8015 out of 19723 positions were excluded from the resulting alignment by BEAST. This did not affect the pairwise distances between Acanthobdellida samples, which remained unchanged before and after exclusion: 2%

between all A. peledina samples and 14% between A. peledina and P. livanowi samples. These distances are lower than 9-20% distances between some Amynthas specimens, which cast doubt on the validity of the Paracanthobdella genus. The BI-phylogeny revealed three main lineages corresponding to polychaetes (P), oligochaetes (O), and leeches together with leech-like worms (A+H). As seen on Figure, Acanthobdellida formed a sister group to Hirudinea. High posterior probabilities of the most nodes on tree indicates high confidence of resulted phylogeny. Thus, our results suggest that the level of genetic differences between A. peledina and P. livanowi is insufficient to assume the existence of two separate genera, which is closer to Epstein's original classification.

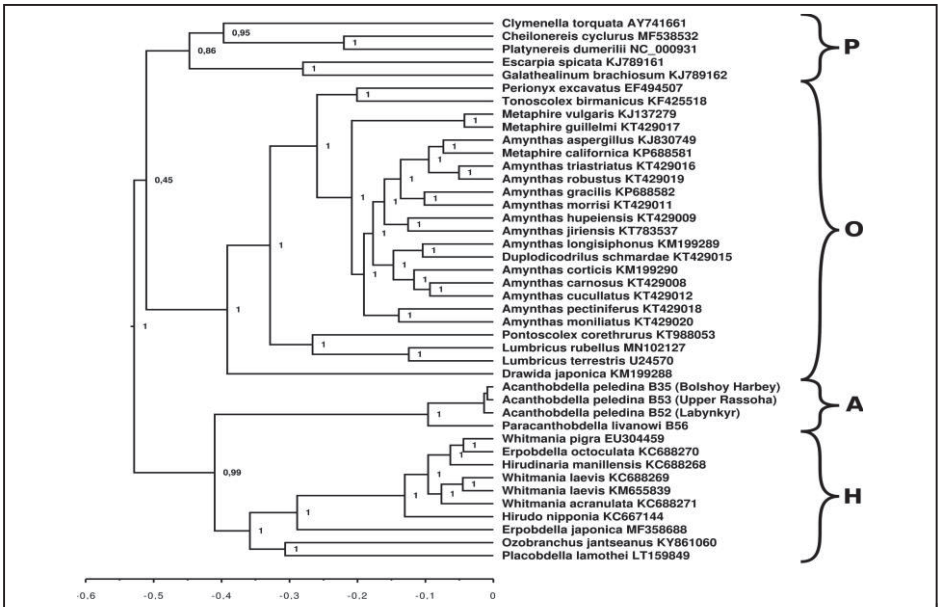


Fig. 1. Bayesian phylogeny based on complete mitochondrions with posterior probabilities. Curly brackets indicate borders of Polychaeta (P), Oligochaeta (O), Acanthobdellea (A) and Hirudinea (H) clades.

Acknowledgment

This work was funded by RFBR grants nos. 17-2905097, 19-34-50072 and 19-34-90011. The sequencing was performed in ICG Center of Genomic Investigation. The genome assemblies were conducted in Siberian SuperComputer Center.

References

[1] N. A. Livanow, "Acanthobdella peledina Grube," 1851 Morphol. Study. vol. 72(5–8), pp. 1–266. 1905.

[2] W. Michaelsen, "Uber die Beziehungen der Hirudineen zu den Oligochaeten," Mitt. Zool. Mus. Hamburg, vol 36. pp. 131–153, 1919.

[3] M. E. Siddal, K. Apakupakul, E. M. Burreson, K. A. Coates, C. Erséus, S. R. Gelder et al., "Validating Livanow: molecular data agree that leeches, branchiobdellidans, and Acanthobdella peledina form amonophyletic group of oligochaetes," Mol. Phylogenet. Evol., vol. 21(3). pp. 346–351, December 2001.

[4] M. Tessler, D. de Carle, M. L. Voiklis, O. A. Gresham, J. S. Neumann, S. Cios, M. E. Siddall, "Worms that suck: phylogenetic analysis of Hirudinea solidifies the position of Acanthobdellida and necessitates the dissolution of Rhynchobdellida," Mol. Phylogenet. Evol., vol. 127, pp, 129–134, October 2018.

[5] A. J. Phillips, A. Dornburg, K. L. Zapfe, F. E. Anderson, S. W. James, C. Erséus, et al., "Phylogenomic analysis of a putative missing link sparks reinterpretation of leech evolution," Genome Biology and Evolution, vol. 11(11), pp. 3082–3093, November 2019.

[6] J. Madill, K. A. Coates, M. J. Wetzel, S. R. Gelder, "Common and scientific names of aphanoneuran and clitellate annelids of the Unuted States of America and Canada," Soil Biol. Biochem. Vol. 24(12), pp. 1259-1262. December 1992.

[7] V. M. Epstein, "Acanthobdella livanowi sp. n. – a new species of ancient leeches (Archihirudinea) from Kamchatka water bodies," Proceedings of AS USSR, vol. 168(4), pp. 955-958, 1966.

[8] V. M. Epstein, Leeches, in Identification Guide to Parasites of Freshwater Fish in the Soviet Union, Bauer, O., Ed., Leningrad: Nauka, 1987, pp. 340—372.

[9] B. Chevreux, T. Wetter, S. Suhai, "Genome Sequence Assembly Using Trace Signals and Additional Sequence Information," Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) 99, pp. 45-56. 1999.

[10] G. Meng, Y. Li, C. Yang, S. Liu, "MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization," Nucleic Acids Research, vol. 47(11), e63 June 2019.

[11] L. A. K. Ayad, S. P. Pissis, "MARS: improving multiple circular sequence alignment using refined sequences," BMC Genomics, vol. 18, e86, 2017.

[12] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," Nucleic Acids Res., vol. 32(5), pp. 1792-1797, 2004.

[13] R. Bouckaert, J. Heled, D. Kühnert., T. Vaughan, C. Wu, D. Xie et al., "BEAST 2: a software platform for bayesian evolutionary analysis," PLoS Comput. Biol., vol. 10(4), e1003537, 2014.

[14] D. Darriba, G. L. Taboada, R. Doallo, D. Posada, "jModelTest 2: more models, new heuristics and high-performance computing," Nat. Methods, vol. 9(8), e772. July 2012.

DOI 10.18699/BGRS/SB-2020-130

Genetic diversity of the flat leeches (Hirudinea, Glossiphoniidae) in Western Siberia

Nadezhda Bolbat
Irkutsk State University
Irkutsk, Russia
nadya.mandzyak@mail.ru

Lyudmila Fedorova
Surgut State University
Surgut, Russia
ludiko@list.ru

Irina Kaygorodova
Limnological institute SB RAS
Irkutsk, Russia
irina@lin.irk.ru

Abstract — **The poor knowledge of the glossiphoniid fauna of Western Siberia is a consequence of the unresolved taxonomy of this group. An incomplete description of the species due to the phylogenetic ambiguity of external morphological characters and high intraspecific variability complicated the identification of these leeches. Using molecular genetic methods in combination with classical morphological analysis, we proved the presence of 7 species of the family Glossiphoniidae in Western Siberia, with interspecific genetic pairwise distances varying from 3.7 to 18.2%. In addition to Theromyzon tessulatum, Helobdella stagnalis and Hemiclepsis marginata, previously detected in the region, four potentially new species (Glossiphonia sp. 1, Glossiphonia sp. 2, Glossiphonia sp. 3, and Alboglossiphonia sp.) were found. The results obtained for the Western Siberian leeches once again confirm the advantage of the DNA barcoding application, as one of the molecular methods for species delimitation, used in an integrated approach to the study of taxonomic composition of a fauna.**

Keywords — species delimitation, mtCOI, DNA barcoding, leeches, Western Siberia.

Introduction

Glossiphoniidae is a family of parasitic annelids that plays an important role in the environment, providing nutrition to aquatic vertebrates [1], being an indicator of environmental stress [2; 3], a regulator of the number of hydrobionts, being parasites of mainly mollusks [4], as well as carriers of helminths and blood parasites of vertebrates [3].

The history of the study of flat leeches in Western Siberia originates in the first half of the 20th century [5; 6; 7]. Recent study of the Ob River macrozoobenthos have revealed six species of flat leeches Glossiphonia complanata (L., 1758), Alboglossiphonia heteroclita (L., 1761), Glossophonia concolor (Apáthy, 1888), Helobdella stagnalis (L., 1758), Hemiclepsis marginata (Müller, 1774), Theromyzon tessulatum (Müller, 1774) [8]. Earlier, in the Kazakhstan part of the Irtysh River Basin, five species of flat leeches were recorded - H. stagnalis, H. marginata, G. complanata, T. tessulatum and Alboglossiphonia sp.; moreover, 3 distinct morphotypes were found within G. complanata [9].

The faunistic studies of flat leeches are complicated because of lack of distinct taxonomic system. Morphologic species distinctions are mainly based hard-to-detect anatomic details of digestive and reproductive systems structure. The scarcity of external morphological features and, at the same time, high intraspecific variability makes it difficult to determine the taxonomic informativity of morphological diagnostics.

To assess the boundaries of taxa, as well as for new species identification, DNA barcoding is increasingly being used [e.g., 10; 11; 12]. The method of DNA barcoding [13] is based on a comparison of the mtCOI gene fragments between homologous sequences, which make it possible to accurately identify species and stable subspecies clusters. Recent studies have shown that genetic barcodes successfully distinguish morphological species in different groups of Glossiphiniidae [14; 15; 16].

Based on the results of similar studies, we decided to use DNA barcoding in conjunction with classical morphological analysis to delimit species of flat leeches from Western Siberia.

Materials and methods

Biological material was collected in freshwater bodies of the Irtysh and the Ob Rivers Basins in 2014-2018. Morphological analysis of flat leeches was carried out in accordance with modern zoological identification keys [1, 17].

Molecular analysis was conducted though a standard phenol-free DNA extraction with subsequent amplification with universal mtCOI primers. Sanger's sequencing was performed at the «Syntol» Company (Moscow, Russia).

Phylogeny reconstruction was carried out with the Tamura-Nei model using the Maximum likelihood (ML) method implemented in MEGA 7.0 [18].

Results and discussion

The COI-sequences of the Folmer's fragment with length of 709 bp were obtained for 54 flat leeches from different reservoirs of Western Siberia, together with the 65 closest homologous sequences from GenBank, were included in the analysis as a comparison group. Sequences of the genera Batracobdella and Placobdella were used as an outgroup.

Sequence analysis revealed that the Western Siberian flat leeches spited into 7 phylogenetic lineages (Fig. 1), genetic pairwise distances between which vary from 3.7% to 18.2%. These values exceed the threshold value for interspecific variability, which may indicate their independent species status taking into account the low level of genetic polymorphism (0.1-0.3 %) within each lineage.

Basing on the COI-phylogeny (Fig. 1), it can be argued that 3 species, which morphologically fit the G. complanata description, are actually distinct species - Glossiphonia sp. 1, Glossiphonia sp. 2, Glossiphonia sp. 3, which indicates the presence of a cryptic species complex. P-distances between them and the G. complanata clade are 8-11 %. Nucleotide sequences of the Western Siberian Alboglossiphonia sp. formed in a separate branch, distant from representatives of the congeneric European fauna, with a significant genetic distance of 12.6%.

The wide-spread Palaearctic species T. tessulatum, H. stagnalis and H. marginata are clustered according to their

morphologic classification. P-distances within each of these species-rank clusters are not higher of 0.4 %, that corresponds to intraspecific variability according to the DNA barcoding theory [13].

Thus, the use of DNA barcoding allowed us to clarify the species composition of the fauna of flat leeches in Western Siberia and to discover four new species, including a cryptic complex consisting of three genetically distinct species, morphologically indistinguishable from G. complianta. Our study confirms the effectiveness of an integrated approach for taxa delimitation. We find flat leeches to be an exemplary object to demonstrate the benefits of the DNA barcoding technique.

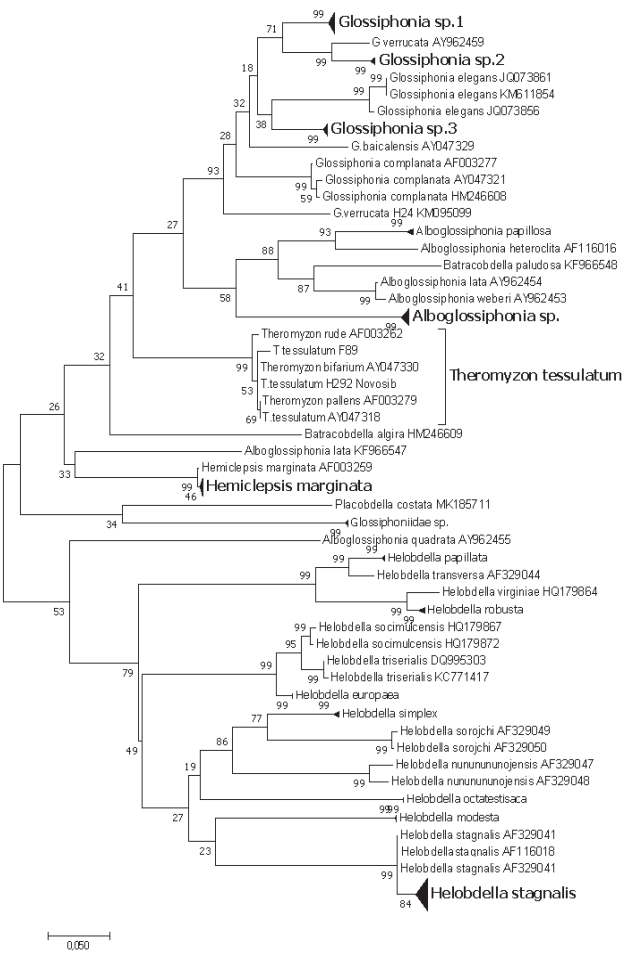


Fig. 1.ML-tree (lnL =-20102,61), based on 119 COI nucleotide sequences

Acknowledgment

This work was supported by RFBR grants. 17-29-05097. The authors are grateful to E.A. Fedorova A.Yu. Suleymanov for help in collecting biological material for research.

REFERENCES

[1] E.I. Lukin, "Leeches of fresh and saline waters (fauna of the USSR. Leeches)," Nauka, Leningrad, pp. 284, 1976.

[2] C. Grosser, B. M. Gerald, D. Heidecke, "Untersuchungen zur Eignung heimischer Hirudineen als Bioindikatoren für Fließgewässer," Hercynia, Vol. 34, pp. 101-127. 2001.

[3] P. Trivalairat, K. Chiangkul, W. Purivirojkul, "Placobdelloides sirikanchanae sp. nov., a new species of glossiphoniid leech and a parasite of turtles from lower southern Thailand (Hirudinea, Rhynchobdellida)," ZooKeys, Vol. 882, pp. 1-24, 2019.

[4] I. N. Bolotov A. L. Klass, M. V. Vinarski, "Freshwater mussels house a diverse mussel-associated leech assemblage," Scientific Reports 9:16449, 2019.

[5] W. Michaelsen, "Eine neue Haplotaxiden-Art und andere Oligochaeten aus dem Telezkischen See im nordlichen Altai, in Verhandlungen des Naturwissenschaft lichen," Vereins zu Hamburg, vol. 3, no. 10, pp. 1–7, 1903.

[6] I. Ioffe, "The benthic fauna of the Ob–Irtysh basin and its fishery significance", Izv. Vses. Nauchno-Issled. Inst. Ozern. Rechn. Rybn. Khoz., vol. 25, no. 1, pp. 113–161, 1947.

[7] I. I. Malevich, "On the fauna of oligochaetes of Lake Teletskoye," Tr. Zool. Inst., vol. 7, no. 4, pp. 119–123, 1949.

[8] E. N. Krylova, D. M. Bezmaternykh, "Fauna, Spatial Distribution, and Ecological Peculiarities of Oligohaetes (Oligochaeta) and Leeches (Hirudinea) in the Ob' River Basin," Biology Bulletin, Vol. 46(9), pp. 1001-1011, 2019.

[9] I. A. Kaygorodova, L. I. Fedorova, "The first data on species diversity of leeches (Hirudinea) in the Irtysh River Basin, East Kazakhstan," Zootaxa, 4144 (2), pp. 287-290, 2016.

[10] K. W. Will, D. Rubinoff, "Myth of the molecule: DNA-barcodes for species cannotre place morphology for identification and classification," Cladistics, V. 20, pp. 47–55, 2004.

[11] D. Pauls, M. Selcho, N. Gendre, F. R. F. Stocker, A. S. Thum, "Drosophila Larvae Establish Appetitive Olfactory Memories via Mushroom Body Neurons of Embryonic Origin," Journal of Neuroscience, 30(32), pp. 10655-10666, 2010.

[12] K. Kiontke, M.-A. Félix, M. Ailion, M. V. Rockman, C. Braendle, J.-B. Pénigault, D. H. Fitch, "A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits," BMC Evol. Biol., Vol.11, pp,339, 2011.

[13] P. D. N. Hebert, A. Cywinska, S. L. Ball, J. R. de Waard, "Biological identifications through DNA barcodes," Proc. R. Soc. Lond, B. 270, pp. 313-321, 2003.

[14] U. Kutschera, H. Langguth, D.-H. Kuo, D. A. Weisblat, M. Shankland, "Description of a new leech species from North America, Helobdella austinensis n. sp. (Hirudinea: Glossiphoniidae), with observations on its feeding behavior," Zoosystematics and Evolution, Vol.89 (2), pp. 239-246, 2013.

[15] M. Reyes-Prieto, A. Oceguera-Figueroa, S. Snell, "DNA barcodes reveal the presence of the introduced freshwater leech Helobdella europaea in Spain," Mitochondrial DNA, Vol. 25(5): pp. 387-393, 2014.

[16] I. A. Kaygorodova, N. B. Mandzyak, "Molecular phylogeny of Siberian Glossiphoniidae (Hirudinea)," Molecular Biology, Vol. 48(3), pp. 452-455, 2014.

[17] H. Nesemann, E. Neubert, "Clitellata, Branchiobdellada, Acanthobdellada, Hirudinea ," Susswasser fauna von Mitteleuropa. Heidelberg, Berlin, Spectrum Akademischer Verlag, V.6 (2), pp.1-178, 1999.

[18] [18] S. Kumar, G. Stecher, K. Tamura, "MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets," Molecular Biology and Evolution, Vol. 33, pp. 1870-1874, 2016.

DOI 10.18699/BGRS/SB-2020-131

The genomes and mechanisms of adaptation to the cold climates in Russian native cattle breeds

Laura Buggiotti
Royal Veterinary College, University of London, London, UK
Institute of Cytology and Genetics, Novosibirsk, Russia

Andrey Yurchenko
Kurchatov Genomics Center, Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia

Denis M. Larkin
Royal Veterinary College, University of London, London, UK
Institute of Cytology and Genetics, Novosibirsk, Russia

Nikolay Yudin
Kurchatov Genomics Center, Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia

Keywords — adaptation, cattle, missense mutation

The genetic structure, history and signatures of selection were revealed using the sequenced genomes of two native Russian cattle breeds: the Yakut cattle and the Kholmogory. Twenty individuals from each breed have been resequenced to ~10-11x filtered coverage and combined with the genome sequences of two most closely related cattle breeds: Hanwoo and Holsten. Our combined dataset contained over 31 million high-quality SNPs. This set has been used to detect copy number variants (CNVs) in the four breed genomes. The CNVs found in the Yakut and breed were enriched in genes from the gene ontology terms *regulation of actin filament-based processes*, *MHC protein complex*, and *antigen processing and presentation* could potentially contribute to adaptation to local environments. There were 299 CNVRs shared between the Yakut and Kholmogory breeds, of which the *ACTG1* (actin gamma 1) was the most frequent CNVR and covered >90% of the gene. Our SMC analysis had demonstrated a different N_e trajectory for the Yakut cattle compared to the other three breeds and additional Chinese and Indian indicine cattle breeds, while the Kholmogory's one was close to Holstein. The demographic history of the Yakut cattle was further analyzed with the TreeMix, *f3* and *D4* statistics and the RFMix. The TreeMix results indicated no admixture between the Yakut cattle and other cattle breeds and related bovinae species. While the RFMix data suggested introgression of indicine cattle in to the Yakut cattle genomes. We further looked into the Yakut genome regions likely being introduced from other cattle populations. We found that the corresponding SNP alleles often are present in other bovinae genomes, with very few of them were indicine-cattle specific. This suggests that the Yakut cattle could have genome regions containing ancestral alleles, removed from the European taurine genomes. We looked into the high-frequency missense mutation found in the Yakut cattle. Out of 48 such mutations 21 were found in the *SPTBN5* gene. Spectrin 5 is an actin-binding gene involved in resistance to parasites in humans, suggesting that it could be related to parasite resistance in the Yakut cattle as well. Spectrin 5 region had the highest F_{ST} between the Yakut cattle and over 3000 cattle individuals from the 1000 bull genomes project. The only exclusive Yakut-specific missense change absent from any other cattle breeds and other bovinae species was found in the *NRAP* gene. Interestingly, the same amino acid change observed in the Yakut cattle was observed in the vast majority of mammals that enter either torpor or hibernation suggesting that this gene, expressed exclusively in the heart and skeletal muscles could be related to mechanisms of slowing metabolism in the Yakut cattle during the cold winter months. This an important step toward revealing the history, and signatures adaptation within the native Russian cattle made to facilitate application of genomic technologies to improve breeding procedures and to increase the effectiveness of genomic selection for livestock in Russia.

Acknowledgment

This study was funded in part by the RFBR grant 18-016-00185 and the RSF grant 19-76-20026.

DOI 10.18699/BGRS/SB-2020-132

The phenotypic manifestation of Wolbachia genetic diversity in host fitness

Elena V. Burdina
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
bella79@list.ru

Nataly Gruntenko
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
ORCID 0000-0003-3272-1518

Petr N. Menshanov
Laboratory of Functional
Neurogenomics
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
eternity@bionet.nsc.ru

Roman A. Bykov
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
bykovra@bionet.nsc.ru

Yury Y. Ilinsky
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
paulee@bionet.nsc.ru

Inga Yu. Rauschenbach
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
iraushen@bionet.nsc.ru

***Abstract* — The maternally inherited bacterial endosymbiont of arthropods Wolbachia is deeply integrated into the biology of its host and can be regarded as a component of the host cell. There is evidence that *Wolbachia* contributes to the fitness of the host species, including the increase in heat stress resistance and resistance to viral infections. However, the exact mechanisms of this contribution, as well as *Wolbachia*-host interactions in total, remain largely unclear. In order to study these mechanisms we composed several *Drosophila melanogaster* lineages with the same nuclear background (wild type *Bi90* line) and different *Wolbachia* genotypes in cytoplasm. Here we showed that the symbiont did not effect on host insect heat stress resistance and fertility with the exception of two unique *Wolbachia* strains: the well-known pathogenic *wMelPop* strain and first discovered here *wMelPlus*.**

Keywords* — *Wolbachia, Drosophila melanogaster, Heat stress, Stress resistance, Juvenile hormone, Fecundity.

Introduction

The *Wolbachia pipientis* is a intracellular α-proteobacterium, which is found in 40% of terrestrial arthropods and is one of the most common prokaryotic invertebrate symbionts [1]. *Wolbachia* is transmitted maternally and can manipulate the reproductive system of the host in order to promote its spread in the host population [2]. At the same time, the infection with *Wolbachia* can be useful for the host [3, 4]. Differences in phenotypic manifestations of infection can be associated with peculiarities of the host organism physiology, including the processes of endocrine regulation of the growth, development, and adaptation [4–6]. However, the mechanisms underlying *Wolbachia*–host interactions, as well as mechanisms by which *Wolbachia* promotes the host organism adaptation, remain poorly understood.

According to the phylogeny reconstruction of full genome sequences, the symbiont diversity in *D. melanogaster* includes the IVI and VIII clades [7, 8]. In terms of polymorphism of certain genome markers the six genotypes were revealed, i.e. *wMel*, *wMel2*, *wMel3*, *wMel4*, *wMelCS*, *wMelCS2* [9, 10]. Regarding the *Wolbachia* effect or its source (fly stock), several strains were investigated but one of them, *wMelPop*, is the most interesting, especially in our discourse. *Wolbachia* diversity can be reduced to unique, monophelytic *wMel* strain [11–13], which can be divided into several genotypes. The *wMel* genotype includes I-V and VIII clades and is the most widespread [7, 9, 14]. The *wMel2* genotype belongs to VIII clade, *wMel4* – to III clade, *wMelCS* and *wMelCS2* belong to VI clade [7-9, 14]. *wMelPop* strain is a variant of *wMelCS* genotype, which is a virulent *Wolbachia* variant that over-replicates massively and shortens the lifespan of its fruit fly host. [15]. *wMelPop* is, hence, an exceptional vertically transmitted symbiont [16]. Here, we found the another unique *Wolbachia* strain crucially altering the host fitness but "with an plus sign" – increasing it.

Results and discussion

To study the effects of various *Wolbachia* genotypes on *D. melanogaster* fitness, we have created a study system that consists of five conplastic lineages carrying the nuclear background of one wild-type lineage, *Bi90*, and cytoplasmic backgrounds with *wMel*, *wMel2*, *wMel4*, *wMelCS* and *wMelPop* genotype variants of *Wolbachia* [17]. *Bi90* strain was established from wild-caught female of “Bishkek 2004” population and interbred for more than 300 generations, thereby it could be considered a nearly isogenic line. This strain was earlier characterized by *Wolbachia* infection (*wMel*) [9, 18, 19]. One pair of flies from *Bi90* strain was isolated to get *Bi90* branch, which was treated with tetracycline for 3 generations to make *Wolbachia*-free *Bi90T* strain [4, 14] and was further used in making conplastic strains and as a control in experiments. *Wolbachia* donor strains were also characterized for infection (*wMel2*, *wMel4*, *wMelPop* and *wMelCS*) [9, 10, 20]. Two independent runs were performed to make each conplastic strain, and finally two strains of ‘certain *Wolbachia’*-cytoplasmic/*Bi90* nuclear background were created.

We demonstrated that two of five investigated *Wolbachia* variants promote changes in *Drosophila* heat stress resistance [17]. What is especially interesting, *wMelCS* genotype of *Wolbachia* increases stress resistance, whereas *wMelPop* strain decreases it. *wMel*, *wMel2* and *wMel4* genotypes of *Wolbachia* do not show any effect on the survival under heat stress [17].

We also found out that the type of *Wolbachia* infection determines the effect on *Drosophila melanogaster* fecundity [21]. The *wMelCS* genotype of *Wolbachia* decreases egg production in infected *D. melanogaster* females in the beginning of oviposition and increases it later (from the sixth day after eclosion), whereas the *wMelPop Wolbachia* strain causes the opposite effect. *wMel*, *wMel2* and *wMel4* genotypes of *Wolbachia* do not show any effect on this trait compared with uninfected *Bi90 D. melanogaster* females [21].

Thus, we found out that *wMelCS* genotype looks like "antipode" to *wMelPop* strain. We wondered if these characteristics were typical to *wMelCS* genotype on the whole, or we had run into a unique strain similar to *wMelPop* only "with a plus sign". To clarify this question we created six more conplastic lineages carrying the *Bi90* nuclear background and cytoplasmic backgrounds with *wMelCS* genotype of *Wolbachia* of different origins.

Testing the viability under heat stress in these lineages carrying different *wMelCS* strains we showed that the one studied previously is truly unique: no other strain of *wMelCS* genotype under study showed any effect on survival under heat stress. So we named this unique *Wolbachia wMelCS* strain *wMelPlus*.

Conclusions

Based on the data obtained, we can conclude that in the vast majority of cases the *Wolbachia* endosymbiont does not affect the fitness of host insect in general. However, a bacteria variant that sharply increases or decreases host fitness can appear as a result of mutation.

Acknowledgment

This research was supported by the Russian Foundation for Basic Research (Project # 20-04-00579).

References

[1] R. Zug, and P. Hammerstein, "Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected", PLoS One, 2012, vol. 7, no. 6. e38544. https://org/10.1371/journal.pone.0038544.

[2] J. H. Werren, L. Baldo, and M. E. Clark, "*Wolbachia*: master manipulators of invertebrate biology", Nat. Rev. Microbiol., 2008, vol. 6, no. 10, pp. 741—751. Doi 10.1038/nrmicro1969.

[3] A. R. Weeks, M. Turelli, W. R. Harcombe, K. T. Reynolds, A. A Hoffmann, "From parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila"*, PLoS Biol., 2007, vol. 5, no. 5. https://org/10.1371/journal.pbio.0050114.

[4] L. Teixeira, A. Ferreira, and M. Ashburner, "The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster"*, PLoS Biol., 2008. https://org/10.1371/journal.pbio.1000002.

[5] T. Ikeya, S. Broughton, N. Alic, R. Grandison, L. Partridge, "The endosymbiont *Wolbachia* increases insulin/IGF-like signalling in *Drosophila"*, Proc. R. Soc. B, 2009, vol. 276, (1674), pp. 3799—3807. doi 10.1098/rspb.2009.0778.

[6] V. G. Faria, N. E. Martins, S. Magalhaes, T. F. Paulo, V. Nolte, Ch. Schlötterer et al, "*Drosophila* adaptation to viral infection through defensive symbiont evolution", PLoS Genet., 2016, vol. 12, no. 9. e1006297. https://org/10.1371/journal.pgen.1006297.

[7] M. F. Richardson, L. A. Weinert, J. J. Welch, R. S. Linheiro, M. M. Magwire, F. M. Jiggins et al., "Population genomics of the *Wolbachia* Endosymbiont in *Drosophila melanogaster"*, Kopp A, ed. PLoS Genet, 2012, 8(12):e1003129.

[8] A. M. Early, A. G. Clark, "Monophyly of *Wolbachia pipientis* genomes within *Drosophila melanogaster*: geographic structuring, titre variation and host effects across five populations", Mol Ecol, 2013, 22:5765–78.

[9] Y. Y. Ilinsky, "Coevolution of *Drosophila melanogaster* mtDNA and *Wolbachia* genotypes", PLoS One, 2013, 8(1):e54373.

[10] M. Riegler, M. Sidhu, W. J. Miller, S. L. O'Neill, "Evidence for a global *Wolbachia* replacement in *Drosophila melanogaster"*, Curr Biol, 2005, 15:1428–33.

[11] C. Paraskevopoulos, S. R. Bordenstein, J. J. Wernegreen, J. H. Werren, K. Bourtzis, "Toward a *Wolbachia* multilocus sequence typing system: discrimination of *Wolbachia* strains present in *Drosophila* species", Curr Microbiol, 2006, 53:388–95.

[12] L. Baldo, J. H. Werren, "Revisiting *Wolbachia* supergroup typing based on WSP: spurious lineages and discordance with MLST", Curr Microbiol, 2007, 55:81–7.

[13] V. I. D. Ros, V. M. Fleming, E. J. Feil, J. A. J. Breeuwer, "How diverse is the genus *Wolbachia*? Multiple-gene sequencing reveals a putatively new *Wolbachia* supergroup recovered from spider mites (Acari: Tetranychidae)", Appl Environ Microbiol, 2009, 75:1036–43.

[14] E. Chrostek, M. S. P. Marialva, S. S. Esteves, L. A. Weinert, J. Martinez, F. M. Jiggins et al., "*Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis", Malik HS, ed. PLoS Genet, 2013, 9(12):e1003896.

[15] K. T. Min, S. Benzer, "*Wolbachia*, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death", Proc Natl Acad Sci U S A, 1997, 94:10792–6.

[16] J. H. Werren, "*Wolbachia* run amok", Proc Natl Acad Sci U S A, 1997, 94: 11154–5. pmid:9326576.

[17] N. E. Gruntenko, Y. Y. Ilinsky, N. V. Adonyeva, E. V. Burdina, R. A. Bykov, P. N. Menshanov et al., "Various *Wolbachia* genotypes differently influence host *Drosophila* dopamine metabolism and survival under heat stress conditions", BMC Evol Biol, 2017, 17(Suppl 2):252.

[18] Y. Y. Ilinsky, I. K. Zakharov, "Cytoplasmic incompatibility in *Drosophila melanogaster* is caused by different *Wolbachia* genotypes", Russ J Genet, 2011, 1:458.

[19] N. Y. Waisman, M. D. Golubovsky, Y. Y. Ilinskii, "Differences in the parameters of longevity and its sex-specificity in human populations and modeling them in *Drosophila"*, Adv Gerontol, 2013, 3(4):268–76.

[20] E. Chrostek, L. Teixeira, "Mutualism breakdown by amplification of *Wolbachia* genes", Malik HS, ed. PLoS Biol, 2015, 13(2):e1002065.

[21] N. E. Gruntenko, E. K. Karpova, N. V. Adonyeva, O. V. Andreenkova, E. V. Burdina, Y. Y. Ilinsky et al., "*Drosophila* female fertility and juvenile hormone metabolism depends on the type of *Wolbachia* infection", JEB, 2019, 222, jeb195347. doi:10.1242/jeb.195347.

DOI 10.18699/BGRS/SB-2020-133

Application of ITS1 and ITS2 for population genetic studies of sturgeons (Acipenseridae)

Guzel Davletshina
IMCB SB RAS, Novosibirsk, Russia
ICG SB RAS, Novosibirsk, Russia
guzel@mcb.nsc.ru

Sergey Kliver
IMCB SB RAS, Novosibirsk, Russia
skliver@mcb.nsc.ru

Elena Interesova
TSU, Tomsk, Russia
e.interesova@ngs.ru

Dmitry Prokopov
ICG SB RAS, Novosibirsk, Russia
dprokopov@mcb.nsc.ru

Vladimir Trifonov
IMCB SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
vlad@mcb.nsc.ru

***Abstract* — The order Acipenseridae is a very interesting group for evolutionary genetics: all species have unique morphology, inter-specific hybrids are widely occurring and there are variations between species in ploidy levels. Most acipenserids are endangered due to poaching and special efforts are required for the maintenance of natural populations. The genetic studies of acipenserids are still limited, although these are needed for successful farming. ITS – is the DNA spacer located between the small subunit and large subunit rRNA genes. The genes encoding ribosomal RNAs are located one after another in tandem and are repeated several hundred times, so we use new generation sequencing to estimate the frequency of occurrence of SNPs in the genome of one organism. ITS1 and ITS2 are used as phylogenetic markers to study the relationships between highly diverged taxonomic groups [1]. Despite high interest to different sturgeon species, acipenserid ITS1 and ITS2 sequences are missing in the GenBank depository, and most sturgeon population studies are performed using mitochondrial markers. Here we study the structure of ITS1 and ITS2 in several sturgeon species and demonstrate efficiency of these nuclear markers for species identification and interspecific hybrids confirmation.**

Keywords* — *acipenser, phylogenetic markers, ITS, NGS, polyploidy, interspecific hybrids.

Motivation and aim

Motivation

ITS1 and ITS2 are phylogenetic markers, widely used for many species of animals, plants, fungi and bacteria, but these important nuclear markers have never been applied for sturgeons. As stugreons are paleopolyploids, their microsatellite analysis is complicated, and mitochondrial markers characterize only maternal lineages.

Aim

Here we study the structure of ITS1 and ITS2 in *Acipenser baerii*, *A. ruthenus* and identify SNPs specific to different species and their populations. Besides, we demonstrate that this nuclear marker facilitates identification of interspecific hybrids.

Methods

Primers for PCR amplification of sturgeon ITS1 and ITS2 were designed using genome sequence of *A.ruthenus*. Obtained PCR products from 21 individual were then sequenced on the Illumina platform. Raw sequencing data were preprocessed using custom pipeline followed by variant calling by Samtools.

Results

We generated consensus sequences of ITS1 and ITS2 for sturgeon species: *A.baerii* and *A.ruthenus*. We found six single nucleotide substitutions differentiating these two sturgeon species. We sequenced ITS1 and ITS2 of interspecific hybrids between *A.baerii* and *A.ruthenus*, and found that indeed there were SNPs in the sequence characteristic for both ITS1 and ITS2. Based on NGS data analysis, we demonstrated that *A.ruthenus* specific SNPs were represented in around 33% reads, while *A.baerii* specific SNPs were found in 67%, which corresponds to firgt generation hybrids subgenome ratio (as Siberian sturgeon genome is twice as large as sterlet genome) (Table 1).

Table 1. The mean frequency of variant occurrence in the reads for all samples characteristic of ITS1 and ITS2 SNPs in two species of sturgeons and their hybrids

Species, number of individuals		***A. baerii**, n=7**		***A. ruthenus**, n=11**		***A.baerii+A.ruthenus*, n=3**	
number of chromosomes		***240***		***120***		***120+60***	
ITS1	1934	**G(0,974)**	A(0,025)	G(0,005)	**A(0,995)**	G(0,668)	A(0,332)
	2454	**T(0,974)**	C(0,025)	T(0,021)	**C(0,979)**	T(0,682)	C(0,318)
ITS2	2898	**T(0,747)**	C(0,253)	T(0,002)	**C(0,998)**	T(0,481)	C(0,519)
	2917	**C(0,972)**	T(0,028)	C(0,002)	**T(0,998)**	C(0,655)	T(0,345)
	3065	**T(0,952)**	C(0,048)	T(0,002)	**C(0,998)**	T(0,579)	C(0,421)
	3103	**T(0,970)**	C(0,030)	T(0,042)	**C(0,958)**	T(0,625)	C(0,375)

*-reference allels for *A.baerii* and *A.ruthenus* are in bold. Allels from both parental species were detected in hybrids.

Acknowledgment

The work was supported by the Russian Science Foundation grant No. 18-44-04007, Guzel Davletshina was additionally funded by the budget project No. 0259-2019-0003-C-01

References

[1] Allard M.W. and Honeycutt R.L. (1991) Ribosomal DNA Variation Within and Between Species of Rodents, with Emphasis on the Genus Onychomys. Mol. Biol. Evol. 8: 71-84.

DOI 10.18699/BGRS/SB-2020-134

Analysis of the associations between missense substitutions in the human MT-ATP6 gene

Maria Golubenko
Laboratory of population genetics
Research Institute of Medical Genetics, Tomsk NRMC
Tomsk, Russia
maria.golubenko@medgenetics.ru

Alexey Zarubin
Laboratory of evolutionary genetics
Research Institute of Medical Genetics, Tomsk NRMC
Tomsk, Russia
aleksei.zarubin@medgenetics.ru

***Abstract* — Mitochondrial DNA encodes 13 subunits of mitochondrial respiratory chain complexes. There is high mtDNA variability in human populations, and substantial part of it is regional specific. It is hypothesized that mtDNA common variants can be adaptive and have some effect on the mitochondrial function. MT-ATP6 gene is particularly variable in human populations in regard of missense polymorphisms. We have analyzed MT-ATP6 missense substitutions in the mtDNA haplogroups which are common for Northern Asia (Siberia). Representatives of mitochondrial haplogroups A, D4, C4, C5, Z, and L (the latter as an example of a "Southern" haplogroup) were selected from the GenBank collection of whole human mtDNA sequences, and amino acid sequences encoded by ATP6 were analyzed. We show that amino acid variability in the ATP6 depends on the substitutions in the roots of these haplogroups. In particular, Ala59Thr in combination with His90Tyr substitution (haplogroup A) are associated with a number of polymorphic amino acids in the first 60 positions of ATP6, whereas ancestral Ala59 variant (haplogroup D), either alone or in the combination with Ala20Thr substitution (haplogroups C4, C5, and Z), seems to have "suppressing" effect on the variability in this region. Comparison of the estimated "coupling values" (which reflect associations between amino acids in the protein) between each of these three positions and all amino acid positions in the 1-59 region has revealed that mean coupling values were comparatively higher for the position 20, lower for the position 59, and the lowest for the position 90 (p=0.0002). In contrast, for the missense substitutions which actually have occurred in these haplogroups, the mean coupling values were close to 0 at all three positions. Our results demonstrate possible epistatic effect of the haplogroup-defining missense variants in the mtDNA encoded protein.**

Keywords* — *mitochondrial DNA, molecular phylogeny, adaptation

Introduction

Mitochondrial DNA (mtDNA) is widely used in the population and evolutionary genetics, due to its uniparental inheritance and high mutation rate. In human populations, mtDNA variation has been extensively studied, and detailed intraspecific phylogeny was reconstructed and is available online [1]. All mtDNA haplotypes can be classified as belonging to different haplogroups which share common nucleotide substitutions. Along with silent and non-coding substitutions, these haplogroup-specific polymorphisms include amino acid changes in mtDNA-encoded subunits of mitochondrial respiratory chain complexes, as well as nucleotide substitutions in the mitochondrial rRNA and tRNA genes. Additional polymorphisms occur at the terminal branches of the tree. MtDNA variation has substantial geographic differentiation, so that main branches of the tree are continent-specific. It was proposed that not the genetic drift alone, but natural selection as well could contribute into the global distribution of the mtDNA variation [2], and that some mtDNA haplogroups might have slightly increased or decreased respiratory chain efficiency, contributing to the adaptation to climate and/or food sources [3].

It is possible that the sequential accumulation of mutations along the mtDNA molecule in the course of human microevolution is not random, because of interactive and cumulative effects of several mutations, but such effects are difficult to estimate. The aim of our study was to search for possible patterns of amino acid replacements in the human mtDNA phylogeny, using the MT-ATP6 gene as an example.

Methods

For the analysis of the intraspecific phylogeny of human mtDNA, a data set was generated which contains all complete human mitochondrial genome sequences (48935 sequences) available in the NCBI Nucleotide database. These sequences were aligned to the rCRS reference sequence using bwa bwasw, and then a list of substitutions was obtained from the sequences. The detected substitutions were used for verification of the haplogroup using HaploGrep [4]. Sequences with low confidence of haplogroup affiliation (less than 0.9) were excluded from the analysis. Then, several haplogroups were chosen which had experienced expansion in Northeastern Eurasia (A, D4, C4, C5, Z), as well as one “Southern” haplogroup L3 which had expansion in Africa. For these haplogroups, amino acid sequences for ATP6 were translated from the DNA sequence data. List of amino acid replacements and their conservation indices and MutPred values were obtained using MtPhyl program [5]. In addition, estimates of mtDNA selection scores [6] and coupling values for amino acid pairs [7] were analyzed. Statistical comparisons were performed in the Statistica 8.0 package (StatSoft Inc.).

Results

ATP6 protein consists of 226 amino acids. The number of polymorphic amino acid positions in the studied data set was 51. Almost all sequences in all haplogroups have the “Thr112Ala” substitution (A8860G), where Alanine is in fact ancestral state, because the reference sequence (rCRS) belongs to the haplogroup H2a2a which has Ala112Thr substitution. Also, all haplogroups except of the haplogroup A have Thr59Ala substitution (A8701G), again with Alanine being the ancestral allele (the Ala59Thr change had occurred in the macrohaplogroup N, from which macrogaplogroup R originated, including H2a2 with the rCRS). Haplogroups C4, C5, Z, D4 belong to the macrogaplogroup M, which originated directly from L3 (the root for non-African haplogroups), therefore they also have this difference from the reference sequence. Haplogroup A belongs to N, so it has threonine in this position. However, almost all haplotypes of haplogroup A have a different substitution in the same part of the protein sequence, His90Tyr (C8794T). Representatives of the M branches (C4, C5, Z – all belonging to the M8 cluster) have another replacement in the N-terminus of ATP6 – Ala20Thr (G8584A).

The total number of polymorphic positions in each haplogroup should depend on its age and prevalence: the

probability of amino acid change is higher in the older and more abundant haplogroups, regardless of their adaptive significance. In the analyzed data set, haplogroup D4 was the most variable, with the coalescence time of about 26 thousand years, [8] and the number of the analyzed sequences (1323) was the largest. Analysis of the haplogroup D4 revealed 24 polymorphic amino acids. The coalescence time of haplotypes of haplogroup L3b was 12-15 thousands years ago [9], and for this cluster, 11 polymorphic positions were observed. The least variability in our dataset was demonstrated for the haplogroups Z and C5, with the coalescence time estimated as 16.5 and 20 thousand years ago [8], and the number of analyzed haplotypes is 175 and 120, respectively. Only one polymorphic amino acid was detected in haplogroup C5 (in addition to Ala20Thr), and in haplogroup Z there were 7 polymorphic amino acids. From the other side, the coalescence time for both haplogroups A (n=697) and C4 (n=570), was estimated at about 21 thousand years [8, 10], but there was significant difference in amino acid ATP6 variability: 22 polymorphic amino acids (in addition to His90Tyr) in the haplogroup A and only 4 (in addition to Ala20Thr) in the haplogroup C4. So, there is decreased protein variation in the haplogroup C4, versus the increased variability in the haplogroup A, despite similar age and prevalence.

When considering the pattern of the ATP6 amino acid sequence polymorphism in the studied haplogroups, it can be noted that the presence of threonine at position 59 (haplogroup A) compared with alanine (the remaining halpogroups) is associated with an increase in variability in the N-terminal part of the protein: in the region from position 1 to 59, there are 11 polymorphic amino acids in haplogroup A, compared to 6 in haplogroup D and 3 in haplogroup C4. It should be noted, however, that all C4 representatives in this area have a replacement Ala20Thr (occurred at the root of M8 haplogroup which encompasses the haplogroup C). The explanation for the observed picture may be the adaptive effect of the Ala59Thr substitution, which weakens the pressure of negative selection. Another explanation could be interaction of this amino acid with the N-terminal part of the protein, i.e. the epistatic effect of substitution in this position. An indirect assessment of epistatic effects can be given by the method of analysis of interactions published in [7]. In that study, based on a comparative analysis, data on the "coupling values" between amino acids at various positions for several thousands known proteins, including the sixth subunit of ATP synthase, were calculated, and the results have been made publicly available. It was assumed that in the absence of interaction between amino acids, the average value of this index should be close to zero. Using these data, we selected and compared the calculated estimates of the coupling values of positions 20, 59, and 90 (where haplogroup-specific amino acid substitutions occurred in the studied haplogroups) with all amino acids in the N-terminal part of the protein where we found haplogroup-specific variability (i.e. amino acids 16-58). The analysis of variance revealed statistically significant differences in the average coupling values between the three positions (p = 0,0002). Maximum average coupling value was estimated for the amino acid at the position 20, the intermediate value for the position 59, and the minimum for the position 90 (Fig.1). In contrast, mean coupling values for the positions 20, 59, 90 and actually occurred polymorphic amino acids did not differ from 0. Thus, the results suggest possible effect of the amino acid changes at the root of a haplogroup on the further protein variability. However, this question requires further research - in particular, a more complete analysis of other mtDNA haplogroups would allow to determine whether the detected association of threonine 59 with higher variability in the N-terminal part of the protein is preserved in the branches of N and R macrohaplogroups. In general, the high number of polymorphic positions in the ATP6 gene in the "northern" haplogroups D and A is consistent with the previously described phenomenon of high amino acid variability of ATP6 in the northern latitudes [2, 11]. The relatively low variability in the haplogroups C and Z can be associated with the presumable effect of the Ala20Thr substitution.

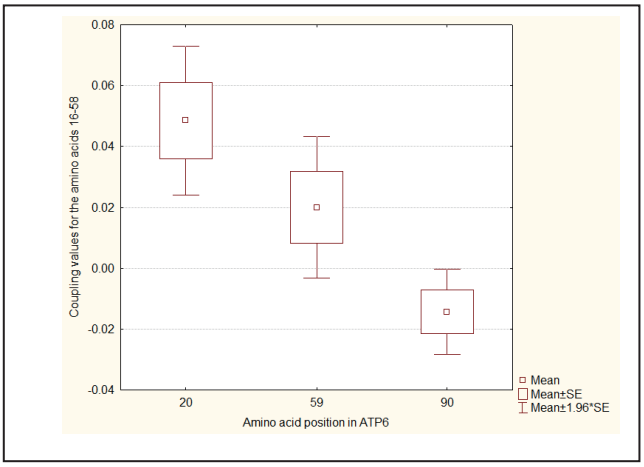


Fig. 1. Mean coupling values for the amino acids 20, 59, and 90 with the amino acids 16-58 in ATP6.

References

[1] M. van Oven and M. Kayser, "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat., vol. 30, pp E386-E394, 2009.

[2] E. Ruiz-Pesini, D. Mishmar, M. Brandon, V. Procaccio, D.C. Wallace, "Effects of purifying and adaptive selection on regional variation in human mtDNA," Science, vol. 303, pp. 223-226, January 2004.

[3] D. C. Walace, "Mitochondrial DNA Variation in Human Radiation and Disease", Cell, vol. 163, pp. 33-38, 2015.

[4] H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H. J. Bandelt, F. Kronenberg, A. Salas, S. Schönherr, "HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing," Nucleic Acids Res., vol. 44 (W1), pp. W58-63, July 2016.

[5] N. Eltsov and N. Volodko, "New method and new tool for human mtDNA phylogeny reconstruction," Europ. J. Hum. Genet., vol 17, suppl. 1, p. 298, May 2009 [Abstracts of the European Human Genetic Conference, Vienna, P 11.079

[6] L. Pereira, P. Soares, P. Radivojac, B. Li, D. C. Samuels, "Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity," Am. J. Hum. Genet., vol. 88, pp. 433-439, April 2011.

[7] T.A. Hopf, J.B. Ingraham, F.J. Poelwijk, C.P. Schärfe, M. Springer, C. Sander, D.C. Marks, "Mutation effects predicted from sequence co-variation," Nat. Biotechnol., vol. 35, pp. 128-135, Februalry 2017.

[8] M. Derenko, B. Malyarchuk, T. Grzybowski, G. Denisova, U. Rogalla, M. Perkova, I. Dambueva, I. Zakharov, "Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia," PLoS One, vol. 5, p. e15214, December 2010.

[9] P. Soares, F. Alshamali, J. B. Pereira, V. Fernandes, N. M. Silva, C. Afonso, M.D. Costa, E. Musilová, V. Macaulay, M. B. Richards, V. Cerny, L. Pereira, "The Expansion of mtDNA Haplogroup L3 within and out of Africa," Mol Biol Evol., vol. 29, pp. 915-927, March 2012.

[10] M. Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, Dorzhu C, Luzina F, Lee HK, Vanecek T, Villems R, Zakharov I. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. Am J Hum Genet. 2007 Nov;81(5):1025-41.

[11] M. Ingman and U. Gyllensten, Rate variation between mitochondrial domains and adaptive evolution in humans. Hum Mol Genet. 2007 Oct 1;16(19):2281-7.

DOI 10.18699/BGRS/SB-2020-135

mtDNA mammalian evolution: mice walk with many little steps while elephants with a few big

Dmitrii Iliushchenko
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal, University
Kaliningrad, Russia Federation
iliushchenkodmitrii@gmail.com

Anastasia Sokol
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal, University
Kaliningrad, Russia Federation
anastasia3sokol@yandex.ru

Konstantin Popadin
School of Life Sciences
Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal, University
Kaliningrad, Russia Federation
konstantinpopadin@gmail.com

Konstantin Gunbin
The Institute of Cytology and Genetics of the SB RAS
Novosibirsk, Russia Federation
genkvg@gmail.com

Abstract — **In this project, we use aligned mitochondrial genomes (353 species, 40 families, 13 genes) to test the hypothesis that, with a decrease in body weight, the rate of evolution will be faster due to very neutral substitutions, and with an increase in body weight there will be an increase in the rate of evolution overdue to more radical substitutions.**
Keywords — mitochondrial DNA, mammalians, evolution rate

Introduction

The ecological properties of species drive different patterns of molecular evolution. Numerous associations of ecological and genetic traits are especially well known for mammalian species - which are well characterized ecologically and deeply sequenced. It is known, for example, that small-bodied short-lived (hereafter 'mice') mammalian species have a higher rate of neutral evolution (i.e. accumulation of synonymous substitutions per the unit of time) as compared to large-bodied long-lived (hereafter 'elephants') species [1]. Most likely this effect is driven by the different number of DNA replications of germ cells per unit of time (higher in mice versus elephants). Also, it is known that slightly-deleterious variants, opposite to the neutral mutations, tend to accumulate faster in elephants versus mice because of their lower effective population size which leads to stronger genetic drift. Interestingly, both these most common categories of mutations (neutral and slightly-deleterious) together very often lead to an approximately constant rate of molecular evolution: mice accumulate numerous neutral (or close to neutral) substitutions, while elephants accumulate more rarer and more deleterious variants. Here we use mitochondrial genomes to test the hypothesis that in physicochemical space "mice walk with many little steps while elephants with a few big ones".

Methods

Mitochondrial genes, which we used in this project allow us to analyze hundreds of mammalian species with a complete sequenced mtDNA and hundreds of species with at least 10 available within-species sequences. We plan to find a physicochemical metrics of mtDNA evolution to describe and analyze mice-specific and elephant-specific patterns of amino-acid substitutions.

Using codon alignment of mitochondrial genes (353 species, 40 families, 13 genes) we generated all possible trios of species within each family. For each species from each trio, we obtained generation length as the main ecological factor, associated with effective population size. The approach is based on the relative ratio test [2]: we compare two species B and C with different generation length. If species B is more long-lived as compared to species C we expect that the rate of neutral evolution will be decreased during the B-specific evolution, while the rate of accumulation of slightly-deleterious variants will be increased. To test it we use an outgroup species A to analyze the pattern of evolution between A-B and A-C. Finally, we work with different outgroups for each pair of species to demonstrate the robustness of the observed effects. Additionally to comparative species scale, we analyze genetic polymorphisms in all mammalian species with sequenced al least 10 different organisms. We expect, that populations of large-bodied mammals would demonstrate an increased frequency of slightly-deleterious variants, segregating within the population.

Results

Our pilot results from comparative-species scale demonstrate the expected trend: long-lived species tend to accumulate more distant amino-acid substitutions (estimated as increased average Grantham distance of the amino-acid substitutions) as compared to short-lived species. Currently, we use additional properties of amino-acids to find a physicochemical signature of substitutions typical for shortlived versus long-lived species. Our pilot results from the intra-species scale also demonstrate expected results: an increased frequency of more deleterious (more biochemically distant) variants segregating in the long-lived species.

Discussion

Both results emphasize relaxed purifying selection in long-lived species because of their low effective population size. Future analyses will uncover a more detail signature of amino-acid substitutions patterns in mammalian species with different generation lengths.

References

[1] K. Popadin et al. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals, PNAS, 2007 104 (33) 13390-13395; https://doi.org/10.1073/pnas.0701256104.

[2] L. Bromham, et al. The Power of Relative Rates Tests Depends on the Data, J Mol Evol (2000) 50:296–301DOI: 10.1007/s002399910034.

DOI 10.18699/BGRS/SB-2020-136

Novel archaeal metagenome assembled genomes from acidophilic microbial community of Parys Mountain copper mine (UK)

Aleksei Korzhenkov
Kurchatov genome center,
NRC Kurchatov Institute,
Moscow, Russia
oscypek@ya.ru

Stepan V. Toshchakov
Winogradsky Institute of Microbiology
FRC "Biotechnology" RAS,
Moscow, Russia
stepan.toshchakov@gmail.com

Ilya V. Kublanov
Winogradsky Institute of Microbiology
FRC "Biotechnology" RAS, Moscow, Russia
kublanov.ilya@gmail.com

Peter N. Golyshin
School of Natural Sciences
Centre for Environmental Biotechnology
Bangor University, Bangor, UK
p.golyshin@bangor.ac.uk

Olga V. Golyshina
School of Natural Sciences
Centre for Environmental Biotechnology
Bangor University, Bangor, UK
o.golyshina@bangor.ac.uk

***Abstract* — Metagenomic analysis is a powerful tool for discovery of novel microbial taxa. Here we report reconstruction of several archaeal metagenome assembled genomes (arMAGs) belonging to uncultivated lineages of Euryarchaeota and DPANN superphylum.**

Keywords* — *extremophiles, acidophiles, archaea, metagenomics.

MOTIVATION AND AIM

Motivation

Extremophilic microbial communities represent an inexhaustible source of novel microbial lineages, enzymes, mechanisms of adaptation and evolution. Recent advances in metagenomic analysis techniques allow us the direct analysis of uncultured extremophiles' genomes and predict their ecological niches and metabolic characteristics.

Aim

Our recent work based on NGS metabarcoding techniques revealed prevalence of Euryarchaeota and presence of DPANN archaea in the acidophilic microbial community of Parys Mountain mine (Anglesey Island, UK) [1]. In this study we aimed to acquire genomic sequences of representatives of earlier detected taxa.

Methods

Metagenomic analysis of acidic mine drainage (AMD) sediment samples was performed as described previously [2]. Metagenomic bins were formed using YAMB [3]. Completeness and contamination of bins were assessed using CheckM [4]. MAG quality score was calculated as *genome completeness - 5 · genome contamination* [5]. MAGs were annotated using RAST web-server with RASTtk pipeline [6]. Preliminary taxonomic assignment of MAGs was performed by aligning 16S rRNA gene sequences against NCBI nr/nt and SILVA reference databases. Multilocus phylogenetic analysis was performed using EPhIMM tool (http://github.com/laxeye/EPhIMM/) with a subset of single copy conserved archaeal genes [4]. Average nucleotide identity (ANI) was calculated using ani.rb script (https://github.com/lmrodriguezr/enveomics/). Functional annotation was made using KAAS server [7], metabolic pathways mapping was made using KEG Mapper [8].

Results

Of 31 reconstructed metagenomic bins eight bins were assigned to Archaea. Six of reconstructed archaeal bins satisfied the high quality MAG criterion – quality score > 50 [5]. Four of identified MAGs were closely related to previously reported 'Ca. Parvarchaeum acidophilus' ARMAN-5 [9], *Cuniculiplasma divulgatum* [10], 'Ca. Mancarchaeum acidiphilum' Mia14 [1], and *Thermoplasmatales* archaeon E-plasma [11] (bin-24, bin-28, bin-31 and bin-r-1, respectively). The other two MAGs show significant remoteness from known genome and MAG sequences on the basis of phylogenetic analysis we could say about relatedness of bin-5 to 'Terrestrial Miscellaneous Euryarchaeal Group' and place bin-30 into cluster with *Cuniculiplasma divulgatum* and several uncultured *Thermoplasmatales* archaea such as UBA582 and E-plasma.

TABLE 1. CHARACTERISTICS OF HIGH QUALITY ARCHAEAL MAGS

ID	MAG size, Mbp	MAG quality score	Nearest cultivable archaea (16S identity, %)	Nearest genome/MAG (ANI, %)
bin-r-1	1.7	88.75	*Cuniculiplasma divulgatum* strain S5 (93.06)	*Thermoplasmatales* archaeon E-plasma (98.96)
bin-5	1.8	78.4	*Methanomassiliicoccus luminyensis* B10 (83.15)	-
bin-24	0.8	67.33	*Methanococcus aeolicus* str. Nankai-3 (74.63)	'Ca. Parvarchaeum acidophilus' ARMAN-5 (99.21)
bin-28	1.9	92.28	*Cuniculiplasma divulgatum* strain PM4 (100)	*Cuniculiplasma divulgatum* strain PM4 (99.80)
bin-30	2.3	93.08	*Cuniculiplasma divulgatum* strain PM4 (94.85)	-
bin-31	0.9	82.4	'Ca. Mancarchaeum acidiphilum' Mia14 (100)	'Ca. Mancarchaeum acidiphilum' Mia14 (99,99)

ACKNOWLEDGMENT

Supported by a grant of Ministry of Science and Higher Education of Russian Federation allocated to the Kurchatov Center for Genome Research (075-15-2019-1659).

References

[1] Golyshina, O. V. et al. (2017). 'ARMAN'archaea depend on association with euryarchaeal host in culture and in situ. Nature communications, 8(1), 1-12.

[2] Korzhenkov, A. A. et al. (2019). Archaea dominate the microbial community in an ecosystem with low-to-moderate temperature and extreme acidity. Microbiome, 7(1), 11.

[3] Korzhenkov, A. (2019). YAMB: metagenome binning using nonlinear dimensionality reduction and density-based clustering. BioRxiv, 521286.

[4] Parks, D. H. et al. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome research, 25(7), 1043-1055.

[5] Parks, D. H. et al. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nature microbiology, 2(11), 1533-1542.

[6] Brettin, T. et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Scientific reports, 5, 8365.

[7] Moriya, Y. et al. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic acids research, 35 (suppl_2), W182-W185.

[8] Kanehisa, M., & Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. Protein Science, 29(1), 28-35.

[9] Baker, B. J. et al. (2010). Enigmatic, ultrasmall, uncultivated Archaea. Proceedings of the National Academy of Sciences, 107(19), 8806-8811.

[10] Golyshina, O. V. et al. (2016). Biology of archaea from a novel family Cuniculiplasmataceae (Thermoplasmata) ubiquitous in hyperacidic environments. Scientific reports, 6, 39034.

[11] Dick, G.J. et al. (2009). Community-wide analysis of microbial genome sequence signatures. Genome biology, 10(8), R85.

DOI 10.18699/BGRS/SB-2020-137

Comparative genomics of heat shock proteins system in extremophile nonbiting midges

Olga Kozlova
Kazan Federal University, Kazan, Russia
olga-sphinx@yandex.ru

Guzel Gazizova
Kazan Federal University, Kazan, Russia
grgazizova@gmail.com

Oleg Gusev
Kazan Federal University, Kazan, Russia
gaijin.ru@gmail.com

Elena Shagimardanova
Kazan Federal University, Kazan, Russia
rjuka@mail.ru

***Abstract* – Here we compare the number and expression profiles of HSP-coding genes in larvae of *Chironomidae* family (Diptera), who are known for their ability to successfully combat abiotic stresses using wide range of behavioral, morphological and biochemical features. In order to perform comparative studies we sequenced and assembled genomes of 4 chironomids from different habitats and also sequenced whole-genome RNA of their larvae in control and stressed conditions. It was shown that compact genome sizes (up to 200 Mb) are typical for *Chironomidae*, while changes in size of a genome are mediated by elongation and shortening of introns length, as well as by changes in quantity and content of dispersed repeats. For all extremophile species under consideration species-specific gene expansion accompanied by formation of compact clusters in a genome was detected. The most amplitudinous reaction towards abiotic stress (desiccation) was shown by anhydrobiotic species *Polypedilum vanderplanki* (Africa). As for HSP-coding genes, it was noticed that genes of HSP20 and HSP70 show the most dramatic and universal up-regulation of expression in response to abiotic stress, while genes of chaperonins (HSP60) tend to be up-regulated in response to desiccation, but not to heat shock. But the most surprising notion was linked to acid-tolerant species *Polipedilum cf. tamanigrum* (Japan), because none of HSP-coding genes in this species showed statistically significant up-regulation, what may be explained by absence of special regulatory sequence – heat shock element (HSE) in their promotor regions.**

Keywords* – *Chironomidae, stress exposure, heat shock proteins, desiccation, genome assembly, transcriptomics

Motivation and aim

Motivation

Most abiotic stresses – such as elevated temperature, desiccation or chemical stress – lead to denaturation and aggregation of wrongly folded proteins. In order to protect their proteins from aggregation and non-native folding, all living systems use wide range of molecular chaperones, including heat shock proteins (HSPs) which present the largest and the most evolutionary conservative class of such chaperones [1]. The regulation of HSPs-coding genes in eukaryotes is performed by specific family of transcription factors – heat shock factors (HSFs), which bind to special regulatory sequence – heat shock element (HSE) in the promotor region of a target gene. Until recently the most well studied aspects of HSP system and its role in adaptation among Diptera belonged to the model genus Drosophila, whose species differ by the level of thermal adaptation and thus can be a convenient object for studying adaptations to elevated temperatures. However, it's well known that HSPs take part in the reaction towards wide range of abiotic stresses – not only to heat shock itself. Taken this into account, it could be interesting to compare features of quantitative composition, genome location and expression patterns of HSPs-coding genes in species, which are remote enough from each other to the evolutionary and geographic point of view, but united by habitation in ecological niches, linked to different extreme impacts. To this point of view the *Chironomidae* (non-biting midges) is considered to be one of the most perspective family of Diptera, because many of its species during their larval stage of development face need to vitality preservation in different extreme and unstable conditions, for what they use specific behavioral, physiologic and molecular-genetic adaptations [2].

Aim

The development of DNA and RNA sequencing technologies opens up broad prospects for studying features of extremophiles lives on the whole-genome level. Using bioinformatic methods of comparative genomics and transcriptomics one can find new genes, predict their functions and study changes of expression during adaptations to extreme conditions. Taking this into account, the aim of the given study was to compare the structure of *de-novo* assembled genomes of extremophile nonbiting midges and to characterize HSP-coding genes and their expression under abiotic stresses.

Methods

In the given research three draft genome assemblies of extremophile *Chironomidae* species were performed. In order to investigate changes in expression levels of HSPs-coding genes we also sequenced RNA of larvae in different stress conditions – heat shock and desiccation for Australian larvae of *Paraborniella tonnoiri*, heat shock for Russian larvae of «Orthocladiinae acuticauda» and heat shock and ion stress (exposition in fresh water) for acid-tolerant larvae of *Polypedilum cf. tamanigrum* from Japan. We also performed brand new variant of genome assembly for unique anhydrobiotic African species *Polypedilum vanderplanki* by sequencing DNA of Pv11 cell line, derived from embryonic mass of the insect [3]. For differential expression analysis of genes of *P. vanderplanki* we used earlier sequenced libraries, which reflect desiccation-rehydration cycle of its larvae [4]. For orthogroups searching we used all predicted aminoacid sequences of species under consideration as well as proteins for 6 other Diptera.

Results

It was found that in genomes of species, which occupy similar ecological niches, even if they are remote from each other geographically and evolutionary, specific orthogroups, which are either totally unique for the species or have significant increase in copy number, can be found. This observation becomes clearer in case of species, which face desiccation in natural habitats, even despite the differences in their physiological, morphological and behavioral adaptations. Thus, a presence of convergent evolutionary changes in genome structure as a result of colonization of new ecological niches, extreme environments of which led to appearance of similar molecular-genetic adaptations could be suggested. Another typical trait of genome structure of extremophile chironomids is compact clusters of paralogous genes, whose expression simultaneously increases after exposure to stress. Among the most obvious examples of this observation there are genes, coding for cysteine-rich secretory proteins of CRISP family in «Orthocladiinae acuticauda» genome (12 gene copies against maximum 4 in other genomes including model Diptera). The compact cluster structures in a genome are also typical for some genes, coding for small HSPs. In contrast to high-molecular HSPs, the number of these genes differs between species: from 10 copies for *P. cf. tamanigrum* till 15 copies for *P. vanderplanki*. It was shown that each extremophile species (except for *P. cf. tamanigrum*) has its own specific cluster with expansion of gene copies, and exactly these species-specific genes show sharp increase of expression after stress. This allows suggesting that the adaptation to extreme environments by using small HSPs walked the path of gene expansion as well as regulation. Among all HSPs-coding genes those corresponding to low-molecular HSP and HSP70 families showed the most dramatic and universal up-regulation of expression in response to abiotic stresses. In contrast to other extremophile species, in case of desiccation-tolerant midge *P. vanderplanki* we noted up-regulation of almost each HSP-coding gene (except for some low-molecular HSPs). Chaperonins (HSP60 family) were specifically up-regulated in case of desiccation, but not after heat shock. But the most surprising observation was linked to acid-tolerant species *P. cf. tamanigrum* which showed significant up-regulation of HSPs-coding genes neither after heat shock nor after ion stress. In order to find a possible reason of this notion we performed motif enrichment analysis in supposed promotor regions of HSPs-coding genes, taking 500 random genes of each genome as a control group. As a result we found statistically significant enrichment of motifs, similar to Drosophila's HSE for genomes of «Orthocladiinae acuticauda», *Paraborniella tonnoiri* and *Polypedilum vanderplanki*, but no enrichment for *Polypedilum cf. tamanigrum*. Thus we could make a conclusion that all HSP system is not involved in reaction to heat shock in *P. cf. tamanigrum*, though for the present it is not clear, how such conservative mechanism of regulation could have been broken.

Acknowledgment

The work was performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References

[1] Chen B. et al. (2018) Evolution of heat-shock protein expression underlying adaptive responses to environmental stress. Molecular Ecology. 27(15): 3040-3054.

[2] Cornette R. et al. (2015) Chironomid midges (Diptera, chironomidae) show extremely small genome sizes. Zoological science. 32(3): 248-254.

[3] Nakahara Y. et al. (2010) Cells from an anhydrobiotic chironomid survive almost complete desiccation. Cryobiology. 60(2): 138-146.

[4] Gusev O. et al. (2014) Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. Nature Communications. 5: 4784

DOI 10.18699/BGRS/SB-2020-138

Comparative genomic analysis of moderate bacteriophages of alfalfa root nodule bacteria

Kozlova A.P.
All-Russian research institute for agricultural microbiology, Pushkin, Saint-Petersburg, Russia
alexsandrak95@mail.ru

Muntyan V.S.
All-Russian research institute for agricultural microbiology, Pushkin, Saint-Petersburg, Russia
vucovar@yandex.ru

Afonin A.M.
All-Russian research institute for agricultural microbiology, Pushkin, Saint-Petersburg, Russia
afoninalexeym@gmail.com

Antonova E.V.
Institute of Plant and Animal Ecology, Ural Division of Russian Academy of Sciences, Ekaterinburg, Russia
selena@ipae.uran.ru

Muntyan A.N.
All-Russian research institute for agricultural microbiology, Pushkin, Saint-Petersburg, Russia
allmuntyan@gmail.com

Kabilov M.R.
ICBFM SB RAS, Novosibirsk, Russia
kabilov@niboch.nsc.ru.

Dzyubenko E.A.
Federal Research Center N. I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR) Ministry of science and higher education, Saint-Petersburg, Russia
elena.dzyubenko@gmail.com

Roumiantseva M.L.
All-Russian research institute for agricultural microbiology, Pushkin, Saint-Petersburg, Russia
mroumiantseva@yandex.ru

***Abstract* — the bacteriophage AP300 infecting Sinorhizobium meliloti was isolated from soil sample from mountainous region of Dagestan, which belongs to the Caucasus Center for Legume Variety in the frame of the grant RSF 17-16-01095. The AP300 genome was annotated and its similarity to bacteriophage vB MloP_Lo5R7ANS (NC_025431; *Podoviridae* family) infecting *Mesorhizobium loti* strains was revealed. In the chromosomes of 4 strains of *Sinorhizobium meliloti* isolated in geographically different regions from soils and nodules of alfalfa plants, intact vB MloP_Lo5R7ANS-like prophages were detected. The 46.39-57.14% ORFs of AP300 and 4 prophages were homologous to ORFs of vB MloP_Lo5R7ANS. The 18 ORFs from mentioned above ORFs were identified in all studied sequences and they were similar to those in vB MloP_Lo5R7ANS. The other 7.22-15.79% ORFs were homologous to other phages. ORFs encoding virion components: head, capsid, tail, virion, integrase were detected in soil bacteriophage AP300 and in all studied prophages determined in rhizobia symbionts of alfalfa.**

Keywords — bacteriophage, prophage, root nodule bacteria, tail tubular protein A, comparative genomic research.

Motivation and aim

Motivation

Bacteriophages are genetically diverse biological objects which are abandoned on the Earth, occurring on average 1 * 10^{31} particles [1]. Bacteriophages can cause lytic infection, accompanied by reprogramming of the host metabolism, destruction of the host cell, or can integrate into the bacterial chromosome in the result of lysogenic infection and transmitted to future generations [2, 3]. It was shown that phages can significantly reduce the number of strains in soil microbiomes as well as in rhizosphere of legume host plant, and affect on the cultural or symbiotic properties of rhizobia [4].

The specificity of phage infection toward bacterial cell is related to tail proteins of the bacteriophage. For example, in vB MloP_Lo5R7ANS the three types of tail proteins involved in the bacterial infection mechanism were annotated. They are tail tubular A protein (TTPA), tail tubular B protein (TTPB), and tail fiber protein (TFP). The latest protein is responsible for recognizing host membrane receptors and for initiating the infection process. The TTPA is a structural protein that is located in the tail of the bacteriophage. It plays an important role in phage adhesion on bacterial cell wall, which is a prerequisite for host infection. For TTP A and B, depolymerization activity in relation to exopolysaccharides of bacteria capsule was shown [6].

For *Sinorizobium* bacteriophages or prophages comparative genomic research and annotation of common specific proteins have not been carried out before.

Aim

The aim of the study was a comparative genomic research of prophages identified in native isolates and phages recovered from soil in order to search common sequences related phage adhesion and infection.

Methods

A pure culture of the bacteriophage *Sinorhizobium meliloti* AP300 was obtained by the Adam's double-layer agar method from a phagolysate obtained from the soils of the mountainous region of Dagestan by *S. meliloti* L5-30 trap strain.

The genomes of the AT7-2 and Ur106 strains and the AP300 bacteriophage were sequenced by short reading method (MiSeq, Illumina, USA). Complete genomes of S. meliloti strains were sequenced by Nanopore sequencing (Oxford Nanopore Technologies, UK). Assembling to circular replicons and annotating genomes was performed by bioinformatics approaches (SPAdes, denovo assembler Flye, Racon and Medaka modules, Pilon software tool, Prokka annotation software package for bacteria). Genome sequences of strains of *S. meliloti* RU11 / 001 and AK21 were obtained by GenBank (NZ_CP021219.1 and NZ_CP026525.1). Prophages were searched on the chromosomes of *S. meliloti* strains by Phaster web-server; comparative analysis was performed by BLASTn, BLASTp, Prosite, I-TASSER, SWISS-MODEL, PDBflex resources.

Results

The four prophages were identified in genomes of *S. meliloti* strains isolated from different geographical areas, and genome of bacteriophage isolated using alfalfa nodule bacteria were screened and sequences homologous to the bacteriophage vB MloP_Lo5R7ANS *Mesorhizobium loti*, were revealed.

The corresponding sequences identified in AP300 phage was shorter (42.8 kb) then genome size of vB MloP Lo5R7ANS (45.7 kb), while homologous sequences in prophages were ranged from 44.1 to 60.2 kb.

The GC content was: 61.02% in AP300 and in prophages it was ranged from 60.65 to 61.13%, in comparison with 61.05% of vB MloP Lo5R7ANS. The number of ORFs was 60 for AP-300, while vB MloP Lo5R7ANS contained 64. In cases of prophages the number of ORFs was varied from 55 to 97. The AP-300 and vB_MloP_Lo5R7ANS genomes carried leucine tRNA, and the sequences of prophages were inserted into leucine tRNA in the chromosomes of the strains. The 46.39-57.14% ORFs of AP300 and prophages were homologous to ORFs of vB MloP Lo5R7ANS (Evalue $<1 * 10^{-10}$). The 18 ORFs were identified in all studied sequences and similar to those of vB MloP_Lo5R7ANS. The other 7.22-15.79% ORFs of bacteriophage and prophages were homologous to other phages ($E_{value} <1 * 10^{-14}$). For the AP300 bacteriophage and three prophages, 19-22 unique ORFs were detected. The prophage of strain Ur106 had 45 unique ORFss, due to the copy number of 28 ORFs (from 2 to 5), which amounted to 46.39% of the ORFs of its entire sequence. In studied sequences were found ORFs encoding virion components: head, capsid, tail, virion, integrase, which were similar to a corresponding set of ORFs in vB MloP Lo5R7ANS. In AP300 as well as in sequences of prophages the enzyme endolysin were annotated, and in case of prophage of Ur106 the transposase TRm23a was identified.

The analysis of amino acid sequences provided the next information that the number of ORFs encoding proteins with known functions was varied from 6 to 8 ORFs in studied sequences.

In each phage related sequences the tail tubular protein A (TTPA) was annotated. A comparative analysis of the amino acid sequences of TTPA of moderate bacteriophages and vB MloP_Lo5R7ANS was done *in silica*. The size of TTPA was 193 a.o. in prophages, and it was 195 a.a. in vB MloP_Lo5R7ANS. The level of homology was 57.29% (for AP300, and prophages of strains RU11/001, Ur106) (I = 57.29%, Cov = 98%, Evalue = $4 * 10^{-78}$ and 57.81-58.33% for prophages of strains AT7-2 and AK21). The amino acid sequences of the TTPA proteins of the prophages and AP-300 lost two amino acid residues: asparagine and serine and had up to 88 amino acid substitutions compared to vB MloP_Lo5R7ANS.

In total 12 -13 different protein modification sites were established for studied sequences. They are: one protein glycosylation site in TTPA, 4 protein kinase C phosphorylation sites, and 5 type II casein kinase phosphorylation sites. It was revealed that TTPA proteins identified in prophages can have additional one phosphorylation site of protein kinase C or a site of modification with myristic acid, or in contract could lose up to two phosphorylation sites of casein kinase II and one site of PKC phosphorylation.

Analysis of the tertiary structure of TTPA proteins vB MloP_Lo5R7ANS and phage related sequences in rhizobia strains and in AP300 showed that the closest structural homologue is protein A of the tail tubule of the bacteriophage KP32 of *Klebsiella pneumonia* (pdb5MU4) (I = 34-37%, Cov = 94-95 %), similarity models 5MU4 and TTPA vB MloP_Lo5R7ANS was 54.44%, while in case of prophages it was higher and amounted to 56.35, but in prophage of soil strain AT7-2 and in bacteriophage AP300 it was 58.19%. Thus, we have determine the presence of sequences the 18 ORFs related to vB MloP_Lo5R7ANS in sequences of prophages of all studied strains as well as in temperate bacteriophage AP300, while the phage-specific parts in were 33.3-46.39%. The prophage of strain Ur106 has up to 5 copies of one third of the ORFs of its genome. Tail tubular protein A in all studied sequences has structural differences from the currently known TTPA bacteriophage KP32 *Klebsiella pneumoniae* (pdb5MU4), it has specific sites of post-translational modification, which can probably determine the specificity of interaction with EPS of nodule bacteria of alfalfa.

Acknowledgment

Supported by the RFBR 18-04-01278a (analysis of ORF and amino acid sequences) and RSF 20-16-00105 (genome sequencing).

References

[1] Clokie MR, Millard AD, Letarov AV, Heaphy S. Phages in nature. Bacteriophage. 2011 Jan; 1(1):31-45.

[2] Fortier LC, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence. 2013 Jul 1; 4(5):354-65.

[3] Krupovic M, Prangishvili D, Hendrix RW, Bamford DH Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. Microbiol Mol Biol Rev. 2011 Dec; 75(4):610-35.

[4] Dziewit L., Oscik K., Bartosik D., Radlinska M. Molecular characterization of a novel temperate Sinorhizobium bacteriophage, ΦLM21, encoding DNA methyltransferase with CcrM-like specificity. J Virol. 2014 Nov; 88(22): 13111–13124.

[5] Veesler D, Cambillau C. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol Mol Biol Rev. 2011 Sep; 75(3):423-33, first page of table of contents.

[6] Pyra A., Brzozowska E., Pawlik K., Gamian A., Dauter M., Dauter Z. Tail tubular protein A: a dual-function tail protein of Klebsiella pneumoniae bacteriophage KP32. Sci Rep. 2017; 7: 2223. Published online 2017 May 22.

DOI 10.18699/BGRS/SB-2020-139

Distribution of Bax protein in the rat hippocampus

Pavel D. Lisachev
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
lisachev@ngs.ru

Anna L. Proskura
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
annleop@mail.ru

***Abstract* — Bcl-2 family protein Bax is involved in mechanisms of synaptic plasticity. The induction of long term potentiation in hippocampal slices leads to increased Bax expression, but the localization of these changes remains unclear. Bax immunoreactivity is visually detected mainly in the layer of pyramidal neurons and is rarely detected in S100B-positive glial cells. However, a quantitative assessment of Bax colocalization with glial and neuronal markers (S100B and NeuN, respectively) showed that although the Bax content in S100B-positive glial cells is really low, the Bax neuronal somatic pool is not the main source of Bax protein in the hippocampus.**

Keywords — hippocampus; Bax; synaptic plasticity

Introduction

Bcl-2 family proteins are the regulators of apoptosis, but also have other functions. The Bax protein belonging to this family forms a channel in the mitochondrial membrane through which, under stressful conditions, cytochrome c exits mitochondria to trigger activation of a cascade of proteolytic enzymes – caspases, which play, in particular, a key role in apoptosis [1]. This mechanism is also involved in the formation of glutamate N-methyl-D-aspartate (NMDA) receptor-dependent synaptic depression in hippocampal CA1 field. The low frequency (1 Hz) activation of the Schaffer collaterals leads to the limited activation of caspases mediated by the Bax protein, which ultimately leads to the internalization of glutamate α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors and, thereby, to a weakening of synaptic efficiency [2]. Though, during the formation of long term potentiation in the synapses of the Schaffer collaterals on the pyramidal neurons of the hippocampal CA1 field, an increase in Bax expression in the CA1 field homogenates is observed [3]. If this increase occurs in neurons, this may indicate that the long term potentiation of synaptic connections is accompanied by the activation of negative feedbacks, creating conditions for limiting growth or increasing a predisposition to reduce the effectiveness of glutamatergic synapses.

A preliminary analysis of the distribution of Bax in the rat hippocampus showed that Bax immunoreactivity is indeed visually detected mainly in the layer of pyramidal neurons and is rarely detected in S100B-positive glial cells [4]. In the present work, a more detailed quantitative assessment of Bax colocalization with glial and neuronal markers (S100B and NeuN, respectively) was carried out.

Methods

Male Wistar rats (7-9 weeks old) were used in the experiments. After decapitation of the animal, the brain was placed in cold aerated saline. Immediately after isolation, fragments of the hippocampus were placed in 4% paraform. Primary anti-S100B and anti-NeuN antibodies from mouse and anti-Bax from rabbit, secondary goat anti-mouse IgG-FITC and anti-rabbit IgG-DyLight594 were used. After immunohistochemical processing, hippocampal slices (20 μm) were examined using a laser scanning microscope LSM 780, ×20 lens, x/y coordinate resolution 0.83 μm/pixel. When analyzing Bax colocalization with cell markers, in each histological slice, the data obtained on the four (2 μm interval) most representative optical sections were summarized. As a minimum estimate of the background fluorescence level, a MIN value was taken such that at least 99.99% of the voxels in the scan area outside the sample had a fluorescence level less than MIN. The colocalization coefficients of Bax with cell markers were calculated at threshold Bax levels Ln, which were determined as follows:

$$Ln = MIN + n(MAX - MIN)/8 \quad (1)$$

Here n = 0, 1, ..., 7, and MAX is such that no more than 0.1% of all voxels in the scan area had a fluorescence level greater than MAX. Average values are presented as "mean ± standard deviation".

Results

The colocalization coefficient of Bax (the ratio of the number of colocalized voxels to the total number of Bax-positive voxels) with the glial cytoplasmic marker S100B, as a rule, did not depend on the threshold Bax level and ranged from 0.02 to 0.05. This parameter reached its maximum value (0.13) in one of six samples for voxels with a maximum level of Bax immunopositivity (at L7 threshold). In all cases, the colocalization coefficients were higher than the expected random values.

Neuron marker NeuN is expressed primarily in the nuclei of neurons. In smaller quantities, it was also present in the cytoplasm of somata and proximal dendrites. Although visually the degree of colocalization of Bax and the neuronal marker NeuN seems rather high, the coefficient of colocalization of Bax with NeuN turned out to be small – no more than 0.3. Moreover, it gradually increased with increasing threshold Bax level – from 0.06 ± 0.02 at L0 to 0.16 ± 0.08 at L7. Obviously, the colocalization coefficient of Bax with NeuN underestimates the Bax content in neurons. Perhaps a significant part of Bax is located in the processes of cells, where it exists mainly in the form of small inclusions, which are hardly perceived visually. However, although the content of Bax in S100B-positive glial cells is low, it cannot be ruled out that a significant amount is present in cells of other types, for example, in blood vessels.

Conclusion

A visual analysis of the distribution of Bax in the hippocampal cells is insufficient to conclude that Bax is predominantly neuronal. To analyze cell specificity and neuronal activity-dependent dynamics of Bax expression, it may be appropriate to use animals with neuron-specific cytoplasmic expression of fluorescent markers.

Acknowledgment

The histological preparations were examined at the Shared Centre for Microscopic Analysis of Biological Objects of the Institute of Cytology and Genetics SB RAS, the authors are thankful to Dr. S.I. Baiborodin for technical support.

References

[1] J. M. Hardwick and L. Soane, "Multiple functions of BCL-2 family proteins," Cold Spring Harb. Perspect. Biol., vol. 5, a008722, February 2013.

[2] Z. Li and M. Sheng, "Caspases in synaptic plasticity," Mol. Brain, vol. 5, 15, May 2012.

[3] V. O. Pustylnyak, P. D. Lisachev, M. B. Shtark, "Expression of p53 target genes in the early phase of long-term potentiation in the rat hippocampal CA1 area," Neural Plast., vol. 2015, 242158, 2015.

[4] P. D. Lisachev and M. B. Shtark, "Long-term potentiation-associated gene expression: involvement of the tumour protein p53," in The Hippocampus – Plasticity and Functions, A. Stuchlik, Ed. London: IntechOpen, 2018, pp. 49–64.

DOI 10.18699/BGRS/SB-2020-140

Conservative mitotic localization and functions of the nucleolar RPF2/Non3 protein in human and *Drosophila* cells

Svetlana V. Maltseva
Institute of Molecular and Cellular Biology SB RAS,
Novosibirsk State University
Novosibirsk, Russian Federation

Lybov A. Yarinich
Institute of Molecular and Cellular Biology SB RAS,
Novosibirsk State University
Novosibirsk, Russian Federation

Anna A. Ogienko
Institute of Molecular and Cellular Biology SB RAS,
Novosibirsk, Russian Federation

Evgenia S. Omelina
Institute of Molecular and Cellular Biology SB RAS,
Novosibirsk, Russian Federation

Alena V. Razuvaeva
Institute of Molecular and Cellular Biology SB RAS,
Novosibirsk, Russian Federation

Alexey V. Pindyurin
Institute of Molecular and Cellular Biology SB RAS,
Novosibirsk, Russian Federation

Julia V. Popova
Institute of Molecular and Cellular Biology SB RAS,
Novosibirsk, Russian Federation

Anastasia A. Yushkova
Institute of Molecular and Cellular Biology SB RAS,
Novosibirsk State University
Novosibirsk, Russian Federation

Evgenia N. Andreyeva
Institute of Molecular and Cellular Biology SB RAS,
Novosibirsk, Russian Federation
andreeva@mcb.nsc.ru

Gera A. Pavlova
Institute of Molecular and Cellular Biology SB RAS,
Novosibirsk, Russian Federation

***Abstract* — The nucleolus is a non-membrane nuclear organelle playing a crucial role in ribosome biogenesis in eukaryotes. Its morphology and function are associated with cell growth and proliferation, as well as with cell cycle regulation. Besides that, a set of nucleolar proteins are also known to be directly involved in mitotic cell division, taking a part in the spindle assembly. Among them is the human RRS1 (Ribosome biogenesis regulator 1 homolog) protein that function in complex with RPF2 (Ribosome production factor 2 homolog) in the ribosome biogenesis pathway. Here, we report the characterization of the mitotic localization and function of the human RPF2 and its *Drosophila* ortholog, the Non3 (Novel nucleolar protein 3) protein. We demonstrated that depletion of RPF2 using siRNA specific to all its splicing variants leads to abnormal mitotic chromosome segregation in HeLa S3 cells. Similarly, the RNAi-mediated depletion of Non3 in *Drosophila* S2 cells also resulted in mitotic spindle defects including those in sister chromatid separation and segregation. Thereby, it will be possible to investigate mitotic functions of the RPF2 protein in *Drosophila* model system.**

Keywords* — *nucleolus, RRS1, RPF2, Non3, moonlighting protein, mitosis, chromosome periphery, chromosome segregation defects

Motivation and Aim

The nucleolus is a non-membrane nuclear organelle playing a crucial role in ribosome biogenesis in eukaryotes. Its morphology and function are associated with cell growth and proliferation, as well as with cell cycle regulation. Besides that, a set of nucleolar proteins are also known to be directly involved in mitotic cell division, taking a part in the spindle assembly. Among them is the human RRS1 (Ribosome biogenesis regulator 1 homolog) protein that function in complex with RPF2 (Ribosome production factor 2 homolog) in the ribosome biogenesis pathway. RRS1 localizes at the periphery of the mitotic chromosomes and its depletion results in impaired chromosome alignment in metaphase and premature sister chromatid separation [1]. Currently, the information about localization of the RPF2 protein during mitosis is scarce [2] and nothing is known about its possible involvement in this process. However, the *Drosophila* ortholog of RPF2, the Non3 (Novel nucleolar protein 3) protein, was revealed to be somehow connected with the regulation of mitotic spindle length despite its exclusive localization in the nucleolus [3].

Results

We thoroughly studied the subcellular localization of all annotated isoforms of the RPF2 and Non3 proteins in human HeLa S3 and *Drosophila* S2 cells, respectively. For that, each protein was expressed as a GFP fusion and the cells were analyzed by confocal microscopy. We found that only the longest isoform (#1) of RPF2 demonstrates specific localization at the chromosome periphery in metaphase. Both shorter isoforms (#2 and #X1) of RPF2, that lack the first 28 and 53 a.a. of the 190-a.a. long BRIX domain required for interaction with rRNAs, were found to form multiple aggregates dispersed between the metaphase chromosomes instead of covering the outer surface of chromosomes as the longer isoform does. Also, we were able to detect an enrichment of the single known isoform of Non3 around the metaphase plate. Next, we demonstrated that depletion of RPF2 using siRNA specific to all its splicing variants leads to abnormal mitotic chromosome segregation in HeLa S3 cells. Similarly, the RNAi-mediated depletion of Non3 in *Drosophila* S2 cells also resulted in mitotic spindle defects including those in sister chromatid separation and segregation. To analyze tissue-specificity of the phenomenon, we generated a number of *Drosophila Non3* null and hypomorphic mutations. Immunostaining experiments performed on diploid *Drosophila* tissues suggested that centromere chromatin assembly is likely to be affected in *Non3* mutants.

Conclusion

The localization of the nucleolar RPF2 and Non3 proteins during mitosis is similar in human and *Drosophila* cells. In addition, mitotic defects observed upon depletion of these proteins are very similar suggesting the existence of functional conservation as well. Thus, *Drosophila* appears to be a relevant model system to study mitotic functions of the RPF2 protein.

Acknowledgments

This work was supported by the grant from the Russian Foundation for Basic Research (project #18-34-00699).

References

[1] Gambe AE, Matsunaga S, Takata H, Ono-Maniwa R, Baba A, Uchiyama S, Fukui K. A nucleolar protein RRS1 contributes to chromosome congression. FEBS Lett. 2009, 583(12):1951-1956.

[2] Hirano Y, Ishii K, Kumeta M, Furukawa K, Takeyasu K, Horigome T. Proteomic and targeted analytical identification of BXDC1 and EBNA1BP2 as dynamic scaffold proteins in the nucleolus. Genes Cells. 2009, 14(2):155-166.

[3] Moutinho-Pereira S, Stuurman N, Afonso O, Hornsveld M, Aguiar P, Goshima G, Vale RD, Maiato H. Genes involved in centrosome-independent mitotic spindle assembly in *Drosophila* S2 cells. Proc Natl Acad Sci USA. 2013, 110(49):19808-19813.

DOI 10.18699/BGRS/SB-2020-141

Phylostratigraphic approach in evolutionary analysis: comparison of methods

Tatiana Martusheva
Novosibirsk State University
NSU, Novosibirsk, Russia
t.maryanovskaya@g.nsu.ru

Zakhar Mustafin
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
mustafinzs@bionet.nsc.ru

Sergey Lashin
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
lashin@bionet.nsc.ru

***Abstract* — Phylostratigraphic analysis allows us to determine the "age of the gene" - evolutionary time of its origin. There are different methods to calculate this age, each of which has its own issues. In this work, we compare several studies on phylostrastigraphic analysis. The correlation coefficient of the results of two research with different phylostratigraphic trees but the same method to count phylostratigraphic age (BLAST-based) was found to be 0.76. Comparison of the results of calculating the phylostratigraphic age by Orthoscape application (used KEGG as data source) and the myTAI package (used BLAST as data source) showed a lower corellation coefficient – only 0.27. It is planned to continue the studying of differences in the results of different methods and their reasons.**

Key words — phylostratigraphy; evolution; gene age; BLAST; KEGG; ortholog

Motivation and Aim

Phylostratigraphic analysis is a promising field of evolutionary analysis that allows us to determine the most important steps in the evolution of organisms [1]. Using this analysis, one can determine the "age of the gene" - the presumed stage of the origin of the gene in the process of evolution of the organism. The age of a gene is based on the analysis of orthologous genes and the search for the last common ancestor for the gene and its orthologs on the evolutionary tree. There are several ways to search for orthologous genes and their positions on that tree, the most popular of which is using the Basic Local Alignment Search Tool, https://blast.ncbi.nlm.nih.gov/Blast.cgi). Though this method is effective, however, it has some drawbacks [2], [3], which encourages the use of other tools. In particular, databases on orthologous genes can be used, for example, KEGG (Kyoto Encyclopedia of Genes and Genomes, https://www.kegg.jp) and EggNOG (http://eggnogdb.embl.de/#/app/home). In this study, we compare the results of using different methods underlying phylostratigraphic analysis.

Methods and Algorithms

We have reviewed scientific papers with published data on the analysis of gene age, as well as programs for this analysis. One of these programs is Orthoscape [4], which is based on the KEGG orthologs database. Yet another phylostratigraphic information was taken from phylotranscriptomic studies in particular used myTAI package [5], which defines orthologs and position on the taxonomic tree using BLAST. Phylostratigraphic ages resulted from the programs mentioned above, as well as taken from published papers, were compared, in particular, in terms of correlation. We have additionally reduced all the taxonomy data to the uniform tree, which allowed us to make such a meta-analysis.

Results

We have compared phylostratigraphic ages indexes of *Arabidopsis thaliana* genes taken from the publications of Ruprecht and co-authors [6] and Quint and co-authors [7]. Both studies used BLAST, but the authors took different taxonomic trees represented key evolutionary points of plants. Correlation coefficient for initial trees was shown to be 0.74 as well as 0.76 for the uniform trees. This suggests that the selection of key taxa in the phylostratigraphic tree did not significantly affect the age of genes when analyzed using the same method. Comparison of the results of calculating the philostratigraphic age of genes using the Orthoscape application, which uses the KEGG database as a source of data on orthologous genes, and the myTAI package, which uses BLAST, showed a significantly lower correlation coefficient - only 0.27, which indicates a fundamental difference in the result when using these methods. Orthoscape found significantly more young genes, while myTAI found older ones. This is due to separate groups of genes that differ significantly in age in different methods, for example, ribosomal genes are as old as possible in the first method, and as young as possible in the second one. This may also be due to fundamentally different parameters for selecting orthologs, in particular, the minimum identity of amino acid sequences of proteins encoded by genes, which is necessary for genes to be considered orthologs. In the future, this issue will be studied in more detail and various options for selecting orthologs will be considered and the results obtained will be compared.

Acknowledgment

The study is supported by RFBR, grant number 20-04-00885 A.

References

[1] T. Domazet-Lošo *et al.* A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages, *Trends Genet.*, vol. 23, no. 11, 2007, doi: 10.1016/j.tig.2007.08.014.

[2] L. B. Koski and G. B. Golding, The Closest BLAST Hit Is Often Not the Nearest Neighbor, *J. Mol. Evol.*, vol. 52, no. 6, pp. 540–542, Jun. 2001, doi: 10.1007/s002390010184.

[3] B. A. Moyers and J. Zhang, Phylostratigraphic Bias Creates Spurious Patterns of Genome Evolution, *Mol. Biol. Evol.*, vol. 32, no. 1, pp. 258–267, Jan. 2015, doi: 10.1093/molbev/msu286.

[4] Z.S. Mustafin *et al.* Orthoscape: a cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles, *BMC Bioinformatics*, vol. 18, no. S1, pp. 1–9, 2017, doi: 10.1186/s12859-016-1427-5.

[5] H.-G. Drost *et al.* myTAI: evolutionary transcriptomics with R, *Bioinformatics*. 2018, doi: 10.1093/bioinformatics/btx835.

[6] C. Ruprecht *et al.* Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules, *Plant J.* 2017, doi: 10.1111/tpj.13502.

[7] M. Quint *et al.* A transcriptomic hourglass in plant embryogenesis, *Nature*. 2012, doi: 10.1038/nature11394.

DOI 10.18699/BGRS/SB-2020-142

Mitochondrial mutational spectrum in poikilothermic versus homeothermic vertebrates: effects of the temperature

Alina G. Mikhaylova
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University
Kaliningrad, Russian Federation
The Vavilov Institute of General Genetics RAS, Moscow, Russian Federation
polarsong4@gmail.com

Victor Shamanskiy
Center for Mitochondrial Functional Genomics ,Immanuel Kant Baltic Federal University
Kaliningrad, Russian Federation
v.a.shamanskiy@gmail.com

Alina A. Mikhaylova
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University
Kaliningrad, Russian Federation
mihailovaalina777@yandex.ru

Konstantin Gunbin
The Institute of Cytology and Genetics of the SB RAS
Novosibirsk, Russian Federation
genkvg@gmail.com

Vsevolod Makeev
The Vavilov Institute of General Genetics RAS, Moscow, Russian Federation
vsevolod.makeev@gmail.com

Kristina Ushakova
ITMO University
Saint Petersburg, Russian Federation
kristina.ushakova@outlook.com

Konstantin Popadin
School of Life Sciences
Ecole Polytechnique Federale de Lausanne
Lausanne, Switzerland
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation
konstantinpopadin@gmail.com

Abstract — **Mutational processes of mitochondrial genome in Vertebrates remain unclear. In order to understand them better we obtained mutational spectra, which contain information about single nucleotide substitutions occurred in neutral within-species polymorphisms, uncovering key mutagens of mitochondrial DNA. Here, we analyzed mutational spectra of Actinopterygii and hibernating mammalian species and observed that frequencies of A>G substitutions are sensitive to temperature.**

Key words — Mutational spectrum, mitochondrial DNA, somatic mutations, Actinopterygii, hibernating mammals

Introduction

The mutation process is one of the key drivers of evolution. By understanding the rules of mutagenesis, we better learn what distribution of mutations selection we can work with. More and more is found out about the factors affecting the mutagenesis of the nuclear genome: there are many discovered specific mutational signatures of various exogenous (UV light, tobacco smoke ...) and endogenous (DNA polymerase errors, repair enzyme activity) factors. However, for the mitochondrial genome (mtDNA), progress is relatively slow: the main factors (exogenous and endogenous) that determine the specific rules of mtDNA mutagenesis are still unknown. The majority of strong mutagens well known for the nuclear genome (for example, UV-light) do not act the same for mtDNA, which makes mtDNA mutagenesis unique and extremely important, because it is associated with an understanding of selection process of different animal species and largely determines the rate of appearance of human somatic mutations. The global goal of our project is to understand the rules of mtDNA mutagenesis in vertebrates and humans.

II. RECONSTRUCTION OF MUTATIONAL SPECTRA IN VERTEBRATE SPECIES

Following the idea that somatic variants in mtDNA are informative to trace different cellular lineages in our body [3] we hypothesized that mtDNA mutational spectrum, defined as a relative frequencies of mtDNA single-nucleotide substitutions, may be used not only for cellular tracing, but may contain information about cellular features and specific functional signatures of a given cell.

In our recent work [1] we firstly analyzed somatic mtDNA mutations from human cancers and healthy human tissues and then focusing on mammalian species we analyzed polymorphic synonymous substitutions in four-fold degenerate sites and found out that mtDNA mutational spectrum indeed is sensitive to several metabolic and ecological properties on all levels (cellular, tissue and organismal). We observed one universal trend: ratio of A to G transitions (A>G, heavy chain notation) positively correlates with cellular and organismal longevity. Focusing on mammals, we observed that the fraction of A>G is higher in long- versus short-lived species. This led us to propose that deamination of adenine (the main cause of A>G transitions) is more intense in long-lived (aged) oocytes, typical for long-lived species.

Results

Here, we extend our analyses towards cold-blooded vertebrates, because substitution frequencies can be affected also by body parameters such as temperature [6], while first dataset represented only uniform species regarding temperature. Analyzing mtDNA mutational spectrum of up to one hundred Actinopterygii species (N=128) with known longevity (mean maturation time, Tm) and median temperature of environment [5] we observed that the frequency of A>G transitions positively correlates with temperature (results are robust to phylogenetic inertia), while the effect of longevity is only marginally significant. (Fig. 1) Taking into account the fact that observed substitution frequencies can shape specific nucleotide content in neutral positions we analyzed additionally synonymous nucleotide frequencies in more than three hundred (N=302) whole mitochondrial genomes of Actinopterygii species. As expected we observed decrease of A fraction with increasing temperature, which can be long-term result of A>G substitutions (Fig. 2).

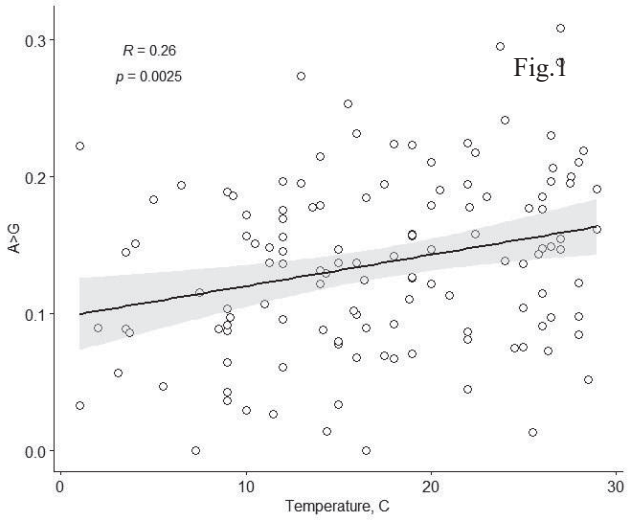


Fig.1

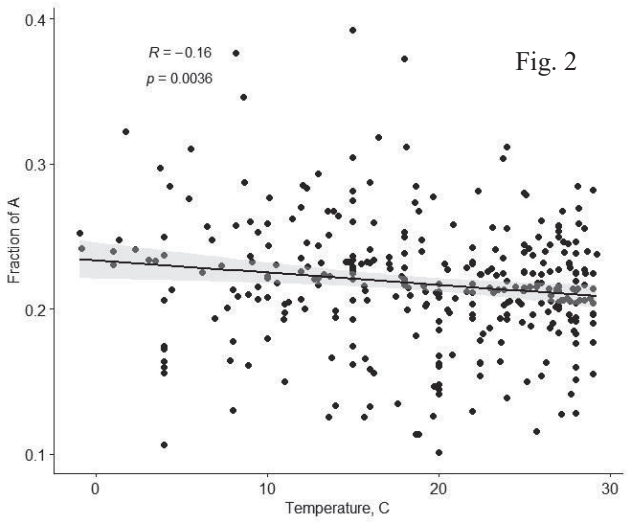


Fig. 2

Altogether, we conclude that A>G is sensitive to both temperature and lifespan, with temperature predominantly shaping variation in mtDNA mutational spectrum of poikilotherms, while longevity is the main factor in homeotherms. To prove this, we compared hibernating mammalian species (HibDay in fig. 3), Monotremata species and Marsupials species (MonoMarsHG in fig. 3) with nonhibernating placental ones (Warm in fig. 3) (N=350), controlling for generation length [2, 4]. We observed a increased fraction of A in hibernating, Monotremata, Marsupial species, in line with their expected decreased temperature.

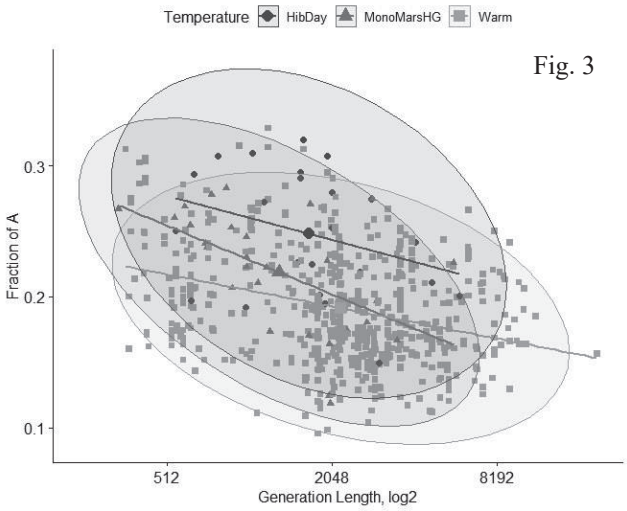


Fig. 3

Multiple models were performed for all analyses to improve results. Taking into account that A>G is one of the most common transitions in mtDNA we conclude that the whole mtDNA mutational spectrum is associated with temperature and longevity. This new information should be taken into account in numerous ecological, population and evolutionary projects where mitochondrial genes are the most frequently used markers.

Acknowledgment.

This work is supported by the 5 Top 100 Russian Academic

Excellence Project at the Immanuel Kant Baltic Federal University, by the Russian Foundation for Basic Research grant 18-29-13055.

References

[1] A. G. Mikhaylova, A. A. Mikhailova, K. Ushakova, E. Tretiakov, A. Yurchenko, M. Zazhytska, D.A. Knorre, E. Zdobnov, Z. Fleischmann, S. Annis, M. Franco, K. Wasko, W.S. Kunz, I. Mazunin, S. Nikolaev, A. Reymond, K. Khrapko, K. Gunbin, View ORCID ProfileK. Popadin, Mitochondrial mutational spectrum provides an universal marker of cellular and organismal longevity", bioRxiv 589168; doi: https://doi.org/10.1101/589168.

[2] J. P. de Magalhaes and J. Costa, "A database of vertebrate longevity records and their relation to other life-history traits.", 2009, Journal of Evolutionary Biology 22(8):1770-1774.

[3] L. S. Ludwig, C. A. Lareau, J. C. Ulirsch, E. Christian, C. Muus, L.

[4] H. Li, K. Pelka, et al. 2019. "Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics." Cell 176 (6): 1325–39.e22.

[5] M. Pacifici, L. Santini, M. Di Marco, L. Santini, D. Baisero, L. Francucci, G. Grottolo Marasini, P. Visconti, C. Rondinini, Generation length for mammals. 2013, Nature Conservation 5: 89-94.

[6] R. Froese and D. Pauly. Editors. 2019. FishBase. World Wide Web electronic publication. www.fishbase.org, (12/2019)

[7] W. Zheng, K. Khrapko, H. A. Coller, W. G. Thilly, W. C. Copeland Origins of human mitochondrial point mutations as DNA polymerase gamma-mediated errors. Mutat Res. 2006 Jul 25;599(1-2):11-20. Epub 2006 Feb 20.

DOI 10.18699/BGRS/SB-2020-143

Sociality affects mutational spectrum of mtDNA in termites versus cockroaches

Alina A. Mikhailova
Center for Mitochondrial Functional Genomics
Immanuel Kant Baltic Federal University
Kaliningrad, Russia
Okinawa Institute of Science and Technology
Okinawa, Japan
ORCID: 0000-0002-6203-9977

Konstantin Gunbin
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
Immanuel Kant Baltic Federal University
Kaliningrad, Russia
genkvg@gmail.com

Konstantin Popadin
School of Life Sciences
Ecole Polytechnique Federale de Lausanne
Lausanne, Switzerland
Center for Mitochondrial Functional Genomics
Immanuel Kant Baltic Federal University
Kaliningrad, Russia
konstantinpopadin@gmail.com

Thomas Bourguignon
Okinawa Institute of Science and Technology
Okinawa, Japan
Thomas.Bourguignon@oist.jp

***Abstract* — Population structure is an important factor driving genetic changes. Comparing social (termites) and non-social (cockroaches) insects we found the difference in their mitochondrial mutational spectrum: transition transversion ratio (ts/tv) is higher in termites compared to cockroaches suggesting potential influence of population structure on mutational patterns.**

***Keywords* — mtDNA, Ts/Tv ratio, termites**

Introduction

Ts/Tv ratio in different species

The transition transversion ratio (Ts/Tv ratio) is very high in human mitochondrial DNA and varies significantly between different species, families, and orders [1], but the reason for this difference remained unknown. Recently it was suggested that the mitochondrial mutational spectrum is associated with cellular and organismal longevity [2]. It was also shown during mutation accumulation experiments that the pattern of mitochondrial mutations is different in four model organisms [3]. So, both polymorphism data and mutation accumulation experiments tell us about differences in the mutational spectrum of mtDNA across species and lineages.

Termites and cockroaches

Termites are eusocial insects that evolved from cockroaches of order Blattodea. They are among the most abundant terrestrial organisms and play an important role in decomposition and nutrient recycling, as they are able to digest lignocellulose [4].

Colonies of termites can range in their size from several hundred to millions of individuals. Fertile females of the colonies called "queens" can live dozens of years. It's especially pronounced in "higher" termites, which tend to have bigger colonies and more long-lived queens.

All that makes termites the perfect object to study the influence of population structure and longevity on the mitochondrial mutational spectrum.

Inter-species comparisons

We collected all mitochondrial protein-coding genes for 98 cockroaches and 402 termites species, reconstructed within-species phylogeny using sequences from sister species as outgroups, reconstructed sequences at each internal node, derived a list of polarized single-nucleotide substitutions between the nodes, and, focusing on the most neutral synonymous fourfold degenerate sites, counted the observed frequencies of twelve types of nucleotide substitutions.

We calculated the Ts/Tv ratio for every species and found an increased ratio in termites compared to cockroaches (Fig. 1), that might be explained by different generation lengths of the compared groups. Moreover when we split termites into the lower and the higher ones we see a tendence for concordant increase in Ts/Tv from short-lived lower termites to more long-lived higher termites.

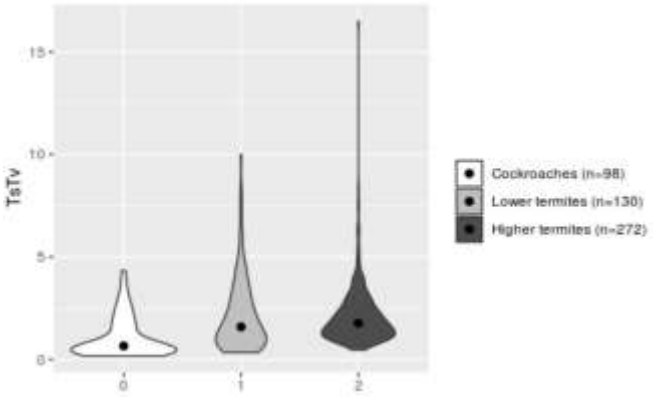


Fig. 1. Ts/Tv ratio for cockroaches, lower and higher termites

Discussion

We analyzed the mutational spectrum of termites and cockroaches and found an increased Ts/Tv ratio in termites. A similar trend was previously found for mammals: long-lived mammals have a higher Ts/Tv ratio compared to short-lived [2]. Indeed, termite queens live longer than cockroaches but the presence of secondary reproductives and social structure make mtDNA dynamics in termites more complicated.

Rising the lower boundary of Ts/Tv ratio from cockroaches to higher termites pushed us to another tempting explanation: this change can be explained by a change in ratio between purines and pyrimidines in nearly

neutral sites. This process can be driven by change in ROS influence. Obviously, ROS is more important, the higher the life expectancy of the body is.

In addition to previous analyzes, we plan to look deeper into different types of mutations, we also intend to analyze nuclear genes to differentiate mitochondria-specific and common mutation patterns.

Acknowledgment

We thank OIST Scientific Computing & Data Analysis Section, as most of the analyzes were made on cluster Sango at OIST.

This work has been supported by the 5 Top 100 Russian Academic Excellence Project at the Immanuel Kant Baltic Federal University, by the Russian Foundation for Basic Research grant 18-29-13055.

References

[1] Belle, E. M., Piganeau, G., Gardner, M., & Eyre-Walker, A. "An investigation of the variation in the transition bias among various animal mitochondrial DNA." Gene 355 (2005): 58-66.

[2] Mikhaylova, Alina G., et al. "Mitochondrial mutational spectrum provides an universal marker of cellular and organismal longevity." bioRxiv (2019): 589168.

[3] Montooth, Kristi L., and David M. Rand. "The spectrum of mitochondrial mutation differs across species." PLoS biology 6.8 (2008).

[4] Jouquet, P., Traoré, S., Choosai, C., Hartmann, C., & Bignell, D. "Influence of termites on ecosystem functioning. Ecosystem services provided by termites." European Journal of Soil Biology 47.4 (2011): 215-222.

DOI 10.18699/BGRS/SB-2020-145

OrthoWeb – web application for macro- and microevolutionary analysis of genes

Zakhar Mustafin
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
mustafinzs@bionet.nsc.ru

Dmitry Afonnikov
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
ada@bionet.nsc.ru

Sergey Lashin
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
lashin@bionet.nsc.ru

Alexey Mukhin
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
mukhin@bionet.nsc.ru

Yury Matushkin
ICG SB RAS, Novosibirsk, Russia
mat@bionet.nsc.ru

***Abstract* — Problems of macro- and microevolution remain key problems of biology. Phylostratigraphic analysis became one of the most popular methods for studying macroevolutionary characteristic of genes. It is based on estimation of the divergence time among orthologous genes. Along with methods of microevolutionary analysis (for example, dN/dS ratio estimation), phylostratigraphic methods are increasingly included into the methodological arsenal of evolutionary bioinformatics. There are some software packages used for computational analysis of the evolution by using different phylostratigraphic and microevolutionary indices, but they require users to know some additional software or programming languages. Here we present OrthoWeb software to calculate such indices using web browser only.**

Key words* — *evolution, phylostratigraphy, ortholog, gene network

Motivation and Aim

Problems of macro- and microevolution remain key problems of biology. Phylostratigraphic analysis proposed over a decade ago [1] became one of the most popular methods for studying macroevolutionary characteristic of genes. It is based on estimation of the divergence time among orthologous genes. Study of such genes may highlight the evolutionary age of genes and help to understand the evolution. Along with methods of microevolutionary analysis (for example, dN/dS ratio estimation), phylostratigraphic methods are increasingly included into the methodological arsenal of evolutionary bioinformatics. There are some software packages (http://apps.cytoscape.org/apps/orthoscape,[2]; https://cran.r-project.org/web/packages/myTAI,[3]; https://github.com/arendsee/phylostratr,[4]), used for computational analysis of the evolution by using different phylostratigraphic and microevolutionary indices, but they require users to know some additional software (like Cytoscape [5]) or programming languages (like R). Here we present OrthoWeb software to calculate such indices using web browser only.

Methods and Algorithms

OrthoWeb is the software to calculate evolutionary analysis of genes, based on evolutionary analysis algorithms used in Orthoscape the software previously developed by our group. There are two groups of indices implemented in OrthoWeb.

PAI

PAI (Phylostratigraphic Age Index) aims at macroevolution analysis. It allows finding the "gene age" as a supposed stage of gene origin (the stage where the earliest duplication of ortholog can be found). The PAI analysis is based on comparison of the sequences of proteins, coded by original gene and its orthologous genes. There are two ways to compare genes and orthologs:

1) By using KEGG database (Kyoto Encyclopedia of Genes and Genomes, https://www.kegg.jp, [6]). There are two main thresholds allow us to check if genes are orthologous or not. The first one is the identity of amino acid sequences (as a part from 0 to 1) and the second is the SW-Score (the result of the Smith-Waterman alignment algorithm). The genes, which values of identity and SW-Score in compare with original genes are higher than user-defined threshold are counted as orthologous genes. There is also featured threshold based on domains comparison, which can be used for more complex analysis. It is possible to specify the domains of proteins coded by analyzing genes.

2) By using BLAST (Basic Local Alignment Search Tool, https://blast.ncbi.nlm.nih.gov/Blast.cgi). This method based on uploading the sequence of gene into NCBI (National Center for Biotechnology Information) and analyzing it using BLAST. User can specify identity and score by the same way as in the method 1.

Both methods return the list of orthologous genes and taxonomic lineages of organisms which genes are presented. The phylum most remote from the root of the taxonomic tree (Cellular Organisms) and common for gene and every ortholog represents the "gene age".

DI

DI (Divergence Index) aims at microevolution analysis. The index based on well-known ratio dN/dS, where dN shows the number of nonsynonymous substitutions and dS the number of synonymous substitutions. The ratio value above 1 indicates the evolution of the gene under positive Darwinian selection. The ratio close to 1 indicates that a gene evolves under neutral regime. The values of the ratio close to 0 indicate strong purifying selection acting on a gene.

DI is an average value of dN/dS for gene and its orthologs. OrthoWeb allows to choose the degree of relationship between organisms but usually the closest

organisms using in analysis (for example *A.lyrata* for *A.thaliana* or Hominidae's orthologs for *H.sapiens*).

To count dN/dS OrthoWeb uses PAML (Phylogenetic Analysis by Maximum Likelihood, http://abacus.gene.ucl.ac.uk/software/paml.html, [7]).

Results

OrthoWeb is available for analysis of gene sets. It is implemented using Java language. The server part is implemented using Spring framework (https://spring.io), the client part – Webix framework (https://webix.com).

Some results obtained with OrthoWeb were present on [8]. Currently is available by request to author.

Acknowledgment

The study is partially supported by RFBR, grant No. 20-04-00885 A, and Budget Project No. 0324–2019-0040-C-01.

References

[1] T. Domazet-Lošo, J. Brajković, and D. Tautz, "A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages," *Trends Genet.*, vol. 23, no. 11, pp. 533–539, Nov. 2007, doi: 10.1016/j.tig.2007.08.014.

[2] Z. S. Mustafin, S. A. Lashin, Y. G. Matushkin, K. V. Gunbin, and D. A. Afonnikov, "Orthoscape: a cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles," *BMC Bioinformatics*, vol. 18, no. S1, pp. 1–9, 2017, doi: 10.1186/s12859-016-1427-5.

[3] H.-G. Drost, A. Gabel, J. Liu, M. Quint, and I. Grosse, "myTAI: evolutionary transcriptomics with R," *Bioinformatics*, vol. 34, no. 9, pp. 1589–1590, May 2018, doi: 10.1093/bioinformatics/btx835.

[4] Z. Arendsee, J. Li, U. Singh, A. Seetharam, K. Dorman, and E. S. Wurtele, "phylostratr: A framework for phylostratigraphy," *bioRxiv*, p. 360164, Jul. 2018, doi: 10.1101/360164.

[5] P. Shannon *et al.*, "Cytoscape: A software Environment for integrated models of biomolecular interaction networks," *Genome Res.*, vol. 13, no. 11, pp. 2498–2504, 2003, doi: 10.1101/gr.1239303.

[6] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: New perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D353–D361, 2017, doi: 10.1093/nar/gkw1092.

[7] Z. Yang, "PAML 4: Phylogenetic analysis by maximum likelihood," *Mol. Biol. Evol.*, vol. 24, no. 8, pp. 1586–1591, 2007, doi: 10.1093/molbev/msm088.

[8] Z. S. Mustafin, V. I. Zamyatin, D. K. Konstantinov, A. V. Doroshkov, S. A. Lashin, and D. A. Afonnikov, "Phylostratigraphic Analysis Shows the Earliest Origination of the Abiotic Stress Associated Genes in A. thaliana," *Genes (Basel).*, vol. 10, no. 12, p. 963, Nov. 2019, doi: 10.3390/genes10120963.

DOI 10.18699/BGRS/SB-2020-146

Circular RNA host gene and orthologue prediction using the self-designed CircParser pipeline

Artem Nedoluzhko
Faculty of Biosciences and Aquaculture
Nord University
Bodø, Norway
artem.nedoluzhko@nord.no

Fedor Sharko
Complex of NBICS Technologies
NRC "Kurchatov Institute"
Moscow, Russia
fedosic@gmail.com

Golam Rbanni
Faculty of Biosciences and Aquaculture
Nord University
Bodø, Norway
golam.rbbani@nord.no

Anton Teslyuk
Complex for Simulation and Data
Processing for Mega-science
NRC "Kurchatov Institute"
Moscow, Russia
anthony.teslyuk@gmail.com

Ioannis Konstantinidis
Faculty of Biosciences and Aquaculture
Nord University
Bodø, Norway
ioannis.konstantinidis@nord.no

Jorge M.O. Fernandes
Faculty of Biosciences and Aquaculture
Nord University
Bodø, Norway
jorge.m.fernandes@nord.no

***Abstract* — Circular RNAs (circRNAs) are long noncoding RNAs that play a significant role in various biological processes, including embryonic development and stress responses. A number of circRNA de novo and host gene prediction tools are available to date, but their ability to accurately predict circRNA host genes is limited in the case of low-quality genome assemblies or annotations.**

Here we describe CircParser, a novel, easy to use and Unix/Linux pipeline for circular RNAs host gene and orthologue prediction using the blastn program and the freely available bedtools software. CircParser is most useful for circRNA host gene prediction analysis in whole transcriptomic datasets for low-quality assembled as well as poorly annotated genomes.

We demonstrate the prediction capacity of CircParser on a recently published transcriptomic data set from the wild and domesticated females of Nile tilapia (*Oreochromis niloticus*) fast muscle using the five most popular circRNAs *in silico* prediction tools – CIRI, CIRI2, CircExplorer2, find_circ, and circFinder (Fig. 1).

The Perl implementation of CircParser is available at https://github.com/SharkoTools/CircParser

Keywords* — *circular RNAs, circRNAs, host gene, tool, Nile tilapa

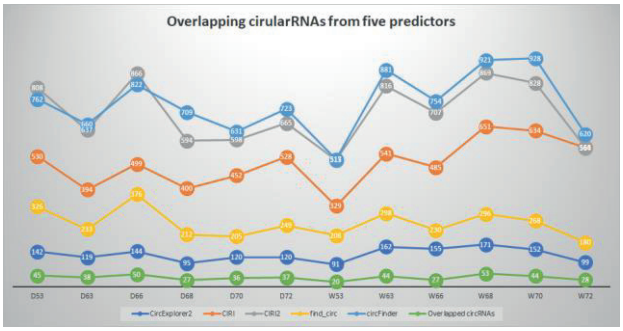


Fig. 1. Number of circular RNAs that have been predicted by CIRI, CIRI2, CircExplorer2, find_circ, circFinder, and that are common between all prediction algorithms.

Acknowledgment

This study has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [grant agreement no 683210] and from the Research Council of Norway under the Toppforsk programme [grant agreement no 250548/F20]. Fedor Sharko was supported by the RFBR (Russian Foundation for Basic Research) Grant No. 19-04-00033.

DOI 10.18699/BGRS/SB-2020-147

Phylogenetic analysis of poxviridae genomes using K-mer approach

Tatyana Nepomnyashchikh
SRC VB "Vector" Rospotrebnadzor,
Koltsovo, Russia
nepom@vector.nsc.ru

Denis Antonets
SRC VB "Vector" Rospotrebnadzor,
Koltsovo, Russia
antonec@yandex.ru

Tatyana Tregubchak
SRC VB "Vector" Rospotrebnadzor,
Koltsovo, Russia
tregubchak_tv@vector.nsc.ru

Alexander Shvalov
SRC VB "Vector" Rospotrebnadzor,
Koltsovo, Russia
shvalov_an@vector.nsc.ru

Elena Gavrilova
SRC VB "Vector" Rospotrebnadzor,
Koltsovo, Russia
gavrilova_ev@vector.nsc.ru

Rinat Maksyutov
SRC VB "Vector" Rospotrebnadzor,
Koltsovo, Russia
maksyutov_ra@vector.nsc.ru

Abstract — **There is a growing interest towards alignment-free approaches that use k-mers as genomic features. Large sizes of poxvirus genomes, their complex organization with repetitive regions and recombination sites, makes it difficult to produce accurate whole genome alignments especially for distantly related species. K-mer based trees built for 663 whole genome sequences were found to be in good agreement with alignment-based ones and correspond well to currently accepted taxonomy and evolution of poxviruses.**

Keywords — phylogenetic analysis, alignment free, k-mer frequency profile, poxviruses

Motivation and aim

K-mer frequency profiles analysis applied to different eukaryotic and prokaryotic sequences was found to be highly consistent with currently accepted taxonomy [1,2]. Recently, k-mer based methods were successively used to build a cladogram of RefSeq viral genomes [3], to analyze phylogeny of Ebola viruses and human papillomaviruses [4], influenza virus A [5] etc. The primary advantages of these methods are their low computational costs, and their ability to seamlessly compare highly divergent sequences, incomplete draft genomes sequences of different lengths, organization etc [3, 6].

Poxviruses are a family of large DNA viruses replicating in the cytoplasm. Poxvirus genomes contain highly conserved central region (~100 kbp) encoding viral proteins essential for RNA and DNA synthesis, protein processing, virion assembly and structural proteins, and highly variable terminal regions, bearing the genes encoding host range proteins, virulence factors and immunomodulators which are non-essential for virus growth *in vitro*. Despite these similarities in genome organization, their length varies from 139 kb in Orf virus to 289 kbp in Fowlpox virus (FPV) and GC-content varies from 25.3% in capripoxviruses to 65% in parapoxviruses [7]. Large sizes of poxvirus genomes and profound length variation, their complex organization with repetitive regions and recombination sites, makes it difficult to produce accurate whole genome alignments and especially for distantly related poxvirus species, thus we decided to use the alignment-free approaches and to compare the results with alignment-based ones.

Methods

Genome sequences were aligned with MAFFT [8] or Mauve software [9]. Phylogenetic analysis and dendrogram construction were performed with R package ape [10]. Dendrogram comparisons were done with R package dendextend [11]. K-mer frequencies analysis was performed with scripts written in Python programming language.

Results

VARV genomes, including three ancient strains recently extracted and sequenced by German scientists, were analyzed with common alignment-based approach. VARV strains were grouped according to geographic regions and isolation times. Two large main groups were distinguished. The largest one contained African strains, in turn, isolated in a separate cluster, strains BSH-74, BSH-75, IND-53, NEP-73. This group also included IND-64, KUW-67, AFG-70, SYR-72, IRN-72, YUSL-72, PKN-69 strains which formed a separate clade, Japanese strains JPN-46, JPN-51 and Indian IND-53 New Delhi formed a separate cluster. The second isolated group was formed by GAR-66, BRZ-66, UK-52 Butler, SL-69, GUI-69, BEN-69 and NIG-69 strains. African strains of this group also formed a separate cluster. Strain V563 (LT706528), isolated from remains about 60 years old, was attributed to the first, the most populated group. Strain V1588 (LT706529), isolated from remains about 160 years old, turned out to be evolutionarily closer to the second group. And the most ancient isolate VD21 (BK010317) extracted from almost 400 years old mummy was found to be isolated from all "modern" strains, although it was more closely related to the second group. The trees built using the k-mer frequency profiles with k ranging from 3 to 9 were found to be highly concordant with those built with alignment-based approach, and even tetranucleotide composition analysis resulted in cladograms representing the same major groups observed in alignment-based trees. The subsequent analysis of all collected poxvirus genomes with k-mer approach demonstrated a good agreement with current taxonomy data, even when k was set to 3. It was observed that cowpox viruses are the most heterogenic group. Camelpox was observed to be the most closely related to variola virus. We also captured other well-accepted evolutionary patterns of Poxviridae family. Thus, we can confirm that alignment-free approaches might be a valuable tool for at least a preliminary phylogenetic analysis of large DNA viruses. Another important point was that this analysis can be performed in a few tens of minutes on a laptop.

References

[1] Sims G.E. et al. (2009) Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions, Proc. Natl. Acad. Sci., 106:2677–2682.

[2] Bonham-Carter O. et al. (2014) Alignment-free genetic sequence comparisons: a review of recent approaches by word analysis, Brief. Bioinform., 15:890–905.

[3] Zhang Q. et al. (2017) Viral Phylogenomics Using an Alignment-Free Method: A Three-Step Approach to Determine Optimal Length of k-mer, Sci. Rep., 7:40712.

[4] Utro F. et al. (2019) A Quantitative and Qualitative Characterization of k-mer Based Alignment-Free Phylogeny Construction. In: Bartoletti M. et al. (eds) Computational Intelligence Methods for Bioinformatics and Biostatistics. CIBB 2017. Lecture Notes in Computer Science, Springer, 10834:19–31.

[5] Zhang Y. et al. (2018) Phylogenetic analysis of protein sequences based on a novel k-mer natural vector method, Genomics, in press (DOI: 10.1016/j.ygeno.2018.08.010).

[6] Han G.-B., Cho D.-H. (2018) Genome classification improvements based on k-mer intervals in sequences, Genomics, S0888-7543(18)30447-6.

[7] Gubser C. (2004) Poxvirus genomes: a phylogenetic analysis, J. Gen. Virol., 85:105–117.

[8] Katoh K., Standley D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability, Mol. Biol. Evol., 30:772–780.

[9] Darling A.C.E. (2004) Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements, Genome Res., 14:1394–1403.

[10] Paradis E., Schliep K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35:526–528.

[11] Galili T. (2015) dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. Bioinformatics, 31:3718–3720.

DOI 10.18699/BGRS/SB-2020-148

Mitogenomics and phylogenetics of vulnerable and endangered birds of genera *Aratinga* and *Psittacus*

Sofia Ochkalova
Applied Genomics Lab, SCAMT,
ITMO University,
Saint-Petersburg, Russia
ochkalova@scamt-itmo.ru

Aleksey Komissarov
Applied Genomics Lab, SCAMT,
ITMO University,
Saint-Petersburg, Russia
komissarov@scamt-itmo.ru

Taras Oleksyk
Biology Department, University of Puerto Rico, Mayaguez, Puerto Rico
Department of Biological Sciences, Oakland University, Rochester, USA
oleksyk@oakland.edu

***Abstract* — Among parrots, there are more threatened and endangered species than in any other avian order. Exploration of these outstanding birds exhibiting advanced cognitive abilities is a relevant scientific problem. In this study, we de novo assembled 19 mitochondrial genomes from Illumina raw reads, integrated them with information obtained from related genomes, and did a phylogenetic reconstruction of species from *Aratinga* and *Psittacus* genera. Additionally, we created a Python pipeline for assembly mtDNA bearing tandemly repeated regions from whole genome sequencing data. Then we annotated genes and finally performed phylogenetic analysis to provide insights into the evolution and demographic history of the species.**

Keywords* — *Aratinga, Psittacus, mitochondrial DNA, phylogenics, parrots

Motivation and aim

Motivation

Since long ago, parrots (*Psittaciformes*) have aroused human attention. Attractiveness of plumages, ability to mimic speech and intelligence make them popular pets. Unfortunately, the consequence is a radical drop in their populations due to the extensive illegal harvest of wild birds in addition to habitat loss. Our work is supposed to contribute to a major project on genome database of endangered parrots, which will help to develop a method to recognize species prohibited for trade using PCR tests.

Aim

We aimed to generate good-quality mitogenomes of the vulnerable tropical parrots from *Aratinga* and *Psittacus* genera fort phylogenetic analysis.

Materials and methods

Samples for DNA sequencing were obtained from ten and nine different individuals of unidentified species from *Aratinga* and *Psittacus* genera respectively. For all samples sequencing was performed using the Illumina HiSeq2000 platform in paired-end mode.

The initial QC of NGS data was performed using FastQC [1]. Reads were pre-processed for downstream analysis using in-house script tool for trimming Illumina adapters.

Mitogenome assembly utilized low-coverage whole genome data. We attempted to reconstruct mitochondrial genome using de novo assembler NOVOPlasty 3.8.3 for organelle genomes [2], but did not succeed. The output contig was only 15450 bp length and did not include 2 protein-coding genes. Assuming that the breakdown of the assembly was due to the inability of NOVOPlasty to master repeats in control region, we developed a pipeline to assemble mitogenomes considering this feature. Annotation was conducted with MITOS Annotator [3].

Phylogenetic tree was reconstructed using different combinations of mtDNA regions of our samples and other species from *Psittaciformes* (*Amazona*, *Ara*, *Psittacara*, *Eupsittula*, *Arapornis*). Extracted genes were aligned using MAFFT [4], processed by Gblocks to remove hypervariable regions and concatenated. Obtained alignments were used to reconstruct a maximum likelihood tree with RAxML v8.2 [5].

Results

The circular assembled mtDNA size is 17,011 bp for *Aratinga* and about 16,850 bp for *Psittacus*. Similar to the existing data, we identified 13 predicted protein-coding genes (PCGs), two ribosomal RNA (rRNA) genes, and 22 tRNAs and one long tandem repeat 2 kb long and located before control region.

Phylogenetic reconstruction revealed that the sampled individuals of *Aratinga* genus refer to *Guaruba guarouba* monotypic group. For samples of *Psittacus* genus we discovered two different subspecies of *Psittacus erithacus*.

Acknowledgment

Russian Foundation for Basic Research Grants 17-00-00144 as part of 17-00-00148.

References

[1] S. Andrews, F. Krueger, A. Seconds-Pichon, F. Biggins, and S. Wingett, “FastQC. A quality control tool for high throughput sequence data. Babraham Bioinformatics,” *Babraham Institute*, 2015. .

[2] N. Dierckxsens, P. Mardulyn, and G. Smits, “NOVOPlasty: De novo assembly of organelle genomes from whole genome data,” *Nucleic Acids Res.*, 2017.

[3] M. Bernt *et al.*, “MITOS: Improved de novo metazoan mitochondrial genome annotation,” *Mol. Phylogenet. Evol.*, 2013.

[4] K. Katoh, K. I. Kuma, H. Toh, and T. Miyata, “MAFFT version 5: Improvement in accuracy of multiple sequence alignment,” *Nucleic Acids Res.*, 2005.

[5] A. Stamatakis, “RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies,” *Bioinformatics*, 2014.

DOI 10.18699/BGRS/SB-2020-149

The genetic component of the human embryonic selection: uncovering of the strength and main targets

Sergey Oreshkov
Center for Mitochondrial Functional Genomics, IKBFU, Kaliningrad, Russia
s.oreshkov@live.com

Elisaveta Zezyulya
Center for Mitochondrial Functional Genomics, IKBFU, Kaliningrad, Russia
liza.zez1@hotmail.com

Konstantin Gunbin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Center for Mitochondrial Functional Genomics
IKBFU, Kaliningrad, Russia
genkvg@gmail.com

Dmitrii Iliushchenko
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University
Kaliningrad, Russian Federation
iliushchenkodmitrii@gmail.com

Evgenii Tretiakov
Center for Brain Research, Medical University of Vienna
Vienna, Austria
Orcid 0000-0001-5920-2190

Konstantin Popadin
School of Life Sciences
Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
Center for Mitochondrial Functional Genomics, IKBFU, Kaliningrad, Russia
konstantinpopadin@gmail.com

Abstract — **Runs of Homozygosity(ROH) are sites in the two homologous chromosomes that are identical by descent in an individual. Excess of ROHs carries a risk associated with various diseases. At the same time, long ROHs in healthy individuals are enriched with slightly deleterious variants, which are constantly filtered out by selection. Hence, distribution and length of Runs of Homozygosity is affected by recombination and purifying selection. Data from families with different levels of consanguinity allows us to access both these effects and their influence on highly autozygoteous pedigree. Understanding connections between runs of homozygosity, recombination and purifying selection will help us to better understand evolutionary processes in highly inbred populations.**

Keywords — Runs of Homozygosity, consanguinity, recombination, purifying selection, in-silico modeling

Introduction

Human embryonic selection is extremely poorly known since the majority of miscarriages occur very early and there is no possibility to collect a representative dataset of miscarried samples for consequent genetic analyses. All such collections will be (i) biased towards late miscarriages and (ii) will not be able to shed light on potential selection during germline differentiation: formation of oocytes and spermatozoids. The best approach to estimate the strength and targets of human embryonic selection is to compare parental and offspring genomes (hereafter "trios") in order to uncover over- or under- inherited alleles. It is rational to start to develop and test tools dissecting the transmission disequilibrium from parents to offspring using collection of trios with expected strong miscarriage rate - such as consanguineous families.

Methods

Consanguinity increases the risk of expression of deleterious autosomal recessive mutations causing a correlation between the coefficient of inbreeding, F [1], and the probability of offspring to survive. The more closely related the parents, the higher the mortality of the children [2]. Importantly, the nominal coefficient of inbreeding, F, derived from pedigree, approximates an expected fraction of loci, identical by descent, however the observed coefficient of inbreeding in each particular case, Fo, can significantly fluctuate between full siblings due to stochasticity in the number and position of recombination events during meiosis as well as random gamete fusion. For example, for siblings from a first cousin marriage, the nominal coefficient of inbreeding is 0.0625 (1/16), however its standard deviation is very high (0.0243), causing great variation in probability of survival. Thus, variation in F provides a material for selection inside each family and we expect that Fo (healthy siblings) < Fo (affected siblings) < Fo (expected siblings).

As a molecular metric approximating the observed coefficient of inbreeding, Fo, we use the portion of genome covered by long Runs Of Homozygosity (ROHs) [3]. Runs of homozygosity (ROH) - stretches of the two homologous chromosomes within the same individual that are identical by descent. Long ROH are enriched in deleterious variants in healthy individuals[4] and an excess of ROHs is a risk factor for Parkinson disease, Alzheimer disease, Schizophrenia, speech delay and autism and the degree of intellectual disability. This implies that long ROHs carry one severe or many mildly-deleterious recessive alleles and thus there is a selection force against long ROHs in the human population.

Datasets

Main dataset is genotype data of 50 consanguineous families from the study of Makrythanasis et al. [5]. Additionally, there is information about pedigree(the tree) and exomes of affected and unaffected siblings.

All families have at least one unaffected offspring and at least two individuals with intellectual disability or development delay.

Results

Using our main dataset we have performed our insilico pipeline and compared the total length of long ROHs in healthy versus affected (with intellectual disabilities, ID) and expected (in-silico generated) siblings. For in-silico kids derived randomized models for expected siblings, which captures distribution properties of heterozigotious recombination in offspring. We produced several models of synthetic offspring, which capture different aspects of the source dataset. In different models we capture

recombination and selection features such as different distribution of ROHs and different recombination rate. We have observed that healthy siblings are indeed characterized by the decreased total length of ROHs as compared to affected. Additionally we have demonstrated an excess of ROH in expected (generated in silico) versus born siblings (both affected and unaffected). Altogether it suggests an action of purifying selection against long ROHs.

Discussion

In this project we concentrated on the estimation of the strength of the short-term selection against the autozygome (the fraction of the genome that has been biparentally inherited from the same ancestor). For this,we compared the fractions of autozygous regions in genomes of unaffected, affected and theoretically expected (in silico modeled) siblings in each consanguineous family. Selection against autozygous regions manifested itself as a decrease of this fraction in unaffected versus affected and expected siblings.

The strength of the decrease can depend on phenotypes of the affected offspring (Mendelian or complex disease) as well as the level of consanguinity of parents and the previous history of the parental consanguinity. We conclude that this approach is effective in uncovering the strength and targets of human embryonic selection and plan to extend it towards analysis of polygenic risk scores (PRS) of parents and offspring as well as analysis of new human trios.

References

[1] Wright S., "Coefficients of inbreeding and relationship" The American Naturalist 56, pp 330-338, 1922.

[2] Newton E. Morton, James F. Crow, and H. J. MullerJ. Clerk Maxwell, "An estimate of the mutational damage in man from data in consanguineous marriages" PNAS, November 1, 42(11), 1956, pp.855-863.

[3] McQuillan R et al. "Runs of homozygosity in European populations." Am J Hum Genet. Sep;83(3), 2008, pp 359-72.

[4] Szpiech ZA at al. "Long runs of homozygosity are enriched for deleterious variation." Am J Hum Genet. Jul 11;93(1), 2013 , pp. 90-102.

[5] Makrythanasis P., et al. "Diagnostic exome sequencing to elucidate the genetic basis of likely recessive disorders in consanguineous families." Hum Mutat. Oct;35(10), 2014 , pp. 1203-10.

DOI 10.18699/BGRS/SB-2020-150

Distribution of Runs Of Homozygosity (ROHs) along the human genome is shaped by recombination and purifying selection

K. Popadin
EPFL, Lausanne, Switzerland
konstantinpopadin@gmail.com

E. Zezyulya
IKBFU, Kaliningrad, Russia
liza.zez1997@gmail.com

D. Iliushchenko
IKBFU, Kaliningrad, Russia
serioussamkrsk@gmail.com

A. Reymond
University of Lausanne, Lausanne, Switzerland
alexandre.reymond@unil.ch

Abstract — **Whereas the higher fraction of the human genome, covered by Runs Of Homozygosity (ROHs), was associated with decreased height, decreased educational attainment and other phenotypes, the mapping position of the contributing ROHs was not assessed in details. Here, analysing genome-wide distribution of ROHs in offsprings of 100 consanguineous families we observed a non uniform distribution. Using multiple linear model we demonstrated that this variation could be explained by (i) recombination level, i.e. the higher the recombination in a given region, the less the number of ROHs; and (ii) selection, i.e. the higher the pLI score (which estimates loss of function intolerance) of the embedded genes the less the number of ROHs.**

The observed selection component might be explained by elimination of carriers of homozygous recessive loss of function mutations in genes with high pLI. To examine the selection component further we stratified ROHs by age and length following a previously suggested classification: class A (old and short: < 0.6 Mb), class B (intermediate in age and length: 0.6 – 1.6 Mb) and class C (young and long: > 1.6 Mb) and rerun our analyses. We observed that while the distribution of all three classes is shaped by recombination, only the class C ROHs showed a strong negative relationship with pLI, suggesting that younger ROHs are under stronger selection. Altogether our results provide additional metrics (properties of affected genes, ROH classes), which should be taken into account inferring deleterious effect of ROHs in GWAS and genetic medicine.

Key words — ROH, Human genomes

Methods

Having 100 consanguineous pedigrees with genotyped parents and offspring we called all ROHs using default PLINK settings (DNA regions with at least 50 consecutive homozygous SNPs). Thereafter we focused on offspring ROHs and analyzed their distribution along the human genome taking into account recombination map and properties of embedded genes.

Results

The distribution of three classes of ROHs (short, intermediate, long) is highly non-uniform along human chromosomes (Fig 4). We hypothesized that this distribution is shaped by recombination, which mechanistically disrupts ROHs and thus decreases their abundance in the recombination hotspots, and by purifying selection, which eliminates ROHs from sensitive genomic regions with embedded haploinsufficient genes. Using multiple linear model we demonstrated that:

Coverage(ROH) ~ -2.27*RecombinationRate-0.6*ProbabilityOfLossOfFunctionIntolerance (all p values are < 10-6; coefficient are scaled). Interestingly, the distribution of the longest ROHs (class C) demonstrates the strongest dependence on both recombination and LoF intolerance. We conclude that selection in line with recombination shapes the ROH distribution along the human genome.

Acknowledgment

We are grateful for the data provided to P. Makrythanasis (Academy of Athens, Athens, Greece), M. Ansar, S.E. Antonarakis (University of Geneva, Geneva, Switzerland).

References

[1] Hamamy, H. (2012) Consanguineous marriages: preconception consultation in primary health care settings. J Community Genet. 3 (3): 185–92.

[2] McQuillan, Ruth, et al. (2008) Runs of homozygosity in European populations. *The American Journal of Human Genetics* 83.3, 359-372.

[3] Pemberton, Trevor J., et al. (2012) Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics* 91.2, 275-292.

DOI 10.18699/BGRS/SB-2020-151

Candidate SNP markers of rheumatoid arthritis changing the affinity of TATA-binding protein for the human gene promoters expo disruptive selection of immunoactivative and immunosuppressive genenets that provoke and prevent this disorder, respectively, as if it could be a self-domestication syndrome

Natalya Klimova
Molecular Genetics Department
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
klimova@bionet.nsc.ru

Irina Chadaeva
Systems Biology Department
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
ichadaeva@bionet.nsc.ru

Evgeniya Oshchepkova
Systems Biology Department
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
nzhenia@bionet.nsc.ru

Dmitry Oshchepkov
Systems Biology Department
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
diman@bionet.nsc.ru

Mikhail Ponomarenko
Systems Biology Department
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
pon@bionen.nsc.ru

academician Vladimir Kozlov
Research Institute of Fundamental
and Clinical Immunology
SB RAS, Novosibirsk, Russia
niiki01@online.nsk.su

Abstract — **Here we conducted a computational genome-wide study of the all known single-nucleotide polymorphism (SNP) of 70 bp proximal promoters of 70 human rheumatoid arthritis (RA)-related genes that displayed disruptive natural selections of immunoactivative or immunosuppressive genes raising or reducing risks of RA, respectively, as if it maybe a domestication syndrome. That is why, we confirmed it using the genome-wide RNA-seq of the differentially expressed genes (DEGs) within hypothalamus of adult male foxes (Vulpes vulpes) of two unique outbred lines bred in aggressiveness and tameness as an animal model of human diseases at the significance levels $p < 0.0025$ of Pearson's $\chi 2$-test and $p < 0.05$ at Fisher's exact test independently from one another as well as in addition to $p < 0.05$ of Bonferroni's correction as the DEG threshold used.**

Keywords *— rheumatoid arthritis, human, gene, promoter, TATA-binding protein (TBP) binding site (TATA box), single nucleotide polymorphism (SNP), candidate SNP marker, verification.*

Background

Rheumatoid polyarthritis (RA) is an autoimmune disease with autoantibodies (e.g., antibodies to citrullant antigens) and proinflammatory cytokines (e.g., TNF-α, IL-6) that are involved in the induction of chronic synovitis, bone erosion, followed by deformity [1]. Currently, it is commonly accepted that immunopathogenesis is mostly contributed by the mechanisms of the breakdown of immune tolerance to its own antigens that is characterized by an increase in the activity of T-effector cells, causing RA symptoms [2]. On the contrary, a low activity of regulatory cells (e.g., regulatory T-cells (Treg) and myeloid suppressor cells) is met in RA [3]. That is why, to say that it is the immunosuppressor cell deficit that is the central feature in RA pathogenesis [4]. It is thought that lifestyle and living conditions define a half of the RA risks and genetic susceptibility to RA do for another half [5]. Additionally, immune cells of adaptive memory in a given individual keep information on all diseases he / she survived that increase his / her resistance to them in future [6]. That is why here using our SNP_TATA_Comparator [7], we predicted RA-related candidate SNP markers, found regularities in their genome-wide frequencies, and tested them in vivo using an animal (rat) model of the human health.

Materials and methods

We extracted 1939 SNPs of 70 bp proximal promoters in front of transcription start sites of 70 RA-related genes using UCSC Genome Browser [8], dbSNP [9] and GRCh38/hg38 assembly of the human reference genome [10] and, next, treated them using our SNP_TATA_Comparator [7]. Finally, we verified the obtained results using the earlier published experimental data on the differentially expressed genes (DEGs) of males tame foxes (*Vulpes vulpes*) as compared to males aggressive foxes, both artificial breeding [11]. To this end, we have independently applied both Pearson's χ^2-test and Fisher's exact test using the standard package Statistica (StatSoft^{TM}, Tulsa, USA).

Results and Discussions

First, using the only clinical SNP marker rs2276109 of RA [12] in *MMP212*, we tested adequacy of candidate SNP markers of RA-risks found by SNP_TATA_Comparator [7] Then, for 646 SNPs in question within 27 human genes clinically associated with disorders comorbid to RA, we found 87 and 77 candidate SNP markers elevating (e.g., rs1185314734) or reducing (e.g., rs1190659847) expression of these genes, respectively, among which 106 and 58 ones can worsen (e.g., rs766797386, *ESR2* [13, 14]) or relieve (e.g., rs549858786, *IL1B*) RA. It means the genome-wide neutral drift [15-17] towards susceptibility to RA in line with the common view that diversity of adaptive memory immune cells of an individual grow with each inflammation he/she survived that improves his/her resistance to them, but this immune memory growth raises risks of impaired autoimmune tolerance (e.g., RA risks). Thus, by the same way we confirmed this result using ten genes most GWAS-

associated with RA [18] (e.g., rs1035054020, *NPY* [19]), and, independently, using eight immunoactivative genes (e.g., rs56317732, *IL9R* [20]). Contrary, we found 66 and 113 candidate SNP markers aggravating and relieving RA, respectively (e.g., rs72990754, *IL1R2* [21] and rs546214861, *TGFB2* [22]) using 25 immunosuppressive genes. This difference immunoactivative *vs* immunosuppressive genes means disruptive natural selection as if RA could be a self-domestication syndrome [23]. That is why, we verified this consequence of our results using the published experimental data on DEGs in hypothalamus of males tame foxes in a comparison with those of aggressive ones, both artificial breeding [11]. Three *Npy*, *Tgfb2* and *Il9* among these five DEGs are superexpressing in tame male foxes, whereas both rest *Esr2* and *Il1r2* are underexpressing there, all of which without exemptions are five clinically proven physiological markers of RA-susceptibility and *vice versa*, as cited above [13, 14, 19 - 22]. So that, the standard 2x2-table has four values: $N_{TAME;\uparrow RA} = 5$, $N_{TAME;\downarrow RA} = 0$, $N_{AGGRESSIVE;\uparrow RA} = 0$, and $N_{AGGRESSIVE;\downarrow RA} = 5$, - which means that the association between tameness and RA-susceptibility is statistically reliable at both $p < 0.0025$ of Pearson's $\chi 2$-test and $p < 0.05$ at Fisher's exact test independently from one another as well as in addition to $p < 0.05$ of Bonferroni's correction as the independent identification threshold used for of these DEGs.

Conclusions

Here we studied 1939 SNPs of 70 human RA-related genes that resulted in 526 candidate SNP markers of RA pinpointing RA as a self-domestication syndrome. Finally, we confirmed this link between domestication and RA using the RNA-seq of tame *vs* aggressive adult male foxes [11].

ACKNOWLEDGMENTS

Manuscript writing was supported by project #0324-2019-0040-C-01 (for MP); data analysis - Russian Federal Science & Technology Program for the Development of Genetic Technologies (for DO, EO); the result interpretation - project #0324-2019-0042-C-01 (for NK, IC).

References

[1] J. Smolen et al., “Rheumatoid arthritis therapy reappraisal: strategies, opportunities and challenges” Nat Rev Rheumatol, vol. 11, pp. 276-289, 2015.

[2] J. Sarkander et al., “Vaccination to gain humoral immune memory” Clin Transl Immunology, vol. 5, pp. e120, . 2016.

[3] O. Alsaed et al., “Seronegative bilateral symmetrical inflammatory polyarthritis: think twice before starting immunosuppression” Eur J Case Rep Intern Med, vol. 5, pp. 000895, 2018.

[4] C. Lawson et al., “Early rheumatoid arthritis is associated with a deficit in the CD4+CD25high regulatory T cell population in peripheral blood” Rheumatology (Oxford), vol. 45, pp. 1210-1217, 2006.

[5] N. Nair et al., “DNA methylation as a marker of response in rheumatoid arthritis” Pharmacogenomics, vol. 18, pp. 1323-1332, 2017.

[6] J. Sarkander, et al., “Vaccination to gain humoral immune memory” Clin Transl Immunology, vol 5, pp. e120, . 2016.

[7] M. Ponomarenko et al.. “How to use SNP_TATA_Comparator to find a significant change in gene expression caused by the regulatory SNP of this gene's promoter via a change in affinity of the TATA-binding protein for this promoter” Biomed. Res. Int., vol. 2015. pp. 35983004625. 2015.

[8] M. Haeussler et al., “Navigating protected genomics data with UCSC Genome Browser in a box” Bioinformatics, vol. 31. pp. 764-766. 2015.

[9] S. Sherry et al., “dbSNP: the NCBI database of genetic variation” Nucleic Acids Res, vol. 29, pp. 308-311, 2001.

[10] D. Zerbino et al., “The Ensembl regulatory build” Genome Biol, vol. 16: pp. 56. 2015.

[11] J. Hekman et al., “Anterior pituitary transcriptome suggests differences in ACTH release in tame and aggressive foxes” G3: Genes, Genomes, Genetics (Bethesda), vol. 8, pp. 859-873, 2018.

[12] G. Hunninghake et al., “MMP12, lung function, and COPD in high-risk populations.” N Engl J Med, vol. 361, pp. 2599-2608, . 2009

[13] S. Philips et al.,. 2012. “Functional characterization of a genetic polymorphism in the promoter of the ESR2 gene”. Horm Cancer, vol. 3, pp. 37-43, 2012.

[14] C. Armstrong et al., “A novel shift in estrogen receptor expression occurs as estradiol suppresses inflammation-associated colon tumor formation”. Endocr Relat Cancer, vol. 20, pp. 515-525, 2013.

[15] J. Haldane. “The cost of natural selection”. J Genet, vol. 55. pp. 511–524. 1957.

[16] M. Kimura. “Evolutionary rate at the molecular level”. Nature, vol. 217. pp. 624-626. 1968.

[17] 1000 Genomes Project Consortium, “An integrated map of genetic variation from 1.092 human genomes”. Nature, vol. 491. pp. 56-65. 2012.

[18] J. Karami et al., “Genetic implications in the pathogenesis of rheumatoid arthritis; an updated review” Gene, vol. 702, pp. 8-16, 2019.

[19] A. Stofkova et al. “Activation of hypothalamic NPY, AgRP, MC4R, and IL-6 mRNA levels in young Lewis rats with early-life diet-induced obesity”. Endocr Regul, vol. 43, pp.:99-106, 2009.

[20] S. Raychaudhuri et al. “IL-9 receptor: regulatory role on FLS and pannus formation” Cytokine vol. 111, pp. 58-62, 2018.

[21] T. Ocsko et al. “Transcription factor Zbtb38 downregulates the expression of anti-inflammatory IL1r2 in mouse model of rheumatoid arthritis” Biochim Biophys Acta Gene Regul Mech, vol. 1861, pp. 1040-1047, 2018.

[22] S. Um et al. “TGF-β2 downregulates osteogenesis under inflammatory conditions in dental follicle stem cells” Int J Oral Sci, vol. 10, pp. 29, 2018.

[23] D.K. Belyaev. “The Wilhelmine E. Key 1978 invitational lecture. Destabilizing selection as a factor in domestication” J Hered, vol. 70, pp. 301–308, 1979.

DOI 10.18699/BGRS/SB-2020-152

Dynamics and hypotheses of gene order shifts in mitochondrial genomes of Baikalian amphipods

Elena A. Sirotinina
Lab of gene systematics
LIN SB RAS
Irkutsk, Russia
haleo.inc@gmail.com

Dmitry Yu. Sherbakov
Lab of gene systematics
LIN SB RAS
Irkutsk, Russia
dysh007@gmail.com

Elena V. Romanova
Lab of gene systematics
LIN SB RAS
Irkutsk, Russia
elena_romanova@lin.irk.ru

***Abstract* — In our study we describe the variabilities in the mitochondrial gene order of amphipod species. Using phylogenetic inference and the methods of gene order rearrangement counting we identify species with mostly rearranged mitochondrial genomes and try to explain the reasons for such variability.**

Keywords* — *Lake Baikal; mitochondrial genomes; amphipods; gene rearrangement

Motivation and aim

Motivation

Gene order in the mitochondrial genomes (mt genomes) of invertebrates is shown to be different. The dynamics of such changes are usually studied by analysis of different organisms' lineages. It is notable that in some animal lineages the mitochondrial gene order maintained the same for the long periods, whereas other lineages demonstrate a significant variation of this feature [1]. Amphipods can be a good model for studying mitochondrial gene order dynamics because of a relatively large number of species with sequenced mt genomes (about 100) available and due to the variability of their mt gene order. Recent studies showed that endemic amphipods from Lake Baikal possess rearranged mt genomes in comparison to the non-Baikalian species [1, 2, 4].

Aim

Amphipod species with currently sequenced mt genomes belong to different taxa and occupy different ecological niches, which makes it possible to assess whether mt gene order rearrangements correlate with ecological features of species, their life-history traits, time of lineage divergence, etc. For instance, there was shown a positive correlation between the rate of gene rearrangements with the level of nucleotide substitutions in their protein-coding genes [5, 6]. We also recently showed the significant acceleration in mutation rate in both lineages of Baikalian amphipods in comparison to species from gen. *Gammarus* [7].

There are two mechanisms of the new gene orders emergence: 1) an erroneous duplication of the whole mt genome sequence during replication by "rolling ring" mechanism, associated with the DNA polymerase slippage on specific sites or repeats, and 2) a partial tandem duplication and random loss (TDRL) of a certain genome region [8]. To measure such gene rearrangements quantitatively it is necessary to determine the number of mutational steps (permutations) from the ancestral gene order pattern to ones for the modern species. An ancestral gene pattern for any lineage is identified using phylogenetic reconstructions.

Methods

We defined an ancestral mt gene order pattern for all amphipods performing phylogenetic inference based on amino acid sequences of 13 mt protein-coding genes. To estimate possible scenarios of gene order changing, including the minimum number of elementary gene permutations, we used CREx and SORT2 [9, 10]. In some cases we assessed the changes manually by taking the number of displaced genes into account individually, and also counting the number of moved gene blocks grouped based on the proximity of their location.

Results

The analysis showed that Pancrustacean pattern (the pattern common for all crustaceans and insects) generally used as a reference in mt gene rearrangements studies [1, 3] differ from deduced Amphipoda pattern by the location of some tRNA genes. The Amphipoda pattern deduced in our study turned out to be the same as one for gen. *Gammarus* species. The phylogenetic analysis showed that this pattern was ancestral for both lineages of Baikalian amphipods. Thus the rearranged patterns of the majority of modern Baikalian amphipods were acquired during the evolutionary process in Lake Baikal. Notably, Baikalian amphipods from the second lineage (the one includes the majority of species which have very divergent morphological and ecological traits) possess gene order alteration affecting mainly tRNA gene positions. Whereas amphipods from the first lineage, which comprise much less number of species, having a smooth body and inhabiting mainly in shallow water, have much more complex alteration in their mt gene order, affecting the positions of the protein-coding and ribosomal genes [2]. We hypothesize that the fixation of highly rearranged mt genome becomes possible due to relaxed purifying selection in populations with low genetic diversity and low estimated effective population size. The shallow-water amphipod species in Lake Baikal could probably be more prone to the dramatic decrease of their population and consequently genetic diversity due to environmental changes in the Lake.

Acknowledgment

The work was supported by the governmentally funded project 0345e2019e0004 (AAAA-A16-116122110060-9).

References

[1] F. Kilpert, L. Podsiadlowski. “The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features”, BMC Genomics, vol.7, p.241, 2006.

[2] E. V. Romanova, V. V. Aleoshin, R. M. Kamaltynov, K. V.Mikhailov, M. D. Logacheva, E. A. Sirotinina et al. “Evolution of mitochondrial genomes in Baikalian amphipods”, BMC genomics, vol. 17, pp. 1016, 2016.

[3] L. Krebes, R. Bastrop. “The mitogenome of *Gammarus duebeni* (Crustacea Amphipoda): A new gene order and non-neutral sequence evolution of tandem repeats in the control region”, Comp Biochem Physiol, vol.7(2), pp.201-11, 2012.

[4] Y. Shen, Q. Kou, Z. Zhong, X. Li, L. He, S .He, X. Gan. “The first complete mitogenome of the South China deep-sea giant isopod *Bathynomus sp.* (Crustacea: Isopoda: Cirolanidae) allows insights into the early mitogenomic evolution of isopods”, Ecol Evol., vol.16, pp. 1869-1881, 2017.

[5] R. Shao, M. Dowton, A. Murrell, S. C. Barker. “Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects”, Mol Biol Evol, vol.20(10), pp.1612-1619, 2003.

[6] S. Fourdrilis, A. M. de Frias Martins, T. Backeljau. “Relation between mitochondrial DNA hyperdiversity, mutation rate and mitochondrial genome evolution in *Melarhaphe neritoides* (Gastropoda: Littorinidae) and other”, Sci Rep., vol. 8 (1), p.17964, 2018.

[7] E. V. Romanova, D. Y. Sherbakov. "Different rates of molecular evolution of mitochondrial genes in Baikalian and non-Baikalian amphipods." Limnology and Freshwater Biology, vol.6, pp.339-344, 2019.

[8] Y. Xia, Y. Zheng, R. W. Murphy, X. Zeng. “Intraspecific rearrangement of mitochondrial genome suggests the prevalence of the tandem duplication-random loss (TDLR) mechanism in *Quasipaa boulengeri*”, BMC Genomics, vol.17, p. 965, 2016.

[9] M. Bernt, D. Merkle, K. Ramsch, G. Fritzsch. “CREx: Inferring Genomic Rearrangements Based on Common Intervals”, Bioinformatics, vol. 23(21), pp.2957-2958, 2007.

[10] Y. L. Huang, C. C. Huang, C. Y. Tang, C. L. Lu. “SoRT2: a tool for sorting genomes and reconstructing phylogenetic trees by reversals, generalized transpositions and translocations”, Nucleic Acids Res, vol.38, pp.221-227, 2010.

DOI 10.18699/BGRS/SB-2020-153

Mitochondrial genetics of amphipods: revealing mechanisms of diversity

Elena V. Romanova
Laboratory of Molecular Systematics
LIN SB RAS, Irkutsk, Russia
elena_romanova@lin.irk.ru

Maria D. Logacheva
Moscow State University
Moscow, Russia, IITP RAS
Moscow, Russia

Yurij S. Bukin
Laboratory of Molecular Systematics
LIN SB RAS, Faculty of Biology and
Soil Studies Irkutsk State University
Irkutsk, Russia

Elena A. Sirotinina
Laboratory of Molecular Systematics
LIN SB RAS
Irkutsk, Russia

Kirill V. Mikhailov
Moscow State University
Moscow, Russia
IITP RAS
Moscow, Russia

Vladimir V. Aleoshin
Moscow State University
Moscow, Russia
IITP RAS
Moscow, Russia

Dmitry Yu. Sherbakov
Laboratory of Molecular Systematics
LIN SB RAS, Irkutsk, Russia
Faculty of Biology and Soil Studies
Irkutsk State University
Irkutsk, Russia

***Abstract* — We performed a thorough analysis of mitochondrial (mt) genomes of amphipods with the focus on endemic Baikalian species. Possible causes and mechanisms explaining an unusually high variability in mt genomes of Baikalian amphipods in comparison to that of non-Baikalian species are discussed.**

Keywords* — *mitochondrial genomes, amphipods, Lake Baikal

Motivation and aim

Motivation

Mitochondrial (mt) genome sequences are widely used in evolutionary and phylogenetic studies. The architecture of mt genomes (i.e. their shape, length, gene order) can significantly vary in different organisms' groups [1]. The studying of the lineages with different mt genome architecture can help to reveal causes and mechanisms leading to such diversity.

Aim

In our previous study, we sequenced mt genomes of nine endemic amphipod (Crustacea) species from Lake Baikal [2]. We further sequenced mt genomes of Holarctic species *Gammarus lacustris* and pelagic Baikalian species *Macrohectopus branickii*. Amphipods is the most numerous and diverse group of Baikalian animals [3]. Genetic studies showed that modern Baikalian amphipod diversity (about 350 species) in the lake appeared by means of adaptive radiation of gammaroidean ancestors which had two independent invasions in the lake [3, 4, 5]. We showed that mt genomes of Baikalian amphipods significantly vary in length, gene order and gene number in comparison to the modern amphipods of gen. *Gammarus*, which are their nearest relatives. Thus we applied different bioinformatics approaches to discover the mechanisms of mt genome architecture diversity in Baikalian amphipods.

Methods

We had performed phylogenetic inference of amphipod species with available mt genomes (100 species in total) based on amino acid sequences of their mt protein-coding genes [6]. The scenarios of mt gene order rearrangements were estimated using CREx [7]. The prediction of tRNA genes was made by MitFi [8]. For this analysis, we developed amphipod-specific models which showed a better performance than the default metazoan models implemented in MitFi. The non-coding parts of amphipod mt genomes were tested to contain repeated sequences and open reading frames. We used a custom R script to assess rates of molecular evolution in Baikalian amphipods and the group of gen. *Gammarus*.

Results

Phylogenetic analysis allowed us to deduce an ancestral mt gene pattern for both lineages of Baikalian amphipods, which appeared to be the same as one for species of gen. *Gammarus* group. Baikalian species show a drastic diversity in their mt genome lengths (from 14370 to 42256 bp.). A pelagic species *M. branickii* has the longest mt genome. The largest part of this genome consists of the numerous repetitive non-coding regions, which, presumably arose due to the proliferation of ribosomal genes. Mt genome sequences of seven out of twelve Baikalian amphipods species possess from one to four additional tRNA genes. Four species have tRNA genes that underwent to remolding (changing tRNA gene identity through a mutation in their anticodon sequences) [9, 10]. Among 88 non-Baikalian amphipod species, only five species have additional tRNA genes in their mt genomes and four of them have tRNA remolding cases. The frequent changes of gene number and order as well as other "peculiarities" in mt genomes of Baikalian amphipods in comparison to species of gen. *Gammarus* may be explained by the frequent sequence duplication and loss events happening in the former group. Thus we may assume that the specific environmental conditions in Lake Baikal allow the appearance and maintenance of such mt genomes without loss of species fitness.

Acknowledgment

The work was supported by the governmentally funded project 0345e2019e0004 (AAAA-A16-116122110060-9).

References

[1] D.V. Lavrov and W. Pett, “Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages, " Genome Biol. Evol., vol. 8, pp. 2896–2913, 2016.

[2] E.V. Romanova, V.V. Aleoshin, R.M. Kamaltynov, K.V. Mikhailov, M.D. Logacheva, E.A. Sirotinina et al. “Evolution of mitochondrial genomes in Baikalian amphipods”, BMC genomics, vol. 17, pp. 1016, 2016.

[3] R.M. Kamaltynov, “Amphipoda: Gammaroidea in Angara and Yenisei rivers,” in Index of animal species inhabiting Lake Baikal and its catchment area, Vol. 2, O.A.Timoshkin, V.I. Proviz, T.Y. Sitnikova, Z.V. Slugina, N.G. Melnik, Eds. Novosibirsk: Nauka, 2009, pp. 297–329.

[4] K.S. Macdonald Iii, L. Yampolsky, J.E. Duffy, “Molecular and morphological evolution of the amphipod radiation of Lake Baikal,” Mol. Phylogenet. Evol., vol. 35, pp. 323–343, 2005.

[5] S.A. Naumenko, M.D. Logacheva, N.V. Popova, A.V. Klepikova, A.A. Penin, G.A. Bazykin et al., “Transcriptome-based phylogeny of endemic Lake Baikal amphipod species flock: fast speciation accompanied by frequent episodes of positive selection, ” Mol. Ecol., vol. 26, pp. 536–553, 2017.

[6] L.T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, “IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies,” Mol. Boil. Evol., vol. 32, pp. 268–274, 2014.

[7] M. Bernt, D. Merkle, K. Ramsch, G. Fritzsch, M. Perseke, D. Bernhard et al., “CREx: inferring genomic rearrangements based on common intervals,” Bioinformatics, vol. 23, pp. 2957–2958, 2007.

[8] F. Jühling, J. Pütz, M. Bernt, A. Donath, M. Middendorf, C. Florentz, et al., “Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements,” Nucleic Acids Res., vol. 40, pp. 2833–2845, 2012.

[9] P. Cantatore, M.N. Gadaleta, M. Roberti, C. Saccone, A. C. Wilson, “Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes,” Nature, vol. 329, pp. 853–855, 1987.

[10] E.V. Romanova, Y.S. Bukin, K.V. Mikhailov, M.D. Logacheva, V.V. Aleoshin, D.Y. Sherbakov, “Hidden cases of tRNA gene duplication and remolding in mitochondrial genomes of amphipods,” Mol. Phylogenet. Evol. vol. 144, pp. 106710, 2020.

DOI 10.18699/BGRS/SB-2020-154

Plastid genome evolution in the genus *Allium*

Victoria Scobeyeva
Lomonosov Moscow State University, Moscow, Russia
Moscow Institute of Physics and Technology, Moscow region, Russia
skobei-khanum@yandex.ru

Denis Omelchenko
Institute for Information Transmission Problems, Moscow, Russia
omdeno@gmail.com

Maria Logacheva
Lomonosov Moscow State University, Moscow, Russia
Skolkovo Institute of Science and Technology, Moscow, Russia
maria.log@gmail.com

Maxim Antipin
Lomonosov Moscow State University, Moscow, Russia
sagefool@yandex.ru

Ilya Artyushin
Lomonosov Moscow State University, Moscow, Russia
sometyx@gmail.com

Andrey Samoilov
Central Research Institute of Epidemiology, Moscow, Russia
andrei.samoilov@gmail.com

Evgenii Konorov
Vavilov Institute of General Genetics RAS, Moscow, Russia
crazydedula@mail.ru

Maxim Belenikin
Moscow Institute of Physics and Technology, Moscow region, Russia
molecular.modeler@gmail.com

Anastasiya Krinitsina
Lomonosov Moscow State University, Moscow, Russia
ankrina@gmail.com

Anna Speranskaya
Lomonosov Moscow State University, Moscow, Russia, Central Research Institute of Epidemiology
Moscow, Russia
hanna.s.939@gmail.com

***Abstract* — Chloroplast genomes are highly conservative and plastomes of Alium specia are not an exception. We assembled and annotated plastomes of 12 specia from genus Allium. The only species with big inversion is A.paradoxa. We identified some regions of Allium plastomes that can possibly be under selection and some genes in these regions. The ycf1 gene is evolving at the highest rate.**

Keywords — Allium, plastome, evolution

Motivation

Plastid genes are often used in plant molecular systematics and species identification due to their conservative sequences and high amount of copies in a single specimen. AtpF-H, matK, psbK-I, rbc ,ropC1,rpoB, trnH-psbA, and trnL-F were used most widely in past decade([1] and two of them – were selected as core plant barcodes. Nevertheless plastid genes code proteins, crucial for photosynthesis and should bear signs not only of neutral, but also of adaptive evolution.

Aim

Specia of Alium genera inhabit wide range of environments with various insolation, humidity, average temperature and other parameters, affecting photosynthesis. We sequenced and assembled plastid genomes of 12 Alium specia and tried to find signatures of natural selection in plastid genes.

Methods

We sequenced and assembled de novo plastomes of the following species: A. fistulosum L., A. macleanii Baker, A. moly L., A. nutans L., A. obliquum L., A. platyspathum Schrenk, A. pskemense B.Fedtsch., A. schoenoprasum L., Allium semenovii Regel, Allium tuberosum Rottl. ex Spreng., A. victorialis L., Allium zebdanense Boiss. & Noë. cpDNA was extracted from living specimen with standard CTAB protocol. DNA samples were sequenced using the Illumina MiSeq high-throughput sequencing platform. The TruSeq protocol (NEBNext® DNA Library Prep Master Mix Set for Illumina, E6040, NEB reagents) was used for preparing the genomic libraries according to manufacturer's recommendation. PE sequences (2x250bp or 2x300bp) using MiSeq. After the quality trimming with Trimmomatic sequencing reads were filtered using chloroplast genome sequences of A. cepa and A. sativum by Bowtie mapper. Dual Organellar GenoMe Annotator (DOGMA) and GeSeq were used for gene annotation. Alignments for evolution rate analysis of each region of chloroplast genome were built using MUSCLE. FUBAR was used for selection seeking. Phylogenetic trees were made with iqtree, using nucleotide alignment. With iqtree Partition finder were determined 3 partitions and evolution rates were calculated independently for each partition.

Results

We have compared the gene order between plastomes of Allium species sequenced de novo and of species sequenced previously The general structure and gene order across all species show high similarity and absence of rearrangements, with A. paradoxum, being the only exception. This species has a major inversion described in our previous work We have not found any large inversions in any other studied species of Allium [2]. So we conclude that plastome of A. paradoxum has a very unusual structure for the genus. Several regions under selection within Allium plastomes were identified. The major region is located from ycf1 (the last gene of the IRa) to ycf2 (the first gene of IRb) and covers the whole SSC region (except only trnL). Partitions evolve with quite different speed from 0.16 (genes clpP rpl2 rps7 rps12 rpl2) to 1.43 (genes psbM psbL ndhD ndhE ndhI ndhA). Alium plastomes are highly conservative. Alignment has 13 sequences with 45030 columns, 614 distinct patterns, 879 parsimony-informative and 1410 singleton sites.

Acknowledgment

Supported by the RFBR (18-04-01203).

References

[1] Dong et al. (2015) Ycf1, the most promising plastid DNA barcode of land plants. Scientific reports. 5 : 8348

[2] Omelchenko et al. (2019) Complete plastome sequencing of Allium paradoxum reveals unusual rearrangements and the loss of the ndh genes as compared to Allium ursinum and other onions. Gene:144–154

DOI 10.18699/BGRS/SB-2020-155

The anatomy of mtDNA of mammalians: the links with organismal longevity

Viktor Shamanskiy
BFU, Kaliningrad, Russia,
v.a.shamanskiy@gmail.com

Kristina Ushakova
BFU, Kaliningrad, Russia,
kristina.ushakova@outlook.com

Konstantin Popadin
EPFL, Lausanne, Switzerland
BFU, Kaliningrad, Russia,
konstantinpopadin@gmail.com

Konstantin Gunbin
ICG SB RAS, Novosibirsk, Russia
BFU, Kaliningrad, Russia,
genkvg@gmail.com

Abstract — **Till now the mtDNA determinants of longevity are poorly known. What is the biological reasons for the effects of mtDNA structures on animal lifespan? In this study in order to answer this question, we correlates various mtDNA properties versus mammals longevity and found interesting answers rooted in mitochondria functional optimization.**

Keywords — mtDNA anatomy, phylogenetic comparative analysis

Motivation and aim

Till now the mtDNA context relation with longevity is poorly understood. Probably, some of these correlations are driven by the decreased somatic mutation rate in long-lived mammals, however, still there is a lack of understanding of the mechanisms of these correlations. There is no common and well established model, explaining causative effects of mtDNA structures on animal lifespan. To answer this question for each species with sequenced complete mitochondrial genome we need to discriminate the all possible anatomical mtDNA features as well as various nucleotide contexts and correlate these versus organisms longevity.

One of the obvious determinants of longevity shortening is the elevated error rate in transcription and translation. The elevated error rate of mitochondrial proteins driven by the oversimplification of the translation machinery – only 22 tRNA genes are responsible for mitochondrial protein synthesis [1] comparing to 41 - 55 nuclear tRNAs with hundreds of nuclear tRNA genes [2]. Thus, mitochondrial ribosome must select the correct aminoacyl-tRNAs from a large pool in very time-efficient manner to sustain an elongation rate and to prevent translation elongation stalling. Is such an optimization observed in nature? It is the second question for answer into this work.

Methods

At the first step, we concatenated mitochondrial protein alignments (generated via PROMALS3D) of 13 mitochondrial proteins, which extracted from whole-genome sequences of mtDNA in more than 3500 Vertebrate species from NCBI GenBank. At the second step, we selected optimal evolutionary model for protein sequences evolution (mtMAM, by iqtree v. 1.5.4). At the third step, we reconstruct final tree and optimized branch lengths using the species tree topology constraints (from TimeTree database; tree reconciliation by treefix v. 1.1.10) and optimal evolutionary model. The branch lengths were calculated based on heterotachy model implemented in iqtree v. 1.6.beta3.

Microsatellites across Vertebrates mtDNA were identified using PERF v0.2.5 [3]. After that, we calculated the fraction of genome covered by microsatellites. Melting, cruciform, or Z-state of mtDNA in Vertebrate species was detected using SIST software [4]. We calculated fractions of mitochondrial genomes characterized by these states. Using pqsfinder R package we extract all mtDNA sequence positions characterized by quadruplex properties. We calculated normalized fraction of quadruplexes in mitochondrial genome by dividing their total length by total mtDNA lengths and by mtDNA GC contents (quadNL). We sleeked for DNA triplexes by two software: triplexator [5] that intended to DNA-RNA triplex identification and triplex R package that intended to DNA-DNA triplex identification. After triplexes identification, we calculated total length of triplexes in mitochondrial genome divided by total mtDNA lengths and by mtDNA GC contents. Using the compseq software from EMBOSS v. 6.6.0 we calculated the observed fractions of codons for all mitochondrial protein coding genes. Using banana (EMBOSS v. 6.6.0) program we calculated average and maximum bend and curve of mtDNA; the btwisted program was used for average stacking energy per dinucleotide (SEpb), average bases per turn, total stacking energy, total twist, and total turns computation for each mtDNA. We used cpgreport (EMBOSS v. 6.6.0) software for calculation the fraction of CpG islands in mitochondrial genome normalized by mtDNA GC content (normCpG) and we used geecee software for mtDNA GC content computation. dan software (EMBOSS v. 6.6.0) were used for average, min, and max mtDNA melting temperatures (Tm) computation; we used "thermo" option for average dG, dH and dS computation. Using Biostrings R package we calculated the fractions of di-, tetra-, hexa- and octa-nucleotides in mitochondrial genomes. We normalized these fractions to the product of nucleotide frequencies in order to take into account possible nucleotide occurrence biases in mtDNA. Additionally, using precalculated data on various degenerate repeats types in Vertebrate mtDNA (ImtRDB [6]) we calculated median, Q1 and Q3 of repeats density per bp.

We intersected our species list in Vertebrate tree with such list in the database on generation length of mammals [7]. In our analyses, we used GenerationLength_d variable from this database. We found 782 species identical in both lists. We performed correlation analysis between the genome characteristics and mammalian generation lengths using phylogenetic independent contrasts implemented in ape R package.

Results

Correlation analysis between the microsatellites fractions in genome, the fractions of melting / cruciform / Z-state DNA and mammalian generation lengths did not result in any statistically significant correlations. Correlation between normalized quadruplexes lengths and generation lengths found positive and statistically significant correlation (Table 1). We did not find any statistically significant relations in correlation analysis between any triplex forms and mammalian generation lengths. The observed statistically significant correlations of codon fractions and normalized fractions of di-, tetra-, and hexa-nucleotides versus mammalian generation lengths are shown in Table 1. We observed that DNA bend and average stacking energy per dinucleotide correlate significantly and positively with mammalian generation lengths, while average mtDNA melting temperature, average dG and dH correlate with generation lengths significantly and negatively (Table 1). High statistical significance and positive sign are also characteristic for correlation of mammalian generation lengths with GC contents, while the normalized fraction of CpG islands correlates at lower statistical significance with the same sign (Table 1). The analysis of degenerate repeats density per bp shown that only Q3 and median of complementary repeats (crQ3 and crMd, respectively) and Q3 of inverted repeats (irQ3) have significant negative correlations with mammalian generation lengths.

TABLE 1. WHOLE GENOME PHYLOGENETIC INDEPENDENT CONTRASTS ANALYSIS. ONLY STATISTICALLY SIGNIFICANT (P<0.01) CORRELATIONS OF MAMMALIAN GENERATION LENGTHS WITH VARIOUS PROPERTIES WERE SHOWN

Properties	t value & positive or negative relation (sign)	p value
quadNL	3.312	0.000969
Bend	4.902	1.15e-06
Tm	4.885	1.25e-06
SEpb	-4.695	3.15e-06
dG	-4.857	1.44e-06
dH	-4.468	9.05e-06
GC content	4.071	5.15e-05
normCpG	3.051	0.00236
irQ3	-3.639	0.000292
crQ3	-3.434	0.000626
crMd	-2.611	0.0092
Codon fractions		
CAC	4.481	8.52E-06
CGG and GGC	>4.115	<4.28E-05
CCT and ACC	>3.978	<7.59E-05
GCC and CCG	>3.655	<0.000274
GTC	3.578	0.000367
CCA	3.486	0.000517
TCG	3.33	0.000908
TCC and CCC	>3.229	<0.00129
GTG	2.879	0.0041
CGC	2.878	0.00411
CGA	2.871	0.00421
ACA	2.678	0.00755
ATA	-3.16	0.00164
CTT and GTT	<-3.338	<0.000885
TAA and ATT	<-3.815	<0.000147
AAT	-4.007	6.74E-05
TAT	-4.126	4.09E-05
TTG	-4.439	1.03E-05
TTT and TTA	<-4.446	<1E-05
Normalized fractions of oligonucleotides		
GC	-2.960	0.00317
TT	-2.818	0.00495
AGCA	-4.287	2.03e-05
GCAG	-4.117	4.24e-05
CGAG	-3.436	0.000622
CGGGTG	-4.894	1.2e-06
CACGAG	-4.533	6.71e-06
TTGCAG	-3.727	0.000207
TGATGC	-3.664	0.000265
CTGTCA	3.459	0.000572
CGGTGA	3.433	0.000629
CACGTC	3.340	0.000876
TTGCTC	-3.331	0.000906

In is of interest that the optimization against translation stalling in long lived mammals is possibly achieved by two ways – via positive selection of codons having hydrogen bonds number ≥6 with anticodon and by negative selection of codons having <6 hydrogen bonds. It is also of greate interest that the optimization of mtDNA nucleotide context (negative selection versus C|G enriched patterns and repeats, positive selection on DNA Bend, Tm) enriched in regions with mtDNA transcriptional and translational functional load.

Acknowledgment

The work supported by the RFBR grant # 18-29-13055.

REFERENCES

[1] Suzuki T, et al. (2011) Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu Rev Genet. 45:299329.

[2] Goodenbour JM, Pan T. (2006) Diversity of tRNA genes in eukaryotes. Nucleic Acids Res., 34:6137-46.

[3] Avvaru AK, Sowpati DT, Mishra RK. (2018) PERF: an exhaustive algorithm for ultra-fast and efficient identification of microsatellites from large DNA sequences. Bioinformatics. 34(6):943–948.

[4] Zhabinskaya D, Madden S, Benham CJ. (2015) SIST: stress-induced structural transitions in superhelical DNA. Bioinformatics. 31(3):421– 422.

[5] Buske FA, et al. (2012) Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. Genome Res. 22(7):1372–1381.

[6] Shamanskiy VA, et al. (2019) ImtRDB: a database and software for mitochondrial imperfect interspersed repeats annotation [published correction appears in BMC Genomics. 20(1):556.

[7] Pacifici M, et al. Database on generation length of mammals. 5427 data records. Nature Conservation 2013; 5:89-94.

DOI 10.18699/BGRS/SB-2020-156

A genetic handicap approach: how to estimate the genome-wide burden of slightly-deleterious variants in a model population

Victor Shamanskiy
Center for Mitochondrial Functional Genomics
Immanuel Kant Baltic Federal University
Kaliningrad, Russia
v.a.shamanskiy@gmail.com

Konstantin Gunbin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Immanuel Kant Baltic Federal University
Kaliningrad, Russia
genkvg@gmail.com

Konstantin Popadin
School of Life Sciences
Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University
Kaliningrad, Russia
konstantinpopadin@gmail.com

***Abstract* — We hypothesized that strongly-deleterious mutation (hereafter 'genetic handicap'), epistatically interacting with the burden of slightly-deleterious variants (SDV), segregating in the genome might be of great importance as a marker of the burden. We expect that handicap-carriers (individuals surviving with the genetic handicap) have decreased burden of SDV as compared to controls. Here to test this hypothesis under different parameter space (mutation rate, severity of SDV and severity of the genetic handicap, effective population size) we use SLim simulation environment. We conclude that the genetic handicap hypothesis indeed works within a broad spectrum of biologically meaningful parameters.**

Keywords* — *SDVs, epistatic interactions, genetic handicap, SLiM

Introduction

All organisms harbor numerous slightly-deleterious variants (SDVs) in their genomes. However, the composition of SDVs, as well as its total effect on fitness, are poorly known due to extreme complexity of the system: thousands of SDVs with very small individual phenotypic effects can interact with each other in various ways and with environment in non-additive manner, making reconstruction of fitness space from genome data almost impossible.

Taking into account widespread negative epistatic interactions between SDVs [1] we introduced a genetic handicap hypothesis: the healthy carrier of a severely-deleterious variant (hereafter "genetic handicap") is expected to have a decreased number of SDVs in genome compared with the organisms without genetic handicap. If so, the genetic handicap can be used in evolutionary and population studies as a signature of genome clearance.

Recently, considering trisomy 21 as a genetic handicap in humans [2] we got the first empirical evidence, supporting our hypothesis: live-born Down Syndrome individuals have decreased the number of SDVs compared to control individuals. Here we computationally investigate the genetic handicap hypothesis in more detail.

Genetic handicap hypothesis

If the biological fitness of an organism is determined as a function of its mutational load, and this mutational load consists mainly of unconditionally deleterious mutations, then we expect to observe a trade-off between the presence of a severely deleterious mutation and the total number of SDVs. Using the term handicap, introduced by Amotz Zahavi as a marker of high genome quality of a carrier [3], and defined in Oxford dictionary as "a circumstance that makes progress or success difficult", we introduce a genetic handicap principle, in which an organism bearing a severely-deleterious mutation (hereafter a 'genetic handicap' in Zahavi's sense) is only fit (viable at birth) if its genome wide load of SDVs is sufficiently low. The rationale for this hypothesis is that only highly fit organisms, with a sufficiently low SDV load, are able to tolerate the effects of severely-deleterious mutations and survive. It is interesting to note, that our approach resembles the 'liability' introduced by Falconer as a single continuous normally distributed factor representing a mixture of environmental and genetic traits and determining a probability to get a complex disease [4].

Here, we define a genetic handicap as a severely-deleterious mutation with very broad effect on cellular metabolism. The broad effect assumes that the genetic handicap is unlikely to be compensated / modified by a few conditional variants and therefore, might only be compensated by a genome-wide decreased load of unconditionally deleterious variants. In this case, we can approximate the total mutational load using the load of unconditionally deleterious mutations.

We assume here that embryonic viability is an important component of fitness, so that the genetic handicap splits the affected population into two groups: survivors at birth and non-survivors, thus providing a simple grouping of affected organisms based on their SDV loads. The main prediction of this hypothesis is that handicap-carriers have a lower number of SDVs as compared to controls (live-born organisms without handicap). Moreover, the stronger the genetic handicap severity, the higher the difference in loads of SDVs between handicap-carriers and controls is expected.

The genetic handicap model

The genetic handicap model is shown on the Fig 1. (A) The distributions of the number of SDVs in control (gray) and affected (red) populations are demonstrated. The genetic handicap mutation (black arrow) is a numerical equivalent of many SDVs. (B) Truncation selection eliminates all organisms with the number of SDVs higher than the given threshold (vertical black line) from both control and affected populations. (C) Handicap carriers have

a decreased number of SDVs (SDVs do not include the genetic handicap per se) compared to controls; this difference represents the handicap effect.

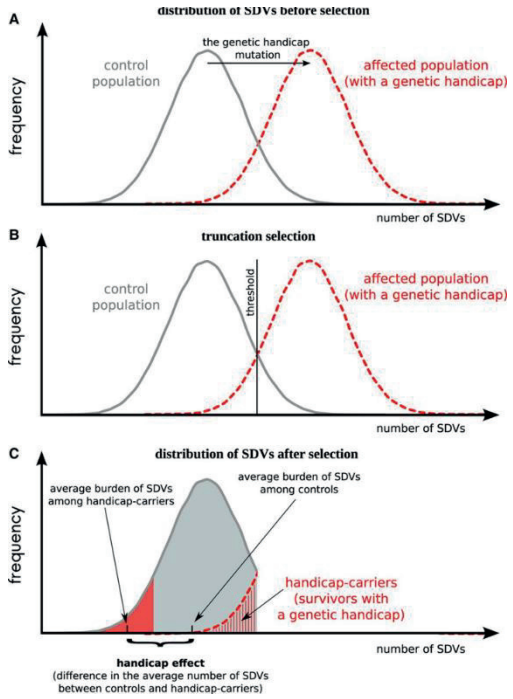


Fig. 1. The genetic handicap model

Forward based simulation using SLIM

SLiM (https://messerlab.org/slim/) is an evolutionary simulation framework that combines a powerful engine for population genetic simulations with the capability of modeling arbitrarily complex evolutionary scenarios. Simulations are configured via the integrated Eidos scripting language that allows interactive control over practically every aspect of the simulated evolutionary scenarios. The underlying individual-based simulation engine is highly optimized to enable modeling of entire chromosomes in large populations.

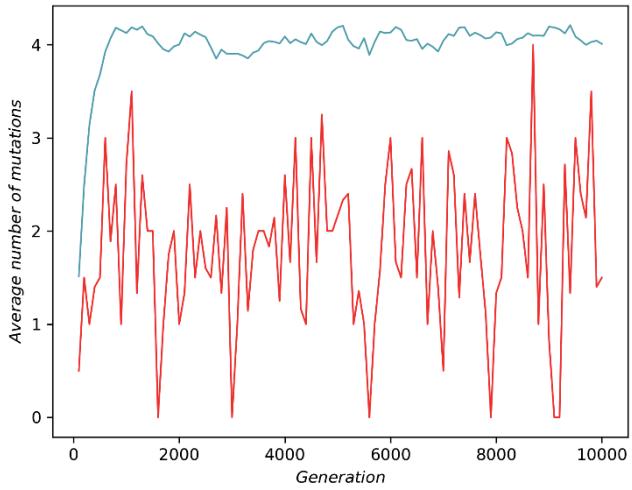


Fig. 2. The Mutation-Selective Balance with handicap-carriers

To test the genetic handicap hypothesis in the SLiM environment we followed two main steps. First of all we established numerous populations under mutation-selection equilibrium (MSE). One example of such a population (effective population size = 100000, genome length = 10000, selection coefficient of each SDV is 0.01, mutation rate of SDV per genome per generation is 0.000001) is provided in Fig 2, where the blue line reflects - how the average number of SDV per genome of average organism (axis Y) is changing with generations (axis X). We can see that the population reaches the MSE during the first 1000 generations. Second, we introduced an additional category of a mutation - the genetic handicap (coefficient of selection = 0.5, mutation rate = 0.000000001) and plotted the burden of SDV among handicap carriers (organisms surviving with the genetic handicap mutation) as red line (Fig 2) We can see that despite strong fluctuations, red line is located effectively below the blue line, confirming that handicap carriers indeed have decreased burden of SDV. (Fig. 2).

Acknowledgment

This work is supported by the 5 Top 100 Russian Academic Excellence Project at the Immanuel Kant Baltic Federal University, and by the Russian Foundation for Basic Research grant 18-29-13055.

References

[1] M. Sohail, “Negative selection in humans and fruit flies involves synergistic epistasis.” Science. (2017): 539-542;

[2] K. Popadin et al, “Slightly deleterious genomic variants and transcriptome perturbations in Down syndrome embryonic selection”, Genome Res. 28 (2018): 1–10;

[3] A. Zahavi, “Mate selection—A selection for a handicap”, Journal of Theoretical Biology 53 (1975): 205-214;

[4] D. S. Falconer, “The inheritance of liability to certain diseases, estimated from the incidence among relatives”, Annals of human genetics 29 (1965): 51-76.

DOI 10.18699/BGRS/SB-2020-157

Multigene phylogenies for the earthworm *Eisenia nordenskioldi* (Lumbricidae, Annelida)

Sergei V. Shekhovtsov
Kurchatov Genomics Center,
ICG SB RAS,
Novosibirsk, Russia
shekhovtsov@bionet.nsc.ru

Alexandra A. Shipova
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
shipova@bionet.nsc.ru

Tatiana V. Poluboyarova
ICG SB RAS, Novosibirsk, Russia
poluboyarova@bionet.nsc.ru

Sergei E. Peltek
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
peltek@bionet.nsc.ru

***Abstract* — *Eisenia nordenskioldi* is a polymorphic earthworm species with two subspecies and high cryptic genetic diversity. We performed transcriptome sequencing for eleven genetic lineages of this species and built phylogenies based on different extracted datasets. We demonstrated that, contrary to previous studies, *E. nordenskioldi* is monophyletic, but both of its subspecies are polyphyletic. *E. nordenskioldi* was split into two groups of lineages, representing the northwestern and southeastern parts of its range.**

Keywords* — *earthworms, Eisenia nordenskioldi, phylogeny, transcriptomes

Motivation and aim

Motivation

Eisenia nordenskioldi was long known to be a polymorphic species with multiple morphological forms and chromosomal races. A series of recent studies uncovered high cryptic DNA diversity within this species [1-2]. However, phylogenetic inferences using conventional approaches were inconclusive.

Aim

Studies that employed a few genes or mitochondrial genomes failed to resolve the relationships among genetic lineages and subspecies constituting *E. nordenskioldi* [2-3]. Therefore, we performed transcriptome sequencing and extracted multigene datasets in order to elucidate phylogenetic relations this species.

Methods

Total RNA was extracted from live specimens; the poly-A RNA was isolated using commercial kits and sequenced by IonTorrent (four specimens) and Illumina HiSeq (seven specimens). The obtained reads were assembled by Trinity. Gene and protein datasets were extracted using the ProteinOrtho and HaMStR programs for ortholog assignment. Phylogenetic trees were built using Bayesian inference and Maximum Likelihood algorithms.

Results

Within the studied dataset *E. nordenskioldi* was monophylic, albeit not with high statistical support. E. n. nordenskioldi was paraphyletic with respect to *E. n. pallida*. Lineages constituting the latter subspecies did not for a monophyletic group. This suggests that the ancestor of E. nordenskioldi was pigmented, and pigmentation was independently lost during the evolution of the species in several distinct lineages. *E. n. pallida* is thus an artificial group. *E. nordenskioldi* was split into two groups of lineages, representing the northwestern and southeastern parts of its range.

Acknowledgment

Supported by the RFBR (19-54-04006_Bel_mol_a, 19-04-00661_a) and by Budget Project of FIZ ICiG SB RAS (0324-2019-0040-C-01).

References

[1] Shekhovtsov S.V. et al. (2013) Cryptic diversity within the Nordenskiold's earthworm, Eisenia nordenskioldi subsp. nordenskioldi (Lumbricidae, Annelida). European Journal of Soil Biology. 58:13-18.

[2] Shekhovtsov S.V. et al. (2016) Cryptic genetic lineages in *Eisenia nordenskioldi pallida* (Oligochaeta, Lumbricidae). European Journal of Soil Biology. 75:151-156.

[3] Shekhovtsov S.V. et al. (2020) Phylogeny of the *Eisenia nordenskioldi* complex based on mitochondrial genomes. European Journal of Soil Biology. 96:103137.

DOI 10.18699/BGRS/SB-2020-158

Polygenic transmission disequilibrium of slightly-deleterious variants in Down syndrome trios

Kseniia Sholokhova
Center for Mitochondrial Functional Genomics
IKBFU, Kaliningrad, Russia

Konstantin Gunbin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Center for Mitochondrial Functional Genomics, IKBFU, Kaliningrad, Russia
genkvg@gmail.com

Konstantin Popadin
School of Life Sciences
Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
Center for Mitochondrial Functional Genomics, IKBFU, Kaliningrad, Russia
konstantinpopadin@gmail.com

Viktor Shamanskiy
Center for Mitochondrial Functional Genomics
IKBFU, Kaliningrad, Russia

Abstract — **A trisomy 21 (T21) is associated with high, up to 80% miscarriage rates. Yet the genetic component of this selection is unknown - why some fetuses with T21 can be live-born, while others are miscarried. Here we hypothesize that genome-wide burden of slightly-deleterious variants in the T21 genome can affect the outcome. To test this hypothesis we analyze genomic data from families with live-born offspring with T21. In order to uncover patterns of embryonic selection against fetuses with T21 we need to analyze which set of alleles is over- or under- inherited from parents to offspring with T21. For this, we calculate individual predisposition of each member of a trio to different phenotypes (mainly different diseases). These genetic risks, or polygenic risk scores (PRS) are calculated by summing up risk alleles, that are weighted by effect sizes derived from GWAS analyzes. Comparing the PRS of parents with the PRS of offspring we can uncover disequilibrium in the transmission of certain risks. Understanding of molecular mechanisms of over- and under-transmission of alleles is of great importance because they can determine the viability of T21 fetuses. Discovered in this way non-randomly transmitted classes of polymorphisms can help us to point out molecular signatures of phenotypes associated with T21 as well as may potentially help to find new ways of treatment.**

Keywords — PRS, T21, Down syndrome, polymorphisms, slightly-deleterious variants, transmission disequilibrium

Introduction

To date, trisomy 21 is the most common chromosomal abnormality occurring in humans. It is caused by a meiotic nondisjunction event with subsequent presence of all or part of the third copy of chromosome 21. Moreover, trisomy 21 is associated with high, up to 80% miscarriage rates [1].

Down syndrome phenotypes are induced by trisomy 21, but they can be modified by numerous slightly deleterious variants located on both chromosome 21 [2] and all other chromosomes [3]. Here using PRS approach we try to find a set of alleles which can modify the effect of T21. Over-transmitted alleles we interpret as protective ones, which can partially compensate the effect of T21, while under-transmitted alleles we interpret as deleterious ones, which aggravate the effect of T21.

Methods

Our main dataset is 628 genotyped trios (two parents and live-born DS offspring in each trio). Using all DNA polymorphisms, associated with 50 different traits (large-scale studies, N > 50,000) such as body mass index (BMI), coronary artery disease, blood pressure, type 2 diabetes, etc.), we calculated a polygenic risk score (PRS) for each member of each trio. Below we describe step by step how we calculated PRS and estimated transmission of these risks from parents to T21 offspring.

(i) PRS

For finding complex diseases which affect DS we first counted polygenic risks scores. A polygenic risk score (PRS) is a sum of trait-associated alleles weighted by effect sizes estimated from a genome-wide association study.

We counted PRS in 4 different ways:

(1) without in-house strict clumping & thresholding the number of SNPs, on a full sample, for all trios - in this case, the soft clumping & thresholding built-in PRSice2.2.8 was used;

(2) by in-house strict clumping & thresholding the number of SNP, calculation of PRS on a complete sample, for all trios by simple summation of the products beta * N, where N is the number of alleles of the alternative genotype;

(3) by in-house strict clumping & thresholding, calculating PRS for each triple using PRSice2.2.8 (clumping & thresholding built into PRSice2.2.8 was disabled by the –no-clump option);

(4) by in-house strict clumping & thresholding, calculating PRS on a full sample, for all triples using PRSice2.2.8 (the built-in clipping & thresholding in PRSice2.2.8 was disabled by the –no-clump option).

The highest qualitative correspondence, according to the variation of CliffDelta and CohenD, with the hypothesis that the median PRS of parents more than in children with Down syndrome was obtained only by using the simplest method (2).

It has to be said that we measured the effect size of the differences between the PRS distributions of parents and children with Down syndrome using Cliff's Delta and Cohen D - all 4 methods yielded negligible differences between the compared PRS distributions.

In addition, it must be said that we tried to calculate the statistical significance of differences in the distributions of PRS of parents and PRS of children with Down syndrome using Welch t-test and U-test. Statistically significant (however marginally significant) differences in the distributions were obtained exclusively by analysis by method (1), however, the variations of Cliff's Delta and Cohen D, not to mention PRS of children with Down syndrome or PRS of parents, in this case were inconsistent with each other.

(ii) Transmission disequilibrium

The transmission of burden of slightly deleterious variants can be estimated by polygenic transmission disequilibrium test (pTDT) [4]. This approach allows to point out how various genetic variants can be over- and under-transmitted from parents to offspring.

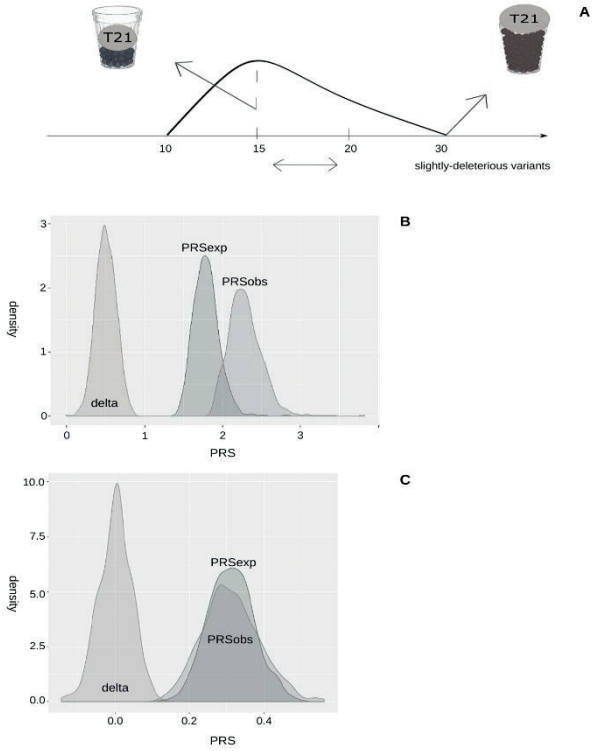


Fig. 1. The concept of inheritance of slightly deleterious variants.

A-a number of slightly-deleterious variants determines whether the fetus with T21 will be live-born (left cup with some space) or a miscarriage (overflowing right cup);

B- fictional model with transmission disequilibrium, where delta is different from 0, PRSexp - expected RPS (based on parents' genotypes), PRSobs - observed PRS (based on an offspring's genotype);

C- results as a PRS for schizophrenia, schizotypal and delusional disorders, where delta equals 0, PRSexp - expected RPS (based on parents' genotypes), PRSobs - observed PRS (based on an offspring's genotype).

When PRSs for both parents and offspring are counted we compare average (mid-parent score) and/or whole distributions with the PRS of offspring. The contrast between the scores will shed the light whether genetic predisposition to a given phenotype was randomly transmitted to offspring (i. e. contrast is close to zero) or over- / under- transmitted.

Results

Currently, analyses of transmission of all calculated risks demonstrated the expected neutral inheritance from parents to the offspring (average delta = 0), as in fig.1C. It is in line with the absence of strong genetic selection for or against the T21 during embryogenesis. This can be explained either by real absence of genetic selection (meaning that live-born or miscarried T21 can be purely random or depends on environment) or by the fact that we didn't find yet T21 - relevant phenotype. Thus we continue to look for polygenic transmission disequilibrium in other phenotypes.

In addition to this, we plan to annotate specific targets of DS embryonic selection in order to point out the strongest genetic variants, genes and gene networks which affect survival probability of individuals with Down Syndrome.

Acknowledgment

This work has been supported by the 5 Top 100 Russian Academic Excellence Project at the Immanuel Kant Baltic Federal University, by the Russian Foundation for Basic Research grant 18-29-13055.

References

[1] Nussbaum RL, McInnes RR, Willard HF. 2004. Thompson & Thompson genetics in medicine, revised reprint, 6th ed. Saunders, Philadelphia, PA

[2] Sailani MR, Makrythanasis P, Valsesia A, Santoni FA, Deutsch S, Popadin K, Borel C, Migliavacca E, Sharp AJ, Duriaux Sail G, et al. 2013. The complex SNP and CNV genetic architecture of the increased risk of congenital heart defects in Down syndrome. Genome Res 23: 1410–1421.

[3] Letourneau A, Santoni FA, Bonilla X, Sailani MR, Gonzalez D, Kind J,Chevalier C, Thurman R, Sandstrom RS, Hibaoui Y, et al. 2014. Domains of genome-wide gene expression dysregulation in Down's syndrome. Nature 508: 345–350.

[4] Weiner et al 2017 Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nature Genetics volume 49, pages 978–998

DOI 10.18699/BGRS/SB-2020-159

Tandem repeats are selfish elements which mark the level of hidden recombination in animal mitochondrial genomes

Aleksandr Smirnov
Center for Mitochondrial Functional Genomics
Immanuel Kant Baltic Federal University
Kaliningrad, Russia
aleksandr.smirnov2009@gmail.com

Alina A. Mikhailova
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University
Kaliningrad, Russia
mihailovaalina777@yandex.ru

Valeria Lobanova
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic *Federal University*
Kaliningrad, Russia
aaren183@gmail.com

Konstantin Gunbin
Institute of Cytology and Genetics SB
RAS, Novosibirsk, Russia
Immanuel Kant Baltic Federal University
Kaliningrad, Russia
genkvg@gmail.com

Konstantin Popadin
School of Life Sciences
Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic *Federal University* Kaliningrad, Russia
konstantinpopadin@gmail.com

Abstract — **Despite the fact that mitochondrial genome is streamlined, its size still varies significantly across all vertebrate taxa (from 15kb in Mammalia to 25kb in Reptilia approximately).**

To uncover the driving reason of this variation we analyzed structures of the whole mitochondrial genomes of 3793 Chordata species belonging to 5 taxa: Actinopterygii (1905 sp.), Amphibia (227 sp.), Reptilia (285 sp.), Mammalia (855 sp.), and Aves (521 sp.). In each of these taxa we observed strong positive correlation between the genome size and the length of non-coding region (D-loop) (all Spearman's rho >= 0.76, all p-values <= 2.2e-16). We assumed that some repeated sequences may contribute in D-loop length variation and looked for five types or repeats: tandem (annotated by Tandem Repeats Finder), direct, symmetrical, inverted and complementary (annotated by our own scripts). Analyzing the impact of the number and the length of each repeat type on D-loop size variation we revealed that the tandem repeats is the main repeat types, shaping the size of the D-loop.

Keywords — Tandem repeats, recombination

Introduction

To reveal potential associations between the abundance of the tandem repeats and species-specific life-history traits we focused on Mammals, for majority of which there is a high quality ecological information such as generation length, body mass and metabolic rate. For 434 species we observed a negative correlation between tandem repeats abundance and generation length (rho = -0.10, p-value = 0.029) indicating that tandem repeats are more common in short-lived species (this correlation is robust if we introduce in our model GC content and PIC).

It is important to take into account not only presence/absence of tandem repeats and their total length but also properties of their motifs: length of the motif, nucleotide content of the motif, copy number of the motif (how many times it is repeated in mtDNA), the level of degradation (average similarity between each repeated copy of the motif), and so on. Using pairwise correlations and principal component analysis we observed several groups of tandem repeats, for example more GC rich motifs tend to be more perfect (less degraded) with more repeated copies. We hypothesized that some of these groups of repeats may be more important than others in the shaping of the correlation with ecological (generation length) and genetical (whole genome size, D-loop size) traits. Indeed we observed that long, GC-rich, low-degraded and many-times-repeated motifs correlate better with generation time.

There are three non-mutually exclusive potential explanations of the observed negative correlation between the tandem repeats abundance and generation time: (i) strong negative selection against tandem repeats in mtDNA of longlived species (within long-lived dormant oocytes)[2], (ii) selfish drive of tandem repeats and (iii) neutral propagation of tandem repeats in mtDNA of species with non-zero level of recombination.

Methods

(i) It is known that in human postmitotic tissues (neurons and skeletal muscles) somatic mtDNA deletions propagate with age leading to many age-related phenotypes. Somatic deletions are very deleterious and are almost never transmitted from mother to offspring but probably the same process is happening on the low scale during evolution. This process is expected to be more powerful in the oocyte, which lives long life - for example in elephants it is expected to be stronger than in mice (20 years within elephant oocyte versus 1 year in mouse oocyte). However, until now there is a lack of evidence, that short size of the genome is associated with clonal advantage.

(ii) Tandem repeats may provide an advantage to the host mtDNA (selfish drive)activating replication of a given molecule. Recently it has been shown that some tandem repeats may increase replication rate of the host mtDNA genome providing it a replication advantage [3]. Interestingly this process may lead to the accumulation of deleterious variants throughout the whole genome (and even to extinction of a line) since replication advantage overcomes some deleterious effects of the linked variants [3]. In this scenario, tandem repeats are active drivers of the increased replication rate of the host molecules.

(iii) Tandem repeats (especially with long motifs) may propagate better in the recombinational environment and thus they might be a neutral marker of the hidden recombination. It is commonly accepted that mtDNA of chordata are mainly non-recombining, maternally inherited genome. However, numerous well-proofed exceptions are known from this rule. We hypothesize, that tandem repeats can be a marker of the low level of recombination, undetectable by many other methods.

Interestingly the number and the length of tandem repeats differ not only between species but also between taxa: on the average tandem repeats are more common in Amphibia and Reptilia while genomes of Mammalia and Aves are more poor in tandem repeats. It is known that exactly Amphibia have the highest frequency of biparental inheritance of mtDNA.

Altogether we conclude that tandem repeats are markers of hidden and rare recombination in mtDNA of vertebrate species Additionally we decided to estimate the probability of recombination in mammalian species using polymorphism data. For 224 mammalian species we were able to find 10 different CytB sequences, aligned them and analyzed a probability of recombination using LDr2 approach (http://www.lifesci.sussex.ac.uk/CSE/test/ld.php). After analyzing polymorphic data of 224 species for 33 of them we found some evidence of recombination (species with nominal p-values less or equal 0.01). Next we tested if recombination is associated with tandem repeats. We observed that among species with more than one tandem repeat the fraction of species with recombination is 21% (17 out of 80) while among the species with one or less tandem repeats this fraction is only 11% (16 out of 128 species). The increased fraction of recombination among the tandem-repeat-rich species is marginally significant ($p = 0.049$, Fisher's odds ratio = 2.15). Interestingly, between class comparison showed that aves have the most frequent recombination (18% of species), actinopterygii - the second most frequent (11%) while reptilia (8%) and mammals (7%) are the classes with the most rare recombination detected (we don't have enough amphibia species to validate it). In the future we plan to use several additional tests on recombination in order to validate the robustness of our findings.

Results

Altogether, using several lines of evidence (association of tandem repeats with mtDNA genome length, with speciesspecific generation length, with Kn/Ks and with recombination events) we conclude that mtDNA tandem repeats indeed mark very rare and very hidden recombination.

Acknowledgment

This work is supported by the 5 Top 100 Russian Academic Excellence Project at the Immanuel Kant Baltic Federal University, and by the Russian Foundation for Basic Research grant 18-29-13055.

References

[1] M. Pacifici, et al. (2013) Generation length for mammals, Nature Conservation, 5:89.;

[2] Y. Kraytsberg, et al. (2006) Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons, Nature genetics, 38.5:518;

[3] H. Ma, and P. H. O'Farrell (2016) Selfish drive can trump function when animal mitochondrial genomes compete, Nature genetics, 48.7:798;

[4] L. Polishchuk, et al. (2015) A genetic component of extinction risk in mammals, Oikos, 124.8:983-993;

[5] D. J. White, and N. J. Gemmell (2009) Can Indirect Tests Detect a Known Recombination Event in Human mtDNA?, Molecular Biology and Evolution, 26.7:1435

DOI 10.18699/BGRS/SB-2020-160

Genetic aspects of internet-dependence in teenagers

Marina Smolnikova
Laboratory of cell-molecular physiology and pathology
Scientific Research Institute of Medical Problems of the North
FRC KSC SB RAS: Krasnoyarsk, Russia
smarinv@yandex.ru

Sergey Tereshchenko
Clinical department somatic and mental health of children
Scientific Research Institute of Medical Problems of the North
FRC KSC SB RAS, Krasnoyarsk, Russia
legise@mail.ru

Abstract — **The rapid emergence and spread of Internet addiction in adolescent populations, combined with a rapid change in consumed content due to mobile access availability and the new access devices, are new challenges for classical psychology and fundamental medicine that require urgent solutions. The presence of the genetic component of Internet addiction was convincingly shown in different populations, but to date the specific genes involved in the mechanisms of such heritability have not been well identified. The aim of the study was to investigate the role of candidate neurotransmitters and to perform a population analysis of candidate genes polymorphisms for the Internet addiction formation. As a result of the analysis, 7 Internet-dependent adolescents and 39 adolescents with a tendency to Internet-dependent behavior were identified. The frequencies of genotype distribution in the population sampling of Caucasians (n = 302) were obtained: * CC 59.6%, * CT 40.4%, * TT 0% (rs1800497 DRD2); * AA 23.8%, AG 52.4%, * GG 23.8% (rs4680 COMT); * CC 37.4%, * CG 46.0%, * GG 16.6% (rs2229910 NTRK3). The results will help to open new perspectives in assessing the fundamental neurobiological causes for the Internet addiction and the personalization of therapeutic approaches in Internet dependent adolescents.**

Keywords *— adolescents, Internet addiction, gene polymorphism*

Introduction

Internet addiction (IA) is a relatively new psychological phenomenon having signs of a social epidemic, specifically attributed to vulnerable groups (adolescents and young adults) in some regions (South Korea, China, Japan). The prevalence of IA among adolescents varies depending on the ethnosocial groups studied and the diagnostic criteria used, from 1% to 18% [1]. In Europe, therefore, the prevalence of IA among teenagers is 1-11%, on average 4.4%, in the USA, the prevalence of IA in the total sampling is 0.3-8.1%. However, the prevalence of IA among teenagers and young adults is significantly higher - 9-37.9% in Asian countries [2].

The genetic component of developing Internet addiction was shown by twin studies using various populations, however, to date, the specific genes involved in the mechanisms of such heritability are not exactly identified. Pilot studies verified some polymorphic sites of five candidate genes: rs1800497 (dopamine D2 receptor (DRD2) gene, Taq1A1 allele) and rs4680 (methionine variant of dopamine degradation enzyme catecholamine-o-methyltransferase (COMT) gene), and the association of minor alleles associated with low dopamine production (rs4680) and a low number of dopamine receptors in the prefrontal cortex (rs1800497) with pathological addiction to Internet games in adolescents in South Korea was shown to occur [3]. An exome analysis including a study of 83 polymorphic sites revealed a statistically significant association only with rs2229910 (neurotrophic tyrosine kinase receptor type 3 (NTRK3) gene), perhaps also associated with anxiety, panic, depression and other disorders [4].

Thus, the emergence and spread of Internet addiction in adolescent, combined with the rapid change in consumable content due to the general availability of mobile access to the network and appearing some new access devices, pose for fundamental medicine some new challenges to be studied. This is particularly true because of the established serious comorbidity of IA (depression, anxiety, suicidal behavior, unmotivated aggression, as well as psychosomatic diseases). An analysis of ethnogeographic differences of Internet addiction, while taking into account the ethnic differences in the prevalence of genotypic characteristics of populations seems to be relevant and that is not sufficiently explored in modern neurogenetics of addiction in adolescents, especially in such a multinational country as Russia. The research results will help to open a prospect in assessing the fundamental neurobiological aspects of developing Internet addiction and individualised therapeutic approach in Internet-dependent teenagers.

Matirials and methods

The overall prevalence assessment of Internet addiction in adolescents of 13-18 years old living in the city of Krasnoyarsk was carried out using the Chen Internet addiction Scale Test (CIAS, S.-H. Chen, 2003) in the adaptation of V. L. Malygin, K.A. Feklisov [5,6] in randomly selected secondary schools. 233 teenagers of Caucasian origin at the age of 13-18 years old were surveyed (average age is 15.3 ± 1.6), with 111 girls and 122 boys being involved in the study. All examined ones or their parents gave written informed consent to participate. DNA was isolated using salting out method. Genotyping of rs1800497, rs4680 and rs2229910 was taken on 302 DNA samples of a population sampling of Caucasians in the Krasnoyarsk Territory using RT-PCR.

Results and discussion

An analysis of the content - structure of Internet addiction among adolescents in Krasnoyarsk resulted in the identification of three groups of adolescents: (1) adolescents with no Internet addiction behavior, a control group (the expected frequency was 60-75%), the detected frequency being 80.3%; (2) adolescents with a tendency for developing Internet addiction/pre-addictive stage (the expected frequency was 20-30%), the detected frequency being 16.7%; (3) adolescents with Internet addiction / behavior with an internet abuse component (the expected frequency was 5-10%), the detected frequency being 3.0%. As a result of the analysis, 7 Internet-dependent adolescents and 39 adolescents with a tendency to Internet-dependent behavior were identified, they were also paired (corresponding to gender, age and nationality) for sampling the biological material, further DNA extraction and genetic screening of polymorphic genes on IA susceptibility. According to the literature, a methodologically verified study of students in 9-11 grades in Moscow using the CIAS, i.e. questionnaire validated by the authors, showed that "11.0% have some signs of Internet addiction, 42.0% abuse the Internet of all the adolescents examined (n = 190)".

Another study by the same research group demonstrated that "4.25% were diagnosed as Internet addicts and 29.33% as Internet abusers of 1084 adolescents (the average age in the sampling was 15.6 years old)" [5].

We carried out a genotype test of the population sampling of Caucasians in Krasnoyarsk (n = 302) rs1800497 DRD2, rs4680 COMT and rs2229910 NTRK3. The frequencies of genotype distribution were obtained: * CC 59.6%, * CT 40.4%, * TT 0%; * AA 23.8%, AG 52.4%, * GG 23.8%; * CC 37.4%, * CG 46.0%, * GG 16.6%, respectively. The frequency of allelic variants of the studied SNPs corresponds to world data on Caucasoid populations (according to www.ensembl.org).

Conclution

As a result of increase in the group of teenagers, a comparative analysis of the distribution of polymorphic sites rs1800497, rs4680, rs2229910, rs25532 (serotonin transporter (SS-5HTTLPR) gene) and rs1044396 (nicotinic acetylcholine receptor subunit alpha 4 (CHRNA4) gene) is expected to carry out in adolescents of Central Siberia in order to identify candidate genes of developing Internet addiction.

Acknowledgment

The study was carried out with the financial support of the Russian Federal Property Fund in the framework of the scientific project No. 18-29-22032 \ 19.

References

[1] L. Cerniglia, F. Zoratto, S. Cimino, G. Laviola, M. Ammaniti, W. Adriani, "Internet Addiction in adolescence: Neurobiological, psychosocial and clinical issues," Neurosci Biobehav Rev, 2017, vol. 76, pp. 174-184.

[2] S. Park, J. Kim, C. Cho, "Prevalence of Internet addiction and correlations with family factors among South Korean adolescents", Adolescence, 2008, vol. 43(172), pp. 895-909

[3] D. Han, Y. Lee, K. Yang, E. Kim, I. Lyoo, P. Renshaw, "Dopamine genes and reward dependence in adolescents with excessive internet video game play", J Addict Med, 2007, vol. 1(3), pp. 133-138.

[4] J. Jeong, J. Rhee, T. Kim, S. Kwak, S. Bang, H. Cho, Y. Cheon, J. Min, G. Yoo, K. Kim, J. Choi, S. Choi, D. Kim, "The association between the nicotinic acetylcholine receptor alpha4 subunit gene (CHRNA4) rs1044396 and Internet gaming disorder in Korean male adults", PLoS ONE, 2017, vol. 12(12), pp. e0188358.

[5] V. Malygin, Yu. Merkurieva, A. Iskandirova, E. Pakhtusova, A. Prokofieva, "Specific features of value orientations in adolescents with Internet addictive behaviour", Medical Psychology in Russia, 2015, vol. 33 (4), pp. 1-20.

[6] S.-H. Chen, L.-J. Weng, Y.-J. Su, H.-M. Wu, P.-F. Yang, "Development of a Chinese Internet Addiction Scale and Its Psychometric Study", Chinese Journal of Psychology, 2003, vol. 45(3), pp. 279-294.

DOI 10.18699/BGRS/SB-2020-161

Homologous series. Law or rule?

Valentine Suslov
Systems Biology Department
ICG SB RAS
Novosibirsk, Russia
valya@bionet.nsc.ru

Mikhail Ponomarenko
Systems Biology Department
ICG SB RAS
Novosibirsk, Russia
pon@bionen.nsc.ru

Dmitry Rasskazov
Systems Biology Department
ICG SB RAS
Novosibirsk, Russia
rassk@bionen.nsc.ru

Abstract — **Despite the centenary, the status of Vavilov law of homologous series (LHS) remains uncertain. Unlike an empirical rule, a law may have no exceptions. Exceptions are indicative of having to abrogate or to ascertain the limits of the law's applicability, the latter being derivable from the ground on which the law stands. As the basis of the law neither the common ancestor heritage, nor the evolutionary enumeration of a limited space of possibilities satisfy this demand. We consider LHS as a consequence of the functional overlap of adaptive characters.**

Keywords *— Vavilov law of homologous series, promoter, TATA box, composition element, coadaptive substitutions, self-adaptive evolution.*

Background

Similarity of homologous series (HS) is not weak, incomplete, or unexplored (in all cases, how much?) resemblance but an independent "third" similarity of characters caused by the principles of their organization[1]. HS does not require, does not exclude, but does not boil down to the identity (homology) and/or common function (analogy) of the characters [1]. The law has no exceptions, but only abrogate (**a**) and ascertain (**b**) limits of applicability. If **a**, then the gaps in the HS are imaginary i.e. an observer error, a member of one HS fell into another, not derived from the first (α). If **b**, then both the exception and the HS are equally inferred from one cause (β). Otherwise, there is empirical rule, no law (γ). Logically, α and β are complementary, both oppose γ: $(\alpha \vee \beta) \wedge \neg\gamma$. For the first time, Cope spoke of HS as $(\alpha \vee \beta) \wedge \neg\gamma$, where **a** it's *characters of genera* without adaptive selection associated with morphogenetic dynamics which based on physicochemical laws; **b** it's *characters of species* that adaptive evolved only when **a** and the selection vector coincide [2]. Thus, any character of genera not only limited the *space of possibilities* (LSP) for characters of species, but also provided correlations between these characters[2]. However, the factual fuzziness of this characters' dichotomy has reduced the Cope laws to γ. In editions of *Law of HS* (LHS) of 1920 and especially 1922, Vavilov [1], having put forward many reasons, also reduced LHS to γ. In 1935 Vavilov [1] raises the question of $(\alpha \vee \beta) \wedge \neg\gamma$, for which he excludes from LHS the similarities whose reasons arise long before (i.e. popular today developmental hourglass and less popular common ancestor heritage) [3] or independently of this taxon (i.e. limitations by LSP or extreme physicochemical laws like Maupertuis principle). He continues to build only combinative (weak epistasis)[4] HS, linking their genesis with *centers of origin* (CO)[5] [1]. But the schools of Filipchenko and Schmalhausen, paradigmatically claiming for LHS in fact returned to Cope (only characters hierarchy instead its dichotomy)[6]. Is LHS the law? Are all HSs deducible from LHS?

Materials and Methods

We studied the phenotypic effect of mutations in the core and flanks of TATA boxes and composition elements (CEs) composed of TATA boxes with their flanking sequences overlapping other regulatory sites. *The character under study was gene expression level*: the service SNP_TATA_Comparator [2] discriminates between the TBP/TATA affinity contribution into the gene expression level and the regulatory site contribution for CE as well as between the contributions of the TATA box core and flanks.

1Hence obligate for Vavilov HSs character transgression [1]: in spite of a rapid divergence of phyla, transgressing characters retain the shared phenotypic aspects and the way the phenotype varies, while the convergence preserves only the phenotype.

2The scheme that somehow underlies all modern evo-devo theories [3]. Since any hierarchy of characters automatically generates a modular organization of ontogenesis and parallelisms as a consequence, the problem of this scheme will not be HSs, but gaps which the scheme cannot deduce from an own foundation. For example [3] several gaps may be reduced by includ embrional forms in HS. But how deep can go down into embryos for it (not to mention that Vavilov used only postnatal characters in LHS [1])? The status of the law has only Stark-Zavarzin theorem [2], where the hierarchy is absolute, since it is given by the logical incompatibility of characters or their biochemistry that identic in adults and young.

3The same problem. Going down into embryogenesis and/or changing the complexity of characters by reducing them to sequences of biopolymers, an observer can arbitrarily get both homology (the postgenome homology crisis [5] foreseen by Meyen) and developmental hourglasses [4].

4 Paradoxically ignoring correlations of characters that he himself appreciated. Including genetic maps, but not ploidy. For example, he built many HSs for separate flower's characters, but not one for flower's formula as a whole [1]. This formula, as known now, relies on a conservative gene network-automat.

5Vavilov spoke about centers of origin [1], but neo- and paleo data do not confirm either the obligatory origin of a taxon in the center, or the supereffectiveness of the primary center for taxa's adaptive radiation [6,7]. Paradoxically, according to Vavilov, for LHS similarity forming it is necessary that the taxon invade into the region of maximum environmental diversity. That is disaccording to traditional theories [3] where similarity in environmental diversity is the least likely.

6According to Filipchenko [2], the hierarchy is determined by the time of the character appeared (from a common ancestor [3,5] or nearest hourglass [2,4] in the last version), to which the Modern Synthesis has added logical compatibility as Dobzhansky-Muller model [3]. According to Schmalhausen synthesis, this compatibility is regulated by feedbacks [2] and/or trade-offs [3].

Results and Discussions

Shown, that *HSs appear due to coadaptive substitutions regardless of common ancestor*: nonhomologous mutation in the TATA box mutually compensate for each other when there is a *functional overlap* (FO) between the flanks and the core. The same is as the TATA box has a FO with CE sites. *Being dependent only on the level of diversity (in our case, on mutability) and the level* of *FO, this evolution will go both in a changeable and even in an unchangeable environments. But in both cases it should be a subpessimal environment. Finally, its own self-adaptive* [7] *trend is forming.* The trend starts with the substitution of homology for being homologous and proceeds with a coordinated divergence of HSs[8]. During such a coordinated divergence, the bundle of closely related phyla carrying the character diverge for this character into bundles, in each of which phyla are differently related. *Any bundle is a case of character transgression; any one-phylum bundle appears as a case of classic Darwinian bifurcation, an "exception" (β_1) to LHS, but, on the other hand, many bundles this is any gap in each (β_2, the main "exceptions" to LHS).* A trend continues if self-adaptation in any bundle leads to the dominance of only one functional contribution to the gene expression level. For CE either the contribution of TBP/TATA affinity dominates, or the contribution of the regulatory site[9]. *Violation of the contribution symmetry forms the Harland scheme*: major gene+gene (s)-modifier(s) [8]. It is limit **b** to LHS. Since there are more than two modifiers and they are structurally separable, the aforementioned bundles coordinated divergence will increase, moreover, with all "exceptions", to which gaps in the number of modifier genes[10], their contribution/effect and other its parameters are added (β3+k) 11 . Bundles coordinated divergence will slow down with decrease in FO (weak epistasis, as Vavilov put it!) and will stop once FOs have been banned; they may be banned for 4 reasons (3 identified for CEs): elimination of one of the characters, their emergent union 12 (high epistasis) or independent operation13. There are limits a to LHS.

Conclusions

LHS is a law. Not all HSs deducible from LHS[14] but only HSs consisting of characters with FO.

Acknowledgments

The work supported by the budget project No. 0324-2019-0040-C01.

References

[1] N. Vavilov. "The law of homologous series in variation" Nauka. 1987. 260 p.

[2] V. Suslov, M. Ponomarenko, D. Rasskazov. "Homology and Vavilov's law of homologous series". http://mccmb.belozersky.msu.ru/2019/thesis/MCCMB2019/pages/author.utf8.html#S.

[3] J. Herron, S. Freeman. "Evolutionary Analysis". Pearson. 2014. 864 p.

[4] I. Grosse. "Entropic Hourglass Patterns and the Emergence of Biodiversity". http://mccmb.belozersky.msu.ru/2019/thesis/MCCMB2019/pages/author.utf8.html#G.

[5] J. Moore, P. Willmer. "Convergent evolution in invertebrates". Biol. Rev. Camb. Philos. Soc. vol. 72. pp. 1-60. 1997.

[6] N. Goncharov. "The centers of cultivated plants origin". Vestnik VOGiS. vol. 11. pp. 561-574. 2007.

[7] D. Begun, M. Nargolwalla, L. Kordos. "European Miocene hominids and the origin of the African ape and human clade". Evol Anthropol. vol. 21. pp. 10-23. 2012.

[8] S. Harland. "The genetical conceptions of the species". Biol. Rev. Camb. Philos. Soc. vol. 11. pp. 81-112. 1936.

[7]What makes self-adaptation be different from co-adaptation is a symmetry: a predator and a prey do not swap their roles, but a contributions to expression (providing that expression is underway), are interchangeable: *in subpessimal environment not only decrease, but also increase in contribution coming from one site favors the fixation of the mutation that compensate for this effect.* Evolution resembles Le Châtelier's principle [2]. In *severe unchangeable environment* a crucially necessary excess of the gene product blocks this symmetry; any *gentle environment* is a neutral [3] non-Vavilovian [1] evolution.

[8] In 1923 Sobolev was the first to observe it – with no comments left, though. *The only case of divergence – seed-shattering tips in rye, wheat and Aegilops plants – observed by Vavilov was mentioned in a comment, with no HSs created* [1]*!*

[9]For TATA box, the contributors are either the core or the flanks, respectively. Since this pair of contributors is both functionally and structurally inseparable, the only possible trend here is the swing between the TATA+ and TATA-less promoters. The destruction of a pair is the death of the proximal promoter.

[10]Selection in favor of major gene means automatic selection in favor of modifier genes. However, among modifier genes, while their effect is evaluated in total, during this selection we, as a side result, will automatically receive advanced and lagging genes (the Abel's crossing of specializations). The contribution of the former is much larger, the contribution of the latter is significantly less than the average contribution for all modifier genes. Finally, lagging modifier genes must become neutral, pseudogenized, or even be eliminated if they cannot change function by Ohno model of Modern synthesis [3] ($\beta_{(4+k)}$ "exception").

[11]Included by Vavilov to LHS in 1935 [1]. In many times used in synchronous criticism of LHS. Finally, it is equally excluded from LHS in Schmalhausen's and Modern syntheses as incompatible with the Cope foundation.

[12] As in advanced Ohno models duplication-degradation-subfunctionalization in Modern synthesis [3]: each character is irreplaceable, but everyone works for common function. Vavilov LHS degenerate into Cope's rule.

[13]The fourth reason is Ukhtomsky dominante-like. Suppose the selection in favor of major gene supports only those modifier genes that enhance the phenotypic effect of major gene expression. Then the selection against antagonistic modifier genes is combined with two main contributions: for major gene and for synergistic modifier genes. Self-adaptation can contrasting or leveling these contributions. In the first case, we obtain an emergent character, in the last case, a death of Harland scheme after complete indistinguishability of these contributions.

[14]Besides the Cope's HSs it will be observed series of genetic maps (but not ploidy), where the violation of similarity is the destruction of the common ancestor heritage by transposons [3], whose activity is weakly associated with organismal characters with FO.

DOI 10.18699/BGRS/SB-2020-162

Resequencing genomes of the Russian native Baikal and Tuva sheep breeds

James Sweet-Jones
Royal Veterinary College, University of London, UK

Nikolay Yudin
Kurchatov Genomics Center, Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Denis M. Larkin
Royal Veterinary College, University of London, UK
Institute of Cytology and Genetics, Novosibirsk, Russia

Twenty animals of the Baikal and Tuva native Russian sheep breeds have been resequenced resulting in ~17 million SNP identified, of which approximately 30% were absent from the SNPdb database at the NCBI. Analysis of the sequencing data has identified genome areas likely to be under selective pressure in these Siberian sheep populations. The Tuva breed showed a much larger number of areas with an increased fixation index (Fst) relative to the Baikal breed when we used as a contrast 14 sheep genomes from the UK. This may be due to the remote phylogenetic position of the Tuva breed from the European sheep. Interestingly, almost every region with an elevated fixation index in both breeds contained a protein-coding gene and about 50% of them contained nonsynonymous nucleotide changes with high Fst values, suggesting possible effect of these substitutions on the protein functions in these breeds. For the Tuva breed, the highest Fst was observed in the PPP1CC gene area, which was shown earlier to be involved in the process of fat storage in livestock. Another gene (HHIP), which affects the differentiation of adipocytes, is under likely selection pressure in the Tuva breed on chromosome 17. In the Baikal breed, the highest fixation index was observed in the area of the gene ZNF280B, the transcription factor that regulates many other genes. The second top significant area of the Baikal breed contained the RALY gene, which, as previously shown, affects the quality of wool in sheep. A number of other genes affect immunity (TARBP1 and CLSTN1) and may thus be associated with adaptation to local conditions. Since we were able to identify nucleotide variations under selective pressure in these two populations, these variants can be tested for effects in genomic selection programmes in Russian sheep breeds.

Acknowledgments

This work was funded by the RSF grant 19-76-20026.

DOI 10.18699/BGRS/SB-2020-163

Heat shock protein 90 as a long-term buffer of mutational burden

Valeria Timonina*
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University
Kaliningrad, Russia
valeratimonina@gmail.com

Anastasia Sokol*
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University
Kaliningrad, Russia
anastasia3sokol@yandex.ru

Dmitry Knorre
Belozersky Institute Of Physico-Chemical Biology
Moscow State University
Moscow, Russia
knorre@belozersky.msu.ru

Evgenii Tretiakov
Department of Molecular Neurosciences, Medical University of Vienna, Vienna, Austria
evgenii.o.tretiakov@gmail.com

Konstantin Gunbin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
genkvg@gmail.com

Konstantin Popadin
School of Life Sciences
Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University
Kaliningrad, Russia
konstantinpopadin@gmail.com

*- *this authors contributed equally to this work*

***Abstract* — Heat shock proteins play an important role in protein homeostasis of cells maintaining the correct folding of disturbed proteins. Several lines of evidence suggest the role of hsp90 in buffering mutational burden in model organisms. Here we propose a hypothesis that hsp90 may act as a long-term buffer of mutational load influencing species evolution. To prove the hypothesis we use interspecies comparison of expression and genomic data.**

Keywords* — *hsp90, mutation buffering, evolutionary constraints

Introduction

Heat shock proteins (hsp) maintain optimal folding of their clients - proteins, affected by environmental (temperature) or genetic (a load of slightly deleterious mutations) perturbations. In several short-term experiments (typically from one to several generations) it has been shown that hsp90 acts as a buffer of the mutational burden influencing the manifestation of slightly-deleterious variants [1–4]. Here, we propose that hsp90 may play also a long-term evolutionary role compensating, for example, an increased mutational burden in species with low effective population size (Ne). Thus hsp90 might be under stronger evolutionary constraints in low-sized versus high-sized populations.

Results

Taking into account that the number of fixed slightly deleterious nonsynonymous variants is higher in species with low effective population size [5,6], we expect that the expression level of hsp90 in these species is increased to partially compensate the high burden of slightly-deleterious nonsynonymous variants (SDNVs). Indeed using comparative species expression level data [8] we found out that expression level of hsp90 is the lowest in Macaque, which is characterized by the highest effective population size among other six analyzed [8] primates (Ne of macaque ~ 80’000, Yuan et al.2012; human: 13’100-16’200; gorilla: 28’400-56’900; chimp: 30’900-61’800; orangutan: 42’300-84’600, data from [9]).

Analyses of hsp90 expression pattern between different human tissues (the highest in the ovary, GTEX portal - https://www.gtexportal.org/) and different stages of mammalian development (the highest expression is observed at early stages of development when the most essential genes are expressed) [9] is in line with supportive role of hsp90 during the development.

Next, we assumed that because of this potential compensatory role, hsp90 might be under stronger evolutionary constraints in low-sized versus high-sized populations. In order to compare the level of relaxation of hsp90 in different mammalian species, we performed a linear regression analysis with scaling of variables between species-specific Kn/Ks of hsp90 (using only 1-to-1 orthologs of human Hsp90) and species-specific generation length (which approximates the effective population size) of 38 mammalian species. We obtained weak positive correlation (slope = 0.2864, p-value = 0.05, R^2 = 0.082) which is in line with our expectations (Fig 1.). In order to make the final conclusion about selection relaxation of hsp90 we compared the hsp90 specific slope with slopes of genes similar to hsp90. To derive a subset of genes comparable to hsp90 we performed Principal Component Analysis taking into account evolutionary characteristics of human genes such as the probability of being Loss of Function intolerant, haploinsufficiency [7], etc. as well as structural features like average exon length, gene length, number of transctripts, etc.

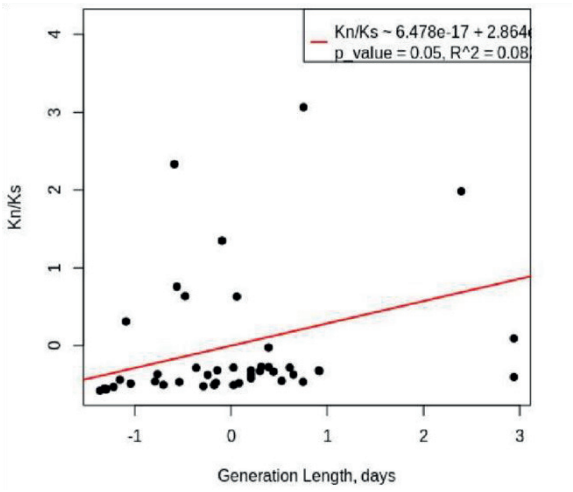


Fig. 1. Linear regression of Kn/Ks vs Generation Length for hsp90

We expect that the hsp90-specific slope should be less steep compared to other genes. Across comparable genes, hsp90 has one of the lowest mean Kn/Ks that emphasizes its high level of evolutionary constraints. However detailed analysis of slopes from linear regression models of Kn/Ks vs Generation length did not show robust signal of the decreased hsp90-specific slope versus the slopes of other genes.

Finally, we analyzed the distribution of cis-eQTLs (GTEX portal - https://www.gtexportal.org/) in hsp90 compared to a subset of similar genes (see the previous section). We observed that hsp90 has a significantly lower assessed allele frequency of cis-eQTLs than in other genes (Mann-Whitney test, W = 88256, p-value = 0.025). This fact suggests that hsp90 due to its role in homeostasis maintenance should keep a constant expression level and so it has a deficit of regulatory expression variants compared to other genes.

Altogether we observed several lines of evidence, confirming the compensatory role of hsp90 but more analyses are needed to confirm it's long-term bufferisation effect.

Acknowledgments

This work has been supported by the 5 Top 100 Russian Academic Excellence Project at the Immanuel Kant Baltic Federal University, by the Russian Foundation for Basic Research grant 18-29-13055.

References

[1] Fares, M. A., Ruiz-González, M. X., Moya, A., Elena, S. F., & Barrio, E. (2002). GroEL buffers against deleterious mutations. *Nature*, *417*(6887), 398-398.

[2] Sangster, T. A., Salathia, N., Lee, H. N., Watanabe, E., Schellenberg, K., Morneau, K., ... & Lindquist, S. (2008). HSP90-buffered genetic variation is common in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *105* (8), 2969-2974.

[3] Rutherford, Suzanne L., and Susan Lindquist. "Hsp90 as a capacitor for morphological evolution." *Nature* 396.6709 (1998): 336. Maisnier-Patin, S., Roth, J., Fredriksson, Å. *et al.* Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nat Genet* 37, 1376–1379 (2005).

[4] Popadin, K., Polishchuk, L. V., Mamirova, L., Knorre, D., & Gunbin, K. (2007). Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proceedings of the National Academy of Sciences*, *104*(33), 13390–13395.

[5] Popadin, K. Y., Gutierrez-Arcelus, M., Lappalainen, T., Buil, A., Steinberg, J., Nikolaev, S. I., Antonarakis, S. E. (2014). Gene age predicts the strength of purifying selection acting on gene expression variation in humans. *American Journal of Human Genetics95*(6), 660–674.

[6] Bartha, I., di Iulio, J., Venter, J. *et al.* Human gene essentiality. *Nat Rev Genet* 19, 51–62 (2018). https://doi.org/10.1038/nrg.2017.75

[7] Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., ... & Albert, F. W. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, *478*(7369), 343-348.

[8] Prado-Martinez, J., Sudmant, P. H., Kidd, J. M., Li, H., Kelley, J. L., Lorente-Galdos, B., ... & Cagan, A. (2013). Great ape genetic diversity and population history. *Nature*, *499*(7459), 471-475.

[9] Cardoso-Moreira, M., Halbert, J., Valloton, D. *et al.* Gene expression across mammalian organ development. *Nature* 571, 505–509 (2019).

DOI 10.18699/BGRS/SB-2020-164

Analysis of sequenced chromosome-specific libraries of gekkonids sheds light to large scale genome reshuffling in reptiles

Katerina Tishakova
Institute of Molecular and Cellular Biology SB RAS
Novosibirsk, Russia
Novosibirsk State University
Novosibirsk, Russia
k.tishakova@g.nsc.ru

Ilya Kichigin
Institute of Molecular and Cellular Biology SB RAS
Novosibirsk, Russia
kig@mcb.nsc.ru

Artem Lisachov
Institute of Cytology and Genetics SB RAS
lisachov@bionet.nsc.ru

Dmitry Prokopov
Institute of Molecular and Cellular Biology SB RAS
Novosibirsk, Russia
dprokopov@mcb.nsc.ru

Anna Molodtseva
Institute of Molecular and Cellular Biology SB RAS
Novosibirsk, Russia
rada@mcb.nsc.ru

Vladimir Trifonov
Institute of Molecular and Cellular Biology SB RAS
Novosibirsk, Russia
Novosibirsk State University
Novosibirsk, Russia
vlad@mcb.nsc.ru

Lukáš Kratochvil
Charles University
Prague, Czech Republic,
lukkrat@email.cz

***Abstract* — Representatives of Infraorder Gekkota demonstrate relatively low chromosomal evolution and rapidly evolved sex chromosome systems. Using low-coverage chromosome sequencing, we sequenced chromosome-specific libraries of *Gekko japonicus* and putative sex chromosome of *Hemidactylus platyurus* to compare with annotated genomes of squamate reptiles. We found evolutionary long conservation of macro-chromosome and high dynamic of micro-chromosomes. Also, we discovered homologues synteny blocks between sex chromosomes of species from different infraorder.**

Keywords* — *Karyotypic evolution, macro- and micro-chromosomes, sex chromosomes, Gekko japonicus

Introduction

Infraorder Gekkota represents a specious infraorder of reptiles, taking a basal position relative to a large clade of squamate reptiles, including snakes, iguanids, scincids, lacertids and anguimorphs. The infraorder is characterized by unimodal karyotype without macro- and micro-chromosomes. The ancestral karyotype is known to have consisted of 38 acrocentric chromosomes, and modern karyotypes differ from it by a relatively small number of chromosomal rearrangements [1]. Also, geckos demonstrate the rich diversity of sex chromosome systems [2]. However, most chromosome-wide assemblies of the genome are available for lacertids, snakes, varanids and iguanids are still missing in Gekkota. Here we applied low-coverage chromosome sequencing to evaluate the genetic content in a representative of geckos – *Gekko japonicus* and putative sex chromosome of *Hemidactilus platyrus* to compare with annotated genomes of squamate reptiles and study homologous synteny blocks and evolutionary breakpoints regions.

Metods

The chromosome-specific libraries of *Gekko japonicus* were prepared from chromosome sorted with a dual laser cell sorter (Mo-Flo, Dako) at the Cambridge Resource Center for Comparative Genomics, Department of Veterinary Medicine, University of Cambridge, Cambridge, UK, as described previously [3]. Libraries were amplified and labeled using DOP-PCR, chromosome specificity of libraries was verified by FISH. To prepare chromosome-specific libraries for sequencing, we used TruSeq Nano DNA Low Throughput Library Prep Kit (Illumina). 300 bp paired-end reads were generated on Illumina MiSeq using the Illumina MiSeq Reagent Kit v3 according to the manufacturer's instructions. Sequencing data were processed using the DOPseq_analyzer pipeline [4]. At first, the cutadapt removes of Illumina adapter sequences and DOP primers. The purified reads are aligned on the *Lacerta agilis* and *Crotalus viridis viridis* genomes to search for target regions and the human genome to remove contamination reads using BWA, low-quality alignments were discarded. Then, by calculation, the density of alignment, the identification of target regions occurs using DNAcopy package.

Results

We compared the obtained sequencing data with the annotated genome of *Lacerta agilis* (Lacertidae, Squamata) and *Crotalus viridis viridis* (Viperidae, Squamata) and found evolutionary long conservation of macro-chromosome along with high dynamics of micro-chromosomes. We refined the breakpoints in reptile genomes and demonstrated the breakpoint re-use. Also, we found homologues of the putative sex chromosome of *Hemidactilus platyurus* in the *G. japonicus* karyotype, which is not a sex chromosome. However, the putative sex chromosome of *H. platyurus* has a synteny block with the sex chromosome of the *Noropus sagrey* (Iguania, Squamata) and contains the RSPO1 gene. This result suggested that in different reptile lines, the same genetic material can independently be involved in the formation of sex chromosomes.

Acknowledgment

This work was supported by RFBR 19-54-26017.

References

[1] Pokorná M.J., Trifonov V., Rens W., Ferguson-Smith M. ., & Kratochvíl L. (2015). Chromosome Research, 23(2): 299-309.
[2] Gamble T. (2010). Sexual Development, 4(1-2), 88-103.
[3] Yang F., Carter N., Shi L., & Ferguson-Smith M. (1995). Chromosoma, 103(9): 642-652.
[4] Makunin A.I., Kichigin I.G., Larkin D.M., O'Brien P.C., Ferguson-Smith M.A., Yang F., ... & Graphodatsky A.S. (2016). BMC Genomics, 17(1): 618.

DOI 10.18699/BGRS/SB-2020-165

FMO superfamily protein phylogeny and the origin of YUCCA family

Igor Turnaev
Systems Biology Department
ICG SB RAS, Novosibirsk, Russia
turn@bionen.nsc.ru

Valentin Suslov
Systems Biology Department
ICG SB RAS, Novosibirsk, Russia
valya@bionen.nsc.ru

Dmitriy Afonnikov
Systems Biology Department
ICG SB RAS, Novosibirsk, Russia
ada@bionen.nsc.ru

Konstantin Gunbin
Systems Biology Department
ICG SB RAS, Novosibirsk, Russia
genkvg@bionen.nsc.ru

***Abstract* — We performed a comparative analysis of phylogeny, conservative sites, protein domains, and the presence in taxa of FMOs and BVMOs proteins. Earlier in the FMOs, the following protein groups were distinguished: type I FMO, type II FMO, YUCCAs. According to the results of our analysis, we propose that the type II FMO group is, in turn, divided into two groups: FMO IIA type and type IIB FMO. Type IIB FMO proteins represent a new special group in a number of parameters (site consensus, composition of protein domains, presence in taxa and, apparently, functions) significantly different from all other FMOs proteins and type IIA FMO proteins, in particular. While type IIA FMO proteins exhibit properties typical of the FMO subclass. The type IIB FMO protein group is likely to represent a new family of proteins. These proteins are of interest for biocatalysis.**

Keywords* — *protein phylogeny; protein evolution; flavin monooxygenase; auxin biosynthesis; YUCCA family; conservative motifs

Background

YUCCA proteins are important plant enzymes involved in the biosynthesis of one of the key hormones, auxin. [1]. It is believed that these genes originated as a result of horizontal transfer from bacteria to plant genome [2].

YUCCA enzymes belong to the superfamily of flavin-dependent monooxygenases (FMO), therefore, studying the position of the proteins of the YUCCA family within the FMO superfamily can shed light on the origin of the auxin biosynthesis pathway in higher plants.

Class B flavin-dependent monooxygenases (EC 1.14.13) include three subclasses: N-hydroxylating monooxygenases (NMOs), Baeyer – Villiger monooxygenases (BVMOs), and flavin-containing monooxygenases (FMOs). The YUCCA protein family belongs to the last subclass of FMOs [3].

In this work, a detailed analysis of the phylogeny of FMOs proteins was carried out, which allowed us to make an assumption that there are two subgroups, type IIA and IIB, in the type II FMO group. Subgroup IIA, but not IIB, is sister to YUCCA. Based on these data, assumptions were made about the origin and evolution of the proteins of the YUCCA family.

Materials and methods

Proteins of YUCCAs family and their closest homologues were extracted from NCBI database using protein BLAST with a threshold e-value < 1e-5. Homologs of protein YUC2 (AT4G13260) of Arabidopsis thaliana were the BLAST queries; only full-size proteins were taking into account. Promals used for alignment reconstruction. The phylogenetic tree was built using Maximum Likelihood methods. Protein motifs were built using WebLogo (HTTPS://WEBLOGO.BERKELEY.EDU).

Results and Discussions

We have built a FMO-like (YUCs homologues) phylogenetic tree (as a target for the search was taken protein YUC2 Arabidopsis thaliana – At4G13260, e-value ≤ 1e-05) (fig. 1). In addition to FMO proteins, Baeyer-Villiger monooxygenases (BVMO) were taken as outgroup.

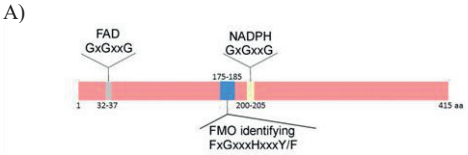
In the phylogenetic tree, we have identified 5 groups of proteins: (1) YUCCAs; (2) type II FMO (on the basis of Riebel's FMO classification – [4]); (3) type I FMO (on the basis of Riebel's FMO classification); (4) NMOs; (5) BVMOs. According to the results of the analysis group (2) - type II FMO is divided into 2 subclusters, which we called type IIA FMO and type IIB FMO.

Comparative analysis of conservative motifs among groups of FMO proteins.

All proteins of type B flavin monooxygenases (FMOs, NMOs and BVMOs) contain three conservative sites (fig. 1A). Two typical Rossman fold motives (GxGxxG/A): the first is FAD binding site located closer to the N-terminus and the second is the NADPH binding site near the middle of the protein [4, 5]. Between them is the FMO identification site (FxGxxxHxxxF/Y/W).

In the sequences of the FMOs subclass, the last amino acid at this site is F/Y, and that of the BVMO subclass is W. NMOs has lost its identification site, leaving only one central histiditis (xxxxxxHxxxx) [6]. In this connection, NMOs proteins are not included in further analysis.

A)



B)

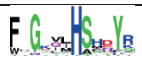
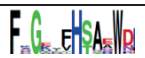
Conservative sites names	FAD binding site	FAD identified site	NADPH binding site
Site motives	GxGxxG	FxGxxxHxxx$^{Y}/_{F}{}^{K}/_{R}$	GxGxxG/A
i. YUCCA			
ii. type IIA FMO			
iii. type IIB FMO			
iv. type I FMO			
v. BVMOs			

Fig. 1. Distribution and alignment of key functional sites of FMOs / BVMOs homolog proteins. A) The localization of motifs along FMO proteins is shown schematically (the amino acid positions in the diagram are indicated by the sequence AT4G13260 - Arabidopsis thaliana YUC2). B) The table, respectively, in columns 2-4 shows the consensus of the three motives in FMOs / BVMOs proteins: FMO-binding, FMO identifying and NADH-binding sites.

We analyzed the sequences of three functional motifs in five protein clusters isolated on the phylogenetic tree: YUCCAs, type IIA FMO, type IIB FMO, type I FMO, BVMOs (fig. 1B). The motives for these three sites were built using WebLogo. The first motive is the FAD binding site - the motif for the sequences of all protein groups is similar and contains 3 highly conserved glycines. The most interesting are the cardinal differences between the sequences on the one hand, type IIB FMO, and on the other, type I, IIA, IIC FMOs, YUCCAs in the second - FMO identification site - and in the third - NADPH binding site - motives. So, in type IIB FMO sequences, in the FMO identified motif, in 70% of cases, instead of Y/F, histidine (H) is located, and in the NADPH motif, in the third position, in 75% case, instead of glycine (G) is asparagine (N). This suggests that type FMO IIB proteins may, like BVMO proteins, belong to a family separate from other FMOs that perform other functions than YUCCAs proteins, type I FMO and type IIA FMO.

Comparative analysis of domains of FMO proteins

The analysis of FMO and BVMO protein domains showed that type FMO IIB proteins contain an N-terminal domain of about 160 aa. not characteristic of either FMO-like (YUCs, type I FMO, type FMO IIA), nor for BVMO proteins.

Characteristic group of proteins type IIB FMO

The type FMO IIB includes proteins of bacteria, fungi, charophyte *Klebsormidium nitens* GAQ32387 protein and 4 proteins from club moss Selaginella moellendorffii. Previously identified by the group Wang et. al. [7] homology of GAQ32387 *K. nitens* protein with land plant YUCs proteins allowed them to hypothesize the earlier origin (TAA-YUCCA) of the auxin biosynthesis pathway already in charophytes, and not in the first land plant, as previously thought [2]. According to our data, GAQ32387 refers to type IIB FMO. An additional search for GAQ32387 homologs among the RNAseq sequences of the base 1000 plant genomes revealed its close homologues among the red algae, lower green algae, moss, club moss, ferns, gymnosperms, monocots and dicots. This implies a significantly wider distribution of type FMO IIB proteins in plants than could be expected.

Conclusions

1) According to the results of our analysis, we propose the group protein type II FMO (according to the classification Riebel et al.'s [4]) divided into two groups type FMO IIA and type FMO IIB; 2) The type FMO IIB proteins seem to possess properties unique to FMO and BVMO (fig. 5), as has already been shown for proteins FMOe, f, g R. jostii RHA1 [8]; 3) The remoteness of type FMO IIB proteins and, accordingly, GAQ32383 *K. nitens* protein from YUCs on phylogeny and their structural and probably functional uniqueness is in favor of the origin of YUCs in first land plants (Yue et. al.'s [2] hypothesis), and not in Charophyceae (Wang et. al.'s [Wang et. al., 2014 7] hypothesis).

Acknowledgment

The work supported by the budget project No. 0324–2019-0040-C-01.

References

[1] K. Mashiguchi, K. Tanaka, T. Sakai, S. Sugawara, H. Kawaide, M. Natsume et al., "The main auxin biosynthesis pathway in Arabidopsis." Proc. Natl. Acad. Sci. U S A, vol. 108(45), pp. 18512-18517, 2011.

[2] Yue J., X. Hu, and J. Huang, "Origin of plant auxin biosynthesis." Trends in Plant Science, vol. 19(12), pp. 764-770, 2014.

[3] M. M. E. Huijbers, S. Montersino, A. H. Westphal, D. Tischler, W. J. H. van Berkel, "Flavin dependent monooxygenases." Archives of Biochemistry and Biophysics, vol. 544, pp. 2-17, 2014.

[4] A. Riebel, G. de Gonzalo, M. F. Fraaije, "In Expanding the biocatalytic toolbox of flavoprotein monooxygenases from Rhodococcus jostii RHA1." Enzymatic, vol. 88, pp. 20– 25, 2013.

[5] A. Riebel, H. M. Dudek, G. de Gonzalo, P. Stepniak, L. Rychlewski, M. W. Fraaije, "Expanding the set of rhodococcal Baeyer–Villiger monooxygenases by high-throughput cloning, expression and substrate screening. Appl Microbiol Biotechnol, vol. 95(6), pp. 1479–1489, 2012.

[6] M. W. Fraaije, N. M. Kamerbeek, W. J van Berkel, D.B Janssen, "Identification of a Baeyer-Villiger monooxygenase sequence motif." FEBS Lett., vol. 518(1-3), pp. 43-47, 2002.

[7] C. Wang, Y. Liu, S.-S. Li and G.-Z. Han, "Origin of plant auxin biosynthesis in charophyte algae." Trends in Plant Science, vol. 19(12), pp. 741-743, 2014.

[8] A. Riebel, M. J. Fink, M. D. Mihovilovic, M. W. Fraaije, "Type II Flavin-Containing Monooxygenases: A New Class of Biocatalysts that Harbors Baeyer–Villiger Monooxygenases with a Relaxed Coenzyme Specificity." ChemCatChem, 2014, vol. 6(4), pp. 1112-1117, 2014.

DOI 10.18699/BGRS/SB-2020-409

A global human mitochondrial tree as a resource for population and evolutionary studies

Kristina Ushakova
ITMO University, IKBFU Center for Mitochondrial Functional Genomics
chubbyprogrammer@gmail.com

Viktor Shamanskiy
IKBFU Center for Mitochondrial Functional Genomics

Alina A Mikhailova
IKBFU Center for Mitochondrial Functional Genomics

Alina G Mikhailova
IKBFU Center for Mitochondrial Functional Genomics

Evgeniy Tretiakov
Medizinische Universität Wien

Ilia Mazunin
Skoltech Center of Life Sciences
Moscow, Russia

Konstantin Popadin
IKBFU Center for Mitochondrial Functional Genomics, Ecole polytechnique Federale de Lausanne

Konstantin Gunbin
IKBFU Center for Mitochondrial Functional Genomics,
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

***Abstract* – The mitochondrial DNA (mtDNA) is a highly streamlined genome with essential role for cellular metabolism. To create a resource for investigation of various aspects of human mtDNA evolution we reconstructed global human mtDNA tree using 40000 complete human mtDNAs. To make our reconstructions robust we used several alternative models of nucleotide substitutions (TN92, GTR); several alternative mitochondrial genomes with more than one million reconstructed single-nucleotide substitutions. We confirmed high quality of our reconstructions, demonstrating: (i) an absence of stop codons on internal nodes, (ii) strong expected excess of C>T and A>G transitions (heavy chain notation) among synonymous substitutions etc. As a first task demonstrating utility of our tree, we reconstructed detailed mtDNA germline mutational spectrum (global, gene-specific and with nucleotide context) based on a collection of more than 300'000 synonymous substitutions. Next we compared it with somatic mtDNA mutational spectrum, derived from cancer data. Globally, two mutational spectra were similar to each other, while local shifts in the frequency of several types of substitutions may be explained by different mutagenic environment between female gamete line and various somatic tissues. We hope that this resource will help to address numerous questions of human mtDNA evolution in the future.**

Keywords* – *Mitochondrial genetics, mutational spectrum, humans, tree reconstruction, ageing, cancer

The mitochondrial DNA (mtDNA) is a highly streamlined genome with an essential role for cellular metabolism. To create a resource for investigation of various aspects of human mtDNA evolution we reconstructed the global human mtDNA tree using 40000 complete human mtDNAs. To make our reconstructions robust we used several alternative models of nucleotide substitutions (TN92, GTR); several alternative approaches to reconstruct ancestral states for each nucleotide in each node (parsimony, empirical Bayesian approach and ML marginal reconstruction) and tested statistical significance of each internal node (SH-like approximate likelihood ratio test, approximate Bayes test and ultrafast bootstrap approximation).

approaches to reconstruct ancestral states for each nucleotide in each node (parsimony, empirical Bayesian approach and ML marginal reconstruction) and tested statistical significance of each internal node (SH-like approximate likelihood ratio test, approximate Bayes test and ultrafast bootstrap approximation). As a result we obtained a detailed phylogenetic tree of human

As a result we obtained a detailed phylogenetic tree of human mitochondrial genomes with more than one million reconstructed single-nucleotide substitutions. We confirmed the high quality of our reconstructions, demonstrating: (i) an absence of stop codons on internal nodes, (ii) strong expected excess of C>T and A>G transitions (heavy chain notation) among synonymous substitutions etc. As a first task demonstrating the utility of our tree, we reconstructed a detailed mtDNA germline mutational spectrum (global, gene-specific and with nucleotide context) based on a collection of more than 300'000 synonymous substitutions. Next we compared it with somatic mtDNA mutational spectrum, derived from cancer data. Globally, two mutational spectra were similar to each other, while local shifts in the frequency of several types of substitutions may be explained by different mutagenic environments between female gamete line and various somatic tissues. We hope that this resource will help to address numerous questions of human mtDNA evolution in the future.

From that reconstructed tree we pulled all of the mutations, annotated them and filtered out a total of 200 thousand synonymous four-fold degenerate substitutions from mtDNA and calculated the spectrum. Analyses showed the following:

- Germline is more protected, so germline vs somatic mtDNA mutational spectrum demonstrates decreased chemical damage asymmetry
- Higher C>T in context of CpG profile in somatic data is a signature of age (experimentally linked to SBS1 signature)
- Higher C>T in context of CpC in germline is a marker of redox stress (supported by experiments on single-stranded yeast DNA).

DOI 10.18699/BGRS/SB-2020-166

Inter- vs. intraspecific genetic variability of morphologically similar *Ligophorus* species

Ekaterina Vodiasova
IBSS RAS, Sevastopol, Russia
eavodiasova@gmail.com

Evgenija Dmitrieva
IBSS RAS, Sevastopol, Russia
genijadmitrieva@gmail.com

Olga Shikhat
IBSS RAS, Sevastopol, Russia
shixatolya@yandex.ru

Alexei Ermolenko
FSC the East Asia Terrestrial Biodiversity, Vladivistok, Russia
ermolenko_alexey@mail.ru

Dmitry Atopkin
FSC the East Asia Terrestrial Biodiversity, Vladivistok, Russia
pan2006_82@mail.ru

***Abstract* — Monogeneans of *Ligophorus* (Platyhelminthes), specific parasites of the fish of Mugilidae, present comprises 60 species based mainly on morphology. The contradictions between morphological and genetic (rDNA fragments) inter- vs. intraspecific variability is shown for some species. Six species of *Ligophorus* newly collected in the Black Sea, South China Sea and Sea of Japan and 80 sequences of ITS1 and 28S were obtained. Two species, *L. satunensis* and *L. fenestrum*, from the three neighboring regions of the Western Pacific (WP) have no morphological differences but diverge from each other in 3.3–5.8% by 28S rRNA. In contrast to this, the low intraspecific genetic variability (0.1–0.6% for 28S) is revealed for sympatric *L. pilengas* and *L. llewellyni*, distinguishing in morphology. Due to the different rates of divergence of species of this genus, the integrated morphological and genetic approach, considering the ecology, phylogeography and isolation of host populations, for the correct phylogenetic reconstruction of this genus are suggested.**

Keywords* — *molecular taxonomy, Ligophorus, phylogeny, morphology, rDNA

Motivation and aim

The traditional taxonomy of *Ligophorus* spp. is based on the morphology of the structures of the posterior attachment organ and male reproductive system. Among dozens of articles dedicating systematics of these monogeneans, only five of them uses rDNA sequences to identify phylogenetic relationships between species. However, a lack of enough data on intraspecific genetic variation leads sometimes to results contradicting with morphology. The most species of *Ligophorus* have only two or three rDNA sequences deposited to GenBank. Moreover, the genetic distances between species varies widely, from 0.4 to 50.2 % for both ITS1 and 28S [1-3].

The aim of this work is to analyze the interspecific and intraspecific genetic variability of *Ligophorus* species with similar morphology.

Methods

Six species of *Ligophorus* newly collected in the Black Sea, South China Sea and Sea of Japan and additional data on some species from GenBank NCBI are included in analysis. 80 sequences of ITS1 and 28S from 40 specimens were obtained. DNA extraction was carried out using innuPREP DNA Mini Kit. The fragment of 28S was amplified using the primers U178 (5'- GCA CCC GCT GAA YTT AAG -3'; [4]) and LSU1200R (5'- GCA TAG TTC ACC ATC TTT CGG -3'; [5]). ITS1 was amplified with the primers Lig18endF (5- GTC TTG CGG TTC ACG CTG CT -3) and Lig5.8R (5- GAT ACT CGA GCC GAG TGA TCC -3) [1]. The reaction conditions consisted of initial denaturation at 95°C for 3 min, followed by 35 cycles of denaturation at 94°C for 40 s, annealing at 56°C for 30 s and extension at 72°C for 45 s, the final extension at 72°C for 4 min. PCR products were sequenced in both directions using the standard BigDye Terminator Cycle Sequencing Ready Reaction Kit and a ABI PRISM 3130 analyzer (Applied Biosystems Inc.). The obtained fragments of rDNA were aligned in the BioEdit software program [7]. Evolutionary analyses were conducted in MEGA7 [8] using a specimen of *Ergenstrema mugilis* from GenBank as a reference sequence for ITS1 (GenBank JN996835) and 28S (GenBank JN996800).

Results

Two species, *L. satunensis* and *L. fenestrum,* from the three neighboring regions of the Western Pacific (WP) have no morphological differences but diverge from each other in 3.3–5.8% by 28S rRNA. Additionally, the high intraspecific genetic variability in this gene is revealed for *L. johorensis* (2.6%) and *L. kedahensis* (2.8%) also from different localitis of WP. The same level of genetic divergence (2.7%) was previously observed between two morphologically very similar sympatric species, *L. uruguayense* and *L. saladensis* [3]. However, inter- and intraspecific distances overlap for 28S for *L. saladensis* and *L. mediterraneus* from different hosts and regions. The low intraspecific genetic variability (0.1–0.6% for 28S) is also revealed for sympatric *L. pilengas* and *L. llewellyni*, distinguishing in morphology.

The revealed contradictions between morphological and genetic variability indicate different rates of divergence of species of this genus. Thus, a comprehensive morphological and genetic study, considering the ecology, phylogeography and isolation of host populations, is necessary for the correct phylogenetic reconstruction of *Ligophorus*.

Acknowledgment

This work is funded by a scientific theme of IBSS number AAAA-A18-118020890074-2, RFBR and Sevastopol project number 20-44-920004, and “Ecolan 3.1” of Vietnam-Russian Tropical Research and Technological Centre.

References

[1] Blasco-Costa I., Miguez-Lozano R., Sarabeev V., Balbuena J.A. (2012) Molecular phylogeny of species of *Ligophorus* (Monogenea: Dactylogyridae) and their affinities within the Dactylogyridae. Parasitology International. 61: 619–627.

[2] Pakdee W, Ogawa K, Pornruseetriratn S, Thaenkham U, Yeemin T (2018). The first record of *Ligophorus* Euzet & Suriano, 1977 (Monogenea: Dactylogyridae) on *Crenimugil buchanani*

(Teleostei: Muglidae) from Thailand based on morphological and molecular analyses. Journal of Helminthology. 93(6):752–762.

[3] Marchiori N.C., Pariselle A., Pereira Jr. J., Agnèse J.-F., Durand J.-D., Vanhove M.P.M. 2015: A comparative study of *Ligophorus uruguayense* and *L. saladensis* (Monogenea: Ancyrocephalidae) from *Mugil liza* (Teleostei: Mugilidae) in southern Brazil. Folia Parasitol. 62: 024. doi: 10.14411/fp.2015.024

[4] Lockyer AE, Olson PD, Littlewood DTJ. Utility of complete large and small subunit rRNA genes in resolving the phylogeny of the Neodermata (Platyhelminthes): im- plications and a review of the cercomer theory. Biological Journal of the Linnean Society 2003;78:155–71.

[5] Littlewood DTJ, Curini-Galletti M, Herniou EA. The interrelationships of Proseriata (Platyhelminthes: Seriata) tested with molecules and morphology. Molecular Phylogenetics and Evolution 2000;16:449–66.

[6] Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser. 41: 95-98.

[7] Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33:1870-1874.

DOI 10.18699/BGRS/SB-2020-167

Genome and karyotype evolution after whole genome duplication in free-living flatworms of the genus *Macrostomum*

Kira Zadesenets
ICG SB RAS, Novosibirsk, Russia
kira_z@bionet.nsc.ru

Nikita Ershov
ICG SB RAS, Novosibirsk, Russia
nikotinmail@mail.ru

Dmitry Oshchepkov
ICG SB RAS, Novosibirsk, Russia
diman@bionet.nsc.ru

Eugene Berezikov
ERIBA, Groningen, The Netherlands
ICG SB RAS, Novosibirsk, Russia
e.berezikov@umcg.nl

Lukas Schärer
Evolutionary Biology, Zoological Institute, University of Basel, Basel, Switzerland
lukas.scharer@unibas.ch

Nikolay B. Rubtsov
ICG SB RAS, Novosibirsk, Russia
rubt@bionet.nsc.ru

***Abstract* — Whole genome duplication (WGD) is a large-scale evolutionary transformation that took place in genome evolution in many taxa of existing animal species. However, the mechanisms underlying the early stages of genome evolution after a WGD event in animals has remained unclear. The study of genome organization of neopolyploid species may shed light on the processes of genome reorganization leading to its re-diploidization after a recent round of WGD.**

Keywords* — *whole genome duplication, hidden polyploidy, karyotype instability, genome evolution, macrostomids, flatworms

Motivation and aim

Motivation and Aim

Earlier we uncovered a group of free-living flatworms in which genomes have likely undergone a recent WGD. We found out that karyotype instability was linked to hidden polyploidy in both species *M. lignano* (2n = 8) and its sibling species *M. janickei* (2n = 10) [1, 2]. Additionally we studied other species of the genus *Macrostomum* and revealed a new species (further called *M. mirumnovem*) with a highly unstable karyotype. We studied the detailed karyotype and genome organization in three *Macrostomum* species.

Methods

To describe karyotype diversity in the *Macrostomum* species (*M. lignano*, *M. janickei*, *M. mirumnovem*) the routine single-worm karyotyping was performed. The detailed cytogenetic analyses were made using a set of different DNA probes (microdissected region- and chromosome specific DNA probes, DNA repeats, unique DNA fragments). Microdissected DNA probes (specific to regions and whole chromosomes) were generated for the *M. lignano* and *M. mirumnovem* chromosomes. The generated chromosome-specific DNA probes for the *M. lignano* chromosomes were sequenced and the obtained NGS data were used for following bioinformatics analyses using the existing genome assemblies for the *M. lignano* genome.

Results

The detailed cytogenetic analyses using a set of different DNA probes (microdissected region- and chromosome specific DNA probes, DNA repeats, unique DNA fragments) revealed the peculiarities of karyotype and genome organization in three *Macrostomum* species with unstable karyotypes. Similar to *M. lignano* and *M. janickei*, our findings suggest that *M. mirumnovem* arose via whole genome duplication (WGD) followed by considerable reshuffling of its chromosomes. The combined approach allowed us to explore the peculiarity of the *M. lignano* genome organization, the strategy included bioinformatics analysis of the existed genome assemblies of *M. lignano* and NGS data for its separate chromosomes. Based on the obtained results, we suggest possible evolutionary scenarios for the emergence and reorganization of the karyotypes and genomes of the post-WGD *Macrostomum* species.

Acknowledgment

This study is supported by grant of the Russian Science Foundation (RSF) (project No. 19-14-00211).

References

[1] Zadesenets K.S. et al. (2016) Evidence for karyotype polymorphism in the free-living flatworm, *Macrostomum lignano*, a model organism for evolutionary and developmental biology. PLoS One. 11(10): e0164915.

[2] Zadesenets K.S., Schärer L., Rubtsov N.B. (2017) New insights into the karyotype evolution of the free-living flatworm *Macrostomum lignano* (Platyhelminthes, Turbellaria). Sci. Rep. 7: 6066.

DOI 10.18699/BGRS/SB-2020-168

Genomic analyses of *Novymonas esmeraldas* and *Ca*. Pandoraea novymonadis

Alexandra Zakharova
Life Science Research Centre
University of Ostrava
Ostrava, Czech Republic
alexandraz.6946@gmail.com

Daria Tashyreva
Institute of Parasitology
Biological Centre
České Budějovice, Czech Republic
tashyreva@paru.cas.cz

Jorge Morales
Heinrich Heine University
Düsseldorf, Germany
jjfreites@hotmail.com

Eva Nowack
Heinrich Heine University
Düsseldorf, Germany
E.Nowack@uni-duesseldorf.de

Julius Lukeš
Institute of Parasitology
Biological Centre
České Budějovice, Czech Republic
jula@paru.cas.cz

Vyacheslav Yurchenko
Life Science Research Centre
University of Ostrava
Ostrava, Czech Republic
Vyacheslav.Yurchenko@osu.cz

***Abstract* — Members of the family Trypanosomatidae are intensively studied because some of them are infectious agents of dangerous human ailments, such as Chagas disease, African sleeping sickness and leishmaniasis. It has been recently demonstrated that representatives of 4 trypanosomatid genera of 2 subfamilies [Leishmaniinae (genus *Novymonas*) and Strigomonadinae (genera *Angomonas*, *Strigomonas* and *Kentomonas*)] harbor bacterial symbionts. The endosymbiont of *Novymonas esmeraldas* is *Ca.* Pandoraea novymonadis. This symbiotic system is of particular interest, since none of the participants has close relatives involved in such relationships, indicating its independent and relatively recent origin. In this work, we have investigated genomes and proteomes of N. esmeraldas and Ca. Pandoraea novymonadis. We found and investigated factors involved in the symbiotic interactions. Biochemical analysis demonstrated that numerous host-encoded proteins were enriched in the bacterial fraction, represented by β-oxidation enzymes, transporters, transmembrane proteins and others. We have found that members of this endosymbiotic system may be a good model to study the establishment and development of symbiosis between bacteria and protists.**

Keywords* — *Trypanosomatidae, Novymonas, bacterial endosymbiont, genome, proteome

Introduction

A new member of the family Trypanosomatidae, *Novymonas esmeraldas*, was discovered in 2008 in the vicinity of Atacames (Esmeraldas Province, Ecuador). The parasite was isolated from the intestine of *Niesthrea vincentii* (Hemiptera: Rhopalidae) [1]. *Novymonas esmeraldas* was assigned to a new genus, which was identified as a member of the subfamily Leishmaniinae.

Further research documented the presence of bacterial endosymbionts inside *N. esmeraldas* cells. Phylogenetic analysis demonstrated that endosymbiont is a member of the genus *Pandoraea*. *Pandoraea* are Gram-negative rod-shaped β-proteobacteria of the family Burkholderiaceae [2]. This was a first example of *Pandoraea* being the intracellular endo, and named *Candidatus* Pandoraea novymonadis [2].

Other previously described cases of endosymbiosis in trypanosomatids have been limited to the representatives of the subfamily Strigomonadinae and *Ca*. Kinetoplastibacterium spp. (Alcaligenaceae) [3]. These relationships appear to have been established earlier in evolution. These 2 endosymbiotic systems differ not only because of the participating members. Other distinctions are: i) uncontrolled division of endosymbionts in *N. esmeraldas*, while strigomonadins control the division of their symbiotic bacteria; ii) localization of symbionts in *N. esmeraldas* (vacuolar) and strigomonadins (cytoplasmic); iii) synchronized division of symbionts in strigomonadins resulting in the presence of one bacterium per host cell versus multiple bacteria per host cell in *N. esmeraldas*.

This symbiotic system is especially interesting because neither the protist host nor the bacterial endosymbiont have close relatives involved in such relationship, further confirming their independent origin. In addition, the phylogenetic positions of both the trypanosomatid and its bacterial partner suggest that their relationship was established relatively recently. Thus, this symbiosis system may become a suitable model for studying the early stages of endosymbiosis between a bacterium and an eukaryote.

Genomic and proteomic analyses

Genomes of *N. esmeraldas* and *Ca*. Pandoraea novymonadis have been sequenced. A total length of *N. esmeraldas* genome is 32 Mb. A total length of *Ca*. P. novymonadis genome is 1,16 Mb, which is smaller than that of the free-living *Pandoraea* spp. (4.46–6.5 Mb), but larger than that *Ca*. Kinetoplastibacterium spp. (~0.8 Mb) [2].

Protein fraction of protist host with endosymbiont was analyzed by mass spectrometry. As a result 4,620 peptides were detected, of which 719 were encoded by the endosymbiont and 3,901 by *Novymonas*. Among the host-encoded proteins found in the endosymbiont fraction were two transporters and trifunctional enoyl-CoA-hydratase/enoyl-CoA-isomerase/3-hydroxyacyl-CoA-dehydrogenase, and a few lipid metabolism enzymes. This indicates that *N. esmeraldas* complements the biosynthesis of fatty acids and β-oxidation in *Ca*. Pandoraea novymonadis.

Conclusion

The symbiotic relationship between *N. esmeraldas* and *Ca.* Pandoraea novymonadis can serve as a model for studying the development of endosymbiosis in parasitic protists and other eukaryotes.

ACKNOWLEDGMENT

We thank all the members of our laboratories for helpful and stimulating discussions. This work was supported by the Grant Agency of the Czech Republic grant 20-07186S.

REFERENCES

[1] A. Kostygov, E. Dobáková, A. Grybchuk-Ieremenko, D. Váhala, D. Maslov, J. Votýpka, ... and V. Yurchenko, "Novel trypanosomatid-bacterium association: evolution of endosymbiosis in action," mBio, vol. 7(2), pp. e01985-15, 2016.

[2] A. Kostygov, A. Butenko, A. Nenarokova, D. Tashyreva, P. Flegontov, J. Lukeš, and V. Yurchenko, "Genome of *Ca*. Pandoraea novymonadis, an endosymbiotic bacterium of the trypanosomatid *Novymonas esmeraldas*," Front. Microbiol., vol. 8, pp. 1-13, 2017.

[3] J. Votýpka, A. Kostygov, N. Kraeva, A. Grybchuk-Ieremenko, M. Tesařová, D. Grybchuk, ... and V. Yurchenko, "*Kentomonas* gen. n., a new genus of endosymbiont-containing trypanosomatids of Strigomonadinae subfam. n.," Protist, vol. 165(6), pp. 825-838, 2014.

Symposium
Cognitive science and genomics

DOI 10.18699/BGRS/SB-2020-169

Possibilities of enhancing the neuroprotective effect of autophagy activation in the brain by stimulation of an mTOR-independent pathway of its regulation in a transgenic mouse model of Parkinson's disease

Anna A. Akopyan
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
annaaleksanovna@mail.ru

Aleksandr B. Pupyshev
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
apupyshev@mail.ru

Maria A. Tikhonova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
tikhonovama@physiol.ru

Abstract — **Autophagy induction promotes cell survival that is especially important for neurons which have a limited proliferative resource. Autophagy is regulated by the classical mTOR-dependent mechanism activated by rapamycin, and also via mTOR-independent pathways triggered by trehalose, lithium, metformin, etc. The neuroprotective effect was shown upon combined activation of these pathways *in vitro*. However, the possibilities of enhancing the therapeutic effect of autophagy activation *in vivo* remain unclear.**

Transgenic mice of B6.Cg-Tg (Prnp-SNCA*A53T)23Mkle/J strain with overexpression of alpha-synuclein were used as a model of Parkinson's disease (PD) in the study. In the striatum and substantia nigra, which are mainly affected in PD, trehalose caused an increase in autophagy, as measured by the expression of the autophagy marker LC3-II. Trehalose in combination with rapamycin increased LC3-II expression by two to three times in comparison with the action of rapamycin alone. In the frontal cortex no changes in LC3-II expression were observed neither under the action of trehalose nor with the combined treatment with trehalose and rapamycin. The most pronounced neuroprotective effect was observed upon the combined use of rapamycin and trehalose by the tyrosine hydroxylase expression in the nigrostriatal system marking the restoration of dopaminergic neurons. In the Rotarod test, the mice were tested for motor function. Animals treated with rapamycin, trehalose, or their combination stayed much longer on a rotating drum, compared to controls. Autophagy contribution to the therapeutic effects was confirmed by administration of autophagy inhibitor 3-methyladenine, which completely blocks the neuroprotective effects of the drugs.

Key words — autophagy, rapamycin, trehalose, mice, alpha synuclein, Parkinson's disease

ACKNOWLEDGMENT

This work was supported by a grant No. 16-04-01423 from the Russian Foundation for Basic Research.

DOI 10.18699/BGRS/SB-2020-170

Pharmacological effects of arecoline on zebrafish behavior, neurochemistry, neurophysiology and brain gene expression

Tamara G. Amstislavskaya
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
amstislavskayatg@physiol.ru

Nazar Serikuly
School of Pharmacy, Southwest University, Chongqing, China
avkalueff@gmail.com

Erik T. Alpyshov
School of Pharmacy, Southwest University, Chongqing, China
avkalueff@gmail.com

Dong Mei Wang
School of Pharmacy, Southwest University, Chongqing, China
avkalueff@gmail.com

Jing Tao Wang
School of Pharmacy, Southwest University, Chongqing, China
avkalueff@gmail.com

Allan V. Kalueff
School of Pharmacy, Southwest University, Chongqing, China
avkalueff@gmail.com

***Abstract* — Arecoline is a naturally occurring psychoactive alkaloid with partial agonism at nicotinic and muscarinic acetylcholine receptors. Arecoline consumption is widespread, making it the fourth (after alcohol, nicotine and caffeine) most used substance by humans. However, the mechanisms of acute and chronic action of arecoline remain poorly understood. Animal models are a valuable tool for CNS disease modeling and drug screening. Complementing rodent studies, the zebrafish (Danio rerio) emerges as a promising novel model organism for neuroscience research. In this study, we assessed the effects of acute and chronic arecoline on adult zebrafish behavior and neurophysiology. Acute and chronic arecoline treatments produce overt anxiolytic-like behavior without affecting general locomotor activity and whole-body cortisol levels. Acute arecoline at 10 mg/L disrupted shoaling, increased social preference, elevated brain norepinephrine and serotonin levels and reduced serotonin turnover. As it also upregulated early protooncogenes c-fos and c-jun, chronic arecoline increased brain expression of microglial biomarker genes egr-2 and ym1, implicating microglial neuroimmune mechanisms in potential effects of long-term arecoline use. These findings suggest that novel anxiolytic drugs can eventually be developed based on arecoline-like molecules, whose integrative mechanisms of CNS action may involve monoaminergic and neuro-immune modulation.**

Keywords* — *zebrafish; arecoline; CNS; anxiety; genomic effects.

ACKNOWLEDGMENT

The study was supported by Russian Science Foundation grant 19‑15‑00053 and budgetary funding for basic scientific research of the Scientific Research Institute of Physiology and Basic Medicine (theme No. AAAA-A16-116021010228-0).

DOI 10.18699/BGRS/SB-2020-171

Study of personal qualities and EEG activity in a stop signal paradigm in residents of northern regions

Tatiana Asakhova
NSU, Novosibirsk, Russia
t.astakhova@g.nsu.ru

Alexander Saprigyn
State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk, Russia
saprigyn@mail.ru

Sergey Tamozhnikov
State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk, Russia
stam@physiol.ru

Alexandra Karpova
North-Eastern Federal University, Yakutsk, Russia
karpova74@list.ru

Natalya Borisova
North-Eastern Federal University, Yakutsk, Russia
borinat@yandex.ru

Elena Afanaseva
North-Eastern Federal University, Yakutsk, Russia
eb.afanaseva@mail.ru

Alexander Savostyanov
State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk, Russia
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
alexander.savostyanov@gmail.com

***Abstract* — It is known that the adaptation to new conditions depends on personality traits. The results obtained showed that the lower level of anxiety and neurotism, the better behavioural control and adaptation. Also, the adaptation level is the lowest right after the move to another place, and it grows by the time along with the growth of behavioral control.**

Keywords — adaptation, personality traits, stop signal paradigm, SSP, EEG.

Aim

The aim of the study was to identify correlates of personality traits in the context of SSP in local people and migrants who live in polar or subpolar regions of the Republic of Sakha (Yakutia).

Methods

Four groups of young healthy people participated in the study: 1) people who permanently live in Novosibirsk (150 people, 70 men, average age 23 ± 4 y.o.); 2) people who permanently live in Yakutsk (85 people, 39 men, 24 ± 3 y.o.); 3) people who permanently live in the Arctic regions of Yakutia (50 people, 21 men, 26 ± 4 y.o.); 4) labor migrants who moved to Yakutia from the southern countries (50 people, all men, 23 ± 2 y.o.). The examination of group of migrants was carried out twice - immediately after the move and six months after. Stop signal paradigm methodology was chosen in order to observe the brain activity in conditions of behavior control. In SSP experiment participants had to quickly press one of the two buttons after the target signal onset (130 trials), or suppress the movement that had already begun (33 trials). EEG was recorded by 64 channels amplifier - Brain Products, Germany. Event related potentials (ERP) were used as a functional changes indicator of tasks performance. Markers of the "Big Five" (D. Goldberg's questionnaire) were considered as personality characteristics indicators.

Results

The questionnaires results analysis showed differences between four groups of participants in terms of neurotism, personal anxiety, extraversion and compliance. Behavioral indicators of motor control were the best in local people of Yakutia (both groups), average for Novosibirsk and the worst in migrants at the first examination. The quality of motor control increased in the second examination. The quality of motor control positively correlated with indices of neuroticism and anxiety among the indigenous inhabitants of Yakutia (both groups), negatively correlated among migrants and didnot correlate with neuroticism among residents of Novosibirsk. EEG in the group of local people of Yakutia positively correlated between the P300 amplitude, the potential of readiness and the level of neurotism and anxiety, while migrants showed negative correlations of these indicators, while reliable correlation was not found in the participants from Novosibirsk.

Conclusion

Acute stress during adaptation to the north climate among migrants caused a decrease in the ability to control their behavior. It also resulted an increase in the neuroticism and anxiety. In contrast, in local Yakuts the increased neurotism and anxiety were associated with the increase in the ability to control behavior.

Acknowledgment

The study was supported by RFFI grants No. 18-29-13027 and 18-415-140021.

DOI 10.18699/BGRS/SB-2020-172

Monoamine signaling gene networks unraveled in mouse social stress model

Vladimir Babenko
Laboratory of Neuropathology Modeling, Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
bob@bionet.nsc.ru

Natalia Kudryavtseva
Laboratory of Neuropathology Mediling Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
natnik@bionet.nsc.ru

Abstract — **Synapse signaling manifests the key mechanism of signal transduction across and within brain regions. The advent of RNA-Seq protocol yielded a unique chance to observe stable states of brain region interaction modes based on synapse genes expression profiles. We proposed the subsets of synapse - related genes for particular neurotransmitter types, in particular, monoamines (dopamine and serotonin) and glutamate/gaba. These include 3 types: a) vesicular monoamine transporters; b) monoamine transporters; c) synapse receptors.**

Upon considering 3 brain regions (dorsal striatum (STR), ventral tegmental area (VTA), and midbrain raphe nuclei (MRN)) we elucidated the specific physiological states of 9 species in the animal stress model published elsewhere (Kudryavtseva et al., 2014) [1].

Keywords — animal stress model, aggression, depression, dorsal striatum, dopamine, serotonin.

Introduction

The phenomenon of depression and addiction are currently one of the major sources of human disabilities in the world. Their physiology is tightly linked to the limbic system of the striatal system. In particular, addiction was shown to affect the Nucleus accumbens (NAcc) region along with dorsal striatum (STR) in mouse, and caudate/putamen regions in human, also comprising the striatum region. The dorsal striatum maintains supervision of motor activity and stereotypical behaviors. It is involved in cognitive, reward and social hierarchy manifestations. STR exemplifies the initiation and execution of movements via regulation of muscle tone, so it is one of the central brain regions to manifest overall behavior patterns.

The major body of STR cells is Medium Spiny Neurons (MSNs), which account for 95% of neural cells in STR. Previously, we showed that STR exhibits specific dopamine – glutamate alternative signaling [2]. In this work we performed RNA-Seq based analysis in conjunction of other brain regions upon STR, namely, Ventral TGegmental Area (VTA) and Midbrain Raphe Nuclei (MRN), the major sources of dopamine and serotonin, correspondingly.

Research

We compiled the synaptic related genes set to assess the signaling VTA<->STR, MRN<->STR. Next we performed Principal Component Analysis on the corresponding expression profiles in these regions. We elucidated that:

a) Dopamine intake is the highest in aggressive species, and lowest in depressive ones.
b) Dopamine expression in VTA is the highest in certain depressive species, but it's essentially re-uptaken back to VTA due to the low D1/D2 receptors density in the depressive STR MSNs.
c) Depressive species uptake serotonin preferably via VTA, while MRN serotonin expression is turned down in their dorsal striatum.

Conclusion

By means of adapted gene subset we can precisely interpret the signal states in synapses VTA->STR, MRN ->STR based on the gene expression profiles. Notably, we identified at least 2 distinct depression – like states in our samples which are featured with distinct gene expression profiles for a range of pathways. It may imply that for certain cases selective serotonin reuptake inhibitor therapy may be non-adequate due to a different etiology of the pathology,

Acknowledgment

This work is supported by Russian Science Foundation (grant No. 19-15-00026, to NNK)

References

[1] Kudryavtseva N.N., Smagin D.A., Kovalenko I.L., Vishnivetskaya G.B. (2014) Repeated positive fighting experience in male inbred mice. Nature Protocol. 9, 11, 2705 - 2717.

[2] Babenko VN, Galyamina AN, Rogozin IB, Smagin DM, Kudryavtseva NN Dopamine response gene pathways in dorsal striatum MSNs from a gene expression viewpoint: cAMP-mediated gene networks. BMC Neroscience, accepted.

DOI 10.18699/BGRS/SB-2020-173

Expression of autophagy genes and markers of inflammation in the brain in a transgenic mouse model of Parkinson's disease

Victor M. Belichenko
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
belichenko@physiol.ru

Alexandra B. Shintyapina
Federal Research Center for Basic and Translational Medicine
Novosibirsk, Russia
Alexandra.Shintyapina@mail.ru

Larisa A. Fedoseeva
Federal Research Center "Institute of Cytology and Genetics"
Novosibirsk, Russia
fedoseeva@bionet.nsc.ru

Anna A. Akopyan
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
annaaleksanovna@mail.ru

Tatiana A. Korolenko
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
t.a.korolenko@physiol.ru

Tamara G. Amstislavskaya
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
Amstislavskaya@yandex.ru

Maria A. Tikhonova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
tikhonovama@physiol.ru

Abstract — **The autophagy-lysosomal pathway is one of the main mechanisms for cleaning of brain tissue from "pathological" proteins that contribute to the neurodegenerative disorders including Parkinson's disease (PD) [1]. Levels of the regulatory factor Beclin1 (encoded by gene *Becn1*) and protein LC3-II are markers of autophagy activation. Cystatin C (encoded by gene *Cst3*) have a wide range of biological effects [2] including the autophagy induction through mTOR inhibition and regulation of inflammation, a marker of which are chitinases (chitoriosidase encoded by gene *Chit1*, and acidic-1 chitinase encoded by gene *Chia1*) upon activation of microglial cells. The aim of the work was to study gene (*Becn1, Cst3, Chia1, Chit1*) and protein (LC3-II) expression in the brain in a transgenic mouse PD model overexpressing the A53T-mutant alpha-synuclein at an early stage of the pathology progression (at the age of 5 months). The levels of gene expression were determined by qPCR. Protein expression was estimated with immunohistochemical analysis in frozen brain sections. Autophagy activity was suppressed in the striatum and amygdala (according to LC3-II marker) and in the frontal cortex (according to mRNA level of *Becn1*) in the transgenic mice compared to wild-type control. No changes were found in the expression levels of chitinases, while the expression of *Cst3* was reduced in the striatum and amygdala of transgenic mice that was associated with a reduced autophagy in these structures. This deficit in the cellular defense response may be a target for effective neuroprotection through therapeutic activation of autophagy at early stages of PD.**

Keywords *— **autophagy, Cystatin C, alpha-synuclein, mice, chitinase, microglia***

ACKNOWLEDGMENT

The study was supported by budgetary funding for basic scientific research of the Scientific Research Institute of Physiology and Basic Medicine (theme No. AAAA-A16-116021010228-0) and the Russian Foundation for Basic Research (grant No. 15-54-52029_HHC-a).

REFERENCES

[1] M. Deleidi and W. Maetzler, "Protein clearance mechanisms of alpha-synuclein and amyloid-β in Lewy body disorders," Int. J. Alzheimers. Dis., vol. 2012, ID 391438, pp. 1-9, 2012.

[2] B. Tizon, S. Sahoo, H. Yu, S. Gauthier, A. R. Kumar, P. Mohan, M. Figliola, M. Pawlik, A. Grubb, Y. Uchiyama, U. Bandyopadhyay, A. M. Cuervo, R. A. Nixon, and E. Levy, "Induction of autophagy by cystatin C: a mechanism that protects murine primary cortical neurons and neuronal cell lines," PLoS One, vol. 5, ID e9819, pp. 1-12, 2010.

DOI 10.18699/BGRS/SB-2020-174

EEG correlates of strategies of emotional regulation during perception of emotional information

Andrey V. Bocharov
Laboratory of Differential Psychophysiology, Scientific Research Institute of Physiology and Basic Medicine, Novosibirsk, Russia
Humanitarian Institute, Novosibirsk State University, Novosibirsk, Russia

Sergey S. Tamozhnikov
Laboratory of Differential Psychophysiology, Scientific Research Institute of Physiology and Basic Medicine, Novosibirsk, Russia

Ekaterina A. Proshina
Laboratory of Differential Psychophysiology, Scientific Research Institute of Physiology and Basic Medicine, Novosibirsk, Russia

Alexander N. Savostyanov
Laboratory of Differential Psychophysiology, Scientific Research Institute of Physiology and Basic Medicine, Novosibirsk, Russia, Humanitarian Institute, Novosibirsk State University, Novosibirsk, Russia
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Ekaterina A. Merkulova
Laboratory of Differential Psychophysiology, Scientific Research Institute of Physiology and Basic Medicine, Novosibirsk, Russia

Gennady G. Knyazev
Laboratory of Differential Psychophysiology, Scientific Research Institute of Physiology and Basic Medicine, Novosibirsk, Russia

Alexander E. Saprigyn
Laboratory of Differential Psychophysiology, Scientific Research Institute of Physiology and Basic Medicine, Novosibirsk, Russia

Abstract — **The aim of research was to study the relation between oscillatory dynamics during the perception of negative, positive and neutral sentences and emotional facial expressions and strategies of perception and processing emotional information.**

Keywords — EEG, reappraisal, suppression, emotional sentences, emotional facial expressions

72 subjects (43 women) took part in EEG study and completed J. Gross questionnaire (2003). In task of the searching of syntax errors, the participants were needed to determine whether the sentence contains a syntax error or not in three types of emotional colored sentences (negative, neutral and positive).

Four types of facial emotional expressions (angry, neutral, sad, and happy) were used. EEG recording was performed using a cap with 127 electrodes. To measure the induced responses, event-related spectral perturbations (ERSP) were calculated.

"Reappraisal emotions" scale positively correlated with an increase in alpha spectral power in the first 200 ms after the presentation of sad facial expressions. According to the Klimesch hypothesis (2011), early alpha oscillations in response to visual stimuli may be associated with inhibition, which is effective during initial access to a complex knowledge system. It could be suggested that an early increase in alpha oscillations in subjects who prefer a strategy of cognitive reappraisal of emotions could reflect a preparation for the perception of emotional information.

Reappraisal emotions positively correlated with an increase in theta rhythm in the time interval from 200 to 400 ms after presentation of negative sentences which described symptoms of depression and after presentation of angry faces. It could be assumed that people who prefer the strategy of Reappraisal emotions, during perception of negative information, pay more attention to the emotional content of stimuli. Emotion suppression scale was positively correlated with ERSP of beta range after presentation of positive sentences which contained syntax errors. It could be related with a decrease of top-down control processes and a decrease of semantic processes during the search of syntax errors in positive sentences.

ACKNOWLEDGMENT

The study was supported by the Russian Foundation for Basic Research (RFBR) under Grants No. 18-313-00174, No. 18-00-00939, No. 20-013-00404, No. 18-29-13027 and partially supported by the federal budget for basic scientific research theme No. AAAA-A16-116021010228-0.

DOI 10.18699/BGRS/SB-2020-175

Dynamic regulation of murine cortical transcriptome by early-life stress: Impairment of myelination and cognitive functions

Natalya Bondar
ICG SB RAS, Novosibirsk, Russia
nbondar@bionet.nsc.ru

Anastasia Shulyupova
ICG SB RAS, Novosibirsk, Russia
shulyupova@bionet.nsc.ru

Polina Kisaretova
ICG SB RAS, Novosibirsk, Russia
kisaretova@bionet.nsc.ru

Nikita Ershov
ICG SB RAS, Novosibirsk, Russia
ershov@bionet.nsc.ru

Elena Antontseva
ICG SB RAS, Novosibirsk, Russia
antontseva@bionet.nsc.ru

Tatiana Merkulova
ICG SB RAS, Novosibirsk, Russia
merkulova@bionet.nsc.ru

***Abstract* — Stress early in life negatively affects the formation of the brain regions. We explored the effects of two types of stress in mice: prolonged repeated separation of pups from their mothers for 3h per day during the first 2 weeks of life (maternal separation, MS), or 24h single maternal separation on 9th day of life (maternal deprivation, MD). RNA-seq analysis revealed weak transcriptome changes in prefrontal cortex of 15-day pups but at the same time had strong delayed consequences on gene expression in adult mice. The observed changes were mainly related to genes associated with myelin sheath development. The direction of the alterations in 15-day pups and adult mice were mainly opposite.**

Keywords — early life stress, RNA-seq, myelin, mice.

Motivation and Aim

Stressful events in early postnatal period have critical consequences for the individuals' life and can increase later risk for the development of psychiatric disorders. The maturation of brain structures continues in the postnatal period in both humans and animals. Early-life stress negatively affects the formation of the brain regions taking part in the implementation of emotion, memory and cognition. Various animal models have been developed to investigate the molecular consequences of early life stressful events. Among them, maternal separation models in rodents are the most commonly used and established ones [1]. Prolonged maternal separation causes significant amounts of stress resulting in negative long-lasting changes of emotion-related behavior, stress reactivity and cognitive functions in rodents [2-5].

The aim of our study was the investigation of the transcriptome profile in the prefrontal cortex both immediately after prolonged maternal separation and through a long period of comfortable life. Comparison of immediate and delayed effects on transcription should contribute significantly to our knowledge of a link between early life exposures and neurobiological and behavioral outcomes in adulthood. Given the fact that the prefrontal cortex is responsible for complex cognitive functions, regulation of emotion, and adaptation to stress, we expected to reveal some alterations in gene expression in this brain region reflecting the long-lasting consequences of stress early in life.

Methods

We used two types of early life stress: prolonged repeated separation of pups from their mothers for 3h per day during the first 2 weeks of life (maternal separation, MS), or 24h single maternal separation on 9th day of life (maternal deprivation, MD). We analyzed the effects of stress on transcriptome of the prefrontal cortex of 15-days pups (at 24h after the last stress exposure) and 3 months old adult mice. To validate the RNA-seq results, the expression of selected up- or down-regulated genes were confirmed by real-time PCR.

Results

We revealed the weak effect of stress on the expression patterns of genes in 15-days pups: only 6 genes were differentially expressed in MS group to compare with controls. Stress early in life resulted in changes in expression of genes whose proteins connected with myelin sheath development (*Mag*, *Mobp*, *Mog*, *Mal*, *Plp1* and *Ugt8a,* all down-regulated). Despite weak changes in pups, transcriptome analysis of the prefrontal cortex in adult mice showed a number of differentially expressed genes in MS and MD groups compared to controls (648 genes in MS group and 192 genes in MD group). In MS group the changes mainly occur in systems associated with the development of the myelin sheath, regulation of trans-synaptic signaling and regulation of membrane potential. All the genes that were changed in prefrontal cortex of 15-days pups also remained altered in adults, but in the opposite direction (Fig. 1). Genes, which were differentially expressed in MD group to compare with controls, relate to different systems, without significant overrepresentation of any one.

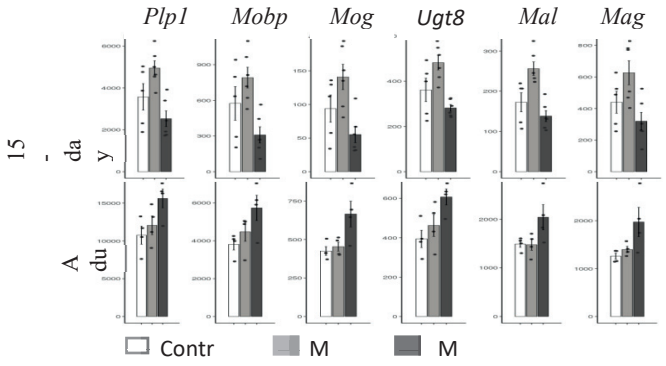


Fig. 1. Level of expression *Plp1*, *Mobp*, *Mog*, *Ugt8a, Mal* and *Mag* genes in adult mice and 15-day pups with history of early-life stress. Data represent as normalized counts.

The obtained results suggest the existence of delayed long-term expression changes in the prefrontal cortex triggered by early postnatal stress. We can suggest that disturbances in myelination can underlie enhanced susceptibility to mental illness.

ACKNOWLEDGMENT

The work was supported by Russian Science Foundation project 16-15-10131.

REFERENCES

[1] Pryce C.R., Feldon J. Long-term neurobehavioural impact of the postnatal environment in rats: manipulations, effects and mediating mechanisms. Neurosci Biobehav Rev. 2003. 27(1-2): 57-71.

[2] Bondar N.P., Lepeshko A.A., Reshetnikov V.V. Effects of early-life stress on social and anxiety-like behaviors in adult mice: sex-specific effects. Behav Neurol. 2018. 2018: 1538931.

[3] Di Segni M., Andolina D., Ventura R. Long-term effects of early environment on the brain: Lesson from rodent models. Semin Cell Dev Biol. 2018. 77: 81-92.

[4] Franklin T.B., Saab B.J., Mansuy I.M. Neural mechanisms of stress resilience and vulnerability. Neuron. 2012. 75(5): 747-761.

[5] Reshetnikov V.V., Studenikina A.A., Ryabushkina J.A., Merkulova T.I., Bondar N.P. The impact of early-life stress on the expression of HPA-associated genes in the adult murine brain. Behaviour. 2018. 155(2-3): 181 – 203.

DOI 10.18699/BGRS/SB-2020-176

Compulsive-like behaviors in *DISC1*-mice

Nadezhda Chizhova
ICG SB RAS, Novosibirsk, Russia
chnadezhda1995@gmail.com

Kristina Smirnova
NSU, Novosibirsk, Russia
PhBMRI, Novosibirsk, Russia
vedelina@mail.ru

***Abstract*— The *DISC1* gene is associated with the development of mental disorders in humans. Currently a number of genetic models of psychopathologies with mutations in the gene have been created, including genetic lines of mice with point mutations *DISC1*-Q31L$^{-/-}$ (model of depression) and *DISC1*-L100P$^{-/-}$ (model of schizophrenia), which are available at the unique scientific installation "Biological collection – genetic biomodels of neuropsychiatric diseases" (State Scientific-Research Institute of Physiology and Basic Medicine, Novosibirsk). In additional, *DISC1*-Q31L$^{-/-}$ mice showed compulsive-like behavior, and in this connection was proposed to study this type of behavior in heterozygotes and diheterozygotes for these mutations. The work performed allows us to conclude that the presence of only one mutant allele (both *DISC1*-Q31L and *DISC1*-L100P) is not sufficient for the expression of compulsive-like behavior in mice. However, mice that combine both mutations are predisposed to this behavior and exhibit it depending on their sex and mother's genotype: this behaviour characterise males whose mothers were homozygous for *DISC1*-Q31L mutation, and females whose mothers were homozygous for *DISC1*-L100P mutation. The conclusion about the influence of homozygosity and, accordingly, the behavior of mothers on the compulsive-like behavior of females and males can be verified in further experiments.**

Keywords*— *obsessive-compulsive behavior; models of psychopathology; behavioral testing; DISC1; mice.

Motivation and Aim

The *DISC1* gene (*Disrapted-In-Schizophrenia-1*) is a multifunctional scaffold protein that regulates many aspects of the development of the nervous system and is involved in signaling and associated with a number of human psychopathologies [1]. Today several lines of model animals with disorders in the *DISC1* have been created. Of particular interest are mice with point mutations in the gene: Q31L, which model a depressive-like state, and L100P, a universally recognized genetic model of schizophrenia [2–4]. Interestingly, these mutations fall into the region of deletions of the second and third exons studied in 2018. It was shown that this deletion led to an increase in performance in the test for instillation of marbles, which is one of the most common tests for obsessive-compulsive-like behavior [5,6]. Our colleagues showed that homozygous males with the Q31L mutation, but not the L100P, bury more balls in the marble burying test, which is a sign of compulsive behavior in these mice [7]. In this regard, it seems promising to study this behavior on heterozygous mice *DISC1*-Q31L$^{+/-}$, *DISC1*-L100P$^{+/-}$ and mice combining both mutations (taking into account the potential influence of the genotype and, accordingly, the behavior of a homozygous mother). The aim of this study was to evaluate the compulsive behavior in *DISC1*-Q31L$^{+/-}$, *DISC1*-L100P$^{+/-}$ and *DISC1*-L100P$^{+/-}$/Q31L$^{+/-}$ mice, as well as to assess the contribution of the maternal genotype.

Materials and methods

In this work we studied the behavior of mice in 10 experimental groups: mice of two sexes and four genotypes (*DISC1*-L100P$^{+/-}$ (hereinafter referred to as "L100P-HET"), *DISC1*-Q31L$^{+/-}$ ("Q31L-HET"), *DISC1*-L100P$^{+/-}$/Q31L$^{+/-}$ - these animals were also divided into two groups depending on the maternal genotype: *DISC1*-L100P$^{-/-}$ ("L100P(f)/Q31L(m)") or *DISC1*-Q31L$^{-/-}$ ("Q31L(f)/L100P(m)"), line C57BL/6NCrl ("WT") provided by UNU "Biological collection - genetic biomodels of neuropsychic diseases" (No. 493387) PhBMRI [8]. Mothers for L100P-HET and Q31L-HET were WT females. All mice were kept in vivarium PhBMRI in plastic cages (OptiMice Biotech AS; 34×29×15 cm), with a light mode of 12:12 (light 6:00 am, dark 18:00 pm) at a temperature of about 23°C. Animals received food and water *ad libitum*.

Testing of the obtained experimental mice at the age of about 3 months was carried out between 9:00 and 16:00 hours. The test for instillation of balls was carried out in the home room according to the protocol [9]. The open field test was carried out in accordance with the description [10]; before this, the mice were placed in the experimental room for 30 min for habituation. All conditions for working with animals were observed in accordance with international standards (Council of the European Communities Directive 86/609/EES).

Results

ANOVA showed the effect of genotype-sex interaction [$F(4,62)=3,2469$; $p<0.05$] the number of fully buried marbles. *Post-hoc* analysis presented significant differences between the following groups (Fig. 1): males Q31L(f)/L100P(m) and females L100P(f)/Q31L(m) instilled more balls than WT mice of the same sex; males Q31L(f)/L100P(m) compared to females buried more balls, and males L100P(f)/Q31L(m), on the contrary, less; while males Q31L(f)/L100P(m) instilled more balls compared to males L100P(f)/Q31L(m)), but females Q31L(f)/L100P(m) burried them less than females L100P(f)/Q31L(m); females L100P(f)/Q31L(m) also instilled more balls than heterozygous females Q31L ($p<0.05$).

It is important to note that in this case the study of the behavior of the same groups of mice in open field showed no differences in horizontal activity ($p>0.05$). MANOVA showed the effect of sex [$F(1,109)=4,2086$; $p<0.05$] for vertical activity, however, *post-hoc* analysis revealed differences only between the males of Q31L(f)/L100P(m), which made more racks compared to heterozygous males of Q31L ($p<0.05$).

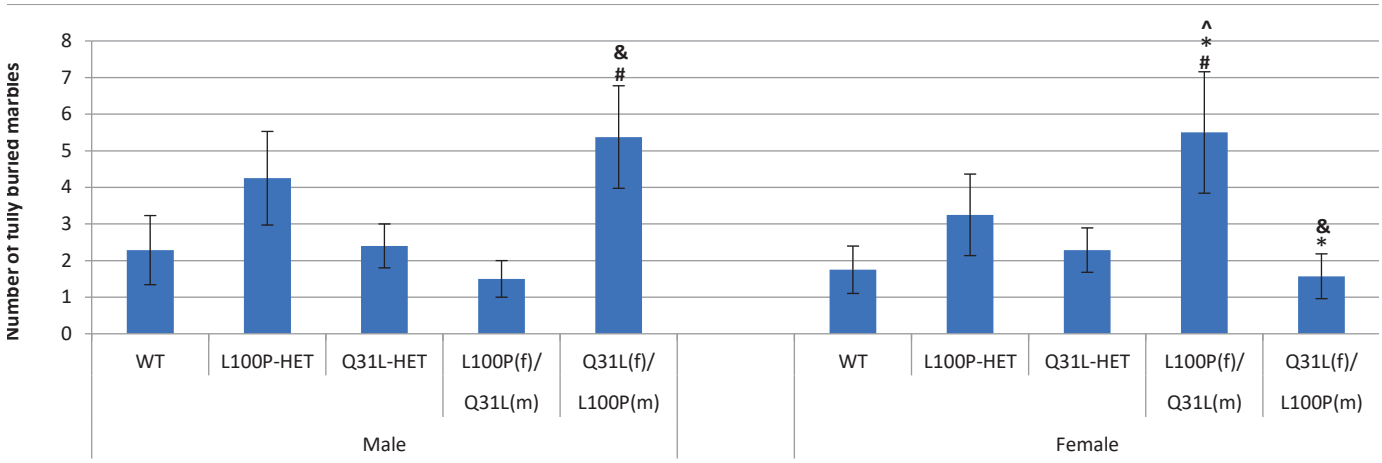


Fig. 1. The number of buried marbles in the test for compulsive behavior in *DISC1*-mutant mice (mean±SEM). * - $p<0.05$ vs same genotype males; # - $p<0.05$ vs same sex WT ; ^ - $p<0.05$ vs same sex Q31L-HET; & - $p<0.05$ vs same sex L100P(f)/Q31L(m).

Discussion

The lack of increased activity in the studied groups of mutant mice allows us to consider the results of the marble burying test to be correct for determining their compulsive-like behavior. Thus, the presence of a single mutation (both Q31L and L100P) in the heterozygous state is not sufficient for the manifestation of compulsive-like behavior. However, in mice with both of these mutations, the manifestation of this behavior depended both on the mother's genotype and on sex: compulsive behavior was characteristic of males whose mothers were homozygous for Q31L mutation, and females with mothers homozygous for L100P. It was shown that the combination of these two mutations in the *DISC1* gene seems to make these mice predisposed to the appearance of certain behavioral disorders, including sex-dependent [11]. Since in the earlier work of our colleagues [7] only males were used, and maternal individuals were homozygous for the corresponding mutation, comparing the results obtained in this work, the authors hypothesize that it is the homozygosity of mothers for the Q31L mutation that leads to a compulsive-like behavior as in males *DISC1*-L100P$^{+/-}$/Q31L$^{+/-}$, and in males *DISC1*-Q31L$^{-/-}$. In this case, females with two mutations are similarly affected by the homozygosity of mothers for the L100P mutation, but not Q31L. This conclusion can be confirmed or disproved in subsequent experiments, where (1) compulsive-like behavior was studied in females homozygous for these mutations and (2) the influence of the mother genotype on compulsive-like behavior in mice heterozygous for Q31L and L100P mutations separately.

Acknowledgment

The study was supported by budgetary funding for basic scientific research of the Scientific Research Institute of Physiology and Basic Medicine (theme No. AAAA-A16-116021010228-0).

References

[1] Brandon N.J., Millar J.K., Korth C., Sive H., Singh K.K., Sawa A. (2009) Understanding the role of DISC1 in psychiatric disease and during normal development. J. Neurosci. 29(41), 12768-12775.

[2] Clapcote S.J. et al. (2007) Behavioral phenotypes of *DISC1* missense mutations in mice. Neuron. 54 (3): 387–402.

[3] Lipina T.V. et al. (2010) Enhanced dopamine function in *DISC1*-L100P mutant mice: implications for schizophrenia. Genes Brain Behav. 9 (7): 777–89.

[4] Lipina T.V. et al. (2011) Genetic and pharmacological evidence for schizophrenia-related DISC1 interaction with GSK-3. Synapse. 65 (3): 234–48.

[5] Wulaer B. et al. (2018) Repetitive and compulsive-like behaviors lead to cognitive dysfunction in *DISC1*$^{\Delta 2\text{-}3/\Delta 2\text{-}3}$ mice. Genes, Brain and Behavior, 17(8), e12478.

[6] Alonso P., Lopez-Sola C., Real E., Segalas C., Menchon J. M. (2015) Animal models of obsessive–compulsive disorder: utility and limitations. Neuropsychiatr. Dis. Treat. 11: 1939–1955.

[7] Смирнова К.В, Морозова М.В., Татаурова С.А., Амстиславская Т.Г., Литвинова Е.А., Кожевникова Е.Н. Изучение влияния мутаций в гене *DISC1* у мышей линий L100P и Q31L на формирование поведения , подобного обсессивно-компульсивному. Unpublished.

[8] Петрова Е.С. и др. (2018) Поддержание генетически модифицированных линий мышей: вклад в развитие биоколлекций в России. ЛЖНИ. 2.

[9] Angoa-Pérez, M., Kane, M. J., Briggs, D. I., Francescutti, D. M., Kuhn, D. M. (2013) Marble burying and nestlet shredding as tests of repetitive, compulsive-like behaviors in mice. J. Vis. Exp. (82), e50978

[10] Seibenhener M. L., Wooten M. C. (2015) Use of the open field maze to measure locomotor and anxiety-like behavior in mice. JoVE. (96), e52434,

[11] Чижова Н.Д., Липина Т.В., Амстиславская Т.Г. (2019) Характеристика поведения гетерозиготных мышей с мутациями L100P и Q31L в гене *DISC1*. ЛЖНИ. 3.

DOI 10.18699/BGRS/SB-2020-177

Neurotransmitter gene network reconstruction and analisis

Ivanov R.A.
Sector of Bioinformatics and Information Technologies in Genetics
Institute of Cytology and Genetics
Novosibirsk, Russian Federation
ivanov.romanart@yandex.ru

Matushkin Y.G.
Laboratory of Molecular-Genetic Systems
Institute of Cytology and Genetics
Novosibirsk, Russian Federation
mat@bionet.nsc.ru

Klimenko A.I.
Sector of Bioinformatics and Information Technologies in Genetics
Kurchatov genomics center
Institute of Cytology and Genetics
Novosibirsk, Russian Federation
klimenko@bionet.nsc.ru

Vasiliev G.V.
Sector for Genomic Studies
Institute of Cytology and Genetics
Novosibirsk, Russian Federation
genn@bionet.nsc.ru

Savostyanov A.N.
Laboratory of Psychological Genetics, Institute of Cytology and Genetics;
Laboratory of Differential Psychophysiology, Institute of Physiology and Basic Medicine,
Novosibirsk, Russian Federation
savostyanov@bionet.nsc.ru

Lashin S.A.
Sector of Bioinformatics and Information Technologies in Genetics
Kurchatov genomics center
Institute of Cytology and Genetics
Novosibirsk, Russian Federation
lashin@bionet.nsc.ru

Abstract — **Research of biological bases of depression is one of the main interests of modern neurophysiology. It is still unknown how complex interaction between genes correlates to human depression behavior. Neurobiological understanding of these interactions could help to reveal the molecular mechanisms of depression. Insight on these interactions is crucial for development of diagnostic and treatment of depression. In our research, we reconstruct networks of protein-protein interaction and transcriptional regulation as well, in attempt to create the model of complex interaction between environmental and genetic risk factors of depression.**

Keywords — neurotransmitter, SNP, gene networks

Motivation and Aim

The main interest of modern neuroscience is aimed at studying the biological bases of individual differences in human behavior. The medical relevance of this topic is due to the fact that development of a number of neurological and psychiatric diseases, as well as the characteristics of their occurrence, are directly related to the psychological characteristics of individuals [1]. Many papers on the mechanisms of regulation of individual behavior indicate the important role of neurotransmitter systems in the mechanisms of its regulation [2]. Also recent studies suggest that genetic factors may be modulated by environmental factors [3] which makes it important to study the environmental influences on genetic factors of human behavior. In our research we chose to focus on depression and mental diseases as the extreme manifestations of human behavior. Today gene network approaches are powerful tools for studying genes associated with human traits. In our research, we are trying to identify genes, which expression differs in different populations, based on analysis of sequencing data from several human populations.

Methods

At first, we formed the set of genes associated with neurological diseases for which the difference in several human population was shown based on an analysis of number of specialized databases, namely SFARI Gene, BDgene. PsyGeNET, epiGAD, MK4MDD and HGNC. From these set of genes, we have reconstruct networks of protein interaction using web-service StringDB. Later on we conduct the Gene Ontology analysis of our network by DAVID GO web-service [4] to select list of gene associated with neurological disease. Based on our gene network we also allocated transcriptional regulation factors using MoLoTool web-service [5] (we used significance threshold $p_{ajusted}$ <10-4, which is recommended by the web-service) and reconstruct networks of transcriptional regulation. After that we build the library for targeting sequencing based on Roche's SeqCap EZ kit and conducted Illumina sequencing experiment on more than 500 people from different populations. Obtained data was analyzed via tools of bcftools to receive the list of population-related SNP for further analysis and comparison with phenotypical data.

Results

Using the methods mentioned earlier, we identified a set of genes associated with anxiety behavior and set of genes associated with mental disorders for which the difference in phenotype manifestation among different human populations was shown. Based on this sets of genes, the gene networks were reconstructed. GO analysis reveals 165 genes that were conducted in DNAseq experiment for further verification of our model. Transcriptional regulation networks revealed that our network contains 3 TF: TP53, SREBF2, SREBF1. From the DNAseq experiment on more than 500 people from various populations, we received SNP data unique for each population and soon we will conduct complex analysis of these SNP with phenotypical data of EEG and psychological questionnaires.

Acknowledgment

The research was partially funded by the RFBR Grant 18-29-13027 and the Budget Project 0259-2019-0008.

REFERENCES

[1] G.G. Knyazev et al. "Resting state connectivity mediates the relationship between collectivism and social cognition," Int. J. Psychophysiol., vol. 123, no. November, pp. 17–24, 2018.

[2] R.P. Ebstein, "The molecular genetic architecture of human personality: Beyond self-report questionnaires," *Mol. Psychiatry*, vol. 11, no. 5, pp. 427–445, 2006.

[3] A. N. Savostyanov et al. "Association of anxiety level with polymorphic variants of serotonin transporter gene in Russians and Tuvinians," *Russ. J. Genet. Appl. Res.*, vol. 5, no. 6, pp. 656–665, 2015.

[4] D. W. Huang et al. "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," no. 2, 2008.

[5] I. V. Kulakovskiy et al., "HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis," Nucleic Acids Res., vol. 46, no. D1, pp. D252–D259, 2018.

DOI 10.18699/BGRS/SB-2020-178

Role of inflammatory and synaptic plasticity genes in individual differences in cognitive abilities

Anastasiya Kazantseva
Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of RAS, Ufa, Russia
Kazantsa@mail.ru

Renata Enikeeva
Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of RAS, Ufa, Russia
Enikeevarf@gmail.com

Yuliya Davydova
Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of RAS, Ufa, Russia
Julia.dmitrievna@list.ru

Zalina Takhirova
Russian Academy of Education, Moscow, Russia
Tahirovazalina@mail.ru

Rustam Mustafin
Bashkir State Medical University, Ufa, Russia
Ruji79@mail.ru

Sergey Malykh
Psychological Institute, Russian Academy of Education, Moscow, Russia
Malykhsb@mail.ru

Marina Lobaskova
Psychological Institute, Russian Academy of Education, Moscow, Russia
Lobaskovamm@mail.ru

Tatiana Tikhomirova
Russian Academy of Education, Moscow, Russia
Tikho@mail.ru

Elza Khusnutdinova
Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of RAS, Ufa, Russia
Elzakh@mail.ru

Abstract — **The mechanisms underlying individual cognitive functioning are complex, and genetic and epigenetic factors play a significant role. Considering the possible involvement of inflammatory pathways and synaptic plasticity in cognitive functioning, the present study aimed to estimate the main effect of *BDNF*, *IL1B*, *CRP*, and *TNF* gene polymorphisms together with gene-environment interactions in manifestation of individual differences in several cognitive abilities in young adults. Multiple linear regression conducted on the data obtained from 897 mentally healthy young adults (79% women; 19.74±1.51 years) of Caucasian origin revealed the association of *TNF* rs1041981 A-allele and ability of 3D mental rotation, and general intelligence (factor "g"). Moreover, significant interactive effects of *TNF* rs1041981 and *CRP* rs3093077 and several peculiarities of childhood rearing on cognitive functioning were demonstrated.**

Keywords* — *inflammation, cytokines, neurotrophic factor, gene association, cognitive abilities, number sense, general cognitive ability, working memory.

Motivation and Aim

Motivation

The study of the productivity of cognitive functions (the level of general intelligence, memory, spatial, mathematical, linguistic, etc. abilities) as an integral part of the individual personal potential is becoming increasingly relevant today, since the level of cognitive functioning is the basis of life success and individual self-realization. In particular, spatial ability, "number sense", general intelligence are predictors of success in life and professional activities, for example, in careers in science, technology, engineering and mathematics (STEM), as well as chemistry, physics, anatomy, psychology [1].

It is known that the mechanisms underlying individual predisposition to individual cognitive functioning are complex, and genetic and epigenetic factors play a significant role. Therefore, the manifestation of cognitive abilities is a very complex process, affecting the fine regulation of gene expression through epigenetic changes, as well as the modulation of the immune, endocrine and nervous systems through the involvement of metabolites in the functioning of brain systems. Together with the involvement of synaptic plasticity genes (including brain derived neurotrophic factor, *BDNF*) in cognitive functioning [2], the pathways involved in immune function and inflammatory response, such as those responsible for cytokine and especially interleukin signalling while studying cognitive abilities/deficit are of great interest [3]. Namely, proinflammatory cytokines including interleukin 1 beta (IL1B) [4], C-reactive protein (CRP), tumor necrosis factor (TNF)) have been implicated in the pathogenesis of various psychiatric disorders; however, their role in cognitive functioning in mentally healthy individuals remains unsufficient. Multiple environmental factors may modulate gene-based inflammatory response via changes in individual epigenetic profile, thus resulting in individual differences in cognitive abilities.

Aim

In accordance with "Universalist genes hypothesis" and considering a role of proinflammatory cytokines and BDNF in cognitive impairment, the present study aimed to estimate the main effect of *BDNF*, *IL1B*, *CRP*, and *TNF* gene polymorphisms together with gene-environment interactions in manifestation of individual differences in several cognitive abilities in young adults without cognitive deficit.

Methods

The study included 897 mentally healthy individuals (79% women; 19.74±1.51 years) of Caucasian origin (428 Russians, 200 Tatars, 117 Udmurts, and 152 of mixed ethnicity) from Russia. The assessment of cognitive abilities ("number sense", 3D mental rotation, general cognitive intelligence assessed with Raven's progressive matrices, Corsi block-tapping test for working memory) was conducted under the Battery of cognitive tests developed at International Laboratory for Interdisciplinary Investigations into Individual Differences in Learning (InLab) (Department of Psychology, Goldsmiths, University of London) and validated at the Laboratory of Age Psychogenetics (Psychological Institute, Russian Academy of Education, Moscow). The following environmental parameters, which were previously reported to affect cognitive functioning, were estimated: birth order, sibship size, place of residence, preterm birth, socio-economic status, rearing in a complete/incomplete family, childhood adversity, rearing in a bilingual family, smoking, physical activity, weight at birth, maternal and paternal age at birth. The genotyping of *BDNF* (rs6265), *IL1B* (rs16944), *CRP*

(rs3093077), and *TNF* (rs1041981) gene SNPs was performed using PCR-based KASP genotyping technology on "CFX96" DNA Analyzer (BioRad, USA). Statistical analysis included multiple linear regression followed by FDR-correction for multiple testing (PLINK v.1.09). Genotypes and 13 environmental parameters served as independent factors and cognitive abilities as dependent variable.

Results

Statistical analysis revealed an association of *TNF* rs1041981 A-allele with increased ability of 3D mental rotation in total sample with sex and ethnicity inclusion as covariates (β=1.61; P=0.003; P_{FDR}=0.017) and among women (β=1.57; P=0.011; P_{FDR}=0.043). Interestingly, subsequent analysis of gene-by-environment interactions (GxE) demonstrated that place of residence (urban/rural status) modulated the association of *TNF* rs1041981 A-allele and the ability of 3D mental rotation (β=-2.78; P=0.020).

With respect to general intelligence (factor "g") we observed a sex-specific association of *TNF* rs1041981 A-allele with increased score on factor "g" only among women (β=2.43; P=0.011; P_{FDR}=0.045), while ethnicity-specific association was shown for *IL1B* rs16944 A-allele and lower factor "g" in Tatars (β=-5.57; P=0.004; P_{FDR}=0.019). This finding seems to be congruent to those reported in elderly individuals without dementia [5]. It should be noted that GxE analysis demonstrated a modulating effect of bilingual rearing on the association of *TNF* rs1041981 A-allele and general intelligence (β=5.33; P=0.015). Recently we also demonstrated sex- and ethnicity specific effect of synaptic plasticity genes polymorphisms in variation in such cognitive trait as mathematical anxiety [6].

The GxE analysis demonstrated that paternal age at individual's birth (β=1.15; P=0.008) and two-child family (sibship size=2) (β=-10.96; P=0.007) modulated the association of *CRP* rs3093077 G-allele with "number sense", which reflects the ability to correctly place the number on the number-line. Recently, this SNP was proved in meta-analysis of depression [7], which evidence in the similar brain networks involved in depression and cognitive development.

In conclusion, the present preliminary study for the first time provides evidence that the genes involved in inflammatory pathways (*TNF*, *CRP*, and *IL1B*) may contribute to individual differences in such cognitive abilities as 3D mental rotation, general intelligence, and "number sense" in mentally healthy young adults. Moreover, our findings indicate significant interactive effects of examined gene polymorphisms and such environmental factors as place of residence, bilingual rearing, sibship size and paternal age at birth affecting cognitive functioning.

ACKNOWLEDGMENT

The present study was supported by the Russian Science Foundation (project No. 17-78-30028).

REFERENCES

[1] Nagy-Kondor R. (2017) Spatial ability of engineering students. Annales Mathematicae et Informaticae. 34: 113-122.

[2] Toh Y.L. et al. (2018) Impact of brain-derived neurotrophic factor genetic polymorphism on cognition: A systematic review. Brain Behav. 8(7): e01009.

[3] McCall M.K. et al. (2018) Symptom Science: Omics Supports Common Biological Underpinnings Across Symptoms. Biol. Res. Nurs. 20(2): 183-191.

[4] Tsai S.J. (2017) Effects of interleukin-1beta polymorphisms on brain function and behavior in healthy and psychiatric disease conditions. Cytokine Growth Factor Rev. 37: 89-97.

[5] Sasayama D. et al. (2011) Association of interleukin-1β genetic polymorphisms with cognitive performance in elderly females without dementia. J. Hum. Genet. 56(8): 613-616.

[6] Kazantseva A. et al. (2020) Stress-associated cognitive functioning is controlled by variations in synaptic plasticity genes. Russian Journal of Genetics. 56 (1): 88–95.

[7] Cohen-Woods S. et al. (2018) Interaction between childhood maltreatment on immunogenetic risk in depression: Discovery and replication in clinical case-control samples. Brain Behav. Immun. 67: 203-210.

DOI 10.18699/BGRS/SB-2020-179

Electroencephalographic correlates of an insight

G.G. Knyazev
State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk

A. V. Bocharov
State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk
Novosibirsk National Research State University, Novosibirsk

A. N. Savostyanov
State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk
Novosibirsk National Research State University, Novosibirsk

***Abstract* — It is well known, the conditional rest state activates the functional rest network, including the brain default system (DMN) and the central executive network (CEN). At this stage of research, it can be assumed that both the process of adaptation to the polar climate and new social conditions, and the gradual reduction of the risk of depression among migrants, are due to functional changes within these two brain systems, as well as their functional relationships among themselves. Hypothetically, the increased activity of CEN is associated with a higher level of stress associated with moving to new living conditions**

Keywords* — *EEG, depression, DMN, CEN, migrants, delta-rhythm, gamma-rhythm

Insight is a quick, sudden penetration of the solution into the essence of the problem, which may be accompanied by a feeling of "insight", or " A-ha experience." The study of insights in the laboratory was fraught with significant difficulties and the results obtained were quite contradictory.

In current study we aimed to identify the dynamics of brain activity in the period, preceding and accompanying the occurrence of insight using the distant association test (TOA) and EEG analysis at the level of sources of electrical activity. A comparison of the options for solving problems that were classified as “an insight” by subjects and those rated as “analytical” revealed a number of differences both at the behavioral level and at the level of brain activity.

Insight solutions were found averagely faster and were mostly more precise. Brain activity in solving problems in an analytical way did not significantly differ from brain activity while no solution was found, which allowed to suggest the presence of an analytical method in the latter case. The brain activity during problems solving using insights was significantly different from unsuccessful trails, and different from trails where an analytical strategy was used.

The early stages of problem solution search showed the differences in the delta range in the cortical areas related to the default system and perfomance control networks. Brain activity, probably associated with the occurrence of an insight solution in the mind, was detected in beta and gamma ranges in the cortical areas of the left hemisphere, which were involved in semantic processes and emotional responses. All in all, the revealed effects showed distinct differences in the dynamics of brain activity with insight and analytical options for finding a solution and testified prospects of this research area.

ACKNOWLEDGMENT

This work was supported by a grant from the Russian Foundation for Basic Research (RFFI) KOMFI No. 18-00-00939 and No. 18-00-00569.

DOI 10.18699/BGRS/SB-2020-408

NRG1, PIP4K2A, and HTR2C contain possible genetic biomarkers of several clinical subphenotypes of depression and bipolar disorder

Anastasia Levchenko
Theodosius Dobzhansky Center for Genome Bioinformatics, Saint Petersburg State University
Saint Petersburg, Russia
a.levchenko@spbu.ru
ORCID 0000-0003-3509-1304

Ivan V. Pozhidaev
Mental Health Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences
Tomsk, Russia
craig1408@yandex.ru

Natalia M. Vyalova
Mental Health Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences
Tomsk, Russia
natarakitina@yandex.ru

German G. Simutkin
Mental Health Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences
Tomsk, Russia
ggsimutkin@gmail.com

Timur Nurgaliev
Saint Petersburg State University
Saint Petersburg, Russia
hmtimurhm@gmail.com

Nikolay A. Bokhan
Mental Health Research Institute, Tomsk National Research Medical Center; National Research Tomsk State University; Siberian State Medical University
Tomsk, Russia
bna909@gmail.com

Svetlana A. Ivanova
Mental Health Research Institute Tomsk National Research Medical Center; National Research Tomsk State University; National Research Tomsk Polytechnic University
Tomsk, Russia
ivanovaniipz@gmail.com

***Abstract* — In the present study six genes were analyzed by sequencing a cohort of patients with affective disorders. A number of genetic variants were found to be associated with antidepressant treatment response, time to recurrence of episodes, and depression severity.**

Keywords* — *neuregulin 1 (NRG1), serotonin 2C receptor (HTR2C), phosphatidylinositol-5-phosphate 4-kinase type 2 alpha (PIP4K2A), depressive episode, bipolar affective disorder, treatment response, severity, time to recurrence, genetic biomarker

Introduction

GSK3B [1], *BDNF* [2], *NGF* [3], *NRG1* [4], *HTR2C* [5, 6], and *PIP4K2A* [7] are genes that showed some importance in the study of molecular etiology of psychiatric disorders. These genes are functionally connected. GSK3B plays a central role in the AKT/GSK3 molecular pathway [1]. The binding of either BDNF, NGF, or NRG1 to their respective receptor tyrosine kinases (TrkB, TrkA, and ERBB4) leads to inhibition of GSK3B via activation of phosphoinositide 3kinase (PI3K), 3-phosphoinositide-dependent protein kinase1 (PDPK1) and AKT1 [1, 4, 8, 9]. Binding of serotonin to HTR2C leads to the opposite effect: GSK3B activation [1]. PIP4K2A codes for phosphatidylinositol-5-phosphate 4kinase type 2α, a kinase that catalyzes the phosphorylation of phosphatidylinositol-5-phosphate (PI5P) to form phosphatidylinositol-5,4-bisphosphate (PIP2) [7]. A link between PIP4K2A and GSK3B involves activation of potassium channels KCNQ by PIP2 and GSK3B (Fig. 1) [10-12]. These channels play a critical role in the regulation of neuronal excitability [13] and are potential treatment targets for manic symptoms [14].

Study subjects. We included the following numbers of patients with mood disorders, diagnosed using the criteria of the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10): 19 patients with bipolar affective disorder, 55 patients with depressive episode, and 41 patients with recurrent depressive disorder. As a negative control, 34 individuals without psychiatric disorders were also recruited into the study.

Russian Foundation for Basic Research (grant number 17-29-02205)

Targeted DNA sequencing. Targeted next generation sequencing was performed using the Ion Torrent semiconductor technology (Thermo Fisher Scientific) at the Center “Medical Genomics” of the Tomsk National Research Medical Center. A custom DNA panel of 86 amplicons, covering 57 targets, was designed using the Ion AmpliSeq Designer. The targets comprised exons and flanking intronic regions of *GSK3B, BDNF, NRG1, NGF, HTR2C,* and *PIP4K2A.*

Sequencing analysis workflow. The unmapped BAM files were converted to the FASTQ format using Picard tools. Quality control of raw sequencing data and trimming of low quality bases and adapters was done using FastQC and Trimmomatic. Discovery of variants was done using GATK. Variants were annotated with rsIDs listed in the Database of Single Nucleotide Polymorphisms build 153 (dbSNP) and the Genome Aggregation Database (gnomAD). In order to assure the extraction of reliable sequencing results, we used hard filtering parameters.

Functional annotation of variants. Functional annotation of all discovered variants was done with ANNOVAR, SnpEff, HumanSplicingFinder, the GeneCards database, and the Genotype-Tissue Expression (GTEx) portal that lists expression quantitative trait loci (eQTLs) in various tissues.

Association study. Tests under a number of statistical models – linear regression and various logistic regressions – were run to evaluate association between variants, discovered with sequencing, and clinical subphenotypes.

Results

Association with clinical subphenotypes

1) *NRG1:* The allele "del" of NC_000008.11:g.32614509_32614510del is associated with absence of response to antidepressants among patients with depression. Longer intervals between episodes in the bipolar and cross-disorder groups are associated with the allele C of rs35641374.

2) *PIP4K2A:* The allele A of rs61731109 is associated with absence of response to antidepressants among patients with

depression. The allele G of rs10508649 is associated with absence of response to antidepressants among patients with depression and with longer intervals between manic or mixed episodes among patients with bipolar disorder.

3) *HTR2C*: The allele A of rs2248440 is associated with higher severity of depression among patients with depression.

Functional annotation of associated variants

1) *NRG1:* no obvious biological function of associated variants.

2) *PIP4K2A:* rs61731109 and rs10508649 may affect splicing of *PIP4K2A*'s transcripts; the second variant is found within the enhancer GH10J022572 for this gene.

3) *HTR2C:* rs2248440 is an eQTL.

Discussion

Possible genetic biomarkers

Rs35641374 and rs10508649 (*NRG1* and *PIP4K2A*) may be prognostic biomarkers of time to recurrence of depressive and manic/mixed episodes among patients with bipolar disorder. NC_000008.11:g.32614509_32614510del, rs61731109, and rs10508649 (also *NRG1* and *PIP4K2A*) seem to be predictive biomarkers of response to pharmacological antidepressant treatment. In particular, the allele G of rs10508649 (*PIP4K2A*) may increase resistance to antidepressant treatment and be at the same time protective against recurrent manic/mixed episodes. These results support previous data indicating a biological link between resistance to antidepressant treatment and mania [15, 16]. Bioinformatic functional annotation of associated variants revealed possible impact for transcriptional regulation of *PIP4K2A*. In addition, the allele A of rs2248440 (*HTR2C*) may be a prognostic biomarker of disease severity among patients with depression. This allele decreases expression in the putamen of the immune system gene *IL13RA2* that codes for a subunit of the interleukin 13 receptor. This variant is in linkage disequilibrium with rs6318 (previously associated with mood disorders and response to antidepressant therapy) that is an eQTL for the same gene and tissue.

Genetic interactions suggested by this study

1) *NRG1–PIP4K2A:* These genes could act in concert via their connections with *GSK3B* (Fig. 1) [1, 4, 10, 11, 13]. ERBB4 and its ligand NRG1 play an important role in neurodevelopment, neurotransmission, and synaptic plasticity and this receptor is present on GABAergic, glutamatergic and dopaminergic neurons [4]. One of its functions is to promote the inhibitory GABA release. Thus, together with KCNQ activated by PIP2 [13], ERBB4 regulates neuronal excitability, an aspect that could be important in the pathogenesis of bipolar disorder [17].

2) *HTR2C–IL13RA2:* Serotonin acts upon the brain and peripheral immune system components, while both types of these components regulate the serotonin neurotransmission [18, 19]. This connection could be the key to the observed immune system abnormalities strongly associated with depression [20] and suggests a possibility of an interaction between *HTR2C* and *IL13RA2*. The immune system could therefore be a biological factor that modulates the severity of depression [20].

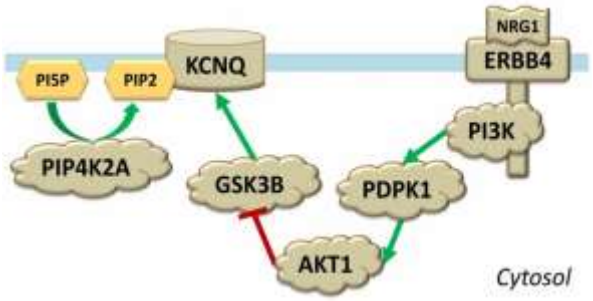


Fig. 1. Shown are molecular interactions involving protein products of *PIP4K2A* and *NRG1* found to be associated with time to recurrence of an episode in bipolar disorder and antidepressant treatment response

Acknowledgment

The work was in part made possible thanks to the Competitiveness Enhancement Program of National Research Tomsk Polytechnic University.

References

[1] Beaulieu, J.M. A role for Akt and glycogen synthase kinase-3 as integrators of dopamine and serotonin neurotransmission in mental health. J Psychiatry Neurosci, 2012. 37(1): p. 7-16.

[2] Lu, B., G. Nagappan, and Y. Lu, BDNF and synaptic plasticity, cognitive function, and dysfunction. Handb Exp Pharmacol, 2014. 220: p. 223-50.

[3] Cirulli, F. and E. Alleva, The NGF saga: from animal models of psychosocial stress to stress-related psychopathology. Front Neuroendocrinol, 2009. 30(3): p. 379-95.

[4] Mei, L. and K.A. Nave, Neuregulin-ERBB signaling in the nervous system and neuropsychiatric diseases. Neuron, 2014. 83(1): p. 27-49.

[5] Chagraoui, A., F. Thibaut, M. Skiba, C. Thuillez, and M. Bourin, 5HT2C receptors in psychiatric disorders: A review. Prog Neuropsychopharmacol Biol Psychiatry, 2016. 66: p. 120-35.

[6] Palacios J.M. et al. A short history of the 5HT2C receptor: from the choroid plexus to depression, obesity and addiction treatment. Psychopharmacology (Berl), 2017. 234(9-10): p. 1395-1418.

[7] Wilcox, A. and K.A. Hinchliffe, Regulation of extranuclear PtdIns5P production by phosphatidylinositol phosphate 4-kinase 2alpha. FEBS Lett, 2008. 582(9): p. 1391-4.

[8] Kim, M.S., L.P. Shutov et al. Nerve growth factor (NGF) regulates activity of nuclear factor of activated T-cells (NFAT) in neurons via the phosphatidylinositol 3kinase (PI3K)-Akt-glycogen synthase kinase 3beta (GSK3beta) pathway. J Biol Chem, 2014. 289(45): p. 31349-60.

[9] Numakawa, T. et al. Actions of brain-derived neurotrophin factor in the neurogenesis and neuronal function, and its involvement in the pathophysiology of brain diseases. Int J Mol Sci, 2018. 19(11).

[10] Jiang, L., A. Kosenko et al. Activation of m1 muscarinic acetylcholine receptor induces surface transport of KCNQ channels through a CRMP-2-mediated pathway. J Cell Sci, 2015. 128(22): p. 4235-45.

[11] Carter, C.J., Multiple genes and factors associated with bipolar disorder converge on growth factor and stress activated kinase pathways controlling translation initiation: implications for oligodendrocyte viability. Neurochem Int, 2007. 50(3): p. 461-90.

[12] Fedorenko, O., N. Strutz-Seebohm et al., A schizophrenia-linked mutation in PIP5K2A fails to activate neuronal M channels. Psychopharmacology (Berl), 2008. 199(1): p. 47-54.

[13] Greene, D.L. and N. Hoshi, Modulation of Kv7 channels and excitability in the brain. Cell Mol Life Sci, 2017. 74(3): p. 495-508.

[14] Grunnet, M. et al. Kv7 channels as targets for anti-epileptic and psychiatric drugdevelopment. Eur J Pharmacol, 2014. 726: p. 133-7.

[15] Hirschfeld, R.M., Differential diagnosis of bipolar disorder and major depressive disorder. J Affect Disord, 2014. 169 Suppl 1: p. S12-6.

[16] Lee, H.J., G.H. Son, and D. Geum, Circadian rhythm hypotheses of mixed features, antidepressant treatment resistance, and manic switching in bipolar disorder. Psychiatry Investig, 2013. 10(3): p. 225-32.

[17] Berridge, M.J., Calcium signalling and psychiatric disease: bipolar disorder and schizophrenia. Cell Tissue Res, 2014. 357(2): p. 477-92.

[18] Robson, M.J., M.A. Quinlan, and R.D. Blakely, Immune system activation and depression: roles of serotonin in the central nervous system and periphery. ACS Chem Neurosci, 2017. 8(5): p. 932-942.

[19] Wu H. et al. Pharmacol Res, 2019. 140: p. 100-114.

[20] Felger, J.C. and F.E. Lotrich, Inflammatory cytokines in depression: neurobiological mechanisms and therapeutic implications. Neuroscience, 2013. 246: p. 199-229.

DOI 10.18699/BGRS/SB-2020-180

Delta- and gamma-activity of resting state EEG as one of markers of risk of depressive disorders in migrants of subpolar and polar regions of Siberia

Natalya Milakhina
ICG SB RAS, Novosibirsk, Russia
tashamilka@mail.ru

Sergey Tamozhnikov
State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk, Russia
stam@physiol.ru

Ekaterina Proshina
State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk, Russia
proshinaea@physiol.ru

Alexandra Karpova
North-Eastern Federal University, Yakutsk, Russia
karpova74@list.ru

Alexander Savostyanov
ICG SB RAS, Novosibirsk, Russia
a-sav@mail.ru

Elena Afanaseva
North-Eastern Federal University, Yakutsk, Russia
eb.afanaseva@mail.ru

***Abstract* — It is well known, the conditional rest state activates the functional rest network, including the brain default system (DMN) and the central executive network (CEN). At this stage of research, it can be assumed that both the process of adaptation to the polar climate and new social conditions, and the gradual reduction of the risk of depression among migrants, are due to functional changes within these two brain systems, as well as their functional relationships among themselves. Hypothetically, the increased activity of CEN is associated with a higher level of stress associated with moving to new living conditions**

***Keywords* — EEG, depression, DMN, CEN, migrants, delta-rhythm, gamma-rhythm.**

Motivation and Aim

A. Motivation

According to the WHO, today depression is one of the most common diseases worldwide, affecting more than 264 million people. Modern neuroscience is focused on the search for markers of mental illness for early diagnosis and rational treatment tactics.

B. Aim.

The aim of the study was to investigate the relationship between the indicators of electroencephalography in functional rest and predisposition to depressive disorders in labor migrants adapting to new climatic and social conditions (Republic of Sakha, Yakutia)

Methods

The experiment involved 50 young migrants and 40 indigenous people who were not previously supervised in psychoneurological dispensers and were not receiving psychological and psychiatric care at the time of the study. All migrants have moved to long-term residence in Yakutia from the southern regions (mainly Central Asia and Africa). The migrants were examined twice: Immediately after the move and six months after a non-visiting stay in Yakutia. In addition, the participants of the experiment filled the psychological questionnaires of Achenbach and the *Beck Depression Inventory* for further evaluation of the predisposition to depression.

EEG was recorded at rest with 128 channels of NeoRec amplifier, Russia. The evaluation of the current source density, connectivity and topological features of the rest EEG networks was made in the sLORETA software package.

Results

When comparing the results of psychological tests between migrants and indigenous people, it was found that migrants have an increased predisposition to develop depression and anxiety disorder immediately after the move, while according to the results of testing of the indigenous population, this trend was not observed. However, in the process of adaptation to new climatic and social conditions, the risk of these diseases in most migrants is reduced, which was revealed by the results of the same psychological tests conducted after six months. The high risk of depressive and anxiety disorders was associated with increased spectral power and connectivity in the delta and gamma rhythms bands in the medial dark and medial prefrontal cortex. Adaptation to the polar climate led to a decrease in the delta and gamma of spectral power, which correlated with a decrease in predisposition to depression.

Conclusion

Adaptation to the circumpolar climatic conditions is accompanied by the functional changes in the resting-state brain networks that can be revealed by the EEG analyses. The first stage of adaptation is accompanied by an increased activity of the central executive network (CEN) that is connected with the increased risk of developing of anxiety disorder. In later stage of adaptation, these changes become less pronounced that correlates with a reduced risk of anxiety and depressive disorders.

Acknowledgment

Supported by the Russian Foundation of Basic Research grants No. 18-29-13027, No. 18-415-140021 and the Budget project No. 0324-2019-0040-C-01.

DOI 10.18699/BGRS/SB-2020-181

Positive effect of joint activation of mTOR-dependent and mTOR-independent pathways of autophagy in the treatment of two experimental models of neurodegeneration

Alexander B. Pupyshev
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
pupyshevab@physiol.ru

Nina I. Dubrovina
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
dubrov@physiol.ru

Maria A. Tikhonova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
tikhonovama@physiol.ru

Anna A. Akopyan
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
annaaleksanovna@mail.ru

Marina V. Ovsyukova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia maryov@ngs.ru

Mikhail V. Tenditnik
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
m.v.tenditnik@physiol.ru

Tatiana A. Korolenko
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
t.a.korolenko@physiol.ru

***Abstract* — Autophagy is a part of the cellular protein quality control system and is able to suppress neurodegeneration via weakening pathogenic proteins accumulation. Meanwhile possibilities of different ways of autophagy induction in treatment of experimental neurodegeneration are poorly known. Here we used pharmacological models of Alzheimer's disease (AD) and Parkinson's disease (PD) induced correspondingly by amyloid-beta and MPTP administration in mice. mTOR-dependent autophagy was induced by rapamycin and mTOR-independent autophagy was activated by trehalose. Both drugs caused increase in autophagy (according to LC3-II level) in the affected brain structures, namely, the striatum and s. nigra in PD model and the hippocampus in AD model. Neuronal loss was reduced according to the expression of tyrosine hydroxylase (PD) or neuronal density (AD). Neuroinflammation was sharply reduced according to IBA1 microglial marker in the hippocampus (AD). Trehalose was more active than rapamycin, but the highest effect was achieved by their combined application. An obvious restoration of cognitive function in the passive avoidance test for memory and learning using all three treatment approaches has been found for both models of neurodegeneration. Recovery of emotional activity, as measured by a decrease in anxiety in AD model, was significant only for the combination of the drugs. Cytoprotective and therapeutic effects of the drugs in AD model were eliminated by blocking autophagy with 3-methyladenine. Thus, a more noticeable neuroprotective effect is produced by trehalose in comparison with rapamycin while a joint activation of the mTOR-dependent and mTOR-independent autophagy pathways, by rapamycin and trehalose, respectively, renders the highest impact.**

Keywords* — *animal models, Alzheimer's disease, Parkinson's disease, neuroprotection, autophagy, mTOR-dependent, mTOR-independent, rapamycin, trehalose, LC3-II, IBA1, passive avoidance test.

Motivation and Aim

Motivation

Autophagy is a part of cellular protein quality control system and renders cytoprotective effect in aberrant misfolded protein accumulation occurring in neurodegeneration processes. Canonical mTOR-dependent pathway of autophagy activation induced by rapamycin is connected with some side effects. mTOR-independent stimulation of autophagy started up by trehalose seemingly is devoided of side impacts and could be a better agonist. The positive effects of the drugs in neurodegeneration are described mostly in *in vitro* models [1].

Aim

Here we tried to estimate neuroprotective effects of rapamycin and trehalose in their separate and joint application in two animal models of neurodegeneration, i.e. Alzheimer's and Parkinson's diseases.

Methods

We used pharmacological models of Alzheimer's disease (AD) and Parkinson's disease (PD). The first was induced by intraventricular injection of amyloid-β 25-35 fragment, and the second was reproduced by systemic MPTP administration into mice C57Bl/6 [2]. Rapamycin was applied i/p in dose 10 mg/kg body weight, trehalose was included in drinking (2 %). Their impacts were evaluated by means of immunohistochemistry of autophagy marker LC3-II, level of amyloid-β, tyrosine hydroxylase, neuronal density (Nissl staining), behavioral tests ("memory of fear", T-maze, "open field").

Results

Both drugs caused increase in autophagy (according to LC3-II level) in the affected brain structures, namely, the striatum and s. nigra in PD model and the hippocampus in AD model. Neuronal loss was reduced according to the expression of tyrosine hydroxylase (PD) or neuronal density (AD). Neuroinflammation was sharply reduced according to IBA1 microglial marker in the hippocampus (AD). Trehalose was predominantly more active than rapamycin, but the highest effect was achieved by their combined application. An obvious restoration of cognitive function in the passive avoidance test for memory and learning using all three treatments has been found for both models of neurodegeneration. Recovery of emotional activity, as measured by anxiety in AD model, was significant only for

the combination of the drugs. Cytoprotective and therapeutic effects of the drugs in AD model were eliminated by blocking autophagy with 3-methyladenine. Thus, a more noticeable neuroprotective effect is produced by trehalose in comparison with rapamycin while a joint activation of the mTOR-dependent and mTOR-independent autophagy pathways, by rapamycin and trehalose, respectively, has the highest impact.

ACKNOWLEDGMENT

The study was supported by budgetary funding for basic scientific research of the Scientific Research Institute of Physiology and Basic Medicine (theme No. AAAA-A16-116021010228-0) and the Russian Foundation for Basic Research (grant No. 16-04-01423-a).

REFERENCES

[1] Sarkar S. et al., (2007). Trehalose, a novel mTOR-independent autophagy enhancer, accelerates the clearance of mutant Huntingtin and alpha-synuclein. J. Biol. Chem. 282 (8): 5641–5652.

[2] Pupyshev A.B. et al., (2019). Therapeutic activation of autophagy by combined treatment with rapamycin and trehalose in a mouse MPTP induced model of Parkinson's disease. Pharmacol. Biochem. Behav.177(1): 1-11.

DOI 10.18699/BGRS/SB-2020-182

Interplay between 5-HT and BDNF system in recombinant mouse strain upon chronic fluoxetine administration

Rodnyy A.Ya.
Institute of Cytology and Genetics
Novosibirsk, Russia
aleksandr1994rodny@gmail.com

Kondaurova E.M.
Institute of Cytology and Genetics
Novosibirsk, Russia
kond_em@bionet.nsc.ru

Antonov Y.V.
Institute of Cytology and Genetics
Novosibirsk, Russia
yegor@bionet.nsc.ru

Ilchibaeva T.V.
Institute of Cytology and Genetics
Novosibirsk, Russia
rbicehok@mail.ru

Tsybko A.S.
Institute of Cytology and Genetics
Novosibirsk, Russia
antontsybko@bionet.nsc.ru

Naumenko V.S.
Institute of Cytology and Genetics
Novosibirsk, Russia
naumenko2002@bionet.nsc.ru

***Abstract* — BDNF plays a key role in the development, differentiation, synaptogenesis and survival of brain neurons and in the processes of their adaptation to external impacts. Serotonergic (5-HT) system is another basic player in brain development and neuroplasticity. The study presents a comparative analysis of chronic fluoxetine treatment in recombinant mice differing in distal chromosome 13 fragment containing *Htr1A* gene of CBA mice strain on C57Bl6 genetic background. The problem here to be studied is mechanism of BDNF and 5-HT systems` interactions in antidepressant insensitivity. We measured mRNA and protein levels of BDNF, $p75^{NTR}$, TrkB, 5-HT1A and 5-HT7 receptors, levels of 5-HT and its primary metabolite 5-Hydroxyindoleacetic acid (5-HIAA) in the brain structures that could have play primary role in mechanism of depression - frontal cortex and hippocampus. At the heart of the discussion are the different changes in BDNF system as well as in 5-HT system which allows us to conclude that the chronic fluoxetine injection increased depressive-like behavior in recombinant mice carrying distal chromosome 13 fragment containing *Htr1A* gene of CBA mice.**

Keywords* — *fluoxetine; BDNF; 5-HT1A; 5-HT7; mice; serotonin

Depression appears to be the most widespread mental illness on the planet and represents one of the most urgent threats to humanity in the 21st century. According to various sources, depression in the stage requiring active medical care affects 10% to 23% of the industrialized countries population, but despite a huge amount of studies, the mechanisms and pathogenesis of depressive disorders appeared to be still far from complete understanding. Disturbance in many neurotransmitter systems underlie the biochemical cause of depression. Among these a special place is occupied by the serotonergic, dopaminergic, cholinergic, GABAergic, norepinephrine systems as well as brain-derived neurotrophic factor system. Both BDNF and 5-HT systems play crucial role in mechanism of neurogenesis and neuroplasticity and are key players in mechanisms of depressive disorders.

Modern antidepressants block the serotonin transporter (5-HTT) suppressing the reuptake of serotonin from the synaptic cleft. However, from 20% to 40% of patients with depression are resistant to this class of antidepressants. Moreover, chronic antidepressant treatment is required to effectively correct depression. A key role in the regulation of the functional state of the brain serotonin system is assigned to the serotonin 5-HT1A receptor, which, through negative feedback, can inhibit the secretion of serotonin and the activity of serotonin neurons. Over the past 20 years, a lot of data has been accumulated confirming the participation of 5-HT1A receptors in the mechanisms of depression and in the action of antidepressants.

Modern neurobiology is actively studying the effect of genetic abnormalities on drug actions, because one or another genetic polymorphism can dramatically change the response to a drug. For example, changing the structure of receptors or transport molecules could essentially mediate signal transmission or drug delivery. It has been known that the same genes the different genetic backgrounds can differentially exhibit their functions, which leads to ambiguous reactions of drug therapy. Of great interest and broad prospects in the study of this problem are recombinant mice, which differ by the genome with the transferred or modified DNA fragment.

Recombinant B6.CBA-D13Mit76C (B6-76C) mouse line carrying distal chromosome 13 fragment containing *Htr1A* gene of CBA on the C57Bl/6 genetic background is characterized by increase in postsynaptic and decrease in presynaptic functional responses mediated by 5-HT1A receptor in comparison with control B6.CBA-D13Mit76B (B6-76B) mouse line. The B6-76C showed decreased metabolism of 5-HT and increased BDNF/proBDNF ratio and these effects are accompanied by increased 5-HT1A receptor gene expression in hippocampus compared with B6-M76B mice. In addition to this after chronic fluoxetine treatment B6-76C demonstrated decreased mobility in forced swim test which is commonly used to evaluate the effectiveness of antidepressants.

The aim of this study was to estimate the effect of chronic fluoxetine injection (20mg/kg, 14 days i.p.) on mRNA and protein levels of BDNF, proBDNF, $p75^{NTR}$, TrkB, 5-HT1A receptors, its transcription factors FREUD-1, FREUD-2 and 5-HT7 receptor, levels of 5-HT and its primary metabolite 5-HIAA in the frontal cortex and hippocampus of B6-76B and B6-76C mouse lines by using Western-Blot analysis, real time PCR and HPLC. These structures assumed to play the main role in pathogenesis of the depressive disorders. The data were analyzed using two-way ANOVA followed by Fisher's LSD post-hoc comparison.

The mRNA level of 5-HT1A receptor gene was decreased in cortex and in hippocampus in the B6-76C mice but not in the B6-76B mice after fluoxetine administration. However, 5-HT1A receptor transcription factors FREUD-1 and FREUD-2 were not altered. Chronic fluoxetine treatment led to decrease in level of 5-HT7 receptor gene expression in the cortex of B6-76C and in the hippocampus of B6-76B mice. The mRNA level of BDNF gene was increased in B6-76B in the cortex and in B6-76C in the hippocampus after fluoxetine treatment compared with control groups. No differences in $p75^{NTR}$ and TrkB receptors genes expression in cortex and hippocampus of the both lines were determined.

We have also found significant changes in serotonin and 5-HIAA upon chronic fluoxetine treatment. Decrease in serotonin and 5-HIAA levels in the cortex of both lines and in the hippocampus of B6-76B mice was shown. The 5HIAA/5HT ratio was not changed in all investigated brain structures in both mouse lines after chronic fluoxetine treatment.

Western-Blot analysis revealed intriguing changes in BDNF system. BDNF protein levels were significantly higher in B6-76C after chronic fluoxetine treatment in frontal cortex and hippocampus. BDNF precursor proBDNF protein level was lower in both lines after fluoxetine in frontal cortex. There no proBDNF protein changes in hippocampus were found. Chronic fluoxetine treatment also failed to affect 5-HT1A, 5-HT7 receptors protein in investigated brain structures and TrkB-FL, TrkB-T1 and $p75^{NTR}$ receptors protein in hippocampus. However, in frontal cortex TrkB-T1 was lower in B6-76B after fluoxetine treatment.

Thus, classic antidepressant increased depressive-like behavior which is accompanied by decreased the expression of post-synaptic 5-HT1A receptor, increased expression of BDNF genes and its product BDNF protein in hippocampus of B6-76C mice and decreased proBDNF level in B6-76C after fluoxetine treatment. It was found that changes in sensitivity of the 5-HT1A receptors results in the significant changes in the response to chronic antidepressant treatment.

ACKNOWLEDGMENT

The work was supported by Russian Scientific Foundation, project No.19-15-00027.

DOI 10.18699/BGRS/SB-2020-183

Altered expression of genes *Npas4* and *Nr1d1* in adult female mice with history of early-life stress

Yuliya Ryabushkina
ICG SB RAS, Novosibirsk, Russia
ryabushkina@bionet.nsc.ru

Vasiliy Reshetnikov
ICG SB RAS, Novosibirsk, Russia
vasiliyreshetnikov@bionet.nsc.ru

Natalya Bondar
ICG SB RAS, Novosibirsk, Russia
nbondar@bionet.nsc.ru

***Abstract* — Stressful events early in life alter neuronal plasticity of the brain regions that regulate social behavior. Previous works have shown that brief and prolonged separation of pups from their mothers leads to enhanced social behavior in adult female mice. In this study, we performed *Egr1*, *Npas4*, *Arc*, and *Homer1* gene expression level analysis in the prefrontal cortex and dorsal hippocampus of adult female mice after exposure to early life stress. Also was evaluated expression level of stress-related genes glucocorticoid and mineralocorticoid receptors (*Nr3c1* and *Nr3c2*) as well as *Nr1d1*, which encodes a transcription factor REVERBα, which regulates sociability and anxiety-related behavior. C57Bl/6 mice were exposed to maternal separation (MS, 3 h once a day) or handling (HD, 15 min once a day) on postnatal days 2 through 14. As adults, female mice behavior was analyzed by behavioral tests, on the day after last testing gene expression level analysis was performed. We found evaluation of *Npas4* expression in prefrontal cortex and *Nr1d1* both in prefrontal cortex and dorsal hippocampus of adult female mice of MS group, not in HD group. The expression of stress-related genes *Nr3c1* and *Nr3c2* was unchanged in both groups. This upregulation of *Npas4* and *Nr1d1* genes in female mice with stressful events early in life and enhanced social behavior may be an adaptation mechanism reversing possible aberrations caused by early-life stress.**

Keywords* — *early life stress, gene expression, prefrontal cortex, dorsal hippocampus.

Motivation and Aim

Stressful events early in life alter neuronal plasticity of the brain regions that regulate social behavior. Molecular mechanisms of this alterations remain unclear. Previous work have shown that brief and prolonged separation of pups from their mothers leads to enhanced social behavior in adult female mice [1].

Immediate early genes are strongly involved in synaptic plasticity [2], their products play a role in several distinct processes required for long-term synaptic changes and memory formation [3].

The specific aim of the present study was to characterize the expression of immediate early genes in the prefrontal cortex and dorsal hippocampus of adult female mice as a marker of modified neuroplasticity elicited both by stress early in life and by previous social interaction.

To exclude effect estrous cycle of female mice on gene expression level, correlational analysis between gene expression and serum estradiol was performed. In addition, correlational analysis between gene expression an behavioral parameters was performed.

Methods

We used brief (15 min per day) and prolonged (3 h per day) maternal separation prosedure for 14 days. As adults, mice were subjected to elevated plus maze, open field, and the social interaction test (one test per day). After last day of test, mice were decapitated and trunk blood was collected. In the same time brains were removed, prefrontal cortex was dissected and snap frozen in liquid nitrogen, the rest of the brain was embedded in the Tissue-Tek O.C.T. Compound. Frozen brains were cut into coronal sections in a cryostat Microm HM 550. Two 150 μm sections were dissected according with Allen Brain Atlas (Bregma from −1.86 to −2.16, levels 73– 76) and dorsal hippocampus was isolated from slices. RNA extraction was performed with usind PureZOL and complementary DNA was synthesized using kit by Syntol in accordance with manufacturer's recommendations. Gene expression level analysis was performed by real-time PCR, evaluation the expression of followed genes: *Egr1*, *Npas4*, *Arc*, *Homer1a*, *Homer1b/c*, *Nr3c1*, *Nr3c2*, and *Nr1d1*. Serum 17β-estradiol was evaluated by using enzyme-linked immunosorbent assay (ELISA) kit following the manufacturer's protocols.

Results

We found that prolonged maternal separation increases expression level of gene Npas4 in prefrontal cortex and Nr1d1 in prefrontal cortex and dorsal hippocampus (Fig.1), expression level of other genes remain unchanged.

Npas4 is a transcription factor, which controls expression of other genes like *Arc*, *Egr1* and *c-Fos*. Usually enhanced *Npas4* expression accompanied with upregulation of genes including *Arc*, *Egr1* и *Bdnf*. In this study, expression of these genes did not exhibit significant change. Despite of lack of significant changes, correlation analysis revealed strong correlation between expression levels of genes *Npas4*, *Arc*, *Egr1* and *Bdnf* in prefrontal cortex. Alteration of expression of these genes has not found in dorsal hippocampus, as well as correlation in expression level. These observed changes in expression level suggest that the observed alterations are specific to the prefrontal cortex. Correlational analysis between expression levels of the studied genes and behavioral parameters revealed positive correlation between genes in prefrontal cortex (*Npas4*, *Egr1*, *Nr3c1* and *Nr3c2*) and one in hippocampus (*Homer1a*) with level of social behavior.

Nr1d1 is a gene encoding the nuclear receptor, also known as REVERBa, which control gene transcription. Its most studied function is regulation of circadian rhythm. Enhances expression of this gene can be related with social behavior through clock genes activity in prefrontal cortex and hippocampus. The stress response system and the circadian clock are closely connected.

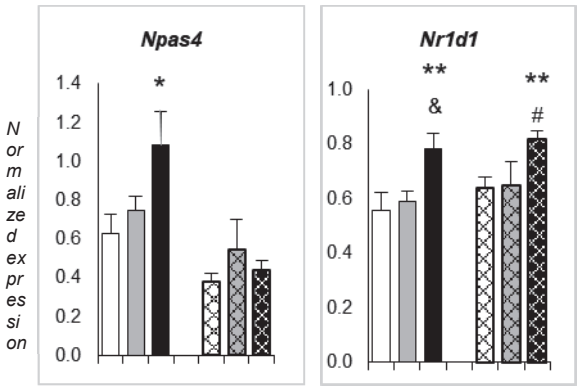


Fig. 1. Effects of early-life stress on gene expression in the prefrontal cortex and dorsal hippocampus. Data are presented as mean ± SEM. HPC: hippocampus; PFC: prefrontal cortex; NC: no-treatment control; HD: handling, MS: maternal separation. ∗p < 0:05, ∗∗p < 0:01 as compared with the NC group; &p < 0:1 (tendency), #p < 0:05 as compared with the HD group (post hoc Fisher's LSD test).

Glucocorticoids may influence on expression of clock genes due to presence of glucocorticoid response elements in their promoters, thus synchronization of peripheral and central circadian oscillators occurs. In this experiment, there is no alterations in *Nr3c1* and *Nr3c2* expression in prefrontal cortex and hippocampus. This corresponds to our previous results [4]. After stress, there was no changes in *Nr3c1* and *Nr3c2* expression in males, but prolonged maternal separation increased *Nr3c2*/*Nr3c1* mRNA ratio in hippocampus and hypothalamus. For most genes, we did not detect a significant correlation between estradiol concentration and gene expression; this finding allows us to compare gene expression levels without considering the stage of the cycle.

ACKNOWLEDGMENT

This study was supported by the Russian Science Foundation (16-15-10131).

REFERENCES

[1] 1Bondar NP, Lepeshko AA, Reshetnikov V V. Effects of Early-Life Stress on Social and Anxiety-Like Behaviors in Adult Mice: Sex-Specific Effects. Behav Neurol. 2018;2018: 32–34.

[2] Okuno H. Regulation and function of immediate-early genes in the brain: Beyond neuronal activity markers. Neurosci Res. 2011;69: 175–186.

[3] Minatohara K, Akiyoshi M, Okuno H. Role of immediate-early genes in synaptic plasticity and neuronal ensembles underlying the memory trace. Front Mol Neurosci. 2016;8: 1–11.

[4] Reshetnikov V V., Studenikina AA, Ryabushkina JA, Merkulova TI, Bondar NP. The impact of early-life stress on the expression of HPA-associated genes in the adult murine brain. Behaviour. 2018;155: 181–203.

DOI 10.18699/BGRS/SB-2020-184

Effects of diets rich in plant polyphenols in mouse models of neurodegenerative disorders

Maria A. Tikhonova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
tikhonovama@physiol.ru

Tamara G. Amstislavskaya
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
amstislavskaya@yandex.ru

Anna A. Akopyan
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
annaaleksanovna@mail.ru

Marina V. Ovsyukova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
maryov@ngs.ru

Michael V. Tenditnik
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
m.v.tenditnik@physiol.ru

Elena K. Khlestkina
N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources
St. Petersburg, Russia
Federal Research Center "Institute of Cytology and Genetics"
Novosibirsk, Russia
khlest@bionet.nsc.ru

***Abstract* — Currently functional nutrition has developed intensively. Functional foods enriched with biologically active substances are an essential part of dietary therapy. Plant polyphenols attract particular attention due to their multiple beneficial properties. This study was aimed to evaluate the effects of diets rich in plant polyphenols in a mouse model of Alzheimer's disease (AD) induced by amyloid-beta and a transgenic model of Parkinson's disease (PD) in mice of B6.Cg-Tg (Prnp-SNCA*A53T)23Mkle/J strain with overexpression of human alpha-synuclein. We used two wheat lines created at the Institute of Cytology and Genetics SB RAS [1] that have almost similar genomes with the exception of a small part of chromosome 2A, which contains a gene regulating anthocyanin biosynthesis, to assess the effects of an anthocyanin-rich grain diet, and a concentrate of grape polyphenols "Enoant" dissolved in the drinking water in addition to standard chow. Mice were kept at those diets for five months. Mice with AD model demonstrated deficits in working and short-term spatial memory. Both grain diets prevented the development of disturbances in working memory but not in short-term spatial memory. Mice of transgenic PD model also showed deficits in working memory that were improved only by the grain diet with high content of anthocyanins. The anthocyanin-rich grain diet enhanced microglial activation and increased the expression of arginase1 marking the M2 microglia that promotes tissue viability and neuronal survival in the brain of transgenic PD-like mice. Although "Enoant" did not produce significant effects in AD and PD models, it substantially prolonged the average life span in mice with natural aging.**

Keywords* — *functional food, wheat grain, anthocyanins, grape polyphenols, mice, Alzheimer's disease, amyloid-beta, alpha-synuclein, Parkinson's disease.

ACKNOWLEDGMENT

The study was supported by Russian Science Foundation (grant No. 16-14-00086).

REFERENCES

[1] E. I. Gordeeva, O. Y. Shoeva, and E. K.Khlestkina, "Marker-assisted development of bread wheat near-isogenic lines carrying various combinations of purple pericarp (Pp) alleles," Euphytica, vol. 203, pp. 469–476, 2015.

DOI 10.18699/BGRS/SB-2020-185

Associations of polymorphic variants of the genes of neurotrophic factors *BDNF, NGF, NRG1* with remission in patients with depressive disorders

Natalia Vyalova
Mental Health Research Institute of TNRMC, Tomsk, Russia
Natarakitina@yandex.ru

German Simutkin
Mental Health Research Institute of TNRMC, Tomsk, Russia
ggsimutkin@gmail.com

Nikolay Bokhan
Mental Health Research Institute of TNRMC, Tomsk, Russia
bna909@gmail.com

Svetlana Ivanova
Mental Health Research Institute of TNRMC, Tomsk, Russia
ivanovaniipz@gmail.com

***Abstract* — The investigation the relationship between *BDNF, NGF* and *NRG1* genotypes with remission in patients with MDD was made in a group of 208 patients. The study showed that rs3924999 of the *NRG1* gene was associated with clinical remission, rated on the HDRS-17 scale on day 28 of antidepressant therapy.**

Keywords* — *depressive disorders, brain derived neurotrophic factor, nerve growth factor, neuregulin 1, gene polymorphism, remission

Motivation and Aim

Motivation

Patients with major depressive disorder exhibit changes in biological biomarkers, such as neurotrophic factors, demonstrating that these molecules are involved in the psychopathology, therapy, and monitoring of major depressive disorder [1]. In recent years, the possibility of using knowledge about gene polymorphisms of neurotrophic factors to predict or evaluate the effectiveness of a response to therapy has been widely studied [2].

Aim.

The aim of our study was to identify possible associations of polymorphic variants of the BDNF, NGF, and NRG1 genes with clinical remission in patients with depressive disorders.

Methods

We examined 208 patients with MDD (as part of a single depressive episode or recurrent depressive disorder (F32 and F33 according to ICD-10)) aged 20 to 60 years. The severity of depressive symptoms was assessed using the Hamilton Depression Scale (HDRS-17) and the Clinical Global Impression Scale - Severity (CGI-S) (before and on the 14th and 28th days of therapy). The presence of remission was evaluated using the HDRS-17 scale (for the presence of remission, a score of 7 or less after 28 days of therapy was taken) and the CGI-S scale (for the presence of remission, a score of 2 or less after 28 days of therapy).

Genotyping of polymorphic variants rs6265 of the BDNF gene, rs6330 of the NGF gene, and rs3924999 of the NRG1 gene was performed by real-time PCR on a Step One Plus™ Real-Time PCR System (Applied Biosystems, USA) using TaqMan1 Validated SNP Genotyping Assay (Applied Biosystems, USA). Statistical processing of the results was performed using the SPSS program, version 20.0.

Results

Our study showed that rs3924999 of the NRG1 gene is associated with clinical remission rated on the HDRS-17 scale on day 28 of antidepressant therapy. In patients without remission, the frequency of occurrence of the AA genotype was statistically significantly higher (27.3%) than in patients who achieved clinical remission (11.6%), OR = 2.86, 95% CI = 1.26-6.48, p = 0.029. Thus, we can conclude that the carriage of the AA genotype of the polymorphic variant rs3924999 of the neuregulin-1 gene NRG1 is a factor preventing the timely formation of clinical remission in patients with depressive disorder receiving antidepressant therapy.

Acknowledgment

Supported by the RFBR No. 17-29-02205.

References

[1] Wagner S. et al. (2018) Plasma brain-derived neurotrophic factor (pBDNF) and executive dysfunctions in patients with Major Depressive Disorder. World J Biol Psychiatry. 15:1-38.

[2] Ochi T., et al. (2019) Investigating the potential role of BDNF and PRL genotypes on antidepressant response in depression patients: a prospective inception cohort study in treatment-free patients. Journal of Affective Disorders. 259:432-439.

Symposium
Bioinformatics
and systems biology of plants

DOI 10.18699/BGRS/SB-2020-186

Simulation climatic model for time to flowering in wild chickpea

Andrey Ageev
SPbPU, St.Petersburg, Russia
andreyageev1@mail.ru

Jens Berger
CSIRO, WA, Australia
Jens.Berger@csiro.au

Eric Bishop-von Wettberg
UVM, VT, USA
Eric.Bishop-Von-Wettberg@uvm.edu

Abdullah Kahraman
Harran University, Sanliurfa, Turkey
kahraman55@hotmail.com

Abdulkadir Aydogan
CRIFC, Ankara, Turkey
akadir602000@yahoo.com

Douglas Cook
UC Davis, CA, USA
drcook@ucdavis.edu

Sergey Nuzhdin
SPbPU, St.Petersburg, Russia
USC, LA, CA, USA
snuzhdin@usc.edu

Maria Samsonova
SPbPU, St.Petersburg, Russia
m.samsonova@spbstu.ru

Konstantin Kozlov
SPbSPU, St.Petersburg, Russia
kozlov_kn@spbstu.ru
snuzhdin@usc.edu

***Abstract* — Accurate prediction of flowering time helps breeders to develop new varieties that can achieve maximal efficiency in changing climate. We have developed a methodology for construction a simulation model for time to flowering in which a function for daily progression of the plant from one to the next phenologic phase is obtained in analytic form by stochastic minimization. For demonstration, we build a simulation model for time to flowering in the dataset of wild chickpeas assembled recently.**

Keywords — wild chickpea, simulation, climatic factors

Motivation and aim

Motivation

Chickpea (*Cicer arietinum* L.), which was originally domesticated in Southeastern Turkey, is the second most cultivated grain legume crop, grown in more than 50 countries of the world. Accurate prediction of crop flowering time is required for reaching maximal farm efficiency. Several models developed to accomplish this goal are based on deep knowledge of plant phenology, requiring large investment for every individual crop or new variety [1-2].

Aim

Mathematical modeling can be used to make better use of more shallow data and to extract information from it with higher efficiency. Earlier we have developed a non-linear regression model that predicted time to flowering based on climatic factors averaged over different periods of time [2]. Here to better interpret factors' impacts we build a simulation model for time to flowering that uses daily values of minimal and maximal temperature, solar radiation, precipitation and day length.

Methods

Plant material

Cultivars of chickpea, *Cicer arietanum*, are currently being improved by introgressing wild *C. reticulatum* biodiversity with very different flowering time requirements. The dataset consists of wild chickpea (*C. reticulatum* L. and *C. echinospermum*). Accessions were collected at 21 sites in five regions in Turkey by von Wettberg et al. [3] and planted in climatically distinct sites in Turkey (Sanliurfa and Ankara). Being grown in contrasting environments the phenotype data on time to flowering is highly diverse – time to flowering ranges from 117 to 221 days. We build a model for samples sown in autumn (2174 samples).

Climatic data

The data on climatic conditions for each day in the period of field experiments were taken from publicly available site https://rp5.ru/Weather_in_the_world and NASA.

Mathematical model

The analytic representation of a function for daily progression of the plant from one to the next phenologic phase is build using a formal approach called Grammatical Evolution (GE) [4] and Differential Evolution Entirely Parallel (DEEP) method as stochastic minimization [5]. Consequently, a function is constructed as a superposition of arithmetic operations of expressions: X, (X - Const), 1/(X – Const), where X is a climatic factor and Const is a threshold value specific to each factor. DEEP is used to minimize the sum of squared differences between model solution and data.

Results

Model adaptation

The cross-validation and testing was performed to select the best and reliable model. The dataset was split into training, validation and test sets. The number of samples was 1482, 492 and 200, respectively. Optimization was performed with training set, then the model with a lowest score (18580.0) for the validation set was selected and run on a test set. According to Mann-Witney-Wilcoxon criterion the difference between mean errors for training and validation sets is not statistically significant (P=0.13). Consequently, the constructed model is validated. Mean deviation of the model solution from data for validation and test sets are 5.0 and 6.1 days, respectively. The analytic form of a function combines linear and multiplicative parts that is similar to the Soltani et al. model [1].

Impacts of climatic factors

To assess the impacts of climatic factors to the model we performed a permutation test with 100 permutations. The results are presented in Fig.1. The minimal temperature T_n has the maximal impact to the model – over 85%, the day length impact is 14.37%. The impacts of maximal temperature, solar radiation and precipitation are almost negligible – 0.02%, 0.08% and 0%, respectively. Day length has high impact on long day plants such as wild chickpea.

The high impact of minimal temperature on time to flowering emphasizes the importance of vernalization.

The obtained results show that developed methodology can be used to model wild chickpea growth.

ACKNOWLEDGMENT

Supported by the RFBR (18-29-13033).

REFERENCES

[1] Soltani, A., and Sinclair, T.R. (2011). A simple model for chickpea development, growth and yield. Field Crops Research 124, 252–260.

[2] Kozlov, K., Singh, A., Berger, J., Wettberg, E.B., Kahraman, A., Aydogan, A., Cook, D., Nuzhdin, S., and Samsonova, M. (2019) Nonlinear regression models for time to flowering in wild chickpea combine genetic and climatic factors. BMC Plant Biology 19, 94.

[3] von Wettberg, E.J.B., Chang, P.L., Başdemir, F., Carrasquila-Garcia, N., Korbu, L.B., Moenga, S.M., Bedada, G., Greenlon, A., Moriuchi, K.S., Singh, V., et al. (2018) Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. Nature Communications 9.

[4] O'Neill, M., and Ryan, C. (2001). Grammatical evolution. IEEE Transactions on Evolutionary Computation 5, 349–358.

[5] Kozlov, K., Samsonov, A.M., and Samsonova, M. (2016). A software for parameter optimization with Differential Evolution Entirely Parallel method. PeerJ Computer Science 2, e74.

DOI 10.18699/BGRS/SB-2020-187

Analysis of repeatomes in Cannabaceae family

Julia Bocharkina
Laboratory of Applied Genomics and Crop Breeding
All-Russia Research Institute of Agricultural Biotechnology
Moscow, Russia
Life sciences
Skolkovo Institute of science and technology
Moscow, Russia
julia.bocharkina@skoltech.ru

Olga Razumova
Laboratory of Applied Genomics and Crop Breeding
All-Russia Research Institute of Agricultural Biotechnology
Moscow, Russia
razumovao@gmail.com

Gennady Karlov
Laboratory of Applied Genomics and Crop Breeding
All-Russia Research Institute of Agricultural Biotechnology
Moscow, Russia
karlovg@gmail.com

***Abstract* — Analysis of repeatomes in the *Cannabaceae* family was done. There were found family-specific, species-specific and sex-specific clusters. Primers were created and checked.**

***Keywords* — NGS, repeatome, plants, Cannabaceae, sex**

Introduction

Plant genomes are characterized by large amount of repetitive DNA and this amount is known as repeatomes of species [1]. There are 3 major types of repetitive DNA: transposable elements (TE), tandem repeats (TR), ribosomal DNA. Transposable elements have a relatively high duplication rate [2]. These copies can impact on transcription, expression of the genes, also they can disrupt genes[3, 4, 5]. Thus, repeats can be very heterogeneous, in addition, they can be unique, low- or high-copy. They may differ in the length of the motifs in the genome [6]. Many families of repeats DNA are not yet known. Also, the role of these sites in the genome is not fully understood. They are known to play an important role in stabilization and maintenance chromosome structures, they are also involved in chromosome recognition and their correct division during mitosis and meiosis. [7]. It is shown that species differentiation in plants often associated with rapid changes in DNA-repeats [2]. Our aim in this research was to study of repeatomes from *Cannabaceae* family.

The family *Cannabaceae* contains at least 3 agricultural-significant species: *Cannabis sativa* L., *Humulus lupulus* L., *Humulus japonicus* L.. Hemp's (*Cannabis sativa* L.) has 2n = 20 and female haploid genome size is 818 Mb, but male haploid genome size is 843 Mb [9]; these facts show us that *Cannabis sativa* L. has heterogenous X and Y chromosomes in such a way that Y chromosome is bigger than X. *Cannabis sativa* L. has male-determining Y chromosome [9,10,11].

There is an interesting situation with hops (*Humulus*). Common hop (*Humulus lupulus* L.) contains 2n = 20 with the XY system, but sex-determination occurs according to the A:X ratio [8]. *Humulus japonicus* L. contains XY_1Y_2 and the female plant has 2n = 16, the male plant has 2n = 17 [12]. Why sex determination is so different within one family?

Therefore, one of the important tasks of modern biology is to study repeatomes from the sex-evolution, structural bioinformatics and cytogenetics points of view.

Materials and methods

For the research, there were used *Humulus lupulus* plant samples from Moscow region, *H. japonicus* plants cultivar "Samurai" Gavrish, *Cannabis sativa* L. plant ("Zenitsa" cultivar) DNA were purified from leaves using the Doyle & Doyle method with modifications [13]. DNA libraries were prepared according to the instructions of commercial kits. For all plants, there were done one full cycle of Illumina MiSeq sequencing.

Sequencing reads were analyzed by quality control tool FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Quality filtering of the resulting raw fastq reads, adaptors trimming, filtering out short or unpaired sequences and trimming reads to lengths using the Trimmomatic tool. [14]. Then, the two-resulting pair-term sequencing files were combined using the FASTQ interlacer tool, which allows you to automatically filter paired and unpaired sequences. A file with paired reads was used for further analysis.

Bioinformatic analysis of the repeatomes was carried out on a local server with 28 nuclear processors and 128 GB of RAM, running the Linux operating system using RepeatExplorer [15]. Clusters containing at least 0.01% of all clustered reads were characterized using RepeatMasker [16] using the Repbase database [17]. The determination of the unit of tandem repeat monomer is performed by RepeatExplorer.

Pairwise scatterplots were built in R using the ggplot2 package using custom scripts [18].

PCRs were done, conditions were optimized: annealing temperature 60°, 35 cycles.

Results

Identification of common and species-specific repeats were done with usage strategy of comparative clusterization according to the following scheme: a set of *Humulus lupulus* L. sequences (female + male plant) and a set of *Humulus japonicus* L. sequences (female + male plant).

The majority of species-specific sequences for both genomes are non-classified repeats-families. And a large amount of known species-specific repeats related to the Ty3/Gypsy family(~8%).

Finally, 88 species-specific clusters for the common hop and 10 species-specific clusters for Japanese hop were identified. The highest-copy clusters were selected as candidates for the creating of species-specific markers.

There were not completely sex-specific clusters. This may be due to several factors, for example, the number of reads used for analysis. We compared two different species, one can notice a stronger divergence of the repeating sequences for *Humulus japonicus* L., which may be evidence of an earlier

discrepancy between the ancestors of *Humulus japonicus* L. compared to *Humulus lupulus* L. As a criterion for the selection of sex-specific clusters, we were guided by two parameters: the copy number of the cluster and the difference in the number of sequences of the male and female genomes in this cluster. So, for *Humulus japonicus* L., 8 of the most replicated clusters were selected with a minimum value of the difference in the number of reads between the male and female genomes. For *Humulus lupulus* L. there were chosen 4 clusters for further work.

Primers were created and, finally, we had 17 pairs of primers. We decided to check our primers with doing PCR and electrophoresis It was shown that most of the primers are not species-specific.

PCR was done and it was shown that we have 3 categories of primers:

1) Universal for *Cannabaceae* family (for both hops and hemp);

2) Universal for *Humulus lupulus* and *Humulus japonicus*;

3) Species-specific for *Humulus lupulus* L. or *Humulus japonicus* L.

This research is in progress.

ACKNOWLEDGMENT

The work is supported by RFBR No. 20-316-70018\19.

REFERENCES

[1] Maumus F., Quesneville H. Deep investigation of Arabidopsis thaliana junk DNA reveals a continuum between repetitive elements and genomic dark matter // PLoS One. – 2014. – T. 9. – №. 4.

[2] Lisch D. How important are transposons for plant evolution? // Nature Reviews Genetics. – 2013. – T. 14. – №. 1. – C. 49-61.

[3] Feschotte C. Transposable elements and the evolution of regulatory networks // Nature Reviews Genetics. – 2008. – T. 9. – №. 5. – C. 397-405.

[4] Faulkner G. J. et al. The regulated retrotransposon transcriptome of mammalian cells // Nature genetics. – 2009. – T. 41. – №. 5. – C. 563.

[5] Fedoroff N. V. Transposable elements, epigenetics, and genome evolution // Science. – 2012. – T. 338. – №. 6108. – C. 758-767.

[6] Wicker T. et al. A unified classification system for eukaryotic transposable elements // Nature Reviews Genetics. – 2007. – T. 8. – №. 12. – C. 973-982

[7] Neumann P. et al. Plant centromeric retrotransposons: a structural and cytogenetic perspective // Mobile DNA. – 2011. – T. 2. – №. 1. – C. 4.

[8] Winge O. The nature of sex chromosomes // Proc Sixth Int Congr Genet. – 1932. – T. 1. – C. 343-355.

[9] Sakamoto K. et al. Characterization; genome sizes and morphology of sex chromosomes in hemp (Cannabis sativa L.) // Cytologia. – 1998. – T. 63. – №. 4. – C. 459-464.

[10] Ming R., Bendahmane A., Renner S. S. Sex chromosomes in land plants // Annual review of plant biology. – 2011. – T. 62. – C. 485-514.

[11] Divashuk M. G. et al. Molecular cytogenetic characterization of the dioecious Cannabis sativa with an XY chromosome sex determination system // PloS one. – 2014. – T. 9. – №. 1.

[12] Aleksandrov O. S., Divashuk M. G., Karlov G. I. Development of a sex-specific molecular marker for Japanese hop Humulus japonicus Siebold & Zucc // Russian Journal of Genetics. – 2011. – T. 47. – №. 8. – C. 1016.

[13] Doyle J. J. Isolation of plant DNA from fresh tissue // Focus. – 1990. – T. 12. – C. 13-15.

[14] Bolger A. M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data // Bioinformatics. – 2014. – T. 30. – №. 15. – C. 2114-2120.

[15] Novák P. et al. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads // Bioinformatics. – 2013. – T. 29. – №. 6. – C. 792-793.

[16] Smit A. F. A., Hubley R., Green P. RepeatMasker Open-4.0. 2013–2015. – 2015.

[17] Jurka J. et al. Repbase Update, a database of eukaryotic repetitive elements // Cytogenetic and genome research. – 2005. – T. 110. – №. 1-4. – C. 462-467.

[18] Wickham H. ggplot2: elegant graphics for data analysis. – Springer, 2016.

DOI 10.18699/BGRS/SB-2020-188

Genome-wide prediction of transcription start site in four conifer species

Eugeniia I. Bondar
Laboratory of Forest Genomics
Siberian Federal University
Krasnoyarsk, Russia
Laboratory of Genomeic Research and Biotechnology
FRC KSC SB RAS
Krasnoyarsk, Russia
bondar.ev@ksc.krasn.ru

Dmitry A. Kuzmin
Laboratory of Forest Genomics
Siberian Federal University
Krasnoyarsk, Russia
Department of High Performance Computing
Siberian Federal University
Krasnoyarsk, Russia
dm.kuzmin@gmail.com

Konstantin V. Krutovsky
Laboratory of Forest Genomics
Siberian Federal University
Krasnoyarsk, Russia
Department of Forest Genetics and Forest Tree Breeding
Georg-August University of Göttingen
Göttingen, Germany
Laboratory of Population Genetics
Vavilov Institute of General Genetics
Moscow, Russia
Department of Ecosystem Science and Management
Texas A&M University
College Station, TX, USA
konstantin.krutovsky@forst.uni-goettingen.de

Vadim V. Sharov
Laboratory of Genomeic Research and Biotechnology
FRC KSC SB RAS
Krasnoyarsk, Russia
Laboratory of Forest Genomics
Siberian Federal University
Krasnoyarsk, Russia
Department of High Performance Computing
Siberian Federal University
Krasnoyarsk, Russia
sharvadim07@ya.ru

Tatiana V. Tatarinova
Department of Biology
University of La Verne
La Verne, USA
Functional Genomics Group
Vavilov Institute for General Genetics
Moscow, Russia
Siberian Federal University
Krasnoyarsk, Russia
Bioinformatics Center of IITP RAS
Moscow, Russia
ttatarinova@laverne.edu

Abstract — **Current draft annotations for sequenced conifer genomes are preliminary and limited, but provide opportunities for further structural and functional analysis. We attempted to improve the existing genome annotations by marking 5'-UTRs in the four conifer species *Pinus taeda, Picea glauca, Picea abies* and *Larix sibirica*. Prediction of transcription start sites (TSS) was performed on the promoter sequences of genes with RNA or protein support using TSS prediction program TSSPlant. The distribution of 5'-UTR lengths from the annotations of several model plants was used to select the best prediction per gene. Frequency of TATA(A/T)A(A/T) motif in the predicted TSS-centered promoter regions showed a pronounced peak around 60 bp upstream of TSS.**

Keywords — 5'-UTR, plant promoters, conifer genome, GC_3 biology

Introduction

Conifers are an ancient group of plants represented by more than 600 species that play major role in boreal forest ecosystems. Due to the enormous size and highly repetitive nature deciphering conifer genomes takes more time and effort than many other plant species. In the last years several mega-genomes of conifer species were sequenced and assembled to the draft state. Their current annotations are preliminary and limited, but provide opportunities for structural and functional analysis. Understanding of regulatory relationships between genome elements requires information on promoter sequences, which are usually located directly upstream or at the 5' end of the transcription start site (TSS). Here we present our attempts to improve the existing genome annotations by predicting TSSs and marking 5'-UTRs in the four relatively recently published species *Pinus taeda, Picea glauca, Picea abies*, and *Larix sibirica*.

Materials and Methods

A. Genome Assemblies and Annotations

The following genome assemblies and annotations of four conifer species were used: Pita v2_01 for *Pinus taeda* [1, 2], PG29 v3.0 for *Picea glauca* [3, 4], Pabies v1.0 for *Picea abies* [5], and genome assembly under GenBank accession NWUY0000000000 (BioProject PRJNA393226) together with the unpublished draft annotation data for *Larix sibirica* [6].

B. Gene Filtering

To filter out possible pseudogenes and putative predicted coding sequences that do not present functional genes all gene models retrieved from genomic annotations were aligned against the database of RNA-seq data, including ESTs and TSAs, of a corresponding species using hisat2 alignment program. To additionally verify selected gene models we aligned corresponding protein products to Refseq plant protein sequences using blastp program.

C. Transcription Start Site (TSS) Prediction

Prediction of putative TSSs was performed in the promoter sequences of selected genes, which were defined as regions of -1000 and +250 bp around the start codon, using TSSPlant program [7] and selection of the best prediction based on the distribution of 5'-UTR lengths from the annotations of several model plants.

D. GC_3, CG-skew and Frequency of TATA motif

Nucleotide frequency analysis of promotors was performed on TSS-centered sequences (-1000 and +200 bp around TSS). Frequencies of CA and TATA motifs were calculated with a sliding window (width = 40 bp, increment step = 10 bp) using stringr package for R. CG-skew of a given sequence was defined as a proportion (C-G)/(C+G) and

calculated with sliding window width of 50 bp and window increment step of 10 bp along the promoter sequence. GC_3 was calculated using gene sequences with removed introns (only joined exons were used) and seqinr package for R.

Results

Aligning to RNA-seq, ESTs, and Refseq protein data allowed to retrieve 9260 evidence supported gene models for *P. taeda*, 16853 for *P. glauca*, 7587 for *P. abies*, and 23077 for *L. sibirica* (Table I). For these genes (both TATA and TATA-less) TSSs were predicted using TSSPlant algorithm. Frequency of TATA(A/T)A(A/T) motif in the predicted TSS-centered promoter regions showed pronounced peak around 60 bp upstream of TSS (Fig. 1A). All four species have similar GC_3 distribution with the average around 0.3 and GC_3 gradient that gradually decreases after 250 bp from the start of coding sequence (Fig. 1B). In contrast to what was observed in *A. thaliana, O. sativa*, and some other plant species, CG-skew in four conifer species exhibited distinct decline right before the TSS (Fig. 1C).

TABLE I. SUMMARY OF ASSEMBLIES AND ANNOTATIONS

Assembly	Assembly N50	Sum length, Gb	Number of genes	RNA/Refseq supported genes
P. taeda v2_01	100 218	20.43	36 732	9 260
P. abies v1.0	7 747	9.9	58 587	10 434
PG29 v3	34 405	20.8	103 694	16 839
L. sibirica	3 098	5.6	50 163	23 077

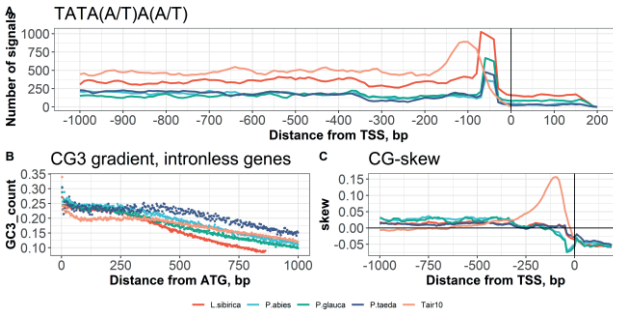


Fig. 1. Frequency of TATA(A/T)A(A/T) motif in the TSS-centered promoter region (A), GC3 gradient of coding sequences (B), and CG-skew around TSS (C).

Acknowledgments

We thank Dr. N. V. Oreshkova and Dr. S. I. Feranchuk for help with sequencing and draft annotation of the *Larix sibirica* genome, respectively.

References

[1] D. B. Neale, et al. "Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies", Genome biology 15 no. 3 , March 2014, p. R59.

[2] A. V. Zimin, et al. "An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing", GigaScience 6, no. 1 January 2017, p. giw016.

[3] I. Birol, et al. "Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data", Bioinformatics 29, no. 12, Jun 2013, pp. 1492-1497.

[4] R. L. Warren, et al. "Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism", The Plant Journal 83, no. 2, May 2015, pp. 189-212.

[5] B. Nystedt, et al. "The Norway spruce genome sequence and conifer genome evolution", Nature 497, no. 7451, May 2013, pp. 579-584.

[6] D. A. Kuzmin, et al. "Stepwise large genome assembly approach: a case of Siberian larch (*Larix sibirica* Ledeb)", BMC bioinformatics 20.1, Feb 2019, 37.

[7] I. A. Shahmuradov, R. K. Umarov, V. V. Solovyev, "TSSPlant: a new tool for prediction of plant Pol II promoters", Nucleic acids research, 5;45(8), May 2017, pp. e65-e65.

DOI 10.18699/BGRS/SB-2020-189

Features of the organization of bread wheat 5BS chromosome region carrying the leaf rust resistance gene Lr52

Maria Bragina
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk,
Russia koltunova@bionet.nsc.ru

Dmitriy Afonnikov
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
ada@bionet.nsc.ru

Elena Salina
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk,
Russia salina@bionet.nsc.ru

***Abstract* — Despite the reference wheat genome sequencing and annotation, the search and chromosome localization of economically valuable traits loci in the wheat genome remains relevant. We reported the sequence structure of the 26.6-Mb wheat chromosome 5BS subtelomeric region associated to leaf rust resistant genes through targeted sequencing. We show that this chromosomal region of leaf rust resistance and susceptible wheat lines were efficiently captured with sufficient coverage for the annotation of the sequenced region.**

Keywords* — *bread wheat genome; chromosome 5BS; Lr52 gene; targeted sequencing

Introduction

The wheat reference cultivar 'Chinese Spring' genome published by the International Wheat Genome Sequencing Consortium (IWGSC, 2018) and SNP markers associated with economically valuable genes in wheat varieties and lines will contribute to solving important problems in genetics and breeding. To date, more than 30 genes that control a number of morphological and quantitative traits, resistance to abiotic and biotic environmental factors have been mapped on wheat chromosome 5B. Of particular interest is the distal region of its short arm (5BS) which bears a number of agronomically important genes, including genes conferring resistance to fungal diseases (*Lr52*, *LrK1*, *Yr47*, *Snn3*, *Pm30*, *H31*, QTL for stem rust, rhizoctonia net blotch and loose smut of wheat). The mapping of DNA markers linked to the resistance genes on the 5BS reference pseudomolecule (RefSeq v1.0) allows the region where these genes reside to be identified. Our study was aimed to approbate targeted sequencing for the structure analysis of chromosome 5BS region with a length of about 26.6 Mb using lines with and without the *Lr52* gene.

Materials and methods

Mapping population F4 (lines from the cross LrW(52) × hybrid_215) was developed for chosen plants which are differ to resistant genes against fungal diseases located on target region of 5BS. We selected 5 plants with *Lr52* and 5 plants without *Lr52* according to data of KASP and SSR genotyping together with screening for resistance (benzimidazole (0.035 % w/v)). Genomic DNA was isolated from the young leaves of individual plants using the Kleargene plant 96-well plate DNA extraction kit (LGC Group), following the manufacturer's protocol.

Target sequencing was performed using the SeqCap EZ Target Enrichment System (Roche) to enrich a 26Mb of 5B chromosome subtelomeric region. The Roche company developed probes to cover the region under study on the basis of the reference genome. As a result, the maximum size of covered regions was about 8 Mb, mainly due to repeats. The 150-bp pairedend sequencing of the obtained library was performed on an Illumina NextSeq 550 platform (ICG SBRAS).

Results and Discussion

The wheat lines of population F4 (lines from the cross LrW(52) × hybrid_215) were captured with probes developed by Roche to cover the region under study on the basis of the reference genome. As a result of sequencing, an average of 4.28 Gb and 3.51 Gb of reads was obtained for the wheat lines with *Lr52* and without *Lr52*, respectively. The total size of data obtained was about 39 Gb. The proportions of sequences with Q>30 were 78.95 and 78.52 %, and the mean quality score Q values were 31.44 and 31.35 for the wheat lines with *Lr52* and samples without *Lr52*, respectively. After preprocessing and quality control using the FASTXtoolkit utility, the reads obtained were mapped onto the studied region of the wheat reference genome using the BWA-MEM program (Li and Durbin, 2009). On average, 70% of the reads were not mapped, 7-9 % were mapped one time and 20% more than once for samples with and without *Lr52*. In the target regions, 10 wheat plants had narrow mean coverage ranges, 40× and 48×, for the leaf rust resistant and susceptible samples, respectively.

To visualize the mapping of reads to the reference genome, the studied region was divided into 100 bases long segments, each of which was assigned a different color depending on the sequence matching (Fig 1).

Fig. 1. Study of the gene *Lr52* localization region.

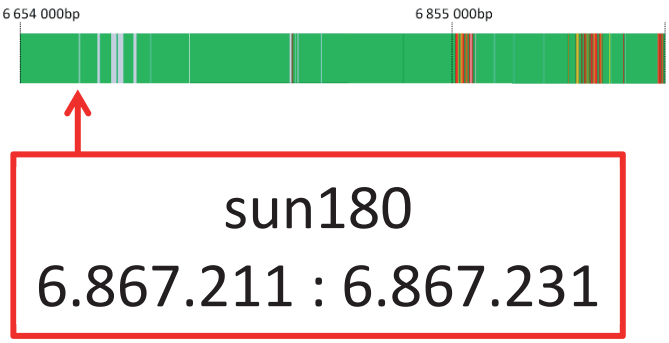


Fig. 2. Example of the reads mapping to the wheat reference genome on a plot with coordinates 6 654 000bp - 6 956 436bp. Green areas indicate the coincidence of the reference genome sequences, plants with Lr52 and 5 plants without Lr52; red color means the coincidence of the reference genome sequences and rust-susceptible plants without Lr52 gene and the absence of mapped sequences in this region in resistant samples; white - the gaps of a reference sequence in this region. Also the sun180 marker, which co-segregates with the Lr52 gene is indicated.

The 26.6Mb region was analyzed in NCBI, Ensembl Plants, GrainGenes, UniProt, PlantPAN 3.0, PlantPRe, PRGdb, psRNAtarget, miRBase, PmiRExAt databases. In result about 20.000 SNP were found to be dispersed in the 5BS distal region About 500 candidates for major resistance genes

with the domains NBS, CC, LRR, Tm and kinase domains were revealed and about 30.000 potential exones, 5.500 putative genes and 400 miRNA were found. In the 6.6Mb – 7Mb region several microRNAs were found only in samples resistant to leaf rust, the sequences of which differed from the sequences of the reference genome and samples without *Lr52* gene by SNP that are located in the miRNA stem-loop. Next, an automatic annotation was carried out using the FGENESH 2.6 program, which revealed differences between the samples in the area with 6.85Mb – 6.86Mb coordinates. For samples without *Lr52* gene, a potential gene is located in this segment and there is no coincidence in databases. In addition, samples with *Lr52* gene contained a deletion in the 6.91Mb – 6.93Mb region in which samples without *Lr52* gene had annotated *TRAESCS5B02G005400* gene that encodes a protein including two putative kinase domains. This gene is homologous to wheat stem rust resistant gene *Sr60* from *Triticum monococcum* and *Sr60* has a deletion including the last two exons in Chinese Spring. So we may propose some versions of leaf-rust resistance mechanism.

Acknowledgment

Supported by the RFBR (17-04-00507).

References

[1] Li H., Durbin R. Fast and accurate short read alignment with BurrowsWheeler transform. Bioinformatics. 2009;25(14):1754–1760.

[2] The International Wheat Genome Sequencing Consortium. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Sci. 2018;361(6403).

DOI 10.18699/BGRS/SB-2020-191

Genetic regulation of wheat inflorescence development

Oxana B. Dobrovolskaya
VNIIKR, Moscow region, Ramenskoe distinct, Bykovo, Russia
ICG SB RAS, Novosibirsk, Russia,
oxanad@bionet.nsc.ru

Alina E. Dresvyannikova
ICG SB RAS, Novosibirsk, Russia
alinka.dresvyannikova@gmail.com

Petr Martinrk
Agrotest Fyto, Ltd,
Kroměříž, Czech Republic
martinek@vukrom.cz

***Abstract* — Identification of genetic factors that control inflorescence development of important crops and study on their structure and functions have fundamental and practical importance. The obtained results showed that different genetic loci take part in the spikelet meristem determinacy of the genuine branching and false-true spike ramification wheat non-standard morphotypes.**

Keywords* — *wheat, non-standard spike morphotypes, inflorescence development, meristem identity, meristem determinacy

Motivation and aim

Motivation

Investigation of the structural and functional organization of genetic determinants of plant development is an actual problem of modern plant genetics and developmental biology.

Aim

The aim of this work is to study on genetic regulation of wheat spikelet and floret meristem activities, namely, the establishing of identity and determinacy of wheat inflorescence meristems.

Methods

Accessions with abnormalities in spike development from the ICG SB RAS wheat collection served as a unique experimental model in this research. A set of classical and modern approaches of genetics and developmental biology, including high-throughput genotyping, light and electron microscopy methods and modern bioinformatic approaches was used.

Results

Using DArTseq genotyping, wheat chromosome maps (genome BBAA) were constructed, quantitative trait loci (QTLs) that determine the formation of non-standard spike morphotypes associated with abnormalities in the establishing of identity and determinacy of inflorescence meristems were identified. The formation of the false true spike ramification morphotype is controlled by the *qshr2-2A* locus of chromosome 2AL, and by two loci (chr.7A) with minor effects. It was found that, along with the *wfzp-A.Bh1* gene, the formation of additional spikelets of the wheat genuine branching morphotype is influenced by QTLs on chromosomes 7AL and 5BL. Therefore, different genetic loci take part in determinacy of the spikelet meristem of the genuine branching and the false-true spike ramification morphotypes.

Acknowledgment

This work was supported by the RFBR (18-04-00483).

DOI 10.18699/BGRS/SB-2020-192

EIN3 binding site architecture guides transcriptional response to ethylene in Arabidopsis

Vladislav Dolgikh
ICG SB RAS, Novosibirsk, Russia
volleydollmc1@gmail.com

Victor Levitsky
ICG SB RAS, Novosibirsk, Russia
levitsky@bionet.nsc.ru

Dmitry Oshchepkov
ICG SB RAS, Novosibirsk, Russia
diman@bionet.nsc.ru

Elena Zemlyanskaya
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
ezemlyanskaya@bionet.nsc.ru

***Abstract* — EIN3 transcription factor is the master regulator of gene expression in response to plant hormone ethylene. The structure of EIN3 binding site (EBS) was determined using *in vitro* binding assays. However, the role of EBS architecture in EIN3 binding at the whole-genome level is poorly understood. Here we demonstrate that an inverted repeat of EBS with the overlap of the motifs is a canonical EIN3 target in the Arabidopsis genome. We show that this EBS architecture provides a more pronounced response to ethylene than either a single EBS or EBS repeats with a spacer. These findings can be used to develop a new genetic sensor for highly sensitive detection of ethylene signaling in Arabidopsis.**

Keywords* — *bioinformatics, transcription factor, ETHYLENE-INSENSITIVE3, EIN3 binding site (EBS), ChIP-seq, RNA-seq

Motivation and aim

Motivation

ETHYLENE INSENSITIVE 3 (EIN3) transcription factor is the master regulator of gene expression in response to plant hormone ethylene that guides plant growth under stress conditions. EIN3 is a transcriptional activator that binds a short nucleotide sequence referred to as EBS to induce transcription [1]. *Arabidopsis thaliana* reporter *EBS:GUS* driven by such EIN3 binding site [2] is widely used as a sensor for detection of ethylene signaling. EIN3 tends to bind DNA as a homodimer and it has been shown recently that EBS inverted repeats with a spacer of 10 bp provide EIN3 binding *in vitro* with a higher affinity than a single EBS [3]. However, the role of EBS architecture in EIN3 functioning was not investigated on the whole genome level.

Aim

Here we accomplish a systematic bioinformatics analysis of EIN3 bound sequences in Arabidopsis genome to shed light on molecular mechanisms utilized for regulation of transcriptional response to ethylene in plants.

Methods

We used publicly available ChIP-seq data on EIN3 binding, RNA-seq data on ethylene-induced transcriptomes in Arabidopsis seedlings [4] and published DAP-seq data [5]. We used Homer [6] for *de novo* motif search in the peaks, and MCOT [7] for enrichment analysis of EBS repeats. Associations of EBS configurations with peaks and genes features were estimated with Fisher's exact test.

Results

We discovered a previously unknown EBS architecture that is enriched in EIN3 bound sequences to a much greater extent than a single EBS motif. This new configuration is a head-to-head inverted repeat of EBS-like sequences with 1 bp overlap referred to as 2EBS(-1). It is noteworthy that none of the EBS repeats with spaced motifs (including the one with 10 bp spacer) was enriched in EIN3 bound sequences compared to a random expectation. We also demonstrated that unlike a single EBS motif, 2EBS(-1) repeat was significantly associated with a sustained profile of EIN3 binding, i.e. it facilitated EIN3 binding regardless of the duration of ethylene treatment. Based on these findings we consider that 2EBS(-1) is a canonical EIN3 binding site in *A. thailana* genome. We further showed that of all EBS configurations under study only 2EBS(-1) was significantly associated with transcriptional response of EIN3 targets to ethylene treatment. Moreover, it tended to cause a more pronounced transcriptional response than other EBS configurations. These findings can be used to design a new genetic sensor for highly sensitive detection of ethylene signaling. Taken together, this work provides new insight on the molecular mechanisms utilized for regulation of transcriptional response to ethylene in plants.

Acknowledgment

Supported by the RSF (20-14-00140) and State Budget Project (0324-2019-0040).

References

[1] Kosugi S. and Ohashi Y. (2000) Cloning and DNA-binding properties of a tobacco Ethylene-Insensitive3 (EIN3) homolog. Nucleic Acids Res. 28(4):960-967.

[2] Stepanova A.N. et al. (2007) Multilevel interactions between ethylene and auxin in Arabidopsis roots. Plant Cell. 19(17):2169-2185.

[3] Song J. et al. (2015) Biochemical and Structural Insights into the Mechanism of DNA Recognition by Arabidopsis ETHYLENE INSENSITIVE3. PLoS One. 10(9):e0137439.

[4] Chang K.N. et al. (2013) Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. Elife. 2:e00675.

[5] O'Malley R.C. et al. (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell. 165(5):1280-1292.

[6] Heinz S. et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. Mol. Cell. 38(4):576-589.

[7] Levitsky V. et al. (2019) A single ChIP-seq dataset is sufficient for comprehensive analysis of motifs co-occurrence with MCOT package. Nucleic Acids Res. 47(21):e139.

DOI 10.18699/BGRS/SB-2020-193

Targeted mutagenesis of the *HvMyc2* and *HvAnt2* genes in *Hordeum vulgare* L.

Anastasiya Egorova
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
IPK, Gatersleben, Germany
egorova@bionet.nsc.ru

Ksenia Strygina
VIR, St.Petersburg, Russia
k.strygina@vir.nw.ru

Olesya Shoeva
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
olesya_ter@bionet.nsc.ru

Christian Hertig
IPK, Gatersleben, Germany
hertig@ipk-gatersleben.de

Iris Koeppel
IPK, Gatersleben, Germany
koeppel@ipk-gatersleben.de

Stefan Hiekel
IPK, Gatersleben, Germany
hiekel@ipk-gatersleben.de

Alexander Vikhorev
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
vikhorev@bionet.nsc.ru

Sophia Gerasimova
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
gerson@bionet.nsc.ru

Jochen Kumlehn
IPK, Gatersleben, Germany
kumlehn@ipk-gatersleben.de

Elena Khlestkina
VIR, St.Petersburg, Russia
khlest@bionet.nsc.ru

***Abstract* — The Myc2 and Ant2 genes of barley encode bHLH transcription factors that activate transcription of anthocyanin biosynthesis structural genes. It is shown, that the HvMyc2 and HvAnt2 genes are implicated in the pigmentation of barley grains. The present work aims to introduce mutations into the HvMyc2 and HvAnt2 genes in order to confirm their function in barley grain color formation.**

Keywords — barley, grain color, targeted mutagenesis

Motivation and aim

Motivation

Grains of barley (*Hordeum vulgare* L.) can have different color which is directly connected to accumulation of different pigments. Blue and purple color are caused by accumulation of anthocyanins in aleurone and pericarp, respectively. In previous studies, candidate genes which could be responsible for barley grain color, were identified; the blue color of grain aleurone depends on the *HvMyc2* gene [1] and the purple color of grain pericarp on the *HvAnt2* gene [2]. These genes encode bHLH/MYC proteins which, together with the MYB and WD40 factors, activate structural genes of the anthocyanin biosynthesis pathway.

Aim

The aim of the present study is to use the Cas9/gRNA site-directed mutagenesis system to generate barley plants bearing mutations in their *Myc2* and *Ant2* genes so as to confirm the functions of these genes in grain color formation and to create pigmented barley featuring competitive agricultural performance.

Methods

Putative target regions were pre-selected using online tools like DeskGen [3], CRISPR-P [4] and WuCRISPR [5]. A final selection was then made based upon modelling of the secondary structure of respective guide-RNAs (gRNAs) using RNAfold [6]. A modular vector system [7] that greatly facilitates complex cloning steps using IIS restriction enzymes was used to create *cas9*/gRNA transformation plasmids to be used for site-directed mutagenesis of the target genes. First, intermediate constructs were created with up to four gRNAs each being driven by the RNA-Polymerase III U6 promoter of wheat. Next, these assemblies were combined with the maize codon-optimized *cas9* nuclease gene being under control of the maize Ubi-1 promoter. A protoplast assay was performed in order to estimate the constructs' efficiency in triggering targeted mutagenesis. To this end, barley leaf mesophyll protoplasts were transfected with the above constructs, protoplast genomic DNA was isolated and deep-sequencing of amplicons from the target regions was performed [8]. Finally, the *cas9*/gRNA-carrying constructs were introduced into the p6i binary vector (DNA Cloning Service, Hamburg, Germany) for subsequent Agrobacterium-mediated stable transformation of barley using immature embryos [9].

Results

Two *cas9*/gRNA plasmids were constructed for two target sites of *HvMyc2* gene, each containing four times the same gRNA. These vectors showed 58% and 76% mutation efficiency in the protoplast assay. Additionally, one construct was generated containing three gRNAs for different target sites in the *HvAnt2* gene. All constructs are planned to be used for stable barley transformation and subsequent regeneration of plants carrying mutations in the two target genes. Barley accessions with colored and uncolored grains will be used as donor material for transformation.

Acknowledgment

This work is supported by the Russian Science Foundation (RSF) grant No. 16–14-00086 and by IPK Visiting Program for Scientists from Transition Countries in Europe and the former Soviet Union.

References

[1] Strygina K.V., Börner A., Khlestkina E.K. (2017) Identification and characterization of regulatory network components for anthocyanin synthesis in barley aleurone. BMC Plant Biol. 17(1):184.

[2] Shoeva O.Y., Mock H.P., Kukoeva T.V., Börner A., Khlestkina E.K. (2016) Regulation of the flavonoid biosynthesis pathway genes in purple and black grains of Hordeum vulgare. PloS ONE. 11(10):e0163782.

[3] Hough S.H. et al. (2017) Erratum to: Guide Picker is a comprehensive design tool for visualizing and selecting guides for CRISPR experiments. BMC Bioinformatics. Vol. 18, № 1. P. 202.

[4] Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K., and Chen, L. (2017) CRISPRP 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. Mol. Plant. 10:530–532.

[5] Wong N., Liu W., Wang X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. Genome Biol. 2015. Vol. 16. P. 218.

[6] Gruber A.R. et al. The Vienna RNA websuite. Nucleic Acids Res. 2008. Vol. 36. Web Server issue. P. W70-74.

[7] Koeppel I., Hiekel S., Kumlehn J. (2019) A versatile modular vector system for RNA-guided Cas endonucleases [abstract]. VISCEA International Conference Plant Transformation & Biotechnology V. 03-04.07.2019. Vienna, Austria. http://viscea.org/wp-content/uploads/2019/08/PTBT-abstract-book.pdf.

[8] Gerasimova S.V., Korotkova A.M., Hertig C., Hiekel S., Hoffie R., Budhagatapalli N., Otto I., Hensel G., Shumny V.K., Kochetov A.V.,Kumlehn J., Khlestkina E.K. (2018) Targeted genome modification in protoplasts of a highly regenerable Siberian barley cultivar using RNA-guided Cas9 endonuclease. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 22(8):1033-1039.

[9] Hensel G., Valkov V., Middlefell-Williams J., & Kumlehn J. (2008). Efficient generation of transgenic barley: the way forward to modulate plant-microbe interactions. J. Plant Physiol. 165. 71-82. 10.1016/j.jplph.2007.06.015.

DOI 10.18699/BGRS/SB-2020-194

Flowering patterns of herbaceous multi-flowered monocarpic shoots of Campanula sarmatica

Eduard Fomin
Systems Biology Department
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
fomin@bionet.nsc.ru

Tatyana Fomina
Laboratory of introduction of ornamental plants
Central Siberian Botanical Garden SB RAS
Novosibirsk, Russia
0000-0003-4724-2480

Abstract — **Flowering analysis of Campanula sarmatica, Campanulaceae family at the level of individual shoots was performed. It has been established that the flowering curves of C. sarmatica shoots are bimodal. The first peak corresponds to the flowering of the main axis, and the second to the lateral axes of the inflorescence. Flowering curves have noticeable individual differences in the heights of the main peaks, due to the difference in the number of flowers on the main axis and lateral axes on different shoots. The found mechanism of the formation of the curves makes it possible to extract the type of inflorescence (botryoid or panicle) and estimate the number of lateral axes directly from the flowering curve for any observed shoot.**

Keywords — flowering patterns, flowering model, monocarpic shoot, Campanula sarmatica, inflorescence, ornamental plants

Introduction

The main component of the reproductive success of a plant in its seasonal development is the synchrony with flowering and environmental conditions favorable for fertilization and seed setting. For plants with multi-flowered inflorescences, another important component of reproduction efficacy is the arrangement of flowers on the shoot and the sequence of their opening [1]. The evolutionary mechanisms underlying the formation of flowering patterns may be different. Most scientists interpret the observed flowering regularities as a trait of adaptation to the totality of abiotic and biotic factors evolved in intra- and interspecific competition [2]. Heterogeneous environment, climatic changes, and the seasonal activity of pollinators and phytophages injuring generative organs are considered determiners of the optimum flowering time. Also, the entrance to the generative phase of the seasonal cycle is influenced by vegetative growth parameters [3].

Studies of flowering at the organ, organismal, populational, or plant cenotic levels pursue different goals and employ different methods [4]. One of the approaches to plant development description involves phenological curves, which represent the time variations of certain traits. Flowering curves have gained the greatest acceptance. They reflect the start, mean date, maximum, duration, flowers opening synchrony, and asymmetry, and these parameters are phylogenetically contingent [5]. Most models applied to various plant taxa and life forms describe their phenology, including flowering, at the population level [6, 7].

The classical study [8] mentions that scientists give insufficient attention to the modeling of individual flowering. Although the environmental conditions modulate flowering timing [9], flowering phenology is determined primarily by genetic (physiological) factors, thus demanding research at the organismal level [6]. The phenology of flowering an individual plant is not the mere sum over its individual flowers; rather, it is determined by regulatory mechanisms more complicated than at the organ level and by more intricate ecological relationships. If only because of the longer duration of the process, it is substantially influenced by the populational and cenotic interactions with pollinators and phytophages [3]. The attention to flowering modeling at the organismal level is also associated with the recent advance of automated phenotyping systems based on pattern recognition [10, 11]. Here we use the new nonparametric computational model proposed by us earlier [12] to describe the flowering of individual shoots, individual plants or their groups in multiflorous species. The model can be regarded as data-driven statistical and/or instance-based. All structural and dynamic relationships in the model are numerically obtained by accumulation and integration of available observations, and these relationships are automatically rebuilt as soon as more data is added.

Materials and methods

The model was tested on Georgian bellflower (Campanula sarmatica Ker-Gawl.) from the collection of ornamental plants of natural flora, Central Siberian Botanical Garden, Novosibirsk (54°49′15.0″N 83°06′16.1″E). This species is an herbaceous polycarpic, which holds promise as a long-living perennial with lovely flowers [13]. It annually forms 36–90 cm high reproductive shoots bearing 15–73-cm terminal inflorescences. The inflorescence is a one-sided leafy-bracted botryoid or panicle with 2nd order lateral axes, bearing 6 to 25 flowers. The terminal flower on the main axis always opens first, and is accompanied by three or four flowers in the lower third of the inflorescence, and subsequent flowers open in a divergent way. Two to six flowers develop on lateral axes, not necessarily starting from the terminal position. The longevity of one flower is two to four days, and the flowering of a shoot lasts for about 15 days.

Phenological observations were conducted throughout the flowering season, generally at three-day intervals. Phases of development — start (petal separation), opened flower, and end (corolla withering and browning) — were recorded for each flower on a shoot. The description of inflorescence structure followed modern work [14]. At the shoot level, flowering dynamics was traced according to the following phenological phases: start (opening of the first flower), complete flowering (about one-half of flowers open and first ovaries appear), and end (the last one or two flowers). The inflorescences were described at the end of the growth season on a sample of ten shoots for three years.

The data missed in the observations were reconstructed by averaging over a large sample of the most probable values generated by the maximum likelihood method [15]. All nonlinear regression functions were obtained by lowess smoothing [16]. The modeling was done with in-house software written in C++ on the OS Linux platform.

Results

To illustrate the model, we select three panicle-type observed shoots of C. sarmatica with increasingly complicated structures of their inflorescence. We show features of flowering on axes of various orders and how the complication of the structure modifies flowering curves. The fig.1 demonstrate some of the detected dependencies. Data on axes of different orders are shown in different colors.

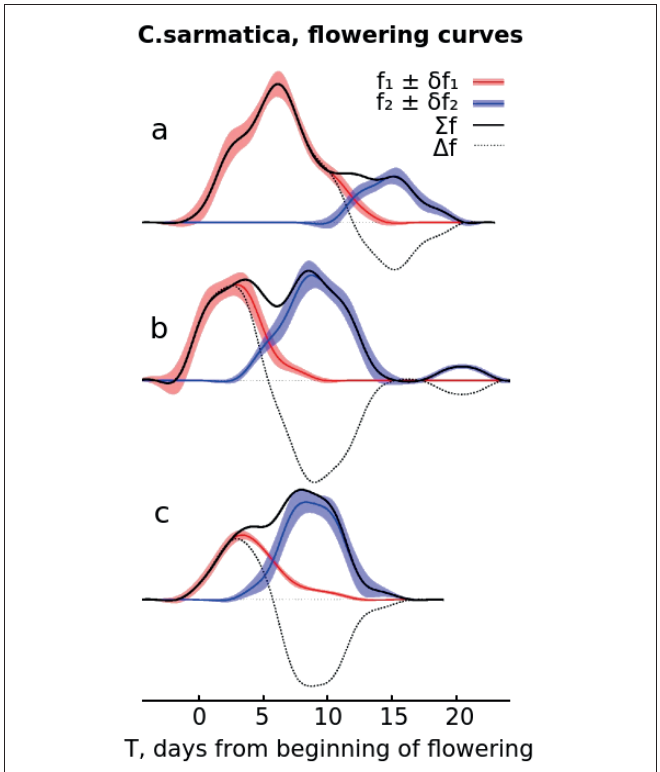


Fig. 1. Flowering curves for 1st and 2nd order axes with moving standard errors $f_i \pm \delta f_i$; the total Σf and difference Δf flowering curves. The shoot with: (a) few 2nd order flowers; (b) few 2nd order axes; (c) many 2nd order axes.

The flowering curves of observed shoots illustrate the following features of *C. sarmatica* flowering:

- The main axis and 2nd order axes flower asynchronously and form a bimodal flowering curve.
- The shape of a flowering curve depends on the degree of inflorescence branching. The curve for one-sided leafy-bracted botryoid or for panicle with occasional 2nd order flowers has a single peak. The curve for an inflorescence with few lateral axes and commensurable numbers of 1st and 2nd order flowers is bimodal.
- The flowering on 2nd order axes starts when the 1st order axis generally withers. The intervals between the peaks of flowering curves of different orders may vary among observed shoots.

Discussion

Using the new model, we confirmed the previously obtained results [15] that the flowering curve of individual C.sarmatica shoots is bimodal. The first peak reflects the opening of flowers on the main axis, and the second, on lateral axes of the inflorescence. Additionally, we found that the main variations in the flowering curve shape are determined by two factors: the degree of inflorescence branching and the time of transfer of flowering from the main axis to 2nd order axes. The former factor determines the amplitudes of the major peaks, and the latter, the distance between them. Thereby, the flowering curve for a one-sided leafy-bracted botryoid and for panicle with occasional 2nd order flowers has one peak, and for a panicle with advanced 2nd order axes, two peaks, the height of the second peak is related to the number of the 2nd order flowers. The latter makes it possible to extract the type of inflorescence (botryoid or panicle) and estimate the number of lateral axes directly from the flowering curve for any observed shoot. It should be mentioned that in the literature there are no examples of extracting any parameters of inflorescence from flowering curves, although the fact that the parameters of flowering curves are due phylogenetically was noted decades ago [5].

Thus, the new model does not only reproduce the natural variability of individual flowering and can be used to describe groups of individual plants in academic and applied research. The model allows us to discover new patterns of flowering, which expands our knowledge of the biology of individual flowering. They can be considered as phenotypic signs of plant life forms with multi-flowered inflorescences.

State Budgeted Projects No. 0312–2016–0003 and No.0324–2019-0040-C-01.

References

[1] Harder LD, Prusinkiewicz P. 2013. The interplay between inflorescence development and function as the crucible of architectural diversity. *Annals of Botany* 112:1477–1493.
[2] Snow DW. 1965. A possible selective factor in the evolution of fruiting seasons in tropical forest. *Oikos* 15:274–281.
[3] Ehrlén J. 2015.Selection on flowering time in a life-cycle context. *Oikos* 124(1):92–101.
[4] Davila-Velderrain J, Martinez-Garcia JC, Alvarez-Buylla ER. 2016. Dynamic network modelling to understanding flowering transition and floral patterning. *Journal of Experimental Botany* 67(9):2565–2572.
[5] Rathcke B., Lacey EP. 1985. Phenological patterns of terrestrial plants. *Annual Review of Ecology and Systematics* 16:179–214.
[6] Normand F, Habib R, Chadoeuf J. 2002. A stochastic flowering model describing an asynchronically flowering set of trees. *Annals of Botany* 90:405–415.
[7] Michalski SG, Durka W. 2007. Synchronous pulsed flowering: analysis of the flowering phenology in Juncus (Juncaceae). *Annals of Botany* 100(6):1271–1285.
[8] Primack RB. 1985. Patterns of flowering phenology in communities, populations, individuals, and single flowers. In: White J, eds. *The Population Structure of Vegetation. Handbook of Vegetation Science*, vol 3. Dordrecht: Springer, 571–593.
[9] Erwin J. 2007. Factors Affecting Flowering in Ornamental Plants. In: Anderson NO, eds. *Flower Breeding and Genetics.* Dordrecht: Springer, 7–48.
[10] Li L, Zhang Q, Huang D. 2014. A review of imaging techniques for plant phenotyping. *Sensors* 14(11):20078–20111.
[11] Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A. 2017. Going deeper in the automated identification of herbarium specimens. *BMC Evolutionary Biology* 17:181.
[12] Fomin ES, Fomina TI. 2020. "A nonparametric model for analysis of flowering patterns of herbaceous multi-flowered monocarpic shoots" unpublished.
[13] Fomina TI. 2012. *Biological characteristics of ornamental plants of natural flora in West Siberia.* Novosibirsk, In Russian
[14] Claßen-Bockhoff R, Bull-Hereñu K. 2013. Towards an ontogenetic understanding of inflorescence diversity. *Annals of Botany* 112:1523–1542.
[15] Fomin ES, Fomina TI. 2018. Patterns and models of flowering of some Campanulaceae Juss. species. *Vavilov Journal of Genetics and Breeding.* 22(7):845–855.
[16] Cleveland WS. 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. *J Am Statistical Association* 74(368): 829–836.

DOI 10.18699/BGRS/SB-2020-195

SeedCounter – mobile application for high throughput grain phenotyping

Mikhail Genaev
Kurchatov Genomic Center,
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
mag@bionet.nsc.ru

Evgeny Komyshev
Kurchatov Genomics Center,
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
delkom07@gmail.com

Dmitry Afonnikov
Kurchatov Genomic Center,
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
ada@bionet.nsc.ru

***Abstract* — Grains morphometry is an important step in selecting new high-yielding plants. The manual assessment of parameters such as the number of grains per ear and their size is laborious. The solution to this problem is image-based analysis, which can be performed using a desktop PC. The effectiveness of this analysis in the field can be improved through the use of mobile devices. We propose a method for the automated evaluation of phenotypic parameters of grains using mobile devices running the Android operational system. The experimental results show that this approach is efficient and sufficiently accurate for the large-scale analysis of phenotypic characteristics in wheat grains.**

Keywords* — *grain, phenotyping, computer image analysis, mobile devices, Android

Introduction

Large-scale breeding experiments require processing substantial phenotypic data, often in field conditions and thus without access to desktop computers and scanners. In this case, a digital camera is a viable option, but the images must be subsequently copied to a laptop or PC.

Modern mobile devices (smartphones and Internet tablets) contain digital cameras with high resolution. Mobile devices have multicore processors with sufficient computational power for image processing and analysis. These features allow users to capture and process images wherever necessary. A number of applications for mobile devices have been developed for the morphometry of plant organs. Leafsnap [1] is able to identify plant species in real time based on their leaf images: a user takes pictures of a plant leaf using a mobile device and sends the images from the camera to a remote server where they are processed. Leaf Doctor [2] is another mobile application that estimates the percentage of disease severity based on leaf images in a semi-automated manner.

In this work, we present a mobile application, SeedCounter [3], for the Android platform that enables the automated calculation of morphological parameters of wheat grains using mobile devices "in the field" (conditions without computer facilities). The application estimates the number of grains scattered on a sheet of paper and morphological parameters such as the length, width, area, and shape indexes: circularity, roundness, rugosity, solidity.

Materials and methods

The program input is a color image of grains placed arbitrarily on a sheet of white paper (A4, Letter, Legal, A3, A5, B4, B5, or B6). We recommend minimizing any contact between grains. To reduce errors, users should provide the following conditions for image capture: the paper sheet should be placed on a dark background and bright side lighting should be avoided.

The boundaries of the paper sheet on the background should be parallel to the sides of the frame. The fixed size of the paper makes it possible to calculate the scale of the image and evaluate the grain sizes in metric units. The SeedCounter application receives images directly from the camera of the mobile device.

The algorithm is implemented using the OpenCV image processing library [4] and consists of several steps:

(1) The paper sheet is recognized as a light area of tetragonal shape on a dark background.

(2) If the paper shape on the image deviates from rectangular, affine transformations convert it to rectangular.

(3) Grains are identified as contours to the image fragment corresponding to the paper sheet.

(4) The marked watershed method [5] is used to resolve the boundaries of seed grains that are in contact with one another. The resulting contours are approximated by grain ellipsoids, allowing for estimates of the size of the major and minor principal axes corresponding to the length and the width of the grain. SeedCounter additionally identifies the grain image area and shape indexes. The circularity and roundness indices show how close the shape of the contour is to the shape of the circle.

$$Circularity = \frac{4\pi \times area}{perimeter^2}$$

$$Roundness = \frac{4 \times area}{\pi\,[Major\ axis]^2}$$

The roughness index is defined as the ratio of the perimeter (Ps) of the contour to the convex perimeter (Pc):

$$Roughness = \frac{Ps}{Pc}$$

The solidity index is the ratio of the contour area to the convex hull contour area:

$$Solidity = \frac{Contour\ Area}{Convex\ Hull\ Area}$$

Grain recognition algorithm implemented on the Java programming language.

Results

The SeedCounter mobile application for Android devices is free to download at the Google Play Store (https://play.google.com/store/apps/details?id=org.wheatdb

.seedcounter) distributed under the BSD (Berkley Software Distribution) license.

SeedCounter is distributed under the BSD (Berkley Software Distribution) license. The grain number estimation accuracies for different experiment series [3] are shown that the MAPE (mean absolute percentage error) of the estimate of the number of grains on the sheet is close to 2%. A more detailed analysis showed that the largest errors in counting the number of grains occur if two or more grains on the paper are in contact and that under poor lighting conditions, the algorithm does not separate most of the grains. If the grains on the sheet are all separated, the seed counting error vanishes. The accuracy of length and width estimation for the grains by different devices in different conditions [3] is shown that the grain size estimation accuracy was approximately 0.30 mm that is approximately 8% of the linear dimensions of the wheat grain.

The work supported by the Kurchatov Genomic Center, Institute of Cytology and Genetics, SB RAS, agreement № 075-15-2019-1662 with the Ministry of Education and Science.

References

[1] Kumar N. et al. Leafsnap: A computer vision system for automatic plant species identification // European conference on computer vision. – Springer, Berlin, Heidelberg, 2012. – С. 502-516.

[2] Pethybridge S. J., Nelson S. C. Leaf Doctor: A new portable application for quantifying plant disease severity // Plant disease. – 2015. – Т. 99. – №. 10. – С. 1310-1316.

[3] Komyshev E., Genaev M., Afonnikov D. Evaluation of the SeedCounter, a mobile application for grain phenotyping // Frontiers in plant science. – 2017. – Т. 7. – С. 1990.

[4] Bradski G. The opencv library // Dr Dobb's J. Software Tools. – 2000. – Т. 25. – С. 120-125.

[5] Roerdink J. B. T. M., Meijster A. The watershed transform: Definitions, algorithms and parallelization strategies // Fundamenta informaticae. – 2000. – Т. 41. – №. 1, 2. – С. 187-228.

DOI 10.18699/BGRS/SB-2020-196

Exploring interaction between metabolic pathways involved in pigmentation of barley spike

Anastasiia Yu. Glagoleva
ICG SB RAS, Novosibirsk, Russia
glagoleva@bionet.nsc.ru

Nikolay A. Shmakov
ICG SB RAS, Novosibirsk, Russia
shmakov@bionet.nsc.ru

Aleksandr V. Vikhorev
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
vikhorev@bionet.nsc.ru

Sergei R. Mursalimov
ICG SB RAS, Novosibirsk, Russia
mursalimov@bionet.nsc.ru

Natalia V. Gracheva
VSTU, Volgograd, Russia
gracheva.tasha@yandex.ru

Tatjana V. Kukoeva
ICG SB RAS, Novosibirsk, Russia
kukoeva@bionet.nsc.ru

Olesya Yu. Shoeva
ICG SB RAS, Novosibirsk, Russia
olesya_ter@bionet.nsc.ru

Elena K. Khlestkina
ICG SB RAS, Novosibirsk, Russia
VIR, Saint-Petersburg, Russia
khlest@bionet.nsc.ru

***Abstract* — The color of barley grain (*Hordeum vulgare* L.) can be caused by synthesis and accumulation of various pigments. Besides green chlorophyll, pigmentation of barley spike can also be caused by two types of phenolic compounds: purple or blue anthocyanins and black melanins. The chemical and genetic nature of anthocyanins are well-known, while the metabolic pathways implicating in melanin formation remains less understood. The aim of the study was to explore the interaction between metabolic processes involved in melanin, anthocyanins and chlorophyll formation using the precise genetic models – near-isogenic lines (NILs) with different composition of pigments. Using marker-assisted selection we have developed two hybrid NILs: i:Bw*Blpalm* with chlorophyll deficiency and melanin accumulation and i:Bw*BlpAnt1Ant2* with simultaneous melanin and anthocyanin accumulation in grain. The analysis of these lines allows us to suggest that melanin biosynthesis affects plant metabolic processes significantly. Although melanin synthesis occurs chlorophyll/photosynthesis independently, it probably has an ability to suppress the photosynthesis in barley spike. Unlike that, the presence of dominant *Blp* allele increases the anthocyanins accumulation in grain pericarp.**

Keywords* — *barley, near-isogenic lines, phenylpropanoids, anthocyanins, melanin, chlorophyll

Motivation and aim

Motivation

The color of barley grain (*Hordeum vulgare* L.) can be caused by synthesis and accumulation of various pigments. Besides green chlorophyll, two types of phenolic compounds are involved in pigmentation of barley spike: purple or blue anthocyanins and black melanins. Different genes control the accumulation of these pigments: the complementary genes *Ant1* and *Ant2* determine synthesis of purple anthocyanins in grain pericarp; the *Blx* genes control accumulation of blue anthocyanins in aleurone layer, while melanins are formed under control of the *Blp* gene in glumes and pericarp tissues [1]. The biochemical pathway leading to anthocyanins is well-known, while melanin remains the less studied pigment in plants and the metabolic pathways involved are not clear enough. The comparative RNA-seq analysis between melanin-accumulating NIL and its parental cultivar Bowman have revealed the influence of the *Blp* locus on expression of more than a thousand genes. Among these, genes belonging to phenylpropanoid and fatty acids biosynthesis pathways were the most represented and upregulated in black-colored NIL [2]. Comparative microscopy analysis of the NILs grains at early, soft and hard dough developmental stages revealed the intracellular dark structures co-localized with the red autofluorescence signal of chlorophyll in cells of pericarp and hulls of the i:Bw*Blp* line starting from early dough stage. Thus, the melanin pigments were showed to be formed and accumulated in chloroplasts [3]. The data served as the basis for studying, on one hand, the interaction between biosynthesis pathways of phenolic compounds – melanin and anthocyanins, which together can be accumulated in barley grain pericarp cells,– and relationships between melanin formation and chlorophyll synthesis (and photosynthesis), which together can be present in chloroplasts pericarp cells, on the other hand.

Aim

The aim of the study was to explore interaction between metabolic processes involved in melanin, anthocyanins and chlorophyll formation using precise genetic models – near-isogenic lines (NILs) with different composition of pigments.

Methods

To investigate melanin/chlorophyll and melanin/anthocyanins interactions two hybrid NILs i:Bw*Blpalm* with chlorophyll deficiency and melanin accumulation and i:Bw*BlpAnt1Ant2* with simultaneous melanin and anthocyanin accumulation in grain were developed using marker-assisted selection. The initial Bowman NILs i:Bw*Blp* (NGB20470), i:Bw*alm* (NGB20419), i:Bw*Ant1Ant2* (NGB22213) were obtained from NordGen (Alnarp, Sweden). The study of metabolic pathways interaction was performed with comparative gene expression analysis, including RNA-seq approach and RT-PCR.

Results

Using microsatellite markers linked to the *Alm* [4] and *Blp* [1] loci, the homozygote hybrid plants for these loci were selected from the F_2 population generated by hybridization of i:Bw*Blp* with i:Bw*alm*. The selected plants are characterized by partial albinism phenotype in stem node accompanying the chlorophyll insufficiency in pericarp and glume, and melanin pigmentation in grain. Inheritance of these traits in homozygote and stability of phenotype were checked in F_3 families. Similar approach was used to create the hybrid line i:Bw*BlpAnt1Ant2* characterized by simultaneous accumulation of melanins and anthocyanins in pericarp

tissues. In this case, the intragenic markers for the *Ant1* [5] and *Ant2* [6] genes were used. The developed lines represent a convenient genetic model for comparative studies of interaction between metabolic pathways underlining pigments synthesis.

In the case of melanin/chlorophyll interaction, the fact of obtaining i:Bw*Blpalm* line allows us to suggest chlorophyll/photosynthesis-independent melanin synthesis. At the same time, the data of comparative transcriptomic analysis demonstrated downregulation of chloroplast genes in NIL accumulated melanin. Thus, it allows us to suggest the tendency to photosynthesis suppression in response to melanin synthesis.

In the case of melanin/anthocyanins pathways relationships, we observed the appearance of anthocyanin pigmentation at an earlier stage of spike development than the appearance of melanin pigment in i:Bw*BlpAnt1Ant2* line, as well as the increased content of anthocyanins in the spike in the hybrid line in compared to the parental NIL. This observation points that the same precursor compounds are involved in the biosynthesis of anthocyanins and melanin. The comparative RNA-seq analysis of i:Bw*BlpAnt1Ant2* and its parental NILs i:Bw*Ant1Ant2* and i:Bw*Blp* at the different stages of spike development confirms upregulation of the key phenylpropanoid biosynthesis genes at the early stage in all NILs in comparison to Bowman. However, it can be concluded that the presence of *Blp* locus significantly affects genes expression in barley spike and activates unique genes of melanogenesis.

Acknowledgment

Comparative transcriptomic and cytological analysis of barley NILs were supported the Russian Science Foundation grant No.16-14-00086, development and analysis of the i:Bw*Blpalm* line were supported by the Russian Science Foundation grant No.19-76-00018, development and analysis of the i:Bw*BlpAnt1Ant2* line were supported by the RFBR grant No. 20-316-80016.

References

[1] Шоева О.Ю. и др. (2018). Гены, контролирующие синтез флавоноидных и меланиновых пигментов ячменя. Вавиловский журнали генетики и селекции. 22(3):333-342.

[2] Glagoleva A.Yu. et al. (2017). Metabolic pathways and genes identified by RNA-seq analysis of barley near-isogenic lines differing by allelic state of the *Black lemma and pericarp* (*Blp*) gene. BMC Plant Biology. 17(Suppl 1):182.

[3] Shoeva O.Yu. et al. (2020). Melanin formation in barley grain occurs within plastids of pericarp and husk cells. Scientific Reports. 10:179.

[4] Shmakov N.A. et al. (2016). Identification of nuclear genes controlling chlorophyll synthesis in barley by RNA-seq. BMC Plant Biology. 16: 245.

[5] Himi E., Taketa S. (2015). Isolation of candidate genes for the barley Ant1 and wheat Rc genes controlling anthocyanin pigmentation in different vegetative tissues. Molecular Genetics and Genomics. 290(4):1287-1298.

[6] Cockram J. et al, (2010). Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proceedings of the National Academy of Sciences USA. 107(50):21611-21616.

DOI 10.18699/BGRS/SB-2020-197

The meta-analysis of transcriptomes of *Arabidopsis thaliana* transgenic plants with altered expression of dual-targeting RNA-polymerase RPOTmp

Igor Vladimirovich Gorbenko
Laboratory of Plant Genetic Engineering
SIPPB SB RAS
Irkutsk, Russia
gravov.chemistry@gmail.com

Vladislav Igorevich Tarasenko
Laboratory of Plant Genetic Engineering
SIPPB SB RAS
Irkutsk, Russia
vslav@inbox.ru

Alexander Igorevich Katyshev
Laboratory of Plant Physiological Genetics
SIPPB SB RAS
Irkutsk, Russia
alex@sifibr.irk.ru

Vadim Igorevich Belkov
Laboratory of Plant Genetic Engineering
SIPPB SB RAS
Irkutsk, Russia
anvad.irk@rambler.ru

Yuri Mikhailovich Konstantinov
Laboratory of Plant Genetic Engineering
SIPPB SB RAS
Irkutsk, Russia
yukon@sifibr.irk.ru

Milana Viacheslavovna Koulintchenko
Laboratory of Plant Genetic Engineering
SIPPB SB RAS
Irkutsk, Russia
mk100171@yahoo.com

Abstract — In dicotyledons, RPOTmp is one of three phage-type RNA polymerases encoded in the nucleus and transported to both mitochondria and chloroplasts. It was established that the main functions of RPOTmp are associated, however, with mitochondria, and not with chloroplasts. The activity of RPOTmp in mitochondria is manifested throughout all stages of plant development, where together with the main RNA polymerase of these organelles RPOTm, it transcribes genes encoding subunits of the I and IV respiratory complexes, the protein synthesis apparatus, as well as a number of mitochondrial proteins with unknown functions. The impairment of RPOTmp functions at the molecular level leads to a significant decrease in the activity of I and IV respiratory complexes, at the organism level, to a delay in germination, a number of phenotypic differences in adult plants and a significant delay in flowering. Many aspects of the organelle gene expression regulation, and in particular RPOTmp, are still unknown. The fact of dual localization indicates a possible role of this NEP in anterograde signaling. Comprehensive analysis of transcriptomes data of several transgenic lines obtained by DNA microarray method was done using the bioinformatic approach. Bioinformatic analysis of transcriptomic and metabolic changes in transgenic plants with altered RPOTmp expression is presented, the presumptive role of this enzyme in organelle gene expression regulation as well as in retrograde signaling regulation in plants is discussed.

Keywords — mitochondria, chloroplasts, Arabidopsis thaliana, transcriptome, gene expression regulation, RNAP, retrograde and anterograde regulation

Introduction

It is well-known that cell organelles produce ATP and a number of vital metabolites which causes their importance for plant growth and development. Despite the fact that most of organelle genome was transferred to nucleus during the evolutional process, mitochondria and chloroplasts saved their own genetic apparatus. Several thousand of proteins encoded in nuclear and organellar DNA define organelles structure and function and emphasize the significance of OGE (Organellar Gene Expression) and NGE (Nuclear Gene Expression) coordination. Up to date a significant bulk of data characterizing chloroplast gene transcription and regulatory pathways that control chloroplast function obtained. On the other hand, mitochondrial gene expression regulation involving nuclear-encoded transcription factors and various cell metabolites is still insufficiently explored in many aspects. We believe that the model developed in our laboratory using transgenic plants of Arabidopsis with complementation or overexpression of RPOTmp functions, one of the organellar NEP (Nuclear Encoded Polymerase), targeting either in mitochondria or in chloroplasts, will allow us to identify and study factors involved in the regulation of OGE, or indirectly, via expression regulation of gene encoding this NEP, in retrograde and anterograde signaling.

To investigate the character of transcriptome changes dependent on altered RPOTmp expression, we carried out DNA-microarray assays and performed bioinformatic analysis of obtained data.

Results

Analysed transgenic plants

Earlier, *Arabidopsis thaliana* transgenic plants with RPOTmp overexpression in mitochondria or chloroplasts were obtained in our laboratory (Tarasenko et al., 2016). Four sets of transgenic lines were obtained on the basis of the Arabidopsis wild type plants (Columbia-0), (Col-M and Col-P lines), and on the basis of the *rpotmp* mutant line with inactivated *RPOT2* gene encoding RPOTmp, (Tmp-M and Tmp-P lines). The obtained transformants are characterized by cytoplasmic expression in their cells of recombinant RPOTmp containing a transit peptide for protein targeting to mitochondria (Col-M and Tmp-M lines) or to chloroplasts (lines Col-P and Tmp-P). The Col-M/Col-P lines and the Tmp-M/Tmp-P lines are characterized by RPOTmp overexpression or by complementation of RPOTmp functions respectively.

For DNA microarray analysis, total RNA was extracted from 12-days old seedlings of the wild type, the *rpotmp* mutant line, Col-M, Col-P, Tmp-M and Tmp-P lines. Extracted DNA microarray data were used to perform a high-throughput full-genome transcriptomic analysis of the studied lines in comparison with the wild-type plants. Three independent biological replicates were analyzed, each represented RNA preparations obtained from several plants of the studied lines.

Data obtained

Using multiple comparison method, 2301 genes were estimated as differentially expressed with $p < 0.05$ and $lfc > 1$ (all data was estimated and statistically tested using limma R package [1], the p-value of each gene was corrected by the Benjamini and Hochberg method). Among the DEG array, 61 records were found to be annotated as transcription factors (TF). There were found 10 ethylene response TFs, 9 TFs with the MADS-box domain, 7 TFs of the WRKY family, 6 TFs with the MYB domain and the bHLH motif. The presence of two heat shock factors was observed, HSFA6A, the expression of which was increased in Col-P12 line and in Tmp-M lines, and HSF30 with reduced expression in Col-M20 and Col-P5 transgenic lines.

To deepen our understanding of cell transcriptome changes caused by modified mitochondrial and chloroplast expression of RPOTmp, we performed co-expression analysis and estimated co-expression modules. RPOTmp was found to be located in module 1. We used INTERACTOME 2.0 [2], a predicted protein-protein interactions database to construct protein-protein interactions network (fig. 1). We found that RPOTmp is probably connected via AT5G52640 (HSP81.1 / 83 / 90.1, cytosolic) with large network nodes: AT3G57150 (putative pseudouridine synthase (NAP57), nuclear and cytosolic), AT3G52140 (FMT, Friendly mitochondria) and AT4G37910 (mtHsc70-1, mitochondrial HSP70-1). HSP90.1 is a stress-induced (heat shock) holdase chaperone that is necessary for cell survival and besides is necessary for mediating brassinosteroid signals [3].

In order to find out which particular cellular processes are affected by modified expression of RPOTmp we carried out Gene Ontology and KEGG pathway analysis.

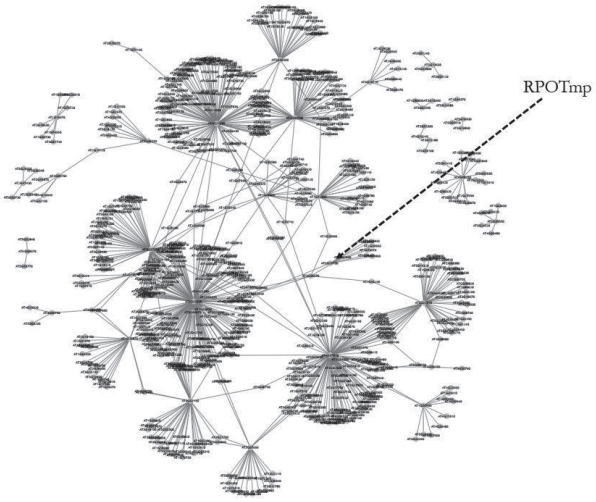


Fig. 1. Protein-protein interactions network of co-expression module I.

Conclusions and perspectives

In our study, it was found that the lack of functional RPOTmp in mitochondria (*rpotmp* and Tmp-P1 lines) caused a significant decrease in HSP90.1 expression, suggesting that these lines may have hypersensitivity to BRs. In the presence of RPOTmp in mitochondria at any level, the expression of HSP90.1 in our samples did not differ from the wild type.

Revealed by co-expression analysis, NAP57 and FMT communicate through a very large number of auxiliary proteins with one of the largest network node, fibrillarin 2 (FIB2, AT4G25630), a key nucleolar protein in eukaryotes that binds to the C / D box of small nucleolar RNAs (snoRNAs), directing methylation of 2'-ribose in rRNA and acts as a bridge for the transfer of information from gene-specific regulatory proteins to the RNA polymerase II transcription apparatus [4, 5]. Expression of FIB2 was found to be reduced ($lfc = –0.95$, $p = 0.018$) in Col-M15, the line with mitochondrial overexpression of RPOTmp.

According to Gene Ontology and KEGG pathway analysis, it was found that the Col-M15 line might have reduced rate of ribosome biogenesis. On the other hand, the *rpotmp* line lacking functional RPOTmp has reduced expression of genes that belong to plant-pathogen interactions pathway and increased expression of endocytosis-associated genes.

References

[1] B. Phipson, S. Lee, I. J. Majewski, W. S. Alexander, G. K. Smyth, “Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression”, Ann Appl Stat, vol. 10, pp. 946-963, July 2016.

[2] J. Geisler-Lee, N. O'Toole, R. Ammar, N. J. Provart, A. H. Millar, M. Geisler, “A Predicted Interactome for Arabidopsis”, Plant Physiology, vol. 145, pp. 317-329, October 2007.

[3] D. Samakovli, T. Margaritopoulou, C. Prassinos, D. Milioni, P. Hatzopoulo, “Brassinosteroid nuclear signaling recruits HSP90 activity”, New Phytol, vol. 203, pp. 743-757, 2014.

[4] F. Barneche, F. Steinmetz, M. Echeverría, “Fibrillarin Genes Encode Both a Conserved Nucleolar Protein and a Novel Small Nucleolar RNA Involved in Ribosomal RNA Methylation in Arabidopsis thaliana”, Journal of Biological Chemistry, vol. 275, pp. 27212-27220, 2000.

[5] D. Rakitina, M. Taliansky, J. Brown, N. Kalinina, “Two RNA-binding sites in plant fibrillarin provide interactions with variousRNA substrates”, Nucleic acids research, vol. 39, pp. 8869-8880, 2011.

DOI 10.18699/BGRS/SB-2020-198

Retrotransposons of *Arabidopsis thaliana* expressed in wild-type plants

Sofya Gvaramiya
Department of marker-assisted and genomic selection of plant
All-Russia Research Institute of Agricultural Biotechnology
Moscow, Russia
sofia.gvaramia@gmail.com

Murad Omarov
Department of marker-assisted and genomic selection of plant
All-Russia Research Institute of Agricultural Biotechnology
Moscow, Russia
muradok98@gmail.com

Ilya Kirov
Department of marker-assisted and genomic selection of plant
All-Russia Research Institute of Agricultural Biotechnology
Moscow, Russia
kirovez@gmail.com

***Abstract* — Mobile genetic elements are widely spread across the genomes of plants and animals. Of them, retrotransposons occupy substantial part of the genome. They had a huge impact on host genome evolution. However, their current activity in plant cell is thought to be very low on both RNA and cDNA (mobilome) levels because multiple silencing pathways. To get deeper insight into the mechanisms of mobilome formation we carried out identification of transcribed TEs of model plant, *Arabidopsis thaliana*. Using bioinformatics and 'wet' approaches we found evidence of expression of 6 TEs of *A.thaliana*. This opens a door for further functional studies of how some TEs can overcome the genome defense against retrotransposon activity.**

Keywords* — *retrotransposons, Arabidopsis thaliana, transposable elements, retrotranscriptome, retrotransposososns mobility

Motivation and aim

Motivation

Repetitive elements constitute a major part of higher plant genome [1]. The ubiquitous class of these elements is retrotransposons (TEs) – mobile genetic elements, which move through host genome via "copy-paste" mechanism, including RNA transcription, reverse transcription with formation of cDNA and its insertion. A large number of active retrotransposons are suppressed by various types of methyltransferases like MET1, CMT3, KYP and by RNA-directed DNA methylation (RdDM) [2]. Therefore, it is hard to study biological processes connected with transposable elements. However, while most of the TEs are silenced in the genomes, the recent studies showed that some of them can form RNAs in somatic tissues under non-stressed conditions raising the question of how some TEs can overcome the genome defense mechanisms. *A.thaliana* being a model plant provides unique set of the tools, but no TEs transcribed in wild-type non-stressed plants have been described so far.

Aims

The aim of our study was to detect retrotransposons expressed in *A. thaliana* WT (Col-0) and ddm1 mutant plants under non-stressed conditions.

Methods

Bioinformatic analysis

Data analysis of LTR-retrotransposons expression was performed using proper pipeline, combined different programs and designed for this research.

Molecular methods

DNA was extracted by CTAB method from leaf tissue and exposed polymerase chain reaction.

RNA was isolated from frozen *A. thaliana* plants with Trizol reagent (Evrogen) according to applied protocol. Extracted RNA was treated with DNAse I (Qiagen). After treatment, RNA was subjected RT-PCR through special kit (Biolabmix).

Results

In our research, we have verified mutant lines of *A. thaliana* by target genes and rated the transcriptional activity of EVD retrotransposon as marker for these plants [3], [4].

We have found six active retrotransposons, which expressed in ddm1 and wild type plants of Arabidopsis by means of our bioinformatic approach and RT-PCR analysis. To confirm their mobility the presence of extrachromosomal circular DNA was analysed. The results of this work provide unique cases for future studying of retrotransposone evolution and functional activity.

Acknowledgment

The work was carried out with the financial support of The Russian Foundation for Basic Research (RFBR grant № 20-34-70032).

References

[1] Slotkin R. K., Martienssen R. A. (2008) Transposable elements and the epigenetic regulation og the genome. Nature Rev. Genet. 8: 272-285.

[2] Oberlin S., Sarazin A., Chevalier C., Voinnet O., Mari-Ordonez A. (2017) A genome-wide transcriptome and translatome analysis of Arabidopsis transposons identifies a unique and conserved genome expression strategy fot Ty1/Copia retroelements. Genome Res. 27: 1549-1562.

[3] Reinders J., Mirouze M., Nicolet J., Paszkowski J. (2009) Parent-of-origin control of transgenerational retrotransposon proliferation in Arabidopsis. EMBO reports. 14(9): 823-828.

[4] Tsukahara S., Kobayashi A., Kawabe A., Mathieu O., Miura A., Kakutani T. (2009) Bursts of retrotransposition reproduced in Arabidopsis. Nature. 461: 423-426.

DOI 10.18699/BGRS/SB-2020-199

Estimation of a joint distribution for several phenotypic traits in breeding or ancient populations

Anna A. Igolkina
SPbPU, St.Petersburg, Russia
igolkinaanna11@gmail.com

Sergey Nuzhdin
SPbPU, St.Petersburg, Russia
UCS, Los Angeles, USA
snuzhdin@usc.edu

Maria G. Samsonova
SPbPU, St.Petersburg, Russia
m.g.samsonova@gmail.com

***Abstract* — Prediction of phenotypic trait values for a population is a challenge in genomic selection and evolutionary studies. We propose a pipeline of developed models, which estimates the distribution of several trait values for a population with known allele (SNP) frequencies. The pipeline consists of two parts: mtmlSEM model, which describes the influences of pleiotropic and single-trait SNPs on genetically correlated traits, and dipophen method, which predicts the distribution of traits in a population, based on mtmlSEM. As an input, the pipeline gets allele frequencies evaluated for a breeding population or SNP ancestral population. The pipeline was applied to two datasets: one from the breeding soybean project and the other from the chickpea "trade routes" project. While allele frequency in the breeding population can be directly calculated, for estimating allele frequencies in the ancestral population, we developed new models.**

Keywords* — *distribution of traits, allele frequencies, pleiotropic SNPs, SEM, Bernoulli random variables, population admixture

INTRODUCTION

Motivation

Significant determinants for many phenotypic traits are stored in the genome. The standard way to identify genetic variants (SNPs) associated with phenotypic traits considers them independently; however, traits are often genetically correlated and should be considered in GWAS analysis simultaneously. Recent methods to cope with correlated traits do not separate SNPs into pleiotropic and single-trait effects; therefore, we propose the mtmlSEM model, which structures hidden biologics mechanism affecting correlated traits and directly searches SNPs of different effects.

The mtmlSEM model not only allows us identifying SNPs associated with related phenotypic variables but also provides the technique to predict phenotypic traits for a new genotype.

The prediction of phenotypic traits for an individual is not as impressive as for the whole population. For the latter, we propose the fast method, dipophen, to predict the joint distribution of several phenotypic traits. This method takes as input allele frequencies in a population and parameter estimated from any linear genotype-phenotype model (e.g., mtmlSEM).

In breeding studies, the evaluation of allele frequencies for a given population is quite simple and requires SNP data for representative samples. In evolutionary studies, estimation of SNP frequencies in ancestral populations requires mathematical modeling. The problem in such modeling is that frequencies belong to a compositional data class (sum of frequencies equal to 1). Still, none of the models, which predict allele frequencies in populations, consider this compositional nature, and new, more correct models are required.

Aim

This study aims to develop the pipeline to estimate the joint distribution of genetically correlated traits in populations characterized by SNP frequency data. To reach the goal, the following list of tasks to develop was set:

- genotype-phenotype model based on structural equation modeling, which describes the biosocial mechanism affecting genetically correlated traits, and separate SNPs of pleiotropic and single-trait effect.
- statistical technique to estimate the joint distribution of the weighted sum of independent and non-identically Bernoulli random variables.
- Bayesian model to estimate allele frequency in the ancestral population, which considers possible ways of dispersal of the population from the center towards cultivation locations.

METHODS

mtmlSEM model

To describe genetically correlated phenotypes we proposed the following mtmlSEM (multi-trait, multi-locus) model (Fig 1A):

$$\eta = B\eta + \Pi g + \varepsilon$$
$$p = \Lambda\eta + \mathrm{K}y + \delta$$

Where p is a vector of phenotypes, η is a vector of latent factors, g and y are SNP variables of pleiotropic and single-trait effects, respectively; Π and K are matrixes of SNP influences on latent factors and phenotypes, respectively Λ is a matrix of factor loadings, Bis a matrix of relationships between latent factors, ε and δ are random errors. We developed an automatic algorithm to specify the model, and Bayesian scheme to get parameter estimates. We utilized the **sympy** package [1] and **mtmlSEM** model [2].

Estimation of the joint distribution of several traits

A mtmlSEM model can be used to predict phenotype from genotype in the following manner:

$$\hat{p} = \Lambda(1 - B)^{-1}\Pi g + Ky,$$

so that each phenotype is a linear combination of SNPs, each taking values from {0,1,2} (additive model). To estimate the distribution of phenotypes in a population, we

can use allele frequencies of g and y and set that random variable $\tilde{p}$ is:

$$\tilde{p} = 2(\Lambda(1-B)^{-1}\Pi\tilde{g} + K\tilde{y}),$$

where $\tilde{g}$ and $\tilde{y}$ are vectors of independent Bernoulli random variable with population allele frequencies. To estimate the joint distribution of $\tilde{p}$, we developed the **dipophen** method, which utilizes the **bernmix** package (https://github.com/iganna/bernmix). The bernmix package allows estimating the distribution of a linear combination of Bernoulli random variables based on Fourier transform and faster than analogs.

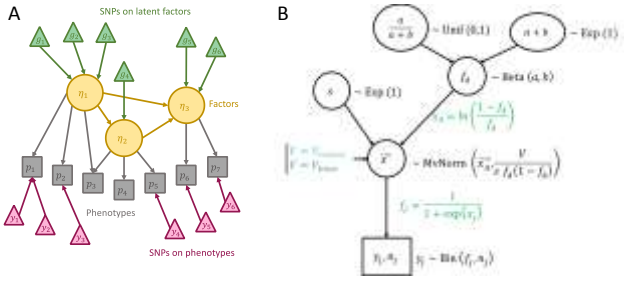


Fig. 1. (A) mtmlSEM model; (B) Bayesian model to predict allele frequencies within clusters

Prediction of allele frequencies of ancestral chickpea populations

We utilized the chickpea genotyped landraces from Vavilov's collection: landraces form ten clusters, each composed of geographic locations with genotypes samples. We examined two hypothetical dispersals of the chickpea population within each cluster and estimated allele frequencies in their centers. We hypothesized that chickpea was spread within a cluster along the binary path either in line with trade routes or by diffusion-like manner. In the model, the contrast between these hypotheses was reflected by different given covariance matrices V (Fig 1B).

The model's important feature is the use of compositional data analysis in the application to allele frequency changes along the binary path from the center towards locations: we considered changes in balances between allele frequencies instead of direct frequencies (Fig 1B).

RESULTS

Soybean: breeding pairs

We analyzed 257 soybean accessions (Glycine max) with known genotypes (4767 SNPs) and values of three traits: productivity, oil and protein contents. We transform ilr-transformation [3] to oil and protein contents, specify the mtmlSEM model with one latent variable, and get parameter estimates. Then we tested all possible pairs of soybean accessions as potential breeding pairs. For each pair, we computed SNP frequencies in F2 and predicted the joint distribution of three phenotypes (Fig 2). The results allowed us to rank all pairs according to the probability of having efficient offspring in F2.

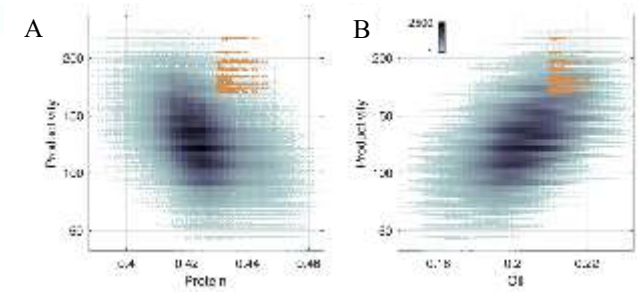


Fig 2. Distributions of trait values in all possible homozygotes. Shades of blue display the density of homozygotes on scatterplots. Orange color denote homozygotes, which satisfy introduced thresholds: productivity should be higher than 170, protein content – higher than 0.43, oil content – higher than 0.21

Chickpea: ancestral populations

The chickpea dataset consisted of 421 samples genotyped on 2579 SNPs and phenotyped on 16 traits. Samples were attributed to ten geographically justified populations. First, we specified mtmlSEM model and obtained 5 latent factors, which describe correlation in 16 traits (Fig. 3).

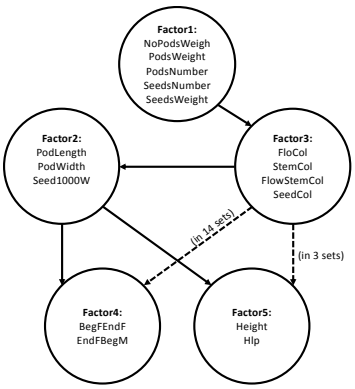


Fig 3. Latent factors joined to form structural part of connected SEM model. Dashed arrows represent relationships, which were not present is all training sets for directed acyclic graph obtained; Solid lanes represent relationships, which were found in each of 20 training sets

Then, we tested historical hypotheses about chickpea dispersals and found that in 9 out of 10 populations, dispersals we likely occurred along with trade routes. We took allele frequency estimated obtained from 'trade route' hypothesis and get distributions off phenotypic traits in ancestral populations (centers of clusters), which were in line with origin of light chickpea subtype (kabuli).

ACKNOWLEDGMENT

The project is supported by the RBFR grant 18-29-13033.

REFERENCES

[1] Igolkina AA, Meshcheryakov G. semopy: A Python Package for Structural Equation Modeling. Struct Equ Model A Multidiscip J. 2020:1–12. doi:10.1080/10705511.2019.1704289.

[2] Igolkina A, Meshcheryakov G, Gretsova M, Nuzhdin S, Samsonova M. Multi-trait multi-locus SEM model discriminates SNPs of different effects. 2020.

[3] Pawlowsky-Glahn V, Buccianti A. Compositional Data Analysis. Chichester, UK: John Wiley & Sons, Ltd; 2011. doi:10.1002/9781119976462.

DOI 10.18699/BGRS/SB-2020-200

The prospects for the study of the avirulence genes characteristic for the West Siberian population of wheat stem rust *Puccinia graminis* f. sp. *tritici*

Vasiliy Kelbin
ICG SB RAS
Novosibirsk, Russia
kelbin@bionet.nsc.ru

Ekaterina Sergeeva
ICG SB RAS
Novosibirsk, Russia
sergeeva@bionet.nsc.ru

Ekaterina Skolotneva
ICG SB RAS
Novosibirsk, Russia
skolotnevaes@bionet.nsc.ru

Elena Salina
Kurchatov Genomics Center, ICG SB RAS,
Novosibirsk, Russia
salina@bionet.nsc.ru

***Abstract* — Wheat stem rust is a plant disease caused by pathogenic fungus *Puccinia graminis* f. sp. *tritici* which leads to significant damage of the crop. During the last decade the disease affected the crops in West Siberia. The plant response to infection is determined by the correspondence of host resistance (*Sr*) and fungus avirulence/virulence (*Avr*/*vr*) genes. We first characterized the West Siberian population of wheat stem rust: the probable sources of infection were traced, the race composition and main avirulence genes were defined. Four fungal lines with contrasting patterns of *Avr* genes were selected for whole genomic sequencing.**

Keywords* — *bread wheat, stem rust, Avr gene, Sr gene, genome sequencing

Motivation

The interaction of bread wheat *Triticum aestivum* L. (2n = 6x = 42) and stem rust *Puccinia graminis* f. sp. (forma specialis) *tritici* is a promising model for the study of "plant host - pathogen" coevolution. This model can elaborate the genetic basis of plant immunity response and pathogenesis, and is also relevant for breeding requirements. During the last decade the extensive outbreaks of stem rust occured at the West Siberia region [1]. Plant breeding for immunity to stem rust requires the extensive data on genetic characteristics of pathogen population. Our study is dedicated to analysis of West Siberian local population of fungal pathogen *P. graminis* f. sp. *tritici*.

Methods

To describe the population we applied the methods of phytopathology and molecular genetics. Initially, we collected the samples of stem rust originated from Novosibirsk, Omsk and Altay regions during the wheat vegetations of 2017-2019 years. Totally we obtained the 94 single pustule isolates (genetically homogeneous lines) that are maintained in laboratory conditions.

Results

First, we identified the type of life cycle for the local (Novosibirsk) stem rust population. The fungus *P. graminis* can realize complete (sexual) or partial (asexual) life cycles. In case of complete cycle, the barberry (*Berberis*) plant is used as alternate host on which the sexual process is carried out [2]. We compared the genotyping profiles for 11 molecular SSR-markers (simple sequence repeats) of 60 rust samples collected from wheat, barberry and wild grasses [3, 4, 5]. Processed using the PAST3 program, SSR profiles obtained from barberry and wild grasses were clustered in the same group and the samples from wheat to distinct one. Therefore, sexual life for *P. graminis* f. sp. *secalis* through barberries and wild grasses and asexual life cycle for wheat stem rust *P. graminis* f. sp. *tritici* were supposed for local in the Novosibirsk region.

Then, we evaluated the avirulence genes (*Avr*) polymorphisms and the race composition of local population of stem rust. The infection types of single pustule isolates *P. graminis* f. sp. *tritici* were tested on special set of rust resistant/susceptible wheat lines (Pgt differential set, AAFC Cereal Research Centre, Canada). In the Novosibirsk region, we identified one predominant race TKRPF and the minor races differing for 1-3 genes. The cluster analysis (PAST3, Multivariate Software) of the samples originated from Novosibirsk, Omsk and Altay regions demonstrated that there are two independent pathogen subpopulations presented in Omsk and Altay regions sharing the races at the contact area of Novosibirsk region. Thus, wheat stem rust in the Novosibirsk region is not endemic but carried by wind from both neighboring regions. This is confirmed by the previous conclusion that barberry is not a source of infection for wheat crops in the Novosibirsk region.

We postulated the profiles of the avirulence/virulence (*Avr*/*vr*) genes and their allelic states in local Novosibirsk population. The 51 single pustule isolates obtained in 2017, 2018 were twice tested by virulence to wheat lines with *Sr* genes (Pgt differential set), and the data were processed with the PAST3 program. The most of *P. graminis* f. sp. *tritici* isolates were virulent to lines with resistance genes *Sr5, Sr9a, Sr9g, Sr10, Sr38, SrMcN*, and avirulent to lines with *Sr24, Sr30, Sr31*. Also we found the changes in the frequencies of some reaction types on lines with *Sr6, Sr7b, Sr8a, Sr9b, Sr9d, Sr9e, Sr11, Sr17, Sr21, Sr36*, and *SrTmp* genes that could be used for tracing the dynamics of virulent/avirulent types in Novosibirsk region. That way, we state the presence of the avirulence genes in local *P. graminis* f. sp. *tritici* population: *AvrSr6, AvrSr7b, AvrSr8a, AvrSr9b, AvrSr9e, AvrSr11, AvrSr17, AvrSr21, AvrSr24, AvrSr30, AvrSr31, AvrSr36, AvrSrTmp*.

As a result of the foregoing we characterized the West Siberia (in particular local Novosibirsk population) for the first time. The data obtained in the study let us select the four single pustule isolates with contrasting patterns of *Avr* genes for further genomic analysis. We performed the whole genome Illumina sequencing for 4 samples in two replications

each at the NextSeq550 capacity of Core Facility of Genomics Investigations (Institute of Cytology and Genetics SB RAS) and obtained from 14,8 to 26 millions of pair end reads (2x150 bp), that corresponds to coverage of 50X and higher. Although the reference sequence of *P. graminis* f. sp. *tritici* is now publicly available, the studies of fungal *Avr* genes structure are at the ongoing stage: proposed numerous candidated for pathogen effector genes [5, 6, 7]. The whole genomic sequencing of isolates (lines) with different *Avr* spactra, further data processing (de novo assembling, reads mapping to reference sequence) and comparative genomic analysis will enable us to develop the molecular SNP (single nucleotide polymorphism) and SSR (simple sequence repeat) markers either for fast efficient pathogen monitoring at the local area or for using in study of interaction "host plant - pathogen" model.

Acknowledgment

The work was supported by RFBR 17-29-08018 and the budgetary project No. 0259-2019-0001-C-01.

References

[1] Skolotneva E.S., Salina E.A. (2016) Features of the causative agent of wheat stem rust in the conditions of Western Siberia. In: Protection of grain crops from diseases, pests, weeds: achievements and problems. Bolshiye Vyazemy. 67-71.

[2] Leonard K.J., Szabo L.J. (2005) Stem rust of small grains and grasses caused by Puccinia graminis. Molecular plant pathology. 6(2): 99-111.

[3] Berlin A. et al. (2012) Genetic variation in Puccinia graminis collected from oats, rye, and barberry. Phytopathology. 102(10): 1006-1012.

[4] Zhong S. et al. (2009) Development and characterization of expressed sequence tag-derived microsatellite markers for the wheat stem rust fungus Puccinia graminis f. sp. tritici. Phytopathology. 99(3): 282-289.

[5] Ellis J.G. et al. (2007) Wheat rust resistance research at CSIRO. Australian Journal of Agricultural Research. 58(6): 507-511.

[6] Rutter W.B. et al. (2017) Divergent and convergent modes of interaction between wheat and Puccinia graminis f. sp. tritici isolates revealed by the comparative gene co-expression network and genome analyses. BMC genomics. 18(1): 291.

[7] Sperschneider J. et al. (2014) Diversifying selection in the wheat stem rust fungus acts predominantly on pathogen-associated gene families and reveals candidate effectors. Frontiers in plant science. 5: 372.

DOI 10.18699/BGRS/SB-2020-201

Detection and investigation of genes with circadian expression pattern in common wheat

Antonina Kiseleva
Kurchatov Genomics Center,
ICG SB RAS
Novosibirsk, Russia
antkiseleva@bionet.nsc.ru

Maria Bragina
Kurchatov Genomics Center,
ICG SB RAS
Novosibirsk, Russia
koltunova@bionet.nsc.ru

Elena Salina
Kurchatov Genomics Center,
ICG SB RAS
Novosibirsk, Russia
salina@bionet.nsc.ru

Abstract — The study of daily oscillations of the wheat transcriptome, the identification of both the core circadian genes and the genes of metabolic pathways under the influence of daily rhythms is the main topic of this work.

Keywords — Triticum aestivum, circadian rythms, gene expression

Motivation and aim

Motivation

Plant circadian rhythms coordinate such physiological processes as growth, transition to flowering, photosynthesis, response to stress, metabolism and phytohormone signal transduction under the daily and seasonal changes in environmental conditions [1]. Such concurrence has a positive effect on plant growth, development and their adaptability. Genes regulated by circadian rhythms include genes for CO_2 assimilation, starch accumulation/degradation in leaves, and storing nutrients (lipids and fatty acids) in seeds [2]. Thus, a large part of the plant transcriptome is controlled by circadian rhythms, which play an important role in the daily regulation of plant physiological processes.

Transcriptome analyses of the circadian mechanisms and its influence on various metabolic processes were performed for such plants as Arabidopsis, Chinese cabbage and soybeans [3]–[5]. The genes of the central oscillator are conservative and their interactions are very similar in different plant species, nevertheless, there are species-specific differences. In addition, in most plant species, the genes of metabolic pathways affected by circadian rhythms are investigated very poorly.

Aim

Common wheat is an allopolyploid (genome AABBDD), so the presence of homoeologous genes in the genome can greatly affect the functioning and regulation of circadian rhythms. Literature analysis and the search in the databases did not find studies exploring the daily expression of the wheat transcriptome, information about the daily expression of the genes and their interaction is practically absent.

Methods

To identify genes that demonstrate rhythmic changes in expression during the day, we used the Chinese Spring variety transcriptome data obtained at four points during the day (0, 3, 9, 16 hours from the moment the light is turned on). Every timepoint was presented by a three biological replicates.

After preprocessing and quality control using the FASTX-toolkit utility the obtained reads were mapped onto the wheat reference transcriptome using Bowtie software. Kallisto v0.42.3 software on default parameters was used to quantifying abundances of transcripts and the trimmed mean of M-values (TMM)-normalized fragments per kb of Transcripts Per Million (TPM) reads mapped for each transcript. To identify rhythmically expressed genes, we applied the JTK-CYCLE circadian transcript analysis software.

Results

System analysis of the daily wheat transcriptome let us to identify genes expressed by the circadian type, and metabolic pathways under their control. It was shown, that significant part of the transcriptome is under the control of daily rhythms, and the expression of these genes can vary greatly depending on time. We detected some prominent expression patterns, characterized by peaks in different time points, and described genes underlying these patterns. We analyzed enrichment of gene ontology terms of different patterns and described major metabolic pathways in every group.

Acknowledgment

The study was funded by RFBR, project No. 20-316-80003.

References

[1] M. F. Covington, J. N. Maloof, M. Straume, S. A. Kay, and S. L. Harmer, “Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development”, Genome Biol., vol. 9, no. 8, 2008

[2] E. Yakir, D. Hilman, Y. Harir, and R. M. Green, “Regulation of output from the plant circadian clock”, FEBS J., vol. 274, no. 2, pp. 335–345, 2007

[3] S. L. Harmer et al., “Orchestrated transcription of key pathways in Arabidopsis by the circadian clock.”, Science, vol. 290, no. 5499, pp. 2110–2113, 2000

[4] B. Y. Chow and S. A. Kay, “Global approaches for telling time: omics and the Arabidopsis circadian clock.”, Semin. Cell Dev. Biol., vol. 24, no. 5, pp. 383–92, May 2013

[5] J. A. Kim et al., “Transcriptome Analysis of Diurnal Gene Expression in Chinese Cabbage”, Genes (Basel)., vol. 10, no. 2, p. 130, 2019

DOI 10.18699/BGRS/SB-2020-202

Identification of an AP2/ERF transcription factor controlling the synthesis of barley epicuticular wax

Ekaterina Kolosovskaya
ICG SB RAS, Novosibirsk, Russia
kolosovskaya@bionet.nsc.ru

Christian Hertig
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
Gatersleben, Germany
hertig@ipk-gatersleben.de

Dmitriy Domrachev
Novosibirsk Institute of Organic Chemistry, SB RAS
Novosibirsk, Russia
dmitry@nioch.nsc.ru

Alexey Kochetov
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
ak@bionet.nsc.ru

Sophia Gerasimova
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
gerson@bionet.nsc.ru

Sergey Morozov
Novosibirsk Institute of Organic Chemistry, SB RAS
Novosibirsk, Russia
moroz@nioch.nsc.ru

Vikhorev Alexander
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
vikhorev@bionet.nsc.ru

Jochen Kumlehn
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)
Gatersleben, Germany
kumlehn@ipk-gatersleben.de

Anna Korotkova
ICG SB RAS, Novosibirsk, Russia
korotkova@bionet.nsc.ru

Elena Chernyak
Novosibirsk Institute of Organic Chemistry, SB RAS
Novosibirsk, Russia
chernyak@nioch.nsc.ru

Nikolay Shmakov
ICG SB RAS, Novosibirsk, Russia
shmakov@ bionet.nsc.ru

Elena Khlestkina
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
Vavilov Institute of Plant Genetic Resources (VIR)
Saint Petersburg, Russia
khlest@bionet.nsc.ru

***Abstract* — Site-directed mutagenesis provides ample opportunity for reverse genetics, which allows us to establish a relationship between a gene and its function. In the present study, the role of the *HvWIN1* barley (*Hordeum vulgare*) gene in the formation of epicuticular wax was identified by taking a targeted knockout approach using RNA-guided Cas9 endonuclease technology.**

Keywords* — *barley, Hordeum vulgare, genome editing, Cas9, gRNA, epicuticular wax

Introduction

Reverse genetics is a set of approaches aiming to establish a relationship between a gene and a trait. As its starting point, this strategy has a gene with an unknown function and finds out how a change in the structure or activity of the gene affects the phenotype. The use of novel methods facilitating site-specific mutagenesis opens up new prospects for solving the problems of reverse genetics. One of the directions in this area is the use of the Cas9/gRNA system to study the functions of regulatory genes. In the present work, the poorly studied *HvWIN1* barley (*Hordeum vulgare*) gene, which encodes a transcription factor associated with resistance to biotic stress, was chosen as the target [1]. This gene belongs to the large family of plant-specific APETALA2/Ethylene-responsive factors (AP2/ERF) which play important roles in the regulation of many processes, including growth, development and stress response. In the present study, the targeted knockout of the *HvWIN1* gene has led to phenotypic abnormalities. Normally, starting from the booting stage, the surface of the flag leaf sheath of barley is covered with a well-visible wax layer. Mutants exhibit a recessive phenotype of epicuticular wax deficiency. To reveal the role of the transcription factor *HvWIN1* in the processes of epicuticular wax formation, a comparative microscopic, biochemical and transcriptome analysis of *win1* mutants vs wild- type plants was performed.

Materials and methods

Primary (M1=T0) barley mutants (cv. «Golden Promise») were obtained by targeted knockout of the *HvWIN1* gene using the Cas9/gRNA system. Three T-DNA-free mutant lines harboring different homozygous mutations in the *HvWIN1* gene were selected in the M4 generation. The ultrastructure of the wax deposited at the plant's surface was studied using scanning electron microscopy. In addition, the biochemical composition of the epicuticular layer was evaluated using gas chromatography coupled with mass spectrometry. In order to identify genetic mechanisms of phenotype formation, a comparative transcriptome analysis of leaf blades and leaf sheaths of *win1* mutant and wild-type plants was performed.

Results

Microphotographs and wax measurements showed that mutant plants and wild-type plants accumulate similar amounts of epicuticular wax on their leaf blades. By contrast, leaf sheath epicuticular wax of mutant plants differs remarkably from that of the wild-type in both total amount and composition. In particular, the amount of β-diketones is significantly reduced in mutants. An analysis of transcriptome data revealed that the mutation leads to a decrease in the activity of 326 genes and an increase of 292 in the leaf blade and a decrease of 63 and an increase of 967 in the leaf sheath. Among the genes showing reduced expression in the mutants were those previously known to be associated with the synthesis of epicuticular wax components. Among the genes that increased their expression in mutants, many are associated with cell wall biogenesis, modification and organization as well as with various responses to stresses, including pathogen attack and impact of reactive oxygen species. Additionally, many genes have been assigned to the gene ontology terms "extracellular space" and "cell membrane".

Conclusion

The nature of the mutant phenotype and the data of comparative transcriptome analysis suggest that the transcription factor *HvWIN1* normally regulates the formation of the cuticular layer at the surface of leaf sheaths of upper barley leaves.

Acknowledgment

Supported by the RSF (16-14-00086). Research was conducted using equipment of The Core Facilities Center "Microscopic analysis of biological objects" at the Central Siberian Botanical Garden SB RAS (Novosibirsk, Russia).

References

[1] A. Kumar, K.N. Yogendra, S. Karre, A.C. Kushalappa, Y. Dion, and T.M. Choo, "*WAX INDUCER1* (*HvWIN1*) transcription factor regulates free fatty acid biosynthetic genes to reinforce cuticle to resist Fusarium head blight in barley spikelets", J Exp Bot., vol. 67(14), pp. 4127-39, July 2016.

DOI 10.18699/BGRS/SB-2020-203

Spikes morphometric characteristics analysis of five species of wheat

Evgeniy Komyshev
ICG SB RAS, Kurchatov Genomic Center, Novosibirsk, Russia
komyshev@bionet.nsc.ru

Mikhail Genaev
ICG SB RAS, Kurchatov Genomic Center, Novosibirsk, Russia
NSU, Novosibirsk, Russia
mag@bionet.nsc.ru

Dmitry Afonnikov
ICG SB RAS, Kurchatov Genomic Center, Novosibirsk, Russia
NSU, Novosibirsk, Russia
ada@bionet.nsc.ru

Yuliya Kruchinina
ICG SB RAS, Kurchatov Genomic Center, Novosibirsk, Russia
kruchinina2014@bionet.nsc.ru

Vasiliy Koval
ICG SB RAS, Kurchatov Genomic Center, Novosibirsk, Russia
kovalvs@icg.sbras.ru

Nikolay Goncharov
ICG SB RAS, Novosibirsk, Russia
gonch@bionet.nsc.ru

***Abstract* — A method for recognition and morphometry of a wheat spike based on the analysis of 2D digital images has been developed. The proposed approach showed high accuracy for determining the qualitative and quantitative characteristics of the spike. An analysis of the obtained characteristics of five species of wheat showed that they are divided into three main clusters characterizing the spike shape.**

Keywords* — *image analysis, pattern recognition, morphometry, spikes of wheat

Motivation and aim

Motivation

The shape and structure of the spike are one of the most important characteristics of cereals associated with their economically valuable qualities such as productivity, lack of fragility of the spike and ease of threshing. Studying the genes that control these characters will allow to purposefully create new varieties with improved characteristics in terms of yield, ease of threshing and resistance to environmental factors [1]. However, the complexity of research in this area is due to the polyploidy of the wheat genome [2] and the need to obtain phenotypic data from thousands of plants. For example, in the work of Boden et al. [3] to identify loci associated with the structure of a spike (inflorescence), more than 13 thousand spikes were analyzed.

Aim

We recently developed method for morphometry of the wheat spike [4] for analysis of the shape of spikes based on the 2D image analysis.

We apply our algorithm for analysis of the five species of wheat: Triticum aestivum, Triticum spelta, Triticum compactum, Triticum sphaerococcum, Triticum antiquorum.

Methods

We propose a wheat spike recognition method based on digital image analysis. Digital images are acquired in two variants: a spike on a table (one projection) or fixed with a clip (four projections). The method identifies spike and awns in the image and estimates their quantitative characteristics (area in image, length, width, circularity, etc.). Models of sections, quadrilaterals, and radial model are proposed for describing spike shape. Parameters of these models are used to predict spike shape type (spelt, normal, or compact) by machine learning.

Results

This method makes it possible to extract several traits associated with spike shape, its awns, and color characteristics. The proposed approach has shown high performance in identifying the image regions pertaining to the spike and its awns. Several models are proposed for describing the shape of wheat spike.

The mean error in spike density prediction for the images in one projection is 4.61 (~18%) versus 3.33 (~13%) for the parameters obtained using four projections. F1 measure in automated spike classification into three types is 0.78 using logistic regression (one projection) and 0.85 using random forest method (four projections).

The analysis of the obtained characteristics of five wheat of wheat showed that they are divided into three main clusters characterizing the linear dimensions of the spike, area projected by the spike and the shape of the spike contour.

Acknowledgment

The work supported by the Kurchatov Genomic Center, Institute of Cytology and Genetics, SB RAS, agreement № 075-15-2019-1662 with the Ministry of Education and Science.

References

[1] Konopatskaia I.D., Vavilova V.Y., Blinov A.G., Goncharov N.P. (2016) Spike morphology genes in wheat species (Triticum L.). Proceedings of the Latvian Academy of Sciences. Section B. Natural, Exact, and Applied Sciences. De Gruyter Open. 70(6): 345-355.

[2] Borrill P., Harrington S. A., Uauy C. (2019) Applying the latest advances in genomics and phenomics for trait discovery in polyploid wheat. The Plant Journal. 97(1): 56-72.

[3] Boden S. A. et al. (2015) Ppd-1 is a key regulator of inflorescence architecture and paired spikelet development in wheat. Nature Plants. 1(2): 14016.

[4] Genaev M.A. et al. (2019) Morphometry of the Wheat Spike by Analyzing 2D Images. Agronomy 9(7): 390.

DOI 10.18699/BGRS/SB-2020-204

Regulation of Transcription Activity of *MAKR4* in *Arabidopsis thaliana* L.

Anastasia Korosteleva
Novosibirsk State University
Novosibirsk, Russia
kartzeva.kar@gmail.com

Daria Novikova
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
da6ik777@gmail.com

Victoria Mironova
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
kviki@bionet.ns.ru

***Abstract* — Auxin has a crucial influence on genes expression. In this work, we show that transcriptional activity of an auxin-sensitive gene *MAKR4 (MEMBRANE ASSOSIATED KINASE REGULATOR 4)* depends on the composite elements in its upstream regulatory region. Analysis of such elements could clarify the understanding of the regulation of this gene and other auxin-sensitive genes.**

Keywords* — *MEMBRANE ASSOCIATED KINASE REGULATORS, lateral roots, auxin, Arabidopsis thaliana

Introduction

Genes of MEMBRANE ASSOCIATED KINASE REGULATORS family play an important role in plant morphogenesis. One of the family members, *MAKR4*, is implicated in lateral root formation. *MAKR4* was shown to convert the prebranch sites into regular lateral organs [1]. It is assumed that the accumulation of MAKR4 on the plasma membrane of pericycle cells triggers the processes which leads to the formation of a lateral root.

In the promoter region of *MAKR4* auxin-associated cis-elements were predicted (Auxin Response Elements, AuxRE). Some of these were predicted to bind ARF transcription factors (AUXIN RESPONSE FACTORS) and some - transcription factors from other families (non-ARF). Possibly, these proteins together provide auxin response of *MAKR4*. ARFs are known to form homodimers and heterodimers with each other and transcription factors from other families [2,3]. The interaction of ARFs and potential partner-proteins in regulation of transcriptional response to auxin could explain formation of the complex *MAKR4* expression pattern in *Arabidopsis thaliana*.

Results

The predicted AuxREs in *MAKR4* upstream regulatory region were annotated with a variety of bioinformatics tools.

We obtained homozygous reporter lines of *Arabidopsis thaliana* plants with *GFP* expressed under the intact promoter of *MAKR4*, as well as under promoter versions with mutated predicted cis-elements.

The expression pattern of GFP was studied with fluorescence microscopy in the control plants and after three-hour auxin treatment. Arabidopsis lines with an intact *MAKR4* promoter had a significant increase in the expression level of GFP. In Arabidopsis lines with mutated predicted cis-elements (AuxRE or non-AuxRE), a significant increase in the expression of GFP did not occur. Lines with two or three cis-elements mutated at the same time had a low level of GFP expression and, thus, did not posess response to auxin.

Analysed cis-elements in *MAKR4* upstream regulatory region have an impact on the auxin-mediated gene expression.

Acknowledgment

This work was supported by RFBR 19-44-543006.

References

[1] Xuan, W., Audenaert, D., Parizot, B., Möller, B. K., Njo, M. F., De Rybel, B., ... & Beeckman, T. (2015). Root cap-derived auxin pre-patterns the longitudinal axis of the Arabidopsis root. Current Biology, 25(10), 1381-1388.

[2] Boer, D. R., Freire-Rios, A., van den Berg, W. A., Saaki, T., Manfield, I. W., Kepinski, S., ... & Weijers, D. (2014). Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. Cell, 156(3), 577-589.

[3] Simonini, S., Deb, J., Moubayidin, L., Stephenson, P., Valluru, M., Freire-Rios, A., ... & Østergaard, L. (2016). A noncanonical auxin-sensing mechanism is required for organ morphogenesis in Arabidopsis. Genes & development, 30(20), 2286-2296.

DOI 10.18699/BGRS/SB-2020-205

New insight on diversity of the Nikita Botanical Gardens plant collections from advanced NGS technology

Irina Mitrofanova
NBG-NSC RAS, Yalta, Russia
irimitrofanova@yandex.ru

Olga Krivenko
NBG-NSC RAS, Yalta, Russia
olkrivenko@gmail.com

Olga Kuleshova
NBG-NSC RAS, Yalta, Russia
v_olgo4ka@inbox.ru

Svetlana Chelombit
NBG-NSC RAS, Yalta, Russia
chelombit@inbox.ru

Valentina Brailko
NBG-NSC RAS, Yalta, Russia
labgennbs@yandex.ru

Olga Mitrofanova
NBG-NSC RAS, Yalta, Russia
invitro_plant@yandex.ru

Abstract — **New scientific program of genomic screening of the Nikita Botanical Gardens plant collections targeted at comprehensive description of their molecular diversity in close connection with morphological, physiological and others phenotype features has started. Strategic goal is to develop and implement advanced molecular approaches for plant biotechnology, breeding, propagation, and conservation routine works. Two whole-genome sequencing projects to produce a high quality assembly of the reference genomes of the common fig *Ficus carica* L. (Moraceae) and Crimean endemic species *Heracleum ligusticifolium* M. Bieb. (Apiaceae) are in progress. Several hybrid assembly approaches using combined data sets with long and short reads obtained by Oxford Nanopore MinION and Illumina NextSeq550 instruments are tested to create accurate and well-resolved reference genome. The first versions of the reference genomes of *F. carica* cultivar 'Sabrutsiya Rozovaya' and *H. ligusticifolium* were made, and their annotations are performing. Subsequent improvement and development of the primary genome assemblies up to "pan-genome" rank will be carried out using advanced bioinformation approaches and computational tools by assimilation of new sequenced genomes and transcriptomes of the model species taken from the NBS-NSC collections and from nature biotopes as well. Such integrative reference genome construction will provide an important resource both for advanced molecular breeding and fundamental researches in plant biology and ecology based on genomic approaches using *F. carica* and *H. ligusticifolium* as model species.**

Keywords — NGS technologies, hybrid genome assembly, Ficus carica, Heracleum ligusticifolium

Introduction

Modern next-generation sequencing technologies (NGS) provide new opportunities for plant genomics research and allow to advance in the field of biotechnology and molecular breeding. Despite a wide range of the targets, methodologies and tools, high quality genome assembly is a key requirement for a successful implementation of the projects associated with whole-genome studies [1]. Additionally, some issues, for example in the field of molecular breeding, require obtaining the most complete information about intraspecies genetic diversity (the so-called "pangenom") or the genomic architecture of the genus ("superpangenom"). The number of plants with a well annotated genome is currently limited to a few dozen of model and / or economically valuable species which significantly limits the use of advanced genomic tools for genotyping, targeted search of quantitative trait locus (QTL) and realization of molecular breeding methodology in general [2]. Development of high-throughput NGS technologies was accompanied by a constant decline of analysis cost. And now they are already affordable for comprehensive scientific researches, so the global screening of plant genomic diversity is on the agenda.

The Nikita Botanical Gardens (NBG) is one of the oldest scientific and breeding centers in Russia with thousands of species and cultivars of plants that are grown as *ex situ* as *in vitro* conditions. The NBG plant collection incorporates extensive gene pool of garden, essential-oil, ornamental and other valuable crops, as well as wild plants, including rare and endemic species. Scientific researches based on these collections have been ongoing for more than 150 years. Various biotechnological methods of plant propagation and conservation have been used for the past three decades. Regular molecular-genetics investigations were started a few years ago and several projects are underway. Genotyping of the valuable crops, phytopathogens analyzing, development of new breeding approaches with searching for QTL and marker selection consist the main applied purposes. In addition, studies in the field of functional genomics are started. Entire transcriptomes for several plant species were sequenced to investigate the molecular mechanisms of plant adaptation to various stress factors, as well as cell regulation of somatic embryogenesis and regeneration based on analysis of differential gene expression. All these studies are integrated in the framework of scientific program of genomic screening of the NBG-NSC plant collections targeted at comprehensive description of their molecular diversity in close connection with morphological, physiological and others phenotype features. Strategic goal of the program is to develop and implement advanced molecular approaches for plant biotechnology, breeding, propagation and conservation routine works. Creation of high-quality reference genomes is the key stage for the subsequent successful implementation of the project aims.

Modeling plant species

Two whole-genome sequencing projects to produce high quality assembly of the reference genomes of the common fig *Ficus carica* L. (Moraceae) and Crimean endemic species *Heracleum ligusticifolium* M. Bieb. (Apiaceae) have started. Common fig is one of the most important subtropical fruits used in food, cosmetic, and soap industries as well as in medicine. The NBG has a large *F. carica* collection which includes 267 cultivars and hybrid forms. A special research program aimed at a comprehensive description of the genetic diversity of the common fig collection, including sequencing of the key cultivars, genotyping of the collection, QTL analysis of the valuable horticultural traits are forming.

Heracleum is the largest genus of the Apiaceae family with a broad spectrum of phytochemistry and biological activities but without any whole-genome sequencing data. *H. ligusticifolium* possesses a number of unique adaptive properties, including those associated with somatic embryogenesis, which ensure the successful survival of this endemic species within a highly restricted distribution area under the extremely adverse environmental conditions. To date assembling of the high-quality reference *Heracleum* genome is the most critical step to advance molecular researches not only for the Crimean endemic *H. ligusticifolium* but also for the genus as whole. Deciphering the biochemical pathways caused by stress and the targeted search for specific genes involved in adaptive reactions are the main tasks for the future investigation after the creation and subsequent annotation of the *H. ligusticifolium* genome.

Results

Leaves of microshoots in *F. carica* cultivar 'Sabrutsiya Rozovaya' and in *H. ligusticifolium* cultured *in vitro* were used to high molecular weight DNA isolation using GeneJET Genomic DNA Kit by standard protocol. The entire genomes have been sequenced using Oxford Nanopore MinION and Illumina NextSeq550 instruments as well. The genomic data were preliminary assembled in Canu v.1.8 software package. Oxford Nanopore technology allows to get extremely long reads, but such readings are known for having too high error rates. Whereas accurate short-reads sequencing by Ilumina does not allow to resolve genome regions with high repeated and duplicated sequences, especially in the complex and polyploid plant genomes. A hybrid assembly using combined data sets with long and short reads is a promising approach to create an accurate and well-resolved reference genome [3]. Several workflows including pipelines for a haplotype-resolved *de novo* genome assembly have been tested. Assessment of assembly genome quality in terms of contiguity, accuracy, and completeness was done. As a result, the first versions of the reference genomes of *F. carica* cultivar 'Sabrutsiya Rozovaya' and *H. ligusticifolium* were made, and their annotations are performing. Subsequent improvement and development of the primary genome assemblies up to "pan-genome" rank will be carried out using advanced bioinformation approaches and computational tools by assimilation of new sequenced genomes and transcriptomes of the model species taken from the NBS-NSC collections and from nature biotopes as well. Such integrative reference genome construction will provide an important resource both for advanced molecular breeding and fundamental researches in plant biology and ecology based on genomic approaches [4] using *F. carica* and *H. ligusticifolium* in our laboratory as model species.

Acknowledgment

The studies are done at the Kurchatov Genomic Center – NBG-NSC of the RAS and supported by the grant № 075-15-2019-1670 of the Ministry of Science and Higher Education of the Russian Federation.

References

[1] Bailey-Serres J. et al (2019) Genetic strategies for improving crop yields. Nature. 575(7781): 109–118. doi:10.1038/s41586-019-1679-0.

[2] Yuan Y. et al. (2017) Improvements in Genomic Technologies: Application to Crop Genomics. Trends in Biotechnology. 35 (6): 547 – 558.

[3] Jung H. et al. (2019) Tools and Strategies for Long-Read Sequencing and De Novo Assembly of Plant Genomes. Trends in Plant Science. doi:10.1016/j.tplants.2019.05.003

[4] Usai G. et al. (2020) Epigenetic patterns within the haplotype phased fig (*Ficus carica* L.) genome. Plant J. doi:10.1111/tpj.14635

DOI 10.18699/BGRS/SB-2020-206

Amyloidogenic properties of the beta-barrel proteins and their involvement in storage of nutrients in plant seeds and bacteria virulence

A.A. Nizhnikov
Laboratory for Proteomics of Supra-Organismal Systems, All-Russia Research Institute for Agricultural Microbiology (ARRIAM), Podbelskogo sh., 3, Pushkin, 196608 St. Petersburg, Russia
Faculty of Biology, St. Petersburg State University, 199034 St. Petersburg, Russia
lab7@arriam.ru

***Abstract* — Amyloid proteins are known to be involved in a number of lethal diseases in humans and perform various vital functions in archaea, bacteria and eukaryote. Here we present recent findings of previously unknown amyloids of plants and symbiotic root nodule bacteria and discuss their physiological roles and association with formation of beta-barrel structures.**

Keywords* — *amyloid, plant, seed, protein, fibril, root nodule bacteria

Amyloids are proteins capable of forming highly ordered protein fibrils with characteristic structure called cross-beta. These fibrils are of significant interest due to their involvement both in various incurable pathogeneses in humans and different vital functions in archaea, bacteria and eukaryotes. Despite their significance, amyloids have not been identified in plants (with exception of several proteins capable of forming amyloid fibrils *in vitro* [1]) and symbiotic bacteria. We have started our studies of plant and symbiotic bacteria amyloids with prediction of potentially amyloidogenic proteins in the proteomes of these organisms. Large scale studies performed on the proteomes of all land plants with annotated genomes and more than 80 species of bacteria belonging to the order *Rhizobiales* have revealed strong enrichment with amyloidogenic regions predicted by Waltz [2] and SARP [3] algorithms of the plant seed storage proteins containing evolutionary conservative Cupin-1 domain [4] and various virulence factors of *Rhizobiales* [5]. Further experimental analyses of the structural properties of the predicted functional groups of plant and bacterial proteins performed by our laboratory with collaborators have confirmed amyloid properties of the RopA and RopB outer membrane protein of the *Rhizobium leguminosarum* root nodule bacterium [6] and Vicilin seed storage protein of garden pea, *Pisum sativum* L. [7] representing the first to our knowledge examples of amyloid proteins that are formed under physiological conditions in symbiotic bacteria and plants, respectively. Notably, despite evolutionary distance, amyloid proteins of plants and root nodule bacteria have exhibited an unusual example of structural convergence being capable of forming beta-barrel structures. Probable biological functions of plant and root nodule bacteria amyloids, their structural properties and interrelations between beta-barrel and amyloid formation are discussed.

Acknowledgment

This work was supported by the Russian Science Foundation, Project 17-16-01100.

References

[1] Antonets K.S., Nizhnikov A.A. (2017) Amyloids and prions in plants: facts and perspectives. Prion. 11: 300-312.

[2] Maurer-Stroh S., *et al.* (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. Nature Methods. 7: 237-242.

[3] Antonets K.S., Nizhnikov A.A. (2013) SARP: a novel algorithm to assess compositional biases in protein sequences. Evolutionary Bioinformatics. 9: 263-273.

[4] Antonets K.S., Nizhnikov A.A. (2017) Predicting amyloidogenic proteins in the proteomes of plants. International Journal of Molecular Sciences. 18: e2155.

[5] Antonets K.S. *et al.* (2018) Exploring proteins containing amyloidogenic regions in the proteomes of bacteria of the order *Rhizobiales*. Evolutionary Bioinformatics. 14: 1-12.

[6] Kosolapova A.O. *et al.* (2019) Two Novel Amyloid Proteins, RopA and RopB, from the Root Nodule Bacterium *Rhizobium leguminosarum.* Biomolecules. 9: e694.

[7] Antonets K.S. *et al.* (2019) Accumulation of storage proteins in plant seeds is mediated by amyloid formation. BioRxiv, Preprint 825091. doi: https://doi.org/10.1101/825091

DOI 10.18699/BGRS/SB-2020-207

Identifying novel elements and regulators in auxin-dependent gene expression

Daria Novikova
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
da6ik777@gmail.com

Nadezda Omelyanchuk
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
nadya@bionet.nsc.ru

Dolf Weijers
Wageningen University and Research
Wageningen, Netherlands
dolf.weijers@wur.nl

Victoria Mironova
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
kviki@bionet.nsc.ru

Abstract — **Auxin is a key phytohormone which controls various aspects of plants growth and development. Basic mechanism of auxin signal transduction was broadly studied; however, it is not sufficient to explain the variety of auxin-mediated transcriptional responses. We proposed a hypothesis that upstream regulatory regions landscape provide the diversity of auxin responses.**

Keywords — cis-regulatory elements, differentially expressed genes, transcription, auxin, Arabidopsis thaliana

Introduction

Phytohormone auxin regulates plants growth and development in a specific manner. Auxin responses are implemented via auxin response factors (ARFs) which bind to TGTC-containing auxin response elements (AuxREs) in upstream regulatory regions of auxin-responsive genes. However, the signaling beyond ARF-dependent response in largely unknown. Non-ARF transcription factors were shown to be involved in auxin-mediated processes regulation [1], but systematic study of the phenomena is still required [2].

Meta-analysis of biological data is widely used nowadays to perform systematic study of different cases. It can be applied to the big data in omics, like transcriptome, metabolome and proteome profiles, etc. Nowadays thousands of transcriptomes are publicly available, many of these study the same phenomena providing the material for the identification of cis-regulatory elements associated with differential expression of genes under specific conditions. Prediction of cis-regulatory elements associations, in turn, leads to the investigation of potential transcription factors related to the process of interest.

Results

We developed an R package *metaRE* which provides search for the cis-regulatory elements of a particular structure enriched in the promoters of differentially expressed genes in response to a stimulus [3]. Meta-analysis of multiple expression profiles provides identification of cis-elements associated with gene up- or down-regulation in response to a stimulus. Meta-analysis of 21 auxin-induced transcriptome datasets was performed to search for single and bipartite elements enriched in upstream regulatory regions of Arabidopsis auxin-responsive genes [3]. A list of cis-regulatory elements associated with auxin up- and down-regulation in early or late response was identified. Interestingly, the most highly overrepresented cis-elements were bHLH- and bZIP- binding sites and AT-rich sequences. For 40 genes driven by the regulatory regions with predicted bipartite elements auxin-inducibility was supported by qPCR. Functionality of six predicted bipartite cis-elements was studied on reporter lines which genomes contain GFP driven by intact promoters of genes and promoters with mutated predicted bipartite elements. Functionality of four predicted bipartite elements was confirmed. Two of these are located in the upstream regulatory region of one gene in close proximity forming cis-regulatory module. We experimentally identified interactions of this regulatory module with transcription factors and tested interactions of these with ARF1 and ARF5 proteins.

Diversity of single and bipartite AuxREs was described and annotated. For some of the predicted cis-elements relation to auxin transcriptional regulation was shown first time. Mechanism underlying two bipartite elements functionality was studied in more detail.

Acknowledgment

This work was supported by RFBR 19-44-543006 and Sandwich PhD scholarship from Wageningen University and Research.

References

[1] Weijers D, Wagner D. 2016. Transcriptional Responses to the Auxin Hormone. Annual Review of Plant Biology 67,539-574.

[2] Lanctot A, Taylor-Teeples M, Oki EA, Nemhauser J. 2020. Specificity in auxin responses is not explained by the promoter preferences of activator ARFs. Plant Physiology. p.01474.

[3] Cherenkov P, Novikova D, Omelyanchuk N, Levitsky V, Grosse I, Weijers, D, Mironova V. (2018). Diversity of cis-regulatory elements associated with auxin response in Arabidopsis thaliana. Journal of experimental botany 69, 329-339.

DOI 10.18699/BGRS/SB-2020-209

Transcripts specifically expressed during secondary vascular development in *Arabidopsis thaliana* L.

Nadezda Omelyanchuk
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
Novosibirsk State University
Novosibirsk, Russia
nadya@bionet.nsc.ru

Evgeniya Pukhovaya
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
Novosibirsk State University
Novosibirsk, Russia
e.pukhovaya@gmail.com

Dmitry Oshchepkov
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
diman@bionet.nsc.ru

Victoria Mironova
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
Novosibirsk State University
Novosibirsk, Russia
kviki@bionet.nsc.ru

***Abstract* — Genes specifically expressed in only limited number of tissues can serve as biomarkers for these tissues and have important functions in development. In this work, we provide the analysis of specifically expressed genes in vascular tissues of *Arabidopsis thaliana*. Since these genes have mostly unknown functions, this methodology can shed light on functional annotation for new regulators of tissue development.**

Keywords* — *Specifically expressed genes, differentially expressed genes, transcription, vasculature, Arabidopsis thaliana

Introduction

As a result of the increase in the accuracy, next generation sequencing technologies allow now studying not only differentially expressed genes (DEGs) but also so-called specifically expressed genes (SEGs) that are not transcribed in one or more tissues but have reads in others [1]. SEGs are becoming widely used in medical research as highly specific and sensitive biomarkers [2], [3], [4]. Also recognition of SEGs starts to be an important issue in developmental biology. Here we have elaborated the methodology for identification of SEGs using RNA-Seq datasets taken from overlapping tissue domains. The methodology was applied to distinguish and study SEGs in the secondary vasculature of Arabidopsis stems.

In Arabidopsis, the primary vasculature consists of bundles where cambial stem cells (fascicular cambium) intervene xylem and phloem, transporting water and nutrients, respectively (reviewed in [5]). In secondary vascular development, interfascicular cambium forms between primary bundles generating altogether with the fascicular cambium the continuous cylinder of vascular cambium, which produces xylem and phloem, inward and outward, respectively, similar to wooden plants.

Vascular tissues have specific markers for the following (in some cases partly overlapping) domains: PXY/TDR, PHLOEM INTERCALATED WITH XYLEM/TDIF RECEPTOR for proximal cambium and the part of differentiating xylem [6]; SMXL5, SMAX1-LIKE 5 for distal cambium and differentiating phloem [6], [7]; APL, ALTERED PHLOEM DEVELOPMENT for differentiated phloem [8]; VND7, VASCULAR RELATED NAC DOMAIN PROTEIN 7 for differentiating xylem vessels [9], [10], and NST3, NAC SECONDARY WALL THICKENING PROMOTING 3 for xylem fibers [11]. Fluorescent reporters driven by the promoters of these markers were employed in fluorescence-activated nucleus sorting followed by RNA sequencing [12]. We used these data for the identification of SEGs.

Results

The raw datasets were preprocessed to calculate FPKM values and estimate the threshold for the presence/absence of transcription. Among 32650 genes transcribed in the vasculature – 84% were expressed in all vascular tissues and 16% were not expressed at least in one tissue and belong to SEGs. Comparison of these two gene groups shows that ubiquitously expressed genes are mainly known genes or members of known families, whereas 57% of SEGs encode unknown and hypothetical proteins and non-coding RNAs. Identification of these genes as SEGs in our study and description of their specific expression domains in the vasculature can be one of the first steps in their functional annotation. Also, markers for specific domains were predicted among known genes including microRNAs and genes involved in hormonal pathways.

Acknowledgment

This work was supported by RFBR-DFG 19-54-12013.

References

[1] Tao, X., Gu, Y. H., Wang, H. Y., Zheng, W., Li, X., Zhao, C. W., & Zhang, Y. Z. (2012). Digital gene expression analysis based on integrated de novo transcriptome assembly of sweet potato [Ipomoea batatas (L.) Lam.]. PloS one, 7(4).

[2] J Tang, Q., Zhang, Q., Lv, Y., Miao, Y. R., & Guo, A. Y. (2019). SEGreg: a database for human specifically expressed genes and their regulations in cancer and normal tissue. Briefings in bioinformatics, 20(4), 1322-1328.

[3] Lv Y, Lin SY, Hu FF, Ye Z, Zhang Q, Wang Y, Guo AY. (2019). Landscape of cancer diagnostic biomarkers from specifically expressed genes. Brief Bioinform. Dec 8. pii: bbz131.

[4] Zhang, Q., Liu, W., Liu, C., Lin, S. Y., & Guo, A. Y. (2018). SEGtool: a specifically expressed gene detection tool and applications in human tissue and single-cell sequencing data. Brief Bioinform, 19(6), 1325-1336.

[5] Fischer, U., Kucukoglu, M., Helariutta, Y., & Bhalerao, R. P. (2019). The dynamics of cambial stem cell activity. Annual review of plant biology, 70, 293-319.

[6] Shi, D., Lebovka, I., Lypez-Salmeryn, V., Sanchez, P., & Greb, T. (2019). Bifacial cambium stem cells generate xylem and phloem during radial plant growth. Development, 146(1).

[7] Wallner, E. S., Lypez-Salmeryn, V., Belevich, I., Poschet, G., Jung, I., Grunwald, K., Sevilem I, Jokitalo E, Hell R, Helariutta Y, Agustí J, Lebovka I, Greb T. (2017). Strigolactone-and karrikin-independent SMXL proteins are central regulators of phloem formation. Current Biology, 27(8), 1241-1247.

[8] Bonke, M., Thitamadee, S., Mahonen, A. P., Hauser, M. T., & Helariutta, Y. (2003). APL regulates vascular tissue identity in Arabidopsis. Nature, 426(6963), 181-186.

[9] Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., Mimura T., Fukuda H. & Demura, T. (2005). Transcription switches for protoxylem and metaxylem vessel formation. Genes & development, 19(16), 1855-1860.

[10] Yamaguchi, M., Mitsuda, N., Ohtani, M., Ohme-Takagi, M., Kato, K., & Demura, T. (2011). VASCULAR-RELATED NAC-DOMAIN 7 directly regulates the expression of a broad range of genes for xylem vessel formation. The Plant Journal, 66(4), 579-590.

[11] Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K., & Ohme-Takagi, M. (2007). NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis. The Plant Cell, 19(1), 270-280.

[12] Shi, D., Jouannet, V., Agusti, J., Kaul, V., Levitsky, V., Mironova V., Sanchez, P. & Greb, T. (2020). Tissue-specific transcriptome profiling of the Arabidopsis thaliana inflorescence stem reveals local cellular signatures. bioRxiv.

DOI 10.18699/BGRS/SB-2020-210

Meta-analysis of transcriptome data clarified hormonal regulation of cold stress response in *Arabidopsis thaliana* L.

Nadezda Omelyanchuk
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
nadya@bionet.nsc.ru

Yana Sizentsova
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
sizentsova.yans@gmail.ru

Victoria Mironova
Institute of Cytology and Genetics SB RAS,
Novosibirsk State University,
Novosibirsk, Russia
kviki@bionet.nsc.ru

***Abstract* — Low positive temperatures not only have damaging effects on plants in various climate from tropics to temperate zones but also provide acclimation to freezing for many plant species. It is why cold stress response is an actual topic for a wide range of experiments, including whole-genome studies. The results of these experiments are diverse and, in many cases, even contradictory. A meta-analysis of whole-genome data is a useful approach to distinguish between the robust effects of low temperature itself and the influence of particular experimental conditions. Here we applied meta-analysis to transcriptome data sets on cold stress response in the model plant species Arabidopsis thaliana, to find out changes in plant hormone metabolism and signaling accompanying this process. We identified that cold stress dramatically changes the plant hormonal status. Here we discuss the dynamics of the changes for each hormone.**

Keywords* — *cold stress, phytohormones, transcriptome, meta-analysis, Arabidopsis thaliana

Introduction

Cold stress greatly limits geographic distribution of agricultural crops. Plants from temperate zones have evolved a number of mechanisms to cope with chilling and freezing stresses, including delay in germination and flowering timing, inhibition of growth etc. Plant hormones are key intermediators between changes in environment and plant developmental reaction to them. Nevertheless, our knowledge on hormonal input into reaction to low positive temperature is enough patchy [1]. For instance, the role of auxin, the major regulator of plant growth and development, has not been studied in response to cold, although cold has been shown to increase expression of auxin-regulated genes and increase auxin accumulation. Brassinosteroids involvement in plant growth and development during cold stress in also unclear. One of the main reasons for this are intensive cross-talks between hormones themselves and their involvement in reactions to other environmental stimuli. The meta-analysis of many transcriptome experiments (microarrays and RNA-seqs) helps to exclude the differences triggered by variations in experimental design and to highlight the most robust pattern for involvement of hormones in cold response.

materials and Methods

We compiled a list of publically available datasets on cold stress response in *Arabidopsis thaliana* (40 datasets in total). Microarray raw data was processed using RMA algorithm implemented in the Bioconductor *affy* [2] and *limma* [3] packages. RNA-Seq data pre-processing and TMM normalization was done using *edgeR* package [4]. As a result, we compiled two matrixes for all 40 datasets with information about DEGs: (1) with log2FC and (2) False discovery rates (FDR). To compare the datasets we applied PCA and hierarchical cluster analysis using *stats* R package [5].

Then we selected the genes that robustly activated or inhibited in response to cold stress over a number of datasets (the threshold was calculated by binomial test).

Functional enrichment analysis was done using DAVID v6.8 [6].

Results

Firstly, we performed functional annotation of the robustly up or down-regulated genes in either early, late or very late response. Among the overrepresented GO categories we paid especial attention to those associated with hormonal metabolism and signaling. The results are summarized on the figure 1. Functional annotation showed that pathways of almost all hormones were activated at early stage of cold stress except auxin and citokinin pathways. Citokinin metabolism and signaling were only activated at very late stages of cold stress, whereas auxin signaling was both up- and down-regulated.

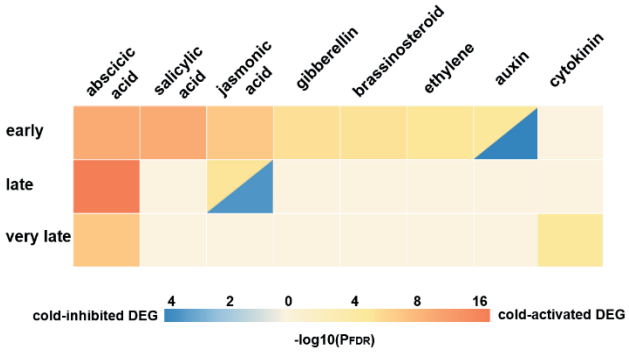


Fig. 1. Functional annotation results for early, late and very late cold response in *Arabidopsis thaliana*. Only GO terms associated with hormonal control are shown.

As Gene Ontology is rather incomplete and redundant, we compiled a list of genes mediating synthesis, degradation, conjugation, transport and synthesis of main plant hormones to better understand the gene expression changes in hormone pathways. We analyzed which of them robustly regulated by cold in early, late or very late response. The analytical results are summed up below.

In the early phase of response to the cold stress, abscisic acid, brassinosteroid and ethylene signaling increases whereas jasmonic acid and auxin signaling, in opposite, decreases. In the late phase fine-tuning of these signaling occurred, which means that each of these pathways has at least one upregulated component along with at least one suppressed. For most of

hormones, fine-tuning of signaling continues in the very late phase, where solely auxin and gibberellin signals increase and decrease, respectively. Cytokinin and salicylic acid signaling starting correspondingly from the early and late phases of cold response have only “fine-tuning” mode of regulation. Similar to salicylic acid, gibberellin signaling begins from the late phase.

Many of these changes in signaling may come from the changes in biosynthesis, degradation and/or conjugation and distribution of hormones. First, in the early phase, cold activates biosynthesis of abscisic and jasmonic acids, ethylene, gibberellin and auxin, and then in the late phase cytokinin biosynthesis is increased. While at the early phase, only activation of hormone biosynthesis takes place at later stages suppression (brassinosteroids) and fine-tuning (abscisic, jasmonic and salicylic acids and gibberellin) are switching on. Ethylene biosynthesis was activated, fine-tuned and suppressed at the early, late and very late phases, respectively.

Both degradation and conjugation are processes leading to decreasing of the hormone level. Degradation and/or conjugation starts in the early phase for abscisic acid, brassinosteroids and cytokinin, and later is inhibited for brassinosteroids and fine-tuned for abscisic and jasmonic acids, cytokinin and auxin. At the early phase of cold response, changes in metabolism occur for all hormones under study except salicylic acid, which react to cold later.

Thus, with some exceptions, cold activates hormone metabolism and/or signaling at the early phase of cold response fine-tuning them at the later stages possibly mainly due to intensive crosstalk between signaling and metabolic pathways of various hormones.

Acknowledgments

This work was supported by Russian Science Foundation 18-74-10008.

References

[1] M. Eremina, W. Rozhon, and B. Poppenberger, "Hormonal control of cold stress responses in plants", (in eng), *Cell Mol Life Sci,* vol. 73, no. 4, pp. 797-810, Feb 2016.

[2] L. Gautier, L. Cope, B. M. Bolstad, and R. A. Irizarry, "Affy-analysis of Affymetrix GeneChip data at the probe level", Bioinformatics, vol. 20, no. 3, pp. 307-15, Feb 2004.

[3] M. E. Ritchie et al., "Limma powers differential expression analyses for RNA-sequencing and microarray studies", Nucleic Acids Res, vol. 43, no. 7, p. e47, Apr 2015.

[4] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data", Bioinformatics, vol. 26, no. 1, pp. 139-40, Jan 2010.

[5] R Core Team, “R: A language and environment for statistical computing”, R Foundation for Statistical, 2019.

[6] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists", Nucleic Acids Res, vol. 37, no. 1, pp. 1-13, Jan 2009.

DOI 10.18699/BGRS/SB-2020-211

Molecular markers based on SNPs in *FAD3* genes for determination of linolenic acid content in flax seed

Liubov Povkhova
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia;
Moscow Institute of Physics and Technology, Dolgoprudny, Russia
povhova.lv@phystech.edu

Parfe Kezimana
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia;
RUDN University, Moscow, Russia
k1par@mail.ru

Tatiana Rozhmina
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia;
Federal Research Center for Bast Fiber Crops, Torzhok, Russia
tatyana_rozhmina@mail.ru

Elena Pushkova
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
pushkova18@gmail.com

Roman Novakovskiy
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
0legovich46@mail.ru

George Krasnov
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
gskrasnov@mail.ru

Alexey Dmitriev
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
alex_245@mail.ru

Nataliya Melnikova
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
mnv-4529264@yandex.ru

***Abstract* — Flax is grown for seeds in many countries. Depending on the ratio of fatty acids, seeds are used in pharmaceutical, food, feed, or polymer production. It is known that genes of *FAD3* family responsible for the desaturation of linoleic (LIO) into linolenic (LIN) fatty acid and some alleles of these genes are associated with a certain composition of flax oil. However, the development of the marker-assisted selection of flax is complicated due to a lack of methods for the identification of these polymorphisms. In the present study, we used 84 flax cultivars/lines with diverse content of LIO and LIN to search for fatty acid associated SNPs and then developed HRM-based system for identification of these SNPs that can increase the efficiency of breeding by obtaining pure improved cultivars with the targeted LIO and LIN content.**

Keywords* — *Flax, fatty acids, FAD, LIN, LIO

Motivation and aim

Motivation

Flax (*Linum usitatissimum* L.) is grown for seeds in many countries. Linolenic (LIN), linoleic (LIO), and oleic (OLE) fatty acids are the main in flax oil, and depending on the ratio of these acids, seeds are used in pharmaceutical, food, feed, or polymer production. Flax genotypes differ in oil composition, and the breeding of cultivars with the particular content of LIN, LIO, and OLE is necessary. It is known that genes of *FAD* (fatty acid desaturase) family play a key role in the synthesis of fatty acids. *FAD3* genes are responsible for the desaturation of LIO into LIN, and some alleles of these genes are associated with a certain composition of flax oil [1]. However, the development of the marker-assisted selection of flax is complicated due to a lack of methods for the identification of DNA polymorphisms affecting the fatty acid content in flax seeds. Besides, cultivar heterogeneity could result in the presence of genotypes with different alleles of *FAD* genes and diverse oil composition in individual flax cultivar, and molecular markers for these genes can be useful to increase the cultivar purity.

Aim

The purpose of our study was to develop a test system for the identification of SNPs (single nucleotide polymorphisms) associated with the fatty acid composition of flax oil.

Methods

For the present study, we used 84 flax cultivars/lines with diverse content of LIO and LIN in oil. Seeds were obtained from the Institute for Flax (Torzhok, Russia). DNA was extracted from flax seedlings using a CTAB protocol. For identification of SNPs, which are associated with fatty acid composition, deep sequencing of *FAD3A* and *FAD3B* was performed on MiSeq (Illumina, USA) as described earlier [2]. VarScan [3] and freeBayes [4] variant callers were used for the identification of polymorphisms in the studied genes. Spearman's and Kendall's rank correlation coefficients between the content of LIO and LIN and polymorphisms in *FAD3A* and *FAD3B* genes were calculated. For identification of SNPs that were associated with oil composition, primers for the HRM (high-resolution melting) analysis were designed and PCR (polymerase chain reaction) conditions were optimized. HRM was performed on a Rotor-Gene Q real-time PCR cycler (Qiagen, Netherlands). Reaction mix contained 1x Taq Turbo buffer (Evrogen, Russia), 400 nM of each primer (Evrogen), 250 μM dNTPs (Evrogen), 1x EvaGreen (Biotium, USA), 1U of HS Taq DNA polymerase (Evrogen), and 40 ng of DNA. The following PCR program was used: 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 60 s. All reactions were performed in three technical replicates. Then, HRM analysis with the ramp from 65°C to 85°C rising by 0.1°C each step was conducted. For data analysis, Rotor-Gene HRM software was used.

Results

In the present study, deep sequencing on the Illumina platform was used to identify polymorphisms in *FAD3* genes in the representative set of 84 flax cultivars and lines. As a result, 300+300-bp reads with an average coverage of 300x enabled accurate assessment of SNPs in *FAD3A* and *FAD3B* genes taking into account cultivar and line heterogeneity. For

FAD3A and *FAD3B* genes, 91 and 62 polymorphisms were identified, respectively. Correlation analysis revealed polymorphisms in *FAD3* genes that were associated with the ratio of fatty acids in flax oil. All except one of the samples with low LIN and high LIO had both the tryptophan to stop codon substitution in *FAD3A* and the histidine to tyrosine substitution in *FAD3B*, while all cultivars and lines with high LIN and low LIO had not these mutations. Samples with only one of the two mutations had medium content of LIN and LIO. The rest low-LIN sample contained another mutation in *FAD3A* resulting in arginine to stop codon substitution that was inherent only to this cultivar from the analyzed collection. Based on data on these three polymorphisms of *FAD3* genes, controlling the LIN/LIO ratio, an HRM-based test system for the determination of linolenic acid content in flax seed was developed. Primers specific to conservative regions were designed taking into account all polymorphisms identified by us and other researchers. The HRM analysis with the pre-amplification method was used to get melt curves for the samples that were used for deep sequencing. Temperature shift in sample melt curves for three analyzed SNPs in homozygous and heterozygous states was distinguished clearly that enabled screening of flax for identification of associated with LIO and LIN content polymorphisms. Molecular markers for the fatty acid composition of flax oil can increase the efficiency of breeding by obtaining pure improved cultivars and are necessary for the development of the marker-assisted selection of flax cultivars with the targeted LIO and LIN content.

Acknowledgment

This work was funded by RFBR according to the research project 17-29-08036.

References

[1] Thambugala D., et al. (2013) Genetic variation of six desaturase genes in flax and their impact on fatty acid composition. TAG Theoretical and applied genetics Theoretische und angewandte Genetik. 126(10): 2627-2641.

[2] Melnikova N.V., et al. (2019) Sex-specific polymorphism of MET1 and ARR17 genes in Populus x sibirica. Biochimie. 162: 26-32.

[3] Koboldt D.C., et al. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22(3): 568-576.

[4] Garrison E., Marth G. (2012) Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907v2 [q-bio.GN].

DOI 10.18699/BGRS/SB-2020-212

Comparative genomic analysis of male and female poplars

Elena Pushkova
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
pushkova18@gmail.com

George Krasnov
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
gskrasnov@mail.ru

Roman Novakovskiy
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
0legovich46@mail.ru

Nadezhda Bolsheva
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
nlbolsheva@mail.ru

Nataliya Melnikova
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
mnv-4529264@yandex.ru

Alexey Dmitriev
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
alex_245@mail.ru

***Abstract* — In the genomic analysis of trees, species belonging to the *Populus* genera serve as model objects. In our work, we performed genome sequencing of male and female plants of *Populus* x *sibirica* that is widespread in the Moscow region. To achieve high-quality *de novo* assemblies, we used a combination of two platforms, namely Oxford Nanopore and Illumina for obtaining long reads and high-accuracy short reads, respectively. In total, using a MinION sequencer, we obtained 18 Gb with average N50 of 25 kb for each genotype, whereas, using a HiSeq 2500 instrument, from 36 to 52 million pair-end 125+125 reads were obtained for male and female plants. The Shasta assembler provided the best results for the assemblies: the N50 values were about 0.5 Mb and the integrity was over 95% according to BUSCO. The analysis of high-quality genomes of male and female *P.* x *sibirica* trees provides an opportunity to reveal the genes associated with poplar sex that will facilitate the specification of its sex-determination mechanism.**

Keywords* — *Populus, de novo genome assembly, Nanopore, Illumina, pure high-molecular-weight DNA

Motivation and Aim

Motivation

Poplar is a model object for genetic studies in trees [1]. In recent studies on the genetics of sex of the genus *Populus*, associated with sex polymorphisms were found and the areas of their localization in the genome were identified. However, many issues related to sex determination in poplar species have not yet been resolved, including the identification of the key genes and mechanisms that are involved in sex determination. *P.* × *sibirica* is one of the most common poplars in the Moscow region, but to date, there is no reference genome assembly for it. High-quality genome assemblies of male and female plants of *P.* x *sibirica* provide the opportunity to compare the results with previous studies of genetically different poplars and to suggest the mechanism of its sex determination.

Aim

This work aims to obtain high-quality assemblies of the genomes of male and female trees of *P.* × *sibirica* using a combination of Oxford Nanopore platform with long reads and Illumina with high-accuracy reads for further identification of genes that play a key role in determining the sex of poplar plants and to reveal the molecular mechanisms of their action.

Methods

Plant material

Leaves were collected from male and female trees of *P.* x *sibirica* grown in different areas of Moscow. Young half-open leaves were gathered in the period of poplar bloom for the proper sex identification.

DNA extraction

For DNA extraction and purification, *P.* x *sibirica* leaves with a mass of 0.2 g were triturated in the mortar with liquid nitrogen. Then, we added 4 ml of the Carlson lysis buffer (100 mM Tris-HCl pH 9.5 (VWR Life Science, USA); 2% CTAB (VWR Life Science); 1.4 M NaCl (Scharlab, Spain); 1% PEG 8000 (PanReac AppliChem, Germany); 20 mM EDTA (Promega, USA)) that was prewarmed to 65°C and supplemented with 12 µl of β-mercaptoethanol (BioRad, USA) and 0.04 g of PVP K30 (PanReac AppliChem). The homogenate was incubated in a "Gnom" thermostat (DNA Technology, Russia) at 65°C for 1 hour, with stirring every 10 min. Next, an equal volume of chloroform (Acros Organics, USA) was added to the homogenate, stirred on a Thermolyne Maxi Mix III Type 65800 shaker (Thermo Fisher Scientific, USA) at 800 rpm for 10 min, followed by centrifugation on a 5418R microcentrifuge (Eppendorf, Germany) using the following parameters: 10000 g, 10 min, 40°C. The aqueous phase was transferred to a clean tube with the addition of 0.2 volume of 5x CTAB buffer (5% CTAB, 350 mM EDTA) and incubated at 65°C for 10 min. After that, an equal volume of chloroform was introduced, stirred on a shaker for 10 min, and centrifuged with the following parameters: 10000 g, 10 min, 40°C. The aqueous phase was transferred to a clean tube with the addition of 0.1 volume of 5x CTAB buffer and incubated at 65°C for 10 min. After that, an equal volume of chloroform was introduced, stirred on a shaker for 10 min, and centrifuged with the following parameters: 10000 g, 10 min, 40°C. The aqueous phase was transferred to clean tubes with the addition of 2 volumes of the buffer for DNA precipitation (1% CTAB, 50 мМ Tris-HCl pH 8.0, 10 мМ EDTA) and incubated at room temperature for 1 hour. It was then centrifuged at 10000 g for 15 min at room temperature. Next, the supernatant was collected gently without disturbing the precipitate. The DNA pellet was air-dried for 5 min, dissolved in 2 ml of prewarmed to 60°C G-buffer from the Blood & Cell Culture DNA Mini Kit (Qiagen, USA), and incubated at 60°C for 10 min. To the DNA sample in the G-buffer, 4 µl of RNase A (100 mg/ml; 7000 units/ml; Qiagen) were added and incubated at 37°C for 30 min. To this, 25 µl of proteinase K (> 600 mAU/ml; Qiagen) were introduced and incubated at 50°C for 40 min. Further, DNA was purified according to the Blood & Cell Culture DNA Mini Kit (Qiagen) protocol. To the DNA

elution, 0.7 volume of isopropanol was added, stirred until DNA strands appeared and neatly wrapped them around a glass rod. Then the DNA was transferred to a tube containing a DNA dilution buffer (Evrogen, Russia) and incubated at 50°C for 60 min. The DNA quality and concentration were evaluated on a NanoDrop 2000C spectrophotometer (Thermo Fisher Scientific). The DNA concentration was also evaluated on a Qubit 2.0 fluorometer (Life Technologies). The assessment of DNA length and control of the absence of RNA were performed by electrophoresis in a 0.8 % agarose gel (Lonza, Switzerland).

DNA library preparation and sequencing of genomes on the Illumina platform

DNA was fragmented on a S220 ultrasonic homogenizer (Covaris, USA), and 1 μg of fragmented DNA was used to prepare the library using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) according to the manufacturer's protocol with the size selection of adaptor-ligated DNA of about 600 bp. The quality and concentration of DNA libraries were evaluated using a 2100 Bioanalyzer instrument (Agilent Technologies) and a Qubit 2.0 fluorometer (Life Technologies), respectively. The resulting DNA libraries were sequenced on a HiSeq 2500 instrument (Illumina) with a read length of 125+125 bp.

DNA library preparation and sequencing of genomes on the Oxford Nanopore platform

To remove short DNA fragments (up to 10 kb), the Short Read Eliminator Kit (Circulomics, USA) was used. Then, the DNA sample was diluted to a concentration of 60 ng/ml in a final volume of 50 μl of DNA dilution buffer (Evrogen) and further purified with AMPure XP beads (Beckman Coulter, USA) in a ratio of 1:0.7 (sample:beads).

Preparation of the libraries was performed using the SQK-LSK109 Ligation Sequencing Kit (Oxford Nanopore Technologies, UK) for 1D genomic DNA sequencing. Minor modifications were introduced to the recommended protocol for library preparation by increasing the incubation time to 20 min at 20°C at the step of combined recovery of DNA using the NEBNext Ultra II End Repair/dA-Tailing Module (New England Biolabs) and NEBNext FFPE DNA Repair Mix (New England Biolabs). Besides, the incubation time at the step of ligation was increased to 60 min. Sequencing was performed on MinION (Oxford Nanopore Technologies) with a FLO-MIN-106 R9.4 flow-cell (Oxford Nanopore Technologies).

Results

We developed and optimized the method of obtaining pure high-molecular-weight genomic DNA from young poplar leaves needed for Nanopore sequencing. According to this protocol, we extracted the DNA of the following quality: A260/280 ratio laying in the range of 1.77-1.80 and A260/230 being in the range of 1.7-2.3. The obtained proximity of concentration values measured with Qubit and Nanodrop served as an additional criterion of DNA purity. The high DNA length (more than 50 kb) and the absence of RNA were confirmed by electrophoresis. After the sequencing of prepared DNA libraries on a MinION instrument, we received 18 Gb with average N50 of 25 kb for each genotype, whereas using a HiSeq 2500 instrument, from 36 to 52 million of pair-end 125+125 reads for male and female plants. The initial genome assembly was performed for the Nanopore data using Shasta, wtdbg2, and Flye tools. Then, these assemblies were consequently polished by Racon (Nanopore data) and Pilon (Illumina data). The Shasta assembler provided the best assembly results: N50 parameter of the assemblies was about 0.5 Mb, while the integrity was over 95% according to BUSCO. The combination of long reads from Nanopore and short high-accuracy reads from Illumina allowed us to achieve a high-quality genome assembly of male and female plants of *P*. x *sibirica*. Besides, the protocol of pure high-molecular DNA extraction from poplar leaves was developed for further sequencing of material on the Nanopore platform. The analysis of high-quality genomes of male and female *Populus* x *sibirica* species provides one with an opportunity to reveal the genes associated with poplar sex that will facilitate the specification of its sex-determination mechanisms.

Acknowledgment

This work was funded by RFBR according to the research project 17-29-08036.

References

[1] Ellis B., Jansson S., Strauss S.H., et al. Why and How Populus Became a “Model Tree”. New York, NY: Springer New York, 2010.

DOI 10.18699/BGRS/SB-2020-213

Comparative analysis of flax (*Linum usitatissimum* L.) genomes and transcriptomes

Elena Pushkova
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
pushkova18@gmail.com

George Krasnov
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
gskrasnov@mail.ru

Roman Novakovskiy
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
0legovich46@mail.ru

Liubov Povkhova
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia;
Moscow Institute of Physics and Technology, Dolgoprudny, Russia
povhova.lv@phystech.edu

Artemy Beniaminov
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
abeniaminov@mail.ru

Nadezhda Bolsheva
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
nlbolsheva@mail.ru

Tatiana Rozhmina
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia;
Federal Research Center for Bast Fiber Crops, Torzhok, Russia
tatyana_rozhmina@mail.ru

Alexey Dmitriev
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
alex_245@mail.ru

Nataliya Melnikova
Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia
mnv-4529264@yandex.ru

***Abstract* — Flax (*Linum usitatissimum* L.) is an agriculturally important plant that has a wide range of applications in industry, and the direction of cultivar application is determined by its genetic characteristics. The present work aimed to obtain high-quality sequences of flax genomes and transcriptomes for genetically diverse cultivars and lines, which have breeding value and different direction of use. Six cultivars/lines (LM98, #3896, Diplomat, Atlant, Universal, Alizee) were selected for the present work. We have developed an optimal method for the extraction of pure high-molecular-weight DNA from flax plants and performed genome and transcriptome sequencing. From 6 to 10 Gb was obtained for each of the studied flax cultivars/lines on the Oxford Nanopore platform and 20-25 million reads on the Illumina platform. Transcriptome sequencing of different flax tissues was also performed. Genome assembly using Flye resulted in the N50 value from 200 kb to 1 Mb depending on the genotype. Genome annotation was then performed. The obtained genome assemblies are the basis for molecular genetic studies in flax and allow assessment of the differences in *L. usitatissimum* cultivars/lines at the genome-wide level.**

Keywords* — *flax, genome assembly, transcriptomes, Nanopore sequencing

Motivation and aim

Motivation

The data on the nucleotide sequences of whole genomes and transcriptomes are the basis for molecular genetic studies in plants, including the analysis of gene expression, polymorphism assessment, development of molecular markers of economically valuable traits, progress in genome selection, and the use of CRISPR/Cas genome editing technologies. The revolution in DNA sequencing technologies that took place in recent decades allowed us to take a study of plant genomes and transcriptomes to a new level and to determine the genome sequences for thousands of genotypes of various species, as well as to resequence genomes for different genotypes of the same species [1]. However, as the number of sequenced plant genomes increases, it becomes clear that there are differences between genotypes of a single species in both single nucleotide polymorphisms and copy number of genes, the presence of extensive insertions/deletions and genome-specific loci, some of which are associated with economically valuable traits [2-5]. Therefore, representative sets of samples should be analyzed to obtain comprehensive data concerning genomes and transcriptomes within the studied species. For the present study, such agriculturally important plant as flax (*Linum usitatissimum* L.) was chosen. Flax seeds are the richest source of omega-3 fatty acids, which reduce the risk of cancer and cardiovascular diseases, and lignans, which have antibacterial, antifungicide, antioxidant, and anti-carcinogenic activity, and also contain easily digestible proteins, dietary fiber, vitamins, and minerals. Flax fiber is used in the production of high-quality textiles, medical devices, and composite materials. Thus, flax has multiple using in different areas and the direction of cultivar application is determined by its genetic characteristics.

Aim

The aim of the present work was to obtain high-quality sequences of flax genomes and transcriptomes for genetically diverse cultivars and lines, which have breeding value and different direction of use.

Methods

When choosing flax genotypes for comparative analysis of genomes and transcriptomes, we took into account their genetic differences, resistance to the most harmful stresses, namely the *Fusarium oxysporum* fungus and adverse soil acidity, and other economically significant characteristics, such as yield, oil composition, fiber quality. As a result of joint research with the Institute for Flax (Torzhok), 6 cultivars/lines were selected (LM98, #3896, Diplomat, Atlant, Universal, Alizee) that differ in valuable traits and are of interest for use in creating improved varieties. To obtain a high-quality assembly of flax genomes, an approach was chosen using a combination of two sequencing platforms, one of which (Oxford Nanopore) allows for reads of large length but insufficient accuracy, and the other (Illumina) – for high-

precision but short reads. For sequencing on the Oxford Nanopore platform, an important task was the isolation of high-quality DNA. We tested numerous approaches, and as a result, developed an optimal method for the extraction of pure high-molecular-weight DNA from flax. The method was based on obtaining the maximum yield of high molecular weight but not very pure DNA with the DNA-EXTRAN-3 (Synthol) kit and the further purification, including precipitation with CTAB containing buffer and the using of ion-exchange columns from the Blood & Cell Culture DNA Mini Kit (Qiagen). HiSeq 2500 (Illumina) and MinION (Oxford Nanopore) were used for sequencing genomes of six selected flax cultivars/lines. NextSeq (Illumina) was used for sequencing the transcriptomes of the roots and shoots of 7-day-old seedlings, leaves and stems of 6-week-old plants, as well as flowers of LM98, #3896, Diplomat, Atlant, Universal, Alizee cultivars/lines (in total, 30 cDNA libraries). For genome assembly, Shasta, wtdbg2, and Flye were used. The polishing of assemblies was performed using Racon and Pilon, and funannotate was used for genome annotation.

Results

From 20 to 25 million reads with a length of 250+250 bp was obtained on the Illumina platform for each of the studied flax cultivar/line (LM98, #3896, Diplomat, Atlant, Universal, Alizee), which after trimming corresponded to 20-25-fold coverage of the flax genome, and from 6 to 10 Gb was obtained on the Oxford Nanopore platform, which corresponds to 15-25-fold genome coverage. From 6 to 16 million reads with a length of 86 bases were received as a result of transcriptome sequencing of the roots and shoots of 7-day-old seedlings, leaves, and stems of 6-week-old plants, as well as flowers of LM98, #3896, Diplomat, Atlant, Universal, Alizee cultivars/lines. Three different assemblers (Shasta, wtdbg2, and Flye) were compared to select the optimal strategy for flax genome assembly. The best results were shown by Flye. Then the polishing of assemblies was performed in two stages – using Nanopore reads (Racon) and using Illumina reads (Pilon). The N50 value ranged from 200 kb to 1 Mb, depending on the genotype, which is a good result. The completeness of the assemblies was more than 90% according to BUSCO metrics. The key point for the assembly quality was the length of the Oxford Nanopore reads – a lower output of reads but of a larger length allowed achieving a better quality of genome assembly. Using funannotate, genome annotation was performed. The obtained genome assemblies are the basis for molecular genetic studies in flax, allows one assessment of the differences in *L. usitatissimum* cultivars/lines at the genome-wide level, and lay the foundation for the development and introduction of marker-assisted and genomic selection of flax and its genome editing. Based on transcriptome sequencing data, gene expression was also evaluated in different tissues of six sequenced cultivars/lines, and genes with the most significant differences in expression between flax genotypes with different characteristics were revealed that is important for identifying associations between gene expression and phenotype.

Acknowledgment

This work was financially supported by the Russian Science Foundation, grant 16-16-00114.

References

[1] Torkamaneh, D., B. Boyle, and F. Belzile, Efficient genome-wide genotyping strategies and data integration in crop plants. Theor Appl Genet, 2018. 131(3): p. 499-511.

[2] Thudi, M., et al., Whole genome re-sequencing reveals genome-wide variations among parental lines of 16 mapping populations in chickpea (Cicer arietinum L.). BMC Plant Biol, 2016. 16 Suppl 1: p. 10.

[3] Bayer, P.E., et al., Assembly and comparison of two closely related Brassica napus genomes. Plant Biotechnol J, 2017. 15(12): p. 1602-1610.

[4] Jiao, Y., et al., Improved maize reference genome with single-molecule technologies. Nature, 2017. 546(7659): p. 524-527.

[5] Schatz, M.C., et al., Whole genome de novo assemblies of three divergent strains of rice, Oryza sativa, document novel gene space of aus and indica. Genome Biol, 2014. 15(11): p. 506.

DOI 10.18699/BGRS/SB-2020-214

Characteristics of interaction of miRNA with mRNA of C2H2, ERF and GRAS transcription factors of arabidopsis, rice and maize

Aizhan Kazievna Rakhmetullina
SRI of biology and biotechnology problems
Al-Farabi Kazakh National University
Almaty, Kazakhstan
zhanullina1994@gmail.com

Svetlana Kazbekovna Turasheva
SRI of biology and biotechnology problems
Al-Farabi Kazakh National University
Almaty, Kazakhstan
svetlana.turasheva@kaznu.kz

Anna Yurevna Pyrkova
SRI of biology and biotechnology problems
Al-Farabi Kazakh National University
Almaty, Kazakhstan
anna.pyrkova@kaznu.kz

***Abstract* — The miRNA binding sites with mRNA of genes encoding C2H2, ERF, GRAS transcription factors (TFs) were identified for Arabidopsis thaliana, Oryza sativa and Zea mays. The free energy (ΔG) of interaction of miRNA with mRNA target genes, the maximum of free energy (ΔGm), the ratio ΔG/ΔGm, and location of the potential binding sites were calculated using MirTarget program. In mRNA of C2H2, ERF, GRAS genes of all studied plants, miRNA binding sites were located in the protein-coding part (CDS) and 5'-untranslated region (5'UTR). The ath-miR5658-5p, ath-miR5021-5p, osa-miR2102-5p, osa-miR5075-3p had binding sites in mRNA of three studied families, with the value of ΔG/ΔGm from 91% to 98%. The miR171a-3p had binding sites in mRNA GRAS transcription factors family of all studied plants, with the value of ΔG/ΔGm equal 100%. The nucleotide sequences of ath-miR171a-3p, osa-miR171a-3p, and zma-miR171a-3p were similar, and their quantitative characteristics of interaction with mRNA of LOC_Os02g44360.1, GRMZM2G037792_P01, and AT2G45160.1 genes, were also similar. The obtained results indicate the dependence of expression TF of C2H2, ERF, GRAS families on miRNA.**

Keywords* — *miRNA, mRNA, gene, binding site, plant, transcription factor

Introduction

Most plant target miRNAs are transcription factors (TFs) that are involved in the regulation of plant growth and development. Studies have shown that miRNAs are able to regulate a range of biological processes in plants, such as: regulation of flower shoots and seeds, formation of nodules, development and responses to environmental stresses [1-3]. miRNAs and TFs provide combinatorial gene regulation that includes a variety of functions that can later be used to improve crop yields. Therefore, it was important to identify *A. thaliana, O. sativa, Z. mays* miRNAs that bind to mRNA of C2H2, ERF, GRAS transcription factors families.

Materials and methods

The nucleotide sequences of *A. thaliana, O. sativa, Z. mays* miRNAs were downloaded from miRBase (http://mirbase.org, release 22.1). The nucleotide sequences of genes of the C2H2, ERF, GRAS family were obtained from Plant Transcription Factor Database v4.0 (http://planttfdb.cbi.pku.edu.cn) and NCBI GenBank (http://www.ncbi.nlm.nih.gov). A search for the miRNA target genes was performed using the MirTarget program. This program defines the following binding characteristics: the start of the miRNA binding site on the mRNA; the location of the miRNA binding site (5'UTR, CDS, 3'UTR); the interaction free energy (ΔG, kJ/mole), the ΔG/ΔGm value (%) [4]. The MirTarget program found hydrogen bonds between adenine (A) and uracil (U), guanine (G) and cytosine (C), G and U, A and C. The numbers of hydrogen bonds in the G-C, A-U, G-U and A-C interactions were found to be 3, 2, 1 and 1, respectively.

Result

The search of interaction of 428 A. thaliana, 738 O. Sativa, 325 Z. mays miRNAs with mRNA of C2H2, ERF, GRAS transcription factors genes was completed. Using the MirTarget program, we found miRNA binding sites in the protein-coding region (CDS) and 5'-untranslated region (5'UTR). Table presents the quantitative characteristics of plant miRNAs interaction with mRNA of the C2H2, ERF, GRAS TF genes. The ath-miR5658-5p, ath-miR5021-5p, osa-miR2102-5p, osa-miR5075-3p had binding sites in mRNA of three studied families. The targets for ath-miR5658-5p were mRNA of AT1G34370.1 (C2H2), AT1G44830.1 (ERF), and AT5G59450.1 (GRAS) genes, with the value of ΔG/ΔGm equal 96% and 94%. The mRNA of AT1G34370.1 and AT5G59450.1 genes contain binding sites for ath-miR5658-5p located in the 5'UTR with the free energy of interaction (ΔG) equal -98 kJ/mole. The mRNA of AT1G44830.1 gene bound to ath-miR5658-5p in the CDS. The ath-miR5021-5p had binding sites in the mRNA of AT1G02030.1 gene of C2H2 family, AT1G22810.1 gene of ERF family and AT4G37650.1 gene of GRAS family with ΔG/ΔGm ratio of 93% or more. The osa-miR2102-5p and osa-miR5075-3p can interact with the free energy (ΔG) more than -109 kJ/mole, which indicates a strong binding to mRNA of LOC_Os09g38610.1, LOC_Os06g10470.1 (C2H2), LOC_Os04g32790.1, LOC_Os04g32790.1 (ERF), LOC_Os12g04380.1, LOC_Os06g40780.1 (GRAS) genes. All detected osa-miR2102-5p, osa-miR5075-3p binding sites were located only in the CDS. ath-miR414-5p, ath-miR867-5p, ath-miR171a-3p had binding sites in mRNA of *AT5G14010.1*, *AT1G53170.1*, *AT2G45160.1* genes with free energy from -98 kJ/mole to -113 kJ/mole and ΔG/ΔGm value equal 92% and 100%. It should be noted that nucleotide sequences of ath-miR171a-3p, osa-miR171a-3p were identical, and their quantitative characteristics in mRNA of *AT2G45160.1* and *LOC_Os02g44360.1* genes were the same: the location of the miRNA binding site (CDS), the free energy of interaction (ΔG) -113 kJ/mole, the ΔG/ΔGm value of 100%. This indicates a strong miRNA interaction with the mRNA. Length of zma-miR171a-3p was less by one nucleotide, and was different by free energy binding in mRNA of *GRMZM2G037792_P01* gene, while the ΔG/ΔGm value remained the same (100%). These genes belong to the GRAS

transcription factor family in *A. thaliana, O. sativa, Z. mays.* The zma-miR159e-5p bound in the CDS mRNA of *GRMZM2G060206_P01* ERF TF family gene, and *GRMZM2G146018_P01* GRAS TF family gene. The remaining zma-miR390a,b-5p, zma-miR169o-3p, zma-miR166n-5p, zma-miR160f-5p, zma-miR529-5p, zma-miR394a,b-5p had one target in the mRNA of *GRMZM2G114660_P01*, *GRMZM2G105224_P01*, *AC211702.2_FGP002* (C2H2), *GRMZM2G474326_P01*, *GRMZM2G060876_P01* (ERF), *GRMZM2G024973_P01* (GRAS) genes, respectively. Discovered binding sites were located in the CDS mRNA target genes, with the ΔG/ΔGm value equal or greater than 90%.

References

[1] A.K. Rakhmetullina, M. Régnier, A. T. Ivashchenko, "The characteristics of miRNA binding sites with mRNA of MYB plant transcription factors," International Journal of Biology and Chemistry, vol.12, pp. 60-68, 2019. DOI: https://doi.org/10.26577/ijbch-2019-1-i14

[2] A. Bari, A. Sagaidak, I. Pinskii, S. Orazova, A. Ivashchenko, "Binding of miR396 to mRNA of genes encoding growth-regulating transcription factors of plants," Russian Journal of Plant Physiology, vol. 61, pp. 807-810, 2014. DOI: 10.1134/S1021443714050033

[3] A. Bari, A. Orazova, A. Ivashchenko, "miR156- and miR171-binding sites in the protein-coding sequences of several plant genes," BioMed Res, vol. 2013, pp.307145, 2013. DOI: 10.1155/2013/307145

[4] A. T. Ivashchenko, A. Y. Pyrkova, R. Y. Niyazova, A. Alybayeva, K. Baskakov, "Prediction of miRNA binding sites in mRNA," Bioinformation, vol. 12, pp. 237-240, 2016.

TABLE - CHARACTERISTICS OF MIRNA BINDING SITES IN MRNA OF C2H2, ERF, GRAS TF GENES OF *A.THALIANA, O. SATIVA, Z. MAYS*

Gene	miRNA	Start of site, nt	Region	ΔG, kJ/mole	ΔG/ΔGm, %	Length, nt
C2H2 family						
AT1G34370.1	miR5658-5p	154	5'UTR	-98	96	21
AT1G02030.1	miR5021-5p	987	5'UTR	-91	93	20
AT5G14010.1	miR414-5p	449	CDS	-98	92	21
LOC_Os09g38610.1	miR2102-5p	229	CDS	-117	96	20
LOC_Os04g39520.1	miR5809-3p	66	CDS	-113	96	20
LOC_Os06g10470.1	miR5075-3p	158	CDS	-115	93	21
GRMZM2G114660_P01	miR390a,b-5p	484	CDS	-108	93	21
GRMZM2G105224_P01	miR169o-3p	1023	CDS	-102	92	20
AC211702.2_FGP002	miR166n-5p	9	CDS	-104	91	21
ERF family						
AT1G22810.1	miR5021-5p	22	5'UTR	-96	98	20
AT1G44830.1	miR5658-5p	133	CDS	-96	94	21
AT1G53170.1	miR867-5p	143	5'UTR	-93	92	21
LOC_Os02g54050.1	miR11343-3p	304	5'UTR	-100	96	21
LOC_Os04g32790.1	miR2102-5p	810	CDS	-113	93	20
LOC_Os04g32790.1	miR5075-3p	340	CDS	-113	91	21
GRMZM2G474326_P01	miR160f-5p	525	CDS	-108	91	21
GRMZM2G060876_P01	miR529-5p	760	CDS	-100	90	21
GRMZM2G060206_P01	miR159e-5p	583	CDS	-102	89	21
GRAS family						
AT2G45160.1	miR171a-3p	1304	CDS	-113	100	21
AT4G37650.1	miR5021-5p	374	CDS	-93	96	20
AT5G59450.1	miR5658-5p	119	5'UTR	-98	96	21
LOC_Os02g44360.1	miR171a-3p	1397	CDS	-113	100	21
LOC_Os12g04380.1	miR5075-3p	1311	CDS	-117	95	21
LOC_Os06g40780.1	miR2102-5p	662	CDS	-110	91	20
GRMZM2G037792_P01	miR171a-3p	1280	CDS	-106	100	20
GRMZM2G146018_P01	miR159e-5p	585	CDS	-106	93	21
GRMZM2G024973_P01	miR394a,b-5p	2039	CDS	-98	90	20

DOI: 10.6026/97320630012237

Conclusion

The presence of miRNA binding sites with strong interaction in mRNA of C2H2, ERF, GRAS TF genes of *A.thaliana, O. sativa, Z. mays* was shown. All binding sites of these miRNAs were located in the CDS and 5'UTR of mRNA target genes. The predicted miRNA binding sites with mRNA of target genes important for the regulatory mechanisms of miRNA in the process of plant growth and development.

Acknowledgements

The work was carried out with the financial support of the Ministry of Education and Science of the Republic of Kazakhstan within the framework of the grant №AP05132460.

DOI 10.18699/BGRS/SB-2020-215

Wheat and maize miRNAs are potential regulators of human genes expression

Aizhan Kazievna Rakhmetullina
SRI of biology and biotechnology problems, Al-Farabi Kazakh National University Almaty, Kazakhstan
zhanullina1994@gmail.com

Anatoliy Timofeevich Ivashchenko
SRI of biology and biotechnology problems, Al-Farabi Kazakh National University Almaty, Kazakhstan
a.iavashchenko@gmail.com

Anna Yurevna Pyrkova
SRI of biology and biotechnology problems, Al-Farabi Kazakh National University Almaty, Kazakhstan
anna.pyrkova@kaznu.kz

Abstract — **With food, a huge variety of biological material gets into the human digestive tract, which the body uses for life support. The variety of food material entering the gastrointestinal tract, especially at the molecular level, cannot be distinguished from endogenous metabolites and these exogenous compounds can significantly alter the body's metabolism. Such compounds include plant miRNAs, which are indistinguishable from endogenous human miRNAs in physicochemical properties. It is necessary to clarify the degree of influence of exogenous plant miRNAs on the expression of human genes, since it is not known in advance what consequences can occur when plant miRNAs enters the human body. A huge amount of research does not allow experiments with all human genes and all plant miRNAs, so we have studied the effect of wheat and maize miRNAs on human genes using computer methods. As a result of studying the binding of 125 tae-miRNAs and 325 zma-miRNAs to mRNAs of 17508 human genes it was revealed that 158 genes were targets for 52 tae-miRNAs and 51 genes for 11 zma-miRNAs. Binding sites in the mRNA of human genes were located in 5'UTR, CDS, 3'UTR.**

Keywords — miRNA, mRNA, binding site, gene regulation, plant, wheat, maize, human

Introduction

Recently, publications have appeared about the ingestion of exogenous miRNAs into the human body. Plant miRNAs that enter the human body along with food are able to penetrate the tissues and affect some metabolic pathways. These extracellular miRNAs are stable and involved in intercellular interactions [1]. Wheat and maize are the most actively cultivated crops worldwide. It is imperative to learn how plant miRNAs can affect the human body and how these miRNAs can be useful or harmful. The MirTarget program used by us with high efficiency determines the quantitative characteristics of the interaction of plant miRNAs with human mRNAs.

Materials and metods

Nucleotide sequences of the mRNAs of human genes were obtained from NCBI (http://www.ncbi.nlm.nih.gov). Nucleotide sequences of wheat and maize miRNAs were downloaded from miRBase v.22 (http://www.mirbase.org/). The miRNA binding sites in mRNA of 17508 human genes were predicted using the MirTarget program, which defines the free energy of interaction miRNA and mRNA (ΔG, kJ/mole), as well as the localization of binding sites and schemes of nucleotide interactions between miRNAs and mRNAs [2, 3]. The ratio of ΔG/ΔGm (%) was determined for each binding site, where ΔGm is equal to the free energy binding of miRNA with its full complementary nucleotide sequence.

Results and discussion

We found that among 325 miRNAs of *Z. mays* only 11 miRNAs were able to bind to mRNA of human genes. These miRNAs interacted with mRNA of 51 different genes. The zma-miRNA binding sites in the mRNA of human genes were located in 5'UTR, CDS, 3'UTR. The largest number of target genes had miR529-3p. The miR529-3p interacted with mRNA of *ATP6V0A4*, *BZRAP1*, *CHD2*, *CNGA2*, *GRM4*, *HCLS1*, *HIVEP2*, *IGFN1*, *LMO7*, *LRRC73*, *LRTOMT*, *LTB4R*, *MAP7*, *MLL*, *NUDC*, *PDE4B*, *POMC*, *SOX9*, *TCEB3*, *TMEM181* genes with the value of ΔG/ΔGm from 91% to 94%. The miRNA binding sites in mRNA of 13 genes were located in CDS, four sites in 5'UTR and 3'UTR. The miR162-3p and miR11969-5p had binding sites in the mRNA of four and six target genes, respectively. The binding sites of these miRNAs were located only in CDS and 3'UTR. The rest miRNAs (miR827-3p, miR529-5p, miR162-5p, miR1432-5p, miR11970-5p, and miR11969-3p) had only one target genes, with the value of ΔG/ΔGm more than 88%. In the group of 11 miRNAs, there were four miR-3p/miR-5p pairs that belong to the same pre-miRNA (miR529-3p/5p miR482-3p/5p, miR162-3p/5p miR11969-3p/5p). The characteristics of the binding sites for miR482-5p and miR482-3p were found for eight and nine mRNA target genes, respectively. Some of the above mentioned 17 genes involved in the development of diabetic retinopathy (*ROBO4*) [4], papillary thyroid carcinoma (*CDKN1C*) [5], myeloid leukemia (*FAM168A*) [6], hypogonadotropic hypogonadism (*FGF17*) [7], pancreatic and gastric cancers (*SLC44A4*) [8], breast cancer (*ATP2A3*) [9], lung cancer (*LARP1*) [10]. The binding sites of these miRNAs were located at 5'UTR, CDS and 3'UTR, and the free binding energy was a -96 kJ/mole and -98 kJ/mole.

As a result of the studies, binding sites were established between the wheat miRNAs and the mRNA of human genes involved in the development of various diseases. Binding sites were found for 52 miRNAs with mRNAs of 158 genes with a ΔG/ΔGm value of 88% or more. The binding sites of the studied miRNAs were located at 5'UTR (9%), CDS (63%) and 3'UTR (28%) mRNA. The highest number of binding sites with high affinity was detected in CDS and 3'UTR mRNA of target genes. Data analysis showed that genes *ADARB2, ADCY6, ALPK3, ANO4, BMS2, EDEM1, ERICH1, FNIP2, KIRREL3, NGB, PAQR6, TJP3,RICH1, RLBP1, SMARCC2, UBE2K, UNG* were targets for miR408-3p, with free energy from -108 kJ/mole to -113 kJ/mole and ΔG/ΔGm value from 91% to 95%. Therefore, miR408-3p can regulate a significant number of genes and further study is required. The miR10521-5p, miR1118-5p, miR1124-3p, miR1129-5p, miR1134-3p, miR1138-3p, miR164-5p, miR5086-5p, miR9652-5p, miR9655-3p, miR9661-5p, miR9773-3p, miR9780-3p, miR9781-3p bound to mRNA of four or more target genes. The functions of the identified target genes were diverse.

The figure shows the interaction schemes of tae-mir408-3p, zma-miR529-3p, and zma-miR482-3p with different regions of mRNA (5'UTR, CDS, 3'UTR) of human genes. The interactions between miRNA and mRNA show the role of noncanonical G-U and A-C pairs in increasing the free energy of interaction between miRNA and mRNA.

Fig. 1. - Schemes of the interaction of tae-miRNA and zma-miRNA with mRNA human genes

Note: The upper and lower nucleotide sequences of mRNA and miRNA, respectively. The bold type indicates the nucleotide of non-canonical pairs U-G; A-C.

It should be noted that miRNA binding sites with mRNAs of target genes had the same ΔG (kJ/mole) and ΔG/ΔGm (%) values for each miRNA: for tae-mir408-3p and mRNA of *ADARB2*, *ANO4*, *FNIP2*, *TJP3* genes ΔG equal 110 kJ/mole, ΔG/ΔGm equal 93%; for zma-miR529 and mRNA of *ATP6VOA4*, *CHD2*, *MAP7*, *MLL*, *PDE4B* genes ΔG equal -106 kJ/mole, ΔG/ΔGm equal 94%; for zma-miR482-3p and mRNA of *ATP2A3*, *CSF1*, *LARP1*, *LMBRD2*, *PTER*, *STK32A*, *TMCO5A*, *TTC25* genes ΔG equal -96 kJ/mole, ΔG/ΔGm equal 92%.

Conclusion

Thus, the consumption of wheat and maize in food not only provides the human body with energy, but also their miRNAs are involved in the regulation of gene expression, affecting human metabolism. The miRNAs significantly affect many genes, so plant miRNAs can be regulators of human gene expression.

Acknowledgment

The work was carried out with the financial support of the Ministry of Education and Science of the Republic of Kazakhstan within the framework of the grant №AP05132460.

References

[1] A. K. Rakhmetullina, A. T. Ivashchenko, “Rice miRNAs are potential regulators of human genes expression,” NEWS of the National academy of sciences of the Republic of Kazakhstan, vol. 5, pp. 24-31, 2019. DOI: 10.32014/2019.2519-1629.44

[2] A. Ivashchenko, O. Berillo, A. Pyrkova, R. Niyazova, S. Atambayeva, “MiR-3960 binding sites with mRNA of human genes,” Bioinformation, vol. 10, pp. 423-427, 2014. DOI: 10.6026/97320630010423

[3] A. Garg, U. Heinemann, “A novel form of RNA double helix based on G·U and C·A+ wobble base pairing,” RNA, vol. 24, pp.209-218, 2018. DOI: 10.1261/rna.064048.117

[4] Q. Gong, J. Xie, Y. Li, Y. Liu, G. Su, “Enhanced ROBO4 is mediated by up-regulation of HIF-1α/SP1 or reduction in miR-125b-5p/miR-146a-5p in diabetic retinopathy,” J Cell Mol Med., vol. 23, pp. 4723-4737, 2019. DOI: 10.1111/jcmm.14369

[5] R. M. de Silva, B. Pupin, T.T. Bhattacharjee, M. A. Vamondes Kulcsar, R. Chammas, R. de Azevedo Canevari, “ATR-FTIR spectroscopy and CDKN1C gene expression in the prediction of lymph nodes metastases in papillary thyroid carcinoma,” Spectrochim Acta A Mol Biomol Spectrosc., vol. 228, pp. 117693, 2020. DOI: 10.1016/j.saa.2019.117693

[6] X. Liu, H. Mai, H. Jiang, Z. Xing, D. Peng, Y. Kong, C. Zhu, Y. Chen, “FAM168A participates in the development of chronic myeloid leukemia via BCR-ABL1/AKT1/NFκB pathway,” BMC Cancer, vol. 19, pp. 679, 2019. DOI: 10.1186/s12885-019-5898-4

[7] M. Men, J. Wu, Y. Zhao, X. Xing, F. Jiang, R. Zheng, J. D. Li, “Genotypic and phenotypic spectra of FGFR1, FGF8, and FGF17 mutations in a Chinese cohort with idiopathic hypogonadotropic hypogonadism,” Fertil Steril., vol. 113, pp. 158-166, 2020. DOI:10.1016/j.fertnstert.2019.08.069

[8] A. L. Coveler, A. H. Ko, D. V. Catenacci, D. Von Hoff, C. Becerra, N.C. Whiting, J. Yang, B. Wolpin, “A phase 1 clinical trial of ASG-5ME, a novel drug-antibody conjugate targeting SLC44A4, in patients with advanced pancreatic and gastric cancers,” Invest New Drugs, vol. 34, pp. 319-28, 2016. DOI: 10.1007/s10637-016-0343-x

[9] E. Izquierdo-Torres, A. Hernández-Oliveras, I. Meneses-Morales, G. Rodríguez, G. Fuentes-García, A. Zarain-Herzberg, “Resveratrol up-regulates ATP2A3 gene expression in breast cancer cell lines through epigenetic mechanisms,” Int J Biochem Cell Biol., vol. 113, pp. 37-47, 2019. DOI: 10.1016/j.biocel.2019.05.020

[10] J. Han, G. Zhan, X. Ma, Q. Dong, H. Zhang, Y. Wang, J. Cui, “CircRNA circ-BANP-mediated miR-503/LARP1 signaling contributes to lung cancer progression,” Biochem Biophys Res Commun., vol. 503, pp. 2429-2435. DOI: 10.1016/j.bbrc.2018.06.172

DOI 10.18699/BGRS/SB-2020-216

Symmetry and asymmetry in bacterial and organellae genomes

Michael Sadovsky
ICM SB RAS, Krasnoyarsk, Russia
msad@icm.krasn.ru

Maria Senashova
ICM SB RAS, Krasnoyarsk, Russia
msen@icm.krasn.ru

Abstract — **Basic mirror symmetry is found, for all three types of genomes (bacterial, chloroplasts and mitochondria). The patterns are arranged by the formally identified fragments of a genome, of the same length converted into triplet frequency dictionaries. Prevalent pattern for all genome types looks like a three ray star with the fragments gathered into the clusters. Some deviations and violations of the typical pattern are observed and discussed.**

Keywords — nucleotide sequence, triplet, frequency, clustering, order, GC-content

Motivation and aim

Motivation

It is well-known that various fragments of a genome differ each other in their role in inherited information processing. We investigated the reciprocal problem of the impact of those functional difference onto the structurally determined difference of the formally identified fragments of a genome. We concentrated on single chromosome genomes (bacteria, chloroplasts and mitochondria).

Aim

The aim of this work is to compare the patterns revealed through the distribution of formally identified fragments of a genome into the triplet frequency space, in terms of the symmetry (or asymmetry) of that former.

Methods

We used two-step transformation of a genome to reveal the inner structuredness. Firstly, each genome was covered with a set of (overlapping) tiles. Each tile is a subsequence of the given length L; they located along a genome with the step T. We used $L = 603$ and $T = 11$; it is important that the length L is odd and divisible by 3, and the step T is indivisible by 3. Each tile is labeled with so called relative phase index. The index marks up the tiles in terms their functional role to be found within a genome, according to an annotation.

To determine the index, we traced the location of the central nucleotide at the tile. If the central nucleotide falls within a non-coding region, it takes index value j. When the central nucleotide falls within a coding region (regardless the inner exon-intron structure of a gene), then the index value is 0, if the remainder of the division of the distance (defined in nucleotides number) from the start position of a coding region to the central nucleotide of the tile position by 3. Correspondingly, the index value is 1 and 2, if the remainder is equal to 1 or 2. These phase indices are determined for the leading strand. Reciprocally, the indices take values j, -0, -1 or -2, for the genes located at the ladder strand. Obviously, the remainder and the distance to the central nucleotide of the tile must determined from the "end" of the annotated fragment, since any gene bank stores a singly strand, only.

As soon, as each tile is labeled, it must be converted into a triplet frequency dictionary $W_l(n)$, where l is the index value, and n is the location of the tile along a genome. Hence, a genome is transformed into an ensemble of the points in 63-dimensional metric space with Euclidean metrics; we eliminated a triplet from the analysis, since the sum of all 64 frequencies yields 1. It should be stressed that the frequency dictionary contains the triplets counted with the reading frame shift equal to 3; in other words, the count of the triplets corresponds to the translation reading frame shift; such design revealed the difference between the coding and non-coding regions, as well as the genes located in leading and ladder strands.

Practically, the triplet with the least standard deviation determined over the entire ensemble was excluded and the motivation is apparent: it contributes least of all into separability of the points. These were the triplets GGG for bacteria, GCG for chloroplasts and ACT for mitochondria. Also, for each tile GC-content has been calculated, as well as the genome-wide one.

Next, we studied the inner structure of the ensemble, through the analysis of the distribution of those points in the metric space provided by frequencies, and principal components [1]. We used freely distributed software *VidaExperts* [2, 3].

The inner structuredness has been studied for a family of a single chromosome genomes, among them are bacteria genomes (168), chloroplast genomes (789) and mitochondria genomes, both animal (463) and plant (45).

Results

Examination of structure produced by the distribution of the tiles converted into triplet frequency dictionary exhibits four-cluster pattern for greater part of genomes. This is a star-like pattern with three rays comprising the tiles of the specific phase indices. Remarkable fact is that outer parts (rays) of the star are comprised with the tiles laid at the coding regions; the inner cluster comprises the tiles belonging to non-coding regions of a genome. The interlocation of the tiles follows the symmetry composition: the first (core) ray comprises the tiles of 2 and -2 indices; two other rays incorporate the tiles of 0 and -1, and -0 and 1 phase indices. The central core of the pattern is occupied with j phase tiles.

The most common fact is that two sets of clusters corresponding to three indices and determined over the leading strand and the ladder one, reciprocally, mirror each other. Mirroring here means that the phase index runs the circuit over the cluster clockwise, for leading strand, and counter clockwise, for ladder strand.

This pattern is absolutely typical for chloroplast genomes: no exclusions have been found in the analysis of 798 entities. It should be noticed that chloroplast genomes indeed have some extra cluster; that latter comprises the tiles covering the area of very densely located tRNA genes in a genome. These

tiles have extremely high local GC-content, thus gathering into an individual cluster. The profiles of GC-content along chloroplast genomes could be seen in [4].

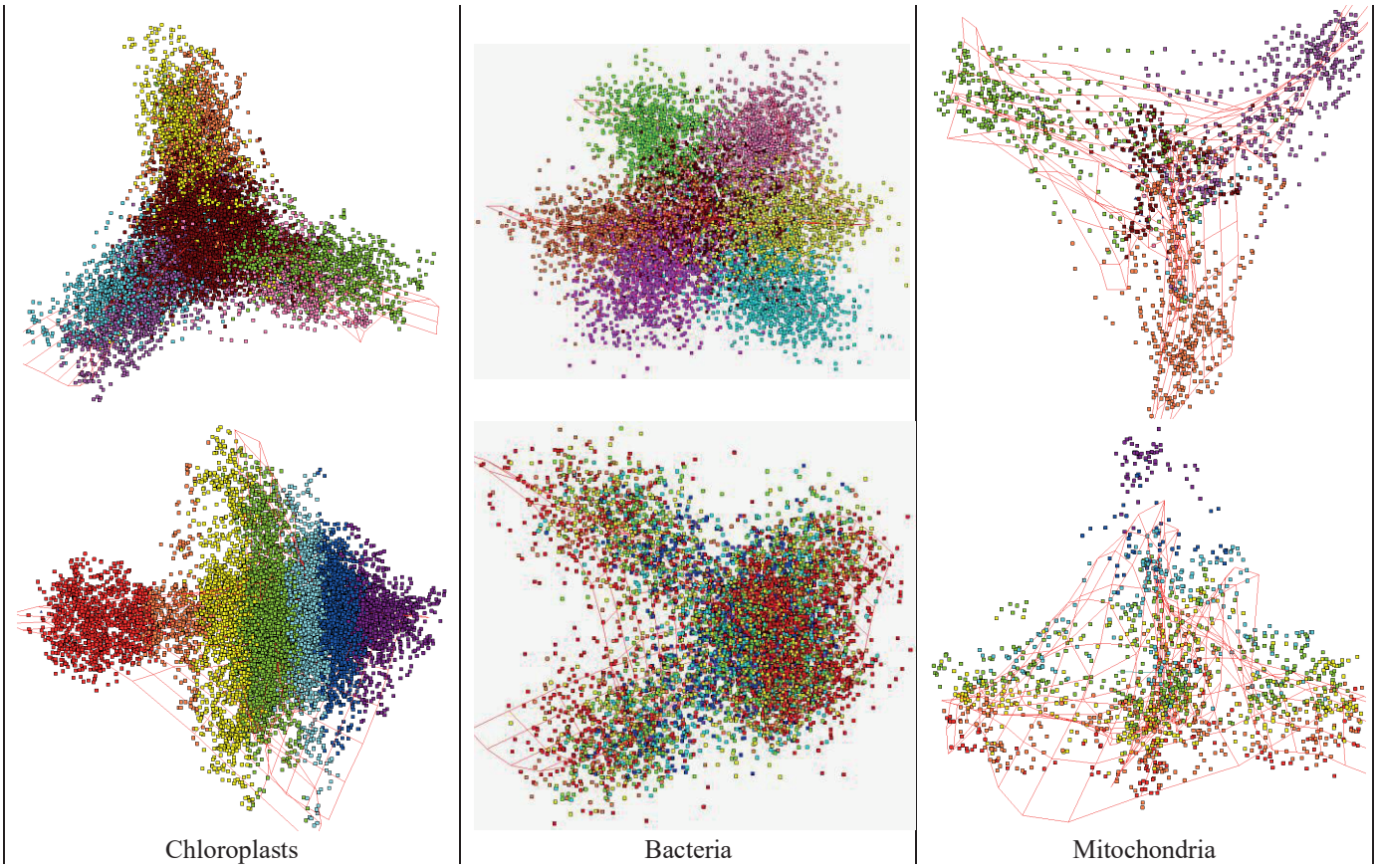


Fig. 1. Typical patterns observed for three types of genomes, see explanations in text.

Bacterial genomes exhibit diverse variety of patterns. Meanwhile, it should be said that three-ray pattern is typical for GC-reach and GC-poor bacterial genomes. Middle level GC-content bacterial genomes may have various patterns, among them is that one shown in Fig.1. Greater majority of genomes shows the three-ray star-like pattern, with the tiles falling into non-coding regions located in the center of the pattern, and the rays comprising the tiles indexed as described above.

The most unexpected and significantly less structured patterns exhibit mitochondrial genomes. To begin with, there is tremendous difference between animal and plant mitochondrial genomes, in terms of the clustering.

Fig.1. shows the examples of the patterns to be observed for the genomes of various origin. Upper row of figures presents the clustering observed in the projection of the second and the third principal components. Coloring in these pictures corresponds to the relative index phases: j is brown, 0 and -0 are blue and pink, 1 and -1 are green and violet, 2 and -2 are orange and yellow. For any genome with mirror symmetry the phases 2 and -2 always form a cluster; for other patterns there might be some other combinations.

Lower row in Fig.1 shows GC-content distribution, over the pattern. That latter is shown in the first and the second principal component projection. Coloring represents the local values of GC-content of each tile, changing in rainbow order from red (the greatest GC-content) to violet (the lowest GC-content).

References

[1] K. Fukunaga, Introduction to Statistical Pattern Recognition (2nd Edition), (1990) Academic Press, 593 p., DOI: 10.1016/B978-0-08-047865-4.50017-X.

[2] Gorban, A., Popova, T., & Zinovyev, A. (2005). Codon usage trajectories and 7-cluster structure of 143 complete bacterial genomic sequences. Physica A: Statistical Mechanics and its Applications, 353, 365-387.

[3] Gorban, A. N., & Zinovyev, A. Y. (2008). Elastic maps and nets for approximating principal manifolds and their application to microarray data visualization. In Principal manifolds for data visualization and dimension reduction (pp. 96-130). Springer, Berlin, Heidelberg.

[4] Sadovsky M.G., Putintseva Yu. A., Senashova M. Yu. (2018) Eight Clusters, Synchrony of Evolution and Unique Symmetry in Chloroplast Genomes: The Offering from Triplets // In: Chloroplasts and Cytoplasm: Structure and Functions (ser. Cell Biology Research Progress), Eds. Carl Dejesus and Lourdes Trask, Nova Science Publishers, Inc., N.-Y., pp.25–95.

DOI 10.18699/BGRS/SB-2020-217

Keeping the gate closed: WOX5 supports the balance between the proximal and distal root meristems via auxin biosynthesis in *Arabidopsis thaliana* L.

Maria S. Savina
ICG SB RAS, Novosibirsk, Russia
savinams1991@gmail.com

Taras Pasternak
Institute of Biology II/Molecular Plant Physiology University of Freiburg, Freiburg, Germany
taras.p.pasternak@gmail.com

Victoria V. Mironova
ICG SB RAS, Novosibirsk, Russia
Novosibirsk State University
kviki@bionet.nsc.ru

Nadezda Omelyanchuk
ICG SB RAS, Novosibirsk, Russia
nadya@bionet.nsc.ru

Viktoriya V. Lavrekha
ICG SB RAS, Novosibirsk, Russia
vvl@bionet.nsc.ru

Abstract — The functioning of the root stem cell niche depends on the auxin maximum formed and maintained by auxin transport and local biosynthesis. The auxin maximum is located in the quiescent center (QC), which separates the mitotically active proximal and distal root meristems. Transcription factor WUSCHEL HOMEOBOX (WOX5) expressed in QC controls stem cell maintenance. In WOX5-overexpression line 35S::WOX5-GR we recorded increase in auxin biosynthesis through TAA1 upregulating and changes in PIN-mediated auxin transport. Altogether it leads to auxin redistribution between proximal and distal root meristems. We show with mathematical modeling that increasing TAA1 expression is sufficient to trigger auxin redistribution. Numerical simulation results reproduce wild type and WOX5 overexpressing columella phenotypes. Key role of auxin biosynthesis was support with L-kynurenine (the competitive inhibitor of TAA1 activity) treatment for 48h, which partially restored the wild type phenotype in 35S::WOX5-GR line. Our model predicts formation of "quiescent domain" by actively dividing columella stem cells and columella stem cell daughters that was proved by 3D image analysis of root tips with traced cell cycle events.

Keywords — RAM, auxin, WOX5, mitotic activity, mathematical model, image analysis, iRoCS toolbox, EdU

INTRODUCTION

Root apical meristem contains a stem cell niche where a quiescent center (QC) promotes maintenance of the adjacent stem cells (initials) [1]. The distal part of the root stem cell niche consists of the columella stem cells (CSCs) producing daughters (CSCDs) further differentiating into columella (central root cap) cells (DCCs). WOX5 (WUSCHEL-RELATED HOMEOBOX 5) transcription factor is the QC signal maintaining the columella stem cells as a part of the distal meristem [2]. Auxin concentration maximum in the QC is also required for maintenance of the adjacent stem cells, including CSCs, and for differentiation of CSC daughters [3]. Tian and co-authors proposed and provide some evidence that WOX5 modulates auxin biosynthesis in the QC [4]. Recent work shows that WOX5 acts mainly in the QC and WOX5 mobility was not required to inhibit CSC differentiation [5]. However, mechanisms underlying this modulation and impact of this regulation on stem cell functions are not fully understood. Here we demonstrate that WOX5 activates the first step of auxin biosynthesis and this is very important for functioning of root stem cells, because disturbance in this regulation (e.g. in the case of WOX5 overexpression) violates root stem cell renewing and differentiation.

METHODS

We create 1D hybrid computation model of auxin distribution in columella cells. In the model cell growth, active and passive auxin transport between cells, and substances degradation are described by continuous functions, but transitions between the cell cycle states are discrete events, which are depend on auxin concentration in the cell.

Recently, iRoCS toolbox was developed [6] for annotation of the root tip organization in three dimensions. Together with the refined experimental procedure for detection of cell cycle progression [7], it provides an unprecedented potential to study the mechanism of cell cycle regulation in an entire organ. Here, using these techniques, we analysed the distributions of the key cell cycle events – DNA replication and mitosis in the root tips of *35S::WOX5-GR* line.

RESULTS

We observe in *35S::WOX5-GR* overexpression transgene upon DEX treatment that columella stem cell daughters do not undergo normal differentiation. We notice that during 48 hours of DEX treatment almost all rows of differentiated columella cells are detached but small cells in the enlarging of upper columella part do not have starch granules and look similar to the wild type CSCs. This means that CSC descendants occurred just after DEX activation of WOX5 expression do not undergo differentiation, but activate mitotic cell cycle progression.

We show that WOX5 modulates TAA1-mediated auxin synthesis and PIN-mediated auxin transport in root apical meristem. We hypothesize that modulation TAA1-mediated auxin synthesis results in observed changes in columella phenotype of WOX5 overexpression line.

We developed 1D computational model of auxin distribution in columella with growth and division of cells. Using the mathematical modeling approach we showed that WOX5-dependent upregulation of TAA1-mediated auxin synthesis are sufficient to reproduce all other changes observes in the *35S::WOX5-GR* plant roots. Incubation of L-kynurenine (the competitive inhibitor of TAA1 activity) provides partially recovery the wild type columella phenotype of WOX5 overexpression line. So, we can conclude that the main role of WOX5 gene for auxin distribution is modulation of TAA1-mediated auxin synthesis.

Moreover, numerical calculations predict that active cell divisions occur only in the lower part of columella while the

cells in the upper part of columella remain quiescent in *35S::WOX5-GR* transgenic line. 3D image analysis of cell cycle events distribution in the root tip confirms the model prediction.

ACKNOWLEDGMENT

The work was supported by the Russian Foundation for Basic Research (19-54-12013).

REFERENCES

[1] Jiang K, Feldman LJ (2005) Regulation of root apical meristem development. Annu Rev Cell Dev Biol 21:485–509.

[2] Zhou, J. J., & Luo, J. (2018). The PIN-FORMED auxin efflux carriers in plants. International journal of molecular sciences, 19(9), 2759.

[3] Pi, L., Aichinger, E., van der Graaff, E., Llavata-Peris, C. I., Weijers, D., Hennig, L., ... & Laux, T. (2015). Organizer-derived WOX5 signal maintains root columella stem cells through chromatin-mediated repression of CDF4 expression. Developmental cell, 33(5), 576-588.

[4] Tian, H., Wabnik, K., Niu, T., Li, H., Yu, Q., Pollmann, S., ... & Friml, J. (2014). WOX5–IAA17 feedback circuit-mediated cellular auxin response is crucial for the patterning of root stem cell niches in Arabidopsis. Molecular Plant, 7(2), 277-289.

[5] Berckmans, B., Kirschner, G., Gerlitz, N., Stadler, R., & Simon, R. (2020). CLE40 signalling regulates the fate of root stem cells in Arabidopsis. Plant Physiology.

[6] Schmidt, T., Pasternak, T., Liu, K., Blein, T., Aubry-Hivet, D., Dovzhenko, A., ... & Ronneberger, O. (2014). The iRoCS T oolbox–3 D analysis of the plant root apical meristem at cellular resolution. The Plant Journal, 77(5), 806-814.

[7] Pasternak, T., Tietz, O., Rapp, K., Begheldo, M., Nitschke, R., Ruperti, B., & Palme, K. (2015). Protocol: an improved and universal procedure for whole-mount immunolocalization in plants. Plant Methods, 11(1), 50.

DOI 10.18699/BGRS/SB-2020-218

Molecular genetic analysis of alloplasmic recombinant lines (*Triticum dicoccum*) – *Triticum aestivum*

Andrey Borisovich Shcherban
Laboratory of Molecular Genetics and Cytogenetics of Plants, Kurchatov Genomics Center, ICG SB RAS
Novosibirsk, Russia
atos@bionet.nsc.ru

Roman Nikolaevich Perfil'ev
Novosibirsk State Agrarian University
Novosibirsk, Russia
pervf.1999@gmail.com

Elena Artemovna Salina
Laboratory of Molecular Genetics and Cytogenetics of Plants, Kurchatov Genomics Center, ICG SB RAS
Novosibirsk, Russia
sunday01@mail.ru

***Abstract* — Molecular markers were used to analyze the mitochondrial genome of the alloplasmic lines of wheat containing the nuclear genome of the hexaploid wheat *T. aestivum* against the background of the cytoplasm of tetraploid wheat *T.dicoccum*. Eight lines showed patterns of PCR and CAPS markers corresponding to the species *T. aestivum*, indicating the substitution of the mitochondrial genome of *T. dicoccum* during backcrossing with hexaploid wheat. In the D-N-05 line, the rps19 and orf256 gene markers corresponded to the parental species *T.dicoccum*. This line is promising in terms of studying the mechanisms of fertility restoration in hybrids with CMS. Also, using molecular marker, the nuclear gene *Dreb-1*, which is a regulator of drought tolerance, was analyzed and it was shown that line D-41-05 had an introgression of this gene from the B genome of *T. dicoccum*, which may lead to the previously established increased drought tolerance of this line.**

Keywords* — *alloplasmic lines, common wheat, mitochondrial genome, drought resistance, gene

Introduction

In connection with the task of increasing the genetic diversity of the main cereal crop, common wheat, *T. aestivum* L. (2n = 42; BBAADD), alloplasmic lines, in which the nuclear genome of this species combines with the cytoplasm of an alien species as a result of remote hybridization and subsequent backcrossing with common wheat, are of great interest. Despite backcrossing, the nuclear genome of such lines may contain multiple introgressions of the maternal genome resulting from recombination. This may be important for breeding, in particular, for the selection of forms with high resistance to stress factors, both biotic and abiotic [1]. Nuclear cytoplasmic interactions, in particular, the influence of the mitochondrial genome, also make a great impact to the phenotypic variability, fertility, and viability of hybrids [2]. The modes of variability of both the nuclear and cytoplasmic genomes in the process of remote hybridization are still poorly understood.

Earlier, fertile alloplasmic wheat lines were obtained at the Institute of Plant Biology and Biotechnology of Kazakhstan. The results of preliminary experiments showed the resistance of some lines to increased salt concentration, as well as to water deficiency [3]. This work is aimed at analyzing the peculiarities in the organization of the mitochondrial genome of these lines, as well as the search for genes associated with various manifestations of drought tolerance, and the study of their organization and expression.

Materials and Methods

As a material we used 9 fertile alloplasmic wheat lines obtained from crossing (♀) *T. dicoccum* Schrank (2n = 28; BBAA) x (♂) *T. aestivum* L. cv. Mironovskaya 808. Total DNA was isolated from 7-day old seedlings according to standard technique with sodium bisulfite.

Mitochondrial genome analysis

PCR markers of the following genes were taken for the analysis of mtDNA of alloplasmic lines: 1) *orf256* - chimeric reading frame near the cytochrome oxidase gene (associated with cytoplasmic male sterility (CMS) [4]; 2) *nad6*-mitochondrial gene encoding NADH-ubiquinone oxidoreductase (subunit 6); 3) *rps19-p*-pseudogene encoding a ribosomal protein [5]. PCR products were separated on a 2% agarose gel with the addition of ethidium bromide. The gel was photographed using GelDoc XR (BioRad, England). The *orf256* PCR product was excised from the gel, purified using a kit (QIAGEN, Germany) and sequenced using a bigdye terminator v3.1 cycle sequencing kit (Applied Biosystems, USA). Sequencing products were analyzed at the Collective Use Center "Genomika" of the SB RAS. Additionally, in the case of *orf256*, a CAPS marker was used, namely: the PCR product was digested with *Taq* I restriction endonuclease (Sibenzyme, Novosibirsk) followed by electrophoresis on a 2% agarose gel.

Analysis of the drought resistance regulator gene Dreb-1

In this work, the marker of the *Dreb-1* gene described in [6] was used. PCR products were analyzed on a 2% agarose gel. To confirm the single-nucleotide substitution specific for the *T. dicoccum Dreb-B1* gene, PCR products were digested with restriction endonuclease *Bst*F5I (Sibenzyme) followed by electrophoresis in agarose gel. A 635 bp fragment specific for *T. dicoccum* was excised from the gel and sequenced as described above.

Results and Discussion

Analysis of the mitochondrial genome of 9 alloplasmic wheat lines using molecular markers for the 3 above genes showed that 8 lines predominantly inherit the type of male parent mtDNA. These lines showed the patterns of PCR products and a CAPS marker identical to those of *T. aestivum* (Mironovskaya 808). Consequently, the restoration of fertility of alloplasmic recombinant lines (*T. dicoccum*) –*T. aestivum* correlates with the substitution of DNA sequences targeting *T. dicoccum* mtDNA. An exception is the D-N-05 line, in which the patterns of PCR markers for the *rps19* and *orf256* genes corresponded to the parental species *T.dicoccum*. Interestingly, the *orf256* sequence in this line has 100% homology with the analogous sequence of the CMS line of *T. timopheevii* (♀) x *T. aestivum* (X56186), however, unlike the latter, the D-N-05 line is fertile. In the future, we plan to identify genes of fertility restoration within the nuclear genome of the D-N-05 line.

As previously shown, the greatest drought tolerance is characteristic of the lines D-d-05b, D-b-05, D-41-05 [3]. The

Dreb (*Dehydration responsive element binding*) family of genes are related to the very important regulatory genes that affect the tolerance to different abiotic stress factors including drought [7]. Using markers for one of the genes of this family, *Dreb-1*, we showed that in the nuclear genome of the drought-tolerant line D-41-05 there is a copy of the *Dreb-B1* gene (Fig. 1), which has a single nucleotide substitution characteristic of the sequence of *Dreb-B1 T. durum* [6], the tetraploid wheat species closest to *T. dicoccum*. Thus, it can be assumed that the increased drought tolerance of the D-41-05 line is associated with introgression of the *T. dicoccum* gene *Dreb-B1* in the genome of common wheat. In the future, we plan to analyze the expression of *Dreb-B1* under the influence of drought in alloplasmic lines differing by drought resistance.

ACKNOWLEDGMENT

The study was supported by the Kurchatov's Genomic Center (grant No. 075-15-2019-1662).

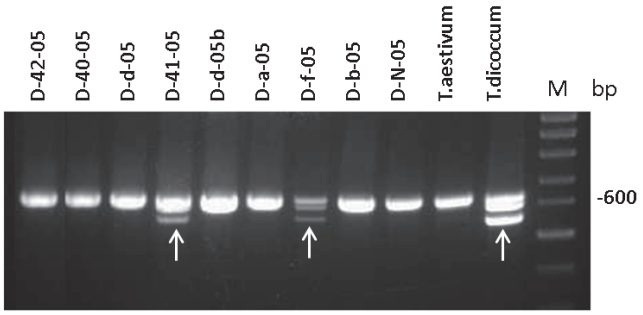


Fig. 1 PCR with a combination of primers to identify the *Dreb-1* gene. The arrows indicate the common fragment for lines D-41-05, D-f-05, and *T. dicoccum*, corresponding to the *Dreb-B1* gene. The upper major fragment corresponds to the *Dreb-A1* and *Dreb-D1* genes.

REFERENCES

[1] V.P. Kholodova, T.S. Bormotova, O.G. Semenov, G. A. Dmitrieva, V. Kuznetsov, "Physiological mechanisms of adaptation of alloplasmic wheat hybrids to soil drought" Russ. J Plant Physiol., vol.54, pp. 480–486, 2007.

[2] K. L. Liberatore, S. Dukowic-schulze, M. E. Miller, C. Chen, S.F. Kianian, "The role of mitochondria in plant development and stress tolerance" Free Radic. Biol. Med., vol. 100, pp. 238–256, 2016.

[3] N. V. Terletskaya, N. A. Khailenko, A. B. Iskakova, "Features of the reaction of alloplasmatic wheat lines seedlings on the effects of osmotic and salt stress" Vestnik Samarskogo Gosudarstvennogo Universiteta. Estestvenno-Nauchnaya Seriya, Issue 2(83), pp. 244–249, 2011.

[4] A.M. El-Shehawi, A.I. Fahmi, S.M. Sayed, M.M. Elseehy, "Genetic Fingerprinting of Wheat and Its Progenitors by Mitochondrial Gene *orf256*" Biomolecules, vol. 2, pp. 228-239, 2012.

[5] A.K. Noyszewski, "Mitochondrial sequence diversity among alloplasmic and euplasmic triticum species" Doctoral dissertation, North Dakota State University, Fargo, 2013.

[6] B. Wei, R. Jing, C. Wang, J. Chen, X. Mao, X. Chang, J. Jia, "Dreb1 genes in wheat (*Triticum aestivum* L.): development of functional markers and gene mapping based on SNPs" Mol Breeding, vol. 23, pp. 13–22, 2009.

[7] S. Lucas, E. Durmaz, B.A. Akpınar, "The drought response displayed by a DRE-binding protein from *Triticum dicoccoides*" Plant Physiol. and Biochem., 49: 346e351, 2011.

DOI 10.18699/BGRS/SB-2020-219

Transcriptomic mechanisms of *Solanum tuberosum* defensive response to golden potato nematode infestation

Alexey Kochetov
ICG SB RAS, Novosibirsk, Russia
ak@bionet.nsc.ru

Egorova Anastasiya
ICG SB RAS, Novosibirsk, Russia
egorova@bionet.nsc.ru

Anastasiya Glagoleva
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
glagoleva@bionet.nsc.ru

Kseniya Strygina
VIR, st. Petersburg, Russia
pushpandjali@gmail.com

Dmitry Afonnikov
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
ada@bionet.nsc.ru

Nickolay Shmakov
ICG SB RAS, Novosibirsk, Russia
shmakov@bionet.nsc.ru

Elena Khlestkina
VIR, st. Petersburg, Russia
director@vir.nw.ru

Sophia Gerasimova
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
gerasimova@bionet.nsc.ru

***Abstract* — Golden potato cyst nematode (GPCN) *Globodera rostochiensis* is one of the major potato pathogens, and has a sophisticated mechanism of interaction with roots of the host plants. Resistance of cultivated *S. tuberosum* forms to GPCN infestation is commonly based on specific R genes introgressed from natural populations of related varieties. In this research, RNA-seq is implemented in order to investigate R genes composition of two *S. tuberosum* varieties with contrasting resistance to GPCN, and transcriptomic response to infestation is studied.**

Keywords* — *transcriptomics, RNA-seq, potato, nematode, defensive response

Motivation and Aim

Motivation

S. tuberosum is a major crop cultivated worldwide and providing a substantial amount of calories in both human food and cattle feed. GPCN is a quarantine pest found responsible for yield losses that can range from 19% to 90%, depending on potato cultivar. Protection of plants against potato cyst nematodes commonly involves the usage of nematicides or trap crops. However, many chemical nematicides are either toxic or have low efficiency.

Alternatively, introduction of R genes associated with nematode resistance is used to protect crops. Main source of such genes are natural populations of *Solanum* species. For example, R genes conferring strong resistance to the pathotype Ro1 of GPCN were introgressed from South America originated species: the H1 gene from the *Solanum tuberosum* subsp. *andigenum* and Gro1–4 gene from the Bolivian species *S. spegazzinii*.

In Russia, GPCN occurs locally in some regions of the European part, southern Siberia, and the Far East. Coupled with importance and wide usage of potato in agriculture in Russian Federation, this facilitates the search for new genes associated with GPCN resistance of *S. tuberosum*, which makes creation of new resistant varieties possible.

Aim

This research aims to investigate transcriptomic response of two *S. tuberosum* cultivars to GPCN infestation. The cultivars used in this study are contrasting in resistance to GPCN. Resistance to nematode infestation is mediated through HR-response and programmed cell death that is commonly connected to NBS-LRR type receptors [1] Thus, one of the main objectives of such transcriptome investigation was to identify R genes, and more specifically, NBS-LRR genes, associated with GPCN resistance specific to the resistant line for further investigation and introgression in potato through conventional breeding or other biotechnological approaches. At the same time, sets of genes differentially expressed between two cultivars at different time stages and after inoculation with GPCN cysts and control treatment with water were identified, and functional annotation of DEGs was performed in order to study general transcriptomic response of *S. tuberosum* to nematode infestation.

Methods

In order to study specifics of plant-nematode interaction and defense mechanisms, two accessions of *S. tuberosum* – k-11291 (collected in Peru) and k-9836 (from Bolivia) – were selected. Accession k-9836 is susceptible to GPCN, while k-11291 is highly resistant, but contains no DNA markers of Gro1-4 and H1, commonly associated with resistance to the nematode. RNA-seq was performed in order to study transcriptome of the two plant cultivars. Total RNA was collected collected before inoculation (0 hours), 24 and 72 hours after inoculation with *G. rostochiensis* (pathotype Ro1). For each genotype, three infected and three control (water-inoculated) plants were taken, which gives us a total of 30 mRNA samples. Illumina NextSeq 550 platform was used for sequencing of 2x150 reads.

After libraries filtering, short reads were aligned to the *S. tuberosum* reference genome using STAR maps. Differential expression search was performed with EdgeR package, and Gene Ontology terms enrichment analysis was carried out.

Additionally, *de novo* transcriptome assembly was performed using Trinity assembler, and obtained transcriptomes were searched for R genes through NBS and LRR domain prediction. Assembled transcriptome was aligned to the reference genome with aid of BLAT aligner in order to identify novel transcript. hmmscan utility from HMMER package was implemented for domain structure prediction of transcripts, and NLR-parser software was used to specifically identify transcripts containing NBS and LRR domains associated with plant immunity and defense responses.

For a set of specific genes of interest, qRT-PCR expression quantification was performed in order to validate predicted differential expression.

Results

Functional annotation of DEGs revealed a strong difference in transcriptomes dynamics in the roots of resistant and susceptible *S. tuberosum* varieties in response to nematode infestation. While susceptible variety's response to GPCN cysts inoculation was similar to the response to simple water treatment performed as a control, and included activation of genes associated with cell cycle and mitosis, resistant variety's response to GPCN was quite different and included up-regulation of genes associated with stress response, and activation of peptidase inhibitor- and glycosylase-encoding genes.

De novo transcriptome assembly revealed a number of transcripts with no annotation in reference transcriptome that are actively expressing in the resistant line. Namely, a contig was identified that has no homology to the annotated genome sequence of *S. tuberosum* nor has any presence in the transcriptome of the susceptible potato variety, and at the same time has NBS and LRR domains. This contig is a promising molecular marker of nematode resistance and can further be implemented in engineering of new agriculturally important varieties of potato.

Acknowledgment

This research is supported by RSF grant 16-16-04073. Plants were cultivated in the IC&G greenhouse facility (supported by budget project 0324–2016-0001)

References

[1] Kochetov AV, Glagoleva AY, Strygina KV, Khlestkina EK, Gerasimova SV, Ibragimova SM, Shatskaya NV, Vasilyev GV, Afonnikov DA, Shmakov NA, Antonova OY, Gavrilenko TA, Alpatyeva NV, Khiutti A, Afanasenko OS. Differential expression of NBS-LRR-encoding genes in the root transcriptomes of two *Solanum phureja* genotypes with contrasting resistance to *Globodera rostochiensis*. BMC Plant Biol. 2017;17(Suppl 2):251

DOI 10.18699/BGRS/SB-2020-220

Construction of a barley gene regulatory network for phenolic metabolism using mutants and near-isogenic lines

Olesya Shoeva
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
olesya_ter@bionet.nsc.ru

Tatjana Kukoeva
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
kukoeva@bionet.nsc.ru

Mats Hansson
Lund University, Lund, Sweden
mats.hansson@biol.lu.se

Ksenia Strygina
VIR, St. Petersburg, Russia
k.strygina@vir.nw.ru

Elena Gordeeva
ICG SB RAS, Novosibirsk, Russia
elgordeeva@bionet.nsc.ru

Elena Khlestkina
ICG SB RAS, Novosibirsk, Russia
VIR, St. Petersburg, Russia
khlest@bionet.nsc.ru

Anastasiia Glagoleva
Kurchatov Genomics Center,
ICG SB RAS, Novosibirsk, Russia
glagoleva@bionet.nsc.ru

Shakhira Zakhrabekova
Lund University, Lund, Sweden
shakhira.zakhrabekova@biol.lu.se

***Abstract* — In the present study, the components of the gene regulatory network governing the anthocyanin and proanthocyanidin biosynthesis in barely grain and leaf sheath were identified based on comparative mapping and transcription analysis approaches. The regulatory relationships between the genes were established using unique barley anthocyanin/proanthocyanidin mutants and near-isogenic lines.**

Keywords* — *anthocyanins, proanthocyanidins, structural and regulatory genes, regulatory loops, Hordeum vulgare L.

Motivation and aim

Phenolic compounds are plant secondary metabolites with important functions in plant physiology and adaptation [1]. They are synthesized in various parts of the plant through phenylpropanoid biosynthesis pathway and different branches of flavonoid biosynthesis pathway. In barley (*Hordeum vulgare* L.), the application of the grain is dependent on the phenolic composition. For example, proanthocyanidins (also known as condensed tannins) are undesirable in barley cultivars used in malt production as they cause chill haze reducing beer quality [2]. In contrast, anthocyanins that can accumulate in aleurone layer and pericarp and cause blue and purple color of grain, respectively, are promising functional food ingredients and are desirable in cultivars for human consumption [3]. More information is required to fully understand how to manage the content of phenolic compounds in the barley grain. Two groups of genes control the biosynthesis of phenolic compounds: structural genes that encode enzymes of the pathway, and regulatory genes that encode MYB, bHLH or WD40 transcriptional factors (TFs). The TFs form complexes that activate transcription of the structural genes in a tissue-specific manner [4].

The aim of the current study was to identify the components of the gene regulatory networks underlining synthesis of phenolic compounds in barley grains and to study regulatory relationships between them in the course of synthesis of anthocyanins and proanthocyanidins.

Methods

A set of barley mutants with lesion in anthocyanin and/or proanthocyanidin biosynthesis and Bowman near-isogenic lines (NILs) differing in anthocyanin pigmentation of grain and leaf sheath were used in the study. The barley lines were obtained from NordGen (Alnarp, Sweden). Nucleotide sequences of the candidate genes for the *Ant1*, *Ant2*, *Ant 5*, *Ant13* loci were determined by Sanger sequencing of the corresponding mutants, NILs and their parental cultivars. Expression of the anthocyanin biosynthesis genes in grain and leaf sheath was explored by qRT-PCR.

Results

The regulatory function of the *Ant1* (mapped to chromosome 7HS), *Ant2* (2HL) and *Ant13* (6HL) loci on anthocyanin biosynthesis were suggested based on the previous studies of the anthocyanin/proanthocyanidin-less barley mutants [5]. Comparative mapping data allowed us to assign molecular functions to the loci. The *Ant1* gene encodes a MYB-like TF, required for the activation of expression of the anthocyanin biosynthesis structural genes in leaf sheath [6,7], while the *Ant2* gene encodes a bHLH-type TF that is required together with *Ant1* for the activation of the anthocyanin biosynthesis in grain pericarp [8]. *Ant13* encodes a WD40 TF. SNPs leading to stop codons and amino acid substitutions were identified in the *ant13* mutants. In the case of the *Ant1* and *Ant2* genes differences in promoter regions were identified and were assumed to prevent expression of the recessive alleles of these genes in non-colored grain pericarp. Bowman and its purple-grained NIL with dominant alleles of the both *Ant1* and *Ant2* genes were used to create new lines with different combinations of the dominant and recessive alleles of these genes [8]. Expression analysis of the structural genes in the obtained NILs revealed that the transcription of the structural gene *Ans* that encodes anthocyanidin synthase is strongly dependent on the dominant alleles of the both *Ant1* and *Ant2* genes. As the analysis of the anthocyanin-less mutants demonstrated that the *Ant1*, *Ant2* and *Ant5* loci controlled together the reaction of conversion of leucoanthocyanidins into anthocyanidins catalyzed by ANS [5], the *Ans* gene was suggested as a candidate gene for the *Ant5* locus. Sequencing of the *Ans* gene in a set of *ant5* mutants and parental cultivars revealed the SNPs causing stop codons and amino acid substitutions destroying the protein function.

The regulatory relationships between the *Ant1* and *Ant2* genes were established. Transcription of *Ant2* was up-regulated in the presence of a dominant allele of *Ant1* and transcription of *Ant1* was increased in the presence of a dominant allele of *Ant2*. The data suggested existence of a

positive regulatory loop between the two genes. The present study is the first report on interaction between the anthocyanins regulatory genes in monocot plant species.

Acknowledgment

The study was supported by RFBR grant №16-34-60052, the Kurchatov Genomics Center of ICG № 075-15-2019-1662, the Swedish Institute Visby Programme 25896/2018, the Swedish Research Council (VR 2018-05117), and the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS 2018-01026).

References

[1] Babenko L.M. (2019). Phenolic compounds in plants: biogenesis and functions. The Ukrainian Biochemical Journal. 91: 5-18.

[2] von Wettstein D. (2007). From analysis of mutants to genetic engineering. Annual Review of Plant Biology. 109: 232-249.

[3] Zhu F. (2018). Anthocyanins in cereals: composition and health effects. Food Research International. 109: 232–249.

[4] Hichri I. et al. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway, Journal of Experimental Botany. 62: 2465–2483.

[5] Jende-Strid B. (1993). Genetic control of flavonoid biosynthesis in barley. Hereditas 119: 187-204.

[6] Shoeva et al. (2015). Barley *Ant1* is a homolog of maize *C1* and its product is part of the regulatory machinery governing anthocyanin synthesis in the leaf sheath. Plant Breeding 134: 400-405.

[7] Zakhrabekova et al. (2015). Genetic linkage facilitates cloning of *Ert-m* regulating plant architecture in barley and identified a strong candidate of *Ant1* involved in anthocyanin biosynthesis. Plant Molecular Biology. 88: 609–626.

[8] Gordeeva et al. (2019). Purple-grained barley (*Hordeum vulgare* L.): marker-assisted development of NILs for investigating peculiarities of the anthocyanin biosynthesis regulatory network. BMC Plant Biology 19: 52.

DOI 10.18699/BGRS/SB-2020-221

Genomic analysis of Vavilov's historic chickpea landraces using GWAS, AMMI and GGE biplot analyses

Alena Sokolkova
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
alyonasok@yandex.ru

Sergey V. Bulyntsev
Federal Research Centre All-Russian N.I. Vavilov Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia
s_bulyntsev@mail.ru

Peter L. Chang
University of Southern California, Program in Molecular and Computational Biology, Dornsife College of Letters Arts & Sciences, Los Angeles, CA 90089 USA
peterc@usc.edu

Noelia Carrasquila-Garcia
University of California Davis, Department of Plant Pathology, Davis, CA 95616 USA
noecarras@ucdavis.edu

Eric von Wettberg
University of Vermont, Department of Plant and Soil Science, Burlington, VT 05405, USA
College of Letters
eric.bishop-von-wettberg@uvm.edu

Margarita A. Vishnyakova
Federal Research Centre All-Russian N.I. Vavilov Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia
m.vishnyakova@vir.nw.ru

Douglas R. Cook
University of California Davis, Department of Plant Pathology, Davis, CA 95616 USA
drcook@ucdavis.edu

Sergey V. Nuzhdin
University of Southern California, Program in Molecular and Computational Biology, Dornsife Arts & Sciences, Los Angeles, CA 90089 USA
snuzhdin@usc.edu

Maria G. Samsonova
Peter the Great St. Petersburg Polytechnic University, Department of Applied Mathematics, St. Petersburg, Russia
m.g.samsonova@gmail.com

***Abstract* — The Vavilov seed bank contains numerous landraces collected nearly hundred years ago, before intensive breeding of most crops, and thus is a potential reservoir of historic crop adaptations. Here we analyze the genomes of 407 of Vavilov's original landraces, sampled from major historic centers of chickpea cultivation and secondary diversification. We performed GWAS to find associations between VIR's chickpea accessions and phenotypic data obtained in Kuban experimental station of VIR in 2017 under infectious conditions. To find out the effects of infectious conditions on phenotypic traits, the data was subjected to AMMI and GGE biplot analyses.**

Keywords — chickpea (Cicer arietinum L.), GWAS analysis, candidate genes, AMMI, GGE biplot

Motivation and Aim

The Vavilov seed bank contains numerous landraces collected nearly hundred years ago, before intensive breeding of most crops, and thus is a potential reservoir of historic crop adaptations. The expanded diversity of these early collections likely contains 'genetic gems' with the potential to enhance modern breeding efforts [1]. Here we analyze the genomes of 407 of Vavilov's original landraces, sampled from major historic centers of chickpea cultivation and secondary diversification.

Earlier we have performed Genome-wide association study (GWAS) to find associations between VIR's chickpea accessions and phenotypic data obtained in Kuban experimental station of VIR in 2016. GWAS analysis identified a large number of genome intervals and potential gene candidates that may affect important agronomic traits. In 2017 the phenotypic data was obtained in Kuban experimental station of VIR under infectious conditions. GWAS analysis was performed with phenotypic data obtained in 2017 and AMMI (Additive Main effects and Multiplicative Interaction) and GGE (Genotype plus Genotype-Environment interaction) biplot analyses were performed to determine the effect of genotype, environment and their interaction for phenotypic traits measured both in 2016 and 2017 under different infectious conditions.

Methods

Genotyping by sequencing (GBS) of VIR's 407 chickpea accessions from different countries identified 56,855 segregating single nucleotide polymorphisms (SNP). For these SNP calls we implemented inclusion criteria: minor allele frequency (MAF) more than 3%, genotype call-rate more than 90%. 2,579 SNPs passed all filters and remained for further analysis. We performed GWAS to find associations between VIR's chickpea accessions and phenotypic data obtained in Kuban experimental station of VIR in 2017 under infectious conditions. To find out the effects of infectious conditions on phenotypic traits, the data was subjected to AMMI and GGE biplot analyses.

Results

GWAS analysis identified three SNPs on chromosomes 2, 7 and 8, respectively, significally associated with days from begginning to end of flowering under infectious conditions. SNP on chromosome 8 was also detected in previous GWAS analysis performed with phenotypic traits obtained in 2016 and with specific climatic variables at collection sites [2]. This SNP is a strong candidate marker for future selection. The results indicated that AMMI and GGE biplot analyses are informative methods to evaluate the effects of infectious conditions on phenotypic traits.

Acknowledgment

Supported by the RSCF (16-16-00007).

References

[1] Plekhanova E. et al. (2017) Genomic and phenotypic analysis of Vavilov's historic landraces reveals the impact of environment and genomic islands of agronomic traits. Sci. Rep. 7:4816. doi: 10.1038/s41598-017-05087-5.

[2] Sokolkova A.B. et al. (2020) Signatures of ecological adaptation in genomes of chickpea landraces. Biofizika. 65(2): 1-4.

DOI 10.18699/BGRS/SB-2020-222

Analysis of agronomic traits of mungbean (*Vigna radiata*) accessions from the World Vegetable Gene Bank (Taiwan)

Alena Sokolkova
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
alyonasok@yandex.ru

Marina Burlyaeva
Federal Research Centre All-Russian N.I. Vavilov Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia
m.burlyaeva@vir.nw.ru

Eric von Wettberg
University of Vermont, Department of Plant and Soil Science, Burlington, VT 05405, USA
eric.bishop-von-wettberg@uvm.edu

Margarita A. Vishnyakova
Federal Research Centre All-Russian N.I. Vavilov Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia
m.vishnyakova@vir.nw.ru

Roland Schafleitner
World Vegetable Center, Shanhua, Tainan 74199, Taiwan
roland.schafleitner@worldveg.org

Tatjana Valiannikova
Kuban Branch of Federal Research Centre All-Russian N.I. Vavilov Institute of Plant Genetic Resources (VIR), Krasnodar region, Russia
fitolaka61@yandex.ru

Chau-Ti Ting
National Taiwan University, Taipei 106, Taiwan
ctting@ntu.edu.tw

Sergey V. Nuzhdin
University of Southern California, Program in Molecular and Computational Biology, Dornsife College of Letters Arts & Sciences, Los Angeles, CA 90089 USA
snuzhdin@usc.edu

Cheng-Ruei Lee
National Taiwan University, Taipei 106, Taiwan
chengrueilee@ntu.edu.tw

Maria G. Samsonova
Peter the Great St. Petersburg Polytechnic University, Department of Applied Mathematics, St. Petersburg, Russia
m.g.samsonova@gmail.com

***Abstract* — The World Vegetable Gene Bank, in Taiwan, houses a unique genebank of mungbean. Here we analyze the genomes of 293 accessions that represent the major market classes. GWAS analysis conducted on phenotypic data obtained in Astrakhan experimental station of VIR in 2019 and multi-environmental GWAS analysis with phenotypic data obtained in different locations and time periods (Taiwan 1984, Taiwan 2018, India 2016, Kuban 2018, Astrakhan 2019) identified a large number of genome intervals and potential gene candidates that may affect important agronomic traits. To find out the effects of different climatic conditions on phenotypic traits, the data was subjected to AMMI and GGE biplot analyses. Phenotyping the mini-core collection of mungbean in different climatic locations can estimate phenology traits such as maturation. Uncovering SNPs for these traits will speed breeding efforts for production of mungbean suitable for cultivation in temperate regions.**

Keywords* — *mungbean (Vigna radiata), GWAS analysis, candidate genes, AMMI, GGE biplot

Motivation and aim

Motivation

Vigna radiata is important Asian pulse crop, important as dried seeds, sprouts, and in the green form as vegetables. With good heat and drought tolerance, this crop is often grown as summer pulses in contrast to winter pulses such as chickpea and lentil, making it important for food security and nutritional diversity in many parts of Asia. The World Vegetable Gene Bank, in Taiwan, houses a unique genebank of mungbean [1]. Here we analyze the genomes of 293 accessions that represent the major market classes.

Aim

Earlier we have performed Genome-wide association study (GWAS) to find associations between accessions of the mini-core collection and phenotypic data obtained in Kuban experimental station of VIR in 2018 under stress weather conditions. GWAS analysis identified a block of four SNPs within the region of strong linkage associated with maturation in stress weather conditions in which genes encoding zinc finger A20 and AN1 domain stress-associated protein are located. SAPs are known to prevent the loss of yield caused by environmental stresses [2]. In 2019 the phenotypic data was obtained in Astrakhan experimental station of VIR. GWAS analysis was performed with phenotypic data obtained in 2019. Previously, the mini-core collection has been grown in common field plots in several locations, including Taiwan (1984, 2018), and India in 2016. Multi-Environmental GWAS analysis was performed with phenotypic data obtained in different locations and time periods and AMMI (Additive Main effects and Multiplicative Interaction) and GGE (Genotype plus Genotype-Environment interaction) biplot analyses were performed to determine the effect of genotype, environment and their interaction for phenotypic traits measured under different climatic conditions.

Methods

Genotyping by sequencing (GBS) of the mini-core collection's 293 mungbean accessions from different countries identified 8,466 segregating single nucleotide polymorphisms (SNP). For these SNP calls we implemented inclusion criteria: minor allele frequency (MAF) more than 3%, genotype call-rate more than 90%. 5,041 SNPs passed all filters and remained for further analysis. We performed GWAS to find associations

between mungbean accessions and phenotypic data obtained in Astrakhan experimental station of VIR in 2019 and multi-environmental GWAS analysis with phenotypic data obtained in different locations and time periods (Taiwan 1984, Taiwan 2018, India 2016, Kuban 2018, Astrakhan 2019). To find out the effects of different climatic conditions on phenotypic traits, the data was subjected to AMMI and GGE biplot analyses.

Results

GWAS analysis conducted on phenotypic data obtained in Astrakhan experimental station of VIR in 2019 and multi-environmental GWAS analysis with phenotypic data obtained in different locations and time periods (Taiwan 1984, Taiwan 2018, India 2016, Kuban 2018, Astrakhan 2019) identified a large number of genome intervals and potential gene candidates that may affect important agronomic traits. The results indicated that AMMI and GGE biplot analyses are informative methods to evaluate the effects of different climatic conditions on phenotypic traits. The mini-core collection of mungbean established by the World Vegetable Center is a valuable resource for mungbean breeding. Phenotyping of this collection in different climatic locations can estimate phenology traits such as maturation. Uncovering SNPs for these traits will speed breeding efforts for production of mungbean suitable for cultivation in temperate regions.

ACKNOWLEDGMENT

Supported by the RSCF (18-46-08001).

REFERENCES

[1] Schafleitner R et al. (2015) The AVRDC – The World Vegetable Center mungbean (*Vigna radiata*) core and mini core collections. BMC Genomics. 16(1):344. doi: 10.1186/s12864-015-1556-7.

[2] Ben Saad R et al. (2019) Functional domain analysis of LmSAP protein reveals the crucial role of the zinc-finger A20 domain in abiotic stress tolerance. Protoplasma. 256(5):1333-1344. doi: 10.1007/s00709-019-01390-2.

DOI 10.18699/BGRS/SB-2020-223

Effects of anthocyanin-rich grain diet on growth and metastasis of Lewis lung carcinoma in mice

Michael V. Tenditnik
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
m.v.tenditnik@physiol.ru

Maria A. Tikhonova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
tikhonovama@physiol.ru

Ekaterina A. Litvinova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
litvinovaea@physiol.ru

Nelly A. Popova
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
nelly@bionet.nsc.ru

Tamara G. Amstislavskaya
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
Amstislavskaya@yandex.ru

Elena K. Khlestkina
N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources
St. Petersburg, Russia
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
khlest@bionet.nsc.ru

Abstract — **Currently functional nutrition has developed intensively. Functional foods are a valuable addition to existing dietary therapy. Such products are enriched with biologically active substances. Anthocyanins attract particular attention due to multiple beneficial properties including antitumor activity. Here we used two wheat near-isogenic lines created at the Institute of Cytology and Genetics SB RAS [1] that have almost similar genomes with the exception of a small part of chromosome 2A, which contains the Pp3/TaMyc1 gene regulating anthocyanin biosynthesis, to assess the effect of an anthocyanin-rich grain diet on the development and metastasis of the Lewis lung carcinoma (LLC), as well as the accompanying immune response in C57BL/6 mice. Mice were kept at a grain or standard diet for four month prior tumor transplantation. A decrease in the number of metastases in the lungs and the size of the tumor in the groups at a grain diet, regardless of the content of anthocyanins, was revealed. However, the highest percentage of animals without metastases was observed at an anthocyanin-rich grain diet. The LLC transplantation caused a significant increase in plasma levels of pro-inflammatory cytokines (IL-6 and LIF) in mice at a standard diet, but not in animals at grain diets. Thus, the anthocyanin-rich diet helps to reduce the severity of the tumor process. The antitumor effects of grain diets are probably mediated by modulation of signaling pathways associated with IL-6 cytokines.**

Keywords — functional food, wheat grain, anthocyanins, tumor, metastasis, mice, immune, cytokine.

ACKNOWLEDGMENT

The study was supported by Russian Science Foundation (grant No. 16-14-00086).

REFERENCES

[1] E. I. Gordeeva, O. Y. Shoeva, and E. K.Khlestkina, “Marker-assisted development of bread wheat near-isogenic lines carrying various combinations of purple pericarp (*Pp*) alleles,” Euphytica, vol. 203, pp. 469–476, 2015.

DOI 10.18699/BGRS/SB-2020-224

MtWOX9-1 gene as somatic embryogenesis stimulator. Search of targets

Varvara Tvorogova
SPSU, St Petersburg, Russia
krubaza@mail.ru

Elizaveta Krasnoperova
SPSU, St Petersburg, Russia
liza_krasnoperova99@mail.ru

Andrei Kudriashov
SPSU, St Petersburg, Russia
a23.10.2012@yandex.ru

Ksenia Kuznetsova
SPSU, St Petersburg, Russia
kskuz95@mail.ru

Elina Potsenkovskaya
SPSU, St Petersburg, Russia
epots556@gmail.com

Ludmila Lutova
SPSU, St Petersburg, Russia
la.lutova@gmail.com

***Abstract — MtWOX9-1* gene was shown to stimulate somatic embryogenesis *in vitro* in *Medicago truncatula.* To understand its functions in this process, we performed transcriptomic analysis of *MtWOX9-1* overexpressing calli, as well as expression correlation analysis. As a result, we obtained a list of possible MtWOX9-1 targets during somatic emyryogenesis. Now we are checking these possible targets using ChIP-seq analysis.**

Keywords — somatic embryogenesis, transcription factors, Medicago truncatula

Motivation and Aim

Motivation

Somatic embryogenesis (SE) is the development of embryo-like structures from somatic plant tissues. SE *in vitro* is widely used for plant propagation and transformation; therefore, the search for SE stimulators and revealing of the mechanisms of their functioning are very important for biotechnology.

Aim

Previously, we have shown that transcription factor MtWOX9-1, belonging to the WOX family, can stimulate SE in the *Medicago truncatula* callus culture. In this research, transcriptomic analysis of highly embryogenic calli with MtWOX9-1 overexpression was performed.

Methods

We used RNA from transgenic calli with *MtWOX9-1* overexpression and wildtype calli of R-108 line as a control to perform trancriptomic analysis. During bioinformatics processing of transcriptomic data, we used Trimmomatic for deleting adapter sequences (Bolger et al., 2014), HISAT2 for alignment with the reference genome (Kim et al., 2015), Stringtie for counting reads (Pertea et al., 2015), DESeq2 for the search for differentially expressed genes (Love et al., 2014), GSEABase for gene enrichment analysis (Morgan et al., 2019) and WGCNA for gene correlation networks search (Langfelder, Horvath, 2008).

Results

It was shown that *MtWOX9-1* overexpression led to the activation of several groups of genes, including genes related to cell division, tissue differentiation, and seed development. Among enriched GO groups, we found several gene sets, related with repressive epigenetic changes. Using Medicago Truncatula Gene Expression Atlas, we also identified a group of genes encoding for transcription factors, that were both coexpressed with *MtWOX9-1* in different plant organs and differentially expressed in our samples. These genes are putative targets of MtWOX9-1, and they may act in the same pathway with this regulator during SE. Now we are checking, if MtWOX9-1 binds indeed to these targets, using ChIP-seq analysis.

Acknowledgment

Supported by the RFBR (19-14-50209, 20-016-00124) and RSF (16-16-10011).

References

[1] Bolger A. M., Lohse M., Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15): 2114–2120 .

[2] Kim D., Langmead B., Salzberg S. L. (2015) HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12(4): 357–360.

[3] Langfelder P., Horvath S. (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9(1): 559.

[4] Love M. I., Huber W., Anders S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12):550.

[5] Morgan M., Falcon S., Gentleman R. (2019). GSEABase: Gene set enrichment data structures and methods. R package version 1.48.0.

[6] Pertea, M. et al. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33(3): 290–295.

DOI 10.18699/BGRS/SB-2020-225

Btr1 genes and the evolution of wheat and *Aegilops* species

Valeriya Vavilova
ICG SB RAS, Novosibirsk, Russia
valeriya-vavilova@bionet.nsc.ru

Irina Konopatskaia
ICG SB RAS, Novosibirsk, Russia
sormacheva@bionet.nsc.ru

Alexandr Blinov
ICG SB RAS, Novosibirsk, Russia
blinov@bionet.nsc.ru

Nikolay P. Goncharov
ICG SB RAS, Novosibirsk, Russia
gonch@bionet.nsc.ru

***Abstract* — *Non-brittle rachis 1* (*Btr1*) gene involves in regulation of the brittle/ non-brittle spike trait, which is one of the key domestication traits in *Triticum* species. In this study genetic variability of *Btr1* genes from di- and tetraploid wheat species and *Aegilops speltoides*, which is donor of B-genome for polyploid wheat species, were investigated.**

Keywords* — *Triticum, Aegilops, spike morphology, brittle rachis, Btr1 gene

Motivation and aim

Motivation

The *Non-brittle rachis 1* (*Btr1*) and *Non-brittle rachis 2* (*Btr2*) genes are known to control important agriculture trait in barley, namely the formation of a non-brittle rachis [1]. Homologues genes *Brt1* and *Btr2* were recently described for wheat species [2,3]. It was shown that one non-synonymous substitution in the coding region of *Btr1-A* gene (A119T) leads to non-brittle rachis formation in accessions of diploid einkorn wheat *Triticum monococcum* [2,4]. The investigation of *Btr1-A* gene variability in polyploid wheat revealed allele which possess the loos-of-function deletion 2bp in length within the coding region [3, 4]. All poliploid wheat species studied so far possess the above mentioned deletion in *Btr1-A*.

Aim

In the present study we investigated genetic variability of *Btr1* genes from di- and tetraploid wheat species and *Aegilops speltoides*, which is B-genome donor of polyploid wheats, in order to clarify the phylogenetic relationships between *Triticum* and *Aegilops* genera.

Methods

Germplasm of di- and tetraploid wheat species and *Ae. speltoides* were grown under standard greenhouse conditions. Brittle/ non-brittle spike trait was determined visually. List of wheat accessions analyzed in this study is presented in Table I. Total DNA was isolated from 100 mg of leaves using the Wizard® Genomic DNA Purification Kit (Promega) according to the manufacturer's protocol. The whole genome sequences (WGS) of *Triticum dicoccoides* (LSYQ02000006), *Triticum aestivum* (OETA01178479) (B-genome) were used to design genome-specific primers to amplify *Btr1-B* and *Btr1-S* genes from tetraploid wheats and *Ae. speltoides*. The primer pair Btr1-BF/ Btr1-BR 5'-GACGAGCTTGACCTCCCATGAG -3'/ 5'-GTGCGAATCGCTACTCCATC -3' amplified partial upstream and downstream parts and coding region *Btr1-B* and *Btr1-S* genes. For PCR amplification of the same *Btr1-A* region primers Btr-A1-F/ Btr-A1-R were used [5]. PCR was performed in 20 µl volume containing 20 ng of genomic DNA, 10 mM Tris-HCl (pH 8.9), 1 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 200 µM dNTPs, 0.5 µM primers, and 0.25 U of Taq DNA polymerase. PCR products were separated by agarose gel electrophoresis and purified using a QIAquick Gel Extraction Kit (QIAGEN). For all wheat accessions purified PCR products were sequenced. Sequencing reactions were performed with 20 ng of the PCR product and ABI BigDye Terminator Kit on an ABI 3130XL Genetic Analyser (Applied Biosystems) in SB RAS Genomics Core Facility (http://www.niboch.nsc.ru/doku.php/corefacility). Nucleotide and amino acid sequences alignments were performed using AliView v. 1.18.1 [6]. Phylogenetic analysis of sequences was performed with the IQ-TREE 1.6.11 program [7].

WHEAT AND AEGILOPS ACCESSIONS USED IN THE STUDY AND THEIR PHENOTYPES

Species	Accession	Phenotype
Triticum monococcum L.	PI 289599	Non-brittle rachis
T. urartu Thum. ex Gandil.	IG-44829	Brittle rachis
T. boeoticum Boiss.	к-40118	Brittle rachis
T. dicoccoides (Körn. ex Asch. & Graebn.) Schweinf.	PI 467027	Brittle rachis
T. dicoccoides (Körn. ex Asch. & Graebn.) Schweinf.	PI 66841	Brittle rachis
T. dicoccum (Schrank) Schuebl.	к-7500	Brittle rachis
T. dicoccum (Schrank) Schuebl.	к-327430	Brittle rachis
T. karamyschevii Nevski	KU190	Brittle rachis
T. turgidum L.	к-16156	Non-brittle rachis
T. durum Desf.	к-17784	Non-brittle rachis
T. durum Desf.	к-17787	Non-brittle rachis
T. turanicum Jakubz.	к-15993	Non-brittle rachis
T. polonicum L.	к-39297	Non-brittle rachis
T. polonicum L.	к-17893	Non-brittle rachis
T. aethiopicum Jakubz.	18999	Non-brittle rachis
T. carthlicum Nevski	к-7106	Non-brittle rachis
T. araraticum Jakubz.	к-30258	Brittle rachis
T. araraticum Jakubz.	к-31628	Brittle rachis
T. timopheevii (Zhuk.) Zhuk.	к-29537	Brittle rachis
T. timopheevii (Zhuk.) Zhuk.	к-29540	Brittle rachis
Aegilops speltoides Tausch	Ae-352	Brittle rachis
Ae. speltoides Tausch	к-1596	Brittle rachis

Results

We obtained the *Btr1* (*Btr1-A* and *Btr1-B*) and *Btr1-S* sequences for 14 wheat species and *Ae. speltoides*, respectively (Table I).

For diploid wheat accessions the length of the *Btr1-A* sequences obtained were 902 bp. In the case of tetraploid wheat accessions, the length of the gene sequences was 899 bp. There were 1 bp deletion in the upstream and 2 bp deletion in the coding region of the *Btr1-A* gene compared with diploids. According to the comparative analyzes, the *Btr1-A* gene sequences of *T. monococcum* (PI 289599), *T. urartu* (IG-44829) and *T. boeoticum* (k-40118) were identical with sequences presented previously [2]. Deletion of 2 bp in the *Btr1-A* coding region of 11 tetraploid wheat species studied was identical to hexaploid wheats [3, 5]. This deletion formed a premature stop codon and resulted in a non-functional protein [3].

Nucleotide sequences of *Btr1-B* and *Btr1-S* genes were detected for the first time. For all tetraploid and *Aegilops* accessions analyzed the length of the *Btr1-B* and *Btr1-S* sequences were variable. This variability was caused by poly-(GA) track in promoter region of the genes. No indels in coding region were detected.

We conducted phylogenetic analysis of *Btr1* genes from di- and tetraploid wheat species and *Ae. speltoides*. The data obtained allowed us to clarify the phylogenetic relationships between *Triticum* and *Aegilops* genera.

Acknowledgment

The study was supported by the Russian Science Foundation (grant №: 16-16-10021).

References

[1] Pourkheirandish M., Hensel G., Kilian B., Senthil N., Chen G., Sameri M., Azhaguvel P., Sakuma S., Dhanagond S., Sharma R., Mascher M., Himmelbach A., Gottwald S., Nair S.K., Tagiri A., Yukuhiro F., Nagamura Y., Kanamori H., Matsumoto T., Willcox G., Middleton C.P., Wicker T., Walther A., Waugh R., Fincher G.B., Stein N., Kumlehn J., Sato K., Komatsuda T. (2015) Evolution of the Grain Dispersal System in Barley. Cell. 162(3):527-539.

[2] Pourkheirandish M., Dai F., Sakuma S., Kanamori H., Distelfeld A., Willcox G., Kawahara T., Matsumoto T., Kilian B., Komatsuda T. (2018) On the origin of the non-brittle rachis trait of domesticated einkorn wheat. Front Plant Sci. 8:1–10.

[3] Zhao Y., Xie P., Guan P., Wang Y., Li Y., Yu K., Xin M., Hu Z., Yao Y., Ni Z., Sun Q., Xie C., Peng H. (2019) *Btr1-A* induces grain shattering and affects spike morphology and yield-related traits in wheat. Plant Cell Physiol. 60:1342–1353.

[4] Vavilova V., Konopatskaia I., Blinov A., Goncharov N.P. (2020) Evolution of *Btr1-A* gene in diploid wheat species of genus *Triticum* L. Russian Journal of Genetics. 56 (5): 1–6 (in Russian).

[5] Vavilova V., Konopatskaia I., Goncharov N.P. (2019) *Non-brittle rachis 1-A* (*Btr1-A*) gene in di- and hexaploid wheat species. Current Challenges in Plant Genetics, Genomics, Bioinformatics, and Biotechnology. 202–204.

[6] Larsson A. (2014) AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics 30(22): 3276-3278.

[7] Nguyen L.T., Schmidt H.A., von Haeseler A., Minh B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1): 268-274.

DOI 10.18699/BGRS/SB-2020-226

Study of the root transcriptome of bread wheat using high-throughput RNA sequencing (RNA-SEQ)

Alexandr Vikhorev
Novosibirsk State University
Novosibirsk, Russia
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
vikhorev@bionet.nsc.ru

Nikolay Shmakov
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia

Anastasia Glagoleva
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia

Olesya Shoeva
Institute of Cytology and Genetics
SB RAS
Nobosibirsk, Russia

Elena Khlestkina
All-Russian Institute of Plant
Recources
Saint-Petersburg, Russia

***Abstract* — Bread wheat (*Triticum aestivum* L.) is the most important crop in the world. It provides about 20% of the total calories consumed by humans. For a long time, wheat selection was mainly based on phenotypic traits of the shoot, but the roots were given little attention. As a result, the root system of modern wheat varieties has weakened. Therefore, the study of genetic control of wheat roots development is an urgent issue. In this study, sequencing of RNA libraries from roots and coleoptile of Russian spring variety "Saratovskaya, 29" was performed. *De novo* transcriptome was assembled. 31,488 up-regulated, 35,851 down-regulated and 18,040 roots-specific transcripts were found. The subsequent analysis of genes with differential expression will allow choosing the candidate genes for development of wheat varieties with resistant root system.**

Keywords* — *wheat, roots, transcriptome, RNA-seq

Introduction

In the middle of 20th century dwarf wheat varieties were obtained, and this event was named the "Green revolution". These varieties had strong, short stem, dark-green erected leaves and were able to produce a large amount of yield under nitrogen fertilizers [1]. But growing population size and climate changes require an increase in grain production not only in well-fertilized, but also in nutrient-poor soils. Therefore, there is necessity in second green revolution, and its source can be roots [2].

Hidden underground, roots perform functions crucial for the plant organism life, such as water acquisition, supply of nutrients and anchoring. Despite that, roots were out of attention of breeders throughout the history of cereal plant breeding. This has led to the weakening of the roots of modern varieties compared to their ancestors. Research shows that the root system of modern wheat varieties may not be able to adequately supply water and nutrients for developing seeds [3].

A convenient method to study the genetic control of roots development is transcriptome profiling using high- throughput RNA sequencing (RNA-seq). The method allows us to evaluate gene expression across the entire genome, as well as to find specific genes responsible for roots development that will be used in the future during marker-assisted selection of wheat varieties with resistant root system.

The aim of the current study is revealing differentially expressed genes between the root and shoot transcriptomes of allohexaploid wheat using RNA-seq analysis and identifying root-specific genes involved in the wheat root system development.

Materials and Methods

Plant material and RNA sequencing

RNA was extracted from 4-days roots and coleoptiles of seedlings of Russian spring wheat cultivar "Saratovskaya 29" in three biological replicas. One sample contained roots or coleoptile from six different plants. Sequencing was performed using Illumina NextSeq 550 platform in Institute of Cytology and Genetics SB RAS, 75-bp reads was obtained.

In silico analysis of RNA libraries

De novo assembly of transcriptome was performed using Trinity software. To evaluate quality of obtained transcriptome Transrate software was used. Quantification of reads was performed using Salmon software. To obtain different expressed transcripts edgeR package was used. For functional annotation Transrate, AgriGO and BlastKOALA services were used.

Results

De novo transcriptome assembly and evaluating

In our study, for the first time, *de novo* transcriptome of the Russian spreeing variety "Saratovskaya 29" was assembled. In the *de novo* assembly, 94.2% of conserved monocotyledonous plant genes were found, as well as 112,934 open reading frames, which is close to the number of reading frames in the annotated reference transcriptome (119,267 open reading frames). This metrics indicate a high quality of the obtained transcriptome.

Differential expand specifically expressed transcripts

67,339 (50% of the total) differentially expressed transcripts were found, of which 31,488 were up-regulated in the roots, and 35,851 were down-regulated. Root-specific expression was found in 18,040 (13%) transcripts.

Functional annotation

The analysis of saturation of gene ontology terms revealed that the most common products of differentially expressed

genes are membrane proteins and proteins associated with phosphorylation. These proteins are probably root-specific receptors and regulatory proteins responsible for root development.

References

[1] G. S. Khush, "Green revolution: The way forward", *Nature Reviews Genetics*, vol. 2, no. 10. Nature Publishing Group, pp. 815–822, Oct-2001.

[2] J. P. Lynch, "Roots of the second green revolution", Australian Journal of Botany, vol. 55, no. 5. pp. 493–512, 2007.

[3] J. G. Waines and B. Ehdaie, "Domestication and crop physiology: Roots of green-revolution wheat", Annals of Botany, vol. 100, no. 5. Oxford University Press, pp. 991–998, Oct-2007.

DOI 10.18699/BGRS/SB-2020-227

Development of DNA markers for identification of a quarantine weed, silverleaf nightshade (*Solanum elaeagnifolium* Cav.), based on chloroplast intergenic spacers

Volodina E.A.
VNIIKR, Bykovo, Russia
jugem14@gmail.com

Kulakova Y.Y.
VNIIKR, Bykovo, Russia
thymus73@mail.ru

Anisimenko M.S.
VNIIKR, Bykovo, Russia
jocus@bk.ru

Dobrovolskaya O.B.
VNIIKR, Bykovo, Russia
ICG SB RAS, Novosibirsk, Russia
oxana-d@yandex.ru

***Abstract* — Silverleaf nightshade is a noxious invasive weed that grows in open habitats in dry forests and scrublands. It became an invasive weed in many countries. In some cases, it is very difficult to identify the *Solanum elaeagnifolium* as individual species. In this work DNA markers were designed for detecting unique SNPs for identification of S. elaeagnifolium Cav.**

Keywords* — *Solanum elaeagnifolium Cav., Solanum, chloroplast genome, SNP

Motivation and Aim

Motivation

Solanum is the largest genera in the Solanaceae family, containing more than 1,400 species that may belong to cultural (*S. melongena* L., *S. lycopersicum* L., *S. tuberosum* L.) and weed plants (*S. nigrum* L., *S. elaeagnifolium* Cav., *S. carolinense* L., *S. rostratum* Dunal, *S. triflorum* Nutt. et al.). One of the invasive weed species is silverleaf nightshade (*S. elaeagnifolium* Cav.). In the first time the species was described by Cavanilles in 1794 from plant specimens grown in the botanical garden of Madrid. *Solanum elaeagnifolium* occurs in both North and South America, and is a harmful invasive weed in dry areas worldwide. It has become adventive on almost all continents and has received the status of a quarantine facility in Australia, Canada, Georgia, Moldova, Ukraine, the EAEU. In the Russian Federation, this species is absent, but there is a risk of penetration with imported seed and food materials.

Silverleaf nightshade characterizes by highly variable morphological traits (shape of leaf, pubescence, etc.), and this variability has led to much synonymy, so in recent years, most researchers have adopted a broad interpretation of the species.

Aim

The aim of this work was to study on the DNA polymorphism of chloroplast intergenic spacers of wild species of the genus Solanum, and design DNA markers for identification of *S. elaeagnifolium* Cav. a quarantine weed species.

Methods

The study was performed on 130 plant samples from the genus Solanum, including *S. elaeagnifolium* Cav., which were represented by 32 samples. The samples for this research were collected in different regions word wide, where silverleaf nightshade grow (Argentina, Mexico, Kenya, Japan, Vietnam, Morocco, etc.), and supplemented by herbarium samples (USA, Australia, Paraguay, etc.). The matK gene and non-coding regions of the chloroplast genome (trnL-trnT, ndhF-rpl32) were Sanger-sequenced and used for further DNA marker design.

Results

This study detected high intraspecific and interspecific variation in some chloroplast regions, and identify SNPs unique to *S. elaeagnifolium* Cav. DNA markers that detect these SNPs were designed. The work will be continued towards the design more species-specific primers and create a test system that allow the identification of quarantine species of the genus Solanum.

DOI 10.18699/BGRS/SB-2020-228

Complete sequencing of barley organellar genomes: new data for intraspecific differentiation

Yermakovich (Makarevich) Anna
IGS NAS of Belarus, Minsk, Belarus
bio.makarevich@gmail.com

Siniauskaya Maryna
IGS NAS of Belarus, Minsk, Belarus
m.sin@inbox.ru

Halayenka Innesa
IGS NAS of Belarus, Minsk, Belarus
goloenkoi@tut.by

Liaudanski Aleh
IGS NAS of Belarus, Minsk, Belarus
666555@tut.by

Davydenko Oleg
IGS NAS of Belarus, Minsk, Belarus
davydenko@tut.by

Abstract — **Organelle genomes are an important tool to investigate domestication, distribution and microevolution of plant species. However, they have found limited use in cereal intraspecific studies so far. In the present study, organelle genomes of wild (*Hordeum vulgare subsp. spontaneum*) and cultivated (*H. vulgare subsp. vulgare*) barley samples were sequenced. We conducted the NGS of isolated chloroplast and mitochondrial DNA mixtures. This non-trivial approach allowed to obtain both genomes for each sample but required some specific steps in the data processing. Comparative analysis of obtained sequences revealed more than 100 polymorphic sites in the chloroplast genome, including new intraspecific SSR-markers, and more than 20 polymorphisms in the mitochondrial genome. We also carry out the phylogenetic analysis of these genomes. Chloroplast and mitochondrial DNA trees were consistent with each other, indicating the presence of two large clades containing both wild and cultivated samples. Our results conform with a hypothesis of several domestication centres of barley. They also provide direct evidence of a higher rate of nucleotide substitutions in the chloroplast genomes as compared to that of mitochondria on a microevolution scale. The revealed high level of variability of chloroplast genomes makes it possible to use them for intraspecific barley differentiation.**

Keywords — barley, NGS, chloroplast genome, mitochondrial genome, phylogeny, intraspecific variability

Motivation and Aim

The barley substituted lines collection was created to study nuclear-cytoplasmic interactions. It is a suitable object for investigating the diversity of organelle genomes within *Hordeum vulgare* species, due to the presence of samples bearing the cytoplasmic genomes from a wide range of wild and cultivated barley.

In the present study, we performed the NGS of chloroplast (cp) and mitochondrial (mt) DNA of 10 wild barley samples (substituted lines with organelle genomes from *Hordeum vulgare subsp. spontaneum*) and 7 cultivar barley samples (5 cultivar varieties and 2 substituted lines with organelle genomes from *H. vulgare subsp. vulgare*).

The study aimed:

- Evaluate the diversity of organellar genomes within the *Hordeum vulgare* species;
- Estimate the prospects of using cp and mt genome sequences for intraspecific barley identification;
- Re-evaluate the phylogenetic structure of *H. vulgare* species using complete cp and mt genome sequences of the presented set of wild and cultivar barley samples.

Methods

The study sample contained 5 barley cultivar varieties (Vezha, Roland, Vizit, Sobolyok, Maresi) and 12 substituted lines whith cp and mt genomes from *H. vulgare subsp. vulgare* (varieties Atlas and Himalaya) and *H. vulgare subsp. spontaneum* (wild samples W1, W3, W4, W8, W9 from Israel). All the substituted barley lines were obtained by seven backcrossings (BC^7).

Organelle DNA of 3-days barley seedlings was extracted by phenol-chloroform method from the cp fraction, obtained by differential centrifugation method [1]. Such fraction usually contains an admixture of mt DNA. This fact allows simultaneous sequencing both cp and mt genomes from one sample but requires to consider areas of homology between these genomes during NGS data processing.

NGS of obtained DNA samples was performed with MiSeq System (Illumina Inc., SanDiego, CA, USA), using NexteraXT library preparation kit and MiSeq Reagent Kits v.3 and v.2. The NGS data processing algorithm was specially optimized for cpDNA and mtDNA mixtures with taking into account the presence of wide homology areas within the genomes and between them. Complete cp (NC008590) and mt (AP017301) genome sequences of *H. vulgare subsp. vulgare* were used as references. The algorithm was tested on artificial Illumina reads synthesized with ART set of simulation tools. The processing included following steps: the raw reads trimming (Trimmomatic); aligning the reads to the "double" reference, containing full sequences of cp and mt barley genomes (Bowtie2); obtaining mapping statistics (bash scripts, BCFtools); visualization of the alignment (Tablet); generating the VCF (Variant Call Format) files (Samtools); filtering of the VCF files (VCFlib)[2]. The resulted VCF files contained all the polymorphic loci of the cp and mt genomes of each sample. Complete FASTA sequences of these genomes were generated on the base of these files.

The phylogenetic analyses of the complete cp and mt genomes was carried out with Bayesian statistics methods (MrBayes) using the GTR (General Time-Reversible) model. The analyses involved sequences obtained in the study (17 cp and 17 mt sequences), as well as accessible in NCBI complete sequences of barley cp (NC008590, KC912688) and mt (AP017300, AP017301) genomes. Complete cp genome sequences of *H. jubatum* (KM974741) and *H. bulbosum* (KY636105) were used as outgroups for the cp tree. There were no accessible complete mt genome sequences of other barley species. So we used Roland(W9) as an outgroup for the mt genome phylogeny because this sample extremely differed from the others both in cp and mt genomes.

Results

We obtained and published in the NCBI GenBank database complete sequences of 17 cp and 17 mt genomes of: 10 samples of *Hordeum vulgare subsp. spontaneum* (substituted lines Roland(W1), Roland(W3), Roland(W4), Roland(W8), Roland(W9), Vezha(W3), Vezha(W8), Vizit(W3), Vizit(W4), Vizit(W8)) – NCBI accession numbers MN171376 – MN171392; 7 samples *of Hordeum vulgare subsp. vulgare* (cultivars: Vezha, Roland, Vizit, Sobolyok, Maresi; substituted lines: Roland(Atlas), Roland(Himalaya)) – NCBI accession numbers MN127966 – MN127982.

The comparative analysis of 22 complete cp genome sequences of barley (17 obtained and 5 previously published) revealed significant variability of this genomes within *H. vulgare species*. There were detected 107 polymorphic loci: 9 INDELs, 79 SNPs and 19 polymorphisms of SSR regions. The vast majority of cp genome polymorphisms lied outside the coding sequences, but there were 20 SNPs detected in genes.Among them, there were 6 amino acid substitution-causing mutations in genes of maturase K protein, translation initiation factor 1 protein, RNA polymerase beta' subunit and NADH dehydrogenase subunits K and 6. Synonymous SNPs were located in genes of RNA polymerase alpha and beta'' subunits, ATP synthase CF1 alpha subunit, photosystem I P700 chlorophyll apoprotein A1 and NADH dehydrogenase subunits 1, 4 and 7.

The analysis of mt genomes included 19 complete mt genome sequences of barley (17 obtained and 2 early published). The variability of mt genomes was noticeably lower than that in cp: we detected only 22 SNPs and 1 INDEL. There were 4 acid substitution-causing mutations: 2 SNPs in the ribosomal protein small subunit 4 gene and 2 SNPs in ribosomal protein S3A pseudogene.

The phylogenetic analysis involved 21 complete cp genome sequences (8 *H. vulgare subsp. vulgare*, 11 *H. vulgare subsp. spontaneum* and 2 outgroups) and 19 complete mt genome sequences (8 *H. vulgare subsp. vulgare*, 11 *H. vulgare subsp. spontaneum*). Roland(W9) showed the maximum number of differences from the other samples in the comparative analysis: 20 unique polymorphisms of cp genome and 10 of mt genome. So, this sample formed a separate clade on the cp phylogenetic tree and also was used as an outgroup in the mt genomes phylogeny.

The phylogenetic trees for both cp and mitochondrial genomes subdivided on two big clades (Fig. 1). Both clades contained as wild (*H. vulgare subsp. spontaneum*), as cultivated (*H. vulgare subsp. vulgare*) barley samples. These results indicated the presence of at least two centres of *H. vulgare subsp. vulgare* origin and conformed with the prevailing hypothesis of several domestication centres of barley. [3].

These results became the starting point for a broader study of the intraspecific diversity of chloroplast and mitochondrial genomes of *H. vulgare*. At the present moment, we are conducting NGS of other substituted lines from our collection and also for some cultivars and wild samples of barley from different countries. We hope to present the new results at the conference.

Acknowledgment

The study conducted as a part of the State program of scientific research "Biotechnology 2016-2020".

References

[1] S.O. Triboush, N.G. Danilenko, O.G. Davydenko, "A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower", Plant Molecular Biology Reporter,. vol. 16, 1998, pp. 183-189.

[2] A. Makarevich, M. Siniauskaya, V. Pankratov et al, "Revealing organelle genomes diversity of barley alloplasmic lines", Book of abstracts of GRC2019, Gatersleben Research Conference: Applied Bioinformatics for Crops: 18–20 March 2019. –Gatersleben, Federal Republic of Germany, 2019, p. 64.

[3] A. Pankin et al, "Targeted resequencing reveals genomic signatures of barley domestication.", The New phytologist, vol. 218, iss.3, 2018,pp. 1247-1259.

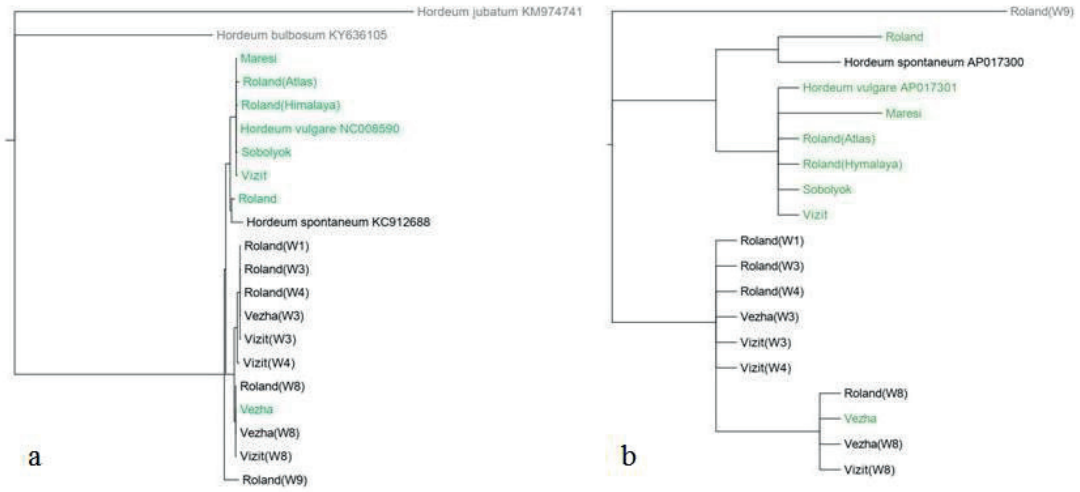


Fig. 1 – The phylogenetic trees of the complete chloroplast (a) and mitochondrial (b) genome sequences of the studied barley samples

Symposium
Animal genetics and genomics

DOI 10.18699/BGRS/SB-2020-231

Differentially expressed genes in longitudinal axis of the fox's hippocampus

Yury Alexandrovich
ICG SB RAS, Novosibirsk, Russia
alexandrovich@bionet.nsc.ru

Larisa Meister
ICG SB RAS, Novosibirsk, Russia
meister@bionet.nsc.ru

Yury Herbeck
ICG SB RAS, Novosibirsk, Russia
herbek@bionet.nsc.ru

***Abstract* — The role of hippocampus in varied neural processes is well known; however, hippocampus is not a uniform structure. Our results demonstrated differences in gene expression between dorsal and ventral hippocampus in foxes. The expression of *CYP26B1* in dorsal hippocampus is increased in tame foxes relatively to aggressive ones.**

Keywords* — *dorsal hippocampus, ventral hippocampus, adult neurogenesis

Motivation and aim

Motivation

Hippocampus is one of the brain structures that regulate stress, emotions, memory and cognitive functions. Moreover, hippocampus is also responsible for adult neurogenesis. It has been shown previously that tame foxes have decreased stress-response and increased adult neurogenesis in hippocampus [1,2]. Stress-response and neurogenesis may be interrelated [3,4]. However, when searching for the molecular mechanisms of these processes and their interrelation in hippocampus, it is important to consider that dorsal and ventral hippocampus are functionally distinct structures with its longitudinal axis being organized along a gradient [5,6]. Particularly, it is often supposed that ventral hippocampus is responsible for stress and emotions, while dorsal hippocampus plays its role in cognitive functions and memory.

Aim

The aim of this study was to test whether hippocampal longitudinal differences by gene expression were present in red foxes (Vulpes vulpes) and to compare the expression of genes between tame and aggressive foxes.

Methods

For the analysis dorsal and ventral hippocampus tissues of male foxes (7-8 months old) were used. Gene expression was evaluated using RT-qPCR analysis. CANX gene was used as a reference gene. Mann-Whitney non-parametric test was used for pair-wise comparison.

Results

Of the genes studied, the gene expression in dorsal hippocampus is increased relatively to ventral hippocampus for *CYP26B1*, *CADM2* and *KCND2* genes. Conversely, in ventral hippocampus the gene expression in increased relatively to dorsal hippocampus for *NTS*, *NR2F2*, *ADRA1A*, *TRHR*, and *CPNE2* genes (all cases $p<0.001$). No significant difference was found for *KCND3* gene.

In addition, the expression of *CYP26B1* gene was demonstrated to be increased significantly in the dorsal hippocampus in tame foxes relatively to aggressive ones ($p<0.01$).

The data obtained conform to the results shown for rodents except for the two genes: *NTS* gene expression is increased in ventral hippocampus in foxes but in dorsal hippocampus in rats [7] and *KCND3* gene expression difference is demonstrated in mice but not in foxes [8].

CYP26B1 is an enzyme responsible for degradation of retinoic acid, which is one of the important participants in the neurogenesis process. The found difference in *CYP26B1* gene expression in dorsal hippocampus indicates the role of retinoic acid pathway in the mechanism of elevation of adult neurogenesis in tame foxes. However, the absence of this difference in the ventral hippocampus indicates, perhaps, more complex causes of changes in neurogenesis.

Acknowledgment

Supported by the RSF (19-74-10041).

References

[1] Trut L. et al. (2009) Animal evolution during domestication: the domesticated fox as a model. Bioessays. 31(3):349–360.

[2] Huang S. et al. (2015) Selection for tameness, a key behavioral trait of domestication, increases adult hippocampal neurogenesis in foxes. Hippocampus. 25(8):963–975.

[3] Sánchez-Vidaña D. I. et al. (2016) Repeated treatment with oxytocin promotes hippocampal cell proliferation, dendritic maturation and affects socio-emotional behavior. Neuroscience. 333:65–77.

[4] Song J. et al. (2012) Modification of hippocampal circuitry by adult neurogenesis. Developmental Neurobiology. 72:1032–1043.

[5] Fanselow M. S., Dong H. W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? Neuron. 65(1):7-19.

[6] Strange B. A. et al. (2014). Functional organization of the hippocampal longitudinal axis. Nature Reviews Neuroscience. 15(10):655-669.

[7] Lee A. R. et al. (2017). Dorsal and ventral hippocampus differentiate in functional pathways and differentially associate with neurological disease-related genes during postnatal development. Frontiers in molecular neuroscience. 10:331.

[8] Cembrowski M. S. et al. (2016). Hipposeq: a comprehensive RNA-seq database of gene expression in hippocampal principal neurons. Elife. 5:e14997.

DOI 10.18699/BGRS/SB-2020-232

Rapidly evolving SNPs feature highly significant trait associations in GWAS SNP hotspots

Roman Babenko
Laboratory of Neuropathology Modeling,
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
babe-roman@yandex.ru

Anton Zhuravlev
Laboratory of Eurasian paleogenetics,
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
zhuravlev@bionet.nsc.ru

***Abstract* — As previously observed, highly significant GWAS related SNPs are regularly clustered with other associated SNPs, implying haplotype mediated mode of association. We compiled a set of enriched GWAS SNPs 'hotspot' regions and worked out a strategy to elucidate population specific haplotype spectra. We present several regions in demonstrating specific effects of these haplotypes. In particular, there is an elevated genetic drift rate in allele frequencies. Asian populations demonstrate the highest homogeneity in haplotype variation, while European populations often fix a unique allele not present in other populations, and usually maintaining deleterious effect, at least considering BMI trait. Also, the regions are non-randomly resided in open chromatin nuclear compartments, and manifest Hi-C interactions in a range of cases.**

Keywords* — *GWAS, SNP, association, obesity, selective sweep, continental populations, 1000G

Introduction

The need for GWAS SNPs annotation remains an acute task in the course of elucidating the casual variants invoking disease pathology and trait manifestation. Many tools have been created for that task: GwasRap, FUMA, SNAP (Nishizaki, Boyle, 2017 [1]). Once the majority of GWAS SNPs reside in non-coding regions, the major approach for elucidating SNPs functional impact on trait currently results in employing many data repositories, such as DNA variation sources based on 1000 Genomes Project, International HapMap Project, annotations of functional elements [2], and conservation information derived from multiple species alignments [2,3]. Chromatin state markers (DNAse accessibility profiles, histone marks landscape, Hi-C maps, etc) are employed by such programs as FunciSNP, Haploreg, GWAS3D [1].

One of the crucial factor of disease related locus is its rapid evolution rate including selection sweeps [4]. It can be assessed by tracing the locus structure dynamics in the course of evolution of modern continental supergroups (EAS, SAS, EUR, AFR). For example, Hapmap project delineates the haploblock structure for each particular supergroup [3]. But the methods for qualitative/quantitative comparison of these cannot be immediately found. We implemented a pipeline to elucidate the most drastic differences between supergroups haploblock structures based on automated selection of the most discriminative GWAS SNPS (evoTAG SNPs) for these. Consequently, based on haploblock dynamics we were able to reconstruct evolutionary events underlining the current alleles structure in the locus.

Research

We constructed the pipeline for extracting the maximal population specific variance SNPs within GWAS hotspots. Based on the methodology we present the analysis of 7 GWAS SNPs hotspot loci with quite rapid fixation of the disease associated SNPs. The proposed method may help in elucidating the casual SNPs, as well as it underscores high disease risk populations across supergroups.

Conclusion

Intriguingly, but expectedly, when we applied the pipeline to the region of chromosome 1 we found that the most informative SNPs from 455 total in the region are 6 GWAS annotated. It underscores the adequacy and utility of the method

ACKNOWLEDGMENT

The work was supported by Russian Foundation for Basic Research grant 19-2015-2099

REFERENCES

[1] Nishizaki SS, Boyle AP. Mining the unknown: Assigning Function to Noncoding Single Nucleotide Polymorphisms. Trends Genet. 2017 Jan;33(1):34-45. doi:10.1016/j.tig.2016.10.008.

[2] Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, Myers Z, Sud P, Jou J, Lin K, Baymuradov UK, Graham K, Litton C, Miyasato SR, Strattan JS, Jolanki O, Lee JW, Tanaka FY, Adenekan P, O'Neill E, Cherry JM. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. 2020;48(D1):D882-D889. doi: 10.1093/nar/gkz1062

[3] 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature. 2010 Oct 28;467(7319):1061-73. doi: 10.1038/nature09534.

[4] Refoyo-Martínez A, da Fonseca RR, Halldórsdóttir K, Árnason E, Mailund T, Racimo F. Identifying loci under positive selection in complex population histories. Genome Res. 2019:1506-1520. doi: 10.1101/gr.246777.118.

DOI 10.18699/BGRS/SB-2020-233

Comparative genomic analysis of *Yersinia pestis* strains of the philogenetic branch 2.MED4

A.N. Balykova
Russian Research Anti-Plague Institute "Microbe", Federal Service for Surveillance in the Sphere of Consumers Rights Protection and Human Welfare
Saratov, Russia
alinabalnik@gmail.com

Zh.V. Al'khova
Russian Research Anti-Plague Institute "Microbe", Federal Service for Surveillance in the Sphere of Consumers Rights Protection and Human Welfare
Saratov, Russia

G.A. Eroshenko
Russian Research Anti-Plague Institute "Microbe", Federal Service for Surveillance in the Sphere of Consumers Rights Protection and Human Welfare
Saratov, Russia

Keywords — Yersinia pestis, strains of medieval biovar, comparative SNP-analysis, SNP-genotyping, phylogenetic analysis

Motivation and Aim

Highly virulent gram-negative bacterium *Yersinia pestis*, the causative agent of plague, is one of the most outstanding example of pathogens that caused huge damage to the health of mankind. Our molecular investigations showed that strains of medieval biovar (phylogenetic branch 2.MED) are the most common in the natural foci of plague with different geographical landscapes and occupy about 80 % of the territory of the natural foci of CIS [1].

In the 18th-19th centuries, numerous outbreaks of plague with high mortality rates were registered in the Caspian-Sea region, Russia. Our research prove that since 1912, outbreaks in the Northern Caspian-Sea region were caused by strains of medieval biovar. We identified 3 *Y. pestis* strains, which were isolated in this region and formed the previously unidentified phylogenetic branch 2.MED4. The strains of branch 2.MED4 have high virulence and epidemic significance and can cause plague in animals and humans [2, 3].

This information determines the importance of further research of 2.MED4 strains. In this regard, it is important to establish their place in the evolutionary scheme of medieval biovar, to determine their role in plague outbreaks in the Caspian-Sea region, and to assess the possibility of their return to these territories. The aim of this study was to identify the specific molecular-genetic features of *Y. pestis* strains of the phylogenetic branch 2.MED4 and to develop a method for their identification by PCR.

Methods and Algorithms

High-performance sequencing was carried out using Ion PGM system (Life technologies). Ion Torrent Suite software 3.4.2 and Newbler gsAssembler 2.6 were applied for data processing. For compara-tive SNP-analysis and phylogenetic reconstruction bioinformatics programs (Wombac 2.0, jModelTest, PAUP 4.0, PhyML 3.1, MEGA X, FigTree 1.4.3.) were used. Fragment sequencing was performed using the ABI PRISM 3500XL platform (Applied Biosystems, USA).

Results

We performed comparative SNP-analysis of 3 *Y. pestis* stains of branch 2.MED4. We found 10 SNPs, which specific to this branch and developed a method for SNP-genotyping of these strains by PCR. We explored 26 *Y. pestis* strains from the plague foci of the Caspian-Sea region by PCR and found and found 5 more *Y. pestis* strains of branch 2.MED4.

We have sequenced and carried out complex analysis of molecular-genetic and phenotypic properties of 5 *Y. pestis* strains of the previously unidentified phylogenetic branch 2.MED4.

According to the complex analysis of phenotypic and genetic properties, all strains are typical of medieval biovar of *Y. pestis*. According to PCR-/SNP-genotyping and phylogenetic analysis, all of 5 *Y. pestis* strains belong to phylogenetic branch 2.MED4, as well as 3 previously identified strains.

Our results may be useful for a retrospective analysis of plague outbreaks and for identifying patterns of evolution of medieval biovar in the Caspian-Sea region in the 18th-20th centuries.

Conclusion

We have sequenced and analyzed 8 *Y. pestis* strains of the previously unidentified phylogenetic branch 2.MED4. We found 10 marker SNPs, specific to this branch and developed a method for SNP-genotyping of these strains by PCR. The data obtained can be used to improve the methods of molecular-genetic identification of *Y. pestis* and to determine patterns of evolution and distribution of medieval biovar in the natural plague foci of the Caspian-Sea region.

References

[1] Kutyrev V. V. et al. (2018). Phylogeny and classification of Yersinia pestis through the lens of strains from the plague foci of Commonwealth of Independent States. Front. Microbiol. 9:1106. DOI: 10.3389/fmicb.2018.01106

[2] Eroshenko G. A. et al. (2019). Circulation of Yersinia pestis in the Volga-Ural Sandy fo-cus: spatiotemporal analysis. Problems of Particularly Dangerous Infections. 2019;3:51–57. DOI: 10.21055/0370-1069-2019-3-51-57.

[3] Eroshenko G.A., Popov N.V., Al'khova Zh.V., Balykova A.N., Kukleva L.M., Kutyrev V.V. (2019). Phylogenetic analysis of Yersinia pestis strains of medieval biovar, isolated in Pre-caspian North-Western Steppe plague focus in the XX century. Problems of Particularly Danger-ous Infections. 2019; 2:55–61. DOI: 10.21055/0370-1069-2019-2-55-61.

DOI 10.18699/BGRS/SB-2020-234

Effect of overexpression of the 5-HT7 receptor gene on behavior and brain serotonin system in ASC mice with predisposition to depressive-like behavior

Baraboshkina I.A.
ICG SB RAS, Novosibirsk, Russia
irina.10.24@yandex.ru

Bazovkina D.V.
ICG SB RAS, Novosibirsk, Russia
daryabazovkina@gmail.com

Ilchibaeva T.V.
ICG SB RAS, Novosibirsk, Russia
rbicehok@mail.ru

Antonov E.V.
ICG SB RAS, Novosibirsk, Russia
a.yegor.v@gmail.com

Kulikova E.A.
ICG SB RAS, Novosibirsk, Russia
kulikova.elisa@gmail.com

Naumenko V.S.
ICG SB RAS, Novosibirsk, Russia
naumenko2002@gmail.com

Abstract — **The serotonin (5-HT) system of the brain plays an important role in controlling various behaviors due to the wide variety of serotonin receptors. The 5-HT7 receptor is of great interest because it is involved in the pathogenesis of depressive disorders. Mice of ASC (Antidepressant Sensitive Catalepsy) line with genetic predisposition to depressive-like behavior were obtained in the Laboratory of Behavioral Neurogenomics of ICG SB RAS (Novosibirsk). In this work, the effect of adenoassociated virus (AAV)-mediated overexpression of the 5-HT7 receptor gene in the midbrain on the behavior and brain serotonin system in ASC mice was comprehensively studied. The ASC mice with overexpression of 5-HT7 receptor gene showed the decrease in depressive-like behavior in the forced swim test compared to control group. The introduction of a vector construct with the 5-HT7 receptor gene affected the expression of the gene encoding the 5-HT7 receptor itself, only in the midbrain. The overexpression of 5-HT7 receptor gene in midbrain led to an increase in the serotonin metabolism index in the cortex, hippocampus, and midbrain of mice, without affecting serotonin levels. The results indicate the effect of overexpression of the 5-HT7 receptor gene on the behavior and serotonin system of the brain in ASC mice with a genetic predisposition to depressive-like behavior.**

Keywords *— serotonin brain system; 5-HT7 receptor; mice; depressive-like behavior; vector construct; overexpression; gene*

Introduction

Currently, an increasing number of people suffer from various psychopathologies. According to World Health Organization, about 25% of the global population suffer from depressive disorders. Thus, at the moment, the problem of finding therapeutic agents and antidepressants for the treatment of diseases of this type is very urgent. It is known that currently existing antidepressants do not have high enough effectiveness, and also have a number of side effects. Thus, when creating new medicines in this direction, first of all, the task is to increase their effectiveness and reduce the possible side effects of their use.

The serotonin system of the brain plays an important role in controlling various behaviors, including pathological ones, due to the large number of types of serotonin receptors and their wide distribution in the brain [1]. The 5-HT1A receptor is involved in mechanisms of response to stress, depression, anxiety, and aggression. Also, 5-HT1A receptors located presynaptically regulate the functional activity of the serotonin system of the brain by the mechanism of negative feedback [2]. 5-HT2A receptors are involved in the mechanisms of development of a number of mental disorders, including depression and schizophrenia, and are also an important link in the mechanisms of action of antidepressant agents [3].

The 5-HT7 receptor is of great interest and has been studied relatively recently. The 5-HT7 receptor is associated with various physiological and pathophysiological processes, including circadian rhythms, thermoregulation, and depression [4]. It was found that it can influence the regulation of the serotonin system of the brain, so its study is a promising direction today. Modern literature suggests that there is a dependence of the presynaptic 5-HT1A receptor on 5-HT7 receptors, which, forming dimer complexes with it, lead to functional inactivation of 5-HT1A receptor [2]. This is a completely new and crucial role of the 5-HT7 receptor and its interaction with the 5-HT1A receptor allows us to take a new look at the problem of the mechanism of depression and the action of antidepressants of the serotonin reuptake inhibitor group.

It is important to select adequate animal models for the study of psychopathological conditions. Mice of ASC (Antidepressant Sensitive Catalepsy) line obtained in the Institute of Cytology and Genetics SB RAS (Novosibirsk) demonstrate pronounced depressive-like behavior and dysfunctions of immune system, which is also observed in depressed patients. Thus, ASC mice meet all the necessary criteria for a laboratory model of depression.

Manipulating the expression of certain genes is a promising area of study and treatment of psychopathologies. This approach is carried out using vectors that ensure delivery of genetic material to the cells of the body and its expression. Recombinant adenoassociated viruses are one of the most promising delivery vectors in gene therapy and neurobiology. Their most important advantages are: non-pathogenicity, low immunogenicity, tropicity to most cells and tissues, as well as high efficiency of transduction and long time of gene expression in vivo without embedding in the host genome.

Purpose of work was to study the effect of overexpression of the 5-HT7 receptor gene in the midbrain on the behavior and brain serotonin system in mice with a predisposition to depressive-like behavior.

In accordance with this, the following tasks were formulated:

1) to investigate the effect of overexpression of the 5-HT7 receptor gene on behavior in the open field and forced swim tests in ASC mice;

2) to study the effect of overexpression of the 5-HT7 receptor gene on the level of mRNA genes encoding key

elements of the serotonin system in various brain structures of ASC mice;

3) to evaluate the effect of overexpression of the 5-HT7 receptor gene on level of 5-HT and and their metabolite 5-HIAA in different brain structures of ASC mice.

Materials and Methods

Experiments were conducted on adult three-month-old male mice of the ASC line (Antidepressant sensitive catalepsy), (n=30) with genetic predisposition to depressive-like behavior. This line was obtained in the Laboratory of Behavioral Neurogenomics of Institute of Cytology and Genetics (Novosibirsk) by selection for high predisposition to catalepsy from the population of backcrosses between the cataleptic CBA and noncataleptic AKR strains.

The animals were divided into 3 groups: the first group of mice (Control) received a buffer into the midbrain (a solvent for the vector construct), the second group of mice (AAV_HTR7_eGFP) received the actual construct with the target gene, and the third group of mice (AAV_eGFP) received the vector construct inserted without target gene.

Behavior was evaluated in the open field and forced swim tests six weeks after the introduction of vector constructs.

The mRNA levels of genes were determined using real-Time PCR. The level of brain serotonin and their metabolite 5-HIAA were measured using HPLC (high-performance liquid chromatography).

Data were expressed as average values with an error of average (x± Sx). All groups were compared using single-factor analysis of variance (one-way ANOVA) followed by multiple comparison using the Fisher post-hoc test.

Results

In mice that were injected with a construct with the target gene, an increase in the mRNA level of 5-HT7 receptor gene in the midbrain was detected, compared to animals of the AAV_eGFP ($p<0.001$) and buffer ($p<0.001$) groups. These data suggest that the viral construct introduced to the animal effectively serves as a vector and the plasmid with the target gene is actually transcribed in neurons.

One-way ANOVA showed the effect of 5-HT7 gene overexpression in the forced swim test ($p<0.05$). There was a significant increase in the mobility of mice from the experimental group compared to the control group ($p<0.01$) and the AAV_eGFP group ($p<0.05$), which indicates a decrease in level of depressive-like behavior in ASC mice with 5-HT7 gene overexpression.

The brain 5-HT level was not differed between all groups. In mice that received a viral construct with the target gene, an increase in the level of 5-HIAA in the cortex was detected, compared to animals from the AAV_eGFP group ($p<0.01$). Also, the significant increase in the 5-HIAA level in hippocampus was found in animals with 5-HT7 gene overexpression compared to the control group ($p<0.001$) and the AAV_eGFP group ($p<0.001$). Moreover, the animals of the experimental group showed an increase in the serotonin metabolism index, defined as the ratio of 5-HIAA/serotonin, compared to the control group ($p<0.01$) and the AAV_eGFP group ($p<0.05$) in all the studied brain structures.

The data indicate that the 5-HT7 gene overexpression in midbrain led to more intense functioning of the brain serotonin system and this was accompanied by decrease in depressive-like behavior in ASC mice.

Acknowledgment

The study was supportet by Russian Science Foundation (grant number 19-15-00025).

References

[1] J. G. Hensler (2012) Serotonin, In: S.T. Brady editor: Basic neurochemistry principles of molecular, cellular, and medical neurobiology. 8th ed. Academic. Waltham. MA.: 300–322.

[2] N. K. Popova, E. G. Ponimaskin, V. S. Naumenko (2015) Cross-talk between 5-HT1A and 5-HT7 receptors: role in the autoregulation of the brain serotonin system and in mechanism of antidepressants action. Russian Journal of Physiology. 101(11): 1270—1278.

[3] B. Dean (2003) The cortical serotonin 2A receptor and the pathology of schizophrenia: a likely accomplice. J. Neurochem. 85: 1–13.

[4] P. B. Hedlund (2009) The 5-HT7 receptor and disorders of the nervous system: an overview. Psychopharmacology (Berl). 206(3): 345–354.

DOI 10.18699/BGRS/SB-2020-235

Genetic structure of breeding pigs of Large White, bred in Russia

Getmantseva L., Bakoev S., Kostyunina O., Traspov A., Prytkov Yu., Bakoev N.
Federal Science Center for Animal Husbandry named after Academy Member L.K. Ernst
Dubrovica, Russia
ilonaluba@mail.ru

***Abstract* — This paper presents a study of the genetic structure of breeding pigs of Russia, belonged to a Large White.**

Keywords — genetic structure, pig, Large White, Porcine HD

In recent decades, the global commercial pig industry has led to significant business consolidation. Mergers and acquisitions of small genetic centers have resulted in a limited number of remaining international breeding companies [1]. Consequently, the breeding lines belonging to these companies also experienced a high degree of consolidation. In the Russian Federation, the Large White represent 54.1% of pigs [2]. The aim of the work is to study the genetic structure of the breeding pigs of the Russian Federation, which belong in the Large White breed.

The main companies supplying genetic material to the world market of pigs are Dan AvI (Denmark), Genesus (Canada), PIC (Great Britain), Hypor (Holland), Nucleus (France), Hermitage Genetics (Ireland). In our work, we investigated pigs from three Selection Genetic Centers (Farm) of Russia. The population of these Farms formed in 2012-2014 by imported genetics of different companies and further improved in accordance with its own strategy. The pigs from Farm1 came from Nucleus (France), Farm 2 - Dan AvI (Denmark) and Farm 3 - Hermitage Genetics (Ireland).

The study was on Large White pigs (n=53) from three Farms of Russia: Farm 1 (n=25), Farm 2 (n=16) and Farm 3 (n=12). Genome DNA extracted from sample (ear plucks) by a set of reagents DNA-Extra-2 (OOO Syntol, Russia). Genotyping of animals was performed using GeneSeek GGP Porcine HD (San Diego, USA), including 68540 SNPs. The functions of the NAM package [3] used to impute missed values using the machine-learning algorithm Random Forest. As a result, 46845 SNPs left for further analysis. The Principal Component Analysis (PCA) method in Plink v.1.9 [4], the ggplot2 package [5], Admixture 1.3 [6] and Ancestry Painter [7] were using the estimation of the population structure.

The results of the study showed a clear division into three clusters of Large White breed (Figure 1). Each cluster corresponded to a specific Farm. Can be clearly identified the genetic structure of the companies: Nucleus (France), Dan AvI (Denmark) and Hermitage Genetics (Ireland) from the data obtain. Among the groups studied, the most consolidated population is determined in Farm 3 based on Irish genetics. In Farm 1 and 2, along with the prevalence of French and Danish genetics, respectively, we can note the presence of a fraction of the genome of different origin, which is probably due to breeding work aimed at improving the breeding stock of pigs and possible formation of their own genetic structure.

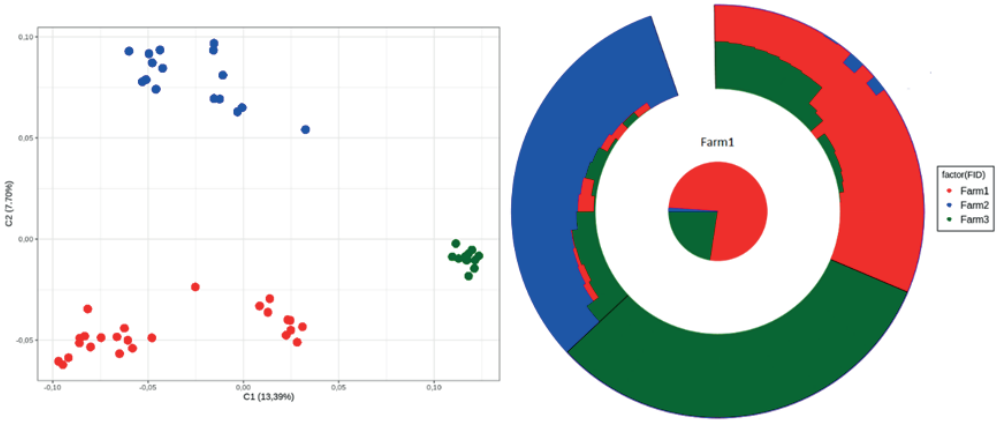


Fig. 1. Genetic structure of breeding pigs of Large White, bred in Russia

References

[1] I. Hulsegge, M. Calus, R. Hoving-Bolink et al. Impact of merging commercial breeding lines on the genetic diversity of Landrace pigs. Genet Sel Evol 51, 60 (2019).

[2] Yearbook on pedigree work in pig husbandry on the farms of the Russian Federation M.: Publishing house FGNU VNIIplem, 2018. - 192 p.

[3] A. Xavier, S. Xu,W.M. Muir, K.M. Rainey NAM: association studies in multiple populations Bioinformatics. 2015. 31(23): 3862-3864.

[4] C.C. Chang, C.C. Chow, L.C. Tellier et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015. 25;4:7.

[5] H. Wickham. ggplot2: Elegant Graphics for Data Analysis. 2009.

[6] D.H. Alexander, J. Novembre, K. Lange. Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19:1655\u20131664, 2009.

[7] F. Qidi. The AncestryPainter software2 is available at http://www.picb.ac.cn/PGG/resource.php

The work is funded by the RSF grant No. 19-16-00109.

DOI 10.18699/BGRS/SB-2020-236

Population genetic variation of serotonin transporter gene (SLC6A4), associated with neurophysiological development

Shyamala Hande
Melaka Manipal Medical College
Manipal, India
shyamala.hande@manipal.edu

Komalroop Kaur Sahote
Melaka Manipal Medical College
Manipal, India

Nirosha Dev
Melaka Manipal Medical College
Manipal, India

Ting Pei Erl
Melaka Manipal Medical College
Manipal, India

Kovindraam Ramakrishna
Melaka Manipal Medical College
Manipal, India

Renuka Ravidhran
Yenepoya Research Centre, Yenepoya
(Deemed to be University), Mangalore
Karnataka, India

Swathy M. Krishna
Yenepoya Research Centre, Yenepoya
(Deemed to be University), Mangalore
Karnataka, India

Ranajit Das
Melaka Manipal Medical College
Manipal, India

Abstract — The serotonergic pathway, including the neurotransmitter serotonin and its cognate transporter protein (5-HTT), is known to influence a wide range of individual behavioral traits and cognitive abilities. The serotonin transporter 5-HTT is encoded by a single gene SLC6A4. Polymorphisms in *SLC6A4* has been associated with a wide variety of neurological and psychiatric disorders including increased risk of post-traumatic stress disorder, higher likelihood for depression, obsessive-compulsive disorder (OCD), increased hostility and criminal behavior. All diseases associated with *SLC6A4* are complex disorders, caused by the interplay of various genetic and environmental factors. It has been shown that genes associated with complex diseases exhibit strong signatures of purifying selection compared to non-disease-causing genes. Purifying selection, in several cases, is associated with background selection, where alongside the deleterious alleles, the linked neutral (or even slightly beneficial) alleles are purged away from the genome. The majority of the genetic variations are rare, arose recently, and are highly ancestry-specific. In this project we aimed to investigate the population genetic variation of the serotonin transporter gene (*SLC6A4*), focusing on the Single Nucleotide Polymorphisms (SNPs). We further aimed to assess whether any modern-day human population show strong signature of natural selection for *SLC6A4* and has significantly differentiated from the rest of the world for this gene, employing 2504 individuals around the globe available in 1000 Genome project Phase III data.

Keywords — Polymorphism in SLC6A4, Ancestry specific variation of natural selection, Signatures of natural selection

Methods

Data collection The lists of SNPs, present in *SLC6A4*, were curated from SNPedia and UCSC genome browser. 25 *SLC6A4* SNPs, associated with several neuropsychological disorders, were obtained from SNPedia and all 147 annotated SNPs of *SLC6A4* were obtained from UCSC genome browser. Whole genome data in the form of VCF files were obtained for 2504 individuals around the world from 1000 Genomes Project database. Populations were broadly classified into five ethnic groups: Americans (AMR), Europeans (EUR), Africans (AFR), East Asians (EAS) and South Asians (SAS). File conversions and manipulations were performed using VCFtools v.0.1.13, PLINK v1.9 and EIGENSOFT v7.2.1. *Clustering of populations based on SLC6A4 SNPs* The fine population structure of the modern-day humans obtained from 1000 Genomes, based on SLC6A4 SNPs was delineated using Principal Component Analysis (PCA) implemented in PLINK v1.9 using --pca command. The two most informative PCs are discussed and plotted in R v3.5.1. *Detection of signature of natural selection on SLC6A4* Cross population extended haplotype homozygosity (XP-EHH) algorithm (http://hgdp.uchicago.edu/Software/) was used for detecting the genetic imprint of recent positive selection by analyzing long-range haplotypes in global human populations obtained from 1000 Genomes. Due to unavailability of updated genetic map information pertaining to the distance between two loci (in centimorgan, cM) and recombination rate for each locus (cM/Mb) for Human Genome Assembly GRCh38, we could not employ all SNPs for XP-EHH analysis. Out of the 25 disease-relevant SNPs and all annotated 147 SNPs of SLC6A4, 20 and 38 SNPs respectively could be used for XP-EHH analysis. *Identical by Descent (IBD) analysis* Populations that have higher IBD sharing for a particular DNA segment can be considered to be under strong and resent selection pressure for that genomic area. Here, IBD sharing was calculated using Pi-Hat statistic implemented in PLINK v.1.9 employing all 147 SNPs of SLC6A4.

Results

Principal Component Analysis (PCA) PCA performed on all 147 SNPs of *SLC6A4* depicted clear and more distinct clustering between Africans (red dots) and the rest of the world along PC1. Additionally, discernible separation among non-Africans was observed along PC2. East Asians (green dots) were found to be present in distinct cluster from other non-African populations. *Detection of signature of natural selection on SLC6A4* For disease-relevant SNPs of SLC6A4, long areas of positive XP-EHH scores was observed among the non-African populations suggesting a recent selective sweep. The highest XP-EHH scores were found among East Asians indicating SLC6A4 is under strong selection among people belonging to this ethnic group. Similar to the disease-associated SNPs of SLC6A4, the rest of the SNPs were also found to be under strong selection especially among East Asian populations with significantly high positive XP-EHH scores compared to Africans. Overall, XP-EHH analysis indicated that SLC6A4 has likely undergone a strong selective sweep among the

East Asians in the recent past. *Identical by Descent (IBD) analysis* Supporting XP-EHH results, the maximum number of individuals with the highest IBD sharing (Pi-Hat = 1) belong to East Asians (53.4%), distantly followed by the native Americans (18.7%). The IBD analysis corroborated that SLC6A4 has likely undergone a strong recent selective sweep among the East Asians.

Discussion

In this study we aimed to assess the ancestry specific variation of signatures of natural selection on SLC6A4. To this end, we obtained the whole genome data of 2504 individuals around the world available in 1000 Genomes Project database. We grouped these 2504 individuals into five broad ancestry categories: Native Americans (AMR), Europeans (EUR), Africans (AFR), East Asians (EAS) and South Asians (SAS). We found that SLC6A4 has likely undergone a strong selective sweep among East Asians in the recent past. East Asians also had the highest IBD sharing for SLC6A4, further indicating to be under strong and resent selection pressure for this genomic area. Recently, Lam et al. identified five SNPs in *SLC6A4* (rs140700, rs4251417, rs6354, rs25528, rs25531) that are likely associated with depression (Lam et al., 2018). However, likely because of the ancestry of the participants (French), they only found nominally significant association between these variants and depression. Our XP-EHH results indicated minimal selection pressure on populations of European ancestry compared to East Asians. The XP-EHH scores for the candidate SNPs for depression (rs140700, rs4251417, rs6354, rs25528) was found to be significantly higher among East Asians (1.49, 0.61, 1.42 and 1.37 respectively) compared to Europeans (0.69, -0.17, 0.59 and 0.54 respectively). Our results indicate that the candidate SNPs for depression are under strong evolutionary constraint governed by purifying selection among East Asians. The evolutionary constraint on *SLC6A4* is highly ancestry specific and likely appeared after the divergence between East Asians and Europeans ~41,000 years ago. Despite not generating any additional genotyping data, our study can help to extend the current knowledge about the association between *SLC6A4* and neurophysiological disorders.

DOI 10.18699/BGRS/SB-2020-237

First evidences of direct interaction between 5-HT2A and TrkB receptors

Tatiana Ilchibaeva
ICG SB RAS, Novosibirsk, Russia
rbicehok@mail.ru

Anton Tsybko
ICG SB RAS, Novosibirsk, Russia
antoncybko@mail.ru

Vladimir Naumenko
ICG SB RAS, Novosibirsk, Russia
naumenko2002@mail.ru

***Abstract* — The existence of 5-HT2A-TrkB heteroreceptor complexes both *in vitro* and *in vivo* was shown for the first time. Also was found that the chronic activation of 5-HT2A receptor with agonists has decreased the total as well as membrane TrkB protein level indicating downregulation of TrkB receptors.**

Keywords* — *5-HT2A, TrkB, agonists, DOI, TCB-2, 25CN-NBOH, co-immunoprecipitation, expression, heterodimerization

Motivation and aim

Motivation

Cross-talk between the brain-derived neurotrophic factor (BDNF) and serotonin 5-HT2A receptors was demonstrated in the number of studies [1]. At the same time an exact mechanism of modulation of BDNF system through 5-HT2A receptors is still unknown. BDNF performs its functions using tropomyosin receptor kinase B (TrkB). And these receptors may be one of the possible targets. The existence of a heteroreceptor complex between the 5-HT1A and FGFR1 (FGFR1 belongs to the receptor tyrosine kinase superfamily) receptors [2] suggest a possible similar complex between 5-HT2A and TrkB receptors.

Aim

The aim of present study was to investigate the presence of 5-HT2A-TrkB heteroreceptor complexes in cell cultures as well as mouse brain tissues.

Methods

Adult males of C57Bl/6 mice were treated with mixed 5-HT2A/2C agonist DOI or selective 5-HT2A agonists TCB-2 or 25CN-NBOH (1 mg/kg, i.p., 14 days). Expression of TrkB receptors was assessed by real-time RT-PCR and Western blot analysis in the frontal cortex, hippocampus, midbrain and striatum. Murine 5-HT2A cDNA was cloned into the pcDNA3.1(+) donor plasmid (Invitrogen) with HA-tag. TrkB_mEGFP was a gift from Ryohei Yasuda (Addgene plasmid #83952; http://n2t.net/addgene:83952; RRID: Addgene_83952). Mouse N1E-115 neuroblastoma cells from the American Type Culture collection (ATCC) were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin at 37 °C under 5% CO2. Co-immunoprecipitation and immunoblotting in N1E-115 cells co-expressing HA-tag_HT2A and TrkB_mEGFP receptors were performed as described [3]. Co-immunoprecipitation from brain homogenates was performed according to [4] with modification.

Results

TrkB mRNA level was reduced after 25CN-NBOH in the midbrain and was elevated after TCB-2 in the striatum. All of used drugs significantly decreased the TrkB total protein level in the midbrain and membrane TrkB protein level in the midbrain and striatum as well. This looks like desensitization of TrkB receptors. Using co-transfection of NIE-115 cells with 5-HT2A and TrkB receptors followed by co-immunoprecipitation the presence of heteroreceptor complexes was shown. Later using co-immunoprecipitation we have found the presence of 5-HT2A-TrkB heteroreceptor complexes in the midbrain and striatum of intact mice that was agreement with the data from chronic 5-HT2A receptor activation. Thus, for the first time the heterodimerization between 5-HT2A and TrkB receptors was found. This phenomenon might underlie the downregulation of TrkB receptors after chronic activation of 5-HT2A receptor with agonists.

Acknowledgment

The work supported by the RFBR (#20-04-00253). The cost of animal maintenance was supported by the basic research project No 0324-2019-0041-C-01.

References

[1] Jaggar M, Vaidya VA (2018) 5-HT_{2A} Receptors and BDNF Regulation: Implications for Psychopathology. In: Guiard BP, Di Giovanni G (eds) 5-HT_{2A} Receptors in the Central Nervous System, The Receptors 32. Humana Press, pp 395-438.

[2] Borroto-Escuela D.O., Tarakanov A.O., Fuxe K. (2016) FGFR1-5-HT1A Heteroreceptor Complexes: Implications for Understanding and Treating Major Depression. Trends Neurosci. 39:5–15.

[3] Kobe, F., Guseva, D., Jensen, T. P., Wirth, A., Renner, U., Hess, D., Müller, M., Medrihan, L., Zhang, W., Zhang, M., et al. (2012). 5-HT7R/G12 signaling regulates neuronal morphology and function in an age-dependent manner. J. Neurosci. 32, 2915-2930.

[4] Bijata M, Labus J, Guseva D, Stawarski M, Butzlaff M, Dzwonek J, Schneeberg J, Böhm K, Michaluk P, Rusakov DA, Dityatev A, Wilczyński G, Wlodarczyk J, Ponimaskin E. Synaptic Remodeling Depends on Signaling between Serotonin Receptors and the Extracellular Matrix. Cell Rep. 2017 May 30;19(9):1767-1782.

DOI 10.18699/BGRS/SB-2020-238

Hippocampal overexpression of the cerebral dopamine neurotrophic factor (CDNF) impaired fear memory formation in rats

Tatiana Ilchibaeva
ICG SB RAS, Novosibirsk, Russia
rbicehok@mail.ru

Elizaveta Zolotenkova
ICG SB RAS, Novosibirsk, Russia
blanka.ardeo@yandex.ru

Dmitry Eremin
ICG SB RAS, Novosibirsk, Russia
dima.969696@mail.ru

Anton Tsybko
ICG SB RAS, Novosibirsk, Russia
antoncybko@mail.ru

***Abstract* — For the first time it was found that AAV-mediated CDNF overexpression in the dorsal hippocampus of rats can impaired formation of cued fear memory which was accompanied by increase in CREB1, c-Fos and 5-HT1B receptor mRNA level.**

Keywords* — *CDNF, AAV-mediated gene transduction, fear conditioning, CREB1, c-Fos, 5-HT receptors, mRNA

Motivation and aim

Motivation

CDNF is considered as a protective factor for the brain dopaminergic neurons and has recently entered clinical trials as an agent against Parkinson's disease [1]. However, the role of CDNF in the regulation of various forms of behavior outside locomotor activity remains unclear. There are some evidences on the implication of CDNF in the formation of spatial memory [2]. It is necessary to establish whether CDNF is involved in the regulation of other forms of memory, e.g. fear memory.

Aim

The aim of present study was to investigate the effect of hippocampal CDNF overexpression on the cued fear memory and neuroplasticity related genes expression in rats.

Methods

We have constructed the pAAV-Syn(0.5)-CDNF_EGFP plasmid and used AAV-mediated gene delivery to induce overexpression of CDNF in the hippocampal neurons of adult males of Wistar rats. Four weeks after stereotactic delivery of AAV-CDNF into dorsal hippocampus the fear conditioning was performed. On the training day 1 rats were placed in a fear conditioning apparatus for 3 min prior to three presentations of auditory cues (80 dB white noise, 30 s), each co-terminating with a foot shock (1.0 mA, 2 s.) and spaced 60 s apart. After the last shock, the animals remained in the chambers for an additional 3 min. Cued fear memory was assessed by exposing the animals to the same context in the presence of tone presentation but without foot shocks on the day 2, 4 and 6 after training. The time spent by each rat freezing was measured. Expression of neuroplasticity-related genes (*Bdnf, Creb1, c-Fos, Ark, Trkb, Ngfr, Htr1a, Htr1b, Htr2a and Htr7*) was assessed by real-time RT-PCR.

Results

We have found that CDNF overexpression impaired cued fear conditioning that was reflected in significantly low total time of freezing as well as freezing time before, during and after tone presentation on the second, fourth and sixth days after training. It should be noted that freezing in experimental and sham-operated animals was not differ on the first training day. The mRNA levels of *Creb1*, *c-Fos* and *Htr1b* genes were significantly increased in the hippocampus of rats with overexpression of CDNF. mRNA levels of another investigated genes were unchanged. Thus, for the first time we have demonstrated that endogenously overexpressed CDNF can negatively regulate the cued fear memory formation involving key transducers of synaptic plasticity CREB1 and c-Fos. 5-HT system also might be affected by CDNF.

Acknowledgment

The work supported by the RSF (#19-75-00016). The cost of animal maintenance was supported by the basic research project No 0324-2019-0041-C-01.

References

[1] Huttunen H.J., Saarma M. (2019) CDNF Protein Therapy in Parkinson's Disease. Cell Transplant. 28(4):349-366.

[2] Kemppainen S. et al. (2015) Cerebral dopamine neurotrophic factor improves long-term memory in APP/PS1 transgenic mice modeling Alzheimer's disease as well as in wild-type mice. Behav Brain Res. 291:1-11.

DOI 10.18699/BGRS/SB-2020-239

Genomic signals of adaptation in the Northern Ural and Western Siberian populations

Gennady Khvorykh
Institute of Molecular Genetics of the Russian Academy of Sciences, Moscow, Russia
khvorykh@img.ras.ru

Giang Vu
Moscow Polytechnic University, Moscow, Russia
vuminhgiang135@gmail.com

Andrey Khrunin
Institute of Molecular Genetics of the Russian Academy of Sciences, Moscow, Russia
khrunin@img.ras.ru

Abstract — **The evolution of humans was influenced by environmental conditions, particularly when they left Africa. Exploration of the genetic variations in the populations leaving in extreme climates will lead to better understanding of molecular-genetics mechanisms of adaptation. In the current study we investigated the genome-wide data of Khahty, Mansi, and Nenets people for signatures of selection using the population branch statistic test and revealed strong signals located in a cluster of alcohol dehydrogenase 1 genes. We hypothesize that it can be due to the limited diversity of food intake the indigenous people of the Northern Ural and Siberian experience.**

Keywords — adaptation, diet, extreme climates, polygenic selection

Motivation and aim

Motivation

Human populations evolved in the tropics and spread all over the climatic zones after they left Africa. Humans adapted to environmental pressures via phenotypic and genetic mechanisms. Knowing the genetic variants involved in adaptive processes can uncover the targets for producing the beneficial phenotypes [1]. There are habitats with exceptional heat, cold, ultraviolet light, or low ambient oxygen. The adaptation to extreme conditions can be of more apparent manifestation on the genomic level and thus more easily detected.

Exploration of genetic variations in populations became available last decade on unprecedented level with genotyping by DNA-microarrays and high throughput sequencing.

The genotypes of 200 individuals from ten indigenous Siberian populations have been scanned for positive selections with iHS, XP-EHH, and PBS tests [2]. The authors reported the loci associated with genes involved in energy metabolism, vascular smooth muscle contraction, and DNA damage response against UV radiation. The other researches [3] revealed the polygenic selection related to fat metabolism in the exome sequence data from Nganasans and Yakuts that was consistent to adoption to a fat-rich animal diet.

We are interested in studying of the neighboring regions with extreme climate. One of such habitats is the arctic and subarctic Transuralic regions. In the Polar Ural the winter lasts nearly seven months and temperature frequently drops below minus 40 degrees of Celsius.

These regions are home to Udmurts, Mari, Komi, Mansi and Khanty people. The indigenous populations of the Polar Urals are the Nenets and Samoyedic peoples. The genetic variations in these populations are not yet studied appropriately. Their genome-wide data on single nucleotide polymorphisms are quite promising for searching for genomic «footprints» of natural selection driven by extreme conditions.

Aim

The aim of the research was to search for signals of natural selection in the genome-wide microarray data of Khanty, Mansi and Nenets peoples with the use of Population Branch Statistics (PBS). The goal was divided into the following tasks: create a pipeline to estimate the PBS values from the genotypic data given in Plink binary format; validate the pipeline using public research data; count the PBS values for Khanty, Mansi, and Nenets peoples; and interpret the genetic signals of selection in terms of biological processes.

Methods

Following the Cardona et al. [2] we developed our own bioinformatics pipeline to estimate the PBS from the genotypic data obtained with DNA-microarrays. Firstly, we applied this pipeline to the data sets of the above mentioned authors who deposited them into NCBI GEO with accession number GSE55586. We estimated the PBS in 100 kb windows across the genome and found that the top 1% of windows obtained by us to be in good agreement with those of published previously. Then we applied the pipeline to estimate the PBS for Khanty, Mansi, and Nenets peoples. The collection of samples, DNA extraction and genotyping procedures for these populations were described previously [4]. The data consisted of 42 (Khanty), 37 (Mansi) and 41 (Nenets) unrelated individuals. CEU population from HapMap III project [5] and a group of Vietnamese people from the study by Cardona et al. [2] were used as outgroup samples during PBS calculations. The number of common SNPs left consisted of 450467 (Khanty-CEU-Viet), 445231 (Mansi-CEU-Viet), and 434777 (Nenets-CEU-Viet).

For each 100 kb window we identified the maximum PBS value but tightened the outlier detection by choosing top 0.1% of such windows. They were further annotated by intersecting with the genomic coordinate of genes using Ensembl gene annotation GRCh37.87 [6]. The intersection was made with bedtools 2.29.2 [7]. The selected genes were further tested for overrepresentation in MSigDB v.7.1 [8].The sets of genes from Gene Ontology and KEGG were used in the tests.

Results

We developed a pipeline to estimate the PBS values and validated it using the data published previously. The application of this pipeline to Kanty, Mansi, and Nenets populations resulted in 26 top ranked windows all over the genome for each tested population. The analysis in MSigDB revealed the genes that are overrepresented in GO:0042572 process, which relates to retinol metabolism. These are ADH1A, ADH1C, ADH1B, and PLB1 (FDR = 1.65E-04) in the case of Khanty and ADH1A, ADH1C, ADH1B, and

ADH6 (FDR = 7.13E-05), in the case of Nenets. The genes ADH6, ADH1A, ADH1B, and ADH1C intersect two adjacent windows starting at genomic coordinates chr4:100100000 and chr4:100200000, which have the PBS values within 0.1% of the top values.

Retinol is one of the three substances of vitamin A. The vitamin A can be effective antioxidant that prevents or improves disease conditions caused by oxidative stress [9].

The indigenous people of Khanty-Mansi Autonomous Okrug (e.g., khanty) were found to have lower content of vitamin A in comparison with those who arrived into this region [10].The authors of this research believe low level of vitamin A can result from inadequate intake. The variety of food resource is limited. The product reach of vitamin A (butter, milk, eggs, cheese, green and yellow vegetables and fruits) are almost absent in the diet of indigenous people of Khanty and Mansi. The only source of this vitamin for them is fish fat.

Thus one can assume that the people from Northern Ural and Western Siberia are adopted to low level of vitamin A and high PBS values annotated with the genes related to the metabolism of retinol can reflect such adoption on the genetic level.

ACKNOWLEDGMENT

The study was funded by RFBR (Russian Foundation for Basic Research) according to the research project No. 20-04-00824.

REFERENCES

[1] Bigham A.W. (2019) Natural Selection and Adaptation to Extreme Environments: High Latitudes and Altitudes, A Companion to Anthropological Genetics, ed by O'Rourke D.H., JOHN WILEY & Sons, 219.

[2] Cardona A., et al. (2014) Genome-wide analysis of cold adaptation in indigenous Siberian populations. PloS one, 9(5): 1-11.

[3] Hsieh P., et al. (2017) Exome Sequencing Provides Evidence of Polygenic Adaptation to a Fat-Rich Animal Diet in Indigenous Siberian Populations, Molecular Biology and Evolution, 34(11): 2913–2926.

[4] Khrunin A., et al. (2020) Genomic landscape of the signals of positive natural selection in populations of Northern Eurasia: A view from Northern Russia, Plos ONE, 15(2): e0228778.

[5] ftp://ftp.ncbi.nlm.nih.gov/hapmap/phase_3. Accessed at December 1, 2019.

[6] ftp://ftp.ensembl.org/pub/grch37/release-87. Accessed at February 5, 2020.

[7] Quinlan A., et al. (2010) BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics, 26(6): 841–842.

[8] Subramanian A., et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles, Proceedings of the National Academy of Sciences, 02 (43) 15545-15550.

[9] Sinbad O.O., et al. (2019) Vitamins as Antioxidants, J Food Sci Nutr Res 2019; 2 (3): 214-235

[10] Korchin V., et al. (2015) A comparative supplement of vitamin A, E, and C of adults in North Region, The Symbol of Science (in Russian), 12: 212-216.

DOI 10.18699/BGRS/SB-2020-240

MtDNA variability in the field vole (*Microtus agrestis* L., 1761), Arvicolinae, Rodentia) in the Urals and adjacent territories

Maria Krokhaleva
Institute of Plant and Animal Ecology
UB RAS,
Ekaterinburg, Russia
krohaleva_ma@ipae.uran.ru

Lidia Yalkovskaya
Institute of Plant and Animal Ecology
UB RAS,
Ekaterinburg, Russia
lida@ipae.uran.ru

Petr Sibiryakov
Institute of Plant and Animal Ecology
UB RAS,
Ekaterinburg, Russia
sibiryakov@ipae.uran.ru

Evgenia Markova
Institute of Plant and Animal Ecology
UB RAS,
Ekaterinburg, Russia
e.markova@ipae.uran.ru

Aleksandr Borodin
Institute of Plant and Animal Ecology
UB RAS,
Ekaterinburg, Russia
bor@ipae.uran.ru

***Abstract* — The data on mtDNA variability (complete cyt *b* gene) in *M. agrestis* in the Cis-Urals, the Southern, Middle, Northern Urals, and Western Siberia obtained for the first time expand significantly the geography of investigations of the species genetic diversity. Analysis of the genetic variability of the field vole with the inclusion of new data does not contradict present views on the genetic structure of the species but reveals genetic heterogeneity of populations within the Eastern subclade, which has not been considered previously. To clarify the genetic structure of the wide-ranging Eastern subclade of *M. agrestis* (from Norway to Lake Baikal), further studies of the species in Northern Eurasia are necessary.**

Keywords* — *Microtus agrestis, genetic variability, cytochrome b, Northern Eurasia

Motivation and Aim

Motivation

The field vole is a widespread Palearctic species, ranging from Western Europe eastwards to Lake Baikal in southeast Siberia [1]. Due to its wide distribution the species is an excellent model for studying genetic variability. Variation of mtDNA based on cyt *b* data in *M. agrestis* has been relatively well-studied [2-6]. However, there are almost no data on the central part of Northern Eurasia, namely on the Urals and adjacent areas. Geographical location of the Ural region determines the diversity of its climatic conditions and the uniqueness of its fauna, comprising both the European and Asian elements as well as widespread species of mammal.

Aim

The aim of the study was to analyze mtDNA variability in the field vole in the Urals and adjacent territories based on cyt *b* data and to determine the position of *M. agrestis* from the studied territory in the phylogeographic structure of the species.

Methods

Total genomic DNA was isolated by the method of salt extraction from muscle tissue specimens fixed in 96% ethanol. PCR and subsequent sequencing of the cyt *b* was carried out with primers L7 and H6. The chromatograms were analyzed using the BioEdit v. 7.2.0 (4.30.2013) software program. Sequence alignments were carried out in MEGA v. 6. Construction of phylogenetic trees using Bayesian inference (BI) was carried out in MrBayes v. 3.2.2. Assessment of nucleotide diversity was performed in the Arlequin v. 3.1 and DnaSP v. 5.10 software programs.

Results

In order to analyze mtDNA variability, complete cyt b (1140 bp) is sequenced from 33 samples of *M. agrestis* from 13 localities in the Cis-Urals, the Southern, Middle, Northern Urals, and Western Siberia. 23 haplotypes are identified, all of them are original and have not been described previously. For comparative analysis, a total of 373 haplotypes of *M. agrestis* are used including our data and 350 haplotypes from GenBank. *M. arvalis* (n=2), *M. rossiaemeridionalis* (n=2), *M. socialis* (n=2), *M. kirgisorum* (n=2) are used as an out-group.

Based on phylogenetic tree constructed using Bayesian inference (BI) three main clades are recognized: South, Portugal, and Northern clade. The Northern clade is widespread and differentiated into 6 subclades. Newly obtained results correspond with the data from previous studies [4, 6]. All haplotypes sequenced for *M. agrestis* in the Urals and adjacent territories are included in the Eastern subclade of the Northern clade.

The Eastern subclade differentiation has not been previously considered despite the significant area of its distribution (from Norway to the eastern boundary of the species range), including regions with different physical, geographical and environmental conditions, geological history. Analysis of the genetic variability of the species with the inclusion of new data from the Cis-Urals, Urals and Western Siberia indicates genetic heterogeneity within the Eastern haplogroup of the field vole. The two major haplogroups are revealed in the Eastern subclade of the Northern clade (BI not less than 0,70). One haplogroup (Northern European group) includes haplotypes from Norway, Finland, and Karelia, and another haplogroup (Ural-Western Siberia group) includes haplotypes from the Urals and Western Siberia. Nucleotide diversity obtained for these two haplogroups suggests their recent expansion.

The data on mtDNA variability (complete cyt *b* gene) in *M. agrestis* in the Cis-Urals, the Southern, Middle, Northern Urals, and Western Siberia obtained for the first time expand significantly the geography of investigations of the species genetic diversity. Analysis of the genetic variability of the

field vole with the inclusion of new data does not contradict present views on the genetic structure of the species, but reveals the genetic heterogeneity of populations within the Eastern subclade, which was not previously considered. To clarify the genetic structure of *M. agrestis* within the wide-ranging Eastern subclade (from Norway to Lake Baikal), further studies of the species in Northern Eurasia are necessary.

ACKNOWLEDGMENT

This study was supported by the Russian Foundation for Basic Research (grant No. 19-04-00966).

REFERENCES

[1] G. I. Shenbrot and B. R. Krasnov. "Atlas of the geographic distribution of the arvicoline rodents of the world (Rodentia, Muridae: Arvicolinae)," Pensoft, 2005.

[2] M. Jaarola and J. B. Searle, "Phylogeography of field voles (Microtus agrestis) in Eurasia inferred from mitochondrial DNA sequences," Molecular ecology, vol. 11, pp. 2613-2621, 2002.

[3] M. Jaarola and J. B. Searle, "A highly divergent mitochondrial DNA lineage of Microtus agrestis in southern Europe," Heredity, vol. 92, pp. 228-234, 2004.

[4] J. S. Herman and J. B. Searle, "Post-glacial partitioning of mitochondrial genetic variation in the field vole," Proceedings of the Royal Society B: Biological Sciences, vol. 278, pp. 3601-3607, 2011.

[5] J. Paupério, J. S. Herman, J. Melo-Ferreira, M. Jaarola, P. C. Alves, and J. B. Searle, "Cryptic speciation in the field vole: a multilocus approach confirms three highly divergent lineages in Eurasia," Molecular Ecology, vol. 21, pp. 6015-6032, 2012.

[6] J. S. Herman, A. D. McDevitt, A. Kawałko, M. Jaarola, J. M. Wójcik, and J. B. Searle, "Land-bridge calibration of molecular clocks and the post-glacial colonization of Scandinavia by the Eurasian field vole Microtus agrestis," PLoS One, vol. 9, 2014.

DOI 10.18699/BGRS/SB-2020-241

Effects of long-term ethanol consumption in mice: interaction between BDNF and brain serotonin systems

Naumenko V.S.
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
naumenko2002@mail.ru

Ilchibaeva T.V.
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
rbicehok@mail.ru

Antonov Y.V.
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
yegor@bionet.nsc.ru

Bazovkina D.V.
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
daryabazovkina@gmail.com

Popova N.K.
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
npopova@bionet.nsc.ru

Abstract — **Brain serotonin (5-HT) and brain-derived neurotrophic factor (BDNF) are principal players in the mechanisms of brain and behavioral plasticity. Here we investigated the effect of chronic (6 weeks) consumption of 10% alcohol on the principal elements of BDNF (BDNF, proBDNF, p75 and TrkB receptors) and 5-HT (5-HT, 5-HIAA, tryptophan hydroxylase-2 (Tph-2), 5-HT transporter (5-HTT), 5-HT1A, 5-HT2A and 5-HT7 receptors) systems in the brain of C57Bl/6 mice. BDNF mRNA level in the midbrain and BDNF protein level in the hippocampus were lowered in ethanol-treated mice. The increase in proBDNF protein level in the midbrain, cortex, and amygdala and the increase of p75 receptor protein level in the midbrain were revealed after ethanol exposure. Alcohol intake reduced the protein level and increased the activity of Tph-2, the key enzyme for serotonin biosynthesis in the brain, and increased the main 5-HT metabolite 5-HIAA level and 5-HIAA/5-HT ratio as well as 5-HT7 receptor mRNA level in the midbrain. In the cortex, 5-HT2A receptor protein level was reduced, and 5-HIAA/5-HT ratio was increased. Alcohol-induced changes in BDNF and 5-HT systems were revealed in the midbrain where the majority of the cell bodies of the 5-HT neurons are localized, as well as in the cortex, hippocampus and amygdala. Our data allow suggesting that BDNF/5-HT interaction contribute to the mechanism underlying chronic alcohol-induced neurodegenerative disorders.**

Keywords — alcohol consumption, brain serotonin system, brain-derived neurotrophic factor, mice

Brain serotonin (5-HT) and brain-derived neurotrophic factor (BDNF) are principal players in the mechanisms of brain and behavioral plasticity. The brain 5-HT system is involved in the regulation of wide range of physiological functions and different kinds of normal and disordered behavior. The incredible polyfunctionality of 5-HT is mediated by 14 different 5-HT receptor subtypes expressed in the mammals. In its turn, BDNF is crucially involved in the mechanisms of growth, development, functioning of the nervous system and in the response to stressful events. BDNF is notable by structural and functional complexity. Mature BDNF is produced by proteolytic cleavage of precursor, proBDNF. BDNF precursor proBDNF is biologically active inducing oppose to the mature BDNF effects via p75 receptor. Whereas mature BDNF plays crucial role in the neurogenesis promoting cell survival and differentiation via TrkB receptors, BDNF precursor via p75 initiates receptors apoptosis. Accumulating evidence indicates the cross-talk between 5-HT and BDNF systems and the crucial role of this interaction in the regulation of brain functioning.

The aim of the present study was to investigate the effect of chronic alcohol consumption on the principal elements of BDNF (BDNF, proBDNF, p75 and TrkB receptors) and 5-HT (5-HT, 5-HIAA, tryptophan hydroxylase-2 (Tph-2), 5-HT transporter (5-HTT), 5-HT1A, 5-HT2A and 5-HT7 receptors) systems in the brain of C57Bl/6 mice.

Experiments were carried out on C57BL/6j mice. Mice were treated with 10% (w/v) ethanol solution in drinking water (one bottle without choice) for a period of six weeks. mRNA levels were assessed using quantitative real-time RT-PCR technique. Protein levels were estimated by Western-Blot analysis. High Performance Liquid Chromatography was applied for 5-HT and its metabolite 5-hydroxiindol acetic acid (5-HIAA) quantification as well as for Tph2 activity estimation. The data were analyzed using one-way ANOVA followed by Fisher's LSD post-hoc comparison.

The chronic alcohol consumption induced reduction of BDNF mRNA level in the midbrain and BDNF protein level in the hippocampus. However, it failed to affect the expression of gene encoding TrkB receptor as well as TrkB-FL and TrkB-T1 receptor protein levels. At the same time, the increase in proBDNF protein level in the midbrain, frontal cortex, and amygdala and the increase of p75 receptor protein level in the midbrain raphe nuclei area were revealed after ethanol exposure. Chronic alcohol intake reduced the protein level and increased the activity of Tph2, the key enzyme for serotonin biosynthesis in the brain. These changes were accompanied by increased 5-HIAA level and 5-HIAA/5-HT ratio as well as 5-HT7 receptor mRNA level in the midbrain. In the frontal cortex, 5-HT2A receptor protein level was reduced, and 5-HIAA/5-HT ratio was increased.

These data showed adverse impact of alcoholization on the BDNF system linked with pro-apoptotic effect of alcohol-induced proBDNF and p75 receptor activation. Pronounced alcohol-induced changes in both BDNF and 5-HT systems were revealed in the raphe nuclei area of the midbrain where the majority of the cell bodies of the 5-HT neurons are localized. Our data allow suggesting that BDNF/5-HT interaction contribute to the mechanism underlying chronic alcohol-induced effects.

The cost of animal maintenance was supported by the basic research project # 0324-2019-0041-C-01; the study was supported by the Russian Scientific Foundation (grant No 19-15-00025).

DOI 10.18699/BGRS/SB-2020-243

Distribution of runs of homozygosity (ROHs) along the human genome is shaped by recombination and purifying selection

K. Popadin
EPFL, Lausanne, Switzerland
konstantinpopadin@gmail.com

E. Zezyulya
IKBFU, Kaliningrad, Russia
liza.zez1997@gmail.com

D. Iliushchenko
IKBFU, Kaliningrad, Russia
serioussamkrsk@gmail.com

A. Reymond
University of Lausanne, Lausanne, Switzerland
alexandre.reymond@unil.ch

Abstract — **Whereas the higher fraction of the human genome, covered by Runs Of Homozygosity (ROHs), was associated with decreased height, decreased educational attainment and other phenotypes, the mapping position of the contributing ROHs was not assessed in details. Here, analysing genome-wide distribution of ROHs in offsprings of 100 consanguineous families we observed a non uniform distribution. Using multiple linear model we demonstrated that this variation could be explained by (i) recombination level, i.e. the higher the recombination in a given region, the less the number of ROHs; and (ii) selection, i.e. the higher the pLI score (which estimates loss of function intolerance) of the embedded genes the less the number of ROHs.**

The observed selection component might be explained by elimination of carriers of homozygous recessive loss of function mutations in genes with high pLI. To examine the selection component further we stratified ROHs by age and length following a previously suggested classification: class A (old and short: < 0.6 Mb), class B (intermediate in age and length: 0.6 – 1.6 Mb) and class C (young and long: > 1.6 Mb) and rerun our analyses. We observed that while the distribution of all three classes is shaped by recombination, only the class C ROHs showed a strong negative relationship with pLI, suggesting that younger ROHs are under stronger selection. Altogether our results provide additional metrics (properties of affected genes, ROH classes), which should be taken into account inferring deleterious effect of ROHs in GWAS and genetic medicine.

Methods

Having 100 consanguineous pedigrees with genotyped parents and offspring we called all ROHs using default PLINK settings (DNA regions with at least 50 consecutive homozygous SNPs). Thereafter we focused on offspring ROHs and analyzed their distribution along the human genome taking into account recombination map and properties of embedded genes.

Results

The distribution of three classes of ROHs (short, intermediate, long) is highly non-uniform along human chromosomes. We hypothesized that this distribution is shaped by recombination, which mechanistically disrupts ROHs and thus decreases their abundance in the recombination hotspots, and by purifying selection, which eliminates ROHs from sensitive genomic regions with embedded haploinsufficient genes. Using multiple linear model we demonstrated that:

Coverage(ROH) ~ -2.27*RecombinationRate-0.6*ProbabilityOfLossOfFunctionIntolerance

(all p values are < 10-6; coefficient are scaled). Interestingly, the distribution of the longest ROHs (class C) demonstrates the strongest dependence on both recombination and LoF intolerance. We conclude that selection in line with recombination shapes the ROH distribution along the human genome.

Acknowledgment

We are grateful for the data provided to P. Makrythanasis (Academy of Athens, Athens, Greece), M. Ansar, S. E. Antonarakis (University of Geneva, Geneva, Switzerland).

References

[1] Hamamy, H. (2012) Consanguineous marriages: preconception consultation in primary health care settings. *J Commun. Genet*. 3(3):185–92.

[2] McQuillan, Ruth, et al. (2008) Runs of homozygosity in European populations. *Am. J. Human Genetics.* 83.3, 359-372.

[3] Pemberton, Trevor J., et al. (2012) Genomic patterns of homozygosity in worldwide human populations. *Am. J. Human Genetics.* 91.2,275-292.

DOI 10.18699/BGRS/SB-2020-244

The impact of early-life stress on the expression of genes associated with the formation of the myelin sheath of neurons in the prefrontal cortex of 15-day-old male mice

Anastasia S. Shulyupova
Laboratory of Gene Expression Regulation
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
shulyupova@bionet.nsc.ru

Arina A. Smelova
Epigenetics laboratory
National Research Novosibirsk State University
Novosibirsk, Russia
smelovarina@gmail.com

Vasiliy V. Reshetnikov
Laboratory of Gene Expression Regulation
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
vasiliyreshetnikov@bionet.nsc.ru

Natalya P. Bondar
Laboratory of Gene Expression Regulation
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
nbondar@bionet.nsc.ru

Abstract — **The early postnatal period is critical for the development of the central nervous system and can predetermine subsequent behavioral patterns, so stress at this age can lead to long-term consequences in adulthood. In our work, we investigated effects early life stress on myelination. We assessed the expression of the genes associated with the myelin sneath (myelin basic protein - Mbp, myelin-associated oligodendrocyte basic protein - Mobp, proteolipid protein 1 - Plp1, proteolipid in compact myelin – Mal, myelin oligodendrocyte glycoprotein – Mog, choline-specific glycerophosphodiester-phosphodiesterase – Enpp6, UDP-glucuronosyltransferase – Ugt8a) in frontal cortex of mice. The greatest difference in the level of expression was found between groups with a maternal deprivation and a maternal separation from mothers: significant changes were shown for the Enpp6, Mal, Ugt8a genes and one of the Mobp gene transcripts. This may indicate a different extent of exposure to stressors. We can conclude that the stress in the early postnatal period affects the myelination process and can lead to impaired transmission of nerve impulses.**

Key words — effects of stress; brain; myelination; mice; pups

Motivation and Aim

The early postnatal period is critical for the development of the central nervous system and can predetermine subsequent behavioral patterns, so stress at this age can lead to long-term consequences in adulthood. Thus, on animal models has been shown that a maternal separation (3 hours) pups from their mothers leads to interruption in the continuity of interaction with the mother and causes severe stress in the pups. Upon reaching adulthood in such animals, an increase in sensitivity to stress, increased anxiety and a number of cognitive disorders (deterioration of spatial memory, ability to recognize objects, etc.) are observed. A previous study of a transcriptome of the prefrontal cortex of 15 day old pups showed that this stress leads to a change in the expression of a large number of genes. Among them, a cluster of genes associated with the formation of the myelin sheath of axons was isolated. The aim of this work was to study the change in the expression of *Mbp, Mobp, Plp1, Mal, Mog, Enpp6, Ugt8a* genes associated with the formation of the myelin sheath of neurons in the prefrontal cortex of 15 day old male mice under the influence of early postnatal stress. In addition, the effect of stress on physiological (animal weight, adrenal gland weight) and biochemical (blood corticosterone level) parameters was evaluated.

Materials and Methods

Materials

To obtain experimental offspring, female mice (28 animals) were planted with males in the ratio of 3 females to 1 male. There were 6 to 9 pups in the litter. Two types of early postnatal stress, conducted from PND2 to PND14, were used in the study: maternal separation of the pups from their mothers (3 hours a day daily, MS) and a maternal deprivation (for a day, MD). The control group did not separate from mothers. The pups were separated daily from 13 to 16 hours. On the 15th day of life, animals were killed by decapitation, blood was collected, the prefrontal cortex was isolated and frozen in liquid nitrogen. The adrenal glands were isolated and weighed. RNA was subsequently isolated from frozen tissue samples, and complementary DNA was synthesized based on it.

Methods

A TaqMan probe-based real-time PCR to evaluate gene expression was used. We evaluated the expression of the main myelin sheath genes: Mbp - encodes a protein that is the main component of the myelin sheath of oligodendrocytes and Schwann cells in the nervous system, Mobp - encodes the main oligodendrocyte protein associated with myelin (evaluated using two pairs of primers covering all known gene transcripts), Plp1 - encodes a proteolipid protein, which plays an important role in the formation and maintenance of the multilayer structure of myelin, Mal - encodes a proteolipid involved in myelin biogenesis. In the studied groups of animals, body weight and organ index of the adrenal glands were evaluated. The organ index was calculated by the formula: OI = adrenal gland weight / body weight * 100%. The level of corticosterone was measured in the blood serum of pups on PND 15 by ELISA.

Results

The effect of stress on the physiological characteristics of pups

The stress factor did not significantly affect the body weight of the pups. However, stress during the first two weeks of life had an effect on the relative weight of the adrenal glands: in the MD group, in comparison with the MS group, the organ adrenal gland index was significantly reduced in pups. The level of corticosterone did not differ significantly between different experimental groups, however, in mice from the MS group there was a tendency ($p < 0.073$) to lower levels compared with the control group.

The effect of stress on gene expression in the frontal cortex

Analysis of gene expression in the frontal cortex showed a significant increase ($p = 0.038$) in the expression level of the *Mal* gene in the MD group compared to the MS group and a tendency to increase ($p = 0.053$) compared to the control group. The *Mobp* gene is characterized by the presence of many alternative transcripts. Since the gene structure does not allow choice of primers covering all isoforms, we estimated the expression level using two pairs of primers. One pair covers the transcripts *Mobp-201, 202, 204, 206, 207*, the second pair is specific only for the transcript ENSMUST00000174193.7 (*Mobp-204*). For the first pair of primers of the *Mobp* gene, the analysis of PCR-RT data did not reveal significant differences between the experimental groups ($p > 0.05$), however, the change in the expression of its transcript ENSMUST00000174193.7 was significantly reduced ($p = 0.03$) in the MS group compared with the MD group. For the *Ugt8a* gene, a tendency towards an increase in the expression level ($p = 0.07$) in the MD group as compared with the control was revealed, and a significant increase in the expression level of the *Ugt8a* gene in the MD group compared to the MS group ($p = 0.04$). Analysis of *Enpp6* gene expression showed that in the MD group the expression level is higher compared to the control ($p = 0.02$) and the MS group ($p = 0.011$). Analysis of the expression of the *Mbp, Plp1*, and *Mog* genes did not show significant differences in the expression level between the groups.

The greatest difference in the level of expression was found between groups with a maternal deprivation and a maternal separation from mothers: significant changes were shown for the *Enpp6*, *Mal*, *Ugt8a* genes and one of the *Mobp* gene transcripts. This may indicate a different extent of exposure to stressors. We can conclude that the experience of early stress affects the myelination process and can lead to impaired transmission of nerve impulses.

Acknowledgment

Supported by the Russian Science Foundation (16-15-10131).

References

[1] Bondar, N. P., Lepeshko, A. A., & Reshetnikov, V. V. (2018). Effects of early-life stress on social and anxiety-like behaviors in adult mice: sex-specific effects. Behavioural neurology, 2018(2), 1-13.

[2] Mehta, M., & Schmauss, C. (2011). Strain-specific cognitive deficits in adult mice exposed to early life stress. Behavioral neuroscience, 125(1), 29.

[3] Sakhai, S. A., Saxton, K., & Francis, D. D. (2016). The influence of early maternal care on perceptual attentional set shifting and stress reactivity in adult rats. Developmental psychobiology, 58(1), 39-51.

DOI 10.18699/BGRS/SB-2020-245

Histological evaluation of postnatal retinal development of senescence-accelerated OXYS rats

Darya V. Telegina
ICG SB RAS, Novosibirsk, Russia
telegina@bionet.nsc.ru

Anna K. Antonenko
NSU, Novosibirsk, Russia
antonenko-98@bk.ru

Oyuna S. Kozhevnikova
ICG SB RAS, Novosibirsk, Russia
oidopova@bionet.nsc.ru

Abstract — *De novo* neurogenesis in the adult mammalian retina very limited. Thereby the structural and functional features formed during the period of maturation and formation of retina can have long-term effects on the further ontogenesis of the tissue, however, the mechanisms of these disorders remain unclear. Using model of premature aging OXYS rats we investigated the early histopathological changes during postnatal retinal neurogenesis. OXYS rats spontaneously develop a retinopathy similar to age-related macular degeneration (AMD). AMD is a complex neurodegenerative disease resulting in a loss of central vision in the elderly. Ganglion, horizontal, and amacrine cells are born in the embryonic phase of rat developmen. Quantitative analysis showed decreasing amacrine cells in OXYS rats as compared Wistar rats (control). At the age of P0 and P1, the number of ganglion and horizontal cells increased in OXYS rats as compared Wistar rats. Bipolar neurons, photoreceptors and Müller glia are born postnatally. We did not find changes in Muller cells. The number of photoreceptor's nuclei per column in Wistar rats increased at the age of P10 and decreased at the age of P14. In OXYS rats, maximum of number of nuclei per column accounted for age of P14 and then decreased. The number of rod bipolar neurons gradually increased by age of P14 in Wistar rats and P10 in OXYS rats. Our results indicating an alteration of retinal formation in OXYS rats during the postnatal period, which may contribute to the early development of their signs of AMD.

Keywords — neurogenesis, retina, AMD, OXYS rats, Wistar rats

Introduction

The retina is a delicate neural tissue responsible for light signal capturing, modulating, and passing to mid-brain. The adult vertebrate retina comprises six major cell classes derived from a common progenitor cell, only one of which, the Müller cell, is a glial cell. It is important that the mammalian neural retina, however, like many other regions of the central nervous system, lacks any significant regenerative potential to replace lost neurons after retinogenesis is complete. Consequently, any damage leading to the death of the light sensing photoreceptor cells or their support cells, either through injury or disease, typically leads to permanent visual impairment. But our understanding of human eye development is very limited, especially the contribution of impaired retinogenesis in development of retinal degeneration.

The retinal neurons include cone and rod photoreceptors and horizontal, amacrine, bipolar, and ganglion cells. There is evidence to suggest that ganglion, horizontal, amacrine, and early-born cone photoreceptor cells originate froma common pool of intermediate progenitors (EBPs). Bipolar, Müller, and late-born photoreceptor cells share a pool of common intermediate progenitors (LBPs) [1]. In rat and mouse EBPs arise from E10 to P4 and peak at E16, while LBPs appear from E13 to P10 and peak at P2. Retinogenesis is complete at P20 in rodents.

Understanding the programs underlying the developmental process not only will provide insights into how each individual cell type is formed and what function it may have, but also will give clues on how to treat retinal diseases in the future such as age-related macular degeneration (AMD). AMD is a complex neurodegenerative disease resulting in a loss of central vision in the elderly. Here, we used a suitable genetic experimental model of AMD is senescence-accelerated OXYS rats, which spontaneously develop a phenotype similar to human age-related disorders including AMD-like retinopathy. Retinopathy that develops in OXYS rats already at a young age corresponds (in terms of clinical manifestations and morphological characteristics) to the dry atrophic form of AMD in humans. Nonetheless, neovascularization develops in some (~ 10–20%) of these rats with age. The clinical signs of AMD-like retinopathy appear by the age of 3 months in 100% of OXYS rats against the background of a reduction in the transverse area of the RPE, impairment of choroidal microcirculation, and retinal thinning. The progression of these abnormalities in OXYS rats until the age of 10 months is accompanied by a significant reduction in thickness of the photoreceptor cell layer and a decrease in the number of photoreceptor cell nuclei of the outer nuclear layer, especially in the central part of the retina [2].

We aimed to give a detailed qualitative and quantitative description on the early histopathological changes in the postnatal development retina of OXYS and Wistar rats (as control) and compare our results with previous studies dealing with the senescence-accelerated OXYS rats.

Methods

All procedures of the present study were performed in concordance with the ARVO statement for the Use of Animals in Ophthalmic and Vision Research and were approved the Institutional Review Board of the Institute of Cytology and Genetics, according to The Guidelines for Manipulations of Experimental Animals. The study was carried out on OXYS and Wistar (control) rats in the critical days of postnatal retinal neurogenesis: 0, 1, 3, 5, 7, 10, 14, 20. We used standard techniques of immunohistochemistry and TUNEL assay [3].

Results

By immunohistochemistry using the specific antibodies carried out a comprehensive assessment of cell populations differentiating in the postnatal period into photoreceptors, Müller cells, amacrine, ganglion, bipolar and horizontal neurons and also evaluated the dynamics of apoptosis using the TUNEL assay.

Müller glia cells were analysed using antibodies against cell type specific cytoskeletal elements: vimentin and nestin. Nestin is an intermediate filament protein that is known as a neural stem/progenitor cell marker. It is expressed in undifferentiated CNS cells during development. Vimentin immunoreactivity throughout the whole depth of the retina of OXYS and Wistar rats was maximum in P0 and was gradually decreased with age. Double labelling with anti - nestin revealed a similar pattern of immunoreactivity in radial cell processes throughout the retina, which colocalized with vimentin immunoreactivity. Comparing sections of OXYS and Wistar rats during postnatal retinal development, we did not find any qualitative or significant quantitative changes in Vimentin and Nestin labelling.

Recoverin is a protein in rods that regulates rhodopsin phosphorylation in visual transduction and enhances visual sensitivity. In the OXYS and Wistar retinas, strong recoverin immunoreactivity was detected in putative differentiating photoreceptors, localized to the apical surface of the neuroblast layer (NBL) and in sparse cells found deeper in the inner NBL (bipolar cells). Recoverin staining was seen throughout the developing outer nuclear layer (ONL) in rat retina. In the ONL, nuclei are typically arranged in columns on well oriented sections. Recoverin-labelled nuclei formed in well-organized columns starting on the fifth postnatal day in OXYS and Wistar rats. Therefore, we counted the average number of recoverin-labelled nuclei per column. In Wistar rat's retina the recoverin-labelled layer was well-formed at the age of P5 whereas in OXYS rat's retina – at the age of P7. Quantitative analysis showed increasing the number of nuclei per column in Wistar rats at the age of P10 and decreased at the age of P14. In OXYS rats, we observed a similar trend but the period time was altered. Therefore, the maximum of number of nuclei per column accounted for age of P14 in OXYS rats and then the number of nuclei per column decreased (at the age of P20). It is important that average number of recoverin-labelled nuclei per column was greater in Wistar rats than OXYS rats at the age of P5 and P7. However, this parameter was greater in OXYS rats than Wistar rats at the age of P10, P14 and P20. Besides the photoreceptors both the ON and OFF cone bipolar cells can be labelled with recoverin antibody. Average number of recoverin-labelled bipolar neurons gradually increased by age of P14 in Wistar rats. In OXYS rats this parameter was drammaticaly increased by age of P10 and then decreased. We proposed that this phenomenon associated with adaptive mechanisms.

Parvalbumin stains dominantly AII amacrine cells, along with a few wide field amacrine cells, a population of bipolar cells and a few cells in the ganglion cell layer [4]. The AII amacrine cells is one well-characterized subset of glycinergic amacrine cells [5]. Regarding AII positive amacrine cell populations, we detected a weakly staining intensity for parvalbumin labelled cells before P3 in OXYS and Wistar rats. After P3 parvalbumin labelled cells are characterized by strongly immunoreactivity and distinctly morphology in both rat strains. Quantitative analysis showed decreasing parvalbumin labelled cells in OXYS rats as compared Wistar rats at the all ages (P0-P20).

The antibody against calbindin, a calcium binding protein involved in calcium transport, stains horizontal cells and a few cells in the INL and GCL under control conditions [4]. We observed a weakly staining intensity for calbindin labelled cells before P5 in OXYS and Wistar rats. After P5 calbindin labelled cells are characterized by strongly immunoreactivity and distinctly morphology in both rat strains. In Wistar rats we detected no significant difference in horizontal cells at the all ages. In OXYS rats average number of calbindin labelled cells was greater that than in Wistar rats at the age of P0. At the age of P1, number og horizontal neurons dramatically decreased. After P0 significant difference between OXYS and Wistar rats is not observe.

In next step we used anti-NeuN antibody as a marker for ganglion cells. Our results showed that NeuN was strongly expressed in ganglion and displaced amacrine cells in the GCL and weakly expressed in amacrine cells and horizontal cells in the INL in OXYS and Wistar rats. The average number of NeuN-positive cells decreased in the age of P7 in Wistar and OXYS rats. We detected increasing number of NeuN labelled ganglion cells in OXYS rats at the age of P0 and P1 as compared Wistar rats.

A TUNEL analysis of retinal cryosections showed a higher apoptosis activity (especcialy P7, P10 and P14 days) in OXYS rats, which indicates an enhanced reduction of neurons.

Summary, we obtained the first results indicating a alteration of retinal formation in OXYS rats during the postnatal period, which may contribute to the early development of their signs of AMD.

Conclusion

Thus, using senescence-accelerated OXYS rats, which spontaneously develop AMD-like retinopathy, we obtained the first results of a study of the contribution of changes in neurogenesis in the retina during the postnatal period to the early development of AMD-like retinopathy in OXYS rats. We hypothesize that the histological features presented in OXYS rats during postnatal neurogenesis could underlie the functional and structural alterations and can have long-term effects on the further ontogenesis of the tissue.

Acknowledgment

This study was supported by the Russian Foundation for Basic Research (18-015-00320) and by a Russian Scientific Foundation Grant (19-15-00044).

References

[1] Jin, K. (2017). Transitional progenitors during vertebrate retinogenesis. Molecular neurobiology, 54(5), 3565-3576. https://doi.org/10.1007/s12035-016-9899-x

[2] Telegina, D., Kozhevnikova, O., Bayborodin, S. et al. (2017). Contributions of age-related alterations of the retinal pigment epithelium and of glia to the AMD-like pathology in OXYS rats. Sci Rep, 7, 41533. https://doi.org/10.1038/srep41533

[3] Kozhevnikova, O. S., Telegina, D. V., Tyumentsev, M. A., Kolosova, N. G. (2019). Disruptions of Autophagy in the Rat Retina with Age During the Development of Age-Related-Macular-Degeneration-like Retinopathy. International Journal of Molecular Sciences, 20(19), 4804. https://doi.org/10.3390/ijms20194804

[4] Szabó, K., et al. (2017). Histological evaluation of diabetic neurodegeneration in the retina of Zucker Diabetic Fatty (ZDF) rats. Scientific reports, 7(1), 1-17. https://doi.org/10.1038/s41598-017-09068-6

[5] Balasubramanian, R., & Gan, L. (2014). Development of retinal amacrine cells and their dendritic stratification. Current ophthalmology reports, 2(3), 100-106. https://doi.org/10.1007/s40135-014-0048-2

DOI 10.18699/BGRS/SB-2020-246

Study of the COI gene fitness for a population-genetic analysis of endemic Baikal Sponges L. Baikalensis

Alena Yakhnenko
LIN SB RAS, Irkutsk, Russia
JINR, Dubna, Russia
yakhnenkoas@gmail.com

Valeria Itskovich
LIN SB RAS, Irkutsk, Russia
itskovich@mail.ru

Abstract — **Freshwater sponges play an important role as filtrators in lake ecosystems. Endemic Baikal sponges make up the bulk of the lake benthos biomass. Events of mass diseases and death of sponges occur on Lake Baikal for the last decade. Due to high morphological plasticity, there is a lack of clear criteria for the species identification of Baikal sponges. However, the development of such criteria is very important for population structure studying purposes and determination of the recoverability of populations. In this work, we first assessed the suitability of the 5'-terminal fragment and the I3M11 fragment of the COI gene for population studies of endemic Baikal sponges, and also examined alternative markers for such studies. It is shown that in the Lubomirskia baikalensis samples, collected in different basins of Lake Baikal, only two different haplotypes were found, which indicates the unsuitability of this fragment for population studies. As an alternative, we propose to use microsatellite markers, which have successfully shown themselves both in studying the structure of populations of marine and freshwater sponges. Currently, work is underway to study the population structure of Baikal endemic sponges using microsatellite markers.**

Keywords — Sponges, population genetics, molecular markers, Baikal endemics, Porifera

Motivation and aim

Motivation

Sponges are ancient multicellular animals leading an attached lifestyle. They are widespread throughout the planet, inhabiting marine and freshwater bodies. According to the method of nutrition, the sponges are filtrators, which contribute to the increased sensitivity of the sponges to contaminants. Thus, sponges are bioindicators of the state of aquatic ecosystems. Freshwater sponges that live in the ancient rift lake Baikal, formed a bouquet of endemic species that have features that clearly distinguish them among freshwater cosmopolitans [1; 2]. Such features include the absence of gemmules and a long life cycle, which, undoubtedly, should affect the structure of populations of endemic species. Unfortunately, genetic structure of Baikal sponges still remains unknown.

Aim

A study of the population structure of sponges will make it possible to understand the causes of their mass disease and death; for marine sponges, such studies have been successfully conducted. In recent years, a process of mass disease and death of endemic sponges occurs on the lake Baikal [3]. Similar events were noted in some marine ecosystems. As shown [4], changes in the sponge population structure are observed precisely in places where their mass death was previously noted. Thus, the study of the population structure of Baikal endemic sponges will allow us to study the influence of sponge diseases on their population structure at the genetic level. Genetic markers successfully used for population genetic studies of sponges are the COI gene and microsatellite markers [5; 6]. Moreover, population analysis using COI is successfully used in the study of sea sponges [7], but there is no such data on freshwater sponges, and microsatellite analysis is used both in the study of sea sponges [8] and freshwater sponges [9, 10].

Methods

Baikal sponge samples were collected by SCUBA diving during expedition in 2018 in the southern, middle and northern basins of Lake Baikal. Samples were fixed in 70% ethanol for morphological and molecular analysis. Total genomic DNA extraction was performed using the CTAB method. The standard 5'-end fragment of COI was amplified with primers previously described by Folmer et al. (1994) (C_1538F: 5'-GGTCAACAAATCATAAAGATATTGG-3') and (C_1539R:5'-TAAACTTCAGGGTGACCAAAAATC A-3′), which amplifed a fragment of 632 base pairs (bp). The I3-M11 fragment was amplifed using primers developed by us (forward primer 5'-GCTGGAGGAGGAGACCCAAT-3') and (reverse primer 5'-TGGAAATCCGAATACCGTCTCG-3'). PCR was carried out in a 20-µl volume of the reaction mixture under the following conditions: preheating to activate the polymerase 2 min at 94 °C, followed by 35 cycles of DNA denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s, and elongation at 72 °C for 45 s, followed by final elongation of 8 min at 72 °C. Each PCR reaction was purifed by electrophoresis in a 0.8% agarose gel, followed by extraction by freezing and thawing. Sequencing of the PCR products was carried out at the service laboratory "Syntol" (Moscow, Russia).

Results

In this work we analyzed 8 samples L.baikalensis species, collected in the Northern, Middle, and Southern hollows of Lake Baikal. During the sequence analysis for the 5'-terminal fragment of the COI, no differences were found among 8 samples, and for the I3M11 fragment, 2 haplotypes differing by 1 nucleotide were detected. The maximum p-distance value was 0.2%, the average value was 0.04%. Values obtained for intraspecific variability indicate the unsuitability of the COI gene for analysis of the population structure of Baikal endemic sponges. Although this gene has been successfully used for population-genetic studies of sea sponges [7]. The reason for the lower variability of the COI gene in Baikal endemic sponges compared with marine sponges could be the later formation of the *L.baikalensis* species. An alternative way to study the population structure of freshwater sponges is microsatellite analysis since it is successfully used both for studying marine [4; 8] and freshwater sponges (*Ephydatia fluviatilis*) [9; 10]. Analysis of the population structure of endemic Baikal sponges using microsatellite markers is being carried out for the first time now.

ACKNOWLEDGMENT

The reported study was funded by RFBR and the Government of the Irkutsk Region, project number 20-44-383010.

REFERENCES

[1] Itskovich V.B. et al. (2006) Monophyletic origin of freshwater sponges in ancient lakes based on partial structures of COXI gene. Hydrobiologia. 568: 155–159.

[2] Itskovich V.B. et al. (2008) Ribosomal ITS sequences allow resolution of freshwater sponge phylogeny with alignments guided by secondary structure prediction. J. Mol. Evol. 67(6): 608–620.

[3] Khanaev I.V. et al. (2018) Current state of the sponge fauna (Porifera: Lubomirskiidae) of Lake Baikal: Sponge disease and the problem of conservation of diversity. J. Great Lakes Res. 44(1): 318-332.

[4] Riesgo A. et al. (2019) Genetic diversity, connectivity and gene flow along the distribution of the emblematic AtlantoMediterranean sponge Petrosia ficiformis (Haplosclerida, Demospongiae). BMC Evol. Biol. 19(1): 24-39.

[5] Gui J. F. Et al. (2010) Genetic basis and breeding application of clonal diversity and dual reproduction modes in polyploid Carassius auratus gibelio. Sci. China Life Sci. 53(4): 409–415.

[6] Vieira M.L.C. et al. (2016) Microsatellite markers: What they mean and why they are so useful. Genetics and Molecular Biology. 39(3): 312–28.

[7] López-Legentil S. et al. (2009) Genetic structure of the Caribbean giant barrel sponge Xestospongia muta using the I3-M11 partition of COI. Coral Reefs. 28(1): 157–165.

[8] Blanquer A. Et al. (2010) Population genetics at three spatial scales of a rare sponge living in fragmented habitats. BMC Evol. Biol. 28(1): 157–165.

[9] Lucentini L. et al. (2013) Spatially explicit genetic structure in the freshwater sponge Ephydatia fluviatilis (Linnaeus, 1759) within the framework of the monopolisation hypothesis. J. Limnol. 72: 172–181.

[10] Li R. et al. (2018) Mesoscale investigations based on microsatellite analysis of the freshwater sponge Ephydatia fluviatilis in the River-Sieg system (Germany) reveal a genetic divergence. Conserv. Genet. 19: 959–968.

[11] Efremova S.M. (2004) New Genus and New Species of Sponges from Family Lubomirskiidae Rezvoj, 1936. Index of animal species inhabiting Lake Baikal and its catchment area. 1(2): 182–192.

[12] Folmer O. et al. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3(5): 294–299.

DOI 10.18699/BGRS/SB-2020-247

Enlarged clinical Belarusians' exomes: opportunities and restrictions of additional analysis

Danat Yermakovich
IGC NAS, Minsk, Belarus
danatyermakovich@gmail.com

Aleh Liaudanski
IGC NAS, Minsk, Belarus
bioinfgroup@igc.by

***Abstract* — The new era of large NGS comes to Belarus. With the quite fast production of big data in wet labs, the problem of processing and analysis it raises up. We as a young bioinformatics group faced the necessarily recycling data for system analysis versus routine clinical investigations on the presence of pathogenic variants. According to the type of data, we chose the population genetics field. After the common variant calling from enlarged clinical exomes NGS Illumina data, we expect to get plausibly infers for the Belarusian population using generally known types of analysis. The obtained already PCA plot shows the distinction of Belarussians from other 1000G populations.**

Keywords — population genetics, targeted NGS

Motivation and Aim

Motivation

Currently 200 enlarged clinical exomes are being sequenced within the institute, and we, a recently assembled bioinformatics group, would like to utilize the data for our own project. We see the restrictions of the data: the sample count, small number of variants, their distribution across the genome, the limitation to exonic variations, and the sample consists only of people with Mendelian diseases. Nevertheless, this will be the first systematically obtained genomic big data for Belarusians.

Aim

We plan to: a) Use this cohort for phasing and to impute from this cohort to other smaller data [1]; b) Visualize the distinction between Belarusian from other 1000G populations via PCA plot [2]; c) Conduct a GWAS. Half of the cohort have CAKUT, the other cardiomyopathies. The diseases seem to be Mendelian, but the rate of pathogenic's mutation finding is not high [3]; d) Infer a demographic history [4]; e) Make admixture with ancient populations [5]. We are considering obtaining microarray data for the cohort and mixing it with clinical exome data in way of enlarging it.

Methods

The targeted Illumina NGS is being carried out by using xGen® Exome Research Panel v1.0 (around 20000 genes). The analysis was made on received at the moment samples. The last pipeline includes the further steps: a) preprocessing: trimmomatic (cutting), bwa (aligning), samtools/picard (converting, marking, sorting), b) variant calling for clinical purposes: samtools/varscan2, freebayes, gatk 4 haplotype caller (calling), bcftools – (filtering, merging callers' results), annovar (annotating). c) variant calling for popgen purposes: freebayes, bcftools. Obtained cohort vcf-file was imputed and phased by shapeit, beagle, eagle in order to compare these tools. Also, the Michigan Imputation Server was tested. Finally, the pca plot was done by using smartpca from eigensoft. The input file for it was converted by plink. Currently, all our pipelines have been written on bash or Snakemake.

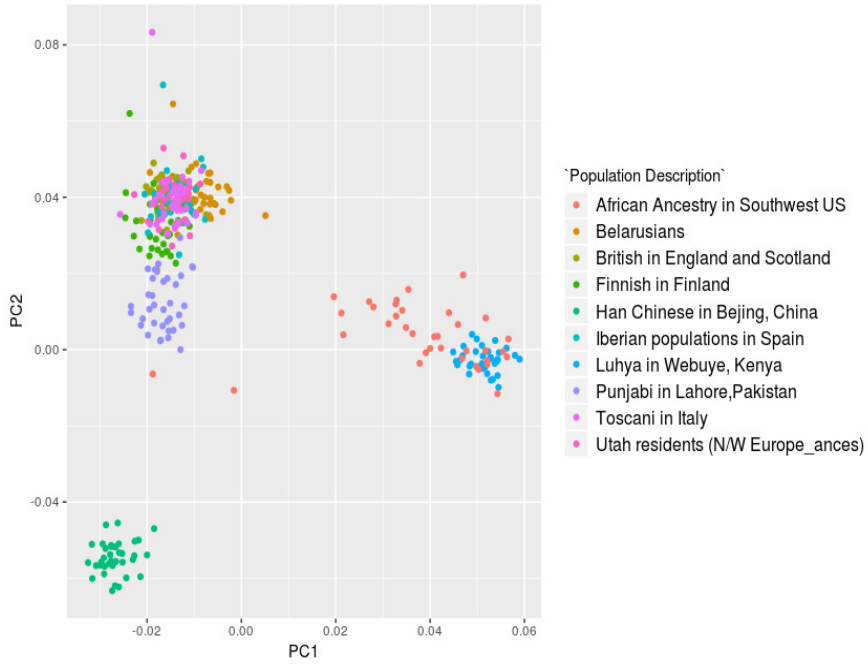


Fig. 1. A PCA plot was made from a subset of 1000G (36 samples in each population; chr 12 13 18 19)

Results

We tested our pipelines on 36 obtained yet samples and GIAB sample. By launching phasing/imputation tools, we found out the complexity of direct results' comparison of such tools due to the process of haplotype's assignment during their work. We're waiting for other samples to explore more aspects of the Belarussian population genetics. The first pca plot shows the relation of Belarusians to other nations Fig. 1.

ACKNOWLEDGMENT

The work is being done as part of the Belarus state research program "Biotechnology", 2019–2020; subprogram 2 "Structural and functional genomics", task 2.45

REFERENCES

[1] Herzig A.F. et al. (2018) Strategies for phasing and imputation in a population isolate. Genet Epidemiol. 42(2): 201-213.

[2] Galinsky K.J. et al. (2016) Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. AJHG. 98(3): 456-472.

[3] Gibson G. (2018) Population genetics and GWAS: A primer. PLoS Biol. 16(3).

[4] Alexander D.H., Novembre J. and Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19: 1655–1664.

[5] Noskova E. et al. (2020) GADMA: Genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. GigaScience. 9(3): 1-18.

DOI 10.18699/BGRS/SB-2020-248

Update status of mouse resources for studies of gene function and disease at RIKEN BRC

Atsushi Yoshiki
Coordinator, RIKEN BioResource Research Center and Head of Experimental Animal Division, Tsukuba, Ibaraki 305-0074, Japan
atsushi.yoshiki@riken.jp

Keywords — genetically-modified, genome editing, mice

Experimental Animal Division of RIKEN BioResource Research Center (BRC) has been designated since 2002 as the core facility of mouse resources in the National BioResource Project (NBRP) of Japan, and collected, preserved and distributed mice as a model system to understand human biology and disease mechanisms. We have so far collected over 9,000 mouse resources generated by Japanese scientists and made the mice available in high-quality to 1,490 organizations around the globe. One of our important challenges is ensuring the quality of genetically-modified mice created by many scientists with various methods including genome editing. We have conducted a strict genetic quality control program for the artificially mutated alleles prior to distribution. To disseminate among biomedical researchers high-quality mouse resources, we have participated in the International Mouse Strain Resource (IMSR) and Asian Mouse Mutagenesis & Resource Association (AMMRA). In addition, we have participated in the International Mouse Phenotyping Consortium (IMPC) to contribute for creation of comprehensive catalogue of mammalian gene functions. We have launched since April 2019 a new web corner to facilitate use of our models for gene functional analysis and human diseases in addition to “Mouse of the Month”. We are willing to collaborate with wider scientific community including clinical researchers and continue our resource infrastructure operation under the founding principles of “Trust, Sustainability, and Leadership”.

DOI 10.18699/BGRS/SB-2020-249

Testing inter-relations between disturbed sleep and sterility in intra-specific hybrids of fruit fly

Lyudmila Zakharenko
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
0000-0002-6341-8522

Dmitriy Petrovskii
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
0000-0002-0623-0363

Arcady Putilov
Research Institute for Molecular Biology and Biophysics
Federal Research Centre for Fundamental and Translational Medicine
Novosibirsk, Russia
0000-0003-2779-9046

Abstract — **Recent research connected sleeping problems and infertility in women. Sleep-wake behaviors of humans and *Drosophila melanogaster* demonstrate similarity in responsiveness to thermal stress. Moreover, studies of dysgenic crosses of fruit flies revealed possibility of inducing sterility in female intra-specific hybrids exclusively in one of two cross directions and only by exposing hybrids to thermal stress. Such facts motivated us to utilize this fruit flies' model of temperature-sensitive sterility for examining inter-relationship between deleterious effects of thermal stress on fertility, survival, and circadian patterns of locomotor activity and sleep. In three identical pilot experiments, these patterns were documented in 285 hybrids kept for, at least, five days in constant darkness at 20°C and 29°C. Additionally, the rhythms of 30 hybrid females of the 70th generation were tested. Sleep-wake patterns in females with temperature-induced sterility (progeny of Canton-S females crossed with Harwich males) significantly differed from the patterns in three other (fertile) crosses (males born from the same parents and offspring obtained by crossing in opposite direction). However, a lower rather than higher sensitivity of infertile females to thermal stress was revealed, and it persisted over generations. Unlike three other crosses, infertile females did not demonstrate such typical for flies and humans' responses to heat as night sleep disturbance and delay of the evening peak of activity. An elongated "siesta" was the only typical response common for this and opposing female cross. Data on survival of 2000 flies indicated that survival rate of female hybrids of the 1st and 70th generations was the highest compared to the rates of parent strains and hybrid males. Future research of fruit flies' models of reproductive health might be purposed on exploring sleep patterns not only in their relation to developmental sterility but also in their relation to acute deleterious effect of thermal stress on fertility.**

Keywords* — *infertility, reproductive health, sleep disorders, circadian rhythm, thermal stress, fruit fly model

Introduction

Relationship of infertility with circadian and sleep disturbances appears to be rather complex. Consequently, it is not well established so far. Complicating the issue is a question of whether circadian and sleep disturbances are the results of infertility, causes of infertility, or whether the relationship between them is one of reciprocal nature [1]. This makes difficult the therapeutic targeting both female reproductive function and sleep disorders. To answering the question of plausibility of causal relationship, it appears to be essential to understand genetic and environmental causes of association between infertility and disruptions of the fundamental circadian and sleep regulation mechanisms.

Given the complexity and limitations of human studies, the fruit fly, *Drosophila melanogaster*, is an ideal model organism for dissecting relationships between genes, environment, sleep-wake behavior, and various diseases [2]. It might, in particular, provide a powerful and rapid platform to study the mechanisms underpinnings the associations between environmental stress, infertility, and sleep-wake disturbances. Hot ambient temperature belongs to the most powerful natural factors causing the profound disturbances of the human sleep-wake cycle. For instance, a much stronger annual rhythmicity of sleeping problems was reported by both newcomers and native residents of a region with hot summer temperature as compared to people living at higher latitudes who are seasonally exposed to extremely long and short photoperiod with mostly cold rather than hot air temperatures [3]. The experimental research indicates that fly's sleep-wake pattern might be reorganized by heat in a way that is very similar to human sleep-wake pattern response to an increase of ambient temperature, i.e., nighttime sleep and daytime activity usually decrease whereas daytime sleep and early night activity usually increase during exposure to thermal stress [4]. On the other hand, *Drosophila melanogaster* can additionally serve as a model of temperature-sensitive agametic sterility that was originally called gonadal dysgenesis sterility [5]. Dysgenesic phenomena can occur exclusively in one of two cross directions and only due to cultivation under elevated temperature (reviewed in [6]). Therefore, the effects of thermal stress on the circadian rhythms in sterile female hybrids might be compared with the effects on these rhythms in the reciprocal (fertile) female hybrids. Consequently, the major aim of our pilot study was to utilize this fruit fly model of temperature-sensitive infertility for delineating possible causal link between the deleterious effects on fertility and sleep-wake behaviors imposed by such environmental factor as hot temperature. The hypothesis was that the circadian patterns of locomotor activity and sleep might be more dramatically disturbed by thermal stress in infertile rather than fertile crosses.

Methods

Hybrids and Conditions

In order to create hybrid females showing intra-specific paternal-maternal hybrid gonadal dysgenesis, females of Canton-S strain were crossed with males of Harwich strain, and F1 hybrids were maintained at high air temperature (i.e., eggs were hatched, larvae were reared, and newly emerged flies were kept permanently at 29 °C). The reciprocal (fertile) female hybrids and fertile hybrid males served as three control crosses cultivated in the same environmental condition. During three different seasons, all flies were cultivated at 29 °C until age of, at least, 5 days in the laboratory to perform three experimental studies of their

circadian patterns of locomotor activity and sleep. Additionally, female hybrids of two cross directions were reproduced for 70 generations at room temperature. In the 4th season they serve two more controls for the experimental study of circadian patterns of F1 female hybrids.

Sleep Recordings

To document circadian rhythmicity and sleep, flies were placed individually in glass locomotor-monitoring tubes. The measurements were obtained simultaneously for equal in size subsamples of flies kept in constant darkness at either 20°C or 29°C during, at least, 5 days. Locomotor activity was monitored in 1-min bins using the Drosophila Activity Monitoring System ("Trikinetics", Waltham, MA, USA). Sleep events were defined as 5 consecutive minutes of absence of any locomotor activity [7]. Parameters of locomotor activity and sleep were calculated using data acquisition software package (www.trikinetics.com). Number of beam breaks and duration of sleep episodes were summed on 30-min intervals of 5-day record to draw illustrations and to perform statistical analyses with the $SPSS_{23.0}$ statistical software package (IBM, Armonk, NY, USA). The first day data were excluded, whereas the following 30-min estimates of activity or sleep were averaged on 4 24-h intervals to obtain measurements at each of 48 time points constituting the 24-hour cycle (Fig. 1).

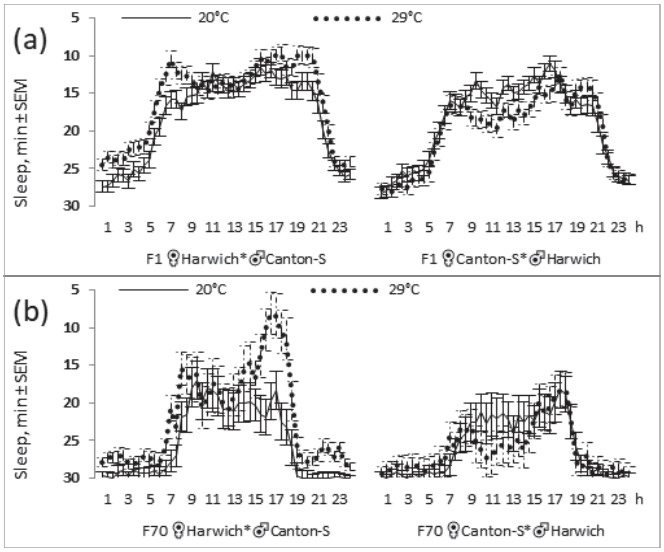


Fig. 1. 24-hour time courses of sleep in hybrid females of opposing cross directions (females from Harwich strain and males from Canton-S strain and females from Canton-S strain and males from Harwich strain). F1 and F70: crosses of either the 1st (a) or 70th generation (b) kept under either low (20°) or high air temperature (29°)

Survival Rate

We also examined differences in survival rate under thermal stress in two parent strains and their F1 or F70 offspring of two cross directions. Flies were placed in tubes and kept at 29°C. Food was replaced with intervals of, at least, 2 days and dead flies were removed and counted. In total, 2000 flies from 100 tubes were counted and mortality analysis was performed using the Kaplan–Meier estimator.

Results

As expected, sterility was found only in F1 females from one of two cross directions: when females from Canton-S strain were hybridized with males from Harwich strain. Moreover, the deleterious effect was found only when eggs were hatched, larvae were reared, and newly emerged flies were permanently kept at 29°C. The main reason for sterility of this female cross was under-development of ovaries.

The crosses demonstrated significant difference in 24-h sleep-wake cycle, and in its response to thermal stress. For example, only fertile F1 females (i.e., progeny of Harwich females crossed with Canton-S males) responded to an elevated temperature by 1) a decrease of night sleep (an insomnia-like symptom), 2) a shift of the second activity peak from evening to early night hours, and 3) an increase of daytime sleep duration (it is typical for males under any temperature). In contrast, sterile F1 females demonstrated only one - the latter - response (see Figure 1, *a*). The same differences in sleep pattern and its stress response were also found in F70 females (see Figure 1, *b*). Longevity analyses yielded significant differences in survival rate indicating that female F1 or F70 offspring of any cross direction were significantly better survivals than females from any of two parent strains and any males.

Discussion

Given the complexity of human reproduction system and numerous limitations imposed on human studies, a fruit fly model was used to investigate the possible association between the predicted effects of thermal stress on the sleep-wake cycle and fertility. Although no evidence was provided for a causal link between temperature-induced sterility and insomnia-like symptoms in females born from females of Canton-S strain and males of Harwich strain, the circadian sleep patterns and their responsiveness to thermal stress were found to be significantly dependent upon cross direction. This dependence was observed in intra-specific hybrids of any sex and persisted in females over generations. Future studies of fruit fly models of temperature-sensitive infertility might be focused on testing possibility of association between the acute rather than developmental thermal effects on fertility and sleep-wake patterns.

Acknowledgment

The fruit fly study of L. P. Z. and D. V. P. was partially funded by the Federal Research Program via project number 0324-2018-0016, and A. A. P. was partially funded by the Russian Foundation for Basic Research through a research grant number 19-013-00424.

References

[1] N. D. White, "Influence of sleep on fertility in women," Am. J. Lifestyle Med., vol. 10, pp. 239-241, 2012.

[2] N. C. Donelson and S. Sanyal, S. "Use of Drosophila in the investigation of sleep disorders," Exp. Neurol., vol. 274 (Pt A), pp. 72–79, 2015.

[3] A. A. Putilov, "A cross-sectional study of retrospectively reported seasonality in native and non-native residents of Chukotka and Turkmenistan," Int. J. Occup. Environ. Health, vol. 24, pp. 17–26, 2018.

[4] K. M. Parisky, J. L. Agosto Rivera, N. C. Donelson, S. Kotecha and L. C. Griffith, "Reorganization of sleep by temperature in *Drosophila* requires light, the homeostat, and the circadian clock", Curr. Biol., vol. 26, pp. 882-892, 2016.

[5] M. G. Kidwell, J. F. Kidwell and J. A. Sved, "Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination," Genetics, vol. 86, pp. 813-833, 1977.

[6] W. R. Engels, "The P family of transposable elements in *Drosophila*," Ann. Rev. Genet., vol. 17, pp. 315-344, 1983.

DOI 10.18699/BGRS/SB-2020-250

A platform for storage and analysis of results of genome-wide association studies of sheep

Alexander S. Zlobin
Kurchatov Genomic Center
of ICG SB RAS
Novosibirsk, Russia
zlobin@bionet.nsc.ru

Anatoliy V. Kirichenko
Laboratory of Recombination and
Segregation Analysis
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia

Tatyana I. Shashkova
Laboratory of Theoretical and Applied
Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Natalya A. Volokova
L.K. Ernst Federal Science Center for
Animal Husbandry,
Dubrovitsy, Moscow Region, Russia

Pavel M. Borodin
L.K. Ernst Federal Science Center for
Animal Husbandry,
Dubrovitsy, Moscow Region, Russia

Lennart C. Karssen
PolyOmica
's-Hertogenbosch, the Netherlands

Yakov A. Tsepilov
Laboratory of Recombination and
Segregation Analysis
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia

Yurii S. Aulchenko
Kurchatov Genomic Center of ICG
SB RAS, Novosibirsk, Russia

Abstract — **We developed a prototype of a platform for aggregation, storage and analysis of the results of genome-wide association scans of different economically important ovine traits.**

Keywords — genome-wide association study; big data; genomic selection; sheep; QTL

Introduction

Sheep breeding is one of the most important branches of agriculture. Currently, the most popular product of sheep breeding is mutton. For this reason, the modern market demands the breeds or breed types of sheep characterized by good viability, high growth rate, good meat qualities, ability to effectively use feed. The solution of the problem of production of pedigree sheep of competitive quality (both on the domestic and world markets) can only be achieved through the shift of selection to a qualitatively new level - from selection by phenotype to the selection at the gene and genomic level. Genome-wide association studies (GWAS) is a modern, agnostic, data-driven method for identifying QTL associated with different traits. The results of genome-wide association studies (RGWAS) are an example of big data that further can be used to build models of marker-assisted and genomic selection and identification of candidate targets for genomic editing. However, the extraction of new knowledge from big data requires its aggregation and the construction of infrastructure for its storage, access and processing. In the context of RGWAS for economically important traits o f ovines, this highly relevant problem has not been solved yet.

Aim: to implement a prototype of a platform for aggregation, storage and analysis of the results of genome-wide association scans (RGWAS) of different ovine traits.

Materials and methods

For searching and filtering papers we used PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) and Google Scholar (http://scholar.google.com). We restricted our search to papers published in English, in peer-reviewed journals since 2013. We chose 2013 as a year when the first biggest GWAS (n>200) for sheep was published [1]. We used a few keyword combinations which included word related to the organism (sheep/ovine/Ovis aires) and word related to the type of study (GWAS). For example, one of the keyword combinations was "sheep GWAS". For each keyword combination, we manually screened and selected for further consideration top 20 most relevant papers. Further, for each study, we have checked the information about the full RGWAS availability. If it was publicly available we have downloaded it. In other cases, we have collected the emails of corresponding authors and contacted them with the request for the full RGWAS.

We used GWAS-MAP platform [4] to consolidate and store ovine RGWAS, and the PheLiGe web-interface to present and visualize data. First, RGWAS are converted into useful format and check correspondence between rs_id and chromosome position based on the reference, and harmonize effect alleles. Second, data should pass quality control (QC). QC contains three steps: comparing allele frequency between RGWAS and reference data, check effect size distribution, and check that reported p-values correspond to z-scores. If QC is fine, RGWAS are saved in the database. For data storage and manipulations on GWAS-MAP platform we use two database management systems, ClickHouse, to store harmonized RGWAS and PostgreSQL for metadata storage. We also present web-interface PheLige to provide user-friendly access to the database and visualize data by interactive plots.

Results

Using 3 different keywords combination (sheep GWAS, ovine GWAS, ovis aries GWAS) we found 86 relevant papers (including duplicates). Next, all papers were manually curated and filtered by relevance. We selected the most relevant studies that led to the final list of 38 publications. About half of the papers described traits related to meat productivity and growth. Other traits were related to wool production, milk production, and some body composition traits (such as root, fat tail, ear size, etc.). Only two studies made the full RGWAS publically available [2,3]. For other studies we have contacted corresponding authors for the data access.

Next, we made the LD reference for sheep based on 96 animals (18 Romanov breed sheep, 6 Katahdin breed sheep, 10 argalis, 48 F1 hybrids Romanov sheep and argali, 14 F1 hybrids Romanov sheep and Katahdin sheep) that were genotyped by Infinium® HD SNP Bead-Chip (Illumina Inc., San Diego, CA, USA). The LD reference contains information about 523578 non-monomorphic SNP.

We have implemented a prototype of a platform for aggregation, storage and analysis of RGWAS of ovine traits. The platform is accessible via public (web-based) and private (ssh-

based) interfaces. At the moment, the database contains 21 millions of associations.

ACKNOWLEDGMENT

AZ analysed the literature, collected the data, performed quality control and developed the content of the database of the results of genome-wide association studies of ovine traits. The work of AZ was funded by the Kurchatov Genomics Center of IC&G (075-15-2019-1662). The work of AK, who was in charge of the software setup, was funded by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genetics SB RAS (project #0324-2019-0040-C-01). The work of TS on GWAS-MAP was funded by PolyKnomics BV and by the Novosibirsk State University via Russian Ministry of Science and Education under the 5-100 Excellence Programme. The work of LK on GWAS-MAP was funded by PolyKnomics BV. The generation of data data for the linkage disequilibrium reference was supported by Russian Science Foundation #18-16-00079 (YT and VIZh). YA and YT jointly supervised the study. The work of YA was funded by the Kurchatov Genomics Center of IC&G. All authors contributed to writing of the manuscript.

REFERENCES

[1] Zhang, L., Liu, J., Zhao, F., Ren, H., Xu, L., Lu, J., Zhang, S., Zhang, X., Wei, C., Lu, G., Zheng, Y. and Du, L.: Genome-wide association studies for growth and meat production traits in sheep., PLoS One, 8(6), e66569, doi:10.1371/journal.pone.0066569, 2013.

[2] Bolormaa, S., Hayes, B. J., van der Werf, J. H. J., Pethick, D., Goddard, M. E. and Daetwyler, H. D.: Detailed phenotyping identifies genes with pleiotropic effects on body composition, BMC Genomics, 17(1), 224, doi:10.1186/s12864-016-2538-0, 2016.

[3] Di Gerlando, R., Sutera, A.M., Mastrangelo, S., Tolone, M., Portolano, B., Sottile, G., Bagnato, A., Strillacc, M.G. and Sardina M.A.: Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. PLoS ONE 14(4): e0215204. https://doi.org/10.1371/journal.pone.0215204, 2019.

[4] Gorev D.D., Shashkova T.I., Pakhomov E., “GWAS-MAP: a platform for storage and analysis of the result of thousands of genome-wide association scans”, Bioinformatics of Genome Regulation and Structure\Systems Biology, The Eleventh International Conference, C. 43. 2018

DOI 10.18699/BGRS/SB-2020-251

Multivariate analysis identify new loci associated with meat productivity and carcass traits in sheeps (*Ovis aries*)

Alexander S. Zlobin
Kurchatov Genomic Center of ICG
SB RAS, Novosibirsk, Russia
zlobin@bionet.nsc.ru

Natalia A. Volkova
L.K. Ernst Federal Science Center
for Animal Husbandry,
Dubrovitsy, Moscow Region, Russia

Pavel M. Borodin
L.K. Ernst Federal Science Center
for Animal Husbandry,
Dubrovitsy, Moscow Region, Russia

Tatiana I. Aksenovich
L.K. Ernst Federal Science Center
for Animal Husbandry,
Dubrovitsy, Moscow Region, Russia

Yakov A. Tsepilov
Laboratory of Theoretical and Applied
Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

***Abstract* — We performed the search of new loci associated with meat productivity and carcass traits in sheep using novel multivariate approaches. We found 4 new loci associated with traits of interest and added more than 30 loci to our previously published database.**

Keywords— sheep, meat productivity traits, carcass traits

Introduction

Sheep breeding is one of the most important branches of agriculture. Currently, the most popular product of sheep breeding is mutton. For this reason, the modern market demands the breeds or breed types of sheep characterized by good viability, high growth rate, good meat qualities, and ability to effectively use feed. The solution of the problem of production of pedigree sheep of competitive quality (both on the domestic and world markets) can only be achieved through the shift of selection to a qualitatively new level - from selection by phenotype to the selection at the gene and genomic level. Genome-wide association studies (GWAS) is a modern, agnostic, data-driven method for identifying QTL associated with different traits. Identification and mapping of genes that affect phenotypic variability in economically important quantitative traits, determining the growth, development and reproductive potential of animals. Unfortunately, today we don't have a full picture of QTLs and genes associated with traits of interest. Filling the gap in our knowledge can be done by performing new GWAS on these traits or reprocessing previously obtained results of GWAS (such as published summary statistics) using novel powerful methods. One of the method to do reprocessing of GWAS results is MANOVA (Multivariate ANalysis Of VAriance). This method allows to find new loci associated with traits of interest using summary statistics. This project focused on application existed methods of multivariate analysis for searching for new loci associated with meat productivity and carcass traits in sheep.

Materials and methods

For this project, we used summary statistics from previously published GWAS for 56 traits associated with meat productivity and carcass traits in sheep.[1] We performed multivariate analysis for 8 new multivariate traits formed from 48 univariate original traits. Each multivariate trait is a combination of several univariate traits grouped based on the similarity effect on the organism. The results of multivariate analysis have greater power that regular univariate analysis and usually depicts loci with high level of pleiotropicity between this group of traits. We used the adapted MANOVA method for z-statistics from Stephens et al. 2013 [2]. Number of SNPs in analisys was more than 450000. The locus was defined as +-250000 bp from top SNP. We considered locus significantly associated in univariate analysis if its p-value was $< 0.05/(N_{snp}*48)$ and in multivariate analysis if its p-value was $< 0.05/(N_{snp}*(48+8))$. We considered locus as new if it didn't present in our previously published database.

Results

We found 4 new loci associated with one multivariate trait related to weight: rs402280290, rs418394153, rs420734786 and rs193632759. These top associated SNPs were missense variant (rs193632759), intron variant (rs420734786) and two - interegenic variants (rs402280290 and rs418394153). The nearest genes were *LOC105603262* (rs402280290), *SETBP1* (rs418394153), *DOCK8* (rs420734786) and *MASP1* (rs193632759). In addition, we found 39 new loci which were associated with multivariate and/or univariate traits and were not presented in our database contained loci associated with meat and growth traits for sheep. [3] As a result of this work we found new loci associated with traits of interest and add these new loci in our database. This information can be useful for further association studies and preliminary estimation of genetic variability for economically important traits in different breeds.

Acknowledgment

The work of AZ was funded by the Kurchatov Genomics Center of IC&G (075-15-2019-1662). The work of NV, PB, TA and YT was supported by Russian Science Foundation #18-16-00079.

References

[1] Bolormaa S. et al. Detailed phenotyping identifies genes with pleiotropic effects on body composition // BMC Genomics. BioMed Central, 2016. Vol. 17, № 1. P. 224.

[2] Stephens M. A Unified Framework for Association Analysis with Multiple Related Phenotypes // PLoS One / ed. Emmert-Streib F. 2013. Vol. 8, № 7. P. e65245.

[3] Zlobin A.S. et al. Recent advances in understanding genetic variants associated with growth, carcass and meat productivity traits in sheep (*Ovis aries*): an update // Arch. Anim. Breed. 2019. Vol. 62, № 2. P. 579–583.

Symposium
Systems biology and biomedicine

DOI 10.18699/BGRS/SB-2020-252

Inflammation is associated with desynchronosis in the immune system (experimental study)

Abdalova A.M.
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
arzuabd@mail.ru

Shurlygina A.V.
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
anna_v_s@mail.ru

Dergacheva T.I.
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
dr-tanja@yandex.ru

Klimontov V.V.
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
klimontov@mail.ru

Letyagin A.Yu.
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
letyagin-andrey@yandex.ru

***Abstract* — The circadian temporary organization of the dynamic processes occurring at the level of the cellular compartment of the immune system ensures their clear time sequence, the optimal ratio of cellular elements in organs and tissues at any given time/ Any pathological process leads to biorhythmological disorders of varying severity, which begin at the very early stages of the disease, when clinical signs are not yet manifested. However, in the literature there is very little information about the daily biorhythms of the immune system in violation of its functions, in particular, in the chronic inflammatory process. We found that with inflammation in the thymus, spleen and lymph nodes, the nature of the diurnal fluctuations in the content of large, medium and small lymphocytes, monocytes / macrophages, CD8 +, CD25 + cells changes. Thus, with the development of inflammation in the uterine mucosa in rats, desynchronosis in the immune system is observed. This allows us to talk about the need to take into account circadian rhythms in the diagnosis and treatment of inflammatory diseases.**

Keywords* — *inflammation, circadian rhythm, thymus, spleen, lymphatic nodes, CD4+ and CD8+ lymphocytes

Introduction

The circadian temporary organization of the dynamic processes occurring at the level of the cellular compartment of the immune system ensures their clear time sequence, the optimal ratio of cellular elements in organs and tissues at any given time [3]. Any pathological process leads to biorhythmological disorders of varying severity, which begin at the very early stages of the disease, when clinical signs are not yet manifested [2]. The study of biorhythms of indicators of the morphofunctional state of immunocompetent cells, their cytokine-producing activity and ability to respond to regulatory factors is of great importance for elucidating the mechanisms of immunopathological processes. However, in the literature there is very little information about the daily biorhythms of the immune system in violation of its functions, in particular, in the chronic inflammatory process.

Methods and Algorithms

In female Wistar rats, a model of experimental endomyometritis was created [1]. On the 13th day after the induction of inflammation, rats were sacrificed by decapitation under ethinal anesthesia at 10.00 and at 20.00. Thymus, spleen, inguinal, iliac and para-aortic lymph nodes were removed. In lymphoid organs, the percentage of CD8 + and CD25 + lymphocytes were determined by flow cytometry with FITC-labeled monoclonal antibodies (BD Pharmingen) using a FACSCalibur flow cytometer (Becton Dickinson). In smears from a cell suspension of lymphoid organs stained by the Romanovsky-Giems method, the percentage of various cell forms — large, medium, small lymphocytes, and monocytes / macrophages — was determined. Statistical processing was performed using the STATISTICA6 software package.

Results

In intact female Wistar rats, the percentage of large, small, CD25 + cells and monocytes / macrophages was increased in the thymus at 10.00 h compared to 20.00 h, the percentage of CD8 + cells were reduced. In the morning, a higher content of large and small lymphocytes and CD25 + splenocytes was noted in the spleen. The percentage of monocytes / macrophages was increased at 20.00 h. Morning-evening differences were revealed in all lymph nodes for the percentage of secondary lymphocytes (Table 1).

With inflammation morning-evening differences were revealed only for the percentage of large and medium thymocytes. In the spleen for large lymphocytes and CD25 + splenocytes an “inversion” of diurnal fluctuations was found, and for the percentage of monocytes / macrophages a decrease in the difference between morning and evening values was found. Significant diurnal fluctuations in the percentage of medium and small lymphocytes disappear in the paraortic lymph nodes, but morning and evening differences for the content of CD8 +, CD25 + cells and monocytes / macrophages appear. In the inguinal lymph nodes, the morning and evening differences in the content of middle lymphocytes are leveled, but they appear for the percentage of large, small lymphocytes and CD8 + cells. Significant diurnal fluctuations in the percentage of large lymphocytes, CD8 + cells and monocytes / macrophages appear in the iliac lymph nodes. Daily variations in the content of middle lymphocytes are inverted (Table 1).

TABLE I. DAILY VARIATIONS IN THE CELLULAR COMPOSITION OF THE LYMPHOID ORGANS OF FEMALE RATS INTACT AND ON THE 13TH DAY OF THE DEVELOPMENT OF EXPERIMENTAL ENDOMYOMETRITIS (M ± SE)

Cells (%)	Int. 10.00 h	Int. 20.00 h	Inf. 10.00 h	Inf. 20.00 h
Thymus				
Large lymphocytes	25,32± 2,21	14,51± 1,87*	21,2 ± 1,02	13,3 ± 1,93*
Middle lymphocytes	21,81± 2,22	40,52± 3,01	17,6 ± 1,43	34,1 ± 2,59*
Small lymphocytes	58,24± 1,28	31,25± 0,26*	54,2 ± 2,02	41,8 ± 3,81
CD8+	48,12± 0,11	69,45± 0,08*	59,6 ± 0,08	70,1 ± 0,07
CD25+	9,21±0,03	4,35± 0,01	6,13 ± 0,02	5,41 ± 0,01
Monocytes/macrophages	4,52±0,12	1,12± 0,02*	3,2 ± 0,54	1,1 ± 0,03
Spleen				
Large lymphocytes	5,81±0,13	2,45± 0,08*	1,0 ± 0,54	6,2 ± 0,43*
Middle lymphocytes	4,87±0,14	9,48±0,95	5,3 ± 2,07	10,1 ± 1,02
Small lymphocytes	88,45± 2,05	65,14± 1,02*	89,0 ± 1,77	79,2 ± 1,65
CD8+	38,51± 1,02	54,25± 0,47	45,4 ± 0,09	49,9 ± 0,09
CD25+	7,54±0,01	3,28±0,01*	4,01 ± 0,001	6,46 ± 0,01*
Monocytes/macrophages	2,45±0,05	6,48±0,08*	3,2 ± 0,81	5,2 ± 0,96*
Paraortal lymph node				
Large lymphocytes	8,45±1,24	6,74±2,14	9,1 ± 0,73	5,6 ± 0,71
Middle lymphocytes	18,36± 1,05	8,64± 0,98*	11,2 ± 1,93	9,9 ± 1,02
Small lymphocytes	32,14± 1,52	69,87± 0,18*	46,0 ± 2,57	63,1 ± 1,93
CD8+	39,18± 0,87	43,87± 5,25	45,2 ± 0,38	37,6 ± 0,74*
CD25+	2,57±0,01	3,12±0,05	4,01 ± 0,11	2,83 ± 0,03*
Monocytes/macrophages	6,14±1,24	8,36±2,78	5,4 ± 0,08	9,1 ± 0,03*
Lymph node inguinal				
Large lymphocytes	8,69±1,32	7,54±2,12	9,6 ± 0,74	4,0 ± 0,94*
Middle lymphocytes	20,87± 1,24	10,14± 1,47*	18,7 ± 1,05	12,5 ± 1,12
Small lymphocytes	58,78± 3,01	54,21± 2,14	36,8 ± 2,95	56,71 ± 1,03*
CD8+	39,18± 2,01	45,11± 3,58	42,8 ± 0,56	34,8 ± 0,08*
CD25+	4,16±0,03	3,87±1,02	5,61 ± 0,06	4,21 ± 0,03
Monocytes/macrophages	15,25± 1,21	22,15± 3,45	12,3 ± 0,78	7,5 ± 1,01
Iliac lymph node				
Large lymphocytes	2,24±0,14	4,12±2,25	1,3 ± 0,05	3,2 ± 0,09*
Middle lymphocytes	9,51±0,14	4,25±0,02*	2,9 ± 0,57	9,2 ± 0,31*
Small lymphocytes	65,13± 3,24	59,25± 1,12	51,9 ± 2,07	68,1 ± 1,12
CD8+	29,15± 0,14	32,12± 1,25	39,7 ± 0,38	41,5 ± 0,18*
CD25+	3,12±0,04	1,25±0,02	5,16 ± 0,08	5,73 ± 0,06
Monocytes/macrophages	10,41±1,12	6,15±2,87	13,0 ± 0,95	3,1 ± 0,07*

Note: Int. - intact, Inf. - inflammation, * - significant differences from 10.00 h of the corresponding group ($p<0,05$; Mann-Whitney test)

Conclusion

With the development of inflammation in the uterine mucosa in rats, desynchronosis in the immune system is observed. This allows us to talk about the need to take into account circadian rhythms in the diagnosis and treatment of inflammatory diseases.

References

[1] E.V. Starkova, T.I. Dergacheva, V.V. Astashov. A method for modeling inflammatory diseases of the female genital organs. Patent for invention RUS 2142163 07/17/1996

[2] V.A. Trufakin, A.V. Shurlygina, S.V. Michurina The lymphoid system is a circadian temporary organization and desynchronosis. Bulletin of the Siberian Branch of the Russian Academy of Medical Sciences. 2012, 32 (1), pp 5-12

[3] K Suzuki, Y Hayano, A Nakai, F Furuta, M Noda. Adrenergic control of the adaptive immune response by diurnal lymphocyte recirculation through lymph nodes. J. Exp. Med, 2016. 213 (12), pp 2567-2574.

DOI 10.18699/BGRS/SB-2020-253

Endoglycosidase expression in pubocervical fascia is up-regulated in menopause patients with severe pelvic organs prolapse

Svetlana V. Aidagulova
Novosibirsk State Medical University
Novosbirsk, Russia
s.aydagulova@gmail.com

Fedor A. Rakitin
Research Institute of Clinical and Experimental Lymphology - Branch ICG SB RAS
Novosibirsk, Russia
rakitinfedorr@mail.ru

Mikhail Yu. Soluyanov
Research Institute of Clinical and Experimental Lymphology - Branch ICG SB RAS
Novosibirsk, Russia
msoluyanov@mail.ru

Vadim V. Nimaev
Research Institute of Clinical and Experimental Lymphology – Branch ICG SB RAS
Novosibirsk, Russia
nimaev@gmail.com

Igor O. Marinkin
Novosibirsk State Medical University
Novosbirsk, Russia
rectorngmu@yandex.ru

***Abstract* — Pelvic organs prolapse in women is very inconvenient age-dependent pathology. One of the causes of pelvic organs prolapse is considered so called undifferentiated dysplasia of connective tissue being genetically determined change in the morphogenesis of connective tissue. The aim of this work was to study heparanase expression in pubocervical fascia of menopause women with pelvic organ prolapse. As shown heparanase protein content was significantly up-regulated in cellular compartment of pubocervical fascia in patients with pelvic organs prolapse and prominent undifferentiated dysplasia of connective tissue degrees.**

Keywords* — *pelvic organs prolapse, undifferentiated dysplasia of connective tissue, endoglycosidase, pubocervical fascia, menopause women

Motivation and Aim

Pelvic organs prolapse (POP) in women is very inconvenient age-dependent pathology. The incidence of women with POP in USA and other countries is reaching up to 50% in patients aged 80 years [1]. The understanding of POP's pathophysiology is important for prevention and treatment. Presumably POP deals with so called undifferentiated dysplasia of connective tissue (UCTD). UCTD is genetically determined change in the morphogenesis of connective tissue due to abnormalities of fibrillogenesis and non-fibrillar extracellular matrix (ECM). In POP it was known the expression changes mainly focusing on the genes/proteins of collagen, elastin, matrix metalloproteinases (MMPs) and their tissue inhibitors [2]. Besides MMPs the ECM also contains at least one's more enzyme involved in connective tissue remodeling, which has an endoglycosidase activity and is named heparanase (HPSE). The aim of this work was to study HPSE expression in pubocervical fascia of menopause women with POP.

Methods and Algorithms

This study was performed on 150 women 60,7±7,73 years old with POP, the diagnosis was based on the POP Quantification (POP-Q ISC, 1996). The average duration of menopause was 10,9±9,1 years. Women with malignances and/or active/latent urinary tract infection were excluded. All 150 patients were suitable for POP-Q III-IV (Ba) and underwent primary scheduled treatment using vaginal access surgery; also in all cases the informed consent of the patient to the examination and treatment of POP was taken in accordance with the directives of the European Community (86/609 / EEC) and the Helsinki Declaration, in compliance with the «Ethical Principles for Scientific Medical Research with Human Participation» and in accordance with the «Rules of Clinical Practice in Russian Federation». Patients were divided into 2 representative by age and by POP's severe groups, depending on the UCTD degree. The 1st group consisted of 48 women with mild degree of UCTD. The 2nd group consisted of 102 women with moderate and severe degree of UCTD. Phenotypic manifestations of UCTD were evaluated on the basis of physical examination and ultrasound study of the heart and pelvic organs, focusing on pathological renal mobility. The minimal manifestations of UCTD included joint hypermobility, mild skin striae and/or myopia. Their combination with signs of the «MASS phenotype» (mitral valve prolapse, additional heart chords, scoliosis, and so on) were interpreted as moderate and severe degree of UCTD [3]. Surgical samples of the pubocervical fascia were fixed in 10% buffered formalin solution. For immunohistochemistry, 5-mkm sections of formalin-fixed, paraffin-embedded tissue sections were deparaffinized and antigens were retrieved in sodium citrate buffer (10 mM sodium citrate, 0,05% Tween-20) at 95-98°C for 20 min. The rabbit polyclonal anti-HPSE (Abcam, cat. № ab85543,1:100) primary antibodies were used. For negative control, the primary antibodies were replaced by 5% bovine serum. HPSE-immunostaining patterns were visualised using Histostain-Plus 3rd Gen IHC Detection Kit (ThermoFisher Scientific). The sections were counterstained with Hematoxylin and observed by light microscopy using Axio Scope.A1 microscope with the camera AxioCam MRc5 and software ZEN blue for the quantitative analysis of 30 images per group with magnification 40x10 (Zeiss, Oberkochen, Germany). Statistical processing of the results was performed using the statistical software application package STATISTICA v.6.0. A value of $p < 0,05$ was considered to indicate a statistically significant difference.

Results

An analysis of the incidence of clinically polymorphic symptoms and diseases inherent to the UCTD in 150 women of 1st and 2nd groups have revealed the genital prolapses and hernia as universal UCTD manifestations in first-line relatives of 40% and 49% cases, respectively. In women of the 2nd group, varicose veins occurred in 79,4% comparing 33,3% cases in the 1st group. And in contrast to the 1st group, in the 2nd group, the biliary dyskinesia was detected in 45% of women, and heart rhythm and conduction disturbances were revealed in 48% ones. The investigation a hypothesis on a potential negative effect of HPSE content on the POP pathogenesis, the dense fibrous connective tissue of pubocervical fascia was studied with emphasis on the intra- and extracellular endoglycosidase localization. HPSE protein content was significantly up-regulated in cellular compartment of pubocervical fascia of the 2nd group with prominent UCTD degrees compared with the 1st one: the average area of total positive immunostaining products amounted as 2,4±0,6 mm2 versus 1,3±0,5 mm^2 ($p < 0,01$).

Conclusion

HPSE known as heparan sulfate cleaver and active participant in different pathological processes such as inflammation and tumorigenesis may affect the ECM properties of pubocervical fascia and should serve as new molecular marker of UCTD and POP.

References

[1] Nygaard I. et al. (2008) Prevalence of symptomatic pelvic floor disorders in US women. JAMA. 300: 1311–1316.

[2] Lim V.F. et al. (2014) Recent studies of genetic dysfunction in pelvic organ prolapse: the role of collagen defects. Aust N Z J Obstet Gynaecol. 54: 198–205.

[3] Smolnova T.Yu. et al. (2003) The phenotypical symptom complex of connective tissue dysplasia in females. Clin. Med. 8: 42-47.

DOI 10.18699/BGRS/SB-2020-254

Computational prediction of miRNA binding sites in mRNA of colorectal cancer candidate genes

Aigul Akimniyazova
SRI of Biology and Biotechnology problems
al-Farabi Kazakh National University
Almaty, Kazakhstan
email: akimniyazova@gmail.com

Abstract — The identification of causative miRNAs in colorectal cancer is unclear due to a lack of understanding of how specific miRNAs affect biological pathways and consequences of disease development. The miRNA is a class of nano-sized non-coding RNAs that regulate many mRNAs, and mRNA can be related to many miRNAs, which make them suitable for diagnosis purposes. Therefore it is required to identify candidate genes of colorectal cancer and to what extent they can interact with miRNA. To determine the important miRNAs biding sites in genes, involved in the development of colorectal cancer, there were used the MirTarget program. The paper presents the results of studying the characteristics of the interaction of miRNAs with mRNAs of 135 candidate genes involved in the development of colorectal cancer. The binding sites of 446 miRNAs have in 113 mRNA of genes at 5'UTR, CDS, and 3'UTR. Found miRNA binding sites with overlapped nucleotide sequences (clusters). The research results are useful for the development of methods for early diagnosis of this disease.

Keywords — miRNA, mRNA, genes, colorectal cancer

Introduction

Colorectal cancer is one of the three most common oncological diseases worldwide and has a high mortality rate [1-2]. The complexity of early diagnosis of the disease lies in its polygenic nature. Colorectal cancer is accompanied by a change in the concentration of miRNAs, which can alter the expression of candidate genes related with the disease. The effect of miRNA on cancerogenesis is described, however it has not been established which target genes these miRNAs can directly affect [3-4]. The purpose of this study was to identify the interactions between human miRNAs and 135 colorectal cancer candidate gene mRNAs.

Materials and Methods

The nucleotide (nt) sequences of 135 colorectal cancer candidate genes were downloaded from GenBank (http://www.ncbi.nlm.nih.gov). The nucleotide sequences of 2,565 miRNAs were taken from miRBase, and 3,707 miRNAs from a previous study [5]. The quantitative characteristics of these interactions were determined using the MirTarget program created in our laboratory [6-7]. This program defines the start of miRNA binding sites in mRNA; localization of binding sites in 5'-untranslated region (5'UTR), protein coding region (CDS), and 3'-untranslated region (3'UTR); free energy of interaction (ΔG, kJ/mole) and scheme of miRNA-mRNA nucleotides (nt) interaction. The ΔG/ΔGm (%) ratio was calculated for each binding site, where ΔGm is equal to the free energy of miRNA interaction with fully complementary nucleotide sequence.

Results

The binding sites of 446 miRNAs in 113 candidate gene mRNAs were determined considering their expression. Significant differences were found in the characteristics of the miRNA interactions in the 5'UTR, CDS and 3'UTR of mRNA candidate genes. The features of the miRNA binding sites have been established depending on their location in the 5'UTR, CDS and 3'UTR. The miRNA binding sites with overlapping nucleotide sequences that form clusters were identified. This organization of binding sites leads to compaction and competition between miRNAs for binding in the cluster. The most effective associations between miRNAs and candidate target genes, which are proposed as markers for the development of methods for the early diagnosis of colorectal cancer, are determined. Studying the interactions between miRNAs and mRNAs was carried out with colorectal cancer candidate genes with RPKM (Reads per kilo base per million mapped reads) expression values up to ten and greater than ten, considering the localization of miRNA binding sites in the 5`UTR, CDS or 3`UTR. The adequate prediction of miRNA binding sites in target gene mRNAs is a key problem for studying the role of miRNAs in the regulation of gene expression.

Characteristics of miRNA interactions with 5`UTR mRNAs of candidate colorectal cancer genes.

Five binding sites in 5'UTR in mRNA of *DCC* gene was revealed. The mRNA of *FMNL3* gene was target of four miRNAs, three of them forms the cluster of binding sites from 51 nt to 77 nt with an average ΔG value equal to -124 kJ/mole. ID01310.3p-miR and ID03332.3p-miR forms a cluster of binding sites in mRNA of *KRAS* gene in position from 17 nt to 60 nt. *SMAD2*, *SMAD4* and *SOCS6* genes had binding sites for miRNAs were located along the entire length of the 5'UTR mRNA of genes without overlapping nucleotide sequences. The miR-6786-5p and ID03064.3p-miR binding sites in the mRNA of the *SMAD7* gene overlap on seven nucleotides, which makes simultaneous binding of these miRNAs impossible. ID03064.3p-miR has the advantage of binding to mRNA, since it has a large free interaction energy compared to miR-6786-5p. mRNA of *CD44* and *CTNNB1* were a targets of two miRNAs for each, that contain independent from each other binding sites that not overlap their nucleotide sequence. ID00894.5p-miR and ID02781.3p-miR forms a cluster of binding sites in mRNA of *CDH1* gene from 57 nt to 79 nt. The mRNA of *FN1* gene had miRNA binding sites that were located in 5'UTR and CDS. These binding sites did not overlap, and each miRNA could independently bind to the mRNA of the *FN1* gene. mRNA of *GSTP1* gene was target only of two miRNAs (ID00267.3p-miR and ID03331.3p-miR), that form a cluster from 77 nt to 113 nt with an average free energy of interaction equal to -122 kJ/mole. miR-6789-5p, ID01675.5p-miR and ID02822.5p-miR forms cluster of binding sites in mRNA of *HIF1A* gene from 54 nt to 89 nt with an average ΔG value equal to -127 kJ/mole. miR-6789-5p in this cluster have a prevalent chance to take this binding site because of higher free energy of interaction equal to -132 kJ/mole and ΔG/ΔGm value equal to 90%. The mRNAs of *SFRP2*, *STAT3* and *TGFBR1* genes had clusters of miRNA

binding sites. ID01777.3p-miR, ID00849.3p-miR and ID01545.3p-miR forms cluster of binding sites in mRNA of *ZEB1* gene from 1 nt to 28 nt with an average ΔG equal to -116 kJ/mole.

Characteristics of miRNA interactions with CDS mRNAs of candidate colorectal cancer genes.

mRNAs of ADAMTS4, BCL2, BIRC5, CYP1B1, DNMT1, ESR1, KSR1, MSH6, PIK3CG, SMAD7 and *TNFRSF4* genes had binding sites only for two miRNAs in CDS. 21 miRNAs form a cluster of binding sites n mRNA of *GSK3B* gene. The *GSK3B* gene mRNA included a cluster of binding sites ranging from 3 nt to 38 nt, with an entire length of 36 nt and an average ΔG of -126 kJ/mole. These 22 binding sites have a total length equal to 750 nt, which is 20 times larger than the cluster length. The formation of these clusters of binding sites in the *GSK3B* gene 5'UTR indicates the greater ability of this gene for compaction, which causes competition between the given miRNAs for the binding site. The mRNA of *ONECUT2, PIK3CA* and *TCF4* also found big cluster of binding sites. These binding sites did not overlap and each miRNA could independently bind to mRNA of given genes. ID01242.3p-miR, ID00306.5p-miR and miR-511-5p had binding sites in mRNA of *ADAMTS5* gene. mRNA of *OGG1* gene was target of three miRNAs, two of which form the cluster of binding sites from 443 nt to 462 nt. CDS mRNA of *POLD1 gene* contain binding sites for five miRNAs, among them, ID03114.3p-miR and miR-629-3p forms a cluster of binding sites from 1069 nt to 1094 nt with an average ΔG= -112 kJ/mole. Three miRNAs had binding sites in CDS mRNA of *POLE* gene. mRNA of *ALDH2* gene contains four miRNAs binding sites, three of them form a cluster in position from 153 nt to 191 nt with an average ΔG value equal to -115 kJ/mole. Four miRNAs had binding sites in mRNA of *CDH1* gene. miR-6798-5p, miR-3909 and ID02852.5p-miR had binding sites in mRNA of *COL3A1* gene. mRNA of *ERBB2* gene have a cluster of binding sites from 3815 nt to 3838 nt with an average ΔG = -116 kJ/mole. mRNA of *HMGB1* gene was target of three miRNAs, two of which forms a cluster from 746 nt to 770 nt. ID01524.3p-miR and ID03238.3p-miR form a cluster of binding sites in mRNA of *MTA1* gene from 406 nt to 444 nt with an average ΔG = -115 kJ/mole and also have binding sites for ID01593.5p-miR and ID01702.3p-miR in CDS. ID00645.5p-miR and miR-6752-5p form a cluster of binding sites in mRNA of *MUC1* gene from 507 nt to 531 nt with an average ΔG = -114 kJ/mole. mRNA of *SEPT9* gene was target of three miRNAs, two of which form a cluster of binding sites from 1350 nt to 1381 nt. *SFRP2*, *THBS2* and *VEGFA* had binding sites for two miRNAs.

Characteristics of miRNA interactions with 3`UTR mRNAs of candidate colorectal cancer genes

The miR-1273g-3p and miR-1273d had binding sites in mRNA of *ALOX15* gene in 3'UTR. These binding sites did not overlap, and each miRNA could independently bind to this mRNA. The mRNA of *FMNL3* gene was target of four miRNAs. The mRNA of *KRAS* gene was a target of four miRNAs, two of which form a cluster of binding sites from 3163 nt to 3176 nt with an average ΔG value equal to -115 kJ/mole. Five miRNAs had binding sites in mRNA of *MTAP* gene, and among them ID01836.5p-miR and miR-1285-5p form a cluster of binding sites in position from 2530 nt to 2566 nt with an average ΔG = -112 kJ/mole. mRNA of *BCL2L1* gene was a target of two miRNAs in 3'UTR. miR-1273c and miR-1273g-3p form a cluster of binding sites in position from 3251 nt to 3291 nt. miR-619-5p has a binding sites in mRNA of *ERBB3* gene. Previously we investigated the variation of nucleotide sequences of these two miR-619-5p binding sites in the 3'UTR of mRNAs of *ERBB3* gene in orthologs [8]. It is found that the binding site with 100% score contains highly conserved nucleotide sequence GGCTCATGCCTGTAATCCCAGC of miR-619-5p binding sites. The mRNA of *FASN* and *GREM1* contains by two binding sites for miRNAs in 3'UTR. ID00791.5p-miR and miR-6829-5p forms a cluster of binding sites in *MTA1* gene from 2360 nt to 2392 nt with an average ΔG= -116 kJ/mole. Four miRNAs had binding sites in mRNA of *PIK3CG* gene. The miR-1277-5p had three binding sites that does not overlap in mRNA of *VEGFA* gene.

Conclusion

It was found that miRNA binding sites in mRNA of target genes were located in all mRNA regions. Clusters of miRNA binding sites have been identified in some genes. The obtained results show that the interaction of the miRNAs with mRNAs can serve as the basis for the selection of associations for the diagnosis of colorectal cancer.

Acknowledgment

The work was performed within the framework of the grant project AP05132460 financed by the Ministry of Education and Science of the Republic of Kazakhstan

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, RL. Siegel, LA. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA Cancer J Clin., vol. 68(6), pp. 394-424, November 2018.

[2] M. Arnold, MS. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, "Global patterns and trends in colorectal cancer incidence and mortality," Gut, vol. 66, pp. 683–91, 2017.

[3] B. Chen, Z. Xia, YN.Deng, et al, "Emerging microRNA biomarkers for colorectal cancer diagnosis and prognosis," Open Biol., vol. 9(1), 2019.

[4] A.W. Tong and J. Nemunaitis, "Modulation of miRNA activity in human cancer: a new paradigm for cancer gene therapy?" Cancer Gene Therapy, vol. 15, pp. 341–355, 2008

[5] E. Londin, P. Loher, A.G. Telonis, K. Quann, P. Clark, Y. Jing et al, "Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs," PNAS USA, vol. 112, pp. 1106-1115.

[6] A. Ivashchenko, O. Berillo, A. Pyrkova, R. Niyazova, Sh. Atambayeva, "The Properties of Binding Sites of miR-619-5p, miR-5095, miR-5096, and miR-5585-3p in the mRNAs of Human Genes," BioMed research international, 2014

[7] A. Ivashchenko, O. Berillo, A. Pyrkova, R. Niyazova, "Binding Sites of miR-1273 Family on the mRNA of Target Genes," BioMed research international, 2014.

[8] S. Atambayeva, R. Niyazova, A. Ivashchenko, A. Pyrkova, I. Pinsky, A. Akimniyazova, S. Labeit, "The Binding Sites of miR-619-5p in the mRNAs of Human and Orthologous Genes," BMC Genomics, vol. 18, pp. 428.

DOI 10.18699/BGRS/SB-2020-255

Associations of methylation level of promoter region of *MLH1* gene and genotypes of the exonic rs1799977

Nadezhda Babushkina
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
nad.babushkina@medgenetics.ru

Anton Markov
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
anton.markov@medgenetics.ru

Irina Goncharova
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
irina.goncharova@medgenetics.ru

Ramil Salakhov
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
ramil.salakhov@medgenetics.ru

Yuliia Koroleva
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
yuliya.koroleva@medgenetics.ru

Anna Postrigan
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
postrigan.anna@medgenetics.ru

Aleksei Zarubin
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
aleksei.zarubin@medgenetics.ru

Aleksei Sleptcov
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
alexei.sleptcov@medgenetics.ru

Aksana Kucher
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
aksana.kucher@medgenetics.ru

Maria Nazarenko
Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia
maria.nazarenko@medgenetics.ru

***Abstract* — The eQTL SNP in exon 8 of the *MLH1* gene – rs1799977, may affect the methylation level of the promoter region of the same gene. When studying both the effects of methylation level and individual SNPs on the formation of phenotypic variability in humans and the risk of developing pathological conditions, it is necessary to use a integrated data analysis approach.**

Keywords — MLH1, SNP, CpG-sites, methylation level

Motivation and Aim

DNA methylation is one of the most studied epigenetic mechanisms regulating the transcriptional activity of chromatin. Through methylation/demethylation of DNA, a number of fundamental biological processes are regulated, so the study of methylation profiles in different tissues, in different pathologies, in dynamics in ontogenesis, etc., is currently actively carried out. DNA methylation levels are known to be sensitive to a wide range of external influences (diet, drug intake, level of physical activity, and various environmental factors). But in addition, DNA methylation may also depend on structural variations of the genome. Several studies have shown that the level of DNA methylation depends, for example, on cis-localized regulatory SNP [1].

In this work, a pilot study of the methylation level of the studied regulatory regions of the *MLH1* gene versus genotypes was conducted for one of the SNP eQTL in the gene. *MLH1* encodes one of the miss-match repair system proteins, involved in many biological processes. Methylation of DNA in the region of the *MLH1* gene, has variable character and significantly changes in response to various substances. The significance of methylation level changes is proved in oncogenesis [2]. A direct relationship between an increase in the level of methylation of the *MLH1* gene and a decrease in expression at the mRNA level has been proven [3].

Studing of the dependence of the he methylation level of the promoter region *MLH1* gene with genotypes on the regulatory exonic SNP of the same gene.

Methods

Methylation was studied in leukocytes of peripheral blood of 22 patients with atherosclerosis of carotid arteries (stenosis level >70%) and 14 healthy individuals (stenosis level <28%). The groups examined are comparable by sex and age, represented by Russian residents of Tomsk. Informed agreement on the purpose and possible risks of the study was signed by all participants.

The *MLH1* gene promoter was studied, which is the “coast” of the CpG-island (within ± 2000 bp from the CpG-island), capturing also a fragment of exon 1 of the *EPM2AIP1* gene (GRCh37/hg19; chr 3: 37033249-37033762).

The genotyping rs1799977 was performed by SNaPshot analysis using Multiplex Kit and Primer Focus kit SNaPshot® ("Applied Biosystems," USA), on the ABI Genetic Analyzer 3730 platform. The analysis of the CpG-sites of the MLH1 gene promoter was performed using the bisulfite method sequencing using the technology of high-performance mass parallel sequencing (massive parallel sequencing, MPS) on the MiSeq platform (Illumina, USA); Bisulfite DNA treatment was performed using the EZ DNA Methylation Kit ("Zymo Research," USA); DNA library preparation for MPS - Nextera DNA Flex Library Prep Kit ("Illuminina," USA) [4]. The study was conducted on the basis of the Center for the collective use of scientific research equipment and experimental biological material “Medical genomics” at the Research Institute of Medical Genetics, Tomsk Scientific and Research Center, RAS.

Bioinformatic processing of the results was carried out using the BWA-Meth programs (for mapping readings) [https://github.com/brentp/bwa-meth], MethylDackel (for extracting information on methylation of CpG-sites), as well as in software methylKit package in the environment R. The

coverage level of CpG-sites with read sequences ranged from 300 to 1100 (with an average reading level of 900). The methylation level of the CpG-site was determined as the percentage of readings corresponding to methylated cytosine residues relative to all readings of cytosine at a given genome position.

Non-parametric tests Kruskall-Wallis, Wilcoxon and Mann–Whitney were used for differentially methylated CpG-sites. Statistically significant differences with achieved level of statistical significance $p<0.05$ were considered, of false-positive control results was carried out by the method of Benjamin-Hochberg.

Results

In the regulatory region of the *MLH1* gene, 9 CpG-sites were analyzed (site coordinates on chromosome 3 (according to assembly GRCh37/hg19): chr3: 37033323-37033662). The "pattern" of methylation of CpG-sites in the studied region of the MLH1 promoter was similar in all samples: the 2nd (chr3: 37033373) and 5th (chr3: 37033562) CpG-sites with a methylation level reduced relative to other CpG-sites flanked areas with higher methylation. This "pattern" approximately repeated the shape of the distribution density of the modified histone H3K27Ac sites in different cell lines.

Significant ($p<0.03$) associations of methylation level with genotypes in rs1799977 were recorded for all analyzed CpG-sites, the differences being more pronounced in the control group. In all cases, carriers of genotype AG had higher levels of DNA methylation (on average about 10%) than carriers of genotype AA. Only one of those examined (from the group of patients with atherosclerosis) was a carrier of the GG genotype, its methylation level of all sites being higher than that of carriers of other genotypes. Single observation cannot make a clear conclusion about the effect of the genotype on the level of DNA methylation. However, this fact is worth noting because it is consistent with the general trend of the level of CpG-site methylation depending on the number of G alleles in the genotype of the surveyed (AA<AG<GG).

When comparing DNA methylation levels between groups of patients and healthy individuals, no statistically significant differences were detected, either in general or taking into account genotypes. At once, for all the studied CpG-sites, AA genotype carriers in the control group had lower methylation levels than patients, while AG genotype carriers in the control group had higher methylation levels than patients. Thus, differences in DNA methylation levels in healthy individuals with different genotypes in rs1799977 significantly exceed those in atherosclerosis patients (5-13% for different sites, $p = 0.00004$).

How the methylation level of the promoter region of a gene (chr3: 37033249-37033762) is associated with genotypes for rs1799977 - a missense substitution in exon 8 (chr3: 37053618), remains unclear. There are two possible explanations for this fact. First, the studied SNP may be in linkage disequilibrium with any SNP in the regulatory region of this gene. Indeed, rs4647200 is resided in the region of the studied CpG-sites (chr3: 37033462). Analysis of bisulfite sequencing data revealed 4 heterozygous and one homozygous carrier of rare allele G by rs4647200 (all in the group of patients with atherosclerosis only). When analyzing bisulfite sequencing data, 4 heterozygous and one homozygous carriers of the rare G allele were identified by rs4647200 (all in the group of patients with atherosclerosis only). There is linkage disequilibrium between the markers rs4647200 and rs1799977 (D'= 1.000, r2 = 0.377). Second, according to the "GeneHancer" database, both the promoter region of the *MLH1* gene and rs1799977 are located in the sites of interaction with the transcriptional factor LRRFIP2 [http://www.genome.ucsc.edu/], which may cause of functional linkage between them. The findings are new information and require further study.

Thus, a pilot study found that rs1799977 in the exon 8 of the *MLH1* gene could have an effect on the methylation level in the promoter region of the same gene. This, in turn, indicates the relevance of using an integrated approach in the analysis both the effects of methylation levels and individual SNPs on the formation of phenotypic variability in humans and the risk of developing of pathological conditions.

Acknowledgment

The work was performed as part of the State Assignment of the Ministry of Science and Higher Education No. 075-00603-19-00.

References

[1] Ahsan M. et al. (2017) The relative contribution of DNA methylation and genetic variants on protein biomarkers for human diseases. PLoS Genet. 13(9): e1007005.

[2] Hitchins M.P. et al. (2011) Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. Cancer Cell. 20(2): 200–213.

[3] Wu C.X. et al. (2018) Changes of DNA repair gene methylation in blood of chronic fluorosis patients and rats. J Trace Elem Med Biol. 50: 223-228.

[4] Masser D.R. et al. (2015) Targeted DNA methylation analysis by Next-generation sequencing. J Vis Exp. – 96: e52488/

DOI 10.18699/BGRS/SB-2020-256

Influence of the immobilized subtilisins on perfomance indicators of the rat heart in experiment

German Baikalov
Department of experimental pharamacology
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences Novosibirsk
Novosibirsk, Russian Federation
Department of pharmacology, clinical pharmacology and evidence based medicine
Novosibirsk State Medical University
Novosibirsk, Russian Federation
gbaikalov@yandex.ru

Roman Knyazev
Research Institute of Biochemistry – Branch of the Federal Research Center of Fundamental and Translational Medicine
Department of pharmacology, clinical pharmacology and evidence based medicine
Novosibirsk State Medical University
Novosibirsk, Russian Federation
knjazev_roman@mail.ru

Pavel Madonov
Department of experimental pharamacology
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russian Federation
Department of pharmacology, clinical pharmacology and evidence-based medicine
Novosibirsk State Medical University
Novosibirsk, Russian Federation
pmadonov@yandex.ru

Konstantin Ershov
Department of pharmacology, clinical pharmacology and evidence-based medicine
Novosibirsk State Medical University
Novosibirsk, Russian Federation
ershov_k@bk.ru

***Abstract* — The immobilized subtilisins are bacterial proteinases with a high thrombolytic activity, capable of improving peripheral blood circulation by dissolving the blood clot and reducing manifestations of endothelial dysfunction. This article presents the materials of the effect of immobilized subtilisins on performance heart indicators in the experiment. The action of these proteinases was studied on a model of isolated rat heart according to Langendorff. Indicators of heart rate (HR), pressure developed by the left ventricle (LVP), heart efficiency (HE) were selected for registration. It was shown when adding the immobilized subtilisins, an increase of the left ventricular contraction force and a decrease in a heart rate were observed. The results indicate that immobilized subtilisins have positive inotropic and negative chronotropic effects.**

Keywords — immobilized subtilisins, isolated heart, heart rate, left ventricular pressure, heart efficiency

Introduction

Currently, the diseases of the circulatory system are an highly important problem and a cause of death worldwide. The largest percentage among them is coronary heart disease, including angina pectoris, myocardial infarction and sudden cardiac death. As a result of this, the search of effective pharmacological preparations for the correction and treatment of cardiovascular diseases is an urgent area.

The proteolytic enzymes have been used by humans for many centuries, but their scientifically reasoned use began in the early twentieth century after the publication of fundamental works by E. Fisher. The proteolytic enzymes can be obtained from plant materials (e.g., papain), from animal raw materials (e.g., trypsin), or from microorganisms. The subtilisins are proteolytic enzymes from the class of hydrolases, belong to the class of serine proteases and they are classified according to genera in accordance with the homology of the amino acid sequence. The name subtilisin comes from the fact that most of these enzymes are produced by the bacteria Bacillus subtilis. The possibility of the clinical use of subtilisins has been studied quite actively in recent years, and studies show their effectiveness not only for the treatment of thrombosis, but also for the prevention of cardiovascular diseases [1, 2].

There is no data in the literature about the effect of subtilisins on heart function. There are various models for experimental studies in the cardiology. One of the relevant models is a perfusion of isolated rat heart according to Langendorff. This model allows to explore and study in detail the physiological and biochemical processes in the heart as well as external influences on it in the absence of neurogenic and humoral factors [3–6].

The purpose of this work is to study the influence of immobilized subtilisins on the performance indicators of isolated rat heart according to Langendorff.

Materials and Methods

The experiments were performed on male Wistar rats weighing 250-300 g. An hour before the experiment, the animals were intraperitoneally injected with heparin (500 units per rat). After decapitation, the heart was quickly removed and placed in a container with perfusion solution at t = 0 °C. A cannula was inserted into the aorta and connected to a perfusion system. A perfusion was performed through the coronary vessels under a constant pressure of 70 mmHg. The

modified Krebs-Henseleit buffer was used as a perfusate. The composition of the modified buffer solution Krebs – Henseleit is shown below (Table 1).

TABLE 1. THE COMPOSITION OF THE MODIFIED BUFFER SOLUTION KREBS-HENSELEIT FOR PERFUSION OF ISOLATED RAT HEARTS

Component	Concentration, mM
NaCl	118
KCl	4,7
$CaCl_2$	3,0
$MgSO_4$	1,2
KH_2PO_4	1,2
$NaHCO_3$	25
EDTA-Na_2	0,5
Glucose	5

We used a carbogen (95% O2 and 5% CO2) for heart saturation and conducted a constant monitoring of pH (7.4). The constant solution temperature was 37.5 ° C. To register the pressure developed by the left ventricle, an incision was made in the left atrium behind the appendage, then a latex balloon was injected through it, subsequently filled with perfusion solution to the required volume and connected to a digital pressure sensor. The balloon was introduced first into the cavity of the atrium, and then through the mitral valve into the cavity of the left ventricle. The frequency and rate of contraction were recorded. The heart efficiency (HE) was defined as the product of the pressure increase to the frequency of contractions per minute.

The number of animals in the study group was 7. Each selected heart worked for at least 15 minutes without recirculating of the perfusion solution until the constant amplitude and frequency rates were established. Then, the investigated component was introduced into the perfusion solution, and the heart worked without recirculation for 40 minutes. The work presents data that is obtained at 5, 10, 20, 30, 40 minutes of perfusion in the presence of the studied drug. The results will reflect the maximum values of performance indicators at the 20 minutes of perfusion. A record of the isolated heart working is shown in Figure 1. The studied preparation of immobilized subtilisins is a pharmaceutical substance of the drug Trombovazim (Producer AO SCPB, Novosibirsk, Russia).

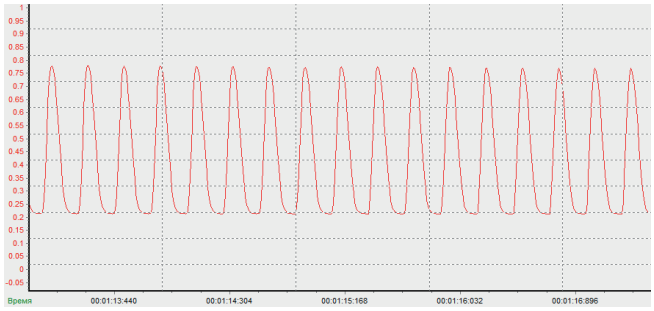


Fig. 1. Record of Isolated Rat Heart Working

The studied component was added to the perfusion solution in 4 concentrations: 150 IU/L, 300 IU/L, 600 IU/L and 1200 IU/L. The initial indicators of the functioning rat heart differed among themselves, therefore, the changes under the influence of the investigated component were evaluated in percent in relation to the initial indicators - "Control". The obtained data were subjected by the statistical analysis using the program StatPlus 2009 Professional (USA). The statistical significance of the results was evaluated using Student's t-test for related samples (before and after administration of the test drug) at a significance level of $p < 0.05$.

Results

A. Study with a dose of 150 IU/L. The heart rate significantly decreased, starting from 20 minutes of perfusion and was 7% in relation to the initial data. In the presence of subtilisins, the maximum increase of the LVP index was noted at the 20 minutes of perfusion by 32% compared with the initial values. The evaluation of the heart efficiency (HE) showed a significant increase by 17% at the 20 minutes of perfusion. The data are summarized in Fig. 2.

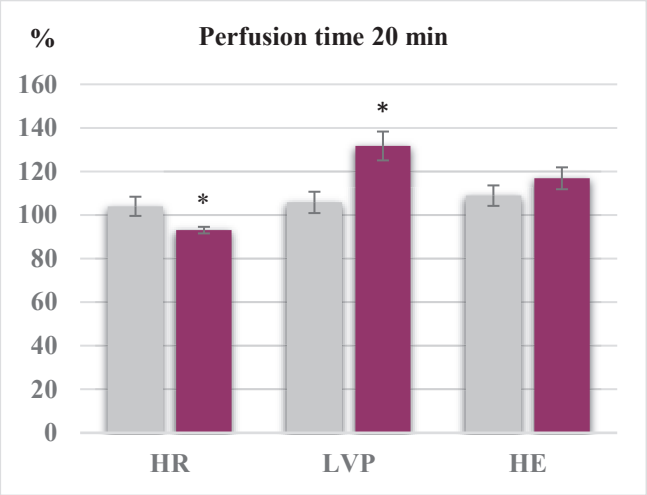


Fig. 2. Perfomance indicators of isolated rat heart, when adding 150 IU/L of immobilized subtilisins (M±m), * - p<0,05 against the control.
Note. Grey column – control (Krebs-Henseleit), lilac column – experiment (Immobilized subtilisins); HR – heart rate, LVP – left ventricular pressure, HE – heart efficiency.

B. Study with a dose of 300 IU/L. The heart rate significantly decreased, starting from 20 minutes of perfusion, and was 5% in relation to the control. During perfusion with subtilisins the maximum increase of the LVP index was observed at the 20 minutes of perfusion by 36%. The evaluation of the heart efficiency (HE) showed a significant increase by 30% at the 20 minutes of perfusion. The data are summarized in Fig. 3.

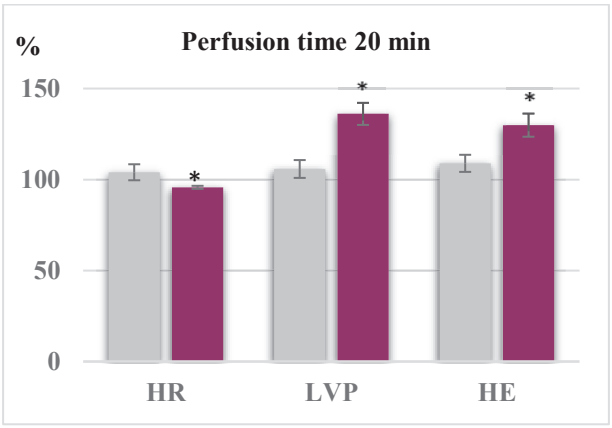


Fig. 3. Perfomance indicators of isolated rat heart, when adding 300 IU/L of immobilized subtilisins (M±m), * - p<0,05 against the control.
Note. Grey column – control (Krebs-Henseleit), lilac column - experiment (Immobilized subtilisins); HR – heart rate, LVP – left ventricular pressure, HE – heart efficiency.

C. Study with a dose of 600 IU/L. The heart rate significantly decreased by 10% in relation to the initial data at 20 minutes of perfusion. There was a maximum increase in the rate of LVP index at 20 minutes of perfusion by 22%, compared with the control. The evaluation of the heart efficiency (HE) showed a significant increase by 18% at the 20 minutes of perfusion. The data are summarized in Fig.4.

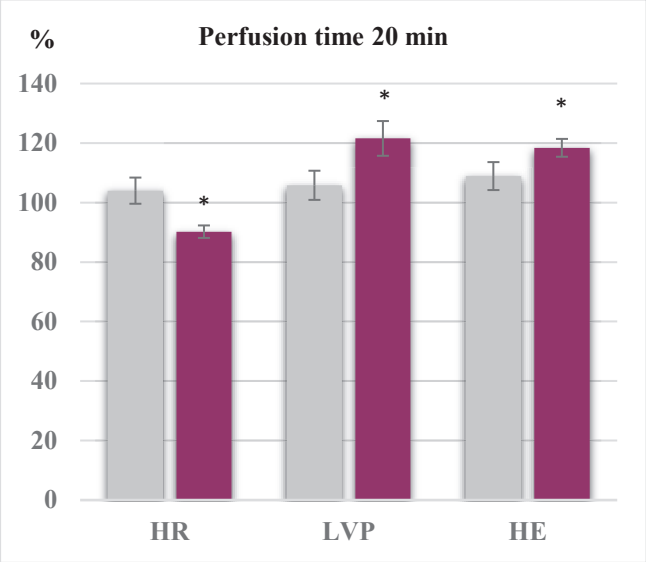


Fig. 4. Perfomance indicators of isolated rat heart, when adding 600 IU/L of immobilized subtilisins (M±m), * - p<0,05 against the control.
Note. Grey column – control (Krebs-Henseleit), lilac column – experiment (Immobilized subtilisins); HR – heart rate, LVP – left ventricular pressure, HE – heart efficiency

Study with a dose of 1200 IU/L. The heart rate was 9% in relation to the control at the 20 minutes of perfusion. The LVP index increased by 25% at 20 minutes of perfusion and relative to the control. The evaluation of the heart efficiency (HE) showed a significant increase by 20% at the 20 minutes of perfusion. The data are summarized in Fig. 5.

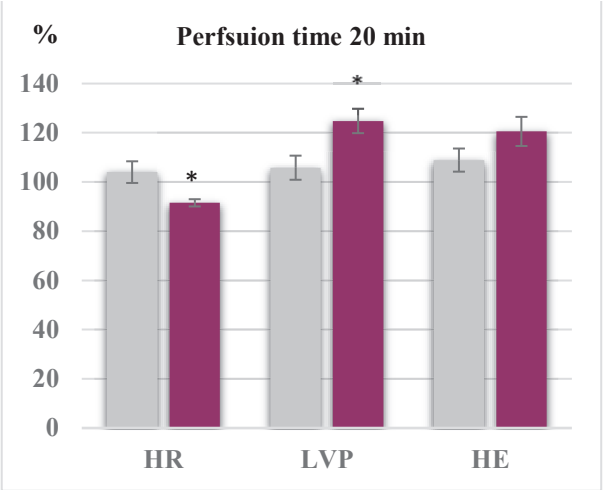


Fig. 5. Perfomance indicators of isolated rat heart, when adding 1200 IU/L of immobilized subtilisins (M±m), * - p<0,05 against the control.
Note. Grey column – control (Krebs-Henseleit), lilac column - experiment (Immobilized subtilisins); HR – heart rate, LVP – left ventricular pressure, HE – heart efficiency.

Consclusion

The studies have shown that immobilized subtilisins have a direct effect on the heart muscle. The immobilized subtilisins increase the force contraction of the left ventricle and reduce the heart rate. There are observed positive inotropic and negative chronotropic effects with perfusion of immobilized subtilisins, most pronounced at a dose of 300 IU/L These circumstances have a scientific and practical interest, since a combination of positive inotropic and negative chronotropic effects establishes a more economical mode of heart function. The immobilized subtilisins have great prospects using in cardiology.

References

[1] P.G. Madonov, S.V. Mishenina, D.N. Kinsht, N.V. Kikhtenko “Chemical and pharmacological properties of subtilisins,” Sibirskij nauchnyj medicinskij zhurnal, vol. 36 (3), pp. 13-22, 2016.

[2] P.G. Madonov, D.N. Kinsht, N.V. Kikhtenko, S.V. Mishenina “Targeted pharmacodynamics of subtilisins,” Sibirskij nauchnyj medicinskij zhurnal, vol. 36 (4), pp. 15-24, 2016.

[3] S.M. Minasyan, M.M. Galagudza, D.L. Sonin “Methods of perfusion of an isolated rat heart,” Regionarnoe krovoobrashchenie i mikrocirkulyaciya, vol. 8(4), pp. 54-59, 2009.

[4] R.G. Merin “The isolated heart preparation,” British Journal of Anesthesia, vol. 60(8), suppl. 1, pp. 28-34. 1988. https://doi.org/10.1093/bja/60.suppl_1.28s.

[5] F.J. Sutherland, K.E. Baker, D.J. Hearse, M.J. Shattock “Mouse isolated perfused heart: characteristics and cautions,” Clinical and Experimental Pharmacology, vol. 30(11), pp. 867-878, 2003. https://doi.org/10.1046/j.1440-1681.2003.03925.x.

[6] R.A. Knyazev, N.V. Trifonova, A.R. Kolpakov, L.M. Polyakov “Investigation of cardiotoxic effect of A-1 apo + D actinomycin complex on isolated rat heart,” Patologiya krovoobrashcheniya i kardiokhirurgiya, vol.20 (4), pp. 88-95. 2016. http://doi.org/10.21688-1681- 3472-2016-4-88-95

DOI 10.18699/BGRS/SB-2020-257

Fibroblast growth factor 21 (FGF21) has a beneficial effect on carbohydrate and lipid metabolism and taste preferences in male and female mice with diet-induced obesity

Balybina N
Department of Physiology
Novosibirsk State University
Novosibirsk, Russia
n.balybina@nsu.ru

Dubinina A, Kazantseva A, Makarova E, Yakovleva T, Bazhan N
Laboratory of Physiological Genetics
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Abstract — **Obesity is a serious problem in modern society, this being the reason for intensive search for drugs to its correction. Fibroblast growth factor 21 (FGF21) is considered as a promising candidate for the development of anti-obesity drugs: it reduces adiposity and increases insulin sensitivity. Preclinical studies of the pharmacological action of FGF21 were conducted only in males. The objective of this study was to investigate the effect of FGF21 administration on taste preferences and carbohydrate and lipid metabolism in male and female mice with diet-induced obesity. Influence of FGF21 administration to obese male and female mice on body weight, glucose and lipid metabolism and choice between standard and high-fat diet was studied. It was shown that despite the significant sex differences in metabolic indices found in mice with diet-induced obesity, the administration of FGF21 had an equal beneficial effect on carbohydrate and lipid metabolism in mice of both sexes. In addition, exogenous FGF21 increased the attractiveness of a balanced standard food compared to a high-fat diet in both male and female mice. These data expand the prospects for the use of FGF21 for pharmaceutical purposes.**

Keywords — fibroblast growth factor, obesity, taste preferences, sex differences

Introduction

Obesity has become one of the main threats to human health. A food abundance and overeating caused by the preference for high-calorie fatty foods, along with a decrease in physical activity are the main factors contributing to the widespread prevalence of obesity. An intensive search for drugs to correct obesity is ongoing. Fibroblast growth factor 21 (FGF21) is considered as a promising candidate for the development of anti-obesity drugs. In laboratory models, monkeys, and humans, administration of FGF21 was shown to reduce body weight and improve blood lipid profile [1]. In experimental animals, FGF21 has an antidiabetic effect: it increases insulin sensitivity and lowers blood glucose [2]. Of special interest is the recently discovered ability of FGF21 to influence taste preferences: the administration of FGF21 reduces the intake of sweets and increases the intake of proteins, is of particular interest. Preclinical studies of the pharmacological action of FGF21 were conducted only in males. However, sex differences in metabolism are so pronounced that NIH recognized sex as an important biological variable that must be considered when conducting preclinical studies. The objective of this study was to investigate the effect of FGF21 administration on taste preferences and carbohydrate and lipid metabolism in male and female mice with diet-induced obesity.

METHODS

C57BL/6J mice were housed in SPF vivarium of the Institute of Cytology and Genetics under a light regime of 13 h light and 11 h darkness, with *ad libidum* access to water and food. Obesity was induced by the addition of a high fat diet (HFD, D12492, protein 24.4%, fats 34.6%) to a standard diet (SD, protein 19%, fats 3.3%). Obese mice aged 7-8 months were placed in the Pheno-Master cells and provided with HFD and SD. After two days of adaptation, body weight, fat and lean mass were measured in all mice and the mice were divided into control and experimental groups: for 7 days, the mice of control group received subcutaneous daily injections of phosphate-buffered saline (PBS), the mice of experimental group received recombinant mouse FGF21 (1 mg / kg) dissolved in the PBS. BW, motor activity, and intake of FHD and SD were measured daily. One day after the last injection, mouse fat and lean masses were measured again, then mice were sacrificed by decapitation and blood and liver samples were collected to analyze biochemical blood parameters and gene expression in the liver. Seven males and five females were administered with PBS, and seven males and five females were administered with FGF21.

Results

In control groups, both calorie intake (ANOVA, $P<0.001$) and consumption of HFD (ANOVA, $P<0.05$) were higher in females than in males, and there were no sex differences in motor activity. At the same time, the rate of obesity development was lower in females: in females, the proportion of fat was lower (ANOVA, $P<0.05$), and the relative weight of lean mass was higher (ANOVA, $P<0.05$) than in males. Administration of FGF21 reduced body weight (ANOVA, $P<0.001$) and the relative proportion of fat (ANOVA, $P=0.05$) with no effect on calorie intake in both males and females (Fig. 1a). Although FGF21 did not affect total calorie intake, it affected the choice between SD and HFD: regardless of sex, it increased the consumption of SD (ANOVA, $P<0.001$) and reduced the consumption of HFD (ANOVA, $P<0.001$) (Fig. 1c, d). This FGF21 effect on food preference enhanced in the course of experiment (ANOVA, $P<0.05$). In control groups, females and males differed in many hormonal and metabolic blood parameters: the blood levels of cholesterol (ANOVA, $P<0.001$), insulin (ANOVA, $P<0.001$) and leptin (ANOVA, $P<0.001$) were lower, and level of adiponectin (ANOVA , $P<0.001$) was higher in females than in males. These differences can be associated with the influence of sex steroids or with different degree of

The Russian Science Foundation (project 17-15-01036) supported the study.

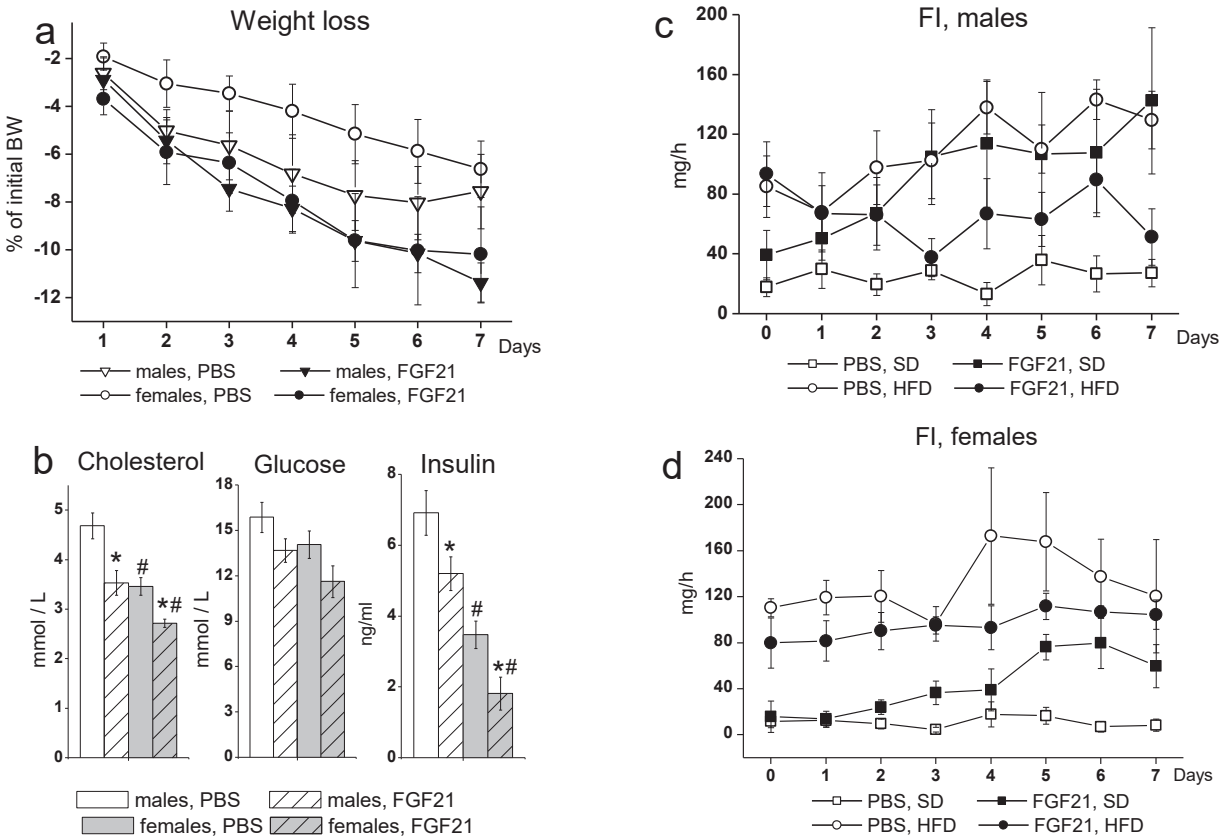


Fig. 1. Influence of FGF21 administration on body weight (a), blood biochemical characteristics (b), and consumption of standard and high-fat diet (c, d) in obese male and female mice. *P<0.05, PBS vs. FGF21; #P<0.05 males vs. females, Post Hoc Fisher LSD test.

adiposity in females and males or with both this factors. FGF21 unidirectionally affected the metabolic parameters in males and females: it lowered cholesterol (ANOVA, P<0.001), glucose (ANOVA, P<0.05) and insulin (ANOVA, P<0.01) levels in the blood (Fig. 1b). A decrease in blood glucose and insulin levels suggests that FGF21 enhances insulin sensitivity, which is consistent with results previously obtained in males [3].

Significant sex differences in liver gene expression were observed: expressions of gene encoding insulin receptor (InsR), genes involved in the activation of fatty acid synthesis (Acca (ANOVA, P<0.001), Accb (ANOVA, P< 0.05)) and oxidation (Pgc1a (ANOVA, P<0.05), Ppara (ANOVA, P<0.05)) were higher in females. Probably, the level of basal metabolism was higher in females than in males. This assumption may explain the mismatch between calorie intake and the degree of adiposity in males and females. Administration of FGF21 unidirectionally affected gene expression in the liver: it reduced the expression of genes encoding fatty acid synthase (Fasn (ANOVA, P<0.05)) and pyruvate kinase (Pklr (ANOVA, P<0.05)) in mice of both sexes. A decrease in the expression of these genes is possibly associated with a decrease in the level of insulin and glucose in the blood, since insulin and glucose activate the expression of Fasn in the liver [4], and glucose activates expression of the pyruvate kinase gene via activation of ChREBP. A decrease in the expression of lipogenic genes in the liver and a decrease in the consumption HFD can contribute to a decrease in blood cholesterol levels under FGF21 administration.

Conclusion

Despite the significant sex differences in metabolic indices found in mice with HFD-induced obesity, the administration of FGF21 had an equal beneficial effect on carbohydrate and lipid metabolism in mice of both sexes. In addition, exogenous FGF21 increased the attractiveness of a balanced standard food compared to a high-fat diet in both male and female mice. These data expand the prospects for the use of FGF21 for pharmaceutical purposes, since they indicate the ability of this drug to correct obesity not only via the influence on metabolic processes, but also due to mental shifts in taste preferences in favor of eating healthy balanced foods.

References

[1] V. M. Jackson, D. M. Breen, J. P. Fortin, A. Liou, J. B. Kuzmiski, A. K. Loomis et al. “Latest approaches for the treatment of obesity”, Expert opinion on drug discovery, vol 10, №. 8, pp. 825-839, 2015.

[2] E. D. Berglund, C. Y. Li, H. A. Bina, S. E. Lynes, M. D. Michael, A. B. Shanafelt et al., “Fibroblast growth factor 21 controls glycemia via regulation of hepatic glucose flux and insulin sensitivity”, Endocrinology, vol. 150, pp. 4084–4093, 2009.

[3] T. Coskun, H. A. Bina, M. A. Schneider, J. D. Dunbar, C. C. Hu, Y. Chen et al., “Fibroblast growth factor 21 corrects obesity in mice”, vol. 149, №. 12, pp. 6018–27, 2008.

[4] Z. Song, A. M. Xiaoli, F. Yang, “Regulation and metabolic significance of de novo lipogenesis in adipose tissues”, Nutrients, vol. 10, №. 10, pp. 1383, 2018.

DOI 10.18699/BGRS/SB-2020-258

Catalytically active bispecific antibodies – new biochemical markers of HIV/AIDS

Svetlana Baranova
ICBFM SB RAS, Novosibirsk, Russia
swb@niboch.nsc.ru

Sergey Sedykh
ICBFM SB RAS, Novosibirsk, Russia
sedyh@niboch.nsc.ru

Georgy Nevinsky
ICBFM SB RAS, Novosibirsk, Russia
nevinsky@niboch.nsc.ru

***Abstract* — Prognosis of the HIV/AIDS infection is an important problem of molecular medicine. Since there were described a number of changes similar to autoimmune processes in patients with HIV/AIDS, such changes might be new markers for disease prognosis.**

Keywords* — *exosomes, milk, microRNA

Motivation and aim

HIV-infection is a viral disease, causing an immune deficiency. In patients with HIV/AIDS there were described a number of changes similar to autoimmune processes. An important problem of HIV/AIDS research is the prognosis of infection. As it was previously shown, autoimmune processes damage the nervous systems, and one of the markers of these processes are catalytic antibodies and bispecific antibodies. Such antibodies *in vitro* can hydrolyze neurospecific substrates.

Methods

Antibodies hydrolyzing neurospecific substrates were described in the blood of healthy donors, patients with systemic lupus erythematosus (SLE) and multiple sclerosis (MS). Proteolytic IgGs hydrolyze myelin basic protein and oligodendrocytic peptide. We have shown that sera of HIV/AIDS, SLE and MS patients contain autoantibodies against histones and myelin basic protein.

Results

Here we show the results of research of natural bispecific catalytically active antibodies isolated from the blood of HIV/AIDS patients. Also, we compared the level of the protease activity of antibodies with the stage and characteristics of the pathological process. Interestingly, the pathological processes in HIV/AIDS are similar with such in MS and SLE. The development of new methods for diagnosing the patient's condition, the effectiveness of therapy, and also predicting disease outcome in the future can be used for personalized therapy and improving the patient's quality of life.

Acknowledgment

The research was supported by Grant of RFBR № 20-34-70115 to Sergey Sedykh.

DOI 10.18699/BGRS/SB-2020-259

The development of compensatory processes in the liver and kidney in conditions of distant tumor growth

Nataliya Bgatova
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
n_bgatova@ngs.ru

Saule Bakhbaeva
Pavlodar State University; Pavladar, Kazakhstan
saule0577@mail.ru

Iuliia Taskaeva
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
email: inabrite@yandex.ru

Asel Rakhmetova
Pavlodar State University; Pavladar, Kazakhstan
asel-rakhmetova@mail.ru

Viktoriia Makarova
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
shedina_vika@mail.ru

***Abstract* — Development of malignant tumor accompanied endogenous intoxication of an organism having a mixed nature, in particular due to dysfunction of organs detoxification and excretion damaging effect of tumor metabolism. The active process of moving toxic and biologically active substances formed during tumor progression from the primary focus by blood and lymph flow leads to damage of distant organ. The aim of this work was to identify structural changes in the liver and kidney under the conditions of modeling distant tumor growth. For modeling tumor growth, hepatocarcinoma-29 (G-29) cells were used. G-29 cells were injected into the muscle of the right thigh of CBA mice [1]. The material for research was collected after 3, 7, 13, and 30 days of the experiment. The structure of the liver and kidney was studied by the method of light and electron microscopy. Both destructive and compensatory changes in the liver and kidney were revealed in conditions of distant tumor growth. In the liver, a decrease a size of hepatocytes and concentration of organelles, as well as development of autophagy, as a mechanism for maintaining intracellular homeostasis, were noted. In the kidney at the early stages of the development of the tumor growth, structural signs of filtration barrier disorde were shown. Subsequent development of compensatory hypertrophy of podocytes and glomerular capillary endothelium, as well as autophagy in the epithelium of the proximal nephron were noted**

Keywords* — *hepatocellular carcinoma, liver, kidney, destructive and compensatory changes

Introduction

Cancer patients often suffer from fatigue, a complex syndrome associated with loss of muscle mass, weakness and depressed mood. Cancer-related fatigue can be detected during diagnosis, manifest during treatment, and persist for many years after treatment. Cancer-related fatigue adversely affects quality of life, limits functional independence, and is associated with decreased survival rates for cancer patients. There are currently no effective treatments for fatigue associated with cancer, as the pathophysiology of this syndrome is poorly understood. One of the causes of cancer-related fatigue is the fact that the development of a malignant tumor is accompanied by endogenous intoxication of the body. Development of malignant tumor accompanied endogenous intoxication of an organism having a mixed nature, in particular due to dysfunction of organs detoxification and excretion damaging effect of tumor metabolism. The active process of moving toxic and biologically active substances formed during tumor progression from the primary focus by blood and lymph flow leads to damage of distant organ. The aim of this work was to identify structural changes in the liver and kidney under the conditions of distant tumor growth.

Methods and algorithms

An experimental study was conducted on CBA male mice weighing 18-20 g. The animals were kept on a standard diet with free access to water and food. Work with animals was carried out in accordance with the "Rules for the work using experimental animals." In the experiment, 2 groups of animals were used. Group 1 included intact mice (n = 5); group 2 - animals, with the development of the tumor process (n = 20). Hepatocarcinoma-29 cells (G-29) were used to simulate tumor growth. G-29 cells were suspended in 10-fold volume of saline and injected 0.1 ml of intact animals in the right thigh muscle [1]. The material for research was collected after 3, 7, 13, and 30 days of the experiment. Animals were removed from the experiment by the cranio-cervical dislocation method. For electron microscopy studies, liver and kidney fragments up to 1 mm3 in size were fixed in a 4% solution of paraformaldehyde prepared on Hanks medium, fixed for 1 hour in a 1% solution of OsO4 in phosphate buffer (pH = 7.4), and were dehydrated in increasing ethyl alcohol concentration and concluded in epon (Serva, Germany). Semi-thin sections with a thickness of 1 μm were prepared from the obtained blocks on a Leica UC7 / FC7 ultratome (Germany / Switzerland), stained with toluidine blue, studied under a LEICA DME light microscope, and the required tissue sections were selected for examination an electron microscope. Ultrathin sections with a thickness of 70-100 nm were obtained from the selected material, contrasted with a saturated aqueous solution of uranyl acetate and lead citrate and studied using a JEM 1400 electron microscope (Center for Collective Use, Federal Research Center for Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences). Morphometric analysis was performed using ImageJ computer software (Wayne Rasband, USA). The average value (M) and standard deviation (SD) were calculated using Microsoft Excel software (Microsoft, USA). The significance of differences between the studied

parameters was determined using the Statistica 6.0 software (StatSoft, USA) using the Mann-Whitney U-test at a confidence level of 95% ($P < 0.05$).

Results

Hepatocarcinoma-29 is an extremely aggressive transplantable cell line. Upon implantation of hepatocarcinoma-29 cells into the thigh region of experimental animals, tumor cells quickly replaced muscle fibers. By 13 days of the experiment, a tumor node was formed, and by 30 days after implantation, mature tumor tissue appeared, having a kind of liver beams surrounded by "sinusoids" [2]. As a result of invasion of tumor cells and damage to membrane structures, there was an increase in the level of secondary products of lipid peroxidation in muscle tissue: on day 7 - 2.1 times, on day 13 - 1.4 times relative to control values [3], which could have a toxic effect on the liver and kidneys. The liver, as the central organ of detoxification and metabolism, is most susceptible to the toxic effects of malignant products. In the dynamics of tumor growth, there was a decrease volume density of mitochondria, endoplasmic reticulum, lipid inclusions and an increasing volume density of lysosomal structures. The development of intracellular autophagic degradation of cytoplasmic organelles was revealed. In autophagosomes, fragments of cytoplasm, glycogen rosette, mitochondria, fragments of the endoplasmic reticulum with ribosomes were observed. The data obtained indicate that in conditions of distant tumor growth in the liver, non-selective autophagy develops to maintain intracellular hepatocyte homeostasis, as well as energy and trophic homeostasis of the body. In the early stages development of the tumor process, destructive disorders in the kidney were noted, indicating a violation of the renal filter - a decrease in the number of fenestra of glomerular capillary endotheliocytes, an increase in the thickness of the glomerular membrane, fusion of the legs of podocytes. Starting from 13 days of tumor development, compensatory processes were noted in the structure of nephron components. Podocyte hypertrophy was identified: the concentration of organelles in the cytoplasm of podocytes and the number of cytopodia were increased. A decrease in the thickness of the glomerular membrane and an increase in the height of the glomerular capillaries were observed. Autophagy in epithelium of the proximal part of nephron was noted.

Conclusion

The development of both destructive and compensatory changes in the liver and kidney in conditions of distant tumor growth was revealed. In the liver, a decrease a size of hepatocytes and concentration of organelles, as well as development of autophagy, as a mechanism for maintaining intracellular homeostasis, were noted. In the kidney at the early stages of the development of tumor growth, structural signs of a disorder of the filtration barrier are shown, subsequent development of compensatory hypertrophy of podocytes and glomerular capillary endothelium, as well as autophagy in the epithelium of the proximal nephron were identified.

Acknowledgment

Supported by the state order No. 0324-2019-0045 Research Institute of Clinical and Experimental Lymphology.

References

[1] Y V. I. Kaledin, N. A. Zhukova, V. P. Nikolin, N. A. Popova, M. D. Belyaev, N. V. Baginskaya, E. A. Litvinova, T. G. Tolstikova, E. L. Lushnikova and D. E. Semenov, "Hepatocellular carcinoma-29 – metastatic transplantable tumor of mice, causing cachexia," Bull. Exp. Biol. Med., vol. 148, pp. 664–669, December 2009.

[2] N. P. Bgatova, Y. I. Borodin, V. V. Makarova, A. A. Pozhidaeva, L. N. Rachkovskaya, V. I. Konenkov « Effects of nanosized lithium carbonate particles on intact muscle tissue and tumor growth», Bull Exp Biol Med., vol. 157, pp. 89-94, May 2014.

[3] V. I. Konenkov, Y. I. Borodin, O. P. Makarova, N. P. Bgatova, L. N. Rachkovskaya «Effects of lithium carbonate nanosized particles on oxidant-antioxidant status in tumor tissue of hepatocarcinoma-29, Patol Fiziol Eksp Ter., vol. 59, pp. 57-64, April -June 2015.

DOI 10.18699/BGRS/SB-2020-260

What will help neurobiology take the next step: rejuvennated classics, connectoms, bioinformatics or artificial intelligence?

Igor Bondar
IHNA & NPh RAS, Moscow, Russia
bondar@ihna.ru

Abstract — **Neurobiology is a rapidly developing field of modern science. New research technologies allow researchers to actively intervene in the work of the whole brain and local neural networks. A true understanding of the fundamental principles of the brain functioning is possible only by combination of systematic approach in the analysis of complex behavior and detailed information on neuronal functions.**

Keywords *— extracellular recordings, connectomics, optogenetic approach, artificial intelligence*

The central issue in neurobiology remains the understanding of how the activity of single neurons, combined into distributed neural networks, provides animals with the ability to adapt to rapidly changing environmental conditions using a specific behavioral pattern. Adaptive behavior can be observed in simply organized living beings who lack a highly developed nervous system, but have the opportunity to feel changes in the physical or chemical properties of the environment, as well as the ability to actively change position in space.

Considering the case of the evolution of the visual system, it was shown that the principle of "division of labor" underlie evolutionary changes [1]. As a result specialized cells emerging in the functional space between the sensitive part of the nervous system and motoneurons that provide response to the stimulus. An "interneuronal explosion" gave the nervous system the capabilities of a finer analysis of changes in environment, but at the same time it complicated the life for modern neurobiologists who risk being buried under the rubble of experimental data.

Currently, neurobiology dominated by an interventionist approach, which involves active intervention in the brain function. It opposed by a systematic approach, whose defenders propose a deeper analysis of animal behavior [2]. This polarization in modern brain research is associated with the progress of sophisticated scientific technologies that provide modern neuroscientists with a wealth of data.

The development of computer technology has given new life to classical brain research methods. The technique for recording extracellular potentials, which was invented about 100 years ago, is used today to record the activity of a large number of single neurons simultaneously. Over the years, the study of the properties of single neurons has been a key research technology in the works that were awarded two times by Nobel Prize. A study of the properties of neuron detectors in the primary visual cortex of a cat's brain and the place cells in a rat's hippocampus has enriched an understanding of fundamental principles organizing the work of the whole brain.

Currently, chronically implanted multiple microelectrodes or special matrices can be used in conjunction with brain activity imaging methods, which makes it possible to determine the degree of involvement of the activity of individual cells in the functioning of distributed neural ensembles. Long-term registrations provide an opportunity to observe plastic changes in the activity of nerve cells associated with the acquisition of experience and learning. Special value this method has for the long-term registration of neurons in experiments connected with the studies of the relationship between neuronal activity and the analysis of sensory information. In this case formation of the internal representation of environmental objects can be investigated [3]. In our experiments we have shown that the prototype, the "averaged image" that forms in specialized parts of the brain, can play a central role in recognition of faces. Thus, updated classical methods can still provide researchers with valuable experimental data. Moreover, methodological approaches tested in animal experiments can become the basis for creating safe and reliable invasive brain-machine interfaces that will provide the ability to restore impaired brain functions of patients [4].

Along with classical methods, new approaches also appear. Some time ago, the idea began to circulate in the research environment that a true understanding of brain functions is possible by detailed description of all its connections. Currently, active work is underway to create a digitalized description of neural networks with a resolution level of a single synapse [5]. Connectoms allow to accurately describe the ultrastructure of the brain, but they are extremely time-consuming and requires large amounts of computing power. The main problem is that at the conceptual level we still do not have an understanding for algorithms that will allow us to move from structure to function. Given that at all levels of the organization, the nervous system demonstrates an amazing ability for constant changes, a detailed "snapshot" of the structure can have only limited value for understanding of fundamental principles of the brain function.

Strengthening the position of the interventionist approach in neurobiology has become possible in connection with the development of genetic and molecular methods. Genome sequencing becomes available for different model organisms, which allows to purposefully introduce new information into them, affecting the functional properties of neurons. The optogenetic approach uses the full power of bioinformatics for targeted transfection of subpopulations of neurons [6]. Subsequent local light exposure of nerve tissue gaves researchers the possibility to change the functionality of neural networks. Optogenetic experiments not only shed light on the fundamental aspects of the brain function, but also allow to restore its lost performance [7].

Artificial intelligence has two possible points of application in modern neurobiology. In some labor-intensive research technologies, it can be a working tool for solving specific problems. Automated analysis of the results of structural and functional brain imaging becomes more efficient when using artificial intelligence. The combination of artificial intelligence with the collective work of a large number of proofreaders forms a unique symbiosis of a person and a computer program in which both sides get advantages [5]. On the other hand, understanding the basics of natural neural ensembles will allow us to implement this knowledge into the functional architecture of artificial neural networks, which will become the basis for creating new assistive technologies that increase the efficiency of human activities. While research in the field of artificial intelligence is based more on a mathematical apparatus, an important task for neurobiology could be to provide this direction of science with content [8].

Successes in various fields of neurobiology in the context of the general development of knowledge about living things allow us to look to the future with optimism. However, it is extremely important to remember that only a systematic approach to the analysis of the entire array of experimental data will allow us to achieve such an understanding of the brain, which in the future will allow us to “repair” it.

ACKNOWLEDGMENT

Supported by the RFBR (19-015-00349) und funds within the state assignment of Ministry of Education and Science of the Russian Federation for 2019-2021 (AAAA-A17-117092040002-6)

REFERENCES

[1] Arendt D. et al. (2009) The 'division of labour' model of eye evolution.

[2] Philos Trans R Soc Lond B Biol Sci. 364(1531):2809-17. doi: 10.1098/rstb.2009.0104.

[3] Krakauer J.W. et al. (2017) Neuroscience Needs Behavior: Correcting a Reductionist Bias. Neuron. 93(3):480-490. doi: 10.1016/j.neuron.2016.12.041

[4] Leopold D.A. et al. (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature. 442(7102):572-5.

[5] Lebedev M.A., Nicolelis M.A. (2017) Brain-Machine Interfaces: From Basic Science to Neuroprostheses and Neurorehabilitation. Physiol Rev. 97(2):767-837

[6] Lee K. et al. (2019) Convolutional nets for reconstructing neural circuits from brain images acquired by serial section electron microscopy. Curr. Opin. Neurobiol. 55:188-198. doi: 10.1016/j.conb.2019.04.001.

[7] Lamprecht R (2019). Regulation of signaling proteins in the brain by light. Prog Neurobiol. 180:101638. doi: 10.1016/j.pneurobio.2019.101638

[8] Simon C.J. et al. (2020) Opsins for vision restoration. Biochem. Biophys. Res. Commun. pii: S0006-291X(20)30074-7. doi: 10.1016/j.bbrc.2019.12.117

[9] Savage N. (2019) How AI and neuroscience drive each other forwards. Nature. 571(7766):S15-S17. doi: 10.1038/d41586-019-02212-4.

DOI 10.18699/BGRS/SB-2020-261

Analysis of methylation of antioxydant-related genes in patiens with common cardio- and cerebrovascular diseases

Olga Bushueva
KSMU, Kursk, Russia
olga.bushueva@inbox.ru

Ekaterina Barysheva
KSMU, Kursk, Russia
ekatbarysheva@yandex.ru

Anton Markov
RIMG, TNRMC RAS
Tomsk, Russia
anton.markov@medgenetics.ru

Yulia Koroleva
RIMG, TNRMC RAS
Tomsk, Russia
yuliya.koroleva@medgenetics.ru

Egor Churkin
RIMG, TNRMC RAS Tomsk, Russia

Maria Nazarenko
RIMG, TNRMC RAS
Tomsk, Russia
maria.nazarenko@medgenetics.ru

Alexey Polonikov
KSMU, Kursk, Russia
polonikov@rambler.ru

Vladimir Ivanov
KSMU, Kursk, Russia
ivanovvp@kursksmu.net

***Abstract* — It is well-known that oxidative stress may contribute to the increased risk of cardio- and cerebrovascular diseases. The aim of our study was to analyze the level of methylation of antioxidant-related genes (*MPO, TXNRD1, GCLM, GSTP1*) in patients with coronary artery disease (N=45), cerebral stroke (N=30) and healthy controls (N=83). DNA methylation level was evaluated by bisulfite pyrosequencing. Analysis of coronary artery disease revealed a significant decrease in methylation level for all genes, however, the most significant differences were found for *GCLM* ($P=3.63\times10^{-8}$) and *MPO* ($P=3.83\times10^{-7}$). An analysis of cerebral stroke revealed decrease in methylation levels in all genes except *TXNRD1*; the most significant differences were observed for *GSTP1* ($P=9.43\times10^{-6}$) and *MPO* ($P=5.25\times10^{-11}$). Molecular-genetic and epigenetic mechanisms of the involvement of redox-homeostasis genes in the development of common cardio- and cerebrovascular diseases are discussed.**

Keywords* — *oxidative stress, DNA methylation, coronary artery disease, cerebral stroke

Motivation and Aim

Cardio- and cerebrovascular diseases (CCVD) is the most common cause of morbidity and mortality worldwide. Previous investigations have found that oxidative stress may contribute to the increased risk of CCVD via a variety of mechanisms, including endothelial dysfunction, smooth muscle cell contraction, inflammatory process, lipid peroxidation, activation of metalloproteinases etc [1]. It is noteworthy that epigenetic studies of genes involved in regulation of vascular redox homeostasis are single.

The aim of the study was to analyze weather the methylation of antioxidant-related genes may contribute to the increased risk of coronary artery disease and cerebral stroke.

Methods

Genomic DNA was isolated from their peripheral blood leukocytes. Then, bisulfite DNA conversion was performed. To assess the status of DNA methylation, the promoter regions of the *TXNRD1* genes (3 CpG-sites), *GSTP1* (2 CpG-sites), *GCLM* (4 CpG-sites), and 4-6 exons of the *MPO* gene (3 CpG-sites) were selected. DNA methylation status was determined in the blood leukocytes of healthy individuals (N=83) and patients with common cardio- and cerebrovascular diseases (coronary artery disease (N=45) and cerebral stroke (N=30)). All patients with coronary artery disease (CAD) and cerebral strike (CS) had a history of arterial hypertension. DNA methylation level was evaluated by bisulfite pyrosequencing using a PyroMark Q24 (Qiagen). To compare the methylation level of CpG-sites, the Mann-Whitney U test was used in SPSS Statistics v23.

Results

A comparative analysis of the methylation level revealed pronounced differences between CVD patients and healthy individuals. Analysis of coronary artery disease revealed a significant decreasing in methylation level for all genes, however, the most significant differences were found for *GCLM* ($P=3.63\times10^{-8}$) and *MPO* ($P=3.83\times10^{-7}$). An analysis of cerebral stroke revealed decrease in methylation levels in all genes except *TXNRD1*; the most significant differences were observed for *GSTP1* ($P=9.43\times10^{-6}$) and *MPO* ($P=5.25\times10^{-11}$) (Table 1).

Table 1 - A comparative analysis of methylation of genes (the region average value) in patients with various cardiovascular phenotypes and in healthy controls

Genes \ Phenotypes		Control	CAD	Control	CS
GSTP1	Me (Q0,25 -0,75)	4.3 [3.5; 6.7]	3.6 [2.6; 5.7]	4.4 [3.6; 7.1]	2.7 [2.3; 3.9]
	P_{adj}	0.01*		$9.43\times10^{-6*}$	
GCLM	Me (Q0,25 -0,75)	7.8 [5.3;10.1]	3.4 [2.7; 4.7]	7.9 [5.3;10.3]	5.8 [4.1; 7.1]
	P_{adj}	$3.63\times10^{-8*}$		0.004*	
TXNRD1	Me (Q0,25 -0,75)	4.0 [3.0; 5.2]	3.1 [2.5; 3.7]	4.1 [3.0; 5.2]	4.2 [3.0; 5.5]
	P_{adj}	0.0002*		0.78	
MPO	Me (Q0,25 -0,75)	35.4 [30.3; 42.6]	26.5 [24.5; 32.3]	35.6 [30.4; 42.6]	24.3 [19.8; 27.8]
	P_{adj}	$3.83*10^{-7*}$		$5.25*10^{-11*}$	

P_{adj}- P-level after adjustment for multiple comparisons;
Note: Bold indicates differences between the groups

However, DNA methylation levels of studied genes, except *MPO*, varied within low ranges (% of methylated cytosines <10%). Analyzed region of *MPO* is located in CpG-island, and according to JASPAR database, contains motifs for several transcription factors (E2F6, TFDP1, NHLNH2,

and others). ChIP-seq based data from ENCODE strongly supports existence of E2F6 binding site, which could be impaired by abnormal DNA methylation in blood cells of patients with CCVD.

Conclusion

We found significant differences in the level of DNA methylation of redox-homeostasis genes between healthy individuals and patients with common cardio- and cerebrovascular disease (coronary heart disease and cerebral stroke).

REFERENCES

[1] Incalza, Maria Angela, et al. (2018) Oxidative stress and reactive oxygen species in endothelial dysfunction associated with cardiovascular and metabolic diseases. Vascular pharmacology. 100 (2018): 1-19.

DOI 10.18699/BGRS/SB-2020-262

Molecular diagnostics of hearing loss due to mutations in the *SLC26A4* gene in indigenous peoples of Southern Siberia (Russia)

Valeriia Yuryevna Danilchenko
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
danilchenko_valeri@mail.ru

Marina Vyacheslavovna Zytsar
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
zytzar@bionet.nsc.ru

Marita Sergeevna Bady-Khoo
Perinatal Center of the Republic
of Tyva
Kyzyl, Russia
marita.badyhoo@mail.ru

Ekaterina Alexandrovna Maslova
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
State University
Novosibirsk, Russia
maslova@bionet.nsc.ru

Olga Leonidovna Posukh
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
State University
Novosibirsk, Russia
posukh@bionet.nsc.ru

***Abstract* — Establishing a genetic diagnosis of hearing loss is of great importance for clinical evaluation of deaf patients and for estimating recurrence risks for their families. Mutations in the *SLC26A4* gene are a common cause of hearing loss in many regions of the world. This study presents the results of the first molecular genetic analysis of the *SLC26A4* gene sequence performed in patients with hearing loss of unknown etiology belonging to indigenous peoples of Southern Siberia - Tuvinians (the Tyva Republic) and Altaians (the Altai Republic). Contrast differences of the *SLC26A4* pathogenic contribution to the etiology of hearing impairment were revealed: 28.2% for Tuvinians and 4.3% for Altaians. Both known and novel pathogenic variants as well as a wide range of polymorphic variants were found in the *SLC26A4* gene sequences of examined patients. High frequency of mutation c.919-2A>G in Tuvinians is probably due to the founder effect.**

Keywords* — *hereditary deafness, SLC26A4, molecular diagnostics, indigenous peoples of Southern Siberia

Introduction

Hereditary hearing loss is a rare monogenic disease with a unique high heterogeneity of genetic control. To date, about 160 genetic loci and 100 different genes associated with nonsyndromic hearing loss have been identified. Establishing a genetic diagnosis of hearing loss is of great importance for clinical evaluation of deaf patients and for estimating recurrence risks for their families. Mutations in the *GJB2* gene have the most pathogenic contribution to development of isolated hearing loss in many populations, but there are significantly less data on the pathogenic contribution of other genes associated with hearing impairments. Mutations in the *SLC26A4* gene (solute carrier family 26, member 4; 7q22.3, OMIM 605646) are the second most frequent cause of hereditary hearing loss, after mutations in the *GJB2* gene, accounting for ~ 10% of all hereditary hearing impairment cases at least in some Asian populations [1, 2] making *SLC26A4* gene testing essential for the establishment of genetic diagnosis of hearing loss.

The *SLC26A4* gene consists of 21 exons and encodes transmembrane protein pendrin (780 amino acids) - a multifunctional anionic transporter that is mainly expressed in the inner ear, thyroid, and kidneys. Mutations in *SLC26A4* lead to recessive nonsyndromic hearing loss (DFNB4, OMIM 600791) usually accompanied by the abnormalities in the inner ear structures such as the enlarged vestibular aqueduct (EVA) etc., which can be detected by computed tomography of the temporal bones, and some forms of Pendred syndrome (PDS, OMIM 274600) which combines hearing loss and goiter [3, 4].

At present, there is no sufficient experimental information on the pendrin structure and precise molecular mechanisms underlying function of this protein. Moreover, no systematic genotype–phenotype correlations have been made so far in deaf patients having hearing impairment presumably associated with *SLC26A4* mutations. Taking also into account a large physical size of the *SLC26A4* gene, the development of routine *SLC26A4* molecular diagnostics for hearing impairments is a nontrivial task.

About 260-500 pathogenic *SLC26A4* variants associated with hearing impairments, according to different sources (ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/; the Human Gene Mutation Database, http://www.hgmd.cf.ac.uk/ac/; the Deafness Variation Database, http://deafnessvariationdatabase.org/), are currently revealed in patients in different populations worldwide but there are still many regions where the *SLC26A4* contribution to deafness remains unknown.

Previously, we found that the proportion of deafness due to the mutations in the *GJB2* gene in deaf patients belonging to indigenous peoples of Southern Siberia reaches to 22.3% among Tuvinian patients (the Tuva Republic) and 15.1% among Altaian patients (the Altai Republic) [5]. However, the cause of hearing loss remains unclear for significant part of patients in these regions.

The aim of this study is to investigate the *SLC26A4* pathogenic contribution in deafness in Tuvinian and Altaian patients with hearing loss of unknown etiology.

MATERIALS AND METHODS

Analysis of the *SLC26A4* complete sequence including all 21 exons and flanking regions was performed by Sanger sequencing in the group of patients who were negative for the mutations in the *GJB2* gene (171 out of 220 Tuvinian patients and 61 out of 93 Altaian patients). To investigate the inner ear abnormalities, computed tomography of the temporal bones (CT) was conducted in patients with detected *SLC26A4* mutations. Analysis of the carrier frequency of *SLC26A4*

mutations was performed in the ethnically matched samples of unrelated healthy individuals (157 Tuvinians, 218 Altaians). Several bioinformatics prediction tools (PolyPhen-2; PROVEAN; FATHM; MutationAssessor; etc.) were applied for evaluation of pathogenicity of novel *SLC26A4* variants. Two-tailed Fisher's exact test with significance level of $p<0.05$ was applied to compare allele frequencies between patients and controls.

RESULTS

A molecular genetic analysis of the *SLC26A4* gene was performed in patients with hearing loss of unknown etiology belonging to indigenous peoples of Southern Siberia: Tuvinians (the Tyva Republic) and Altaians (the Altai Republic). Both known: c.170C>A (p.Ser57Ter), c.919-2A>G, c.2027T>A (p.Leu676Gln), c.2034+1G>A, c.2168A>G (p.His723Arg) and novel (c.1545T>G (p.Phe515Leu), c.1717G>T (p.Asp573Tyr) pathogenic recessive variants as well as a wide range of benign variants were found in the *SLC26A4* sequences of patients. Probable pathogenicity of novel *SLC26A4* variants was established by bioinformatics prediction tools, and for variant c.1545T>G was confirmed by its segregation with hearing loss revealed by the analysis of patient's pedigrees. Significantly higher pathogenic contribution of the *SLC26A4* gene in deafness (estimated by the presence of biallelic recessive *SLC26A4* mutations) was found in Tuvinian patients (28.2%) in contrast to Altaian patients (4.3%) ($p<0.05$) (Table 1).

TABLE I. *SLC26A4* GENOTYPES FOUND IN EXAMINED PATIENTS

SLC26A4 genotypes	Tuvinian patients (n=220)	Altaian patients (n=93)
Homozygous and compound heterozygous *SLC26A4* genotypes		
c.[919-2A>G];[919-2A>G]	30	-
c.[2027T>A];[2027T>A]	4	-
c.[2168A>G];[c.2168A>G]	-	2
c.[170C>A];[170C>A]	1	-
c.[919-2A>G];[2027T>A]	14	2
c.[919-2A>G];[1545T>G*]	8	-
c.[170C>A];[919-2A>G]	3	-
c.[919-2A>G];[2034+1G>A]	1	-
c.[1545T>G*];[2027T>A]	1	-
Total	62 (28.2%)	4 (4.3%)
Heterozygous *SLC26A4* genotypes		
c.[919-2A>G];[?]	9	-
c.[170C>A];[?]	1	-
c.[1545T>G*];[?]	1	-
c.[2027T>A];[?]	1	1
c.[1717G>T*];[?]	-	1
Total	12	2

* - novel variant

Additional investigation of the *SLC26A4* gene sequence is necessary to perform in group of deaf patients with monoallelic *SLC26A4* mutations (Table 1). The enlargement of vestibular aqueduct (EVA) (varying from 1.5 mm to 5.1 mm) was revealed in most patients homozygous or compound heterozygous for *SLC26A4* mutations who passed CT (n=27). To note, the degree of EVA was characterized by sufficient intra and interfamily variability.High frequency of mutation c.919-2A>G in Tuvinians (69.9% of all mutant alleles detected in patients and carrier frequency reaching to 5.1% in Tuvinian control sample) is probably due to the founder effect.

CONCLUSION

This is the first study to address the *SLC26A4* mutations contribution in deafness in indigenous populations of Southern Siberia. Contrast differences in the proportion of deafness caused by the *SLC26A4* mutations were revealed in Tuvinians and Altaians despite of related ethnicity and geographically close territories of residence of these indigenous peoples of Southern Siberia. Novel data on spectrum and prevalence of pathogenic and benign variants of gene *SLC26A4* in Siberian populations significantly contribute to the *SLC26A4* allelic diversity worldwide.

ACKNOWLEDGMENTS

Study was supported by the Projects #0259-2019-0010-C-01, #0324-2019-0041-C-01, and the RFBR grant #17-29-06016_ofi-m.

REFERENCES

[1] K. Tsukamoto, H. Suzuki, D. Harada, A. Namba, S. Abe, S. Usami. Distribution and frequencies of *PDS* (*SLC26A4*) mutations in Pendred syndrome and nonsyndromic hearing loss associated with enlarged vestibular aqueduct: a unique spectrum of mutations in Japanese. Eur J Hum Genet. 2003, vol. 11(12), pp. 916–922.

[2] Q. J. Wang, Y. L. Zhao, S. Q. Rao, Y. F. Guo, H. Yuan, L. Zong, et al. A distinct spectrum of *SLC26A4* mutations in patients with enlarged vestibular aqueduct in China. Clin Genet. 2007, vol. 72(3), pp. 245–254.

[3] N. Chen, L. Tranebjærg, N. D. Rendtorff, I. Schrijver. Mutation analysis of *SLC26A4* for Pendred syndrome and nonsyndromic hearing loss by high-resolution melting. J Mol Diagn. 2011, vol. 13(4), pp. 416–426.

[4] L. A. Everett, B. Glaser, J. C. Beck, J. R. Idol, A. Buchs, M. Heyman, et al. Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS). Nat Genet. 1997, vol. 17(4) pp. 411–422.

[5] O. L. Posukh, M. V. Zytsar, M. S. Bady-Khoo, V. Y. Danilchenko, E. A. Maslova, N. A. Barashkov, et al. Unique mutational spectrum of the *GJB2* gene and its pathogenic contribution to deafness in Tuvinians (Southern Siberia, Russia): a high prevalence of rare variant c.516G>C (p.Trp172Cys). Genes (Basel). 2019, vol. 10(6) p. 429.

DOI 10.18699/BGRS/SB-2020-263

Functional study of potential regulatory SNPs (rs590352, rs11542583, rs3829202, rs78317230, rs2072580, rs4796672)

Arina Degtyareva
Laboratory of gene expression regulation
ICG SB RAS
Novosibirsk, Russia
degtyareva_rso@mail.ru

Elena Leberfarb
Laboratory of gene expression regulation
ICG SB RAS
Novosibirsk, Russia

Ilya Brusentsov
Laboratory of gene expression regulation
ICG SB RAS
Novosibirsk, Russia

Tatiana Kuzina
Department of natural sciences
NSU
Novosibirsk, Russia

Tatiana Merkulova
Laboratory of gene expression regulation
ICG SB RAS
Novosibirsk, Russia

Abstract — **In a previous study using a new bioinformatic approach based on Chip-Seq data with antibodies against various transcription factors and histone modifications, ChiA-PET, RNA-seq and ICGC data, 32 potentially regulatory single -nucleotide polymorphisms (rSNPs) were associated with colorectal cancer. In this study, a functional analysis of 6 of these polymorphisms was performed using EMSA and the luciferase reporter system. These nucleotide substitutions have been shown to alter the binding of nuclear extract proteins and influence the expression of the reporter gene.**

Keywords — regulatory SNP (rSNP), gene expression, Electrophoretic Mobility Shift Assay (EMSA), luciferase assay

Introduction

Determining the molecular basis of genetic predisposition to various diseases is a fundamental task of medical genetics. Recently the main research area is to establish an association between variants of nucleotide sequences and a particular pathology, and the main tool is a Genome-Wide Association Studies (GWAS). However, the study of associations does not distinguish the polymorphism actually involved in pathology from the marker one (linkage disequilibrium) [1]. Moreover, GWAS does not provide information about the functionality of these variants. So the molecular mechanisms of pathology development remain unknown [2]. This is especially true for SNPs located in non-coding regions, whose share in GWAS-associated variants is about 90%. The functional interpretation of these polymorphisms is the most difficult task [3, 4].

To understand the molecular sense of GWAS- associated polymorphisms, different annotations are used. For example, transcription factor binding motifs, histone modifications, promoter, enhancer and super enhancer landscape in the genome extracted from different functional genomics databases, such as JASPAR, HOCOMOCO, ENCODE and etc. In addition, alternative approaches to search for regulatory polymorphisms (rSNPs) are being developed, for which the primary goal is to determine the functionality of genetic variants. For example, our laboratory has developed a bioinformatic approach that allows to detect rSNPs. This approach is based on the analysis of data on allelic asymmetry of chromatin protein and transcription factors binding and allelic asymmetry of gene expression. As a result, about 1,500 rSNPs were identified. Using data from the ICGC (International Cancer Genome Consortium), 32 rSNPs were associated with colorectal cancer (CRC) [5].

The aim of the present study was to study the functional significance of 6 polymorphisms from these 32 (rs590352 G>C, rs11542583 A >G, rs3829202 T>C, rs78317230 T>C, rs2072580 A>T, rs4796672 C>T) using EMSA and luciferase reporter system.

Materials and Methods

Cell cultures

Human hepatoma cells HepG2, erythromyeloblastoid leukemia cells K562, cervical cancer cells Hela were cultivated in DMEM/F12 medium (GIBCO) with 10% fetal bovine serum (Thermo Scientific HYCLONE). Colorectal adenocarcinoma cells Caco2 were cultivated in DMEM/F12 medium (GIBCO) with 20% fetal bovine serum. Growth medium contained 100 units/ml penicillin, and 100 mg/ml streptomycin (Sigma). Growth conditions were 5% CO2 at 37°C. Confluence of the cell layer was controlled regularly and cells were passaged at about 80% confluence to preserve optimal growth conditions.

Isolation of nuclear extract proteins from HepG2, Caco2, K562, HeLa cell lines

All manipulations were performed on ice; centrifugations were performed at 12000 rpm and 4°C. 5×10^5 cells were washed in 1ml ice-cold of PBS. Pellet was resuspended in a volume of hypo-osmotic lysis buffer (0,3 M sucrose; 2% Tween 40; 10 mM Hepes-KOH pH 7.9; 10 mM KCl; 1.5 mM $MgCl_2$; 0.1 mM EDTA; 0.5 mM DTT). Cell lysate was overlaid on 1 ml of 1.5 M sucrose buffer (1.5 M sucrose; 10 mM Hepes-KOH pH 7.9; 10 mM KCl; 1.5 mM $MgCl_2$; 0.1 mM EDTA; 0.5 mM DTT) and centrifuged for 10 min. Nuclear pellet was resuspended in 1 ml of low-salt wash buffer (10 mM Hepes-KOH pH 7.9; 10 mM KCl; 1.5 mM $MgCl_2$; 0.1 mM EDTA; 0.5 mM DTT) and centrifuged for 30 sec. Washed nuclear pellet was resuspended in 50 µl of high-salt extraction buffer (20 mM Hepes-KOH pH 7.9; 420 mM NaCl; 1.5 mM $MgCl_2$; 0.2 mM EDTA; 0.5 mM DTT; 25% glycerol). Following 20 min of extraction, the sample was centrifuged for 20 min and the supernatant was retained as nuclear proteins. All buffers contained Protease Inhibitor Cocktail (100X) (Sigma). Nuclear extract proteins concentration was measured using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific).

Electrophoretic Mobility Shift Assays

For each of both alleles of SNPs forward and reverse oligonucleotides were synthesized by BIOSSET Ltd. 31 bp DNA-probes were obtained by annealing.

Kinase reaction was conducted for 30 min at 37 °C in 30 µl reaction mixture contained 300 ng DNA-probe, 3 µl 10x T4 polynucleotide kinase buffer (SibEnzime); 35x104 Bq [γ - 32P] ATP; 5 active unit of T4- polynucleotide kinase (SibEnzime) with following inactivation for 15 min at 65 °C. The nuclear extract was incubated with sheared salmon sperm DNA (1 µg of DNA per 7 µg of total protein) for 15 min at 0°C. After that 4 µg of extract was added to the probes which contain 50 pM of labeled oligonucleotide. After incubation at room temperature for 15 min the mixture was subjected to electrophoresis in 4.5% PAAG in 0.5×TBE (89 mM Tris-borate, 89 mM H3BO3, 2 mM EDTA) at 4°C. Electrophoresis was conducted for 45 min at a voltage of 10 V/cm. Then glass plates containing polymerized gel were treated with fixing solution (20% ethanol, 10% glacial acetic acid), dried and scanned with a PharosFX system. Two to three independent replicates were performed for each EMSA experiment.

Plasmid constructs

Choice of vector depended on location of polymorphism in the gene. So, for SNPs located in the promoter regions, a promoter-less plasmid was used, and for SNPs in the protein-coding regions, a plasmid with a minimal promoter was used. For each of both alleles of rs11542583 and rs590352 localized in first exons (*COG8* and *ATXN7L3B* respectively) forward and reverse single, double and triple oligonucleotides were annealed. The size of the single insert was 31 bp, double insert – 62 bp, and triple – 93 bp. Then oligonucleotides were cloned into a linearized pGL4.23[minP/luc]. For rs2072580 (-200bp upstream of *SART3* TSS) and rs78317230 (-200bp upstream of *U2AF2* TSS) the DNA fragments (~ 500-bp) containing promoter regions were amplified using the DNA samples from patients homozygous for both alleles at first [6]. Then amplicons were cloned into a linearized pGl3-Basic. All constructs were confirmed by DNA sequencing.

Luciferase reporter assay

HepG2 cells were used for luciferase analysis. The plasmids under study were co-transfected with pRL-TK using Screen FectA (Incella GmbH, Germany). Luciferase activity was measured 24 h after transfection using the dual-luciferase reporter assay kit (Promega, USA). Relative light units from firefly luciferase were normalized to renilla luciferase activity. Differences in gene expression between plazmids carrying alternative alleles were determined using t-test. All experiments were conducted at least in technical and biological triplicates.

Results

Using EMSA, it was shown that all nucleotide substitutions G>C (rs590352), A>G (rs11542583), (T>C) rs3829202, T>C (rs78317230), A>T (rs2072580), C>T (rs4796672) change the pattern of nuclear extract proteins binding. This happens at least for one of cell lines. For example, in the case of rs78317230 DNA-probes containing T allele (Fig.1A) or C allele (Fig.1B) bind nuclear proteins differently. In Fig.1C one can see, that in contrast of weakening some bands, others are intensified and new ones appear. Also, this electrophoregram illustrates that DNA-probes bind nuclear extract proteins of K562 and HepG2 more actively in comparison to Caco2 and protein binding pattern (band position and intensity) for each cell line is unique. These data may indicate in tissue specific effects.

a

5′-GCCAGGGGGCAGTAA**T**GCCAGATCATCTGTC-3′
3′-CGGTCCCCCGTCATT**A**CGGTCTAGTAGACAG-5′

b

5′-GCCAGGGGGCAGTAA**C**GCCAGATCATCTGTC-3′
3′-CGGTCCCCCGTCATT**G**CGGTCTAGTAGACAG-5′

c

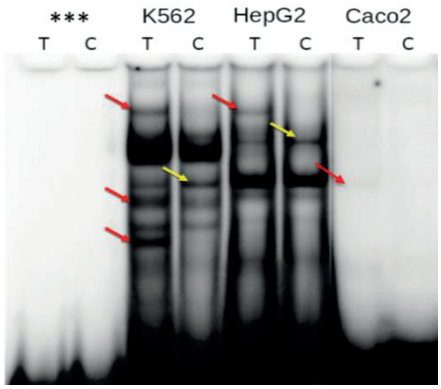


Fig. 1. *a* – nucleotide sequence of DNA-probe contaning allele T rs78317230; *b* – nucleotide sequence of DNA-probe contaning allele C rs78317230; *c* – EMSA results: *** – free DNA-probes; red arrows – intensification/ appearance of bands in the case of T allele; yellow arrows – intensification of bands in the presence of C allele.

As another example, result of EMSA for rs2072580 T>A is presented (Fig.2C). There are bands (red arrows in Fig.2C) that indicate that nuclear extract proteins of HepG2 and Caco2 are bound better with DNA-probe containing A allele. As for the nuclear extract proteins of K562, there are bands demonstrating the better binding to both DNA-probes carrying the A (red arrow) or T (yellow arrow) alleles (Fig. 2C).

a

5′-ATGGCATGCGTTCCG**A**CACAACTTCGGACAG-3′
3′-TACCGTACGCAAGGC**T**GTGTTGAAGCCTGTC-5′

b

5′-ATGGCATGCGTTCCG**T**CACAACTTCGGACAG-3′
3′-TACCGTACGCAAGGC**A**GTGTTGAAGCCTGTC-5′

c

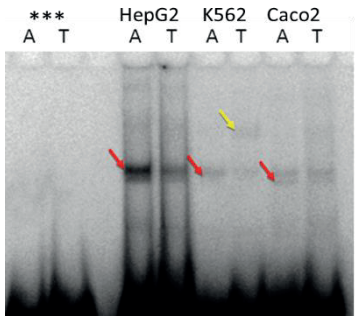


Fig. 2. *a* – nucleotide sequence of DNA-probe contaning allele A rs2072580; *b* – nucleotide sequence of DNA-probe contaning allele T rs2072580; *c* – EMSA: *** – free DNA-probes; red arrows – intensification/ appearance of bands in the case of A allele; yellow arrow – appearance of bands in the presence of T allele.

Reporter constructs containing Luciferase gene were created for 4 (rs590352, rs11542583, rs78317230, rs2072580) of 6 polymorphisms. They were of two types: the first type are constructs based on the plasmid with minimal promotor, because SNPs (rs590352, rs11542583) are located in exon. Constructions of this type included 3 subtypes - these are constructions with a single insert, double and triple. Constructions of the second type are constructs based on promotor-less plasmid, because SNPs (rs78317230, rs2072580) are located in the promoter region

It was shown that all studied nucleotide substitutions influenced the expression of the reporter gene. So, in the case of inserting the promoter region of *U2AF2* containing allele C (rs78317230) the expression of the reporter gene was increased in comparison to allele T (Fig.3A). In the case of inserting the *SART3* promoter region with the allele T (rs2072580) the expression of the reporter gene was decreased in comparison to allele A (Fig.3B.) As for rs11542583, in the case of single and double inserts with A or G allele there was no significant difference in luciferase expression. However, increase of expression was indicated with triple insert of A allele (Fig.3C). Substitution G>C (rs590352) in the case of double and triple oligonucleotide inserts leads to reduced luciferase gene expression.

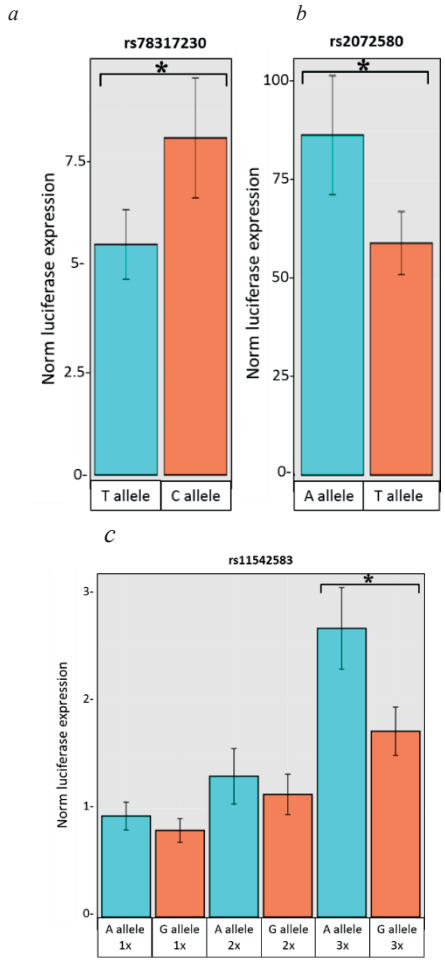


Fig. 3. Effect of nucleotide substitutions on reporter gene expression: *a* – normalized expression of firefly luciferase in the presence of T allele (blue column) or C allele (orange column) rs78317230; *b* – normalized expression of firefly luciferase in the presence of A allele (blue column) or T allele (orange column) rs2072580; *c* – normalized expression of firefly luciferase in the presence of single, double and triple insert with A allele (blue column) or G allele (orange column); n=3, * – $p<0.05$.

Thus, based on the results of EMSA and the luciferase reaction, it can be proposed that the studied polymorphisms can affect the expression of the genes to which they relate. 2 polymorphisms (rs590352 and rs11542583) are located in the the protein-coding regions of the genes (*ATXN7L3B* and *COG8* respectively) and are synonymous substitutions. ATXN7L3B reduces the activity of the SAGA complex that realizes cotranscriptional modifications of histones participating in the regulation of gene expression [7]. COG8 is a component of the conserved oligomeric Golgi complex and is involved in intracellular membrane trafficking and glycoprotein modification. *KLF6* gene (rs3829202) encodes a member of the Kruppel-like family of transcription factors. It is a transcriptional activator, and functions as a tumor suppressor. *KRT15 (*rs4796672) encodes a member of the keratin gene family, coding for intermediate filament proteins of epithelia [8]. According to literature, a considerable increase in the KRT15 is observable in hepatocellular carcinoma [9], squamous cell carcinoma in non-small cell lung cancer [10], a subset of urothelium cellcarcinomas [11], and cervical cancer [12]. *U2AF2* (rs78317230) encodes the U2AF large subunit which is necessary for splicing. *SART3* (rs2072580) encodes a multifunctional protein involved in splicing and transcription regulation [13]. Thus, the listed genes encode proteins that perform important functions, therefore expression change of these genes can lead to serious consequences.

Conclusions

As a result, it was shown that all single nucleotide substitutions (rs590352 G>C, rs11542583 A>G, rs3829202 T>C, rs78317230 T>C, rs2072580 A>T, rs4796672 C>T) affect the binding of nuclear proteins. This indicates their regulatory potential. For 4 polymorphisms (rs590352, rs11542583, rs78317230, rs2072580), it was confirmed in the luciferase reaction.

Acknowledgment

This work was supported by the RFBR grant 18-29-09041 and Budget Project 0259-2019-0010-C-01.

References

[1] S.L. Edwards, J. Beesley, J.D. French, and A.M. Dunning, "Beyond GWASs: Illuminating the Dark Road from Association to Function", Am J Hum Genet. 2013. vol. 93, N. 5, pp. 779–97

[2] T. Lappalainen, " Functional genomics bridges the gap between quantitative genetics and molecular biology", Genome Res. 2015. vol. 25, N 10, pp.1427-31.

[3] D.Welter, J. MacArthur, J. Morales, T.Burdett, P. Hall, H. Junkins, et al, "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations", Nucleic Acids Res. 2014, 42(Database issue):D1001-6.

[4] K. K. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W.J. Housley, S. Beik, et al, "Genetic and epigenetic fine mapping of causal autoimmune disease variants", Nature. 2015. vol.7539, N 518. pp. 337-43.

[5] E.E. Korbolina, I.I Brusentsov., L.O Bryzgalov., E.Yu Leberfarb., A.O Degtyareva and T.I Merkulova, "Novel Approach to Functional SNPs Discovery from Genome-Wide Data Reveals Promising Variants for Colon Cancer Risk", Hum Mutat. 2018. vol. 39, N. 6. pp. 851–59.

[6] E.Y. Leberfarb**,** A.O. Degtyareva, I.I. Brusentsov, V.N. Maximov, M.I. Voevoda, A.I. Autenshlus, et al, "Potential regulatory SNPs in the ATXN7L3B and KRT15 genes are associated with gender-specific colorectal cancer risk", Per Med. 2020. vol.17, N 1. pp. 43-54

[7] W. Li, B.S. Atanassov, X. Lan, R.D. Mohan, S.K. Swanson, A.T. Farria, et al, "Cytoplasmic ATXN7L3B Interferes with Nuclear Functions of the SAGA Deubiquitinase Module", Mol. Cell. Biol. 2016. vol.36. N. 22. pp. 2855–2866

[8] R. Moll, M. Divo and L. Langbein, “The human keratins: biology and pathology”, Histochem. Cell Biol. 2008. vol. 129. N. 6. pp. 705–733

[9] A-RN Zekri, ER El-Sisi, ZF Abdallah, A Ismail, et al, “Gene expression profiling of circulating CD133+ cells of hepatocellular carcinoma patients associated with HCV infection”, J.Egypt. Natl. Canc. Inst. 2017. vol.29, N.1.pp. 19–24

[10] A. Sanchez-Palencia, M. Gomez-Morales, JA Gomez-Capilla, V. Pedraza, L. Boyero, R. Rosellet, al, “Gene expression profilingreveals novel biomarkers in nonsmall cell lung cancer”, Int. J. Cancer. 2011. vol.129, N.2. pp. 355–364.

[11] G. Tai, P. Ranjzad, F. Marriage, S. Rehman, H. Denley, J. Dixon, et al, “Cytokeratin 15 marks basal epithelia in developing ureters and is upregulated in a subset of urothelial cell carcinomas”, PLoS One. 2013. vol.8, N.11. p. e81167

[12] F. Smedts, F. Ramaekers, R.E. Leube, K. Keijser, M. Link, P. Vooijs, “Expression of keratins 1, 6, 15, 16, and 20 in normal cervical epithelium, squamous metaplasia, cervical intraepithelial neoplasia, and cervical carcinoma”. Am. J. Pathol. 1993. vol.142, N.2. pp. 403–12

[13] A. Whitmill, K.A. Timani, Y. Liu and J.J. He, “Tip110: Physical properties, primary structure, and biological functions”, Life Sci. 2016. vol.149. pp. 79–95

DOI 10.18699/BGRS/SB-2020-264

Influence of the factors of maternal milieu on taste preferences and metabolic parameters in mouse male and female offspring

Elena Denisova
Laboratory of Physiological Genetics
Institute of Cytology and Genetics SB
RAS Novosibirsk, Russia

Elena Makarova
Laboratory of Physiological Genetics
Institute of Cytology and Genetics SB
RAS Novosibirsk, Russia

Maria Savinkova
Department of Physiology
Novosibirsk State University
Novosibirsk, Russia

Keywords — leptin, developmental programming, obesity, taste preferences

Introduction

Obesity is now a leader among noncommunicable diseases. Calorie overload induces the obesity development, and the preference and availability of sweet and fat foods contribute to obesity spread. It was shown that prenatal and early postnatal conditions affect the susceptibility to obesity and may influence on taste preferences [1], however, the mechanisms mediating maternal effects on taste preferences in offspring are unknown. The adipose tissue hormone leptin may be one of the factors mediating maternal influence on offspring phenotype. It was shown that an increase in the level of leptin in the blood of pregnant mice counteracts the development of diet-induced obesity in the offspring [2, 3], and the programming leptin effects may be different in the offspring of different sexes. It is not known whether the beneficial programming effect of maternal leptin on the susceptibility to obesity is associated with its effect on taste preferences. The aim of this study was to evaluate the effect of leptin administration to pregnant mice on metabolic phenotype, the rate of diet-induced obesity and taste preferences in offspring of different sexes.

Methods

C57BL/6J mice were housed under a 12:12-h light-dark regime and were provided *ad libitum* access to commercial mouse chow and water. The females were mated with the males at 12-15 weeks of age and were housed individually from the day a copulatory plug was detected (gestational day 0, GD0). The females received subcutaneous injections of recombinant murine leptin (2.0 μg per g BW) during 3 days of pregnancy (GD11, GD12, GD13). Female body weight (BW) and food intake (FI) was measured daily during all period of pregnancy. At birth, the pups were weighed, and the litters that contained more than seven pups were reduced to seven pups. The pup BWs on postpartum days 1, 7, 14, 21, and 28 were measured. On postpartum day 28, one male and one female from each litter were separated from their mothers and housed individually, and their BWs and FI were measured once a week until the age of 10 weeks. During this period, all animals were fed a standard chow diet *ad libitum*. From the age of 10 weeks the mice began to receive sweet biscuits and lard in addition to the standard chow. This mixture mimics the cafeteria diet and potentiates the rapid development of obesity in mice [5]. Mouse BWs and consumption of biscuits, lard and standard chow were measured over the course of 10 weeks. At the end of the experiment, the animals were sacrificed by decapitation, and samples of liver were collected to measure gene expression.

Results

Leptin administration to pregnant mice reduced their FI by 18-20%, and did not affect their BWs and fetal viability, as average litter size in females treated with leptin or saline did not differ. Leptin administration did not affect the weight of pups during lactation, but had an impact on BW and FI in female offspring on a standard diet, and in male offspring on a sweet and fat diet. Beginning from the age of 8 weeks (the age of sexual maturation), the females born to leptin-treated mothers consumed less standard food and gained less weight than the female born to saline-treated mothers (repeated measures ANOVA, $P < 0.05$). However, female offspring did not differ in the rate of development of diet-induced obesity, and the body weight differences observed previously were maintained when female offspring consumed sweet and fat food. Leptin administration did not affect FI and BW in male offspring on a standard diet, and decreased the rate of weight gain on sweet and fat diet (repeated measures ANOVA, $P < 0.05$), Fig. 1. These results are consistent with previous reports [2, 3] indicating the protective effect of maternal leptin on susceptibility to diet-induced obesity in mouse offspring.

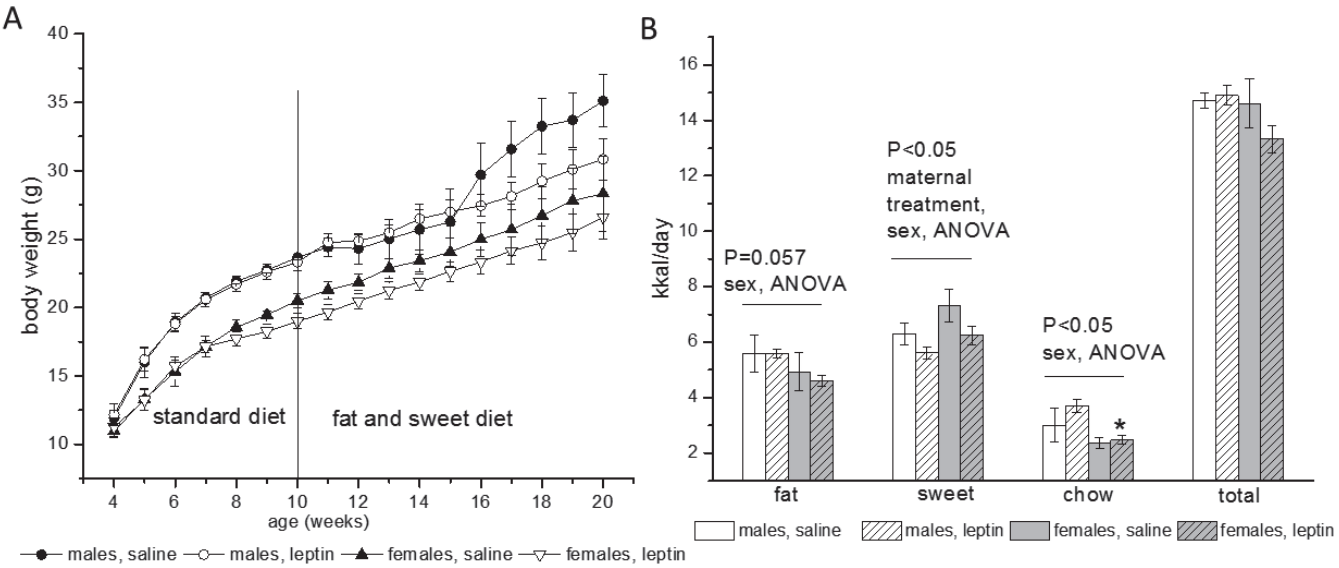


Fig. 1. Influence of leptin administration to pregnant mice on body weight changes during standard and fat and sweet diet feeding (A), and on taste preferences (B) in male and female offspring. B – Calorie intake per day from fat (lard), sweet (biscuits), and standard pelleted chow and total intake. Male and female mice born to saline- and leptin-treated mothers were fed with standard diet until the age of 10 weeks and with high-calorie fat and sweet diet after that during 10 weeks. The results are presented as an arithmetic mean ± SE. *P<0.05 between males and females born to leptin-treated mothers, Post Hoc Duncan test

When fed a sweet and fat diet, males and females differed in the taste preferences: females consumed less standard pelleted chow and more sweet biscuits than males (ANOVA, $P < 0.05$). Leptin administration to pregnant mothers reduced the consumption of sweet biscuits in offspring of both sexes (ANOVA, $P < 0.05$), but did not significantly affect the calorie intake (Figure 1B). The results suggest that programming of taste preferences may be one of the mechanisms mediating the influence of maternal leptin on susceptibility to obesity in progeny, however, this issue requires further study, since no maternal effect on the energy intake has been found.

In the offspring of saline-treated mothers, sex differences in the liver expression of Fgf21 gene were observed. Fgf21 mRNA level was higher in males, which is in line with the data obtained earlier in mice with diet-induced obesity [4, 5]. Leptin administration to pregnant mice reduced the level of FGF21 mRNA in male offspring, and sex differences in Fgf21 expression disappeared. FGF21 is one of the potent regulators of carbohydrate and lipid metabolism and insulin sensitivity. The results suggest that maternal leptin can differently effect the metabolism of male and female offspring via sex-specific programming of FGF21function.

Conclusion

Increased level of leptin during pregnancy can be a factor that reduces the risk of diet-induced obesity in offspring of both sexes, but the mechanisms mediating programming effect of leptin can vary in offspring of different sexes. The long-lasting effect of maternal leptin on offspring taste preferences and liver FGF21 expression may be one of the reasons for its beneficial effect on the susceptibility to obesity in offspring.

ACKNOWLEDGMENT

The study was supported by Russian Foundation for Basic Research (project 17-04-01357-a and 20-015-00469-a) and Budget project 03424-2019-0041-c-01.

REFERENCES

[1] B. S. Muhlhausler and Z. Y. Ong, “The fetal origins of obesity: early origins of altered food intake”, Endocr Metab Immune Disord Drug Targets, vol. 11(3), pp. 189-197, Sep 2011.

[2] E. N. Makarova, E. V. Chepeleva, P. E. Panchenko, and N. M. Bazhan, “Influence of abnormally high leptin levels during pregnancy on metabolic phenotypes in progeny mice”. *Am J Physiol Regul Integr Comp Physiol.,* vol.305, pp. R1268–R1280, 2013.

[3] K. E. Pollock, D. Stevens, K. A. Pennington, R. Thaisrivongs, J. Kaiser, M. R. Ellersieck, D. K. Miller, and L. C. Schulz, “Hyperleptinemia during pregnancy decreases adult weight of offspring and is associated with increased offspring locomotor activity in mice”, Endocrinology, vol. 156, pp. 3777–3790, 2015.

[4] N. Bazhan, T. Jakovleva, N. Balyibina, A. Dubinina, E. Denisova, N. Feofanova, and E. Makarova, “Sex dimorphism in the Fgf21 gene expression in liver and adipose tissues is dependent on the metabolic condition”, Online J. Biol. Sci., vol. 19(1), pp. 28–36, 2019.

[5] F. R. S. Gasparin, F. O. Carreño, J. M. Mewes, E. H. Gilglioni, C. L. S. Pagadigorria, M. R. M. Natali, K. S. Utsunomiya, et al., “Sex differences in the development of hepatic steatosis in cafeteria diet-induced obesity in young mice”, Biochim. Biophys. acta. Mol. Basis Dis., vol. 1864(7), pp. 2495–2509, 2018.

DOI 10.18699/BGRS/SB-2020-265

Hydrogels for tissue regeneration

Timothy E.L. Douglas
Engineering Department and Materials
Science Insitute, Lancaster University,
United Kingdom
t.douglas@lancaster.ac.uk

Abstract — **Mineralization of hydrogels, or 3D polymer networks containing entrapped water, is considered desirable for applications in bone regeneration. We will present several strategies to mineralize hydrogels, including addition of pre-formed mineral particles, enzymatic mineralization and incorporation of calcium-binding substances.**

Keywords — hydrogel, mineralization, bone, composite

Motivation and aim

Biomaterials for bone regeneration have predominantly been fabricated from inorganic substances such as various forms of calcium phosphate (CaP), e.g.hydroxyapatite, tricalcium phosphate and brushite. CaP materials are mechanical stable and bioactive, i.e. they form a direct bone with surrounding bone tissue. However, such pure CaP materials have certain drawbacks. They are brittle, difficult to handle in granulate form and difficult to shape in block form. Furthermore, the incorporation of biologically active substances is not easy.

Hydrogels are highly hydrated three-dimensional polymer networks that are formed by crosslinking of polymer chains in solution. Hydrogels have been widely used as vehicles for drug delivery and are being used increasingly as biomaterials for tissue regeneration. As their main component is water, they have many advantages over pure inorganic materials. Firstly, the incorporation of water-soluble biologically active substances to promote tissue growth (e.g. growth factors) or to combat infection (e.g. antibiotics) is straightforward. Secondly, they are much less brittle. Thirdly, they can be implanted in a minimally invasive manner by injection, as they can undergo gelation, i.e. the transition from liquid to solid, after injection. However, their main disadvantage also stems from the fact that the mail component is water: hydrogels are mechanically weak. In order to combine the advantages of inorganic and hydrogel biomaterials, attention has recently been focused on the development of composites on the basis of mineralized hydrogels.

Methods

Several strategies have been tried [1]. The most common strategy is the addition of preformed inorganic particles to the polymer solution before gelation, after which the particles remain entrapped in the crosslinked polymer network. Ideally, the particles can be distributed homogeneously in the hydrogel. The gelation process can be induced by addition of inorganic particles. For example, the addition of bioactive glass particles to a solution of the anionic polysaccharide gellan gum results in hydrogel formation due to release of ions from the particles [2]. In other words, the particles serve as an "ion-delivery system" to provide homogeneous gelation. Another strategy is to promote precipitation of the inorganic phase in the hydrogel by increasing the concentration of ions. This can be achieved biomimetically using the enzyme alkaline phosphatase (ALP) which is responsible for the mineralization of bone tissue in vivo by cleaving phosphate ions from organophosphate and thus increasing the local phosphate concentration, which in turn promotes CaP precipitation [3]. Yet another strategy is the incorporation of calcium- or phosphate-binding molecules in the hydrogel, in order to increase local ion concentrations and promote CaP precipitation. Once such biomolecule is polydopamine, which binds calcium ions [4]. An added flexibility of mineralized hydrogels is the possibility of manipulation of either the hydrogel phase, or the inorganic phase, or both. For example, in the case of a hydrogel mineralized with CaP, the inorganic phase may be modified by incorporation of magnesium in order to promote adhesion and proliferation of bone-forming cells [5], or by incorporation of zinc in order to endow antibacterial activity [6]. Alternatively, the hydrogel phase may be modified by incorporation of biologically active molecules such as polyphenols, which both bind calcium ions and exhibit antibacterial activity [7].

Results

All the aforementioned strategies have been successfully applied to mineralize hydrogels in previous work [1-8]. More recently research in our group has focussed on a novel material commonly used in food industry, namely whey protein isolate (WPI), which can be combined with particles to improve mechanical properties and cell behaviour [9-11].

Acknowledgment

FWO, Belgium and N8 Agrifood grant "Food2Bone", UK.

References

[1] Gkioni K, et al. (2010) Mineralization of hydrogels for bone regeneration. Tissue Eng Part B Rev. 16(6): 577-85
[2] Douglas TE et al. (2014) Injectable self-gelling composites for bone tissue engineering based on gellan gum hydrogel enriched with different bioglasses. Biomed Mater. 9(4): 045014.
[3] Douglas TE, et al. (2012) Enzymatic mineralization of hydrogels for bone tissue engineering by incorporation of alkaline phosphatase. Macromol Biosci.12(8): 1077-89.
[4] Douglas TE, et al. (2014) Enzymatic mineralization of gellan gum hydrogel for bone tissue-engineering applications and its enhancement by polydopamine. J Tissue Eng Regen Med. 8(11): 906-18.
[5] Douglas TE, et al. (2016) Generation of composites for bone tissue-engineering applications consisting of gellan gum hydrogels mineralized with calcium and magnesium phosphate phases by enzymatic means. J Tissue Eng Regen Med. 10(11): 938-954.
[6] Douglas TE, et al. (2017) Composites of gellan gum hydrogel enzymatically mineralized with calcium-zinc phosphate for bone regeneration with antibacterial activity. J Tissue Eng Regen Med. 11(5): 1610-1618.
[7] Douglas TE, et al. (2016). Biomed Mater. 11(4): 045015.
[8] Douglas TE, et al. (2018). J Tissue Eng Regen Med. 12(6): 1313-1326
[9] Dziadek M, et al. (2019) Novel multicomponent organic-inorganic WPI/gelatin/CaP hydrogel composites for bone tissue engineering. J Biomed Mater Res A. 107(11): 2479-2491.
[10] Dziadek M, et al. (2020). Mater Lett. 261: 127115
[11] Gupta D, et al. (2020) Novel naturally derived whey protein isolate and aragonite biocomposite hydrogels have potential for bone regeneration. Materials and Design. 188: 108408.

DOI 10.18699/BGRS/SB-2020-266

Association between total plasma N-glycan levels and chronic back pain: a prospective study

Elizaveta E. Elgaeva
Laboratory of Glycogenomics, Institute of Cytology and Genetics
Novosibirsk, Russia
elgaeva@bionet.nsc.ru

Yakov A. Tsepilov
Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University
Novosibirsk, Russia
tsepilov@bionet.nsc.ru

Irena Trbojevic Akmacic
Genos Ltd.
Zagreb, Croatia
iakmacic@genos.hr

Andrea Skelin
St Catherine Hospital
Zabok, Croatia
andrea.skelin@svkatarina.hr

Concetta Dagostino
University of Parma
Parma, Italy
cdagostino@parmanesthesia.com

Massimo Allegri
Pain Therapy Department, Policlinico Monza Hospital
Monza, Italy
allegri@italianpaininstitute.com

Jan Van Zundert
Ziekenhuis Oost-Limburgl
Genk, Belgium
Jan.VanZundert@zol.be

Dragan Primorac
St Catherine Hospital
Zabok, Croatia
draganprimorac2@gmail.com

Gordan Lauc
Genos Ltd.
Zagreb, Croatia
glauc@genos.hr

Frances M. K. Williams
Kings College London
London, UK
frances.williams@kcl.ac.uk

Maxim B. Freidin
Kings College London
London, UK
maxim.freydin@kcl.ac.uk

Abstract — **Current prospective study identifies variation of N-glycosylation patterns in total blood plasma associated with chronic back pain.**

Keywords — low back pain, N-glycans, total plasma N-glycome, glycosylation

Introduction

Low back pain (LBP) is one of the most prevalent and disabling health problems worldwide, incorporating different pain syndromes [1]. One of the probable causes of LBP is local inflammation,. Inflammatory reactions have been reported to be associated with alterations in N-glycosylation of several serum glycoproteins [2] and total blood plasma protein [3]. Given that glycans modify the protein structure and affect their function [4], changes in glycosylation are implicated in the pathogenesis of different diseases and may be considered as biomarkers in some [5–8].

Our previous retrospective case-control study identified an association between total plasma N-glycome and chronic LBP (CLBP) [9]. The current prospective study addresses the question of temporal variation of total plasma N-glycome in the development of CLPB and whether plasma N-glycome can serve as a predictive biomarker for CLBP.

Subjects and methods

Clinical samples and data

The sample comprised 1,114 Europeans recruited at baseline (T0) when they experienced a serious episode of acute LBP and followed-up at 3 months (T1). Total sample size was 1,085 people at T0 and 549 people at T1. For n=520 participants data were obtained at both time points.

Blood samples and clinical data were collected in five centers located in the United Kingdom, Belgium, Croatia (2 centers) and Italy in the EU FP7 PainOmics project [10, 11].

Glycan profiles were measured for 39 original glycan traits and summarised into 14 derived traits. Cases and controls were defined according to presence or absence of LBP at T1: those who reported LBP at T1 were considered to have evolved into CLBP. Information about pain type (spontaneous or after injury), probable pain generator and neuropathic pain was also collected.

Statistics

The analysis was carried out in each center cohort separately followed by the meta-analysis. For each cohort, glycan profiles were adjusted for age, sex and body mass index (BMI) and the inverse rank-transformation to normality was performed. Case-control comparisons were performed using generalized linear model. To estimate the predictive power of total plasma N-glycome we calculated the area under the ROC-curve, AUC value. Meta-analysis of the results was performed using metaphor v2.1.0 R package. Paired t-test was applied to detect alterations in glycan levels between T0 and T1 in cases and controls; the results were meta-analyzed using metap v1.1 R package. Additionally, we ran canonical correlation analysis to obtain correlation matrices within and between cases and controls in each cohort for both time points.

We checked for pain generator effect using ANOVA and repeated all the analyses in the subsets of data split by different pain generators and pain types. Some tests were also run on principle components of glycan level traits and neuropathic pain data. All the analyses were performed in R programming language in RStudio, R versions 3.5.3 and 3.6.1. Significance was taken as $p\text{-}value<0.05$ and correction for multiple testing was $0.05/14\approx0.00357$.

Results

1. We addressed three questions: 1) is there a difference between cases and controls at T0?; 2) is there a difference in glycome profile between cases and

controls at T1?; and 3) is there a difference between T0 and T1 in cases and controls?

2. No statistically significant association between levels of total plasma N-glycan at T1 and CLBP was revealed; however, we detected negative association under nominally significant threshold for neutral, monogalactosylated and high-mannose glycans, consistent with our previous retrospective study [9].
3. We detected statistically significant differences in N-glycosylation pattern between T0 and T1 in controls for low- and high-branching glycans, trisialylated, di- and trigalactosylated glycans. No such differences were found for cases, suggesting that people who develop CLBP keep their pro-inflammatory profile of plasma N-glycans.
4. We did not find evidence of predictive capacity of N-glycan profile for future development of CLBP. We also did not reveal interactions between pain-generator and N-glycome of plasma. Results obtained on principle components and in different subsets of data split by pain types and generators were consistent with those obtained in original data.

ACKNOWLEDGMENT

The study was funded by EU FP7 PainOmics project (602736). The work of EEE and YAT was supported by the Russian Foundation for Basic Research (project 19-015-00151).

REFERENCES

[1] T. Vos, R.M. Barber, B. Bell, A. Bertozzi-Villa, S. Biryukov et al., "Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013", Lancet. 386 (2015) 743–800. doi:10.1016/S0140-6736(15)60692-4.

[2] O. Gornik, G. Lauc, "Glycosylation of serum proteins in inflammatory diseases", Dis. Markers. 25 (2008) 267–278. doi:10.1155/2008/493289.

[3] M. Novokmet, E. Lukić, F. Vučković, Ž. Đurić, T. Keser et al., "Changes in IgG and total plasma protein glycomes in acute systemic inflammation", Sci. Rep. 4 (2014) 1–10. doi: 10.1038/srep04347.

[4] G.P. Subedi, A.W. Barb, "The Structural Role of Antibody N-Glycosylation in Receptor Interactions", Structure. 23 (2015) 1573–1583. doi: 10.1016/j.str.2015.06.015.

[5] R. Goulabchand, T. Vincent, F. Batteux, J.F. Eliaou, P. Guilpain, "Impact of autoantibody glycosylation in autoimmune diseases", Autoimmun. Rev. 13 (2014) 742–750. doi: 10.1016/j.autrev.2014.02.005.

[6] F. Vučkovïć, J. Krištïć, I. Gudelj, M. Teruel, T. Keser et al., "Association of systemic lupus erythematosus with decreased immunosuppressive potential of the IgG glycome", Arthritis Rheumatol. 67 (2015) 2978–2989. doi: 10.1002/art.39273.

[7] I. Trbojević Akmačić, N.T. Ventham, E. Theodoratou, F. Vučković, N.A. Kennedy et al., "Inflammatory Bowel Disease Associates with Proinflammatory Potential of the Immunoglobulin G Glycome", Inflamm. Bowel Dis. 21 (2015) 1237–47. doi: 10.1097/MIB.0000000000000372.

[8] F.H. Routier, E.F. Hounsell, P.M. Rudd, N. Takahashi, A. Bond et al., "Quantitation of the oligosaccharides of human serum IgG from patients with rheumatoid arthritis : a critical evaluation of different methods", J. Immunol. Methods. 2013 (1998) 113–130. doi: 10.1016/s0022-1759(98)00032-5.

[9] I. Trbojević-Akmačić, F. Vučković, M. Vilaj, A. Skelin, L.C. Karssen et al., "Plasma N-glycome composition associates with chronic low back pain", Biochim Biophys Acta Gen Subj. 1862 (2018) 2124-2133. doi:10.1016/j.bbagen.2018.07.003.

[10] M. Allegri, M. De Gregori, C.E. Minella, C. Klersy, W. Wang et al., ""Omics" biomarkers associated with chronic low back pain: protocol of a retrospective longitudinal study", BMJ Open. 6 (2016) e012070.

[11] C. Dagostino, M. De Gregori, C. Gieger, J. Manz, I. Gudelj, G. Lauc et al., "Validation of standard operating procedures in a multicenter retrospective study to identify -omics biomarkers for chronic low back pain", PLoS One. 12 (2017) e0176372. doi: 10.1371/journal.pone.0176372.

DOI 10.18699/BGRS/SB-2020-267

Immunoglobulins with proteolytic activity as a biomarker of impaired humoral immune system in schizophrenia

Evgeny Ermakov
ICBFM SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
evgeny_ermakov@mail.ru

Valentina Buneva
ICBFM SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
buneva@niboch.nsc.ru

Georgy Nevinsky
ICBFM SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
nevinsky@niboch.nsc.ru

Abstract — **Schizophrenia is known to be associated not only with imbalance in the neurotransmitter systems but also with immune system dysregulation. In this work, we found that pure polyclonal IgG preparations of patients with schizophrenia possess histone hydrolyzing proteolytic activity. The presence of proteolytic activity of IgGs is evidence of impaired humoral immune system in schizophrenia, which can be used as a biomarker for immunophenotyping of patients.**

Keywords — immunoglobulins, IgG, proteolytic activity, histones, schizophrenia

Motivation and aim

There is a plethora of evidence that inflammation and dysregulation of the immune system is associated with schizophrenia [1]. One of the important signs of impaired humoral immunity is the generation of catalytic antibodies (Abs) that not only binding antigens but also hydrolyze these molecules [2]. Catalytic Abs were found among the total pool of immunoglobulins in autoimmune and viral diseases. Recently, we detected catalytic IgGs hydrolyzing DNA and RNA in serum of patients with schizophrenia [3, 4]. The formation of such Abs can be associated with immunization with nucleic acids and their complexes with nucleosome and histones. Autoantibodies to DNA were found in patients with schizophrenia [5]. Therefore, we suggested that such Abs can hydrolyze not only DNA and RNA, but also histones.

The aim of this work was to investigate the ability to hydrolyze various human histones by polyclonal IgGs from serum of patients with schizophrenia.

Methods

We recruited 25 patients (11 men and 14 women) with a verified diagnosis of paranoid schizophrenia and 20 healthy donors (9 men and 11 women) in the study. IgG preparations were obtained by Protein-G affinity chromatography of the serum proteins. The purity and homogeneity of obtained IgG preparations was verified using SDS-PAGE with silver staining, western blotting and MALDI mass-spectrometry. Catalytic activity of IgGs were revealed by the degree of hydrolysis of human histones as a substrate. IgGs were incubated with the substrate for 20 hours. The products of the hydrolysis were revealed by 20 % SDS-PAGE and Coomassie G250 staining as in [6]. The gels were analyzed by scanning and quantified using Image Lab 6.0 software (Bio-Rad, USA). Statistical analysis was carried out using the Mann-Whitney U-test and Spearman's rank correlation in the Statistica 10 software (StatSoft, USA).

Results

Screening analysis showed that IgG preparations of patients with schizophrenia possessed histone-hydrolyzing activity and are able to hydrolyze various histones with different efficiencies. Using strict generally accepted criteria, we proved that the activity was due to IgGs, and was not a consequence of the impurities of various canonical proteases. The level of histone hydrolyzing activity of IgG varied over a wide range (from 0 to 100%). The level of histones hydrolysis by IgGs of patients with schizophrenia was statistically significantly higher ($p<0.02$) than in healthy donors. Also, the level of histone hydrolysis by IgGs was significantly different ($p<0.05$) in patients with positive and negative symptoms. According to the results of a retrospective analysis, a positive correlation ($R>0.45$) was found between the level of histone hydrolysis and the level of DNA-hydrolyzing activity of antibodies in schizophrenia.

Thus, we found new evidence of impaired humoral immune system in schizophrenia. We suggest that histone-hydrolyzing antibodies may play a positive role in schizophrenia pathogenesis because histone removal from circulation minimizes the immune response. But the catalytic activity of IgG reflects the degree of derangement of immune system; therefore, it can be used for immunophenotyping and stratification of patients with schizophrenia. Identification of subgroups of patients with severe immunological disturbances will allow them to recommend additional anti-inflammatory therapy.

Acknowledgment

This work was supported by RFBR (20-015-00156), by Russian State funded budget project of ICBFM SB RAS (AAAA-A17-117020210023-1) and scholarship of the President of the Russian Federation (SP-2258.2019.4).

References

[1] Miller B.J., Goldsmith D.R. (2017). Neuropsychopharmacology, 42(1), 299-317.
[2] Nevinsky G.A., Buneva, V.N. (2003). Catalytic antibodies in healthy humans and patients with autoimmune and viral diseases. Journal of cellular and molecular medicine, 7(3), 265-276.
[3] Ermakov E.A., et al. (2015). DNA-hydrolysing activity of IgG antibodies from the sera of patients with schizophrenia. Open biology, 5(9), 150064.
[4] Ermakov E.A., et al. (2018). Hydrolysis by catalytic IgGs of microRNA specific for patients with schizophrenia. IUBMB life, 70(2), 153-164.
[5] Smirnova L.P., et al. (2016). Zhurnal nevrologii i psikhiatrii imeni SS Korsakova, 116(4), 47-51.
[6] Baranova S.V., et al. (2019). Antibodies from the Sera of Multiple Sclerosis Patients Efficiently Hydrolyze Five Histones. Biomolecules, 9(11), 741.

DOI 10.18699/BGRS/SB-2020-268

Bone remodeling in men with type 2 diabetes: is it just the same thing as in women?

Olga Fazullina
Laboratory of Endocrinology
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences
(RICEL – Branch of ICG SB RAS)
Novosibirsk, Russia
ORCID: 0000-0002-5868-579X

Maksim Dashkin
Laboratory of Endocrinology
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences
(RICEL – Branch of ICG SB RAS)
Novosibirsk, Russia
ORCID: 0000-0002-5099-5144

Vadim Klimontov
Laboratory of Endocrinology
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences
(RICEL – Branch of ICG SB RAS)
Novosibirsk, Russia
ORCID: 0000-0002-5407-8722

Abstract — **Background and aim: The mechanisms of reducing the bone mineral density (BMD) in men with type 2 diabetes are poorly understood. The aim of our study was to determine the relationships between the markers of bone remodeling and BMD in men with type 2 diabetes.**

Materials and Methods: The study included 59 men with type 2 diabetes, from 50 to 75 years of age. BMD and T-score were determined by dual-energy X-ray absorptiometry. A serum levels of parathyroid hormone (PTH), free testosterone, osteocalcin, osteoprotegerin, sclerostin, and urinary excretion of C-terminal telopeptides of type I collagen (CTX-I) were determined by ELISA. Control group comprised of 21 healthy subjects with normal BMD, matched by sex and age.

Results: A reduced BMD was revealed in 29 patients, including 4 individuals with osteoporosis and 25 subjects with osteopenia. The levels of osteocalcin were decreased and the levels of osteoprotegerin and sclerostin were increased in observed diabetic subjects as compared to control ($p=0.02$, $p<0.001$ and $p=0.02$ respectively). The excretion of CTX-1 was reduced in patients with diabetes ($p<0.001$). There were no differences in PTH and free testosterone concentrations between control and diabetic subjects. In stepwise multivariate regression analysis, sclerostin was the most significant predictor for lumbar spine T-score ($\beta=0.496$, $R^2=0.23$, $p=0.00007$), the level of PTH influenced the femoral neck T-score ($\beta=-0.29$, $R^2=0.26$, $p=0.005$).

Conclusions: The obtained results suggest that the bone remodeling in men with type 2 diabetes is reduced due to the inhibition of osteoblastogenesis and decrease in the bone formation and resorption.

Keywords — diabetes; bone mineral density; osteoporosis; bone remodeling

Introduction

A decrease in bone mineral density (BMD) is observed in more than 20% of patients with type 2 diabetes [1,2]. The previous studies assessed the risk factors for postmenopausal osteoporosis in diabetic subjects mainly [3], while the mechanisms for reducing BMD in men with type 2 diabetes are poorly understood. As compared to women, osteoporosis in men is less common, but mortality in the first year after a hip fracture in men is 51% higher [4]. A disturbance of bone remodeling plays an important role in pathogenesis of decline BMD in patients with type 2 diabetes [5-7]. In postmenopausal women with type 2 diabetes, osteoporosis is associated with decrease in bone formation and acceleration of bone resorption [8]. Thus, the aim of our study was to determine the relationships between the markers of bone remodeling and BMD in men with type 2 diabetes.

Matherials and methods

This cross-sectional observational study was included 59 men with type 2 diabetes, from 50 to 75 years of age (median 63 years). The duration of diabetes was at least 1 year (median 12 years). Patients with known risk factors for secondary osteoporosis and those taking medications that affect the bone remodeling, were not included. Control group comprised of 21 healthy subjects with normal BMD, matched by sex and age.

Dual-energy X-ray absorptiometry was applied for BMD assessment ("Lunar Prodigy" densitometer, *GE, USA)*. The BMD and lumbar spine T-score (L1-L4), proximal and femoral neck, and radius of non-dominant forearm were estimated.

The serum levels of parathyroid hormone (PTH), a bone turnover marker, osteocalcin, a marker of bone formation, osteoprotegerin and sclerostin, markers of bone metabolism, and free testosterone were assessed by ELISA. Urinary concentrations of C-terminal telopeptides of type I collagen (CTX-I), a bone resorption marker, was assessed by ELISA and adjusted to urinary creatinine levels.

Statistical data processing was performed using the STATISTICA software package (*StatSoft.Inc,* USA).

Results

Most of the patients were obese (n=31) or overweight (n=22). Hypoglycemic therapy included metformin (n=46), insulin (n=36), sulfonylurea (n=17), DPP-4 inhibitors (n=8), and SGLT2 inhibitors (n=8), mostly in combinations. The mean levels of HbA1c was 8.2 (7.2; 9.1)%. The estimated glomerular filtration rate (eGFR) was 71 (64; 86) ml/min/1.73 m^2; patients with eGFR less than 30 ml/min/1.73 m^2 were not included.

Based on the T-score, patients were divided into groups with normal BMD (n=30) and reduced BMD (n=29, including 25 subjects with osteopenia and 4 individuals with osteoporosis). Patients with reduction in BMD were older as compared to those without ($p=0.04$). There were no

differences between the groups in body mass index and eGFR.

The parameters of bone remodeling are summarized in Table 1. The levels of osteocalcin were decreased and the levels of osteoprotegerin and sclerostin were increased in observed diabetic subjects as compared to control (p=0.02, p<0.001 and p=0.02 respectively). The excretion of CTX-1 was reduced in patients with diabetes (p<0.001). No differences in PTH and free testosterone concentrations were revealed. There were no significant differences in bone remodeling markers between patients with normal and reduced BMD.

TABLE I. CONCENTRATIONS OF MARKERS OF BONE REMODELING IN MEN WITH TYPE 2 DIABETES

Parameter	Control group (n=21)	Normal BMD (n=30)	Decreased BMD (n=29)
Free testosterone, pg/ml	7.4 (5.6; 8.7)	7.1 (5.8; 9.9)	6.1 (4.8; 8.4)
PTH, pg/ml	56.3 (47.5; 90.2)	48.4 (37.9; 62)	49.8 (33.2; 63.6)
Osteocalcin, ng/ml	18.9 (16.2; 24)	11.2 (7.6; 14.7)*	10.9 (7.8; 13.3)*
Osteoprotegerin, pg/ml	15.9 (14.1; 24.5)	22.4 (20.2; 29.9)*	21.1 (18.6; 25.7)*
Sclerostin, pmol/l	23 (17.5; 27.1)	31.5 (21.4; 56.5)*	31.6 (22.1; 42.7)*
CTX-I, ng/mmol	465 (276.4; 639)	203 (116.8; 282.5)*	238.4 (149.1; 331)*

* significant difference with control

There was a positive correlation between free testosterone level and T-score of lumbar spine (r=0.28; p=0.03). Meanwhile, PTH demonstrated the correlations with T-score at femoral neck (r=-0.29, p=0.002) and FRAX (10-year probabilities of major osteoporotic fracture) score (r=-0.34, p=0.002). Osteoprotegerin and sclerostin correlated with age (r=0.32, p=0.01 and r=0.35, p=0.006, respectively) and eGFR (r=-0.39, p=0.002 and r=-0.45, p=0.0003, respectively). A negative relationship was found between osteocalcin and the hip circumference (r=-0.45, p=0.02). Besides, osteocalcin correlated positively with CTX-I (r=0.44, p=0.0005). The levels of PTH correlated positively with osteoproteregin levels (r=0.36, p=0.005). We found no significant correlations between HbA1c and studied markers of bone remodeling.

In stepwise multivariate regression analysis, sclerostin was the most significant predictor for lumbar spine T-score (β=0.496, R^2=0.23, p=0.00007), the level of PTH influenced the femoral neck T-score (β=-0.29, R^2=0.26, p=0.005).

Conclusion

The obtained results results suggest that the bone remodeling in men with type 2 diabetes is reduced due to the inhibition of osteoblastogenesis and decrease in the bone formation and resorption.

REFERENCES

[1] A. Palermo, L. D'Onofrio, R. Buzzetti, S.Manfrini, N. Napoli' "Pathophysiology of Bone Fragility in Patients with Diabetes", Calcif Tissue Int. 2017, Vol. 100, N. 2, pp. 122-132. doi: 10.1007/00223-016-0226-3.

[2] L. Oei, M. Zillikens "High Bone Mineral Density and Fracture Risk in Type 2 Diabetes as Skeletal Complications of Inadequate Glucose Control: The Rotterdam Study" Diabetes Care 2013, vol.36, N6, pp. 1619–1628. doi: 10.2337/dc12-1188.

[3] V.V. Klimontov, O.N. Fazullina. "The relationship of total body composition with bone mineral density in postmenopausal women with type 2 diabetes". Diabetes Mellitus. 2015. N1, pp. 65-69. doi: 1 0.14341/DM2015165-69

[4] O.A. Nikitinskaya, N.V. Toroptsova "Risk factors for osteoporosis and osteoporotic fractures in men aged 50 years and older", Osteoporosis and osteopathy. 2017, Vol. 20, N. 1, pp. 7-10. doi: 10.14341/osteo201717-11. 5.

[5] MR Rubin "Bone cells and bone turnover in diabetes mellitus", Curr Osteoporos Rep. 2015, Vol 13, N3, pp. 186-191. doi: 10.1007/s11914-015-0265-0.

[6] J. Starup-Linde et al. "Differences in biochemical bone markers by diabetes type and the impact of glucose". Bone. 2016, Vol. 83, pp. 149-155. doi: 10.1016/j.bone.2015.11.004.

[7] G.M. Nurullina, G.I. Akhmadullina. "Features of bone turnover in diabetes mellitus". Osteoporosis and bone diseases. 2017, Vol. 20, N3, pp. 82-89. doi: 10.14341/osteo2017382-89

[8] V.V. Klimontov, O.N. Fazullina, A.P. Likov, V.I. Konenkov, "The relationship between bone turnover markers and bone mineral density in postmenopausal type 2 diabetic women", Diabetes Mellitus. 2016. N5, pp. 375-382. doi: 10.14341 / DM8008

DOI 10.18699/BGRS/SB-2020-269

The Siberian multimodal brain tumor image segmentation dataset

Sergey Golushko
Novosibirsk State University
Novosibirsk, Russia
s.k.golushko@gmail.com

Mikhail Amelin
Radiodiagnosis division
ORCID: 0000-0002-5099-5144
FSBI "Federal Neurosurgical Center"
Novosibirsk, Russia
amelin81@gmail.com

Bair Tuchinov
SDA&ML lab.
Novosibirsk State University
Novosibirsk, Russia
bairts@gmail.com

Evgeniya Amelina
SDA&ML lab.
Novosibirsk State University
Novosibirsk, Russia
amelina.evgenia@gmail.com

Nikolay Tolstokulakov
SDA&ML lab.
Novosibirsk State University
Novosibirsk, Russia
n.tolstokulakov@g.nsu.ru

Evgeniy Pavlovskiy
SDA&ML lab.
Novosibirsk State University
Novosibirsk, Russia
pavlovskiy@post.nsu.ru

Vladimir Groza
Median Technologies
Valbonne, France
vladimir.groza@gmail.com

***Abstract* — Automatic brain tumor segmentation from CT or MRI scans is one of the crucial problems among other directions and domains where daily clinical workflow requires to put a lot of efforts while studying patients with various pathologies. In this paper, we report the results of the research project "Brain Tumor Segmentation" organized in conjunction with the Federal Neurosurgical Center. Several state-of-the-art tumor segmentation algorithms were applied to a set of 100 MRI scans of meningioma, neurinoma and glioma patients - manually annotated by up to three raters - and to 100 comparable scans obtained using the automated tumor multi-region segmentation. Quantitative evaluations revealed a considerable agreement between the human raters in segmenting various tumor sub-regions (Dice scores in the range 85-90%). We found that different algorithms worked best for different sub-regions, but no single algorithm ranked in the top for all subregions simultaneously.**

Keywords* — *Neural Network, Deep Learning, Segmentation, Medical Imaging

INTRODUCTION

Brain tumors are usually very heterogeneuous by its structure and edema is simillar to the healthy or non-enhanced tumor tissues. Within the recent achievements the problem of segmentation of the brain tumors can be effectively solved with the use of 2D and 3D deep learning (DL) based methods. However, none of structural magnetic resonance imaging (MRI) such as T1, T2, contrast-enhanced T1 (T1-CE), and T2 Fluid Attenuated Inversion Recovery (FLAIR) sequences or other permits to observe and identify all tumor composition elements by eyes or with the use of existing (semi-) automated techniques. Using the thorough analysis one could detect and classify diseased tissue and different pathological processes such as edema, necrosis, tumor etc. It is well known that being a crucial task for the effective diagnostics and differentiation between pathological processes the proper segmentation is also a source of recommendations for further treatment and follow-up.

METHODS

We used several approaches and data processing techniques. In the core of this framework we use the "LinkNet-like" networks with two different backbones from Se-Resnext50 and Se-Resnext101. Additionally we tried methods including following techniques: mixture of T1, T1-CE, and T2 FLAIR sequences in a pseudo-RGB image; combination of three T1C images as another pseudo-RGB image including the neighboring slices; wide input where all three sequences are used to create 9-channel input images and others. In order to compare obtained results we define the baseline method using only the T1C images, and also studying the difference between standard ResNet networks and its shift-invariant version with the blur levels.

RESULTS

Results presented in this work were obtained on the private dataset contained 100 full brain scans - manually annotated by up to three raters, 100 comparable scans obtained using the automated tumor multi-region segmentation. Annotations comprise labels of the peritumoral edema, the non-enhancing tumor, GD-enhancing tumor and the necrotic tumor core. We set-up the prototype the Siberian Brain Tumor Image Segmentation dataset (SBT). Quantitative evaluations revealed a considerable agreement between the human raters in segmenting various tumor sub- regions (Dice scores in the range 85-90%). We found that different algorithms worked best for different sub-regions, but that no single algorithm ranked in the top for all subregions simultaneously.

ACKNOWLEDGMENT

The reported study was funded by RFBR according to the research project No. 19-29-01103.

REFERENCES

[1] Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., et al. Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation. MICCAI Brainlesion Workshop. pp. 450–462. Springer, 2017.

[2] Myronenko, A. 3D MRI Brain Tumor Segmentation using Autoencoder Regularization. MICCAI Brainlesion Workshop. pp. 311-320. Springer, 2018.

[3] Letyagin, A. Y., Golushko, S. K. et al. Artificial Intel- ligence for Imaging Diagnostics in Neurosurgery. In 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON) (pp. 336-337). IEEE-INST. ELECTRICAL ELECTRON - ICS ENGINEERS INC.

DOI 10.18699/BGRS/SB-2020-270

Gut microbiome and inflammatory bowel disease (IBD) – is there a link?

Tatiana Grigoryeva
KFU, Kazan, Russia
tata.grigoryeva@gmail.com

Giuseppe Lo Sasso
PMI R&D, Neuchatel, Switzerland
Giuseppe.Losasso@pmi.com

Julia Hoeng
PMI R&D, Neuchatel, Switzerland
Julia.Hoeng@pmi.com

Nikolai Ivanov
PMI R&D, Neuchatel, Switzerland
nikolai.ivanov@pmi.com

Manuel Peitsch
PMI R&D, Neuchatel, Switzerland
manuel.peitsch@pmi.com

Natalya Danilova
KFU, Kazan, Russia
danilova.natalya.87@mail.ru

Rustem Abdulkhakov
KSMU, Kazan, Russia
rustemabdul@mail.ru

Alfiya Odintsova
RCH, Kazan, Russia
odincovaa@yandex.com

Maria Markelova
KFU, Kazan, Russia
mimarkelova@gmail.com

Mariia Siniagina
KFU, Kazan, Russia
marias25@mail.ru

Dilyara Khusnutdinova
KFU, Kazan, Russia
dilyahusn@gmail.com

Ilya Vasiliev
KFU, Kazan, Russia
mepk_m6@mail.ru

Sayar Abdulkhakov
KFU, Kazan, Russia
sayarabdul@yandex.ru

Abstract — **The pathogenesis of IBD is not fully known, but one of the factors predisposing to the development of inflammation is probably an imbalance between commensal bacteria and pathogens in the intestinal lumen, a decrease in microbial diversity, and impaired functional metabolism of bacteria. Fecal microbiota whole-genome sequencing confirmed the presence of specific IBD dysbiotic features at the phylum level, with increased abundance of Proteobacteria, Actinobacteria, and Fusobacteria and decrease in Firmicutes, Bacteroidetes, and Verrucomicrobia. At the genus level, the abundance of both fermentative (SCFA-producing and H2-releasing) and hydrogenotrophic (H2-consuming) microbes was affected in IBD patients. This imbalance was confirmed by the decreased abundance of SCFA species in the feces of IBD subjects as well as the change in anaerobic index, which mirrors the redox status of the intestine. Conclusion: Our analyses highlighted how IBD-related dysbiotic microbiota— which are generally mainly linked to SCFA imbalance—might affect other important metabolic pathways, such as H2 metabolism, that are critical for host physiology and disease development.**

Keywords – intestinal microbiota, shotgun sequencing, 16SrRNA metagenome, inflammatory bowel diseases

Motivation and aim

Inflammatory bowel diseases (IBD) are a group of disorders characterized by chronic inflammation of the gastrointestinal tract. IBD present with two main manifestations: ulcerative colitis and Crohn's disease, each having distinctive clinical and pathological features. The etiology of IBD is not fully understood, and the term is used to describe a collection of chronic immune-mediated diseases of unknown, multifactorial etiology with complex interactions between genetic and environmental factors.

Aim of the study revealed differences in microbiome composition among the ulcerative colitis, Crohn's disease, and healthy control subjects.

Methods

A total of 126 subjects (42 healthy volunteers and 41 Crohn's disease and 43 ulcerative colitis patients) were studied. IBD was diagnosed on the basis of standard clinical, endoscopic, and histological criteria. Fecal microbiota composition was analyzed by 16S and whole-genome shotgun sequencing.

Results

Changes in the intestinal microbiota in patients with UC and CD are characterized by an increase in the presence of bacteria of the fila Proteobacteria and Bacteroidetes against the background of a decrease in the relative number of bacteria fila Firmicutes and Archaea Euryarchaeota; a decrease in the index of alpha-diversity of bacteria, a decrease in the relative representation of butyrate-producing, hydrogen-utilizing bacteria, Methanobrevibacter smithii, which may be a prerequisite for a violation of colonization resistance. An increase in the relative representation of Ruminococcus gnavus and Akkermansia muciniphila may be an additional, characteristic of patients with UC and CD, a sign of dysbiosis. A decrease in the representation of the Butyryl-CoA: acetate CoA transferase gene in patients with CD, a decrease in the absolute content of both individual fatty acids and their total amount, and a change in the ratio of the main fatty acids in patients with IBD may indicate a decrease in the functional activity and amount of anaerobic microflora and / or change in utilization of fatty acids by colonocytes. Metabolic dysbiosis associated with an increase in the relative representation of pathways in patients with IBD. The revealed changes can be considered as characteristic shifts in gut microbiome under IBD and potential targets in the selection of personalized therapy.

DOI 10.18699/BGRS/SB-2020-271

GenCoNet: medical information system to support the treatment of co-morbid diseases

Ralf Hofestädt
University Bielefeld
AG Bioinformatics
Bielefeld, Germany

Elena Bragina
Tomsk NRMC
Institute of Medical Genetics
Tomsk, Russia

Victor Dosenko
National Academy of Sciences
Bogomoletz Institute of Physiolog
Kiev, Ukraine

***Abstract* — During the realization of the "Trilateral Partnership – Cooperation Project of the VolkswagenStiftung" between Scholars and Scientists from Ukraine (Kiev), Russia (Tomsk and Novosibirsk) and Germany (Bielefeld) (2016-2019) we have obtained important data and results concerning pathogenetic factors of such important comorbidity as combination of hypertension and asthma. Based on this project the medical information system GenCoNet was developed. During the last months an extended version of GenCoNet could be implemented.**

Keywords — medical information system, co-morbid diseases, GenCoNet

INTRODUCTION

Hypertension and bronchial asthma are a major issue for people`s health. As of 2014, approximately one billion adults, or ~22% of the world population, have had hypertension. As of 2011, 235 - 330 million people globally have been affected by asthma and approximately 250,000 - 345,000 people have died each year from the disease. The development of the effective treatment therapies against these diseases is complicated by their comorbidity features.

Based on the medical data and knowledge of the project partner from Tomsk National Research Medical Center we could start to identify relevant genes and drugs for asthma and hypertension. Based on lists of genes associated with asthma and hypertension obtained using the HuGENavigator resource and patient drug lists, Bielefeld and Novosibirsk computed and analyzed first relevant metabolic networks. Finally, the project partner form Novosibirsk identified important genes that are involved in comorbid links between asthma and hypertension by using text- and data- mining techniques. The experimental analysis of the detected gene candidates showed that *IL10* silencing leads to decrease of severity of asthma and to positive changes in cardiohaemodynamic parameters. Furthermore, based on the clinical data and semi-automatic data mining approaches a new database was developed and implemented, which presents the positive and negative drug list for asthma and hypertension. A web based implementation of this data base allows the access to this information via internet (https://genconet.kalis-amts.de). A total of 126 subjects (42 healthy volunteers and 41 Crohn's disease and 43 ulcerative colitis patients) were studied. IBD was diagnosed on the basis of standard clinical, endoscopic, and histological criteria. Fecal microbiota composition was analyzed by 16S and whole-genome shotgun sequencing.

WORKFLOW

The investigation of multifactorial diseases requires the consideration of the most important biomedical entities: diseases, genes, variants, and drugs. The data on these entities is spread across several data sources. These data sources need to be researched, identified, and integrated into a suitable database format. For this purpose, a workflow system was developed to extract, transform, and load reliable data on these entities into a uniform graph database eGenCoNet for network analysis of gene-disease associations. The workflow system is divided into four steps as illustrated in Figure 1.

First, the comorbidities of high interest in molecular medicine were determined. Essential hypertension and bronchial asthma are considered as an example of common comorbid diseases. Disease Ontology terms associated with genes were obtained from Osborne et al. Second, human genes that are associated with diseases and variants were obtained from Human Phenotype Ontology (HPO), MalaCards, DisGeNet, and OMIM. In particular, human genes were extracted from these databases that may cause familial syndromes (Mendelian forms). HPO and DisGeNet provide several database subsets from which the "FREQUENT_FEATURES" marked data for HPO and "curated" marked data for DisGeNet were used. MalaCards does not provide a download of the database, so information was manually extracted and integrated from the website. From the OMIM gene map, information was extracted with the phenotype "Asthma, susceptibility to", "Hypertension, essential, susceptibility to" and adjacent annotation. In addition, altered expression data of genes associated with high blood pressure or severe asthma were manually curated and integrated.Third, genes that are controlled by eQTL in blood, codes variants, and gene associations supported by at least two studies were obtained from GWAS Catalog, dbSNP, and DisGeNet. These genetic variants were in turn associated with the comorbid diseases. Fourth, drugs and their target genes in humans were extracted from DrugBank using the full database XML version 5.1.0 including meta information like knownactions. Additional drugs, that are indicated, contraindicated or induced in asthma and hypertension were extracted from the DrugCentral 2017-08-29 database dump.

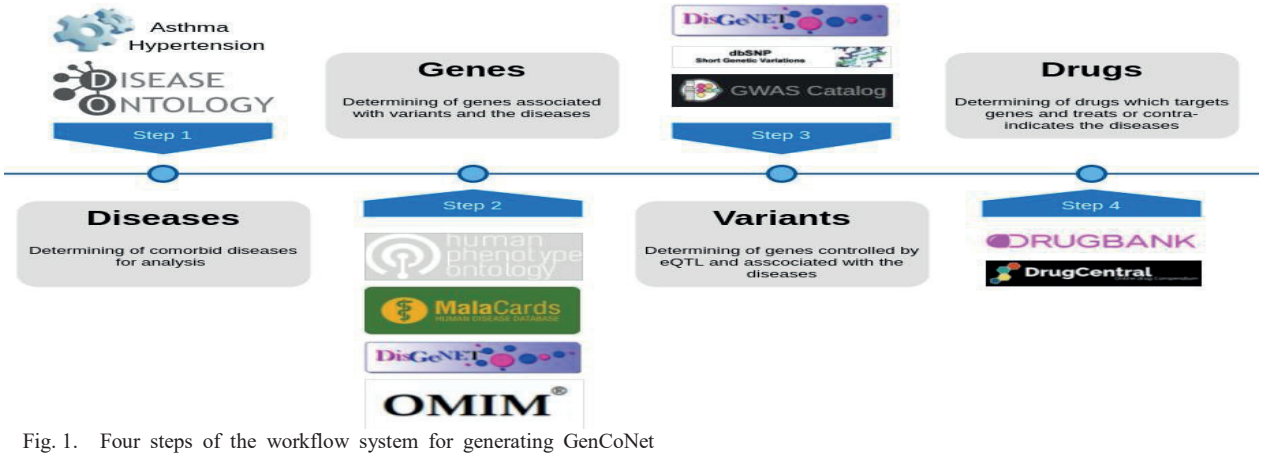


Fig. 1. Four steps of the workflow system for generating GenCoNet semi-automatically.

IMPLEMENTATION

The process of integrating all data sources and merging of the entities of the aforementioned workflow system is an iterative process. Therefore, a semi-automatic pipeline was implemented for the import, fusion, and analysis of data in a highly connected database structure. This pipeline provides custom and very fast import Python scripts and Cypher3 queries for generating a new Neo4j database from large data. TheNeo4j instance is running in a Docker4 container to simplify the setup process. Each step of the workflow system is executed in a separate Python script in order to have a better overview and to be able to execute specific stepson demand. Some of the data sources need to be preprocessed due to the higher information density or more complicated file formats than others. Afterwards, the data sources are processed and the basic connections are formed. Following nodes representing the same kind of entities are merged into fusion nodes. Finally, the results were imported into the new GenCoNet database

APPICATION

The prevalence of comorbidity is increasing and leads to a corresponding polypharmacy, which in turn is the prime risk factor for drug-related problems. In particular, drug contraindications with any disease and drug-induced diseases have to be considered in the context of treatment by healthcare professionals. Using GenCoNet, these risks may be detected (https://genconet.kalis-amts.de). The web interface represents the database entities (diseases, drugs, and proteins) as nodes in a network with edges representing the relationships between them. The vizualistion of networks show drugs (violet) which are prescribed for the treatment of the comorbid diseases (green) but also target genes (red) which maycause the diseases. For instance, the anti-asthmatic drug “Ibudilast” may induce hypertension by targeting thegene PDE3A. As a consequence, these drugs have to be avoided in order to reduce the risk of induced diseases.

DISCUSSION

The discovery of shared molecular players and mechanisms in the pathogenesis of comorbid diseases is still complicated and nevertheless, necessary for decision-making of the most appropriate treatment strategy. To address this ongoing need, the Neo4j database GenCoNet was developed which integrates various associations between diseases, genes, variants and drugs for the diseases bronchial asthma and essential hypertension. While information of highest quality is preferred for integration in GenCoNet, false positives, although rare, cannot be ruled out. However, the use cases and lists already emphasized the potential of applicability in daily medical practice. GenCoNet is meant to be a qualitative resource that facilitates researchers access to the relevant information for the net-work analysis of comorbidities. Therefore, GenCoNet is planned to be extended by data on further diseases.

REFERENCES

[1] A. Shoshi, R. Hofestädt, O. Zolotareva, M. Friedrichs, A. Maier, V. A. Ivanisenko, V. E. Dosenko, E. Yu. Bragina. GenCoNet - A Graph Database for the Analysis of Comorbidities by Gene Networks. Journal of Integrative Bioinformatics 15(4), 2018. DOI: 10.1515/jib-2018-0049.

DOI 10.18699/BGRS/SB-2020-272

Reconstruction and analysis of regulatory gene networks involving human genes associated with main forms of pathozoospermia

Elena V. Ignatieva
The Federal Research Center Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
eignat@bionet.nsc.ru

Alexander V. Osadchuk
The Federal Research Center Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
osadchuk@bionet.nsc.ru;

Maxim A Kleshev
The Federal Research Center Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
max82cll@bionet.nsc.ru

Ludmila V. Osadchuk
The Federal Research Center Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
losadch@bionet.nsc.ru

Abstract — **The study of the molecular-genetic mechanisms predisposing to decline in male reproductive potential (spermatogenic failure) is an actual problem of reproductive biology. Most often in laboratory studies evaluating male fertility, a study of the quality of ejaculate is used. Thus, a pathological condition called pathozoospermia can be detected if the quality indicators of the ejaculate are decreased. Pathozoospermia can manifest itself in several distinct forms, may occur in many diseases and can be caused by many factors, including genetic ones. To reveal regulatory interactions between genes associated with pathozoospermia, we reconstructed gene regulatory network involving genes harboring allelic variants associated with pathozoospermia. Regulatory network comprised seven genes encoding transcription factors (TFs) for which a set of target genes were predicted by MoLoTool web-service. We identified three key regulatory transcription factors (WT1, AHR, NR0B1) that have the greatest number of target genes in the network. Genes encoding these factors can be considered as the most promising candidate genes for identifying genetic variants associated with pathozoospermia.**

Keywords *— male infertility, spermatogenesis, pathozoospermia, genetic association, gene network, transcription factor, target gene*

INTRODUCTION

The decline in male reproductive potential, identified in many countries of the world, is an important scientific problem and serious social issue [1,2]. Most often in laboratory studies evaluating male fertility, a study of the quality of ejaculate is used. Thus, a pathological condition called pathozoospermia may be revealed if the quality indicators of the ejaculate (defined as the reference values of WHO [3]) are decreased. The most common forms of pathozoospermia are azoospermia, asthenozoospermia, oligozoospermia, teratozoospermia. Pathozoospermia is a syndrome that can manifest itself in several distinct forms, may occur in many diseases and can be caused by many factors, including genetic ones. To reveal regulatory interactions between genes associated with pathozoospermia, we reconstructed gene regulatory network and searched for its key regulators.

MATERIALS AND METHODS

The list of 68 human genes associated with specific forms of pathozoospermia was formed as it was described in [4]. To identify genes that encode transcription factors, we used AnimalTFDB [5]. DNA sequences of promoter regions (500 bp and 1000 bp) were extracted from the GRCh38.p12 assembly for all genes using BioMart tool. Transcription factor binding sites (TFBSs) were predicted using MoLoTool web-service (https://molotool.autosome.ru/). MoLoTool enables to scan DNA sequences for TFBSs with position weight matrices presented in HOCOMOCO model collection [6]. We used significance threshold pajusted < 10-4, which was recommended by the web-service. Thus, the information about the hypothesized transcriptional regulation of genes associated with pathozoospermia was obtained. Networks were visualized using Cytoscape [7]

RESULTS

Using AnimalTFDB we identified that 10 out of 68 genes associated with pathoszoospermia encoded transcription factors: WT1, AHR, NR0B1, ETV5, AR, NFE2L2, SOX5, AHRR, YBX2, SOX6. Three of them (AHRR, YBX2, SOX6) had no motif information in the HOCOMOCO model collection [6]. Then we scanned promoter regions (500 bp and 1000 bp) of all 68 genes with MoLoTool and identified a set of TFBSs for each promoter. Next we reconstructed two transcriptional regulatory networks built on data on binding sites for TFs in upstream (500 bp and 1000 bp) gene regions. These networks contained 60 nodes, 123 edges (when 500 bp regions were analyzed) and 65 nodes, 178 edges (when 1000 bp regions were analyzed). Numbers of TFBSs of each type and numbers of target genes for each TF were calculated for each network (Table 1). We found that the greatest number (46 for 500 bp analyzed, and 58 for 1000bp analyzed) of target genes had WT1. The other two top transcription factors (AHR and NR0B1) had less regulated genes (for AHR there were 29 and 38 target genes, and for NR0B1 there were 25 and 33 genes). Our findings are in good agreement with results obtained in mice, indicating that WT1 is critical for spermatogenesis via regulation of sertoli cell polarity via Wnt signaling pathway [8]. Furthermore, we revealed three

regulatory self-loops involving transcription factors WT1, NR0B1, and AR.

TABLE 1. Ranking transcription factors according to the number of target genes in the networks

Transcrip-tion factor	Network based on analysis of			
	500 bp promoter regions		1000 bp promoter regions	
	Number of binding sites	*Number of target genes*	*Number of binding sites*	*Number of target genes*
WT1	344	46	660	58
AHR	46	29	76	38
NR0B1	41	25	59	33
ETV5	19	11	60	24
AR	9	7	19	14
NFE2L2	2	2	6	6
SOX5	3	3	5	5
Total number	**123**	**60**	**178**	**65**

CONCLUSION

Using MoLoTool web-service that scans DNA sequences for transcription factor binding sites we revealed potential target genes for seven TFs and reconstructed transcriptional regulatory networks involving 60 and 65 human genes associated with pathozoospermia. We identified three key regulatory transcription factors that had the greatest number of target genes: WT1, AHR, and NR0B1. We expect that further experiments to test binding ability of the predicted binding sites will verify the interactions between these three TFs and their target genes. We propose to keep in mind genes encoding these transcription factors as most promising candidates for investigating the genetic factors predisposing to pathozoospermia.

ACKNOWLEDGMENT

This study was supported by the grant from the Russian Science Foundation (project No. 19-15-00075).

REFERENCES

[1] C. Krausz, F. Cioppi, A. Riera-Escamilla, “Testing for genetic contributions to infertility: potential clinical impact”, Expert Rev Mol Diagn, Vol. 18(4), pp. 331-346, 2018.

[2] C. Krausz, A. Riera-Escamilla, “Genetics of male infertility”, Nat Rev Urol, Vol. 15(6), pp. 369-384, 2018.

[3] WHO, “WHO laboratory manual for the examination and processing of human semen”, 5-th edition, Cambridge Univer. Press. 2010. https://apps.who.int/iris/bitstream/handle/10665/44261/9789241547789_eng.pdf

[4] E.V. Ignatieva, A.V. Osadchuk, M.A. Kleshev, L.V. Osadchuk, “Towards a Comprehensive Catalog of Human Genes Associated with Main Forms of Pathozoospermia and its Functional Annotation” Systems Biology and Biomedicine (SBioMed-2020), 2020

[5] H.M. Zhang et al., “AnimalTFDB 2 0:A resource for expression, prediction and functional study of animal transcription factors”, Nucleic Acids Res, Vol. 43, pp. 76–81, 2015.

[6] I.V. Kulakovskiy et al., “HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via largescale ChIP-Seq analysis”, Nucleic Acids Res, Vol. 46, pp. D252–D259, 2017.

[7] P. Shannon et al., “Cytoscape: a software environment for integrated models of biomolecular interaction networks”, Genome Res, Vol. 13(11), pp. 2498-504, 2003.

[8] X.N. Wang et al., “The Wilms tumor gene, Wt1, is critical for mouse spermatogenesis via regulation of sertoli cell polarity and is associated with non-obstructive azoospermia in humans”, PLoS Genet, Vol.9(8), p.e1003645, 2013.

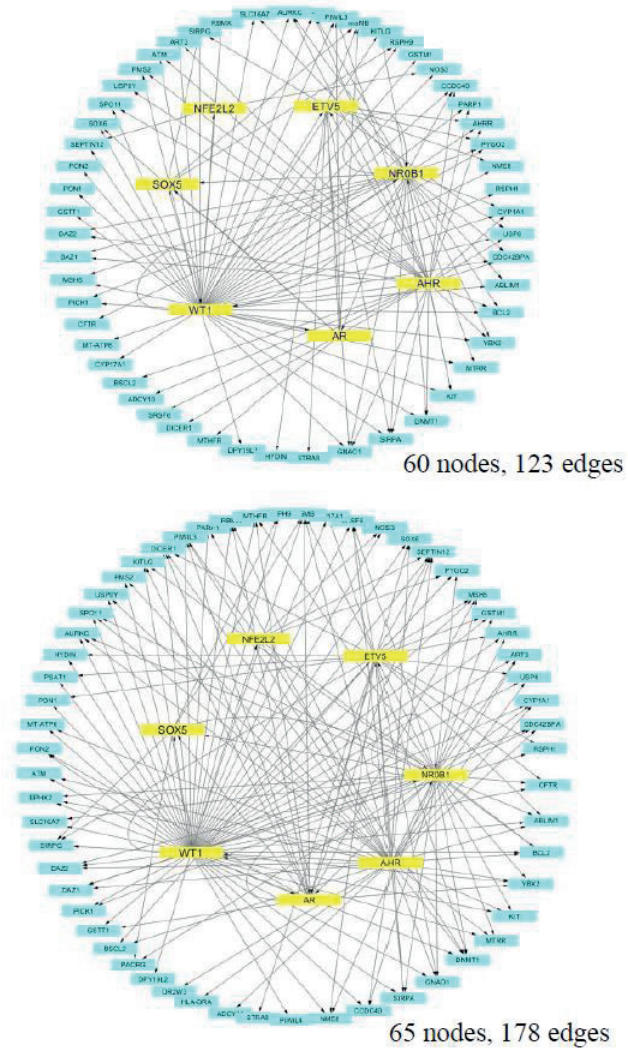


Fig. 1. Gene regulatory networks formed by regulatory interactions between transcription factors and target genes, associated with pathozoospermia. Networks were constructed based on the analysis of 500 bp (top) and 1000 bp (bottom) promoter regions. Yellow nodes correspond to both genes encoding transcription factors and transcription factors themselfs, blue nodes denote target genes

DOI 10.18699/BGRS/SB-2020-273

Towards a comprehensive catalog of human genes associated with main forms of pathozoospermia and its functional annotation

Elena V. Ignatieva
Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
eignat@bionet.nsc.ru

Alexander V. Osadchuk
Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
osadchuk@bionet.nsc.ru;

Maxim A. Kleshev
Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
max82cll@bionet.nsc.ru

Ludmila V. Osadchuk
Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
losadch@bionet.nsc.ru

***Abstract* — The genetic causes of the global decline in male fertility is among the hot spots of scientific research in reproductive genetics. The most common way to evaluate male fertility in clinical trials is to determine the quality of the ejaculate. A pathological condition characterized by lower quality indicators of the ejaculate compared to the norm and leading to impaired fertility is called pathozoospermia. Pathozoospermia is a syndrome that occurs in many diseases and can be caused by many factors, including genetic ones. To systematize information about genes associated with main forms of pathozoospermia, we created a catalog of such genes and analyzed their functional characteristics. On the basis of a manual analysis of scientific publications we collected data on 70 genes harboring more than one hundred allelic variants associated with pathozoospermia. Gene ontology enrichment analysis revealed that genes were significantly associated with spermatogenesis and male gamete generation. In addition it was found that genes from the catalog were associated with more than seventy human diseases that were both reproductive system pathologies and others, such as Parkinson's and Alzheimer's diseases as well as various cancer types.**

Keywords* — *male infertility, spermatogenesis, pathozoospermia, genetic polymorphism, associations, database, functional annotation

Introduction

The decline in male reproductive potential, identified in many countries of the world, is an important scientific problem and serious social issue [4, 5]. The question of the genetic causes of the global decline in male fertility is among the hot spots of scientific research in reproductive genetics. It is believed that in 20-25% of cases of severe reproductive dysfunction in men can be due to genetic causes [4]. Such a high frequency of genetic disorders of male reproductive function indicates its multigenic control. The most common way to evaluate male fertility in clinical trials is to determine the quality of the ejaculate. For this purpose, a number of quantitative indicators are used, defined as the reference values of WHO [6]. A pathological condition characterized by lower quality indicators of the ejaculate compared to the norm and leading to impaired fertility is called pathozoospermia. The most common forms of pathozoospermia are azoospermia, oligozoospermia, asthenozoospermia teratozoospermia. Pathozoospermia is a syndrome that occurs in many diseases and can be caused by many factors, including genetic ones. Despite the urgency of the problem of male infertility, there is currently no specialized information resource accumulating information about the genes associated with specific forms of pathozoospermia. These data are dispersed in a large number of scientific publications, and such database on associations between genes and diseases as ClinVar, OMIM and DiGeNet, contain data by no means on all forms of pathozoospermia. To systematize information about genes associated with main forms of pathozoospermia, we created a catalog of such genes and analyzed their functional characteristics.

Materials and Methods

Data on associations between allelic variants of genes and specific forms of pathozoospermia were extracted from scientific publications that were selected using ANDSystem [3]. Search terms that refer to specific forms of pathozoospermia were obtained from [6]. Gene ontology (GO) and human disease ontology enrichment analysis were performed using DAVID tool [2]. The enriched GO terms (FDR<0.05) from the biological processes vocabulary GOTERM_BP_5 and enriched diseases (FDR<0.05) from the genetic association database [1] were considered in our study.

Results

First we formed a following dictionary of keywords denoting specific forms of pathozoospermia: (1) azoospermia; (2) non-obstructive azoospermia; (3) cryptozoospermia; (4) oligozoospermia; (5) severe oligozoospermia; (6) asthenozoospermia; (7) teratozoospermia; (8) oligoasthenoteratozoospermia (OAT); (9) oligoasthenozoospermia; (10) oligoteratozoospermia; (11) globozoospermia. Using these key words as a search terms we collected data on 70 genes harboring more than one hundred allelic variants associated with pathozoospermia. The largest portions of genes were associated with oligozoospermia and non-obstructive azoospermia (Fig.1A). Most genes (63 out of 70) were located on autosomes, X and Y chromosomes had three genes each, and one gene was located on mitochondrial DNA (Fig. 1B). encoded transcription factors: WT1, AHR,

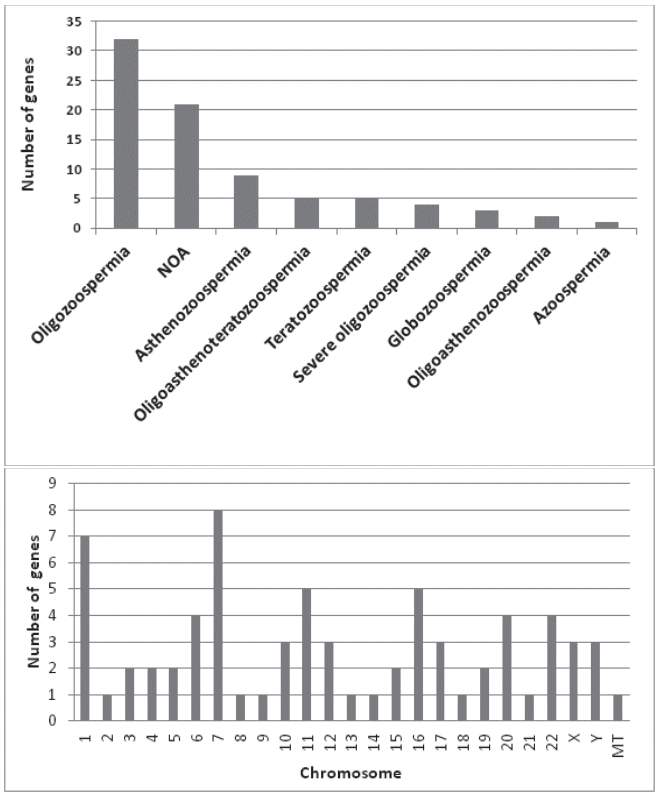


Fig. 1. Information content of the catalog. A - the number of genes associated with a specific form of pathozoospermia; B – the number of genes from the catalog identified on each chromosome. NOA - non-obstructive azoospermia; MT - non-nuclear(mitochondrial)

Enrichment analysis performed using the DAVID-tool revealed 17 significantly enriched (FDR<0.05) biological processes from the GOTERM_BP_5 vocabulary. The most representative gene groups (more than seven genes each) were annotated by enriched terms associated with spermatogenesis and male gamete generation (Table 1). In addition we identified that genes from the catalog were associated with more than seventy human diseases from [1] (data not shown). Among these diseases were both reproductive system pathologies and others, such as Parkinson's and Alzheimer's diseases as well as various cancer types.

Conclusion

We have created a catalog of genes and their allelic variants associated with various forms of pathozoospermia. This information was obtained on the basis of a manual analysis of scientific publications, which indicates a high level of reliability. The data we have accumulated and the results of functional gene annotation can be useful both for interpreting the results of genome- and exome-wide association studies, as well as for developing new approaches to the diagnosis and treatment of male infertility.

ACKNOWLEDGMENT

This study was supported by the grant from the Russian Science Foundation (project No. 19-15-00075).

REFERENCES

[1] K.G. Becker, K.C. Barnes, T.J. Bright, S.A. Wang, “The genetic association database”, Nat Genet, Vol. 36(5), pp. 431-432, May 2004.

[2] D.W. Huang, B.T. Sherman, Q. Tan, J.R. Collins, W.G. Alvord, J. Roayaei, R. et al., “The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists”, Genome Biol, Vol. 8(9), p. R183, 2007.

[3] V.A. Ivanisenko, P.S. Demenkov, T.V. Ivanisenko, E.L. Mishchenko, O.V. Saik, “A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression”, BMC Bioinformatics, Vol. 20(Suppl 1), p. 34, 2019.

[4] C. Krausz, F. Cioppi, A. Riera-Escamilla, “Testing for genetic contributions to infertility: potential clinical impact”, Expert Rev Mol Diagn, Vol. 18(4), pp. 331-346, 2018.

[5] C. Krausz, A. Riera-Escamilla, “Genetics of male infertility”, Nat Rev Urol, Vol. 15(6), pp. 369-384, 2018.

[6] WHO, “WHO laboratory manual for the examination and processing of human semen”, 5-th edition, Cambridge Univer. Press. 2010. https://apps.who.int/iris/bitstream/handle/10665/44261/9789241547789_eng.pdf

TABLE 1. OVERREPRESENTED GO TERMS THAT CHARACTERIZE GENES FROM THE CATALOG

Term[a]	Number of genes	Fold Enrichment	FDR
Gamete generation	25	10.5	1.5E-15
Cell development	22	3.0	6.8E-3
Spermatogenesis	21	11.0	1.1E-12
Male gamete generation	21	11.0	1.1E-12
Germ cell development	15	17.2	1.6E-10
Gonad development	9	11.2	1.9E-3
Development of primary sexual characteristics	9	10.9	2.3E-3
Spermatid development	8	16.3	8.4E-04
Spermatid differentiation	8	15.7	1.1E-3

[a] Terms associated with more than 7 genes are presented

DOI 10.18699/BGRS/SB-2020-274

LYVE-1 expression in liver cells of mice with functional pinealectomy

Irina Yurievna Ishchenko
Group of experimental pharmacology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
irenisch@mail.ru

Svetlana Viktorovna Michurina
Group of experimental pharmacology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
s.michurina@ngs.ru

Sergey Alekseevich Arkhipov
Group of experimental pharmacology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
arhipowsergei@yandex.ru

Andrey Yurievich Letyagin
Laboratory of pharmaceutical technologies
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
letyagin-andrey@yandex.ru

Maxim Aleksandrovich Korolev
Laboratory of connective tissue pathology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
kormax@bk.ru

Evgenii Leonidovich Zavjalov
Center of genetis resources of laboratory animals
ICG SB RAS
Novosibirsk, Russia
zavjalov@bionet.nsc.ru

***Abstract* — The expression of lymphatic vessel endothelial hyaluronan receptor-1 (LYVE-1) in liver cells of C57BL/6 mice males kept 14 days under 24-hour lighting (mice with functional pinealectomy) was studied. The immunohistochemical analysis (indirect avidin-biotin ABC-peroxidase method) and morphometric assessment were used. A 2-fold decrease in the area of LIVE-1 expression was found against the background of a 2% increase in the optical density of LIVE-1 staining in mouse liver cells after chronic continuous illumination. The weak LYVE-1 expression on the membranes of endothelial cells of liver sinusoids may indicate a violation of the functioning of the fenestration fields of these cells. This can lead to a decrease in the endocytotic activity of the latter, difficulty in hemato-tissue exchange, lymphatic drainage deterioration and tissue hypoxia development in the liver of mice with functional pinealectomy.**

Key words* — *continuous illumination, functional pinealectomy, liver, LIVE-1

Introduction

An integral factor of modern life is the use of continuous artificial lighting and constant changes in light mode. As a result of the long existence of a person under such conditions, the production of the pineal gland hormone melatonin significantly decreases/stops in his body (functional pinealectomy), there is an imbalance of cyclicity in the work of body systems, desynchronosis develops.

It is known that the liver is not only a unique metabolic and immunological organ, it is the largest source of lymph production in the body, and it accounts for up to 50% of the lymph entering the thoracic duct [1]. The lymphatic system of any organ helps maintain tissue fluid balance and remove cellular waste products. It was found that with liver diseases its lymphatic system changes significantly. An increase in hepatic lymphatic vessels with cirrhosis, viral hepatitis and hepatocellular carcinoma has been reported [2]. It was found that under the influence of constant lighting in the liver and its lymphatic region, circulatory and lymphatic flow disorders develop. These processes lead to the development of hypoxia, which negatively affects the structure and function of the organ parenchyma cells [3]. The aim of this study was to evaluate the effect of 24-hour illumination (24hL) on the expression of the LIVE-1 marker (lymphatic vessel endothelial receptor-1) in liver cells of male C57Bl / 6 mice.

Materials and Methods

The males of C57Bl/6 mice aged of 10-12 weeks were kept in the SPF-vivarium center of ICG SB RAS. Food and water were provided to animals *at libitum*. The intact animals (the group "Control", n=6) were kept under the standard lighting mode - photoperiod light/dark: 14/10 hour (h). The mice of the experimental group (the gr. "24hL", n=6) were kept for 14 days under conditions of 24-hour lighting - photoperiod light/dark: 24/0 h. Animals were removed from the experiment by the cranio-cervical dislocation method and liver samples were taken for light-optical and immunohistochemical studies.Liver samples were fixed in 10% buffered formalin for 48 hours, dehydrated in a series of alcohols of increasing concentrations and embedded in Histomix. A light-optical study of the liver was performed on sections stained with Mayer hematoxylin and eosin.

Immunohistochemical detection of the LYVE-1 marker was performed in sections with a 3 μm thickness using the indirect ABC-peroxidase method using the VECTASTAIN Universal ABC-Peroxidase Kit (PK-7200, Vector Laboratories, USA). Endogenous peroxidase blocking was performed by incubation of sections in 0.3% H_2O_2 solution for 10 min, then sections were incubated with normal blocking non-immune horse serum for 20 min. Next, the sections were incubated for an hour at room temperature with antibodies to LYVE-1 (Isotype: Rabbit polyclonal, bs-1311R; Bioss) using the final dilution of 5 μg/ml, then they were washed in 3 shifts of phosphate buffer (FB) for 3 min and were incubated for 30 minutes at room temperature with the second biotinylated antibodies followed by washing in 3 shifts of FB for 5 minutes. Incubation with the ABC-peroxidase complex was carried out at room temperature for 30 min, the sections were washed in 3 shifts of FB for 5 minutes. At the last stage immunohistochemical staining of the sections was performed in the chromogenic substrate (ImmPACT DAB, Vector Laboratories, USA). For quantification of LIVE-1 expression in the mouse liver, there was performed computer morphometric analysis of digital photographs at ×400 magnification. The relative areas and optical densities of

zones staining for LIVE-1 in digital images were determined using the program "Image J". A comparison of the results between the groups "24hL" and "Control" was performed using the Mann-Whitney U-test for abnormally distributed characters in independent samples. Differences of compared values were considered statistically significant at $P<0.05$.

Results and Discussion

In the liver of mice after two weeks of round-the-clock lighting (24/0 h), dilatation of the spaces of the Malla and lymphatic vessels, expansion of the bile ducts in the portal tracts were detected. These changes were accompanied by edema of the interlobular connective tissue layers. There was a stasis of red blood cells in the interlobular, central and sublobular veins. The intralobular sinusoid capillaries were also expanded. Lymphoid nodules were found in the parenchyma.

The study of the expression of the LIVE-1 marker in the liver of intact mice under standard lighting mode revealed a pronounced immunohistochemical staining of sinusoid cells of blood capillaries in the periportal regions and especially in the perineal zones of the hepatic lobes. In the parenchyma, individually colored hepatocytes were sometimes found. In animals kept under round- the-clock lighting , a weaker expression of LIVE-1 was detected in the endothelium of sinusoid capillaries, especially in the intermediate zones . Morphometric analysis of mouse liver preparations after two weeks of continuous illumination revealed a two-fold decrease in the area of expression of the LIVE-1 marker against the background of an increase in its optical density of LIVE-1 staining by 2%.

LYVE-1 is known to be a transmembrane protein receptor in endothelial cells of lymphatic vessels, for hyaluronic acid(hyaluronan), an extracellular matrix mucopolysaccharide [4]. Hyaluronan is the most powerful water-binding molecule in our body; it can absorb up to 1000 times its mass in water [5]. The pronounced expression of LIVE-1 in the endothelium of liver sinusoids in intact mice is confirmed by the results of other researchers. It was found that LYVE-1 is expressed in the liver not only by endothelial cells of lymphatic vessels but also by endothelial cells of blood sinusoid capillaries, as well as in some macrophages. As in our study, the primary staining on LIVE-1 of portal and especially middle zones of liver acini in intact animals was previously revealed. It is known that parenchyma cells in the intermediate zones are most sensitive to oxidative stress [6].

In liver diseases, the expression of LIVE-1 may be reduced. Thus, in cirrhosis and hepatocellular carcinoma, LYVE-1 expression is suppressed [6]. In chronic hepatitis and cirrhosis of the liver, a decrease in the expression of LYVE-1 in the endothelium of sinusoids was detected, especially in areas adjacent to active inflammatory or fibrotic lesions. In this case, the sinusoidal endothelium in the affected areas, demonstrating the weakening of LYVE-1, loses fenestration, which the authors attribute to the progression of liver disease [7]. Disorders in the lymphatic vasculature that cause inadequate drainage can lead to a condition characterized by tissue swelling, which is the result of the accumulation of protein and fluid in the interstitium [8].

In this regard, the weak expression of LYVE-1 detected by us on the membranes of endothelial cells of sinusoidal capillaries in mice kept for a long time under continuous illumination may indicate a violation of the functioning of the fields of fenestration of these cells. This can lead to a decrease in the endocytotic activity of the latter, difficulty in blood-tissue exchange, deterioration of lymphatic drainage in the liver, and development of tissue hypoxia.

ACKNOWLEDGMENT

The study is supported by the budget project (No. AAAA-A17-17072710029-7) and implemented using the equipment of the Center for Genetic Resources of Laboratory Animals at ICG SB RAS, supported by the Ministry of Education and Science of Russia (Unique identifier of the project RFMEFI62119X0023).

REFERENCES

[1] M. Tanaka and Y. Iwakiri, “The Hepatic Lymphatic Vascular System: Structure, Function, Markers, and Lymphangiogenesis,” Cell. Mol. Gastroenterol. Hepatol., vol. 2, № 6, pp. 733-749, 2016.

[2] M. Tanaka and Y. Iwakiri, “Lymphatics in the liver,” Curr. Opin. Immunol., vol. 53, pp. 137-142, 2018.

[3] S. V. Michurina, Yu. I Borodin., I. Yu. Ishchenko, A.D. Belkin, A.V. Shurlygina and P.M. Larionov, “The lymphatic region of Wistar rat liver under conditions of combined influence of alcohol intoxication and round-the-clock lighting,” Bulletin of SB RAMS, vol. 5, pp. 44-49, 2008. (in Russian).

[4] J.W. Breslin, Y. Yang, J.P. Scallan, R.S. Sweat, S.P. Adderley and W.L. Murfee, “Lymphatic Vessel Network Structure and Physiology,” Compr. Physiol., vol. 9, № 1, pp. 207-299, 2018.

[5] A.J. Day and J.K Sheehan, “Hyaluronan: polysaccharide chaos to protein organisation,” Curr. Opin. Struct. Biol., vol. 11, № 5, pp. 617–622, 2001.

[6] C. Carreira Mouta, S.M. Nasser, E. di Tomaso, T.P. Padera, Y. Boucher, S.I. Tomarev and R.K. Jain, “LYVE-1 is not restricted to the lymph vessels: expression in normal liver blood sinusoids and down-regulation in human liver cancer and cirrhosis,” Cancer Res., vol. 61, № 22, pp.8079-8084, 2001.

[7] J. Arimoto, Y. Ikura T., Suekane, M. Nakagawa, C. Kitabayashi, Y. Iwasa, K. Sugioka, T. Naruko, T. Arakawa and Ueda M., “Expression of LYVE-1 in sinusoidal endothelium is reduced in chronically inflamed human livers,” J. Gastroenterol., vol. 45, № 3, pp. 317-325, 2010.

[8] P. Saharinen, T. Tammela, M.J. Karkkainen and K. Alitalo, “Lymphatic vasculature: development, molecular regulation and role in tumor metastasis and inflammation,” Trends Immunol., vol. 25, № 7, pp.387-395, 2004.

DOI 10.18699/BGRS/SB-2020-275

Why we need more complex gene diagnostic: the case study of exome from patient with congenital glaucoma

Dinara Ivanoshchuk
ICG SB RAS, Novosibirsk, Russia
dinara@bionet.nsc.ru

Mikhail Voevoda
ICG SB RAS, Novosibirsk, Russia
voevoda@bionet.nsc.ru

Natalia Konovalova
TSMA, Tumen, Russia
doctork@bk.ru

Konstantin Gunbin
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
genkvg@bionet.nsc.ru

***Abstract* — Usually, a detailed analysis of well-known diseases-related genes for the presence of pathogenic variants does not give any meaningful result, despite the good agreement between the patient's diagnosis and the available data on the mutations manifestation in these genes. This study, we try to disentangle such a cases using gene network aware exome analysis and variants prioritization.**

Keywords* — *gene network based mutation prioritization, exome analysis, glaucoma

MOTIVATION AND AIM

One of the major challenge in the interpretation of exome-wide results is the absence of any information about the interaction of weakly damaging mutations (likely-benign and or benign) and mutations with an unknown effect in the compounds. Earlier huge efforts directed towards disentangling biological reasons of single nucleotide variations expression and penetrance of genes implicated in complex diseases determination were taken [1, 2]. Most of approaches intended to solve this problem are based on the analysis of topologies gene network of trait and/or on the data on gene-gene or protein-protein functional interaction. This study, we try to disentangle such a type of cases using our own simplified gene network aware exome analysis and variants prioritization. To do so, we used data from anonymous patient with healthy parents and evidences of congenital glaucoma.

METHODS

Paired reads exome library of anonymized patient with signatures of congenital glaucoma was mapped onto the reference genome using bowtie2-align-s. SNVs were called using Platypus v. 0.8.1 software, we used all default filtering options, except one considering duplicated reads that we switched in two alternative states. As a result, two VCF files were obtained, containing SNVs detected based on all mapped reads and mapped reads without duplicated ones. Next, SNVs from these two VCF files were additionally filtered taking into account the quality of the nucleotides confirming them. First, all SNVs marked by Platypus with the badReads label were disregarded; these are SNVs supported by reads from only one DNA strand having low base calling quality close to the SNV (Phred quality <20). Second, using the GATK v. 4.1.4.1 software, we selected out all SNVs that did not satisfy the lowest local coverage threshold. This work, we did not take into account deletion / insertion variants (INDELs), because we identified that overwhelming number of INDELs (> 96%) was marked with the alleleBias label by the Platypus; this label indicates that the allele frequency is statistically significantly lower than expected in the heterozygous state. At the final, we got two cleaned VCF files containing SNVs, supported by all mapped reads and mapped reads without duplicated ones.

The functional annotation of known missense SNVs and SNVs located in known splice sites was done using dbNSFP v. 4.0a and GnomAD v. 2.1.1 databases. In-house Perl scripts were implemented for variants extraction using Tabix indexing technology. Information on SNV functional significance was taken in the form of 40 normalized rank scores from the dbNSFP database. For each SNV, the maximum occurrence frequency was taken from the GnomAD database.

After SNVs annotating, we prioritized the genes containing these SNVs. For the constructing gene scoring formula for gene prioritization, we used both previously obtained variables for individual SNVs (dbNSFP_boot_median and GnomAD_max_freq), and the gene-wise characteristics, indicating functional significance of the gene. We used the following gene-wise characteristics: 1) the pLI score, the probability of being loss-of-function intolerant (the pLI_value), 2) the average occurrence frequency of de novo mutations in the gene from the mirDNMR database (the DNMR); 3) the sum of the maximum length of the protein sequence and the number of exons in the gene (the AA_SS_Length).

This work, we additionally take into consideration the GADO score [2] reflecting the gene significance in the gene network relating to glaucoma phenotype (the GADO_value) and the the normalized level of the gene expression in the affected tissues – cornea, iris and anterior chamber of the eye (the TISS_value).

The final simple gene scoring formula was as follows: $S(g)=TISS\_value*GADO\_value*pLI\_value*DNMR\_coeff * \Sigma \{dbNSFP\_boot\_median*(1-GnomAD\_max\_freq)*AC\}$, where AC is the number of alternative alleles at the SNV locus (1 - heterozygote; 2 - homozygote), Σ is a sum across the SNV loci containing alternative variants in the gene, $DNMR\_coeff = \log10\{AA\_SS\_Length/DNMR\}$.

RESULTS

A simple and effective guideline for searching and prioritizing genes and exome SNVs associated with a risk to diseases was developed. Using this pipeline, we were able to uncover small number of mutated protein-coding genes

involved in cornea and eye chamber functioning as well as in whole eye (Table 1). Despite very small number of genes involved in cornea and eye chamber functioning, this gene list significantly enriched with Visual phototransduction pathway in REACTOME (p=2.3E-05) and with basolateral cell membrane in GO CC (p=5.4E-04).

ACKNOWLEDGMENT

This work supported by the RFBR grant # 18-315-00297.

REFERENCES

[1] Oulas A, et al. (2019) Selecting variants of unknown significance through network-based gene-association significantly improves risk prediction for disease-control cohorts. Sci Rep. 9(1):3266.

[2] Deelen P, et al. (2019) Improving the diagnostic yield of exome-sequencing by predicting gene-phenotype associations using large-scale gene expression analysis. Nat Commun.10(1):2837

Table 1. The list of all genes associated with glaucoma and related to cornea and eye chamber functioning as well as whole eye functioning

Mutated top genes in whole eye	Mutated top genes functioning in cornea	Mutated top genes functioning in eye chamber
CFB	CFB	CFB
GUCY2D	GUCY2D	GUCY2D
ELAC2	CNGB1	CNGB1
CCDC170	MYO7A	CNGB3
CNGB1	BBS2	
MYO7A	CNGB3	
HNF1A	HNF1A	
CNGB3	DAG1	
BBS2		
HEXB		
DAG1		

DOI 10.18699/BGRS/SB-2020-276

Title Application of a multi-layer systems toxicology framework for in vitro assessment of the biological effects of liquids and corresponding aerosols

Nikolai V. Ivanov
PMI R&D, Philip Morris
Products S.A., 2000
Neuchâtel, Switzerland
nikolai.ivanov@pmi.com

Abstract — **Description We previously proposed a systems toxicology framework for in vitro assessment of e-liquids. The framework starts with the first layer aimed at screening the potential toxicity of e-liquids, followed by the second layer aimed at investigating the toxicity-related mechanism of e-liquids, and finally the third layer aimed at evaluating the toxicity-related mechanism of the corresponding aerosols. In this work, we applied this framework to assess the impact of the e-liquid MESH Classic Tobacco and its aerosol compared with that of cigarette smoke (CS) from the 3R4F reference cigarette. The results showed that the cytotoxicity of the MESH Classic Tobacco liquid was similar to the Base liquid but lower than 3R4F CS TPM at comparable nicotine concentrations. Relative to 3R4F CS exposure, MESH Classic Tobacco aerosol exposure did not cause tissue damage and elicited lower changes in the mRNA, microRNA, and protein markers. In the context of tobacco harm reduction strategy, the framework is suitable to assess the potential reduced impact of electronic cigarette aerosol relative to CS.**

Keywords — Systems Toxicology, in vitro, e-liquid

DOI 10.18699/BGRS/SB-2020-277

How miRNAs can protect humans from coronaviruses COVID-19, SARS-CoV, and MERS-CoV

Anatoliy Timofeevich Ivashchenko
Al-Farabi Kazakh National University
Almaty, Kazakhstan
a.iavashchenko@gmail.com

Aizhan Kazievna Rakhmetullina
Al-Farabi Kazakh National University
Almaty, Kazakhstan
zhanullina1994@gmail.com

Aigul Nurlankyzy Akimniyazova
Al-Farabi Kazakh National University
Almaty, Kazakhstan
akimniyazova@gmail.com

Dana Evgenievna Aisina
Al-Farabi Kazakh National University
Almaty, Kazakhstan
dana.aisina03@gmail.com

***Abstract* — Viral diseases cause significant harm to human health and often cause high mortality. In the past twenty years, humanity has undergone infection by SARS-CoV (severe acute respiratory syndrome), MERS-CoV (Middle East respiratory syndrome) and COVID-19 coronaviruses, which spread from animals to humans and from person to person. These diseases have led to large economic losses. To fight coronaviruses and other viruses, it is proposed to use miRNAs, which regulate protein synthesis at the translational level. Out of 2565 miRNAs, miR-4778-3p, miR-6864-5p and miR-5197-3p were identified as the most effectively interacting with the gRNA of SARS-CoV, MERS-CoV and COVID-19, respectively. Based on the miR-4778-3p, miR-6864-5p and miR-5197-3p sequences, complete complementary miRNA (cc-miR) binding sites in the gRNA coronaviruses were created. Detected binding sites of these cc-miRs did not form intramolecular complexes in the 2D structure of the gRNA of SARS-CoV, MERS-CoV, and COVID-19 with a value of more than 85%. Therefore, cc-miRs will bind gRNA at these sites without competition. The cc-miRs for SARS-CoV, MERS-CoV, and COVID-19 did not have target genes among the 17508 human coding genes with a ΔG/ΔGm of more than 85%, which implies the absence of side effects of cc-miRs on the translation of human mRNAs. cc-miRs can be used as therapeutic agents by incorporating them into exosomes or other vesicles and introducing them into the blood or lung by inhalation. The introduction of cc-miR into the blood will suppress the reproduction of the virus in the blood and in all organs into which it can enter.**

Keywords* — *SARS-CoV, MERS-CoV, COVID-19, virus, mRNA, miRNA, disease, treatment

Introduction

Viral diseases are difficult to predict, which requires prompt responses to their occurrence. Unfortunately, at present, specific drugs do not exist for all viruses. Coronaviruses COVID-19, SARS-CoV (severe acute respiratory syndrome) and MERS-CoV (Middle East respiratory syndrome) are highly pathogenic, and currently, there are no effective drugs to treat them [1]. Therefore, new approaches to the treatment of viral diseases are required. All viruses multiply using a nucleic acid synthesis system and a protein synthesis system in an infected cell. In animal cells, protein synthesis is controlled by miRNA (**m**RNA-**i**nhibiting **RNA**). These nanoscale molecules specifically interact with the mRNAs of target genes and can reversibly suppress translation or contribute to the degradation of mRNA. This biological role of miRNAs can help in the fight against viral reproduction, since the synthesis of viral proteins occurs in the host cell. The use of miRNAs for this purpose requires with the fulfillment of a number of conditions. The miRNA must highly specifically bind to the mRNA of the viral genome (gRNA) or parts of its genome, which can be translated as a whole or in fragments. An arbitrary set of nucleotides in synthetic miRNAs that bind with complete complementarity to the RNA genomes of viruses can have many side effects consisting of the binding of these miRNAs to human mRNAs. To identify target miRNAs of genes, a program that adequately evaluates the interaction of this miRNA with all genes of the human genome is required. The MirTarget program we created successfully determines the characteristics of the interaction of known miRNAs with any RNA nucleotide sequences. Using this method, the effect of known human miRNAs on the mRNA of plant and animal genes was evaluated. It is currently relevant to identify human miRNAs that can affect the expression of the COVID-19, SARS-CoV and MERS-CoV coronavirus genomes at the translation stage. It is necessary to select miRNAs that can most effectively bind to the nucleotide sequence of the RNA genomes of these coronaviruses. The aim of this work was to construct miRNAs specific for coronaviruses COVID-19, SARS-CoV, and MERS-CoV that will not affect the expression of human or animal genes.

Materials and Metods

The nucleotide sequences of 2565 miRNAs were downloaded from miRBase (http://mirbase.org, release 22.1). The nucleotide sequences of genes and the coronavirus genome COVID-19, MERS-CoV, SARS-CoV were obtained from NCBI (http://www.ncbi.nlm.nih.gov). A search for the target genes of miRNAs was performed using the MirTarget program [2]. The MirTarget program works with the miRNA and mRNA nucleotide sequences in a special format. This program determines the following binding characteristics: the start of the miRNA binding site on the mRNA; the location of the miRNA binding site (3'UTR, 5'UTR, CDS); the interaction free energy (ΔG, kJ/mole); and the nucleotide interaction schemes between miRNAs and mRNAs. The ratio of ΔG/ΔGm (%) was determined for each binding site, where ΔGm is equal to the free energy binding of miRNA with its full complementary nucleotide sequence. The MirTarget program found hydrogen bonds between adenine (A) and uracil (U), guanine (G) and cytosine (C), G and U, A and C. The distances between A and C were equal to 1.04 nanometers, between G and C, and between A and U were equal to 1.03 nanometers, and between G and U equal to 1.02 nanometers. The numbers of hydrogen bonds in the G-C, A-

U, G-U and A-C interactions were found to be 3, 2, 1 and 1, respectively.

Results and Discussion

To compile the nucleotide sequence of a fully complementary miRNA (cc-miR) to its binding site in the gRNA of COVID-19, it was necessary to find the cc-miR precursor. Despite the large size of the COVID-19 genome, in comparison with the human protein coding genes, only a few human miRNAs could bind to the viral genome with a ΔG/ΔGm value of 89%. We chose this value as a criterion for the adequacy of determining binding sites based on the requirement that different miRNAs with a length of 22 nt differ in two or more nucleotides, which allows them to act specifically. A decrease in this criterion, for example, by 5%, leads to an exponential increase in the number of putative target genes of a particular miRNA, which results in an uncertainty in the selectivity to the target genes of this miRNA. To create cc-miR, we selected miR-5197-3p, which out of the 2565 known miRNAs (miRBase), binds to the gRNA of the COVID-19 virus with the largest ΔG/ΔGm value of 89% and a miRNA length of 23 nt.

Of the 17,508 human genes available in our gene database, miR-5197-3p bound to the mRNA of only a few genes with similar characteristics. Therefore, the use of miR-5197-3p as a therapeutic agent will cause side effects on the identified genes.

Based on this scheme, a structure for cc-miR2 (cc-miR for the gRNA of COVID-19) with a length of 25 nt was proposed with the replacement of non-canonical nucleotide pairs by canonical ones and the addition of G and U nucleotides at the ends of miR-5197-3p to increase the free interaction energy of cc-miR2 with the gRNA of COVID-19. The miRNAs with a length of 25 nt are found among natural miRNAs (miRBase) and can interact with gRNA as part of RISC (RNA-induced silencing complex). The created cc-miR2 was completely complementary to the gRNA and interacted weakly with mRNA of 17508 genes. This result gives confidence that cc-miR2 can interact with gRNA without side effects on the human protein-coding genes.

Similarly, cc-miRs were designed to bind to the MERS-CoV (cc-miRm) and SARS-CoV (cc-miRs) genomes. Of the 2565 known human miRNAs, miR-6864-5p and miR-4778-5p could interact more strongly than other miRNAs with the gRNA of MERS-CoV and SARS-CoV, respectively. The cc-miRm and cc-miRs do not bind to mRNAs of the 17,508 studied human genes with ΔG/ΔGm values above 85%.

Therefore, these synthetic miRNAs are not expected to have any side effects upon interacting with human genes. The nucleotides of the cc-miR2, cc-miRm, cc-miRs binding sites during intramolecular interactions in the COVID-19, SARS-CoV, and MERS-CoV genomes can bind to other gRNA sites with a value of ΔG/ΔGm less than 71%. Consequently, synthetic cc-miR2, cc-miRm, and cc-miRs will interact with the binding sites in the genomes of the three coronaviruses without competition with other gRNA sites.

Generating cc-miR is an inexpensive and relatively short procedure, similar to the synthesis of primers. In cell culture, the effects of cc-miR and miRNA on coronaviruses need to be tested. The proposed hypothesis can be confirmed in the laboratory with the approval and ability to conduct inexpensive and time-saving tests of the proposed cc-miR as a means of combating coronavirus. Since the size of cc-miR is approximately 10 nm, it can be delivered with blood to many organs as a component of ordinary exosomes in human blood, which measure 30-60 nm. The cc-miR contained in exosomes can be introduced into the lungs by inhalation. The proposed method for combating coronaviruses using miRNA does not have side effects and is economically inexpensive. Like all human miRNAs, cc-miR is susceptible to degradation by nucleases, and its removal from the body is not difficult.

A well-known observation that COVID-19 infects children under the age of 15 less than adults over 60 can be explained by the expression of piwi-interacting RNA (pi-RNA) and miRNA, whose concentrations change during ontogenesis: the proportion of pi-RNA decreases with age, and the proportion of miRNA increases. It is possible that the expression of one or more pi-RNAs that can bind to the viral gRNA decreases during ontogenesis, and the virus begins to multiply. Similarly, the concentration of miRNAs capable of binding to the gRNA and are expressed in the early stages of ontogenesis can decrease, and the virus begins to multiply.

Conclusion

The results show that it is not difficult to select the nucleotide composition of miRNA, siRNA or pi-RNA for target binding to mRNA. The challenge is to predict and verify the absence of side effects of the small RNA on other genes. Unfortunately, this mandatory analysis is performed only at the stage of experimental testing, which will take a long time and is not guaranteed to end successfully. Our proposed method of searching for synthetic miRNA and evaluating its side effects can significantly simplify the generation of the miRNA of interest.

ACKNOWLEDGMENT

We thank Pyrkova A. Yu for the help in the data analysis and interpretation.

REFERENCES

[1] K. Alburikan, H. Abuelizz, “Identifying factors and target preventive therapies for Middle East Respiratory Syndrome sucsibtable patients,” Saudi Pharm J, vol. 28, pp. 161-164, 2020. DOI: 10.1016/j.jsps.2019.11.016.

[2] A. T. Ivashchenko, O. A. Berillo, A. Y. Pyrkova, R. E. Niyazova, S. A. Atambayeva, “MiR-3960 Binding Sites With mRNA of Human Genes,” Bioinformation, Vol. 10, pp. 423-427. 2014. DOI: 10.6026/97320630010423.

DOI 10.18699/BGRS/SB-2020-278

Micrornas (221, 429) are correlation with lymphocytes of axillary lymph nodes in experimental breast cancer

Aleksey Kabakov
RICEL-Branch of
IC&G SB RAS,
Novosibirsk, Russia
kabakov_av85@mail.ru

Alexandr Lykov
RICEL-Branch of
IC&G SB RAS,
Novosibirsk, Russia
aplykov2@mail.ru

Oleg Kazakov
RICEL-Branch of
IC&G SB RAS,
Novosibirsk, Russia
kazakoff_oleg@mail.ru

Alexandr Poveshchenko
RICEL-Branch of
IC&G SB RAS,
Novosibirsk, Russia
poveshchenkoa200@mail.ru

Olga Poveshchenko
RICEL-Branch of
IC&G SB RAS,
Novosibirsk, Russia
poveschenkoov@yandex.ru

Dmitriy Strunkin
RIMBB FRC FTM
Novosibirsk, Russia
strunkind@mail.ru

Lyudmila Gulyaeva
RIMBB FRC FTM
Novosibirsk, Russia
gulyaeva@soramn.ru

Andrey Letyagin,
RICEL-Branch of
IC&G SB RAS,
Novosibirsk, Russia
letyaginay@bionet.nsc.ru

Vladimir Konenkov
RICEL-Branch of
IC&G SB RAS,
Novosibirsk, Russia
vikonenkov@gmail.com

Abstract — **The study of concentration miRNA (21, 221, 222, and 429) in serum, lymph of the thoracic duct and breast cancer tissue, as well as the cellular composition of axillary lymph node in Wistar female rats in chemically induced breast cancer. Concentration microRNA (-221, -429) in the tissues breast cancer and lymph are correlated with a decrease in number of lymphocytes in medullary cords of axillary lymph nodes.**

Keywords - microRNA, breast cancer, lymphocytes

INTRODUCTION

In women all over the world, breast cancer (BC) is one of most common pathologies [4]. One of new important diagnostic markers of tumor growth, attracting the attention of researchers, are microRNAs. MicroRNAs (18–25 nucleotides) small non-coding RNAs take part in regulation of vital functions of cells (proliferation, differentiation, apoptosis), expression of which is often disturbed during pathological processes in organism, including cancer pathology [6], for example, breast cancer [1, 7]. The study of microRNA correlations (-21, -221, -222, -429) in breast cancer with morphological changes in regional lymph nodes of a tumor is of scientific interest and can serve as a basis for the development of new diagnostic methods and therapy.

Purpose of the study identification of correlation relationships of microRNA levels in serum, lymph and breast tissue with axillary lymph node cells in breast cancer in Wistar rats.

MATERIALS AND METHODS

In the experiment were used rats-female Wistar 10-12 week-old, weighted 250-300 g. 1st group - control (20 intact animals), 2nd group - breast cancer. Breast cancer (BC) was induced in 20 Wistar female rats at the age of 3 months by 5-fold administration with an interval of 7 days of N-methyl-N-nitrosourea (MNU) (Sigma, USA) subcutaneously into the area of the 2nd mammary gland on the right. 24-weeks after administration MNU breast tissue, BC tissue, blood [5], and lymph specimens (thoracic duct lymph) were collected [2]. For histological examination, axillary lymph nodes were taken (caudal of 4 nodes) [3]. Total RNA from blood serum, lymph and breast tissue specimens was extracted using a kit of reagents (Vector-Best, Russian Federation) according to the instruction. To obtain the cDNA was carried out reverse transcription on the matrix of microRNAs. To determine the number of microRNAs-21, -221, -222, -429 in biological samples, QT-PCR was performed in real time using reagents on the amplifier CFX96 (Bio-Rad Lab., USA), small RNA U6 (Vector-Best reagents, Russian Federation) was used as a comparison gene. Statistical data processing was performed using the Statistica 10 software package, central tendency and dispersion measures were described by the median (Me), lower (Q1) and upper (Q3) quartiles, the significance of the difference was calculated using the Mann-Whitney U-test, and was taken with $p < 0.05$. The correlation relationships between the studied parameters were estimated by the Spearman rank correlation coefficient.

RESULTS

Induction of breast cancer led to a decrease concentration of tumor-suppressive microRNA-429 in tumor ($p<0.05$). In lymph of animals with BC, concentration of pro-oncogenic microRNA-221 increased in comparison with intact animals ($p<0.05$). Concentration of tumor-suppressive microRNA-429 in lymph in female rats with BC was decreased compared of intact animals ($p<0.05$).

In study of the cellular composition of axillary lymph nodes in Wistar rats with BC revealed changes compared with intact animals. In study of the cellular composition of axillary lymph nodes in Wistar rats with BC revealed changes compared with intact animals. It was shown that in medullary cords, there was a decrease in number of medium lymphocytes

(p = 0.0005) and mature plasma cells (p = 0.013), an increase in number of small lymphocytes (p = 0.0001) and reticular cells (p = 0.0001).

When analyzing the correlation relationships in animals with BC, the number of mature plasma cells of the axillary lymph node medullary cords was correlated with concentration of miR-429 in tumor tissues (R = 0.51; p = 0.0033). In lymph of animals with BC, the number of medial lymphocytes of medullary cords of axillary lymph nodes had a reverse correlation with concentration of microRNA-221 in the lymph (R = –0.62; p = 0.028). At the same time, in Wistar rats with BC the number of medium lymphocytes of medullary cords of axillary lymph nodes had a direct and moderate grade of correlation with concentration of miR-429 in the lymph (R = 0.62; p = 0.028).

CONCLUSION

Thus, concentration of microRNA (-221,-429) in breast cancer tissue and lymph are correlated with a decrease in number of lymphocytes in medullary cords of axillary lymph nodes.

REFERENCES

[1] Grishina K.A., Khailenko V.A., Khailenko D.V., Karpukhin A.V. Role of microRNA in the development of breast cancer and their potential as biomarkers of this disease. // Tumors of the female reproductive system. - 2018. - T. 15. - № 3. - P. 40-47.

[2] Kuznetsov, A.V. New method of collecting lymph from animals. // Bulletin of experimental biology and medicine. - 1993. - № 9. - p. 329–331.

[3] Nozdrachev A.D., Polyakov E.L. Rat Anatomy (Laboratory animals). - SPb .: Lan, 2001. - 464 p.

[4] Ferlay J., Soerjomataram I., Dikshit R., Eser S., Mathers C., Rebelo M., Parkin D.M., Forman D., Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. // Int J Cancer. – 2015. – No. 136(5): E359-86. doi: 10.1002/ijc.29210. Epub 2014 Oct 9.

[5] Harikai N., Fujii T., Onodera S., Tashiro S. Increases in plasma dihydroxyphenylacetic acid levels in decapitated mice after exposure to various stresses. // Biol. Pharm. Bull. – 2002. – Vol. 25. – No. 7. – P. 823–826.

[6] Li Z., Branham W.S., Dial S.L., Wang Y., Guo L., Shi L., Chen T. Genomic analysis of microRNA time-course expression in liver of mice treated with genotoxic carcinogen N-ethyl-N-nitrosourea. // BMC Genomics. – 2010. – Vol. 11. – P. 609. http://www.biomedcentral.com/1471-2164/11/609.

[7] Waters P.S., McDermott A.M., Wall D., Heneghan H.M., Miller N., Newell J., Kerin MJ., Dwyer R.M. Relationship between circulating and tissue microRNAs in a murine model of breast cancer. // PLoS One. – 2012. – Vol. 7, no. 11. e50459. doi:10.1371/journal.pone.0050459.

DOI 10.18699/BGRS/SB-2020-279

Serum actin-binding proteins as markers of metastasisof larynx and laryngeal pharynx cancer

Kakurina G.V.
Cancer Research Institute, Tomsk
National Research Medical Center of
the Russian Academy of Sciences
Tomsk, Russia
kakurinagv@oncology.tomsk.ru

Kolegova E.S
Cancer Research Institute, Tomsk
National Research Medical Center of
the Russian Academy of Sciences
Tomsk, Russia
biochem@oncology.tomsk.ru

Staheeva M.N.
Cancer Research Institute, Tomsk
National Research Medical Center of
the Russian Academy of Sciences
Tomsk, Russia
biochem@oncology.tomsk.ru

Cheremisina O.V.
Cancer Research Institute, Tomsk
National Research Medical Center of
the Russian Academy of Sciences
Tomsk, Russia
biochem@oncology.tomsk.ru

Kondakova I.V.
Cancer Research Institute, Tomsk
National Research Medical Center of
the Russian Academy of Sciences
Tomsk, Russia
biochem@oncology.tomsk.ru

Choynzonov E.L.
Cancer Research Institute, Tomsk
National Research Medical Center of
the Russian Academy of Sciences
Tomsk, Russia
biochem@oncology.tomsk.ru

***Abstract* — To determine the markers of metastasis, it is important to study the molecular characteristics of tumor progression among patients with Squamous cell carcinoma of the larynx and larynxopharynx (L/L SCC). The aim of the study was to study the relationship between the level of actin binding proteins (ABS) (adenylyl cyclase-associated protein 1 (CAP1), cofilin 1, ezrin, fascin 1 and profilin 1) in the blood serum of patients with L/L SCC with the main clinical and morphological characteristics and to assess the possibility of their use as markers of metastasis. Several serum ABPs (CAP1, profilin1 and fascin 1) have been identified as potential markers of metastasis of L/L SCC. For the presented sample of patients a metastasis prognosis model was obtained. For further conclusions about the clinical value of the proposed metastasis markers, further studies with an expanding sample of patients are needed.**

Keywords* — *actin binding proteins, squamous cell carcinoma of the larynx and larynxopharynx, metastases, prognosis

Motivation and Aim

Squamous cell carcinoma of the larynx and larynxopharynx (L/L SCC) is one of the most aggressive types of cancer. High one-year mortality among patients with L/L SCC is associated with a prolonged asymptomatic course and frequent metastasis [1]. The solution to the problem of predicting metastasis of L/L SCC using minimally invasive techniques is relevant. To determine the markers of metastasis, it is important to study the molecular characteristics of tumor progression. The metastasis process is closely related to the remodeling of the cytoskeleton in tumor cell, which is mediated by actin binding proteins (ABPs): adenylyl cyclase-associated protein 1 (CAP1), cofilin 1 (CFN1), ezrin (EZR), fascin (FSCN1) and profilin 1 (PFN1) [2; 3]. At the moment, there is no data on the possibility of using the above listed ABPs circulating in the blood of patients with L/L SCC as markers for the prognosis of this disease. Earlier, we showed the possibility of determining the level of CAP1 in the tumor tissue and blood serum of patients with L/L SCC as markers for predicting the course of the disease.

Therefore, the aim of the study was to study the relationship between the level of CAP1, cofilin 1, ezrin, fascin 1 and profilin 1 in the blood serum of patients with L/L SCC with the main clinical and morphological characteristics and to assess the possibility of their use as markers of metastasis.

Methods

The material for analysis was the сыворотка крови of 46 patients with L/L SCC (T1-4N0-2M0) with histologically verified diagnosis. Material for the study was taken before the special treatment. The average age of patients were 54±5,3 years. Work was carried out in compliance with the principles of voluntariness and confidentiality in accordance with the "Fundamentals of the legislation of the Russian Federation on health care" (Decree of the President of the Russian Federation 39 from 12.24.93 number 2288), the permission of the ethical committee of the Institute was obtained. Blood serum analysis of 42 patients with L/L SCC and 15 healthy volunteers was performed using ELISA kits (Human CFL1, FSCN1, EZR, PFN1 and CAP1) on a Microplate reader Multiskan FC 100 (ThermoFisher Scientific). The results in tables present as level of studying protein. The statistical analysis was carried out by software package Statistica 6.0. The results presented in tables and figures are expressed as Me (Q1; Q3), where Me - median, Q1 and Q3 - upper and lower quartiles, n – number of persons. R is the Spearman's rank correlation coefficient, P is the statistical significance level.

Results

In the serum of patients with L/L SCC the level of CAP1 was almost 5 times higher ($p = 0.01$) relative to the control group. The increase in the levels of FSCN1 and PFN1 was at the level of trends (table 1). The metastatic process was also characterized by a change in the level of serum ABPs (Table 1).

The level of FSCN1 was higher by almost 10 times ($p = 0.05$) and the level of CAP1 was 60% higher ($p = 0.03$) in the group with regional metastases. The level of PFN1 in patients with L/L SCC with metastases was lower by 30% ($p = 0.05$). Using the discriminant analysis method on the presented sample of patients, a metastasis prognosis model was obtained and is described by equation (1):

$$y = -23 + 3{,}7*CAP1 + 1{,}43*FSCN + 68*PFN1 \quad (1)$$

The sensitivity is 89%, the specificity is 62%; prognostic significance is 77%; p≤0,001.

Correlation analysis revealed a medium strength relationship between the size of the tumor node (T) and the levels of CAP1 (r=0.6; p=0.05) and PFN1 (r=0.3; p=0.05), and weak correlations between the presence of metastases (N) and levels of CAP1 (r=0.3; p=0.05) and FSCN1 (r=0.3; p=0.05).

Conclusion

The data obtained in the presented sample of L/L SCC patients indicate a relationship between the level of serum ABPs and the stage of the disease, which may have prognostic value. The results, in general, do not contradict the literature data and indicate the possibility of using serum markers in clinical practice [4; 5]. Currently, several serum ABPs (CAP1, PFN1 and FSCN1) have been identified as potential markers of metastasis of L/L SCC. For further conclusions about the clinical value of the proposed metastasis markers, further studies with an expanding sample of patients are needed.

ACKNOWLEDGMENT

The work was carried out under the financial support of the Russian Foundation for Basic Research (project No. 20-015-00151 A).

REFERENCES

[1] T. Oki, T. Hishida, J. Yoshida, M. Goto, K. Sekihara, T. Miyoshi, K. Aokage, G. Ishii, M. Tsuboi. Survival and prognostic factors after pulmonary metastasectomy of head and neck cancer: what are the clinically informative prognostic indicators? Eur J Cardiothorac Surg. 2019; 55(5):942-947. doi: 10.1093/ejcts/ezy384.

[2] G.V. Kakurina, I.V. Kondakova, O.V. Cheremisina, D.A. Shishkin, E.L. Choinzonov. Adenylyl Cyclase-Associated Protein 1 in the Development of Head and Neck Squamous Cell Carcinomas. Bull Exp Biol Med. 2016;160(5):695-7

[3] G.V. Kakurina, E.S. Kolegova, I.V. Kondakova. Adenylyl cyclase-associated protein 1: Structure, regulation, and participation in cellular processes. Biochemistry (Moscow). 2018; 83(1), c. 45-53

[4] L. Yang, Y. Teng, T.P. Han, F.G. Li, W.T. Yue, Z.T. Wang. Clinical significance of fascin-1 and laminin-5 in non-small cell lung cancer. Genet Mol Res. 2017;16 (2). doi: 10.4238/gmr16029617.

[5] M. Satoh, S. Takano, K. Sogawa, K. Noda, H. Yoshitomi, M. Ishibashi, K. Mogushi, H. Takizawa, M. Otsuka, H. Shimizu, M. Miyazaki, F. Nomura. Immune-complex level of cofilin 1 in sera is associated with cancer progression and poor prognosis in pancreatic cancer. Cancer Sci. 2017. 108(4):795-803

TABLE I. Serum level of actin binding proteins depending on the presence of the tumor process and availability of lymphogenic metastases

ng\ml	Контроль	L/L SCC (T1-4N0-2M0)	P, U-test	N0, n=23	N1-2, n=19	P1, U-test
CFL1	0.84 (0.77; 0.93)	0.80 (0.63; 1.14)	0.41	0.83 (0.60; 1.03)	0.9 (0.65;1.28)	0.42
FSCN1	5.46 (3.65; 6.90)	1.8 (0.43; 8.10)	0.07	0.64(0.34;7.43)	6.73 (6.14;8.30)	0.05
CAP1	0.03 (0.02; 0.03)	0.11 (0.08; 1.15)	0.001	0.09 (0.07; 0.14)	0.14 (0.07; 0.20)	0.05
EZR	2.30 (1.90; 2.50)	2.10 (1.69; 2.56)	0.89	2.12 (1.54; 2.60)	2.20 (1.89; 2.50)	0.83
PFN1	0.26 (0.22; 0.29)	0.28 (0.23; 0.38)	0.06	0.34 (0.28; 0.41)	0.26 (0.23; 0.33)	0.05

Note: P - significant level of differences with group «control»; P1 - significant level of differences with group with lymph node metastasis N1-2

DOI 10.18699/BGRS/SB-2020-280

Correlations between lymph concentrations of cytokine and cell morphometric parameters of mesenteric lymph nodes in rats with chemically induced breast cancer

Oleg Kazakov
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
kazakoff_oleg@mail.ru

Alexandr Poveshchenko
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
poveshchenkoa200@mail.ru.

Nikolai Orlov
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
nbo@ngs.ru

Alexey Kabakov
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
kabakov_av85@mail.ru

Alexandr Lykov
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
aplykov2@mail.ru

Dmitry Strunkin
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
Strunkind@mail.ru

Vladimir Konenkov
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
vikonenkov@gmail.com

***Abstract* — The correlation analysis between morphometric parameters and number of immunocytes of mesenteric lymph nodes with the thoracic duct lymph cytokines levels on chemically induced rat model of breast cancer (BC) was done. It is shown that in BC group the activity of the local immune response in the lymph nodes is aimed at antitumor protection. In BC group the area of the paracortical zone was similar as on intact group, the area of lymphoid nodules with germinative centers and the area of medullary substance are increased, the number of macrophages in the thymus-dependent zone and zone responsible for humoral immunity are increased. In BC group was observed positive correlation between: dividing cells in germinative and medullar substance and IL-5; between number of the medium size of lymphocytes and MIP-1α; between the number of immunoblasts in germinative centers and GRO/KC; between the number of macrophages and MCP-1; between the number of reticular cells and IL-6, and M-CSF; between the number of small size lymphocytes and the number of mature plasma B cells and Gro/KC, that can be caused by their migration from lymph nodes.**

Keywords — breast cancer, Wistar rat, mesenteric lymph node, cytokines, morphometry

Introduction

One of the pathogenetic mechanisms of tumor development and progression is protein mediators - cytokines, chemokines and growth factors. It has been shown that cytokines secreted by both lymphoid and tumor cells affect various target cells and play an important role in the pathogenesis of tumor growth and metastasis, which proceeds mainly lymphogenously [1]. The aim of the study was to assess the correlation between thoracic lymphocytokines and morphological parameters of the structural and functional zones of the mesenteric lymph nodes in a model of rat breast cancer.

Materials and Methods

The study was done on 40 female Wistar. Animal was divided into: control (intact) group and BC group. BC was induced by injection of N-methyl-N-nitrosourea (MNU) 5 times with an interval of 7 days, subcutaneously in the region of the 2nd breast on the right side. Rats were removed from the experiment 6.5 months after induction of BC under anesthesia. Lymph was taken from the cistern of the thoracic lymphatic duct. Concentrations of 24 cytokines in lymph were assessed by flow fluorimetry using Bio-Plex Pro Rat Cytokines 24-Plex Assay (Bio-Rad, USA). Spearman r rank was done to estimate correlation.

Results

Table 1 summarized the estimated correlations between the concentrations of lymph cytokines and the morphological parameters of the mesenteric lymph node. It was found that the increased area of the germinal centers of lymphoid nodules and medullary substance positively correlates with the number of mitotic cells in germinal centers (the number of which is increasing) and medullary substance with IL-5. IL-5 produced by Th2 and stimulates proliferation and differentiation of

activated B-cells. Correlation of middle size lymphocytes of germinative centers and medullary cords with chemokine MIP-1α and correlation of the number of immunoblasts with GRO/KC cytokine, which determines the chemotaxis of immunocompetent cells, may also indicate the preservation of the activity of local immune response. Structural changes

TABLE I. THE PARAMETERS OF SPIRMEAN R-RANK ESTIMATED BETWEEN THORACIC LYMPH CYTOKINES LEVELS AND MORPHOLOGICAL AND CELLULAE PERAMETERS OF MESENTERIAL LYMPH NODE ON RAT BREAST CANCER MODEL

Parameter	Group	IL-4	IL-5	IL-6	IL-12	GRO/KC	IFNγ	M-CSF	MIP-1a	MIP-3a	MCP-1
Germinative centers of secondary lymphoid nodules											
Immunoblasts	BC	-	-	-	-	0.9	-	-	-	-	-
	BC	-	-	-	-	0.95	-	-	-	-	-
Medium size lymphocytes	In	-	-	-	-	-	0.9	-	-	-	0.9
	BC	-	-	-	-	-	-	-	0.9	-	-
Reticular cells	BC	-	-	-	0.95	-	-	-	-	-	-
Mitosis	In	-	-	-	-	-	0.95	-	-	-	-
	BC	-	0.98	-	-	-	-	-	-	-	-
Paracortical zone											
Macrophage	In	-	0.97	-	-	-	-	-	-	-	0.97
	BC	-	-	-	-	-	-	-	-	-	0.95
Reticular cells	BC	-	-	0.9	-	-	-	0.9	-	-	-
Mast cells	BC	-	0.95	-	-	-	-	-	-	-	-
Medullary substance											
Medium size lymphocytes	BC	-	-	-	-	-	-	-	0.9	-	-
Small size lymphocyte	In	0.95	-	-	-	-	-	-	-	0.97	-
Immunoblasts	In	0.89	-	-	-	-	-	-	-	-	-
Macrophage	In	-	-	-	-	-	-	0.95	-	-	-
Mitosis	BC	-	0.9	-	-	-	-	-	-	-	-
Medullary sinuses											
Small lymphocyte	In	-	-	-	0.9	-	-	-	-	-	-
	BC	-	-	-	-	0.9	-	-	-	-	-
Immunoblasts	BC	-	-	-	-	-	0.89	-	-	0.89	-
Immature plasma B cells	In	-	-	-	-	0.9	-	-	-	-	-
Mature plasma B cells	In	-	-	-	0.9	-	-	-	-	-	-
	BC	-	-	-	-	0.97	-	-	-	-	-
Reticular cells	In	-	-	-	-	-	-	-	0.97	-	-

Note. In, control (intact) group; BC, breast cancer group.

in the lymph nodes can also indicate maintaining activity of the local level of the immune response. In the paracortical zone, which plays a key role in the antitumor immune response, structural signs of maintaining the local immune response are also noted: the area of the paracortical zone remains at the level of the intact group, the number of macrophages is increased, and positively correlated the number of macrophages with chemokine MCP-1 is revealed. It is known that MCP-1, produced by mesenchymal stem cells, promotes the migration and metastasis of BC cells. In addition, mesenchymal stem cells also produce IL-6, which induces the migration and invasion of BC cells [2]. Apparently, the positive correlation of reticular cells with IL-6 may be due to its production by tumor cells and serve as one of the factors of growth and progression of the tumor. This can be indicated by the correlation of the number of reticular cells with M-CSF, which affects phagocytic activity, and may also be due to the growth of the primary tumor, and metastases in the lymph nodes. The revealed correlation of the number of mast cells, as factors contributing to homeostasis in the immune system, with IL-5, may also indicate a continuing activity of the local immune response aimed at antitumor protection. The positive correlation of the number of small lymphocytes and the number of mature plasma B cells with GRO/KC chemokine, noted in the medullary sinuses, may be associated with the migration of small lymphocytes and mature plasma B cells from the lymph node. Increased production of IFNγ is positively correlated with the number of immunoblasts in the medullary sinuses. In addition, the number of immunoblasts in the medullary sinuses in the group of animals with BC also significantly correlates with the chemokine MIP-3α.

Conclusion

Thus, with chemically induced BC a study of the correlation of cytokines of the lymph of the thoracic duct with morphological changes in the mesenteric lymph nodes revealed a number of dependencies that may be due to a local immune response in the lymph nodes directed on antitumoral protection.

References

[1] M. I. Harrel, B. M. Iritani, A. Ruddell, "Tumor-induced sentinel lymph node lymphangiogenesis and increased lymph flow precede melanoma metastasis", Am. J. Pathol., vol 170(2), pp. 774-86, February 2007. DOI: 10.2353/ajpath.2007.060761

[2] A. P. Molloy, F. T. Martin, R. M. Dwyer, T. P. Griffin, M. Murphy, F. P. Barry, T. O'Brien., M. J. Kerin "Mesenchymal stem cell secretion of chemokines during differentiation into osteoblasts, and their potential role in mediating interactions with breast cancer cells", Int J Cancer, vol 124(2), pp. 326-32, January 2009. DOI: 10.1002/ijc.23939

DOI 10.18699/BGRS/SB-2020-281

Correlations between lymph concentrations of cytokines and morphometric parameters of mesenteric lymph nodes after breast cancer chemotherapy

Kazakov Oleg
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics,Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
kazakoff_oleg@mail.ru

Poveshchenko Alexandr
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics,Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
poveshchenkoa200@mail.ru

Orlov Nikolai
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics
Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
nbo@ngs.ru

Kabakov Alexey
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics,
Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
kabakov_av85@mail.ru

Lykov Alexandr
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics,Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
aplykov2@mail.ru

Strunkin Dmitry
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics,
Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
Strunkind@mail.ru

Konenkov Vladimir
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics,Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
vikonenkov@gmail.com

***Abstract* — Correlation was found between the morphometry of the mesenteric lymph nodes and the concentration of cytokines in the lymph of the thoracic duct in breast cancer (BC) caused by intramammary administration of N-methyl-n-nitrosourea. In animals with BC after chemotherapy according to the CMF scheme compared with breast cancer without treatment, positive correlation relationships were found, which may indicate an increase in the immunomodulating and antitumor effects of cytokines. Correlation of IFNγ interferon with small lymphocytes (the number of which increased) and macrophages in germination centers and mitotic dividing cells in the medullary substance, correlation in germinative immunoblast centers with MIP-1α and an increase in the number of small lymphocytes in the thymus-dependent area of the lymph nodes, correlation in the medulla interleukin IL-17 with mature plasma cells (the number of which is increased), the correlation of interleukin IL-18 with mature plasma cells in the brain sinuses.**

Keywords* — *breast cancer, chemotherapy, Wistar rat, mesenteric lymph node, cytokines, morphometry

Introduction

At metastasizing, the main, and often the only method of cancer treatment is chemotherapy, which exacerbates the existing imbalance in the immune system, having a damaging effect on lymphoid tissue, which is one of the central problems of tumor therapy. The study of the relationship of cytokine content in lymph with structural changes in mesenteric lymph nodes at chemically induced breast cancer, which has a lot in common with breast cancer in humans [1] and after chemotherapy of breast cancer will allow us to assess the state of the local immune response in this method of treatment. The aim of the study was to study the relationship of various functional groups of lymph cytokines of the thoracic lymphatic duct with morphological parameters of mesenteric lymph nodes rats after chemotherapy of breast cancer.

Materials and Methods

The experiment was conducted on 60 female Wistar rats weight 250-300 g, in compliance with the principles of humanity. Two groups of animals were formed: group 1 – control (intact animals); group 2 – breast cancer; group 3 – breast cancer chemotherapy (after 6 months from the moment of induction of breast cancer). Animals were removed from the experiment 6.5 months after induction of breast cancer under anesthesia (i.p., 40 mg/kg Nembutal). Breast cancer was induced by the introduction of N-methyl-N-nitrosourea (MNU) 5 times with an interval of 7 days, subcutaneously in the region of the 2nd breast on the right. The course of chemotherapy included: 15 mg/kg of 5-fluorouracil and 2.5 mg/kg of methotrexate (Ebewe, Austria, intraperitoneal on days 1 and 8), 3 mg/kg of cyclophosphane (Biochemistry, Russia, intraperitoneal daily once, 14 days). Lymph was taken from the cistern of the thoracic lymphatic duct. Concentrations of 24 cytokines in lymph were assessed by flow fluorimetry using a 2-beam laser automated analyzer (Bio-Plex Assay System (Bio-Rad, USA) using a commercial test system (determined dynamic range 2-32000 pkg/ml) in accordance with the instructions of the manufacturer Bio-Plex Pro Rat Cytokines 24-Plex Assay (Bio-Rad, USA). Differences

between the groups were considered significant at $p<0.05$. Correlation between the studied parameters (cytokines and cells of the structural zones of lymph nodes) were estimated by the Spearman rank correlation coefficient. Significance of the differences was calculated by U-criterion the Mann-Whitney.

Results

A study of the relationship of lymphatic cytokines with morphological changes in the lymph nodes revealed a direct correlation between the cytokines of the lymph and mesenteric lymph node (table).

In contrast to the breast cancer group without treatment, a correlation between IFN interferon and small lymphocytes (the number of which is increased) and macrophages in germinal centers and mitotic dividing cells in the medullary substance was revealed, which may be associated with the action of IFNγ, which has an immunomodulating and antitumor effect , enhancing cytotoxic reactions mediated by T-lymphocytes. This can also be indicated by a correlation in the germinating centers of immunoblasts with MIP-1α and an increase in the number of small lymphocytes in the thymus-dependent zone of the lymph nodes against the background of a decrease in its area.

An effect on the immune system can also be indicated by a correlation in the medullary substance of interleukin IL-17 with mature plasma cells (the number of which is increasing). It should be noted that the main effect of interleukin IL-17 is the activation of neutrophils and macrophages at the site of inflammation, as well as an increase in the activity of most cytokines, especially pro-inflammatory. Pro-inflammatory cytokines also include interleukin IL-12, which correlates with neutrophils in the medullary substance and is a key cytokine for enhancing the cellular immune response. Interleukin IL-18 is directly involved in the pathogenesis of breast cancer and is one of the main immunoregulatory cytokines involved in the local reaction of the body to tumor formation processes [2]. At the same time, after chemotherapy for breast cancer, interleukin IL-18 has a direct relationship with mature plasma cells in the medullary sinuses, and is a functional doubler and synergist of interleukin IL-12 in terms of biological effects [3]. Like interleukin IL-12, interleukin IL-18 promotes the preferential differentiation of T-helpers 0 into T-helpers 1. In addition, IL-18 leads to the formation of GM-CSF and, thus, enhances leukopoiesis.

Conclusion

Thus, the study of the correlation of the concentration of cytokines lymph of the thoracic duct with structural changes in the mesenteric lymph nodes revealed dependencies aimed at increasing the immunomodulatory and antitumor action of cytokines.

REFERENCES

[1] G. Esendagli, G. Yilmaz, H. Canpinar, A. Gunel-Ozcan, M. Guc, D. Guc. “Coexistence of different tissue tumourigenesis in an N-methyl-N-nitrosourea-induced mammary carcinoma model: a histopathological report in Sprague-Dawley rats”, Lab. Animals, vol 43(1), pp. 60-4, January 2009. DOI:10.1258/la.2008.007076.

[2] R. A. Merendino, S. Gangemi, A. Ruello, A. Bene, E. Losi, G. Lonbardo, F. Purello-Dambrosio. “Serum levels of interleukin-18 and sICAM-1 in patients affected by breast cancer: preliminary considerations”, Int J Biol Markers, vol 16(2), pp. 126–9, April-June 2001.

[3] S. Sugama, B. Conti. “Interleukin-18 and stress”, Brain research reviews, vol 58(1), pp. 85-95, June 2008. DOI: 10.1016/j.brainresrev.

TABLE I. CORRELATION BETWEEN CYTOKINES AND MESENTERIAL LYMPH NODE (SPIRMEAN R-RANK)

Parameter	**Group**	**IL-5**	**IL-12**	**IL-17**	**IL-18**	**GRO/KC**	**IFNγ**	**M-CSF**	**MIP-1a**	**MIP-3a**	**MCP-1**
Герминативные центры вторичных лимфоидных узелков											
Immunoblasts	bc	-	-	-	-	0.95	-	-	-	-	-
	ct	-	-	-	-	-	-	-	0.98	-	-
Medium lymphocytes Small lymphocyte	bc	-	-	-	-	-	-	-	0.9	-	-
	ct	-	-	-	-	-	0.9	-	-	-	-
Macrophage	ct			-	-		0.98				
Reticular cells Mitosis	bc	-	0.95	-	-	-	-	-	-	-	-
	bc	0.98	-	-	-	-	-	-	-	-	-
Паракортикальная зона											
Macrophage Reticular cells	bc	-	-	-	-	-	-	-	-	-	0.95
	bc	-	-	-	-	-	-	0.9	-	-	-
Mast cells	bc	0.95	-	-	-	-	-	-	-	-	-
Мозговые тяжи											
Medium lymphocytes	bc	-	-	-	-	-	-	-	0.9	-	-
Immunoblasts	ct	-	-	-	-	-	-	0.98	-	-	-
Mature plasma cells	ct			0.98	-						
Mitosis	bc	0.9	-	-	-	-	-	-	-	-	-
	ct	-	-	-	-	-	0.98	-	-	-	-
Neutrophils	ct	-	0.9	-	-	-	-	-	-	-	-
Мозговые синусы											
Small lymphocyte	bc	-	-	-	-	0.9	-	-	-	-	-
Immunoblasts	bc	-	-	-	-	-	0.89	-	-	0.89	-
Mature plasma cells	bc			-	-	0.97					
	ct	-	-	-	0.9	-	-	-	-	-	-

Note. bc, rats with breast cancer; ct - rats after breast cancer chemotherapy.

DOI 10.18699/BGRS/SB-2020-282

Pharmacological effects of recombinant FGF21 in ovariectomized mice C57Bl/6J

Antonina Kazantseva
ICG SB RAS, Novosibirsk, Russia

Tatyana Yakovleva
ICG SB RAS, Novosibirsk, Russia

Elena Makarova
ICG SB RAS, Novosibirsk, Russia

Nadezhda Kriklivaya
ICG SB RAS, Novosibirsk, Russia

Nadezhda Bazhan
ICG SB RAS, Novosibirsk, Russia

Abstract — **Postmenopausal women have an increased risk of the development of the metabolic syndrome and type 2 diabetes. In pharmacological experiments it was found that hepatic hormone fibroblast growth factor 21 (FGF21) acts as a potent metabolic regulator. But the normalizing metabolic effects of FGF21 have been described only for males. Ovariectomized female mice may be considered as a model of decreased levels of estradiol in women during postmenopause. In this study we investigated the pharmacological effects of recombinant FGF21 in ovariectomized female mice. Ovariectomized animals received daily FGF21 injections (1 mkg/g body weight) for 7 days. On the 7th day, mice were tested in a glucose tolerance test. Then in mice blood levels of insulin, glucose and adiponectin were determined. Our results showed that FGF21 did not reduce body weight and adiposity, but improved carbohydrate lipid metabolism in estrogen-deficient female mice. These data allow considering FGF21 as a potential regulator for the correction of metabolic syndrome caused by a lack of estrogen in female mice.**

Keywords — FGF21, ovariectomy, glucose tolerance, female mice

Motivation and Aim

Postmenopausal women are at increased risk of body weight gain, adiposity, and development of type 2 diabetes. Estradiol usually used for therapy of postmenopausal syndrome. Estradiol treatment improves carbohydrate lipid metabolism, but has a number of negative side effects such as malignant tumors, cardiovascular diseases, and vascular thrombosis. Currently, work is actively underway to discover new pharmacological agents that can normalize metabolic disorders caused by a decrease in estrogens in women after menopause.

Two decades ago, a new hepatic hormone fibroblast growth factor 21 (FGF21) was discovered [1] which acts as a potent metabolic regulator. In pharmacological experiments recombinant FGF21 reduces body weight and significantly improves insulin sensitivity: reduces blood levels of insulin, glucose, free fatty acids, triglycerides, increases glucose tolerance [3]. The normalizing metabolic effects of FGF21 have been described only for males. It is unknown whether FGF21 will inhibit the development of the metabolic syndrome forming in the females due to estrogen deficiency. Ovariectomized female mice may be considered as a model of decreased levels of estradiol in women during postmenopause. Therefore the aim of the study was to investigate the pharmacological effects of FGF21 in ovariectomized female mice.

Methods

C57BL/6J female mice were kept in the SPF vivarium of the Institute of Cytology and Genetics. All experiments were performed according to the European Convention for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes (Council of Europe No. 123, Strasbourg 1985) and Russian national instructions for the care and use of laboratory animals. The protocols were approved by the Independent Ethics Committee of the Institute of Cytology and Genetics (Siberian Branch of the Russian Academy of Sciences).

Three experimental groups were formed: sham surgery animals (SHAM) and ovariectomized animals that received injections of solvent (OVX), and ovariectomized animals that received daily recombinant FGF21 injections (1 mkg/g body weight) (OVX-FGF21). The injections lasted 7 days. A day 7 after a night of fasting the animals were tested in a glucose tolerance test. The animals were decapitated a day after the last injection.

Blood glucose concentration was determined using the OneTouch Select glucometer. Plasma hormone and metabolites concentrations were determined by using commercial kits according to manufacturer's instruction.

The effect of ovariectomy and FGF21 on the studied parameters in females was determined using a single-factor MANOVA variance analysis (gradations of the factor "experimental group": SHAM, OVX, OVX-FGF21) with multiple comparisons using the post hoc Tukey HSD test.

Results

Ovariectomy increased body weight and fat mass (Fig. 1A, B). Administration of FGF21 did not cancel this effect of ovariectomy. Ovariectomy did not affect the blood level of insulin and glucose (Fig. 1D, E). The blood level of adiponectin was higher in OVX females than in SHAM females, but the differences did not reach the significance level ($p=0.08$) (Figure 1F). Ovariectomy reduced glucose tolerance (Fig. 1C). In glucose tolerance test, the area under the curve (AUC) of glucose in the OVX females was higher than in the SHAM females.

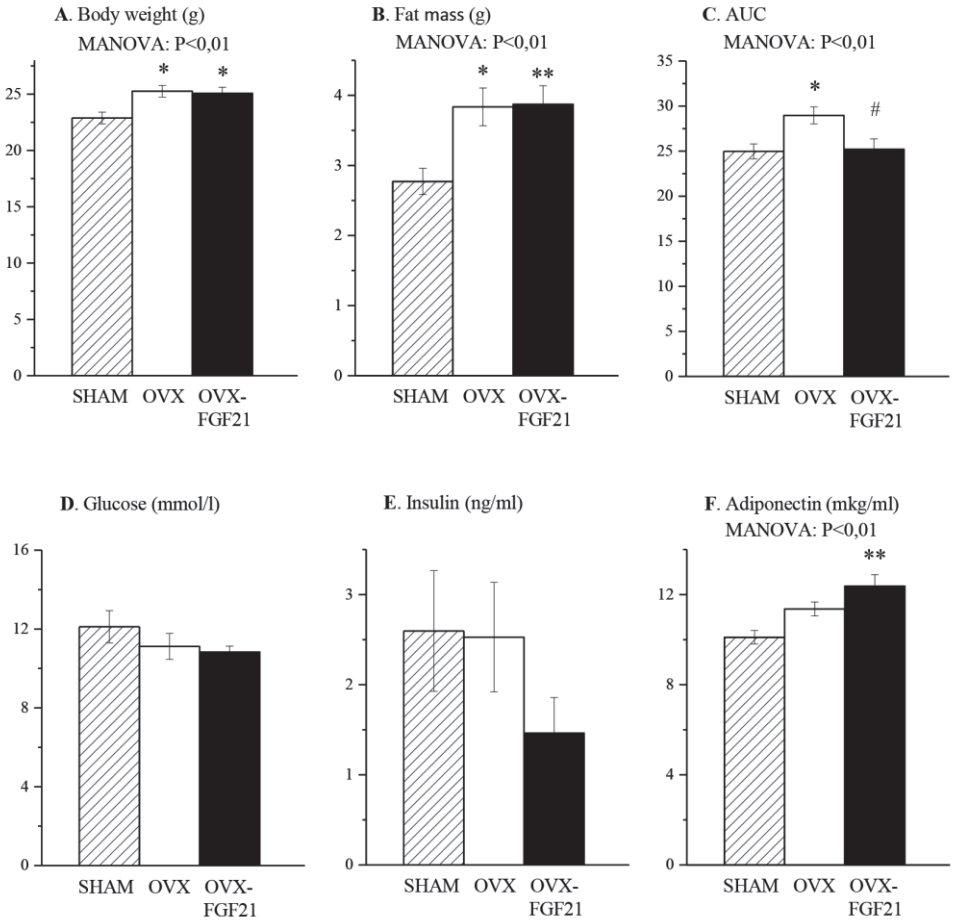


Fig. 1. Body weight (A), fat mass (B), glucose tolerance (C), plasma concentrations of glucose (D), insulin (E) and adiponectin (F) in females of C57Bl/6J mice of three groups: sham surgery animals (hatched column), gonadectomized animals that received injections of solvent (white column), and gonadectomized animals that received injections of recombinant FGF21 (1 mkg/g body weight) (black column).
* p<0.05, ** p<0.01 compared with SHAM; # p<0.05 compared with OVX (post hoc Tukey HSD test)

In ovariectomized females administration of FGF21 improved carbohydrate lipid metabolism due to effect on the adiponectin blood level and glucose tolerance (Fig. 1C, F). In the OVX-FGF21 females adiponectin blood levels were higher than in the SHAM females and AUC was lower than in the OVX females.

Obviously one of the reasons for the increased glucose tolerance in mice treated with FGF21 is an increase in insulin sensitivity. Adiponectin in known to enhances whole-body insulin sensitivity due to FFA oxidation [2]. An increase in insulin sensitivity in OVX-FGF21 mice is also indicated by the fact that the blood insulin level in the FGF21 females was 2 times lower than in the OVX females, although the differences did not reach the significance level (Fig. 1E).

Thus, our results showed that FGF21 did not reduce body weight and adiposity, but improved carbohydrate lipid metabolism in estrogen-deficient female mice. These data suggests that FGF21 reduced body weight and improved glucose tolerance by two different mechanisms. All together, our results allow considering FGF21 as a potential regulator for the correction of metabolic syndrome caused by a lack of estrogen in female mice.

ACKNOWLEDGMENT

The study was supported by the Russian Science Foundation, project 17-15-01036. Use of the equipment of the Center for Genetic Resources of Laboratory Animals at ICG SB RAS was supported by the Ministry of Education and Science of Russia (unique project identifier RFMEFI62117X0015).

REFERENCES

[1] Li et al., "Fibroblast growth factor 21: a novel metabolic regulator from pharmacology to physiology," Front Med, vol. 7(1), pp. 25-30, 2013.

[2] Hojlund et al., "Reduced 540 plasma adiponectin concentrations may contribute to impaired insulin activation of glycogen 541 synthase in skeletal muscle of patients with type 2 diabetes," Diabetologia, vol. 49, pp. 1283–1291, 2006.

[3] Mottillo et al., "FGF21 does not require adipocyte AMP-activated protein kinase (AMPK) or the phosphorylation of acetyl-CoA carboxylase (ACC) to mediate improvements in whole-body glucose homeostasis," Mol Metab, vol. 6(6), pp. 471-481, 2017.

DOI 10.18699/BGRS/SB-2020-283

The sbvIMPROVER metagenomics diagnostics for inflammatory bowel disease challenge: results and lessons learned

Lusine Khachatryan
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
Lusine.Khachatryan@pmi.com

Carine Poussin
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
Carine.Poussin@pmi.com

Yang Xiang
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
Yang.Xiang@pmi.com

Adrian Stan
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
Adrian.Stan@pmi.com

James Battey
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
James.Battey@pmi.com

Giuseppe Lo Sasso
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
Giuseppe.LoSasso@pmi.com

Stephanie Boue
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
Stephanie.Boue@pmi.com

Nicolas Sierro
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
Nicolas.Sierro@pmi.com

Nikolai V. Ivanov
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
Nikolai.Ivanov@pmi.com

Manuel C Peitsch
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
Manuel.Peitsch@pmi.com

Julia Hoeng
PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
Julia.Hoeng@pmi.com

***Abstract* — A growing number of reports indicate the potential benefit of exploiting metagenomics for noninvasive IBD diagnostics. In order to in investigate the diagnostic potential of metagenomics data for IBD patients, we conducted the crowdsourced sbv IMPROVER Metagenomics Diagnosis for Inflammatory Bowel Disease Challenge.**

Keywords— IBD, metagenomic, crowdsourcing

Motivation and Aim

Inflammatory bowel diseases (IBD) constitute a spectrum of chronic inflammatory disorders that affect the gastrointestinal tract. Ulcerative colitis (UC) and Crohn's disease (CD) are the two main clinically defined manifestations of IBD, each with distinctive clinical and pathological features. A growing number of reports showing changes in gut microbiota in subjects with IBD indicate the potential benefit of exploiting metagenomics for noninvasive IBD diagnostics.

Methods

In order to investigate the diagnostic potential of metagenomics data to discriminate IBD patients from non-IBD subjects and, within the IBD group, to separate UC and CD subjects, we conducted the crowdsourced sbv IMPROVER Metagenomics Diagnosis for Inflammatory Bowel Disease Challenge, which was open to the worldwide scientific community from September 2019 to March 2020.

Results

Participants were provided either raw (sub-challenge 1) or processed (sub-challenge 2) independent train and test metagenomics data to develop and apply models for classifying metagenomics fecal samples from CD, UC, and non-IBD subjects. The challenge results will help answer a set of scientific questions, which include: whether metagenomics data are sufficiently informative for predicting IBD status with potential applications for IBD diagnostics; what is the nature of the most discriminative metagenomic features (taxonomy, pathway, or other), and are they distinct in UC and CD; and which predictive computational approach is the most accurate for each of the classification tasks?

In conclusion, the outcomes of the challenge will be summarized and shared with the scientific community to advance the research in the fields of IBD and diagnostics.

DOI 10.18699/BGRS/SB-2020-284

Study of serotonin transporter gene polymorphism Stin2 in two Siberian indigenous populations

Mira Khantemirova
Laboratory of computer genomics
Novosibirsk State University
Novosibirsk, Russia
hantemiramira@mail.ru

Daria Lichman
Laboratory of computer genomics
Novosibirsk State University
Novosibirsk, Russia
daria.lichman@gmail.com

Daria Bazovkina
Laboratory of behavioral neurogenomics
Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
daryabazovkina@gmail.com

Anatoly Bragin
Laboratory of computer genomics
Novosibirsk State University
Novosibirsk, Russia
ibragim@bionet.nsc.ru

Vladimir Naumenko
Laboratory of behavioral neurogenomics
Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
naumenko2002@bionet.nsc.ru

Ludmila Osipova
Laboratory of computer genomics
Novosibirsk State University
Novosibirsk, Russia
ludos77@yandex.ru

***Abstract* — Serotonin (5-HT) participates in regulation of emotions, mood, sleep, appetite, and cognitive processes. Serotonin transporter (5-HTT, SERT) modulates serotoninergic neurotransmission through reuptake of 5-HT from the synaptic cleft. A variable number tandem repeat polymorphism found in intron 2 of the 5-HTT gene (Stin2 polymorphism) influences 5-HTT expression and is associated with anxiety, depression, suicidal behavior, obsessive-compulsive disorder. The aim of the present study was to investigate Stin2 allele frequencies in populations of Tundra Nenets and Nganasans inhabiting north and east Siberia. The frequency of the STin2.10 allele was 72.1% in Tundra Nenets and 89.4% in Nganasans. The data obtained are consistent with the uniform geographic gradient of STin2 alleles, with low frequency of the STin2.10 allele in East Asia and its increase towards west.**

***Keywords* —serotonin, serotonin transporter, 5-HTT, SERT, Stin2, polymorphism, Tundra Nenets, Nganasans**

Introduction

Serotonin (5-hydroxytryptamine, 5-HT) is a signaling molecule widely distributed in the mammalian central nervous system (CNS) and in peripheral tissues. It acts both as a hormone and a neurotransmitter. As a neurotransmitter, 5-HT participates in regulation of emotions, mood, sleep, appetite, and cognitive processes [1]. Dysfunction of the central serotonin system is involved in the etiology of various mental disorders, including depression, suicidal behavior, excessive aggression, anxiety, addiction, obsessive-compulsive disorder, attention deficit/hyperactivity disorder and autism [1, 2].

Serotonin transporter (5-HTT, SERT) is localized in the membranes of presynaptic neurons and is responsible for reuptake of 5-HT from the synaptic cleft thus modulating serotoninergic neurotransmission [3]. Accordingly, 5-HTT is the major target of the selective serotonin reuptake inhibitors (SSRIs), a set of medications widely used to treat depression, anxiety, ADHD and other emotional and behavioral disorders [4].

The 5-HTT protein is encoded by a single gene SLC6A4 (Soluble carrier family 6, member 4) [5]. There are two variable number tandem repeat (VNTR) polymorphisms described for this gene. One is a 43 bp insertion/deletion ~1.2 kb upstream of the transcription start site, referred to as the 5-HTT gene-linked polymorphic region (5-HTTLPR) [6]. It has been shown that 5-HTT expression and 5-HT uptake in the cells of L/S and S/S genotypes are 1.5-2-fold lower than in the cells of L/L genotype [7]. The second VNTR found in intron 2 contains most commonly 9, 10 or 12 copies of a 16–17 bp repeat (Stin2.9, Stin2.10 and Stin2.12 alleles respectively) [8]. It has been shown that Stin2 polymorphism influences 5-HTT expression and response to transcription factors [9, 10]. There is evidence of association of these polymorphisms with anxiety [7, 11], depression [8, 12], alcohol dependence [13], suicidal behavior [14], obsessive-compulsive disorder [15].

Allele frequencies for 5-HTTLPR and Stin2 polymorphisms have been studied in different populations [16-18], including Russians [11] and other people living in Russian Federation [16, 19]. However, to date, there was no information about these polymorphisms in Tundra Nenets and Nganasans inhabiting north Siberia. Accordingly, the aim of the present study was to investigate Stin2 allele frequencies in the samples of Tundra Nenets and Nganasans.

Methods

Sample collection

Blood samples were collected during expeditions to Yamalo-Nenets Autonomous Okrug and Taymyrsky Dolgano-Nenetsky District from 1988 to 2009. Each participant was informed of the purpose of the study, and written informed consent was obtained from all participants. Each volunteer filled out a questionnaire to collect personal data, including their nationality and the nationalities of their parents and grandparents. Subjects of mixed ethnicity or close relatives were excluded from this study. The final samples included 179 Tundra Nenets and 109 Nganasans.

Genotyping

Genomic DNA was extracted from peripheral blood using standard phenol–chloroform method. Genotyping was performed using polymerase chain reaction (PCR) followed by separation of fragments in a 2% agarose gel stained with ethidium bromide. Primer sequences were 5'-GTCACTATCACAGGCTGCGAG-3' and 5'-TCTTCCTAGTCTTACGCCAGTG-3'. The PCR mixture (20 μl) contained 10 ng of DNA, 2 pM of primers, 200 μM of each dNTP, 1 U of Taq polymerase and 1X PCR buffer (67 mM Tris-HCl pH 8.8, 16mM (NH4)2SO4, 0.1% Tween 2.0,

1.0 mM MgCl2). The PCR conditions were as follows: initial denaturation for 3 min at 94 °C; 35 cycles of 15 s at 94 °C, 30 s at 64 °C, 40 s at 72 °C; and final incubation at 72 °C for 2 min.

Statistical analysis

Correspondence of genotype frequencies to the Hardy-Weinberg equilibrium was assessed using Pearson's chi-squared ($\chi 2$) criterion. In the Nganasan sample, we found no 10/10 genotypes. So these homozygotes were combined with the heterozygotes to assess the correspondence to Hardy-Weinberg equilibrium. Differences in allele frequencies between the populations were estimated using a $\chi 2$ test.

Results

Genotype counts and allele frequencies for STin2 are shown in Table 1. The distribution of genotypes was in accordance with Hardy-Weinberg equilibrium in the samples of Nenets ($p=0.065$) and Nganasans ($p=0.771$). We did not find carriers of the 9-repeat allele in either of the two samples.

Both indigenous populations studied have higher frequencies of Stin2.12 allele compared to Russians (60% for the published Russian sample [18], $p<0.001$).

Conclusion

The data obtained are consistent with the uniform geographic gradient of STin2 alleles, with low frequency of the STin2.10 allele in East Asia and its increase towards west. The information on the Stin2 allele frequency in indigenous populations may be further used in personalized medicine in order to choose the optimal treatment options for psychiatric disorders.

TABLE I. Genotype And Allele Distribution For Stin2

Population		**Genotype count**			**Alleles, %**	
	N	***12/12***	***12/10***	***10/10***	***12***	***10***
Nenets	179	88	82	9	72.1	27.9
Nganasans	109	86	23	0	89.4	10.6

ACKNOWLEDGMENT

The work was supported by the Russian Scientific Foundation, project 19-15-00219.

REFERENCES

[1] I. Lucki, "The spectrum of behaviors influenced by serotonin," Biol. Psychiatry, vol. 44, pp. 151-162, August 1998.

[2] B. Olivier, "Serotonin: A never-ending story," Eur. J. Pharmacol., vol. 753, pp. 2-18, April 2015.

[3] R. D. Blakely, L. J. De Felice, and H. C. Hartzell, "Molecular physiology of norepinephrine and serotonin transporters," J. Exp. Biol., vol. 196, pp. 263-281, November 1994.

[4] J. Hyttel, "Pharmacological characterization of selective serotonin reuptake inhibitors (SSRIs)," Int. Clin. Psychopharmacol., vol. 9 Suppl. 1, pp. 19-26, March 1994.

[5] S. Ramamoorthy, A. L. Bauman, K. R. Moore, H. Han, T. Yang-Feng, A. S. Chang, et al., "Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cloning, expression, and chromosomal localization," Proc. Natl. Acad. Sci., vol. 90(6), pp. 2542-2546, March 1993.

[6] A. Heils, A. Teufel, S. Petri, G. Stöber, P. Riederer, D. Bengel, et al., "Allelic Variation of Human Serotonin Transporter Gene Expression," J. Neurochem., vol. 66(6), pp. 2621-2624, November 2002.

[7] K.-P. Lesch, D. Bengel, A. Heils, S. Z. Sabol, B. D. Greenberg, S. Petri, et al., "Association of Anxiety-Related Traits with a Polymorphism in the Serotonin Transporter Gene Regulatory Region," Science (80-), vol. 274(5292), pp. 1527–1531, November 1996.

[8] A. D. Ogilvie, S. Battersby, V. J. Bubb, G. Fink, A. J. Harmar, G. M. Goodwim, et al., "Polymorphism in serotonin transporter gene associated with susceptibility to major depression," Lancet (London, England), vol. 347(9003), pp.731–733, March 1996.

[9] E. A. Lovejoy, A. C. Scott, C. E. Fiskerstrand, V. J. Bubb, and J. P. Quinn, "The serotonin transporter intronic VNTR enhancer correlated with a predisposition to affective disorders has distinct regulatory elements within the domain based on the primary DNA sequence of the repeat unit," Eur. J. Neurosci., vol. 17(2), pp. 417–420. January 2003.

[10] E. Klenova, A. C. Scott, J. Roberts, S. Shamsuddin, E. A. Lovejoy, S. Bergmann, et al., "YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders," J. Neurosci., vol. 24(26), pp. 5966–5973, June 2004.

[11] A. V. Kazantseva, D. A. Gaysina, G. G. Faskhutdinova, T. Noskova, S. B. Malykh, and E. K. Khusnutdinova, "Polymorphisms of the serotonin transporter gene (5-HTTLPR, A/G SNP in 5-HTTLPR, and STin2 VNTR) and their relation to personality traits in healthy individuals from Russia," Psychiatr. Genet., vol. 18(4), pp. 167–176, August 2008.

[12] N. Goldman, D. A. Glei, Y.-H. Lin, and M. Weinstein, "The serotonin transporter polymorphism (5-HTTLPR): allelic variation and links with depressive symptoms," Depress. Anxiety, vol. 27(3), pp. 260–269, March 2010.

[13] M. Laucht, J. Treutlein, B. Schmid, D. Blomeyer, K. Becker, A. F. Buchmann, et al., "Impact of Psychosocial Adversity on Alcohol Intake in Young Adults: Moderation by the LL Genotype of the Serotonin Transporter Polymorphism," Biol. Psychiatry, vol. 66(2), pp. 102–109, July 2009.

[14] C. L. de Lara, A. Dumais, G. Rouleau, A. Lesage, M. Dumont, N. Chawky, et al., "STin2 Variant and Family History of Suicide as Significant Predictors of Suicide Completion in Major Depression," Biol. Psychiatry, vol. 59(2), pp. 114–120, January 2006.

[15] E. Baca-Garcia, C. Vaquero-Lorenzo, M. Diaz-Hernandez, B. Rodriguez-Salgado, H. Dolengevich-Segal, M. Arrojo-Romero, et al., "Association between obsessive–compulsive disorder and a variable number of tandem repeats polymorphism in intron 2 of the serotonin transporter gene," Prog. Neuro-Psychopharmacology Biol. Psychiatry, vol. 31(2), pp. 416–420, March 2007.

[16] J. D. Murdoch, W. C. Speed, A. J. Pakstis, C. E. Heffelfinger, and K. K. Kidd, "Worldwide population variation and haplotype analysis at the serotonin transporter gene SLC6A4 and implications for association studies," Biol. Psychiatry, vol. 74(12), pp. 879–889, December 2013.

[17] J. Gelernter, H. Kranzler, and J. F. Cubells, "Serotonin transporter protein (SLC6A4) allele and haplotype frequencies and linkage disequilibria in African- and European-American and Japanese populations and in alcohol-dependent subjects," Hum. Genet., vol. 101(2), pp. 243–246, December 1997.

[18] D. Gaysina, A. Zainullina, R. Gabdulhakov, and E. Khusnutdinova, "The serotonin transporter gene: polymorphism and haplotype analysis in Russian suicide attempters," Neuropsychobiology. vol. 54(1), pp.70–74, October 2006.

[19] A. S. Gureyev, A. A. Kim, Y. D. Sanina, V. I. Shirmanov, V. A. Koshechkin, O. P. Balanovskiy, et al., "Serotonin transporter gene 5-HTTLPR VNTR allele frequency distribution in Africa and Eurasia," Ekol. Genet., vol. 12(3), pp. 71-86, 2014.

DOI 10.18699/BGRS/SB-2020-285

Proteomic analysis of extracellular vesicles produced by placental mesenchymal stem cells

Anastasia Khutornenko
Laboratory of Cell Technologies, Research Center for Obstetrics, Gynecology and Perinatology, Moscow, Russia
bioingenier@gmail.com

Anastasia Zharikova
Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia
azharikova89@gmail.com

Vasily Popkov
The A.N. Belozersky Institute Of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia
popkov.vas@gmail.com

Sergey Kovalchuk
Laboratory of Bioinformatic methods for Combinatorial Chemistry and Biology, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia
xerx222@gmail.com

Kirill Goryunov
Laboratory of Cell Technologies, Research Center for Obstetrics, Gynecology and Perinatology, Moscow, Russia
kirishgor@gmail.com

Yulia Shevtsova
Laboratory of Cell Technologies, Research Center for Obstetrics, Gynecology and Perinatology, Moscow, Russia
yulshevtsova@yandex.ru

Egor Plotnikov
A.N. Belozersky Institute of Physico-Chemical Biology; Research Center of Obstetrics, Gynecology and Perinatology, Moscow, Russia
plotnikov@belozersky.msu.ru

Dmitry Zorov
A.N. Belozersky Institute of Physico-Chemical Biology; Reserch Center of Obstetrics, Gynecology and Perinatology, Moscow, Russia
zorov@belozersky.msu.ru

Denis Silachev
A.N. Belozersky Institute of Physico-Chemical Biology; Research Center of Obstetrics, Gynecology and Perinatology Moscow, Russia
d_silachev@oparina4.ru

Abstract — **Mesenchymal stem/stromal cells (MSCs) have therapeutic potential in many pathological conditions, that is explained mostly by their paracrine action. MSCs secrete extracellular vesicles (EVs) packed with proteins, DNA, RNA and lipids. One of the easily accessible and reach sources of MSCs is placenta. In order to reveal possible pathways, by which EVs produced by placenta-derived MSCs (EVs-PMSC) could realize their therapeutic potential we conducted proteomic analysis of EVs-PMSC and subsequent bioinformatic analysis of identified proteins using Reactome pathway database. Identified peptides were mapped to 2093 proteins. Among the 2093 proteins, 972 proteins were identified in at least four LC-MS/MS analyses out of eight. According to Reactome the most significant 'top three' 1st level pathways for detected EVs-PMSC proteins were 'Cellular responses to external stimuli'; 'Immune system'; 'Developmental biology'. Some of the most enriched pathways in the group 'Cellular responses to external stimuli' were: 'Response of EIF2AK4 (GCN2) to amino acid deficiency'; 'Cellular response to hypoxia'; 'Detoxification of Reactive Oxygen Species'. Some of the most enriched pathways in the group 'Immune System' were: 'Class I MHC mediated antigen processing & presentation'; 'Neutrophil degranulation'; 'Interferon Signaling'. The most enriched pathway in the group 'Developmental biology' was 'Axon guidance'. Thus far, concerning most enriched pathways EVs-PMSC proteome is expected to: 1) provide protection against stress (amino acid deficiency, hypoxia, ROS) to the damaged tissues; 2) possess immunomodulation properties; 3) facilitate axon guidance. The result (the top list of enriched pathways) leads to the notion that EVs-PMSC could be effective in terms of neuroprotection.**

Keywords — extracellular vesicles, mesenchymal stem cells, placenta, proteomic analysis, reactome analysis

Introduction

It is widely accepted that the therapeutic potential of mesenchymal stem/stromal cells (MSCs) is attributed mostly to their paracrine activity [1], [2]. MSCs secrete extracellular vesicles (EVs) (exosomes, microvesicles, apoptotic bodies) packed with signaling molecules, proteins, DNA, RNA, lipids, all necessary for short and long distance intercellular communication [1]. One of the easily available and reach sources of MSCs is placenta tissue. In order to reveal possible pathways, by which EVs produced by placenta-derived MSCs (EVs-PMSC) could realize their therapeutic potential we conducted proteomic analysis of EVs-PMSC and subsequent bioinformatic analysis of identified proteins using Reactome pathway database. Identified peptides were mapped to 2093 proteins. Among the 2093 proteins, 972 proteins were identified in at least four LC-MS/MS analyses out of eight. Most enriched pathways revealed by Reactome analysis for identified proteins were related to cell stress response (response to hypoxia, reactive oxygen species, heat), immune system, axon guidance. The result leads to the notion that EVs-PMSC could effective in terms of neuroprotection.

Materials and Methods

MSC culture

Fresh human preterm placenta samples (n=4) were obtained from healthy women 25 to 30 years old at the V.I. Kulakov National Medical Research Center for Obstetrics, Gynecology, and Perinatology, according to the World Medical Association Declaration of Helsinki and with the permission of the local ethics committee, informed consent was obtained from all subjects.

MSCs were isolated from tissue samples and cultivated as previously described [2]. MSCs were cultivated as adhesive cultures in DMEM/F12 medium (Paneco, Moscow, Russia) containing 7% FBS (Biosera, Nuaille, France).

Isolation of EVs

EVs were isolated from MSCs conditioned medium by differential centrifugation, and analyzed by nanoparticle tracking analysis (NTA), and transmission electron microscopy (TEM) as described previously [2].

Proteomic analysis

Sample preparation was performed as in [3] with modifications. LC-MS analysis was performed on an Ultimate 3000 RSLCnano HPLC system connected to a QExactive Plus mass spectrometer (ThermoFisher Scientific) in DDA mode. Each sample was analyzed in two technical repeats. Protein identifications and data analysis were done in MaxQuant and Perseus.

A set of identified 972 proteins was used for Reactome analysis (www.reactome.org).

Results

MSCs were isolated from human preterm placenta samples. The cell viability was assessed by trypan blue exclusion (generally >95%). MSCs used in our study were positive for mesenchymal stem cell markers (CD73, CD90, CD105) and negative for hematopoietic cell markers (CD14, CD20, CD45, CD34). MSCs at the third passage were used for EVs isolation. According to NTA, EVs showed a broad size distribution (40 – 250 nm), which indicates the presence of both exosomes and microvesicles. TEM of EV preparations confirmed two types of particles (exosomes and microvesicles). The majority of objects showed a cup-shape morphology characteristic of EVs, with sizes ranging from 40 to 300 nm. A minor fraction of the smaller objects (27–95 nm) had a smooth or angulated shape.

Isolated EVs-PMSC were used for proteomic analysis. 972 proteins were identified and analyzed using Reactome pathway database. Top three enriched pathways are provided in Table I (dark grey).

TABLE I. Reactome analysis of EVs-PMSC proteome

Pathway identifier	*FDR*	*pValue*
Cellular responses to external stimuli	3.89E-15	1.11E-16
Response of EIF2AK4 (GCN2) to amino acid deficiency	3.89E-15	1.11E-16
Cellular response to hypoxia	2.41E-06	3.01E-07
Detoxification of Reactive Oxygen Species (ROS)	1.49E-03	3.72E-04
Immune system	3.89E-15	1.11E-16
Class I MHC mediated antigen processing & presentation	3.89E-15	1.11E-16
Neutrophil degranulation	3.89E-15	1.11E-16
Interferon Signaling	3.89E-15	1.11E-16
Developmental biology	3.89E-15	1.11E-16
Axon guidance	3.89E-15	1.11E-16

The group 'Response of EIF2AK4 (GCN2) to amino acid deficiency' includes ribosomal proteins and GCN1 (eIF-2-alpha kinase activator GCN1), a positive activator of the EIF2AK4/GCN2 protein kinase activity in response to amino acid starvation (www.uniprot.org).

The group 'Cellular response to hypoxia' includes some components/proteins required for HIF1a proteasomal degradation, i.e. proteasomal subunits, elongin B and C, polyubiquitin B and C.

The group 'Detoxification of ROS' includes Peroxiredoxins (PRDX1, 2, 3, 5, 6), Thioredoxin (TXN), Thioredoxin reductases 1 and 2 (TXNRD1, 2), Glutathione S-transferase P (GSTP1), protein disulfide isomerase (PDIA1), superoxide dismutase (SOD1).

The group 'Axon guidance' includes 178 of identified proteins, and can be subdivided into following enriched pathways: 'Signaling by ROBO receptors' (p-value 1,11E-16); 'L1CAM interactions' (p-value 7,77E-16); 'EPH-Ephrin signaling' (p-value 3,48E-13); 'Semaphorin interactions' (p-value 4,01E-06). Identified proteins are implicated in cytoskeleton dynamics/organization (necessary for growth cone formation), synaptic plasticity, adhesion to extracellular matrix.

The group 'MHC-I mediated antigen processing and presentation' among 114 proteins contains HLA-F, non-classical major histocompatibility class Ib molecule postulated to play a role in immune surveillance, immune tolerance and inflammation; may play a role in inflammatory responses in the peripheral nervous system; through interaction with KIR3DL2, may protect motor neurons from astrocyte-induced toxicity (www.uniprot.org); HLA-C; HLA-A; HLA-B; HLA-H; B2M.

The group 'Neutrophil degranulation' includes 117 of identified proteins involved in the regulation of neutrophil functioning. As an example of protein influencing neutrophil transendothelial migration is Annexin A2 (ANXA2).

The group 'Interferon signaling' includes 89 of identified proteins, all of them downstream of IFN, with STAT1 as a main player of IFN-induced genes expression.

Discussion

In terms of neuroprotection activation of EIF2AK4/GCN2 [4], degradation of HIF1a [5], detoxification of ROS, axon guidance are favorable processes. Immune reactions, like MHC-I expression [6], neutrophil degranulation [7], interferon signaling [8], may have dual effect: protective by modulating synaptic plasticity, axonal regeneration, tissue repair and regeneration; destructive by playing neuroinflammatory role.

Acknowledgment

This study was supported by the Russian Foundation for basic research 20-015-00414.

References

[1] K.-S. Park et al. *Stem Cell Res Ther*, vol. 10, no. 1, p. 288, Sep. 2019, doi: 10.1186/s13287-019-1398-3.

[2] D. N. Silachev et al. *Cells*, vol. 8, no. 3, 19 2019.

[3] S.I. Kovalchuk et al. *Mol Cell Proteomics*, vol. 18, no. 2, pp. 383–390, Feb. 2019, doi: 10.1074/mcp.TIR118.000953.

[4] D.W. Gietzen and S.M. Aja. *Mol. Neurobiol.*, vol. 46, no. 2, pp. 332–348, Oct. 2012, doi: 10.1007/s12035-012-8283-8.

[5] H. Shi. *Curr. Med. Chem.*, vol. 16, no. 34, pp. 4593–4600, 2009, doi: 10.2174/092986709789760779.

[6] C. Cebrián, J. D. Loike, and D. Sulzer. *Front Neuroanat*, vol. 8, p. 114, 2014, doi: 10.3389/fnana.2014.00114.

[7] J. Wang. *Cell Tissue Res*, vol. 371, no. 3, pp. 531–539, Mar. 2018, doi: 10.1007/s00441-017-2785-7.

[8] A. McDonough, R. V. Lee, and J. R. Weinstein. *Neurochem. Res.*, vol. 42, no. 9, pp. 2625–2638, Sep. 2017, doi: 10.1007/s11064-017-2307-8.

DOI 10.18699/BGRS/SB-2020-286

Whole-exome sequencing association studies on impaired spermatogenesis in different ethnic groups in Russia

Semyon K. Kolmykov
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
kolmykovsk@gmail.com

Gennadiy V.Vasiliev
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
genn@bionet.nsc.ru

Mikhail P. Ponomarenko
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
pon@bionet.nsc.ru

Maxim A. Kleshev
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
max82cll@bionet.nsc.ru

Aleksandr V. Osadchuk
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
osadchuk@bionet.nsc.ru

Ludmila V. Osadchuk
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
losadch@bionet.nsc.ru

***Abstract* — In this study we identified single nucleotide polymorphisms in whole-exome sequencing data for different ethnic groups in Russia: Slavs, Yakuts and Buryats, and investigated its associations with impaired *spermatogenesis*.**

Keywords* — *WES, SNP, association analysis, impaired spermatogenesis , reproductive potential, infertility

Introduction

The decline in male reproductive potential, marked in various regions of the world, raises the question of the causes of this global phenomenon. Currently, there is a significant gap in our knowledge of the male reproductive potential in the Russian Federation, trends in its regional and ethnic variation and genetic control. Meanwhile, these issues become especially relevant in connection with increased attention to the demographic situation in the country and increased prevention of reproductive disorders.

In approximately 40% of infertile men, the etiology of infertility and subfertility remains unclear [1], and modern molecular genetic approaches, in particular, whole-exome sequencing (WES), expanding the possibilities of genome research, can reveal new genes involved in controlling male fertility. In this study we identified single nucleotide polymorphisms (SNPs) in whole-exome sequencing data for different ethnic groups and investigated its association with spermatogenesis defects.

Materials And Methods

Study sample

The study sample included 62 participants from 3 ethnic groups Slavs, Yakuts, and Buryats (see Tab. 1).

TABLE I. SPERMATOGENIC AND ETHNIC CHARACTERISTICS OF THE INVESTIGATED MALE SAMPLE

Ethnic group	*Normospermia, num*	*Pathospermia, num*
Yakuts	10	9
Buryats	10	9
Slavs	12	12
All	32	30

a.

WES data processing and quality control

The raw WES data were aligned to the reference genome (GRCh38) using BWA-MEM [2]. Optical and PCR duplicates were removed using Picard [3]. Further steps for detecting SNVs and INDELs were carried out using HaplotypeCaller according to GATK Best Practices [4]. Consequently, sets of variations were obtained (562018, 581274, and 519824 variations in the Yakut, Slavic, and Buryat cohorts, respectively).

The obtained variation sets were converted to PLINK format. Further quality control analysis was carried out using PLINK software [5]. At the first stage, poorly presented (<2% of the number of samples) polymorphisms were filtered out, as well as ones deviating from the Hardy-Weinberg equilibrium with a threshold of 1e-6. Subsequently polymorphisms with a Minor Allele Frequency <0.5% were removed. The heterozygosity rate of the samples participating in the study was also analyzed. According to the results, none of the samples deviated ±3SD from the heterozygosity rate mean. As a result of the quality control, sets of 104113, 100202, and 90061 variants were selected in the representatives of the Yakut, Slavic, and Buryat populations, respectively for further analysis.

The analysis of the population structure of the obtained data sets yielded the values of Genomic control inflation factor λ equal to 1.03745, 1.00119 and 1.06702, which indicates a low level of population stratification of the studied samples.

Moreover, we applied the aforementioned pipeline to identify the set of single nucleotide polymorphisms for the jointed data set (62 patients: 32 patients with normospermia and 30 patients with patospermia). As a result, 103983 SNPs were obtained for further analysis of association.

Association analysis

An analysis of the association of identified polymorphisms with spermatogenesis defects was performed using $\chi 2$-test of association and logistic regression using PLINK [5].

Results And Discussion

According to the results of the association analysis, none of the associations reached Bonferroni-corrected significance levels (0.05 / number of SNPs): 4.8e-7, 4.99e-7, 5.55e-7 and 4.81e-7 for Slavs, Yakuts, Buryats and jointed groups respectively. A partial intersection of the top-20 lists of polymorphisms associated with impaired spermatogenesis was observed and analysed. In particular, the presence of an association of polymorphism in KLHL10 gene in the jointed group has been shown. An association of other SNPs in this gene with spermatogenesis disorders was shown previously [6]. For further more significant analysis of associations, it is planned to increase the size of the studied sample.

ACKNOWLEDGMENT

This study was supported by the Russian Science Foundation, grant No. 19-15-00075.

REFERENCES

[1] Krausz C, Cioppi F, Riera-Escamilla A. Testing for genetic contributions to infertility: potential clinical impact. Expert review of molecular diagnostics. 2018 Apr 3;18(4):331-46.

[2] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997. 2013 Mar 16.

[3] Broad Institute. Picard Toolkit: Java command line tools for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. http://broadinstitute.github.io/picard.

[4] Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K. Scaling accurate genetic variant discovery to tens of thousands of samples. BioRxiv. 2018 Jan 1:201178.

[5] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015 Dec 1;4(1):s13742-015.

[6] Yatsenko AN, Roy A, Chen R, Ma L, Murthy LJ, Yan W, Lamb DJ, Matzuk MM. Non-invasive genetic diagnosis of male infertility using spermatozoal RNA: KLHL10 mutations in oligozoospermic patients impair homodimerization. Human molecular genetics. 2006 Dec 1;15(23):3411-9.

DOI 10.18699/BGRS/SB-2020-287

The changes in the panel of circulating cytokines in patients with type 2 diabetes and chronic kidney disease

Anton Korbut
Laboratory of Endocrinology
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia
ORCID: 0000-0003-3502-5892

Nikolaj Orlov
Laboratory of Clinical Immunogenetics
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia
ORCID: 0000-0002-3437-7151

Maksim Dashkin
Laboratory of Endocrinology
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia
ORCID: 0000-0002-5099-5144

Ilya Vinogradov
Clinical Laboratory
"MBU-Technology" Ltd.,
Novosibirsk, Russia
ORCID: 0000-0002-1499-0995

Vyacheslav Romanov
Clinical Laboratory
"MBU-Technology" Ltd.,
Novosibirsk, Russia
ORCID: 0000-0003-1953-2536

Abstract — **The study estimates the changes in the panel of circulating cytokines in patients with type 2 diabetes (T2D) and chronic kidney disease (CKD). We examined 130 patients with T2D duration ≥10 years, including 65 individuals with estimated glomerular filtration rate (eGFR) <60 ml/min×1.73 m^2, 64 patients with increased albuminuria, and 30 healthy subjects as control. Serum IL-1β, IL-1ra, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12 (p70), IL-13, IL-15, IL-17A, bFGF, eotaxin, G-CSF, GM-CSF, IFN-γ, IP-10, MCP-1, MIP-1α, MIP-1β, PDGF-BB, RANTES, TNF-α, and VEGF were assessed by multiplex analysis. Serum high-sensitivity C-reactive protein (hsCRP), urinary nephrin and podocin, podocyte injury markers, and WFDC-2, a marker of tubulointerstitial fibrosis, were assessed by ELISA. Patients with T2D and reduced eGFR had increased levels of hsCRP, IL-1ra, IL-8, IL-13 and decreased IL-7, IL-9 and IL-15 concentrations when compared to those with preserved renal function and control (all p<0.05). Excretion of WFDC-2, but not podocyte markers, was increased in these patients (p<0.02). The elevation of albuminuria was associated with enhanced excretion of nephrin and podocin (p<0.0001) without difference in the levels of measured cytokines. The levels of IL-2, IL-4, IL-5, IL-17A, bFGF, G-CSF, IP-10 and MIP-1α correlated negatively with eGFR; whereas WFDC-2 positively correlated with hsCRP and MIP-1α and negatively correlated with IL-10 and IL-15. These results demonstrate that decreased eGFR and increased urinary excretion of WFDC-2, a marker of tubulointerstitial fibrosis, are associated with predominance of pro-inflammatory and fibrogenic molecules in the panel of circulating cytokines in T2D subjects.**

Keywords *— type 2 diabetes, chronic kidney disease, albuminuria, cytokines, inflammation, fibrosis.*

Introduction

A number of recent studies showed that non-albuminuric chronic kidney disease (CKD) became wide-spread complication in patients with type 2 diabetes (T2D) [1,2] with higher mortality risk in some cohorts as compared to albuminuric patients [3]. It has been suggested that albuminuric and non-albuminuric CKD patterns could be different in their phenotypes and pathogenic mechanisms. The growing body of evidences indicates that chronic low-grade inflammation is involved in development of CKD in obesity [4,5] and T2D [6-10]. This study aims to estimate the differences in a panel of circulating cytokines mediating low-grade inflammation and fibrosis in patients with T2D and different patterns of CKD.

Matherials and Methods

The single-center cross-sectional study included 130 adult male and female patients with T2D duration ≥10 years. Non-diabetic CKD, end-stage renal disease, urinary tract infection, treatment with DPP-4 inhibitors, GLP-1 receptor agonists and/or SGLT-2 inhibitors for three months prior to inclusion were considered as principal exclusion criteria.

Individuals with eGFR ≥60 mL/min×1.73 m^2 and urinary albumin-to-creatinine ratio (UACR) <3.0 mg/mmol were recorded as patients without chronic kidney disease (CKD) signs (CKD– group, N=33). Those with eGFR <60 mL/min×1.73 m^2 and UACR <3.0 mg/mmol were assigned to the non-albuminuric chronic kidney disease group (NA-CKD group, N=33). Patients with eGFR ≥60 mL/min×1.73 m^2 and UACR ≥3.0 mg/mmol were defined as albuminuric with preserved renal function (A-CKD– group, N=32). Individuals with eGFR <60 mL/min×1.73 m^2 and UACR ≥3.0 mg/mmol comprised the albuminuric CKD group (A-CKD+ group, N=32). Thirty healthy normoglycemic subjects (11 men and 19 women) without obesity and cardiovascular disease and corresponded to participants by age were acted as control.

Fasting serum and morning urine samples were used for laboratory investigations. Serum IL-1β, IL-1ra, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12 (p70), IL-13, IL-15, IL-17A, bFGF, eotaxin, G-CSF, GM-CSF, IFN-γ, IP-10, MCP-1, MIP-1α, MIP-1β, PDGF-BB, RANTES, TNF-α, and VEGF were assessed by multiplex analysis. Serum high-sensitivity C-reactive protein (hsCRP), urinary nephrin and podocin, as podocyte injury markers, and urinary WAP four-disulfide core domain protein 2 (WFDC-2), a marker of tubulointerstitial fibrosis, were assessed by ELISA. The results were adjusted to the urinary creatinine levels.

Results

Urinary nephrin and podocin excretions were increased significantly in all diabetic groups (all $p \leq 0.02$). Both markers were higher in A-CKD– and A-CKD+ groups when compared to CKD– and NA-CKD subjects (all $p \leq 0.04$). Patients from CKD– and NA-CKD as well as patients from A-CKD– and A-CKD+ groups did not differ by nephrin and podocin excretion. Urinary WFDC2 concentrations were markedly higher in men than in women ($p<0.000001$). Men from CKD–, NA-CKD and A-CKD+ groups demonstrated increased WFDC-2 excretion compared to control ones ($p=0.04$, $p=0.01$ and $p=0.009$, respectively), without significant differences between diabetic groups. Women with NA-CKD or A-CKD+ had elevated excretion of WFDC-2 as compared to control and CKD– women (all $p \leq 0.01$).

Patients with T2D had higher serum concentrations of CRP when compared to non-diabetic control ($p=0.009$); the changes were more prominent in A-CKD+ group ($p=0.03$ *vs.* control). Patients with T2D demonstrated elevated serum levels of IL-1ra, IL-2, IL-4, IL-13, IL-17A, eotaxin, bFGF, G-CSF, IP-10 and MIP-1α and reduced serum levels of IL-9, IL-12, IL-15, GM-CSF, INF-γ, and VEGF as compared to control (all $p<0.05$).

Patients with NA-CKD had higher levels of IL-2, IL-4, IL-5, IL-8, IL-17A, bFGF, G-CSF, MIP-1α and lower level of VEGF when compared to control and CKD– group (all $p<0.05$). The A-CKD– group was characterized by increased levels of IL-4, IL-7, IL-12, IL-17, G-CSF, IP-10 and MIP-1α (all $p<0.05$). Moreover, A-CKD– patients had higher levels of IL-1ra, bFGF, GM-CSF and lower concentration of VEGF compared to control ($p<0.05$). Patients of NA-CKD group showed higher levels of bFGF, G-CSF, GM-CSF, MIP-1α, IL-4, IL-8 and IL-17 when compared to patients with A-CKD+ (all $p<0.05$).

In a multiple logistic regression analysis, reduced eGFR was associated with hsCRP, IL-4, IL-10, and MIP-1α levels (all $p<0.05$). However, no associations between albuminuria and a panel of cytokines or hsCRP were found. The levels of hsCRP, bFGF, G-CSF, IP-10, MIP-1α, IL-2, IL-4, IL-5 and IL-17 correlated negatively with eGFR. The excretion of WFDC-2 correlated positively with MIP-1α and hsCRP levels; whereas, negative correlations were found between WFDC-2 and IL-10, IL-15 and VEGF. No correlations were observed between cytokines and markers of glomerular dysfunction: UACR, nephrin, and podocin.

Conclusion

The obtained results demonstrate that both albuminuric and non-albuminuric pattern of CKD are associated with chronic low-grade inflammation in T2D patients. The decreased eGFR and increased urinary excretion of WFDC-2, a marker of tubulointerstitial fibrosis, are associated with a predominance of pro-inflammatory and fibrogenic regulators in the panel of circulating cytokines in T2D subjects.

References

[1] F. Viazzi, G.T. Russo, A. Ceriello, P. Fioretto, C. Giorda, S. De Cosmo, R. Pontremoli. "Natural history and risk factors for diabetic kidney disease in patients with T2D: lessons from the AMD-annals", J Nephrol 2019, vol. 32(4), pp. 517–525. doi:10.1007/s40620-018-00561-3.

[2] V.V. Klimontov, A.I. Korbut. "Albuminuric and non-albuminuric patterns of chronic kidney disease in type 2 diabetes", Diabetes and Metabolic Syndrome: Clinical Research and Reviews 2019, vol. 13(1), pp. 474-479. doi: 10.1016/j.dsx.2018.11.014.

[3] G. Penno, A. Solini, E. Orsi, E. Bonora, C. Fondelli, R. Trevisan, et al. "Non-albuminuric renal impairment is a strong predictor of mortality in individuals with type 2 diabetes: the Renal Insufficiency And Cardiovascular Events (RIACE) Italian multicentre study", Diabetologia 2018, vol. 61(11), pp. 2277–2289. doi:10.1007/s00125-018-4691-2.

[4] I.A. Bondar, V.V. Klimontov, A.I. Simakova. "Obesity and chronic kidney disease", Therapeutic Archive 2011, vol. 83(6), pp. 66-70.

[5] J.I. Lakkis, M.R. Weir. "Obesity and Kidney Disease", Prog Cardiovasc Dis 2018, vol. 61(2), pp. 157–167. doi:10.1016/j.pcad.2018.07.005.

[6] I.A. Bondar, V.V. Klimontov. "Immune inflammatory mechanisms in development of diabetic nephropathy", Probl. Endokrinol. 2007, vol. 53(2), pp. 34-40.

[7] V.I. Konenkov, V.V. Klimontov, N.E. Myakina, N.V. Tyan, O.N. Fazullina, V.V. Romanov. "Increased serum concentrations of inflammatory cytokines in type 2 diabetic patients with chronic kidney disease", Therapeutic Archive 2015, vol. 87(6), pp. 45-49. doi: 10.17116/terarkh201587645-49.

[8] V.V. Klimontov, N.V. Tyan, O.N. Fazullina, N.E. Myakina, A.P. Lykov, V.I. Konenkov. "Clinical and metabolic factors associated with chronic low-grade inflammation in type 2 diabetic patients", Diabetes Mellitus 2016, vol. 19(4), pp. 295-302. doi: 10.14341/DM7928.

[9] R. Pichler, M. Afkarian, B.P. Dieter, K.R. Tuttle. "Immunity and inflammation in diabetic kidney disease: translating mechanisms to biomarkers and treatment targets", Am J Physiol Renal Physiol 2017, vol. 312(4), pp. F716–F731. doi:10.1152/ajprenal.00314.2016.

[10] R.E. Pérez-Morales, M.D. Del Pino, J.M. Valdivielso, A. Ortiz, C. Mora-Fernández, J.F. Navarro-González. "Inflammation in Diabetic Kidney Disease", Nephron 2019, vol. 143(1), pp. 12–16. doi:10.1159/000493278.

DOI 10.18699/BGRS/SB-2020-410

Analysis of glioblastoma gene ontology categories and gene network structure using online bioinformatics tools

Sergey S. Kovalev
Novosibirsk State University
Novosibirsk, Russia
kovalev@bionet.nsc.ru

Elvira R. Galieva
Novosibirsk State University
Novosibirsk, Russia
e.galieva@nsu.ru

Anna D. Panova
I.M. Sechenov First Moscow State Medical University
Moscow, Russia
ainushgnomello@gmail.com

Yuriy L. Orlov
I.M. Sechenov First Moscow State Medical University
Moscow, Russia
Novosibirsk State University
Novosibirsk, Russia
orlov@d-health.institute

***Abstract* — Glioblastoma is a dangerous disease with a low survival rate. We aimed reveal set of genes associated with glioblastoma using on-line bioinformatics tools. First, we used OMIM (Online Mendelian Inheritance in Man) tool to list the human genes associated with glioblastoma. Then we analyzed the gene ontology categories of this gene set online using bioinformatics resources PANTHER and DAVID. In addition we reconstructed the gene network of this disease using tools GeneMANIA and STING-DB. We found significant enrichment of the ontology categories related to protein binding, membranes, cell adhesion. The network connectivity was increased relative to random network structure. Thus, we observe frequent protein-protein interaction for glioblastoma genes that might affect the disease progression.**

Fig. 2. ***Keywords — bioinformatics, glioblastoma, human genes, gene network, cancer, gene ontologies, bioinformatics tools***

MOTIVATION AND AIM

Motivation

Glioma is the most common primary brain tumor. In additional, glioma enters into a heterogeneous group and has a neuroectodermal origin. This disease leads to multiple mutations where the leading mutations in the pathogenesis of gliomas are loss of heterozygosity in the long arm of chromosome 10 (LOH 10q), mutation of the PTEN gene (10q23.3), mutation in various exons of the tumor suppressor gene p53, amplification of the EGFR gene, deletion or inactivating mutations of the p16 gene, as well as hypermethylation of the MGMT gene promoter and 1p / 19q codeletion. The study of gene expression in glioblastoma cells, the search for candidate genes for therapeutic effects are relevant in modern high-tech medicine [1].

Aim

The task is to create a list of genes associated with the development of glioblastoma, analyze the categories of gene ontologies of this list and reconstruct the gene network. For key disease genes obtained by analyzing the structure of the gene network, drug search options (substances that interact with this protein) will be considered.

METHODS

We used the online resource OMIM (Online Mendelian Inheritance in Man) (https://omim.org/) to analyze the genes of Mendelian inheritance in humans [2]. A keyword search for glioblastoma yielded 244 genes. As a result, 243 gene names were obtained (GENE SYMBOL - alphanumeric designations of human genes) - ADAM10, ADAM17, ADAM9, ADGRB1, etc.

To analyze the categories of gene ontologies, we used DAVID resources (https://david.ncifcrf.gov/summary.jsp/) [3][4]. A list of 243 human genes was downloaded via the DAVID interface to search for significant categories of gene ontologies for this group of genes. 230 identifiers were recognized.

The Functional Annotation Chart option was used.

The table was limited to the values of the unnormalized p-value <0.0001 (or in the exponential notation <1.0E-04 (9.9E-05). Categories (rows of the table) for groups of less than 8 genes were deleted.

Gene ontologies were analyzed for the generated gene list using the PANTHER resource (http://pantherdb.org/) [5]. Thus, an ontology table was constructed for categories of biological processes using PANTHER. From the created gene list, 206 identifiers were recognized, 37 were not recognized or could not be unambiguously mapped. A total of 20,996 genes were used in the PANTHER reference genome. We limited ourselves to p-value values up to E-12, to represent the most informative results. The categories of processes related to the development of the nervous system are also presented.

To reconstruct the gene network of glioblastoma gene interactions, we used the resources GeneMANIA (https://genemania.org/) [6] and STRING-DB (https://string-db.org/). 226 genes were recognized for GeneMANIA, and 211 genes were recognized for STRING-DB.

RESULTS

The most significant categories for glioblastoma genes are protein binding, positive regulation of protein phosphorylation, membrane location, enzyme binding. Based on the created table, membrane proteins play an

important role, apparently because of signal transmission to the cell, although there is transcription regulation.

In the table created by PANTHER, the most significant categories for glioblastoma genes are organelles, intracellular organelles (such as mitochondria), intercellular compounds, which may be associated with signal transmission between cells and refers to neurons. Note a category such as synapse, although with a relatively low level of significance.

Thus, both DAVID and PANTHER for glioblastoma genes confirm the categories of gene ontologies for protein binding, membranes, cell adhesion, and intercellular contacts.

STING-DB statistics show that a network has a large number of connections with significance <1.0e-16, the average degree of connectivity of a network node (protein) is 10, and the clustering coefficient is 0.493

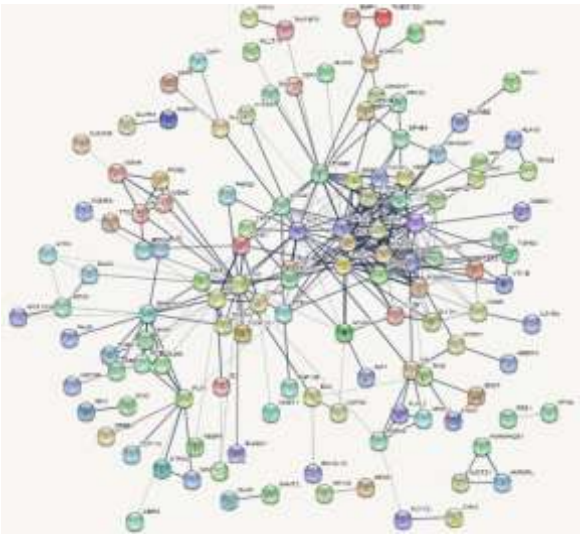


Fig. 1. Gene network for glioblastoma genes reconstructed by STING-DB

Figure 1 present gene interaction network reconstructed for the gene list. One can see several gene (protein) clusters.

The largest network cluster includes the genes AKT1, MTOR, KRAS, STAT3. The largest number of bonds is in the known TP53 oncogene.

In general, analysis of the gene network structure for glioblastoma genes shows the existence of a dense, coherent, large cluster of genes including known oncogenes, such as TP53.

The network connectivity was increased relative to random network structure. Thus, we observe frequent protein-protein interaction for glioblastoma genes that might affect the disease progression [7].

Standard bioinformatics tools appear to be effective for the gene list analysis.

ACKNOWLEDGMENT

The work was supported by the Russian Science Foundation (19-15-00219).

REFERENCES

[1] V. N. Babenko *et al.*, 'Computer Analysis of Glioma Transcriptome Profiling: Alternative Splicing Events', *J. Integr. Bioinform.*, vol. 14, no. 3, 2017.

[2] J. S. Amberger, C. A. Bocchini, A. F. Scott, and A. Hamosh, 'OMIM.org: Leveraging knowledge across phenotype-gene relationships', *Nucleic Acids Res.*, vol. 47, no. D1, pp. D1038–D1043, 2019.

[3] G. Dennis *et al.*, 'DAVID: Database for Annotation, Visualization, and Integrated Discovery.', *Genome Biol.*, vol. 4, no. 5, 2003.

[4] D. Wei Huang, B. T Sherman, and R. A Lempicki, 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.', *Nat. Protoc.*, vol. 4, no. 1, pp. 44–57, 2009.

[5] H. Mi, A. Muruganujan, and P. D. Thomas, 'PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees', *Nucleic Acids Res.*, vol. 41, no. D1, 2013.

[6] D. Warde-Farley *et al.*, 'The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function', *Nucleic Acids Res.*, vol. 38, no. SUPPL. 2, 2010.

[7] R.C. Gimple *et al.* 'Glioblastoma stem cells: lessons from the tumor hierarchy in a lethal cancer', Genes Dev. ;33(11-12):591-609, 2019

DOI 10.18699/BGRS/SB-2020-288

Investigation of the role of alternative splicing in glioblastoma by RNA-seq data

Sergey S. Kovalev
Novosibirsk State University,
Novosibirsk, Russia
kovalev@bionet.nsc.ru

Nataliya V. Gubanova
Institute of Cytology and Genetics,
Novosibirsk, Russia
nat@bionet.nsc.ru

Elena Yu. Leberfarb
Novosibirsk State Medical University,
Novosibirsk, Russia
Institute of Cytology and Genetics
SB RAS,
Novosibirsk, Russia
lieberfarb@bionet.nsc.ru

Vladimir N. Babenko
Novosibirsk State University,
Novosibirsk, Russia
Institute of Cytology and Genetics
SB RAS,
Novosibirsk, Russia
bob@bionet.nsc.ru

Yuriy L. Orlov
I.M.Sechenov First Moscow State
Medical University
Moscow, Russia
Novosibirsk State University,
Novosibirsk, Russia
orlov@d-health.institute

***Abstract* — Glioblastoma is the most aggressive type of glioma. It is classified as a tumor of the astrocytic type of the IV degree of malignancy and is characterized by extremely high heterogeneity in the composition of cell populations and the degree of differentiation of cells, the random arrangement of tumor cells, and polymorphism of their nuclei. Life expectancy in people with this disease remains low. Slow progress in increasing the effectiveness of treatment, life expectancy and improving its quality in gliomas is associated with insufficiently complete ideas about the mechanism of origin and development of these tumors. The pathogenesis of gliomas needs to be actively studied using modern cellular technologies, genome-wide technologies for high-throughput sequencing, determination of gene expression on microarrays, and methods of modern bioinformatics. The main ways of activating genes in tumor cells have already been sufficiently studied - a more detailed study of individual transcript isoforms, non-coding RNA, and alternative splicing, as one of the mechanisms of tumor development, is needed.**

Keywords* — *bioinformatics, cancer, glioblastoma, structural bioinformatics, gene expression, alternative splicing, TP53

Motivation and Aim

Motivation

Glioblastoma, along with astrocytoma, are the most common primary tumors of the central nervous system and make up approximately 60-70% of all pure brain tumors. Primary glioblastoma in 60% of cases occurs in people over 50 years old, while secondary glioblastoma is more characteristic for people under 45 years old [1][2]. Glioblastoma is characterized by rapid infiltrative growth, the presence of foci of necrosis and a change in blood vessels. She is prone to metastasis, has a bright invasive phenotype. The lack of clear boundaries and the ability to relapse are a particular problem for surgical removal of the tumor, and for immunotherapy - the specificity of the brain in relation to the immune system [3].

The main ways of activating genes in tumor cells have already been sufficiently studied and in recent years, scientists have been actively studying the role of alternative splicing in the formation of tumors of different localization [4]. Alternative splicing is important for the regulation of gene expression and the protein to perform appropriate functions in the development of cancer. Changes in alternative gene splicing are observed with cancer quite often and are already becoming one of the important putative markers of tumor growth. It has been established that for different types of cancer, the genes of splicing factors often mutate, leading to a disruption in the normal course of the process, and also that anomalies of alternative splicing contribute to the development of the body's resistance to chemotherapy [5][6].

Aim

Despite the development of new technologies for imaging neoplasms and modern methods of therapy, the overall survival of patients remains very low to this day. To date, the main problem in the diagnosis and choice of therapy is the lack of histological criteria for typing tumors and predicting its further passage. Gliomas of even one histological series are genetic and epigenetic disorders, RNA and protein expression profiles. For an object such as gliomas, it is necessary to conduct new studies based on modern cellular technologies, genome-wide technologies for high-performance sequencing, and the integration of available information from international databases and genomic projects [7].

There is an assumption that alternative splicing factors may become promising targets in the treatment of cancer patients and detailed molecular characterization of glioma stem cells is important for the development of more effective treatments [8, 9].

Methods

Earlier, a prototype of the database of differential alternative gene splicing - “Differential alternative splicing of human genes in secondary glioblastoma (DASGG)”, with the ability to work through the website and search for expression levels of individual isoforms in a glial tumor, was presented at the Institute of Cytology and Genetics SB RAS. The data on differential alternative splicing presented in this database can be used in basic research on glioma stem cells and in the development of cancer treatment diagnostics [10][11]. We used bioinformatics methods described in [12].

The work used data on the primary culture of glioma cells and normal brain obtained because of studies conducted in conjunction with the Federal State Budgetary Institution National Medical Research Center named after Academician

E. N. Meshalkin, Novosibirsk. Primary culture samples were obtained from the primary culture of secondary glioblastoma, and the primary culture of the normal human brain [13].

It is known that epigenetic changes can be observed in the brain, leading to a change in the expression profile. Taking into account this fact, when choosing cultural isolation and further sequencing, RNA takes into account the age and age of patients. The malignant properties of glioblastoma were also taken into account.

There are open international databases on gene expression in various organs (including microarrays and transcript data of GEO NCBI, http://www.ncbi.nlm.nih.gov/geoprofiles/), expression in various types of tumor cells (The Cancer Gene Atlas, cancergenome.nih.gov), gene expression for brain compartments (Allen Brain Atlas), protein interactions databases such as HPRD (http://hprd.org/), KEGG biochemical reactions (http://www.genome.jp/kegg/), Interactome (http://interactome.org/), sequenced tumor genomes, including gliomas and glioblastomas (https://cghub.ucsc.edu/). The Allen Institute has developed the Ivy Glioblastoma Atlas Project database (http://glioblastoma.alleninstitute.org/) according to patients with glioma.

Results

Advances in sequencing and bioinformatics technologies have enabled the analysis of genomic sequences in many cancers, including glioblastoma. Sequencing data for multiple GBM samples has been archived in open access databases such as the Chinese Glioma Genome Atlas (CGGA, http://www.cgga.org.cn/) and the Cancer Genome Atlas (TCGA, https: //portal.gdc.cancer.gov/). The result of a study based on these data is the determination of a prognostic gene signature that can complement commonly accepted clinical prognostic factors and additionally provide the possibility of personalized therapy. The gene signature included genes CD79B, MAP2K3, IMPDH1, SLC16A3, MPZL3, and APOBR [5].

In the process of differential splicing analysis, significant differences in splicing profiles in three genes associated with possible proliferation between normal brain and glioblastoma cells were revealed: amyloid beta precursor protein, CASC4 cancer gene (cancer susceptibility candidate 4) and the well-known oncogene, transcription factor TP53. In particular, the non-coding isoform NR_015381 with a significantly higher frequency in glioblastoma cells was observed in the TP53 gene.

The short isoform of the APP gene (NM_201413) was highly expressed in glioblastoma cells, while its longest isoform (NM_000484) was specifically expressed in normal brain cells.

For the CASC4 gene, its short isoform (NM_177974) was expressed 6 times higher in glioblastoma cells, while the longest isoform (NM_138423) also had a high level of expression. Overexpression of CASC4 is associated with increased expression of the Her2 proto-oncogene in ovarian and breast cancer. For TP53, the entire spectrum (37 of the 43 known isoforms visible on our data) isoforms observed in glioblastoma cells differed in expression level from isoforms in normal brain cells. In particular, the non-coding isoform NR_015381 applies only to glioblastoma cells.

A computer database of oncogene target genes determined from ChIP-seq experiments (including transcription factors R, MYC, TP53) was developed. The database presents the marking of the binding sites of transcription factors, as well as the genes whose expression changes upon activation of the transcription factor in tumor cells are determined.

Data on transcription factor binding sites were obtained from open publications and GEO NCBI resources. Information was obtained on the differential expression of genes. The database was replenished with its own experimental data.

Acknowledgements

The work was supported by the Russian Science Foundation (19-15-00219).

References

[1] H. Ohgaki and P. Kleihues, 'Genetic pathways to primary and secondary glioblastoma', *Am. J. Pathol.*, vol. 170, no. 5, pp. 1445–1453, 2007.

[2] R. L. Bowman, Q. Wang, A. Carro, R. G. W. Verhaak, and M. Squatrito, 'GlioVis data portal for visualization and analysis of brain tumor expression datasets', *Neuro. Oncol.*, vol. 19, no. 1, pp. 139–141, 2017.

[3] T. McGranahan, G. Li, and S. Nagpal, 'History and current state of immunotherapy in glioma and brain metastasis', *Ther. Adv. Med. Oncol.*, vol. 9, no. 5, pp. 347–368, May 2017.

[4] S. Li *et al.*, 'Identification of a gene signature associated with radiotherapy and prognosis in gliomas', *Oncotarget*, vol. 8, no. 51, pp. 88974–88987, 2017.

[5] Q. Wang *et al.*, 'Tumor evolution of glioma intrinsic gene expression subtype associates with immunological changes in the microenvironment', *Tumor Evol. Glioma-Intrinsic Gene Expr. Subtypes Assoc. with Immunol. Chang. Microenviron.*, p. 052076, 2016.

[6] X. Xue *et al.*, 'LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer', *Oncogene*, vol. 35, no. 21, pp. 2746–2755, 2016.

[7] V. N. Babenko *et al.*, 'Computer Analysis of Glioma Transcriptome Profiling: Alternative Splicing Events', *J. Integr. Bioinform.*, vol. 14, no. 3, 2017.

[8] S. C. W. Lee and O. Abdel-Wahab, 'Therapeutic targeting of splicing in cancer', *Nat. Med.*, vol. 22, no. 9, pp. 976–986, 2016.

[9] G. D. A. Guardia *et al.*, 'Proneural and mesenchymal glioma stem cells display major differences in splicing and lncRNA profiles', *npj Genomic Med.*, vol. 5, no. 1, 2020.

[10] A. V. Baranova, V. V. Klimontov, A. Y. Letyagin, and Y. L. Orlov, 'Medical genomics research at BGRS-2018', *BMC Med. Genomics*, vol. 12, 2019.

[11] E. N. Voropaeva, T. I. Pospelova, M. I. Voevoda, V. N. Maksimov, Y. L. Orlov, and O. B. Seregina, 'Clinical aspects of TP53 gene inactivation in diffuse large B-cell lymphoma', *BMC Med. Genomics*, vol. 12, 2019.

[12] V. N. Babenko *et al.*, 'Analysis of differential gene expression by RNA-seq data in brain areas of laboratory animals', *J. Integr. Bioinform.*, vol. 13, no. 4, p. 292, 2016.

[13] V.N. Babenko *et al*. 'Computer Analysis of Glioma Transcriptome Profiling: Alternative Splicing Events', *J Integr Bioinform*; 14(3). pii: /j/jib.2017.14.issue-3/jib-2017-0022/jib-2017-0022.xml 2017.

DOI 10.18699/BGRS/SB-2020-289

The functionalized graphene oxide as new anti-cancer therapeutics

Natalia Krasteva
Electroinduced and Adhesive Properties
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences
Sofia, Bulgaria
nataly@bio21.bas.bg

Milena Keremidarska-Markova
Electroinduced and Adhesive Properties
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences
Sofia, Bulgaria
m_keremidarska@abv.bg

Kamelia Hristova-Panusheva
Electroinduced and Adhesive Properties
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences
Sofia, Bulgaria
kameliahristova@abv.bg

Bela Vasileva
Laboratory of Molecular Genetics
Institute of Molecular Biology "Acad. R. Tsanev", Bulgarian Academy of Sciences
Sofia, Bulgaria
belavas@outlook.com

George Miloshev
Laboratory of Molecular Genetics
Institute of Molecular Biology "Acad. R. Tsanev", Bulgarian Academy of Sciences
Sofia, Bulgaria
karamolbiol@gmail.com

Milena Georgieva
Laboratory of Molecular Genetics
Institute of Molecular Biology "Acad. R. Tsanev", Bulgarian Academy of Sciences
Sofia, Bulgaria
milenageorgy@gmail.com

Abstract — **Clinically, there is an urgent need to identify new therapeutic strategies for selectively treating cancer cells. One of the directions in this research is the search for biocompatible therapeutics which selectively target cancer cells. Here, we show that aminated graphene oxide (GO-NH_2) nanoparticles obtained by two different methods - ammonia modification and hydroxylamine functionalization, demonstrate selectivity toward human colon cancer and hepatocellular cells. The amination by the two different methods leads to decrease in the size of GO-NH_2 nanoparticles and their zeta potential thus assuring easier penetration through the cell membrane. After characterization of the biological activities of pristine and both aminated GO we have demonstrated strong cytotoxicity of ammonia-modified GO-NH_2 toward colon cancer cells in a dose-dependent manner while in hepatic cancer cells, HepG2, hydroxylamine-functionalized GO induced a weaker toxicity effect. We have presented evidence that the common cytotoxic effects of the aminated by the two protocols GO-NH_2 on hepatic cancer cells and colon cancer cells were due to cell membrane damage and increased ROS production while in colon cancer cells the ammonia modified NPs additionally induced DNA damage which was not detected in hepatic cancer cells. In the latter severe mitochondrial dysfunction was detected. Our studies discuss the possibilities of exploiting aminated graphene oxide as an anti-cancer therapeutic and as a drug delivery agent with certain selectivity toward different tumours depending on the mode of functionalization applied.**

Keywords — cytotoxicity, genotoxicity, apoptosis, ammonia, hydroxylamine, HepG2, Colon 26 cells, nanoparticle functionalization, GO, GO-NH_2

Introduction

Graphene oxide (GO) is one of the most explored 2D nanomaterials in the last years for cancer therapy [1]. GO has an extra-high capacity to bind aromatic drugs due to its large specific surface area and abundant oxygen-containing groups [2]. Besides, GO may induce the generation of reactive oxygen species (ROS) in cells, which is one of the main toxicological mechanisms of action of nanomaterials [3]. ROS generated from GO can alter biological macromolecules including proteins, cell membrane lipids, DNA and RNA resulting in initiation of numerous signal transduction pathways linked to inflammation, malignant transformation, proliferation, and apoptosis. Thus, exposure of cancer cells to GO can significantly contribute to cancer cell killing by enhancing the cytotoxicity effect exerted through the induction of DNA damage. Therefore, GO not only can function as an effective drug carrier but also can potentially exert inhibitory effects on tumour cells when used by itself. To improve the therapeutic effect of GO-based cancer therapy, the surface properties of GO can be modulated by chemical functionalization with different functional groups, biomolecules, polymers, etc. However, depending on the synthesis method, the physico-chemical properties of the functionalized GO nanoparticles vary and could exert a palette of diverse effects on cells and tissues [4]. Here, we represent evidence for differential potential of two differently aminated GO nanoparticles (NPs) - by ammonia modification (GO-NH_2) and by hydroxylamine functionalization (haGO-NH_2) versus pristine GO, to induce toxicity in two cancer cell types – Colon 26 (colon cancer) and HepG2 (hepatic cancer), examining their effect on cell proliferation ability, cell membrane, ROS production, DNA damage and cell cycle arrest.

Experimental

A. NPs characterization: the nitrogen content of NPs was determinated by X-ray photoelectron spectroscopy (XPS) and particle size distribution and zeta potential (ZP) - by a Nano Zetasizer.

B. Cell culture: A colorectal cancer cell line, Colon26 and a hepatocellular cancer cell line, HepG2 were purchased from American Tissue Culture Collection (ATCC) and were maintained in DMEM medium supplemented with 10% fetal calf serum and antibiotic/antimycotic solution at 37°C in a fully humidified atmosphere at 5% CO_2.

C. Toxicity assessment of pristine and aminated GO NPs: Cells were seeded in a concentration of $2x10^4$ cell/ml in 24-well plates (for cytotoxicity assessment) and $1x10^5$ cell/ml ml in 6-well plates (for genotoxicity assessment) and incubated 24 and 48 hours before to be exposed to pristine (Graphenea, Spain) and aminated GO NPs with increasing concentrations

of 1, 10, 20 and 50 µg/ml. At the 24h the cells were stained with FDA to visualize cell morphology; LDH assay was performed to detect cell membrane damage. Cell Counting Kit (CCK-8, Sigma–Aldrich) was carried out to measure cell proliferation at the 48h of incubation with NPs. ROS production was measured at 24h of NPs exposure by DCFA-DA analysis. FACS and Comet assay was used to determine DNA damage and cell cycle arrest.

Resultats and Discussion

The results from XPS measurements reveal 3.47% nitrogen content in ammonia-modified GO-NH_2 versus 1.86% in hydroxylamine-functionalized haGO-NH_2 in addition to carbon, oxygen and traces of sulfur and zinc while in pristine GO were detected only carbon and oxygen. Zetasizer assay data showed that pristine GO NPs are a heterogeneous population including a main fracture with NPs size of 1,5-3.6 μm and a small percentage of NPs with size of 250-515 nm and ZP of -24.5 mV. Aminated GOs were more homogenous with particles mean size of 578 nm for GO-NH_2 and 594 nm for haGO-NH_2 as well as ZP of 38.5 mV and -12.28mV, respectively.

The results from the FDA staining showed that GO and GO-NH_2 NPs altered the morphological appearance of Colon 26 cells only at the highest concentrations of 50 μg/ml. In $HepG_2$ cells no visible alterations in cell morphology were observed regardless the concentrations of haGO-NH_2. Analysis of CCK data demonstrated a dose-dependent cytotoxic effect of aminated GO NPs in Colon 26 cells and HepG2 cells, much stronger in $HepG_2$ cells (appr. 2 times). The lower concentrations of pristine GO (4- and 10 μg/ml) have demonstrated a stimulating effect on HepG2 proliferation which could be referred to the "hormesis effect" but surprisingly, a stimulating effect on Colon26 cell proliferation was noticed after exposure of the cells to the highest concentrations of pristine GO NPs (20-and 50 μg/ml).

The results from evaluation of the ability of GO NPs to generate ROS after 24 hours exposure showed that in Colon 26 cells GO and GO-NH_2 generated statistically significant amounts of ROS compared to the control at all tested concentrations, however the ROS production induced from GO-NH_2 was slightly higher compared to GO, while in HepG2 cells all tested concentrations of haGO-NH_2 induced much higher production of ROS levels than those measured in non-treated cells and in GO-treated cells.

Analysis of LDH leakage revealed that only aminated GO NPs regardless of the amination method induced damage in the cell membrane integrity by which possibly induce cytotoxicity in tumor cells. In Colon26 cells certain genotoxicity of both types of GOs (at a concentration of 10 μg/ml for GO and 50 μg/ml of amGO-NH_2) was detected. Further, FACS analysis of the probable mechanisms of this observed genotoxicity led to detection of high level of apoptosis in all tested probes, regardless the treatment of cells with pristine or animated GO nanoparticles while necrosis was at a very low rate. DNA damage of HepG2 cells treated with pristine GO and haGO-NH_2 was very faint to almost insignificant after 24h exposure to the nanomaterials. This points out to the fact that the mechanisms through which haGO-NH_2 NPs exert their biological activities on HepG2 cells are not centred in maintenance of the stability of the genome an its integrity but rather elsewhere. The last made us probe the functions of mitochondria in these cells. Results unambiguously demonstrated compromised mitochondrial functions as a result of haGO-NH_2 treatment.

GO and ammonia modified GO-NH_2 were found to inhibit the proliferation of Colon 26 cells. A reduction in the percentage of cells in the G0-G1 phase of the cell cycle after incubation with GO-NH_2 was observed in these tumour cells. The last showed a concentration-dependent reduction in the number of cells in all cell cycle phases with the most pronounced reduction in the number of cells in the phase G0-G1 at the higher used GO-NH_2 concentrations. These results suggest strong cytotoxic and to some extent slight cytostatic effect of the amimated GO, especially at the concentration of 50 μg/ml. Similar effects were not observed in HepG2 cells treated with haGO-NH_2.

TABLE 1. SUMMARIZES SOME PHYSICO-CHEMICAL CHARACTERISTICS OF GO-NH_2 AND HAGO-NH_2 AND COMPARES SOME OF THE OBSERVED BIOLOGICAL EFFECTS OF THE TWO TYPE OF AMINATED GOS ON TWO DIFFERENT TUMOR MODELS

Sample	Mean size	ZP (mV) ± SE	Polarity	N1s TOT.(%)	Cell adhesion at 50µg/ml (%) HepG2	Cell adhesion at 50µg/ml (%) Colon26	IC_{50} HepG2	IC_{50} Colon26	Inhibition of cell proliferation (%) HepG2	Inhibition of cell proliferation (%) Colon26
GO (ref. 21)	250 ± 68nm 1.5 ± 0.7µm	−24,5 ± 0.4 mV	negative	0-0.99		42.6		1.71 µg/ml		78
(this study)	515 ± 50 nm 3,6 ± 0.5 µm	-33.7 ± 0.4 mV			63		62.97 µg/ml		17	
GO-NH_2 (ref. 21)	560 ± 300 nm	38.5± 2.8 mV	positive	3.47		22.5		1.26 µg/ml		78
haGO-NH_2 hydroxylamine modified (this study)	594 ± 270 nm	−12,28 ± 0.6 mV	negative	1.86	21		3.4 µg/ml		50	

Conclusions

GO-NH_2 and haGO-NH_2 have different physico-chemical properties as well as diverse modes of cyto- and genotoxicity in regard to the two types of tumour cells.

ACKNOWLEDGMENT

Funds from the Bulgarian Science Foundation – under Grants No. DNTS/ 01/6/2016 and KP06-N31/15.

REFERENCES

[1] G. Gonçalves, et al, "Nano-graphene oxide: a potential multifunctional platform for cancer therapy," Advanced Healthcare Materials, vol. 2, no. 8, pp. 1072–1090, 2013.

[2] M. Simsikova and T. Sikola, "Interaction of graphene oxide with proteins and applications of their conjugates," J of Nanomed Res., vol. 5, no. 2, 2017.

[3] M. Pelin, et al., "Graphene and graphene oxide induce ROS production in human HaCaT skin keratinocytes: the role of xanthine oxidase and NADH dehydrogenase, Nanoscale, vol. 10, no. 25, pp. 11820–11830, 2018.

[4] M. Xu, et al., "Improved in vitro and in vivo biocompatibility of graphene oxide through surface modification: poly(acrylic acid)-functionalization is superior to PEGylation, ACS Nano, vol. 10, no. 3, pp. 3267–3281, 2016.

DOI 10.18699/BGRS/SB-2020-412

The role of host genetics in severity of COVID-19

Ivan A. Kuznetsov
Center of Life Sciences
Skolkovo Institute of Science
and Technology
Moscow, Russia
Laboratory of Theoretical and Applied
Functional Genomics
Novosibirsk State University
Novosibirsk, Russia
Ivan.Kuznetsov@skoltech.ru

Tatiana I. Shashkova
Laboratory of Theoretical and Applied
Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Yurii S. Aulchenko

Laboratory of recombination
and segregation analysis
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia

***Abstract* — Coronavirus SARS-CoV-2 caused the pandemic of COVID-19, a respiratory disease with relatively high mortality. It is important to understand the extent to which health-related factors influence on the risk of severe COVID-19 symptoms development. It can help the health care system to save more lives, and governments to take reasonable actions. We aimed to estimate the associations between the severity of COVID-19 infection and previously suggested health-related risk factors using the methods of human quantitative genetics. First, we collected summary statistics of one GWAS for COVID-19 severity and of 6 GWASes for previously suggested risk factors. We unified summary statistics from these analyses, performed quality control, and imported the data to the GWAS-MAP platform. Then, using these summary statistics we performed Mendelian randomization analysis and obtained the estimates of associations between COVID-19 severity and several previously suggested risk factors. Although our estimates of Odds Ratio are directionally consistent with previous estimates obtained from retrospective epidemiological studies, they are not significant.**

Keywords* — *COVID-19, risk factors, Mendelian randomization

INTRODUCTION

Coronavirus SARS-CoV-2 caused a pandemic in the beginning of 2020. COVID-19 is a disease related to this virus. It can cause severe pneumonia and respiratory failure, which can lead to death. While the majority of people have the disease in a mild or asymptomatic form, several percent of the infected population have a critical course of the disease. From studies of other viral infections, we know that the severity of an infectious disease can depend on four groups of factors: mode of transmission or infectious dose, virus genotype, host genotype, host immune status. Specific retrospective epidemiological COVID-19 studies demonstrate that the course of the disease heavily depends on age and other health-related factors such as the presence of chronic diseases or lifestyle. In the current situation retrospective observational study is the only available design to directly estimate the effects of suggested health risk factors. Such studies are not able to distinguish the separate effects of different factors.

Our aim was to estimate the associations between the severity of COVID-19 infection and previously suggested health-related risk factors. To achieve this aim we used the methods of human quantitative genetics.

METHODS

In our work we use Mendelian randomization to estimate Odds Ratio for associations between a COVID-related phenotype and suggested health risk factors of severe course of the disease. The outcome we considered was hospitalization with respiratory failure and SARS-CoV-2 viral RNA confirmed with PCR test. We used the results of the study conducted on Italian and Spanish patients by Ellinghaus et al. [1]. We used summary statistics adjusted for sex, age and 10 principal.

The previously suggested health risk factors of severe COVID-19 we considered in our study were body mass index, coronary artery disease, hypertension, diabetes, asthma and smoking. Sources of summary statistics were powerful genome-wide association studies conducted on dozens and hundreds of thousands of individuals. For each exposure a set of nearly independent SNPs (instrumental variables) was selected.

In our study we used the GWAS-MAP platform that allows us to store and analyse genome-wide association study summary statistics. Results of high-quality GWASes for the exposures had already been imported to the database. We uploaded COVID-19 summary statistics to the platform according to the protocol of GWAS-MAP.

RESULTS

We applied the Mendelian randomization method to unified and validated summary statistics to estimate Odds Ratio for associations between six previously suggested health risk factors and COVID-19 severity. All estimates were higher than one, which is consistent with the results of most retrospective epidemiological studies. However, these estimates were not significantly different from one that doesn't allow us to make a strong claim about the actual effects of the considered factors on the COVID-19 severity.

CONCLUSION

We obtained independent estimates of Odds Ratio for six for previously suggested risk factors. The estimates are directionally consistent with previous estimates obtained from retrospective epidemiological studies. However, they are not significantly different from one.

REFERENCES

[1] D. Ellinghaus, et al., "Genomewide Association Study of Severe Covid-19 with Respiratory Failure," *N. Engl. J. Med.*, 2020.

DOI 10.18699/BGRS/SB-2020-290

Chemotherapeutic agent cisplatin suppresses stabilin-1-mediated clearance of EGF by tumor-associated macrophage

Irina Larionova
National Research Tomsk State University,
Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences,
Tomsk, Russia
larionova0903irina@mail.ru

Elena Kazakova
National Research Tomsk State University
Tomsk, Russia
kazakova.e.o@mail.ru

Artyem Kiselev
Almazov National Medical Research Centre,
Saint Petersburg, Russia
artem.kiselyov@gmail.com

Marina Patysheva
National Research Tomsk State University,
Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences,
Tomsk, Russia
starin5@ya.ru

Julia Kzhyshkowska
National Research Tomsk State University,
Tomsk, Russia,
German Red Cross Blood Service Baden-Württemberg – Hessen,
Mannheim, Germany
julia.kzhyshkowska@googlemail.com

***Abstract* — In this study we have demonstrated the suppression of receptor-mediated clearance of tumor-supportive factors EGF by cisplatin in tumor-associated macrophages. These results may be useful for the determining of the mechanisms of the interaction of TAMs and chemotherapy and the mechanism of macrophage-mediated tumor chemoresistance.**

Keywords — tumor-associated macrophages, cisplatin, EGF, stabilin-1.

Motivation and aim

Tumor is a complex system of transformed cells interacting with the surrounding heterogeneous cellular and molecular microenvironment. Tumor microenvironment affects key activities of cancer cells and contributes to the tumor response to chemotherapy. Tumor-associated macrophages (TAMs) are a major component of innate immunity supporting primary tumor growth and metastasis [1]. Numerous evidence indicate that TAMs can accumulate in tumors after chemotherapy and contribute to chemoresistance [2]. In this way, it is important to investigate how the functional program of TAMs changes under chemotherapy treatment, and what are the possible mechanisms of the modulating effect of TAMs on the treatment effectiveness. Earlier we have shown that cisplatin may affect the transcriptional program of TAMs, and that in tumors of different localization the effect of the CT agent on the macrophage functions may be different. The aim of this study was to identified the thinner mechanisms of the interaction of TAMs and cisplatin.

Methods

For the endocytic uptake of EGF we used flow cytometry analysis. Confocal microscopy was used for the analysis of stabilin-1-mediated internalization and endocytic trafficking of EGF in CHO cells and in modeled TAMs differentiated in the presence of conditioned supernatants of breast cancer (MCF-7) and colorectal cancer (Colo206F) cell lines. We performed next-generation sequencing (NGS) of RNA samples obtained from our modeled TAMs. Validation of sequencing data by real-time PCR was performed for selected genes implicated in the endocytic uptake: DNM3, STX8, DENND1A and EHD1.

RESULTS

For the first time we have demonstrated that stabilin-1 ectopically expressed in CHO cells mediates endocytic uptake of EGF, key growth factor stimulating progression of breast and colorectal cancer. In the model of primary human TAMs, we have demonstrated that cisplatin decreases stabilin-1-mediated internalization and endocytic trafficking of EGF, without significant change in gene expression of scavenger receptor stabilin-1. Molecular mechanisms of cisplatin effects on TAMs were analysed using high throughput RNA sequencing. Gene set enrichment analysis identified that cisplatin contributes to defects in endocyting machinery reducing membrane biogenesis and vesicular transport. Significant suppression of DNM3, STX8, DENND1A and EHD1 genes expression by cisplatin was confirmed by RT-PCR. Suppression of receptor-mediated clearance of tumor-supportive factors, such as EGF, by chemotherapeutic drugs may enhance tumor-supporting effect of TAMs creating the microenvironment supporting tumor chemoresistance or relapse after chemotherapy course.

CONCLUSION

We suggested that suppression of receptor-mediated clearance of tumor-supportive factors by chemotherapeutic drugs may lead to the imbalance in tumor microenvironment and as a result to the chemoresistance.

ACKNOWLEDGMENT

Supported by the grant RSF No. 19-15-00151.

REFERENCES

[1] C.E. Lewis, and J.W. Pollard, Cancer Res, vol. 66, pp. 605-612, January 2006.

[2] B.Z. Qian, and J.W. Pollard, Cell, vol. 141, pp. 39-51, April 2010.

DOI 10.18699/BGRS/SB-2020-291

Mass spectrometry, Ribo-seq, and RNA-seq integration reveals hormone-dependent translation of short open reading frames in human long non-coding RNAs

Leonard Lipovich, Ph.D.
Professor, Cener for Molecular
Medicine and Genetic,s
Wayne State University
Detroit, Michigan,
United States of America
LLipovich@med.wayne.edu

Abstract — **More than two-thirds of human genes do not encode proteins. Long noncoding RNA (lncRNA) genes are the most abundant class of non-protein-coding genes in humans, and their pivotal importance in both normal cellular functions and disease is becoming increasingly appreciated. These functions, rooted in RNA-driven biology independent of protein translation, have been amply documented. LncRNA genes are typically defined as those with all transcripts lacking any predicted Open Reading Frames (ORFs) over 100 amino acids in length and/or with homologies to known proteins. However, lncRNAs are often found in the cytoplasm, where, in principle, their short nonconserved ORFs - if any - may be translated by ribosomes, in a stochastic or perhaps a regulated fashion. Historically, two methods have been used to test whether lncRNAs are endogenously translated: direct discovery of lncRNA ORFs in protein mass spectrometry data (the original method; expensive, less common) and ribosome profiling (Ribo-Seq) (a newer approach, indirect, now commonly used). In the ENCODE (Encyclopedia of DNA Elements) Consortium, we were the first group to empirically test for lncRNA translation in human cells using mass spectrometry, finding that most lncRNAs are not translated. We have now deployed the two methods concurrently in a key nuclear hormone receptor pathway cancer model: human estrogen receptor positive MCF7 breast cancer cells before and after estrogen stimulation. Combining ribosome profiling with mass spectrometry, we identified four lncRNAs with short open reading frames that appear to be erroneously and persistently translated into unexpected peptides. One of these lncRNAs exhibits systematic in-frame mis-translation of multiple stop codons into amino acids, demonstrating an apparent gene-specific violation of the human genetic code. We also identified numerous estrogeninduced and estrogen-repressed lncRNAs jointly supported by Ribo-Seq and mass spectrometry, as well as estrogenresponsive differences in ribosome binding and translation that are distinct from estrogen-regulated transcriptional programs. Funding: National Institutes of Health – Director's New Innovator Award 1DP2-CA196375 (2014-2019) to L.L.**

Keywords — long non-coding RNA, lncRNA, mass spectrometry, proteomics, cancer, human

Funding: National Institutes of Health – Director's New Innovator Award 1DP2-CA196375 (2014-2019) to L.L.

DOI 10.18699/BGRS/SB-2020-292

The effect of lithium carbonate on angiogenesis of hepatocellular carcinoma-29

Viktoriia Makarova
Laboratory of ultrastructural research, Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
shedina_vika@mail.ru

Nataliya Bgatova
Laboratory of ultrastructural research, Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
n_bgatova@ngs.ru

Iuliia Taskaeva
Laboratory of ultrastructural research, Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
inabrite@yandex.ru

***Abstract* — Hepatocellular carcinoma is one of the most aggressive human tumors with a high prevalence and mortality rate. At the same time, there is a problem of HCC resistance to standard therapy. The search for drugs that block the growth and metastasis of the tumor is relevant. There is evidence of antitumor properties of lithium carbonate. The aim of our study was to determine the effect of lithium on the development and structure of blood vessels of experimental hepatocellular carcinoma-29 (G-29). The study was performed on male CBA mice. Hepatocarcinoma-29 cells were used to induce the tumor process. The expression of the blood vessel endothelial marker (CD-31) and the ultrastructural organization of tumor endothelial cell vessels were studied. It was shown that lithium carbonate decrease volume density of blood vessels G-29. Ultrastructural analysis of the tumor vessels revealed the atypical nature of their structure. The walls of the vessels were lined with both endothelial and tumor cells.**

Keyword s— hepatocellular carcinoma, vasculogenic mimicry, angiogenesis, lithium

INTRODUCTION

Hepatocellular carcinoma is one of the most aggressive tumors with high mortality [1, 2]. Neoangiogenesis is known to contribute to cancer progression, relapse, and metastasis. But most modern drugs aimed at inhibiting angiogenesis are ineffective [3]. Recent studies have shown that in highly aggressive tumors not only endothelial cells can be used to create vascular structures, but also alternative ways of forming the vasculature, in which vascular walls are formed by tumor cells. This is the so-called process of vasculogenic mimicry [4]. Therefore, it remains relevant to study the mechanisms of blocking vascular structures that provide trophic support for tumor growth and ways of its metastasis. There are various ways of influencing lithium on many intracellular processes, including in tumor cells [5, 6]. For example, by inhibiting regulatory GSK3β, induce apoptosis of cancer cells and thereby inhibit tumor growth [7]. Based on this, it is proposed to use lithium as an antitumor agent. The aim of our study was to study the effect of lithium carbonate on the development and structure of the vessels of experimental hepatocellular carcinoma-29.

METHODS AND ALGORITHMS

The study was conducted on male mice of the CBA line. Work with animals was carried out in accordance with the "Rules for the work using experimental animals." G-29 cells were used to induce the tumor process. GK-29 was obtained and verified by the Institute of Cytology and Genetics SB RAS. A suspension of GK-29 cells was injected into 0.1 ml of intact animals into the muscle of the right thigh (3). The animals were divided into 3 groups: I - intact animals, 2 - animals with tumor growth, 3 - animals with GK-29, treated with lithium carbonate as a treatment. The material for research was collected on the 30th day of the experiment. Animals were removed from the experiment under ether anesthesia using the cranio-cervical dislocation method. Tumor fragments were fixed in formalin and processed according to standard methods to obtain paraffin blocks. Paraffin sections were dewaxed, rehydrated and washed with distilled water. Antibodies anti-CD31antibody (Abcam, UK) were used to determine vascular endothelial cells. Histological preparations were stained with Mayer hematoxylin. Evaluation of the results was carried out using a light microscope (LEICA DME, Germany), with a 400-fold increase. Photos were obtained using the Avigion computer program. For electron microscopic examination, tumor tissue samples were fixed in a 1% solution of OsO4 in phosphate buffer (pH = 7.4), dehydrated in increasing concentration of ethanol, and enclosed in epon (Serva, Germany). Semi-thin sections with a thickness of 1 μm were stained with toluidine blue and studied under a light microscope (LEICA DME, Germany). Ultrathin sections with a thickness of 70-100 nm were contrasted with a saturated aqueous solution of uranyl acetate and lead citrate and studied using a JEM 1010 electron microscope (JEOL, Japan). Morphometric analysis was performed using Image J software (Wayne Rasband, USA).

RESULTS

Histological analysis of tumor tissue revealed a high density of vascular structures with narrow lumens. It was found that the volume density of blood vessels stained on a blood vessel marker CD-31 was significantly lower than the total volume density of all vascular structures. Ultrastructural analysis showed that part of the walls of such vessels is lined not only with endothelial cells, but also with G-29 cells. It was difficult to differentiate such vessels since they had a structure similar to both blood and lymph vessels. This phenomenon is defined as vasculogenic mimicry, which is widespread in aggressive tumors and is a poor prognostic indicator of patient survival [8, 9].

In the group of animals that were injected with lithium carbonate, a significant decrease in the volume density of vascular structures was noted. At the same time, the density of CD-31-positive vessels decreased to a greater extent.

Apparently, lithium carbonate blocks the angiogenesis of true blood vessels.

CONCLUSION

The data obtained show that experimental hepatocarcinoma-29 has good vascularization. Neoangiogenesis in tumor tissue follows the path of vasculogenous mimicry, which greatly complicates the standard antitumor therapy. Lithium carbonate is a promising means of blocking tumor growth and metastasis.

REFERENCES

[1] Y. Shen, D. Cao, "Hepatocellular carcinoma stem cells: origins and roles in hepatocarcinogenesis and disease progression," Biosci., vol. 1(4), pp. 1157–1169, January 2012.

[2] B. Qu, G. Sheng, L Guo, F. Yu, G. Chen, Q. Lu, R. Wang, B. Han, Y. Lu, "MIG7 is involved in vasculogenic mimicry formation rendering invasion and metastasis in hepatocellular carcinoma," Oncol Rep., vol. 39(2), pp. 679-686, February 2018.

[3] Y. Zeng, X. Yao, X. Liu, X. He, L. Li, X. Liu, Z. Yan, J. Wu, B.M. Fu, "Anti-angiogenesis triggers exosomes release from endothelial cells to promote tumor vasculogenesis," J Extracell Vesicles., vol. 8(1), pp. 1629865, January 2019.

[4] J. Zhang, L. Qiao, N. Liang, J. Xie, H. Luo, G. Deng, J. Zhang, "Vasculogenic mimicry and tumor metastasis," J BUON., vol. 21(3), pp. 533-541, May-June 2016.

[5] Iu. Taskaeva, N. Bgatova, I. Gogaeva, "Lithium effects on vesicular trafficking in hepatocellular carcinoma cells," Ultrastructural Pathology, vol. 43(6), pp. 301-311, 2019.

[6] Y.S. Taskaeva, N.P Bgatova, "Cytological characteristics of a heterogeneous population of hepatocellular carcinoma-29 cells after injection of lithium carbonate in the experiment," Bulletin of Experimental Biology and Medicine, vol. 167(6), pp. 779-783, 2019.

[7] J.S. Wang, C.L. Wang, J.F. Wen, Y.J. Wang, Y.B. Hu, H.Z. Ren, "Lithium inhibits proliferation of human esophageal cancer cell line Eca-109 by inducing a G2/M cell cycle arrest", World journal of gastroenterology, vol. 14, pp. 3982–3989, July 2008.

[8] M.C.M. Prado, S.A.L. Macedo, G.G. Guiraldelli, P. de Faria Lainetti, A.F. Leis-Filho, P.E. Kobayashi, R. Laufer-Amorim, C.E. Fonseca-Alves, "Investigation of the Prognostic Significance of Vasculogenic Mimicry and Its Inhibition by Sorafenib in Canine Mammary Gland Tumors," Front Oncol., vol. 19(9), pp. 1445, December 2019.

[9] V.V. Makarova, N.P. Bgatova, Yu.S. Taskaeva, Yu.I., "Borodin Angiogenesis in experimental hepatocarcinoma-29," The 22nd International Charles Heidelberger Symposium on Cancer Research Proceedings of the International Symposium, pp. 67-68, 2018.

DOI 10.18699/BGRS/SB-2020-293

Generation of the panel of transgenic human cell lines with stable expression of mutant variants of the *GJB2* gene associated with hearing loss for comparative *in vitro* studies

Ekaterina Alexandrovna Maslova
Federal Research Center Institute of Cytology and Genetics, SB RAS
Novosibirsk State University
Novosibirsk, Russia
maslova@bionet.nsc.ru

Marina Vyacheslavovna Zytsar
Federal Research Center Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
zytzar@bionet.nsc.ru

Valeriia Yuryevna Danilchenko
Federal Research Center Institute of Cytology and Genetics, SB RAS
Novosibirsk, Russia
danilchenko_valeri@mail.ru

Konstantin Evgenyevich Orishchenko
Federal Research Center Institute of Cytology and Genetics, SB RAS
Novosibirsk State University
Novosibirsk, Russia
OrishchenkoKE@icg.sbras.ru

Olga Leonidovna Posukh
Federal Research Center Institute of Cytology and Genetics, SB RAS
Novosibirsk State University
Novosibirsk, Russia
posukh@bionet.nsc.ru

***Abstract* — Assessment of pathogenicity of novel variants is a primary task for molecular diagnostics of hereditary diseases, and comprehensive evidences, including functional *in vitro* studies, are necessary to support their involving in pathology. Mutations in the *GJB2* gene encoding transmembrane protein Cx26 are the common cause for hearing loss worldwide. About 400 pathogenic and also not yet classified variants with ambiguous functional significance are described in the *GJB2* gene sequence. In this study, for the first time, the panel of transgenic HeLa cell lines stably expressing of different *GJB2* variants was generated for subsequent comparative analysis of structure and functioning of the Cx26 protein. This panel was successfully applied for confirmation of the pathogenicity of novel *GJB2* missense variant c.516G>C (p.Trp172Cys) found with high frequency in deaf patients from Southern Siberia.**

Keywords* — *hereditary deafness, GJB2 (Cx26), transgenic human cell lines

Introduction

Assessment of pathogenicity of novel variants in genes involved in the pathogenesis of monogenic diseases plays a crucial role for medical genetics. To establish the pathogenic significance of such variants, they must meet following criteria: obvious segregation of variant with pathology; a significant difference of its frequency in patient's samples compared to the controls; rarity or absence in world human genomic databases; multiple computational predictions of deleterious effects of variant, and, finally, confirmation of its pathogenic effects in functional studies [1].

The most common cause for hearing loss in human populations is mutations in the *GJB2* gene (gap junction protein, beta-2, MIM 121011, 13q12.11) which encodes a transmembrane protein connexin 26 (Cx26) [2]. Six Cx26 molecules associate to form transmembrane hexameric hemichannels (connexons) which dock with connexons of adjacent cells forming gap junctions that are essential for the transport of ions and other low-molecular-weight components between cells. Cx26 plays a crucial role in potassium homeostasis in the inner ear and disruption of the Cx26 channels leads to irreversible deafness [3]. To date, about 400 pathogenic (recessive and dominant mutations leading to deafness) and also not yet classified variants with ambiguous functional significance are described in the *GJB2* gene sequence (Human Gene Mutation Database: www.hgmd.cf.ac.uk). Truncating *GJB2* mutations (nonsense, splice and frameshift mutations) result in defective protein synthesis. Nontruncating mutations (missense mutations, in-frame deletions and insertions) can affect Cx26 function in various ways depending on the position and type of amino acid substitution. Many extensive *in vitro* studies using different expression models were performed for elucidate the functional characteristics of several dozen distinct disease-causing Cx26 mutants. However, to our knowledge, there are no model cell lines with unified experimental design having stable expression of mutant *GJB2* variants that could be used for precise comparative analysis of mutant Cx26 molecules.

This study aims to generate the panel of transgenic human cell lines with stable expression of mutant variants of the *GJB2* gene associated with hearing loss and to perform the complex assessment of pathogenicity of novel *GJB2* variant c.516G>C (p.Trp172Cys) found in association with deafness in indigenous peoples (Tuvinians and Altaians) of Southern Siberia (Russia).

Materials and Methods

The *GJB2* knockout HeLa cell line was obtained using CRISPR/Cas9 genome editing system based on plasmid pX330-U6-Chimeric_BB-CBh-hSpCas9 which was a gift from Feng Zhang (Addgene plasmid #42230; http://n2t.net/addgene:42230; RRID:Addgene_42230). Guide RNAs (gRNAs) located to the right and left from the *GJB2* coding region (exon 2) was selected via software platform Benchling (https://benchling.com).

To assemble genetic constructs carrying different variants of the *GJB2* coding region we used plasmid vector pSBbi-GP which was a gift from Eric Kowarz (Addgene plasmid #60511; http://n2t.net/addgene:60511; RRID:Addgene_60511). Then, based on *GJB2* knockout HeLa, we generated several transgenic cell lines with stable

expression of the *GJB2* variants using the Sleeping Beauty transposase system [4].

Complex assessment according to [1] was applied for the establishment of the pathogenicity of novel *GJB2* missence variant c.516G>C (p.Trp172Cys). Genetic analysis was performed in the group of deaf patients (182 Tuvinians, 74 Altaians) and in the ethnically matched control samples (157 Tuvinians, 218 Altanians). Two-tailed Fisher's exact test with significance level of $p<0.05$ was applied to compare the allele frequencies between patients and controls. Several prediction tools (PolyPhen2, PROVEAN, PANTHER, MutationTaster, FATHMM etc.) were applied for *in silico* analysis. We also carried out comparative functional *in vitro* study of c.516G>C (p.Trp172Cys) and other *GJB2* variants. Cellular localization of different forms of protein Cx26 was investigated by immunocytochemistry (ICC) using primary antibodies against the C-terminus of Cx26 (#51-2800, Thermo Fisher Scientific) and secondary antibodies conjugated to Alexa Fluor Plus 555 (#A32732, Thermo Fisher Scientific). Permeability of the Cx26 channels was analyzed by dye (PI, propidium iodide) loading assay [5].

RESULTS

Generation of transgenic human cell lines with stable expression of mutant GJB2 variants

For comparative *in vitro* analysis of functional effects of different variants in the *GJB2* gene which the most common cause for hearing loss in human populations, we generated, on the basis of HeLa cell line, the panel of transgenic human cell lines with stable expression of several *GJB2* variants. Current data on the level of endogenous expression of the *GJB2* gene in HeLa cell line are ambiguous. To avoid any endogenous *GJB2* expression, we primarily obtained the *GJB2* knockout HeLa cell line for subsequent establishment of transgenic cell lines. Seven transgenic cell lines carrying different *GJB2* variants were generated: 1) novel missense variant c.516G>C (Cx26-p.Trp172Cys) with yet unclear functional significance; 2-4) obvious pathogenic deletions c.35delG (Cx26-p.Gly12Valfs), c.235delC (Cx26-p.Leu79Cysfs), c.313_326del14 (Cx26-p.Lys105Glyfs); 5) known dominant mutation c.224G>A (Cx26-p.Arg75Gln); 6) *cis*-configuration of two variants c.79G>A (p.Val27Ile) and c.341A>G (p.Val27Ile) of unproven pathogenicity (Cx26-p.[Val27Ile;Glu114Gly]); and 7) wild type of the *GJB2* coding region (Cx26-wt).

Complex assessment of functional significance of novel GJB2 missence variant c.516G>C (p.Trp172Cys)

Novel nonsynonymous missence variant c.516G>C (p.Trp172Cys) in the *GJB2* gene was found with high frequency in deaf patients belonging to indigenous peoples of Southern Siberia (Tuvinians and Altaians) [6]. Segregation of a homozygous or compound heterozygous c.516G>C (p.Trp172Cys) with hearing loss was established in analysis of pedigrees of deaf patients. Frequency of c.516G>C (p.Trp172Cys) was significantly higher in the patient's groups than in the controls ($p<0.05$). Moreover, c.516G>C (p.Trp172Cys) was not found in the world human genomic databases (ClinVar, dbSNP138, 1000 Genomes Project, Genome Aggregation Database) and most of *in silico* programs predicted a likely damaging effect of p.Trp172Cys. Pathogenicity of c.516G>C (p.Trp172Cys) was also evidenced by our *in vitro* assessment. ICC showed predominantly cytoplasmic localization of mutant protein Cx26-p.Trp172Cys in contrast to Cx26-wt and mutants Cx26-p.Arg75Gln and Cx26-p.[Val27Ile;Glu114Gly] which represented distinct conglomerates on cell membrane as described in literature [5, 7]. Dye (PI) loading assay revealed lower PI loading efficiencies in cells expressing mutant Cx26-p.Trp172Cys compared to Cx26-wt while the absence of PI loading was shown for other mutant Cx26 variants with known pathogenic effects (Cx26-p.Arg75Gln, Cx26-p.Gly12Valfs, Cx26-p.Leu79Cysfs, Cx26-p.Lys105Glyfs). Thus, the pathogenicity of novel *GJB2* missense variant c.516G>C (p.Trp172Cys) was proven by several lines of evidences including comparative *in vitro* analysis.

Conclusion

For the first time, the panel of transgenic HeLa cell lines stably expressing of different *GJB2* variants was generated for subsequent analysis of structure and functioning of Cx26 protein. The pathogenicity of novel *GJB2* missense variant c.516G>C (p.Trp172Cys) was proven by several lines of evidences including comparative *in vitro* analysis with using constructed panel of transgenic HeLa cell lines.

ACKNOWLEDGMENTS

Study was supported by the Project #0324-2019-0041-C-01, the RFBR grant #17-29-06016_ofi-m (molecular genetic studies) and by the State Budget Project of the Laboratory of structural and functional organization of genome, NSU) (cell engineering).

REFERENCES

[1] S. Richards, N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015, vol. 17(5) pp. 405–424.

[2] D. K. Chan, K. W. Chang. *GJB2*-associated hearing loss: systematic review of worldwide prevalence, genotype, and auditory phenotype. Laryngoscope. 2014, vol. 124(2) pp. E34–E53.

[3] R. Bruzzone, T. W. White, D. L. Paul. Connections with connexins: the molecular basis of direct intercellular signaling. Eur J Biochem. 1996, vol. 238(1) pp. 1–27.

[4] E. Kowarz, D. Löscher, R. Marschalek. Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. Biotechnol J. 2015, vol. 10(4) pp. 647-653.

[5] S. Y. Choi, K. Y. Lee, H. J. Kim, H. K. Kim, Q. Chang, H. J. Park et al. Functional evaluation of *GJB2* variants in nonsyndromic hearing loss. Mol Med. 2011, vol. 17(5-6) pp. 550–556.

[6] O. L. Posukh, M. V. Zytsar, M. S. Bady-Khoo, V. Y. Danilchenko, E. A. Maslova, N. A. Barashkov et al. Unique mutational spectrum of the *GJB2* gene and its pathogenic contribution to deafness in Tuvinians (Southern Siberia, Russia): a high prevalence of rare variant c.516G>C (p.Trp172Cys). Genes (Basel). 2019, vol. 10(6) p. 429.

[7] S. W. Yum, J. Zhang, S. S. Scherer. Dominant connexin26 mutants associated with human hearing loss have trans-dominant effects on connexin30. Neurobiol Dis. 2010; vol. 38(2) pp. 226–236.

DOI 10.18699/BGRS/SB-2020-294

Markers of oxidative and atherogenic processes in individuals with hand-arm vibration syndrome and metabolic disorders

Liudmila Masnavieva
ESIMER,
Angarsk, Russia
Masnavieva_Luda@mail.ru

Irina Kudaeva
ESIMER,
Angarsk, Russia
Kudaeva_Irina@mail.ru

Nadezhda Chistova
ESIMER,
Angarsk, Russia

Abstract — **Changes in the activity of lipid peroxidation processes are observed when exposed to vibration, as well as metabolic syndrome and type 2 diabetes, which are characterized by metabolic disorders. This study is devoted to the assessment of atherogenic and oxidative processes in individuals with hand-arm vibration syndrome in combination with metabolic disorders. We examined the levels of cholesterol, cholesterol in high and low density lipoproteins, oxidized low density lipoproteins, antibodies to them and the thiol status in 59 men with hand-arm vibration syndrome, in 73 people with vibrational disease in combination with metabolic syndrome and in 35 individuals with hand-arm vibration syndrome and type 2 diabetes. It was found that the average value of total cholesterol was higher than the reference level in all groups. A decrease in cholesterol in high density lipoproteins and an increase in cholesterol in low density lipoproteins have been observed in individuals with metabolic disorders. Levels of oxidized low-density lipoproteins and antibodies to them were increased in all groups. Correlation existed between the level of total cholesterol and cholesterol in low density lipoprotein, between the content of oxidized low density lipoproteins and total cholesterol, between the concentration of oxidized low density lipoproteins and cholesterol in low density lipoprotein in all groups. The correlation between the cholesterol level in high lipoproteins and thiol status was disrupted in patients with hand-arm vibration syndrome and metabolic disorders.**

Keywords — lipids, oxidized lipoproteins, antibodies to oxidized lipoproteins, hand-arm vibration syndrome, metabolic disorders

Motivation and Aim

Vibration pathology occupies one of the leading positions in the structure of occupational morbidity. A complex of regulatory disorders, changes in oxidative metabolism and oxidative stress develop during hand-arm vibration syndrome (HAVS) and play an important role in its pathogenesis. [1]. Diabetes mellitus (DM) is one of the most common noncommunicable diseases; the number of patients is increasing annually both in Russia and around the world [2]. A complex of disorders in lipid, carbohydrate and other metabolism, which is called metabolic syndrome (MS), appears during the formation of diabetes. It is known that metabolic changes and an increase in the activity of peroxide processes occur in type 2 diabetes and in MS. Lipoproteins and their oxidized forms play an important role in atherogenic processes and in the development of vascular pathology. [3].

This study is devoted to the assessment of oxidative and atherogenic processes in individuals with hand-arm vibration syndrome in combination with metabolic disorders.

Methods

In total, we examined 167 men with HAVS aged 40-66 years. Group I consisted of 59 patients with HAVS, group II - of 73 patients with HAVS in combination with MS, group III included 35 individuals with HAVS and type 2 diabetes. The level of total cholesterol and cholesterol in high density lipoproteins (HDL cholesterol) was determined spectrophotometrically. The content of cholesterol in low density lipoproteins (LDL cholesterol) was calculated using the Friedwald formula. We examined the thiol status, levels of oxidized lipoproteins and antibodies to them (oxLDL and antibodies to oxLDL, respectively) in blood serum using enzyme-linked immunosorbent assay. Statistical processing of the results was performed using nonparametric tests (Kruskell-Wallis, Mann-Whitney, Spearman rank correlation). The research results are presented in the form of a median and interquartile range.

Results

When analyzing the lipid profile, it was found that in group II the values of total cholesterol were slightly higher than in groups I and III ($p = 0.038$ and $p = 0.079$, respectively) (Table 1). The average values of this indicator were higher than the upper reference limit (5.17 mM/L) in all groups. HDL cholesterol levels in the group of individuals with HAVS without metabolic disorders were higher than in groups II and III ($p < 0.001$ and $p < 0.001$). The concentration of LDL cholesterol in patients with HAVS and MS was higher than in individuals of group I ($p = 0.006$) and tended to differ compared to group III ($p = 0.026$). The proportion of individuals with increased LDL cholesterol was 23.3% in the group of patients with HAVS without metabolic disorders, the frequency of increased levels of this indicator in groups II and III was 38.3% ($p = 0.06$) and 30.3% ($p = 0.46$), respectively. The content of oxLDL and the levels of antibodies to them in the blood serum of the subjects did not statistically significantly differ. It should be noted that in all groups the levels of oxLDL and antibodies to oxLDL were elevated (reference values were 117 IU / L and 30 IU / ml, respectively). This may be due to the binding of antibodies to oxLDL, which is aimed at neutralizing the pathological effect of the last lipoproteins on the vascular wall. The redox status, which was assessed by the level of thiol groups, depending on the presence of metabolic disorders, was not established, it did not differ in the groups. We conducted an analysis of the correlation relationships below.

TABLE I. Parameters of lipid and thiol status in individuals with hand-arm vibration syndrome and metabolic disorders

Indicator	Group			
	I	*II*	*III*	*p*
Total cholesterol, мМ/л	5.3 (4.6-5.9)	5.6 (5.0-6.1)	5.5 (4.7-5.8)	0.065
HDL cholesterol, мМ/л	1.3 (1.1-1.6)	1.0 (0.8-1.4)*I	0.9 (0.8-1.2)*I	0.000
LDL cholesterol, мМ/л	3.2 (2.3-3.8)	3.5 (3.2-4.2)*I	3.2 (2.7-3.9)#II	0.009
Thiol status, мкМ/мл	471.2 (418.3-517.4)	457.6 (392.5-521.5)	425.05 (392.5-482.1)	0.596
oxLDL мЕ/л	131.1 (104.0-168.5)	135.0 (111.3-160.3)	131.7 (87.7-152.7)	0.474
Antibodies to oxLDL, мЕ/мл	41.9 (32.1-50.8)	45.4 (39.0-78.8)	41.1 (28.2-55.6)	0.138

Notes: p – is the level of statistical significance of the differences according to the Kruskell-Wallis test; * I - differences are statistically significant compared with group I, p <0,016; #II - tendency to differences compared with group II, 0.016 <p <0.033.

It was found that the correlation was between the level of total cholesterol and LDL cholesterol (R = 0.77, R = 0.84 and R = 0.81; p <0.05 for groups I-III, respectively), the content of oxLDL and total cholesterol (R = 0.50, R = 0.47 and R = 0.67, p <0.05 for groups I, II and III), the concentration of oxLDL and LDL cholesterol (R = 0.68, R = 0.40 and R = 0.83, p <0.05 for groups I-III, respectively) in all groups. We found a negative relationship between the level of HDL cholesterol and thiol status (R = -0.63; p <0.05) in the group of individuals without metabolic disorders. This correlation was absent in groups II and III. This may indicate changes in the antioxidant defense system. It should be noted that weak negative associations between the content of oxLDL and antibodies to oxLDL (R = -0.33; p <0.05) were detected in group II. Increased levels of LDL cholesterol, oxLDL, and the relationship of these indicators may indicate the development of atherogenic processes and increased peroxidation, since oxidized modifications of LDL have proatherogenic properties [4]. The oxLDL can act as autoantigens [5]. A negative correlation between oxLDL and antibodies to them may be due to the fact that a high content of antibodies regulates the level of antigens and can block the absorption of oxLDL by macrophages [6].

Thus, the levels of total cholesterol, oxLDL, and antibodies to oxLDL are increased in individuals with HAVS both with and without metabolic disorders. HDL cholesterol is reduced and LDL cholesterol is elevated in patients with HAVS and metabolic disorders The correlation between the level of HDL cholesterol and thiol status in patients with HAVS and metabolic disorders is disrupted.

REFERENCES

[1] H. Sakakibara, S. Maeda, Y. Yonekawa Thermotactile threshold testing for the evaluation of sensory nerve function in vibration-exposed patients and workers. Int. Arch. Occup. Environ. Health. 2002. vol. 75, N 1-2, pp. 90-96*)*

[2] E.Z. Golukhova, E.V. Kuznetsova Myocardial revascularization in patients with type 2 diabetes mellitus: An overview of modern techniques. Diabetes mellitus. 2016. vol. 19, N 5, pp. 406-413. (references)

[3] S. Ayilavarapu, Diabetes-Induced Oxidative Stress Is Mediated by Ca+-Independent Phospholipase A2 in Neutrophils. J. Immunol. 2010. .vol. 184, pp. 1507-1515.

[4] J.A. Berliner, G. Subbanagounder, N. Leitinger, A.D. Watson, D.Vora. Evidence for a role of phospholipid oxidation products in atherogenesis. Trends Cardiovasc Med. 2001. Vol. 11, N 3-4, pp.142-147.

[5] S. Tsimikas, W. Palinski, J.L. Witztum Circulating autoantibodies to oxidized LDL correlate with arterial accumulation and depletion of oxidized LDL in LDL receptor-deficient mice. Arterioscler Thromb Vasc Biol. 2001. Vol. 21, N 1, pp. 95-100.

[6] P.X. Shaw, S. Hörkkö, S. Tsimikas, M.K. Chang, W. Palinski, G.J. Silverman, et al Human-derived anti-oxidized LDL autoantibody blocks uptake of oxidized LDL by macrophages and localizes to atherosclerotic lesions in vivo. Arterioscler Thromb Vasc Biol. 2001. Vol. 21, pp.1333–1339.

DOI 10.18699/BGRS/SB-2020-295

Amino acid and acylcarnitine levels relate with chronic schizophrenia

Irina Mednova
Mental Health Research Institute,
TNRMC, Tomsk, Russia
irinka145@yandex.ru

Alexandr Chernonosov
Institute of Chemical Biology and
Fundamental Medicine SB RAS,
Novosibirsk, Russia
alexander.chernonosov@niboch.nsc.ru

Marat Kasakin
Institute of Chemical Biology and
Fundamental Medicine SB RAS,
Novosibirsk, Russia
kassakinm@gmail.com

Elena Kornetova
Mental Health Research Institute,
TNRMC, Tomsk, Russia
kornetova@sibmail.com

Arkadiy Semke
Mental Health Research Institute,
TNRMC, Tomsk, Russia
asemke@mail.ru

Nikolay Bokhan
Mental Health Research Institute,
TNRMC, Tomsk, Russia
mental@tnimc.ru

Vladimir Koval
Institute of Chemical Biology and
Fundamental Medicine SB RAS,
Novosibirsk, Russia
koval@niboch.nsc.ru

Svetlana Ivanova
Mental Health Research Institute,
TNRMC, Tomsk, Russia
ivanovaniipz@gmail.com

***Abstract* — Amino acids and acylcarnitines play important role as substrates and intermediate products in the most pathways related with schizophrenia. We found a significant decrease in the concentration of valine, aspartate, citrulline, glycine, arginine, glutamate and ornithine as well as C14, C14OH, C16OH, C16:1, C16:1OH, C18, C18OH, C18:1, C18:1OH, C18:2OH and C5:1 and increase C4DC in patients with schizophrenia in comparison to the controls.**

Keywords* — *amino acid; acylcarnitine, schizophrenia; biomarker; metabolomics

Motivation and Aim

Schizophrenia is a disease with unknown etiology and pathogenesis. To understand the pathogenesis of schizophrenia, abundant factors and mechanisms were discussed e.g. mitochondrial dysfunction, inflammation, lipid oxidation, DNA damage, oxidative stress and apoptosis. These abnormalities of the metabolic pathways in patients with schizophrenia may well be reflected in metabolomic profiles [1]. Amino acids and acylcarnitines play important role as substrates and intermediate products in the most pathways mentioned above [2].

The present study aims to investigate serum amino acid and acylcarnitine levels in schizophrenic patients compare with healthy donors using 'omics' technology.

Methods

The study included 37 persons with schizophrenia (F20 according to ICD-10, median age 35 [31; 39] years, 18 women, mean duration of disease 15 ± 8,5). Healthy controls (n = 36) included an age- and sex-matched cohort without known disease symptoms. Quantification of amino acids and acylcarnitines was carried out using isotope labeled standards from an Amino Acids and Acylcarnitines kit #55000 for newborn screening (Chromsystems Instruments & Chemicals, Germany). Serum from patients and healthy donors was spotted on ProteinSaver cards 903 ™ (Whatman, USA) and samples were prepared according to the kit manufacturer's protocol. The mass spectrometric analysis was carried out in the dynamic multiple reaction monitoring (Dynamic MRM) mode using an Agilent 6410 QQQ mass spectrometer (Agilent Technologies, USA). Peak alignment, integration, and quantification were performed using MassHunter Quantitative Analysis software. Further data processing, analysis and visualization were carried out in the scripting language R v. 3.6.1 with RStudio environment v. 1.2.5001 using the packages "dplyr", "caret", "factoextra", "gglplot2" and others.

Results

We found a significant decrease in the concentration of valine, aspartate, citrulline, glycine, arginine, glutamate and ornithine as well as C14, C14OH, C16OH, C16:1, C16:1OH, C18, C18OH, C18:1, C18:1OH, C18:2OH and C5:1 and increase C4DC in patients with schizophrenia in comparison to the controls. According to cluster analysis, among acylcarnitines of long chain fatty acids, a group of metabolites was isolated: C18, C18OH, C18:1OH, C18:2OH carnitines (stearic FA derivatives), C16OH, C16:1OH carnitines (palmitic FA derivatives), C14OH carnitine (derivative myristic FA), whose concentrations were reduced in the group of patient compared with the control group. At the same time, C4DC showed an inverse relationship - in the group of patients with schizophrenia, it's concentration was higher. According to the literature data these disturbances in the level of amino acids and acylcarnitines are suggested have an impact in pathogenesis both schizophrenia and metabolic imbalance.

Acknowledgment

Supported by the the Russian Scientific Foundation, grant # 18-15-00011.

References

[1] Mednova I. et al. (2019) Amino acids and acylcarnitines as potential metabolomic markers of schizophrenia: new approaches to diagnostics and therapy. Bulletin of Siberian Medicine. 18(4): 197-208. (In Russ.)

[2] Hisamatsu T. et al. (2012) Novel, objective, multivariate biomarkers composed of plasma amino acid profiles for the diagnosis and assessment of inflammatory bowel disease. PloS one. 7(1): e31131.

DOI 10.18699/BGRS/SB-2020-296

Laser 3D-modeling in research of molecular features of skin lymphatic vessels in the patients with urticaria pigmentosa

Svetlana Viktorovna Michurina
Group of experimental pharmacology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
s.michurina@ngs.ru

Svechnikova Natalia Nikolaevna
Laboratory of clinical immunogenetics
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
n.svechnikova@ngs.ru

Andrey Yurievich Letyagin
Laboratory of pharmaceutical technologies
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
letyagin-andrey@yandex.ru

Irina Yurievna Ishchenko
Group of experimental pharmacology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
irenisch@mail.ru

Sergey Alekseevich Arkhipov
Group of experimental pharmacology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
arhipowsergei@yandex.ru

Solovieva Anastasia Olegovna
Laboratory of pharmacological active compounds
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
solovevaao@gmail.com

Konenkov Vladimir Iosifovich
Laboratory of clinical immunogenetics
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
vikonenkov@gmail.com

***Abstract* — Congestion in the blood and lymphatic vessels is detected in the papillary and reticular layers in the skin dermis of patients with pigment urticarial. This promotes the development of microcirculatory disorders and the formation of perivascular infiltrates. LYVE-1 positive staining was detected not only in the endothelium of the lymphatic vessels (both in the skin papillary layer and in the deeper dermis layers), but also in the prickle cells of the stratum spinosum and (less pronounced) in the cells of the stratum basal in the epidermis. Our 3D modeling of the LYVE-1 marker expression in skin epidermis suggests that mainly the prickle cells of the epithelium stratum spinosum and to a lesser extent the cells of the epithelium basal layer in patients with cutaneous mastocytosis participate in the lymphatic transport of hyaluronan and the spaces between these cells can be considered as prelimfatiks of the skin epidermis.**

Key words — urticaria pigmentosa, LYVE-1, prickle cells of the stratum spinosum

Introduction

About 85% of adult patients with mastocytosis have skin lesions (urticaria pigmentosa). The disease is classified as a rare dermatosis, the pathogenesis of which has not been studied enough. Along with comprehensive studies of the state of mast cells in the development of clinical manifestations of the disease, the lymphatic system of the human skin remains unexplored. This can be explained by the complexity of methods for studying the morphology of small lymphatic vessels. However, researchers often overlook the skin lymphatic vessels, since they are small slit-like structures above the hair follicles, resembling artifact-like separation of collagen bundles [1]. The structure of the lymphatic system of the skin was first described in Russia by D. A. Zhdanov in 1952 [2]. A little later, the skin lymphatic vessels were determined using electron microscopy. Selective markers for lymphatic endothelium were until recently rare or unavailable. Currently, lymphatic vessels are easily detected by the immunohistochemical method using LYVE-1, which is a cell surface receptor for lymphatic endothelial cells [3].

Materials and Methods

The paper presents clinical observations of 2 patients with a cutaneous form of mastocytosis, histological examination, immunohistochemical study. For histological examination, skin samples were fixed in 10% buffered formalin in 48 hours, dehydrated in a series of alcohols with increasing concentrations, enclosed in histomix; sections were obtained by the microtome LEICA RM2155. Immunohistochemical detection of the LYVE-1 marker was performed in sections with a 3 μm thickness using the indirect ABC-peroxidase method using the VECTASTAIN Universal ABC-Peroxidase Kit (PK-7200, Vector Laboratories, USA). Endogenous peroxidase blocking was performed by incubation of sections in 0.3% H_2O_2 solution for 10 min, then sections were incubated with normal blocking non-immune horse serum for 20 min. Next, the sections were incubated for an hour at room temperature with antibodies to LYVE-1 (Isotype: Rabbit polyclonal, bs-1311R; Bioss) using the final dilution of 5 μg/ml, then they were washed in 3 shifts of phosphate buffer (FB) for 3 min and were incubated for 30 minutes at room temperature with the second biotinylated antibodies followed by washing in 3 shifts of FB for 5 minutes. Incubation with the ABC-peroxidase complex was carried out at room temperature for 30 min, the sections were washed in 3 shifts of FB for 5 minutes. At the last stage immunohistochemical staining of the sections was performed in the chromogenic substrate (ImmPACT DAB, SK-4105, Vector Laboratories, США). During all immunohistochemical reactions, “negative controls” were carried out, replacing incubation with primary antibodies by incubation with non-immune sera of the corresponding animal (rabbit serum). Digital images of human skin preparations were taken with the AxioImager.M2 Zeiss microscope (Axiocam 305 Color camera). 3 D modeling of

skin prelimphatics was performed using the laser holographic confocal microscope Nanolive (3D Cell Explorer).

Results and Discussion

Patient G., a man, 35 years old, was on outpatient observation with a cutaneous form of mastocytosis. Patient T., a woman, 41 years old, had a diffuse cutaneous form of mastocytosis. Hyperpigmentation of basal layer cells, mild epidermal edema and focal hyperkeratosis with loosening of the horny masses due to edema were detected in histological preparations of the skin in patients. Vacuolization was observed in individual cells of the stratum basale and the stratum spinosum (or the prickle cell layer) in the skin epidermis, that is a manifestation of vacuole hydropic dystrophy, characterized by intracellular edema of keratinocytes with the formation of vacuoles in their cytoplasm, which is an inducer of apoptosis. Individual pictures of acantholysis were identified. Immunohistochemical staining for LYVE-1 revealed that lymphatic vessels were located in the horizontal plane of the papillary layer of the skin dermis, often under the epidermis basal layer and they were wider than blood vessels. Then they plunged into a deeper dermis, where they formed a network. LYVE-1 positive staining was also detected in the sebaceous glands of the skin. In addition, tagging of skin for LYVE-1 revealed a pronounced positive signal for the marker in the cells of the stratum spinosum in the epidermis and a weaker signal for this studied marker in the cells of the epidermis stratum basale in patients with cutaneous mastocytosis. LYVE-1 is a hyaladherin, containing the Link domain - a protein capable of binding to hyaluronic acid. In humans it is encoded by the *LYVE1* gene. LYVE-1 is a type I integral membrane glycoprotein. It binds to both soluble and immobilized hyaluronan. Our 3D modeling of the LYVE-1 marker expression in patients' skin epidermis revealed clear boundaries of prelimphatics and identified that intercellular edema was most pronounced in the stratum spinosum of the epidermis, that led to the loss of communications between epithelial cells in limited areas.

Acknowledgment

The study is supported by the budget project of the Research Institute of Clinical and Experimental Lymphology – branch of ICG SB RAS (No.0324-2019-0046-C-02). This work involved the use of equipment from the Center for Genetic Resources of Laboratory Animal at the Institute of Cytology and Genetics, Siberian Branch, Russian Acadymy of Sciences.

References

[1] J.P. Sundberg, C.H. Pratt, K.A. Silva, V.E. Kennedy, T.M. Stearns, B.A. Sundberg, L.E. King and H. HogenEsch, “Dermal lymphatic dilation in a mouse model of alopecia areata,” Exp. Mol. Pathol., vol. 100, № 2, pp. 332-336, 2016.

[2] D.A. Zhdanov, General anatomy and physiology of lymphatic system. Moscow: Medgiz, 1952, 336p. (In Russian).

[3] C.H. Tripp, B. Haid, V. Flacher, M. Sixt, H. Peter, J. Farkas, R. Gschwentner, L. Sorokin, N. Romani and P. Stoitzner, “The lymph vessel network in mouse skin visualised with antibodies against the hyaluronan receptor LYVE-1,” Immunobiology, vol. 213, № 9-10, pp. 715-728, 2008.

DOI 10.18699/BGRS/SB-2020-297

The expression of Bcl-2 family proteins in liver cells of *C57Bl/6* mice under conditions of functional pinealectomy

Svetlana Viktorovna Michurina
Group of experimental pharmacology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
s.michurina@ngs.ru

Irina Yurievna Ishchenko
Group of experimental pharmacology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
irenisch@mail.ru

Sergey Alekseevich Arkhipov
Group of experimental pharmacology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
arhipowsergei@yandex.ru

Andrey Yurievich Letyagin
Laboratory of pharmaceutical technologies
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
letyagin-andrey@yandex.ru

Maxim Aleksandrovich Korolev
Laboratory of connective tissue pathology
RICEL – Branch of ICG SB RAS
Novosibirsk, Russia
kormax@bk.ru

Evgenii Leonidovich Zavjalov
Center of genetis resources of laboratory animals
ICG SB RAS
Novosibirsk, Russia
zavjalov@bionet.nsc.ru

***Abstract* — The expression of antiapoptotic Bcl-2 protein and the proapoptotic Bad protein in liver cells of C57BL/6 mice males kept 14 days under 24-hour lighting (24hL) (photoperiod light/dark 24/0 hours) was studied. Control animals kept under standard lighting conditions (photoperiod light/dark 14/10 hours). The immunohistochemical analysis (indirect avidin–biotin-streptavidin method) and morphometric assessment were used. We found in the liver cells of mice under round-the-clock lighting the increase in the area of Bcl-2 protein expression without changing the area of Bad protein expression, that led to a decrease in the ratio of Bcl-2/Bad. This result indicated weakening of the antiapoptotic protection of organ cells and creates conditions for activation of the "mitochondrial branch" of apoptosis in animal liver cells with functional pinealectomy.**

Key words* — *round-the-clock illumination, functional pinealectomy, liver, Bad, Bcl-2

Introduction

It was found that blood circulatory and lymphatic flow disturbances develop in the liver lymphatic region under the influence of continuous lighting, which is an integral factor of modern life. These processes lead to the development of hypoxia, which adversely affects the structure and functions of the cell mitochondrial apparatus [1, 2]. Apoptosis helps the body to get rid of unnecessary and defective cells. Most forms of apoptosis are realized not through cell death receptors, but by the mitochondrial pathway with the participation of Bcl-2 family proteins. It is believed that the ratio of the antiapoptotic Bcl-2 protein and proapoptotic Bad, Bax proteins is a "molecular switch", which determines whether tissue growth or atrophy will occur [3]. The purpose of this study was to evaluate the influence of 24-hour lighting (24hL) on the expression of the proapoptotic Bad protein and the antiapoptotic Bcl-2 protein in liver cells of C57Bl/6 mice male, as well as to determine the ratio of Bcl-2 and Bad protein expression areas.

Materials and Methods

The males of C57Bl/6 mice aged of 10-12 weeks were kept in the SPF-vivarium center of ICG SB RAS. Food and water were provided to animals *at libitum*. The intact animals (the group "Control", n=5) were kept under the standard lighting mode - photoperiod light/dark: 14/10 hour (h), at the same time a smooth increase in illumination to daytime values within 1 hour (dawn) and a smooth decrease in illumination values until complete shutdown within 1 hour (sunset) were assigned to the light phase of the day. The mice of the experimental group (the gr. "24hL", n=6) were kept for 14 days under conditions of 24-hour lighting - photoperiod light/dark: 24/0 h. Animals were removed from the experiment by the cranio-cervical dislocation method and liver samples were taken for light-optical and immunohistochemical studies. Liver samples were fixed in 10% buffered formalin for 48 hours, dehydrated in a series of alcohols of increasing concentrations and embedded in Histomix.

The immunohistochemical study of the expression of Bcl-2 and Bad proteins was performed on liver paraffin sections with a thickness of 3 μm by means of indirect avidin–biotin-streptavidine method (ABC-method) using the Vectastain Universal Elite ABC Kit ("Vector Laboratories", Catalog Number PK-7200). At the last stage, immunohistochemical staining was carried out in a chromogenic substrate containing diaminobenzidine (the solution is prepared *ex tempore* from the components of the set "ImmPACT DAB"; "Vector Laboratories", Catalog Number SK-4105). For quantification of Bcl-2 and Bad expression in the mouse liver, there was performed computer morphometric analysis of digital photographs at ×400 magnification. The average areas of zones staining on Bcl-2 and Bad in digital images were determined using the program "Image J". The ratio of the Bcl-2 expression area to the Bad expression area was calculated. A comparison of the results between the groups "24hL" and "Control" was performed using the Mann-Whitney U-test for abnormally distributed characters in independent samples. Differences of compared values were considered statistically significant at $P<0.05$.

Results and Discussion

A study of Bcl-2 family protein expression in the liver of C57Bl/6 mice males kept under 24-hour lighting revealed the pronounced immunohistochemical Bad staining of sinusoidal cells of blood capillaries. The bad-positive signal was detected in the endothelium of interlobular veins and in the ductal epithelium of triad bile ducts, and it was also sometimes found

in hepatocytes located mainly in periportal and pericentral zones of hepatic lobes. At the same time, weak immunohistochemical staining for the antiapoptotic Bcl-2 protein was revealed in sinusoidal liver cells and in single hepatocytes. Staining for Bcl-2 was not determined in the ductal epithelium of triad bile ducts.

Morphometric analysis of liver preparations in the group "24hL" showed an increase in the expression area of the proapoptotic Bad protein compared to animals exposed to standard light (light/dark: 14/10 h). At the same time, the expression area of the Bcl-2 protein in this group did not change significantly compared to control animals, but tended to decrease. The Bcl-2/Bad expression area ratio decreased significantly in the «24hL» mice compared with the «Control» animals due to an increase in the Bad expression area.

It is known that in modern life, a violation of the light regime, and in particular long-term round-the-clock lighting, leads to a decrease in melatonin production resulted to the development of desynchronosis followed by the development of organic pathology [2,4]. Due to apoptosis, the body gets rid of defective cells without starting the process of inflammation. Moreover, most of its forms are realized not through cell death receptors, but by the mitochondrial pathway with the participation of Bcl-2 family proteins. Currently, numerous data indicate that mitochondria are the site of melatonin synthesis itself. These organelles contain the enzymes N-acetyltransferase and hydroxyindole-O-methyltransferase. Melatonin is a mitochondrial antioxidant, protecting these organelles by binding reactive oxygen species [5]. It is known that the mitochondrial dynamics exhibit an oscillatory character which corresponds to the circadian secretory rhythm in the cells [6]. Studies have shown that melatonin reduces the rate of apoptosis, prevents the opening of mitochondrial pores, has a stabilizing effect and supports the mitochondrial membrane potential, prevents the release of cytochrome C, prevents electron leakage and preserves mitochondrial functions. In addition, mitochondrial biogenesis and dynamics are also regulated by melatonin [5]. Since apoptosis is triggered by the inactivation of Bcl-2 when it binds to the Bad protein, the increase in the staining area of the proapoptotic protein Bad set by us indicates a decrease in antiapoptotic protection and in the apoptosis development along the mitochondrial pathway in liver cells. This is also confirmed by a decrease in the ratio of Bcl-2/Bad expression areas in the liver of mice kept under 24-hour lighting (24/0 h) compared to control animals (14/10 h).

ACKNOWLEDGMENT

The study is supported by the budget project (No. AAAA-A17-17072710029-7) and implemented using the equipment of the Center for Genetic Resources of Laboratory Animals at ICG SB RAS, supported by the Ministry of Education and Science of Russia (Unique identifier of the project RFMEFI62119X0023).

REFERENCES

[1] S. V. Michurina, Yu. I Borodin., I. Yu. Ishchenko, A.D. Belkin, A.V. Shurlygina and P.M. Larionov, “The lymphatic region of Wistar rat liver under conditions of combined influence of alcohol intoxication and round-the-clock lighting,” Bulletin of SB RAMS, vol. 5, pp. 44-49, 2008. (in Russian)

[2] K.L.G. Russart and R.J. Nelson, “Light at night as an environmental endocrine disruptor,” Physiol. Behav., vol. 190, pp. 82-89, 2018.

[3] S.Y. Jeong and D.W. Seol, “The role of mitochondria in apoptosis,” BMB Rep., vol. 41, № 1, pp. 11-22, 2008.

[4] S. V. Michurina, D. V. Vasendin and I. Yu. Ishchenko, “Physiological and biological effects of melatonin: some results and prospects of study,” Ross. Fiziol. Zh. Im. I. M. Sechenova, vol. 104, № 3, pp. 257-271, 2018. (in Russian).

[5] R.J. Reiter, D.X. Tan, S. Rosales-Corral, A. Galano, X.J. Zhou and B. Xu, “Mitochondria: Central Organelles for Melatonin′s Antioxidant and Anti-Aging Actions,” Molecules, vol. 23, № 3, p. E509, 2018.

[6] A. Kalsbeek, F.A. Scheer, S. Perreau-Lenz, S.E. La Fleur, C.X. Yi, E. Fliers and R.M. Buijs, “Circadian disruption and SCN control of energy metabolism,” FEBS Lett., vol. 585, № 10, pp.1412-1426, 2011.

DOI 10.18699/BGRS/SB-2020-298

Genetic polymorphism associated with infectious pulmonary diseases in siberian populations and among patients with community acquired pneumonia

Svetlana Vladimirovna Mikhailova
Lab. of human molecular genetics
ICG SB RAS
Novosibirsk, Russia
ORCID 0000-0002-0897-5473

Irina Ivanovna Logvinenko
Lab. of preventive medicine
IIPM – a branch of the ICG SB RAS
Novosibirsk, Russia
ORCID *0000-0003-1348-0253*

Liliya Valeryevna Shcherbakova
Laboratory of Clinical-Populational and Prophylactic Studies on Internal and Endocrine Diseases
IIPM – a branch of the ICG SB RAS
Novosibirsk, Russia
ORCID 0000-0001-9270-9188

Mikhail Ivanovich Voevoda
Dep. of Human Molecular Genetics
ICG SB RAS
Novosibirsk, Russia
ORCID 0000-0001-9425-413X

Nadejda Ivanovna Logvinenko
Dep. of therapy, hematology and transfusiology
Novosibirsk State Medical University
Novosibirsk, Russia
nadejda-logvinenko@yandex.ru

Abstract — **Innate immune system is the first to respond to pathogen invasion; it activates adaptive immunity and regulates the intensity of inflammation. Two single nucleotide polymorphisms of the innate immunity genes: rs5743708 of the *TLR2* and rs8177374 of the *TIRAP* were shown to be associated with pneumonia and tuberculosis, but the data are contradictory in different ethnic groups. We assessed the prevalence of these variants in Caucasoid and Asian population samples, and among patients with community acquired pneumonia (CAP). Carriage of the rs5743708 A allele was found to be predisposed to severe CAP (OR=2.77, p=0.021), GG/CT genotype for rs5743708 / rs8177374 is proved to be protective against it (OR=0.478, p=0.022) in Caucasoid patients. Both Caucasoid and Asian studied populations were turn to be different from European and neighboring Asian populations on the genotypes prevalence. Differentiation of CAP by the causative pathogen could help to eliminate the currently observed contradictions between different research groups**.

Keywords — genetic polymorphism, genetic predisposition, community acquired pneumonia, pulmonary tuberculosis, TLR2 gene , TIRAP gene

Motivation and Aim

Community acquired pneumonia (CAP) and pulmonary tuberculosis (TB) are the diseases characterized by high mortality; according to WHO, they are consistently among the ten leading causes of death in the world [1]. CAP is lower respiratory tract inflammatory disease caused mainly by viruses and bacteria. Improved diagnostic methods have recently shown that viruses, not *Streptococcus pneumoniae*, are the main cause of CAP. However severe pneumonia is more likely to develop in the case of combined infections and a seasonal increase in pneumonia coincides with seasonal outbreaks of influenza [2]. TB is chronic infectious disease, caused mainly by *Mycobacterium tuberculosis*. About a quarter of the world's population is infected with this bacterium and is at risk of developing the disease. Pathogenesis of these diseases is based on different mechanisms: excessive inflammation for CAP and the inhibition of acute immune response by *M. tuberculosis* for TB. Interestingly, for some genes of innate immunity in humans, an association with both CAP and pulmonary TB has been shown [3-6]. However, the data are contradictory in different racial and ethnic groups. In this work, we analyzed the frequencies for the rs5743708 (R753Q) of the *TLR2* gene and rs8177374 (S180L) of the *TIRAP* gene in Caucasian and Asian samples of Russia, including long-lived people sample. In addition, we estimated the frequencies of these polymorphisms among patients with severe and non-severe CAP.

Methods

The adolescent sample was selected from the inhabitants of Novosibirsk city. Yakutian Caucasoid sample was collected in Tommot, Neryungri, Ust-Nera, and Yakutsk towns. These were people who lived in Yakutia for more than 16 years or were born there Asian sample consisted of Chukchis from Kanchalan village and Yakuts from Tommot, Neryungri, Ust-Nera towns and Kylayy village. Long-lived people sample was collected in West Siberia. CAP patients sample was collected in offices of pulmonary hospitals of Novosibirsk and Yakutsk. CAP with various pathogenesis was regarded as severe if the CURB65 rating scale index was 4-5 points. All patients signed an informed consent to participate in the study. Genotyping was performed by PCR with analysis of restriction fragment length polymorphism.

Results and Discussion

Genotypes amount and frequencies for rs5743708 of the *TLR2* gene and rs8177374 of the *TIRAP* gene are presented in Table1.

TABLE 1. GENOTIPES FREQUENCIES FOR RS5743708 / RS8177374 IN POPULATION SAMPLES OF NORTHERN ASIA AND AMONG CAP PATIENTS

Samples \ genotypes	n[a]	GG/CC		GG/CT		GG/TT		AG/CC		AG/CT		AA/CC		AA/CT	
	total	*n*	*p*[b]	*n*	*p*	*n*	*p*	*n*	*p*	*n*	*p*	*n*	*p*	*n*	*p*
Adolescents, Novosibirsk	451	335	0.74	72	0.16	4	0.01	29	0.06	8	0.02	1	0.002	1	0.002
Caucasoids, Yakutia	277	219	0.79	43	0.16	2	0.01	10	0.04	3	0.01	0	0	0	0
Long-lived peoples, West Siberia	188	150	0.80	25	0.13	2	0.01	11	0.06	0	0	0	0	0	0
CAP total, Caucasoids	406	307	0.76	71	0.18	2	0.004	22	0.05	3	0.01	0	0	0	0
Severe CAP	120	93	0.78	13	0.11	1	0.01	11	0.09	2	0.02	0	0	0	0
Non- severe CAP	286	215	0.75	58	0.20	1	0.004	11	0.04	1	0.004	0	0	0	0
Chukchis	91	75	0.82	16	0.18	0	0	0	0	0	0	0	0	0	0
Yakuts	120	111	0.93	6	0.05	0	0	3	0.03	0	0	0	0	0	0
Severe CAP, Yakuts	33	28	0.85	3	0.09	1	0.03	1	0.03	0	0	0	0	0	0

a number of people, b genotype frequency

The genotype distribution in all populations was in Hardy-Weinberg equilibrium. No differences in the alleles and genotypes were found between male and female. No significant difference was found between Russian adolescences and long-lived people samples. Apparently, the carriage of the studied alleles is not associated with a predisposition to longevity. There was significant difference in rs5743708 frequency between Novosibirsk adolescents and Caucasoid sample from Yakutsk (χ^2 =4.647, p= 0.032). No one allele or genotype was associated with CAP predisposition among Caucasoid in total. In case of pneumonia, carriage of the rs5743708A allele of the *TLR2* gene predisposed to severe CAP (AG/CC +AG/CT vs all OR=2.77, 95% CI: 1.227-6.272, p=0.021); heterozygous genotype for rs8177374 of *TIRAP* gene in a combination with GG of *TLR2* gene had protective effect against severe CAP (GG/CT vs all OR=0.478, 95% CI: 0.251-0.909, p=0.022). However, similar association was not observed in Yakuts, possibly due to a small sample size. In Russian, frequency of "protective" rs8177374T allele was lower than in European populations (0.10 vs 0.22); in North Asians, it was higher than in neighboring Chinese populations (0.03-0.09 vs 0.01) [7].

Toll-like receptor 2, TLR2 heterodimerizes with TLR1 or TLR6 after recognition of pathogen-associated molecular patterns. TIRAP (TIR-domain containing adaptor protein) mediates downstream signaling of the TLRs resulting in the initiation of cytokine genes transcription. Both studied substitutions change proteins structure, but have opposite effect on the disease predisposition. This may be due to location of the rs5743708 in a binding site for TLR6 [8]. This substitution cans potentially increaseTLR2:TLR1 concurrent binding, while exactly TLR1 is responsible for NF-κB inflammatory signal cascade activation. The rs8177374T on the contrary, weakens signal transmission. The observed differences in the frequencies of variants of innate immunity genes in Asian populations can reflect the history of human impact with various pathogens during resettlement in Eurasia. "Pathogenic" for CAP and TB variant rs5743708A of the *TLR2* gene e.g. was shown to be protective against Lyme disease [9].

Conclusions

We found carriage of the rs5743708 A allele to be predisposed to severe CAP and GG/CT genotype for rs5743708 / rs8177374 was protective against it in Caucasoid patients with pneumonia.

Differentiation of CAP by the causative pathogen could help to eliminate the currently observed contradictions between different research groups.

ACKNOWLEDGMENT

Supported by the SB RAS fundamental research program AAAA-A17-117072710029-7.

REFERENCES

[1] https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

[2] J.A. McCullers, "The co-pathogenesis of influenza viruses with bacteria in the lung", Nat Rev Microbiol, vol 12, pp. 252-262, April 2014.

[3] I. Patarčić, A. Gelemanović, M. Kirin, I. Kolčić, E. Theodoratou, K.J. Baillie et al., "The role of host genetic factors in respiratory tract infectious diseases: systematic review, meta-analyses and field synopsis" Sci Rep, vol 5:16119, Nov 2015.

[4] T.V. Smelaya, O.B. Belopolskaya, S.V. Smirnova, A.N. Kuzovlev, V.V. Moroz, A.M/ Golubev, et al., "Genetic dissection of host immune response in pneumonia development and progression", Sci Rep, vol 6:35021, Oct 2016.

[5] L. Hu, H. Tao, X. Tao, X. Tang, and C. Xu, "TLR2 Arg753Gln gene polymorphism associated with tuberculosis susceptibility: An updated meta-analysis", Biomed Res Int, vol 2019:2628101, Jan 2019.

[6] C.C. Khor, S.J. Chapman, F.O. Vannberg, A. Dunne, C. Murphy, E.Y. Ling, et al, "A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis", Nat Genet, vol 39, pp.523-528, April 2007.

[7] https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes

[8] Y. Xiong, C. Song, G.A. Snyder GA, E.J. Sundberg, and A.E. Medvedev, "R753Q polymorphism inhibits Toll-like receptor (TLR) 2 tyrosine phosphorylation, dimerization with TLR6, and recruitment of myeloid differentiation primary response protein 88", J Biol Chem, vol 287, pp. 38327-38337, Nov 2012.

[9] N.W. Schröder, I. Diterich, A. Zinke, J. Eckert, C. Draing, V. von Baehr et al., "Heterozygous Arg753Gln polymorphism of human TLR-2 impairs immune activation by Borrelia burgdorferi and protects from late stage Lyme disease", J Immunol, vol 175, pp. 2534-2540, Aug 2005.

DOI 10.18699/BGRS/SB-2020-299

Cluster organization of miRNA binding sites in mRNA of atherosclerosis candidate genes

Dina Mukushkina
SRI of biology and biotechnology problems
Al-Farabi Kazakh National University
Almaty, Kazakhstan
dina.mukushkina@gmail.com

Dana Aisina
SRI of biology and biotechnology problems
Al-Farabi Kazakh National University
Almaty, Kazakhstan
dana.aisina03@gmail.com

Anatoliy Timofeevich Ivashchenko
SRI of biology and biotechnology problems
Al-Farabi Kazakh National University
Almaty, Kazakhstan
a.iavashchenko@gmail.com

Abstract — **Many candidate genes are involved in the development of atherosclerosis, which are manifested in a variety of metabolic processes, disturbances of which are observed in this disease. Many publications describe the involvement of miRNAs in the development of atherosclerosis. Therefore, it is necessary to establish which miRNAs can affect the expression of candidate genes. Quantitative characteristics of the interaction of miRNAs with mRNAs candidate genes were determined using the MirTarget program, which detects localization of miRNA binding sites in 3'UTR, 5'UTR and CDS; free energy interaction miRNA with mRNA; schemes of nucleotide interaction between miRNA and mRNA. In the mRNA of atherosclerosis genes, the binding sites of two or more miRNAs located in 3'UTR, 5'UTR and CDS are established. The organization of miRNAs binding sites with overlapping nucleotide sequences forming clusters was revealed. Such an organization of miRNAs binding sites leads to their compaction several times and causes competition among miRNAs for binding to mRNA. The features of the interaction of miRNAs with mRNA of candidate genes depending on their expression are established. Associations of miRNAs and candidate atherosclerosis genes are proposed for the early diagnosis of this disease.**

Keywords — miRNA, mRNA, gene, association, marker, atherosclerosis

INTRODUCTION

Candidate genes are participated in the development of atherosclerosis, which are implicated in a variety of metabolic processes that are disturbed in this disease. It has been found that miRNAs, which are nanoscale regulatory biomolecules, are involved in many biological processes at all stages of the development of atherosclerosis, from early endothelial dysfunction to the erosion and rupture of an unstable atherosclerotic plaque. The miRNAs are able to regulate gene expression at the levels of translation binding to target mRNAs. Many publications have described the involvement of miRNAs in the development of atherosclerosis. It is necessary to establish which miRNAs affect the expression of candidate genes.

METHODS AND ALGORITHMS

The nucleotide sequences of 2565 miRNAs were downloaded from the miRBase (http://mirbase.org, Release 22.1) and 3707 miRNAs obtained from a report by Londin E. et al. [1]. The nucleotide sequences of mRNA genes were obtained from GenBank (http://www.ncbi.nlm.nih.gov). A database of 91 candidate genes including the names of the genes and publication sources (PubMed database) was compiled, confirming the associations of these genes with atherosclerosis. A search for the target genes of miRNAs was performed using the MirTarget program [2]. This program determines the following binding characteristics: the start of the miRNA-binding site of mRNA; the locations of miRNA-binding sites (3'UTR, 5'UTR, CDS); the interaction free energy (ΔG, kJ/mole); nucleotide interaction schemes between miRNAs and mRNAs.

RESULT

Binding sites for 24 mRNAs genes with miRNAs in the 5'UTR region were identified. The binding sites of the miRNAs and mRNAs of the target genes were not uniform along the length of an mRNA. Both multiple and single binding sites were identified. The mRNA regions containing overlapping binding sites are referred to clusters [3]. The results regarding the location of miRNA-binding sites in mRNAs with overlapping nucleotide sequences in a cluster leads to the compaction and causes competition among miRNAs. The mRNA of *GAS6* gene forms a large cluster with 22 different miRNAs, half of which form two or three binding sites, while the rest represent single binding sites. The cluster size is 37 nucleotides. The total binding sites length in cluster is 858 nucleotides, where the degree of compaction is 23.2. This compaction allowed these miRNA binding sites to be located in 5'UTR with a length of 153 nucleotides. For eight associations of miRNAs with mRNA *GAS6* gene the free interaction energy of miRNAs with mRNA have been more than -130 kJ/mole, which allows us to recommend these associations as markers of atherosclerosis. The cluster of binding sites of 18 miRNAs в the *NFE2L2* gene 5'UTR.mRNA have size of 45 nucleotides (Table). The total size of all binding sites is 521 nucleotides, which means that the degree of compaction is 11.6. Nine miRNAs interacted with the mRNA of the *NFE2L2* gene with a free energy of more than -130 kJ/mole are recommended as markers of atherosclerosis. The 41 genes exhibited one binding site in CDS mRNAs for one miRNA. In CDS of *IRS2* mRNA the binding sites of 28 miRNAs were located in cluster with a length of 41 nucleotides. The total cluster size was 1147 nucleotides, while the degree of cluster compaction was 28.0. The binding sites of 12 miRNAs were located in other cluster with a length of 35 nucleotides. The *PDE4D* gene forms two clusters but smaller.

TABLE CHARACTERISTICS OF MIRNA BINDING SITES IN 5'UTR AND CDS OF MRNA OF *NFE2L2* AND *PDE4D* CANDIDATE GENES

Gene	miRNA	Region of RNA	Start of site, nt	ΔG, kJ/mole	ΔG/ΔGm, %	Length, nt
NFE2L2	ID00522.5p-miR	5'UTR	438	-125	89	23
	ID02187.5p-miR	5'UTR	439÷445(2)	-123	89	23
	ID03367.5p-miR	5'UTR	441÷453(2)	-115	92	20
	ID01804.3p-miR	5'UTR	442÷444(2)	-132	90	23
	ID02294.5p-miR	5'UTR	443	-132	90	24
	ID00061.3p-miR	5'UTR	444÷450(3)	-125÷-134	91÷97	22
	ID00457.3p-miR	5'UTR	444	-123	91	22
	ID01041.5p-miR	5'UTR	444÷445(2)	-129÷-134	88÷91	24
	ID01873.3p-miR	5'UTR	444÷447(2)	-121÷-123	92÷94	21
	ID00296.3p-miR	5'UTR	448	-140	89	25
	ID01641.3p-miR	5'UTR	448	-136	91	24
	ID01702.3p-miR	5'UTR	448	-138	92	24
	ID02890.3p-miR	5'UTR	458	-119	89	23
	ID02770.5p-miR	5'UTR	462	-115	92	20
PDE4D	ID00061.3p-miR	CDS	391÷413(3)	-125÷-129	91÷94	22
	ID00457.3p-miR	CDS	392	-123	91	22
	ID01315.3p-miR	CDS	392	-115	92	20
	ID01377.3p-miR	CDS	394	-121	95	20
	ID01705.3p-miR	CDS	398	-117	92	21
	ID00296.3p-miR	CDS	404÷411(4)	-140÷-142	89÷91	25
	ID01641.3p-miR	CDS	407÷410(2)	-132÷-132	89	24
	miR-3960	CDS	408÷415 (3)	-115÷-117	92÷93	20
	ID01458.5p-miR	CDS	410	-134	91	23
	ID01702.3p-miR	CDS	411	-138	92	24
	ID02064.5p-miR	CDS	415	-129	90	23
	ID01184.3p-miR	CDS	419	-117	93	20

What is characteristic is that miRNA ID01641.3p-miR, ID00061.3p-miR, miR-3960 and ID01702.3p-miR are found in both clusters. The first cluster consists of four single binding sites and one multiple binding site ID01702.3p-miR, cluster size is 35 nucleotides, from the position of 335 nucleotides and to 369 nucleotides. The average ΔG value is -130 kJ/mole. The second cluster consists of four multiple binding sites and six single binding sites, the cluster size is 49 nucleotides from 391 to 439 nucleotides (Table). The total cluster size is 447 nucleotides and degree of compaction was 9.1.

Clusters consisting of single miRNA-binding sites were found in the mRNA following genes: *ACE*, *ADAMTS13, ADRB3*, *ICAM1* and *FASLG*. In 3'UTR mRNA of 45 genes were characterized binding sites. In *ADRB3, CD36, FASLG, F11R, FLT1, ICAM1, PLA2G7, PPARGC1A* mRNAs the clusters of miR-466, ID00436.3p-miR, ID01030.3p-miR binding sites were established. The binding of 53 miRNAs with mRNAs of 14 candidate genes with free energy interaction more than -130 kJ/mole were established. The miR-619-5p was fully complementary to *ADAM17* and *CD3*6 mRNA genes, ID01593.5p-miR to *ANGPTL4* mRNA gene, ID01935.5p-miR to *NFE2L2* gene and miR-5096 to *IL18* mRNA gene.

CONCLUSION

The miRNA associations with atherosclerosis candidate genes have been established, which consist of: one gene and several miRNAs; one gene and one miRNA; one miRNA and several genes; two or more miRNAs with two or more candidate genes. sites in mRNA candidate genes contributesto a more accurate diagnosis of the participation of competing miRNAs in the development of atherosclerosis.

ACKNOWLEDGEMENTS

The work was carried out with the financial support of the Ministry of Education and Science of the Republic of Kazakhstan within the framework of the grant №AP05132460. We are grateful to Pyrkova A.Yu. to performing calculations on the program MirTarget.

REFERENCES

[1] E. Londin, et al. "Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs," PNAS, pp. E1106-E1115, 2015.

[2] A. T. Ivashchenko, A. Y. Pyrkova, R. Y. Niyazova, A. Alybayeva, K. Baskakov, "Prediction of miRNA binding sites in mRNA," Bioinformation, vol. 12. pp. 237-240. 2016.

[3] D. Aisina, R .Niyazova, S. Atambayeva, A. Ivashchenko, "Prediction of clusters of miRNA binding sites in mRNA candidate genes of breast cancer subtypes," PeerJ. vol. 7. pp. e8049, 2019. doi: 10.7717/peerj.8049

DOI 10.18699/BGRS/SB-2020-300

Experimental substantiation of transdermal transport of photosensitizers by fractional laser photothermolysis(FLP)

Sergey Nikonov
Institute of Medicine and Psychology
Novosibirsk State University
Professor, the chair of Surgical Diseases
Novosibirsk, Russia
sibnovomed@mail.ru

Danila Chernopyatov
Institute of Medicine and Psychology
Novosibirsk State University
Student
Novosibirsk, Russia
d9132069178@gmail.com

Vadim Nimaev
Surgical Lymphology Laboratory
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
nimaev@gmail.com

***Abstract* — as known photodynamic therapy and photothermolysis are used to treat skin diseases and to correct signs of aging. These methods are minimally invasive but their disadvantages are related to skin damage, prolonged phototoxicity with intravenous photosensitizer, and the insufficient penetration of the drug through the skin by local application. The possibility of minimally invasive transdermal drug transport by using photothermolysis and photodynamic therapy technologies on the example of the delivery of the Radagel photosensitizer has been demonstrated. The method of transdermal delivery of a photosensitizer through the skin micropores, which are formed by focused energy rays of a non-ablative laser has been developed for the side effects prevention of laser techniques with the summation of their advantages.**

Keywords* — *photodynamic therapy; photosensitivity; photothermolysis; laser energy; skin disease;transdermal transport

BACKGROUND

In the treatment of benign, precancerous and malignant skin diseases, the technology of photodynamic therapy (PDT) is actively introduced, which based on the selective damage of a pathological substrate by a drug photosensitizer (PS) in combination with electromagnetic radiation of the visible red range in the presence of tissue oxygen. The traditional method of intravenous photosensitization of the body involves the introduction of high doses of expensive PS (Radachlorin, Photoditazin, Verteporfin, Alasens), which is associated with long-term retention of drugs in healthy and pathologically altered cells, which leads the for patients to observe the light protection regimen in the post-treatment period to avoid skin burns, photodermatitis, prolonged erythema and phototoxic face edema. In this regard, for the necessities of outpatient dermatology, dermato-oncology and cosmetology, the local application photosensitization of pathological changes by gel forms of FS with subsequent PDT light energy sources is the most acceptable. Despite the fact that when applied to the skin, PS shows a tropism to the tumor, inflamed, and dystrophic cells, the distribution of the drug in the layers of the dermis and the depth of its transdermal penetration is unknown.

We have developed a targeted local photosensitization of the deep layers method of the skin by fractional laser photothermolysis using focused pulsed-periodic radiation from a fiber laser. Microthermal zones which were formed on the skin of experimental mice have provided the efficient transdermal transport of FS, which was proved by pathomorphological studies of skin layers with confocal microscopy. Based on the fluorescence characteristic of PS, the distribution from the epidermis through the stratum corneum, dermis to subcutaneous adipose tissue was shown. All in all, the FLP method for local PS transdermal transport can be highly effective in outpatient PDT in dermatology and cosmetology.

PURPOSE

To substantiate experimentally the applicability of photodynamic fractional laser photothermolysis (PFLP) for creation new treatment technologies in dermatocosmetology and dermato-oncology.

WORK TASKS

To study the transport possibilities of photothermolysis for transdermal delivery of a photosensitizer to the laboratory animals skin.

To study the range of application of “Chlorin e6” fluorescence effects for evidance the usefulness of its transdermal transport through non-ablative laser photothermolysis.

METHODS AND MATERIALS

In the work, 3-month-old male mice were used. They were divided into four groups:

The first group (16 individuals) - fractional laser photothermolysis (FLP) of the skin of the abdominal wall with a “Lakhta Milon” semiconductor laser (λ = 970nm) with an average power of 4W in a scanning pulse-periodic mode (t = 10ms / 10ms) through a quartz-polymer fiber with a focuser to form the epidermis of microthermal zones (MTZ) with a density of 16 MTZ / cm^2 with a dose density of energy in the pulse of 0.25 J / mm^2

Thus, 16 microthermal zones (MTZ) with a diameter of 0.4 mm are formed in the epidermis over an area of 1 sq cm.

The second group (16 individuals) - is performed by FLP, after which the photosensitizer Radagel {0.5% -0.1g (n = 10)} is applied to the skin for 20 minutes. Then, a “Lakhta Milon” semiconductor laser (λ = 662nm) is irradiated with an average power of 2 W in a pulse-periodic mode (t = 100 / 100ms) through an optical fiber with a light spot size of 1, 2 mm^2 on the skin.

The third group (16 individuals) - the photosensitizer Radagel {0.5% -1g (n = 10)} is applied to the skin of the animal for 20 minutes. Then, a “Lakhta Milon” semiconductor

laser (λ = 662nm) is irradiated with an average power of 2 W in a pulse-periodic mode (t = 100 / 100ms) through an optical fiber with a light spot size on the skin of 1, 2 mm^2

The fourth group (4 individuals) - control without exposure.

After the completion of laser irradiation, one individual from every experimental group undergoes intravital fluorescence diagnostics in the blue spectrum of 406 nm and in the red spectrum of 630 nm on a “Fluovisor” instrument.

Then the animals were euthanized by the method of cervical vertebrae dislocation, the anterior abdominal wall was dissected, 4 skin flaps were taken, one was sent for electron microscopy, one for histological examination, and transverse sections was made from two flaps using a microtome (cryotome) for subsequent confocal laser microscopy LSM 510 scanning microscope to record Radagel fluorescence effects.

A similar sequence of studies was carried out on the 7th, 14th and 28th days.

It complies with the provisions of the Federal Law on Amendments to Article 245 "Cruelty to Animals." Entered into force on December 20, 2017. Russia.

RESULTS

On the results of fluorescence diagnostics, in conclusion we can say that the highest fluorescence of the Radagel preparation was recorded in the skin after fractional photothermolysis followed by application photosensitization, which was indicated a pronounced intradermal accumulation of the photosensitizer in comparison with the usual application to the skin.

Laser confocal microscopy was carried out at the optimal wavelengths for investigation: excitation of fluorescence at 458 nm, registration of fluorescence at 675 nm.

In reseach of an individual with FLP, no changes in the structure of the skin were detected, except for light loosening of collagen fibers. Also, the absence of autofluorescence of skin structures is diagnostically important.

According to the results individuals from the group prepared for PFLP, it can be seen that the largest part of the photosensitizer accumulated in the epithelium and papillary layer of the skin, but the intense fluorescence recorded by us in collagen fibers, skin structures (hair and sebaceous follicles), hypodermis and muscle fibers indicates that preparation of the skin by fractional laser photothermolysis creates the conditions for active transdermal transport of the photosensitizer. In the traditional application the drug Radagel on the skin surface, its penetrating ability is limited by the epithelial layer, which implies that the induced photodynamic reaction will have exclusively surface effects.

CONCLUSION

The idea of controlled transdermal photosensitizer transport for the implementation of medical technology of photodynamic fractional laser photothermolysis was theoretically substantiated.

For the first time in the conditions of an experiment, the possibility of creating transport channels in the skin for intra- and transdermal transport of drugs was successfully demonstrated using the drug photosensitizer Radagel.

The ability of non-invasive fractional laser photothermolysis to provide conditions for efficient transdermal transport of the photosensitizer to all skin layers during application is proved experimentally.

This option of local photosensitization of the skin is manageable and, depending on the dose of the photosensitizer used, can be used in cosmetology, dermatology, gynecology, oncology.

DOI 10.18699/BGRS/SB-2020-301

Nonvascular pathways of aqueous humor outflow in the choroid of the human eye

Sabina Nogovitsina
Laboratory ultrastructural research
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russian Federation
nogovitsina.niikel@gmail.com

Nataliya Bgatova
Laboratory ultrastructural research
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russian Federation
n_bgatova@ngs.ru

Alena Eremina
Science department
The academician S.N. Fyodorov Federal State Institution National Medical Research Center «Intersectoral Research and Technology Complex «Eye microsurgery» Ministry of Health of the Russian Federation», Novosibirsk Branch
Novosibirsk, Russian Federation
sci@mntk.nsk.ru

Valeriy Chernykh
The academician S.N. Fyodorov Federal State Institution National Medical Research Center «Intersectoral Research and Technology Complex «Eye microsurgery» Ministry of Health of the Russian Federation», Novosibirsk Branch
Novosibirsk, Russian Federation
rimma@mntk.nsk.ru

Aleksandr Trunov
Science department
The academician S.N. Fyodorov Federal State Institution National Medical Research Center «Intersectoral Research and Technology Complex «Eye microsurgery» Ministry of Health of the Russian Federation», Novosibirsk Branch
Novosibirsk, Russian Federation
trunov1963@yandex.ru

Vladimir Konenkov
Laboratory of clinical immunogenetics
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russian Federation
konenkov@soramn.ru

***Abstract* — The structural organization of lymphatic-like channels in the human eye choroid was studied by using immunohistochemistry and transmission electron microscopy. It was shown that these channels lining fibroblast-like cells that express markers of lymphatic endothelium. (*Abstract*)**

Keywords — fibroblast-like cells, conjunctiva, choroid, lymphatic endothelium, lymphatic channels, human eye

INTRODUCTION

The question about presence of lymphatic outflow from human eye is uncertain until now. For a long time it was supposed that there were no lymphatics in the eye. But after the revelation of molecular markers of lymphatic endothelial cells it became possible to detect the lymphatic capillaries and vessels in the various structures of the human eye in normal and pathological conditions. Nowadays there are some data about lymphatic structures in the anterior eye segment, including ciliary body [1, 2]. This structures possibly takes a part in unconventional pathway of aqueous humor drainage. However, other investigators were not able to reproduce these results [3, 4].

Currently, the question of the presence of lymphatic vessels in the posterior eye segment remains debatable. There are several researches about lymphatic structures in the choroid of the human eyes [5, 6]. Despite the evidence of the existence of lymphatic vessels in the choroid, the structural organization of these lymphatic vessels is not described.

Therefore, the purpose of this research was to study the structural organization of the lymphatic pathways of aqueous humor outflow in the choroid.

METHODS AND ALGORITHMS

The object of study were fragments of enucleated eyes (n = 10), harvested from patients with normal intraocular pressure. For morphological study of the samples of the conjunctiva and the choroid were treated according to standard procedures for light and transmission electron microscopy. Paraffin sections were immunostained with the lymphatic specific endothelial markers Podoplanin (Monosan, Netherlands), Prox-1 (Covance, Germany), and LYVE -1 (Abcam, England), the vascular specific endothelial marker CD31 (Abcam, England), a marker for fibroblast growth factor receptors FGFR-1 (Abcam, England) and fibroblast marker TE-7 (Novus Biologicals, USA). Stained paraffin sections were examined with a light microscope "Leica DME» (Germany). Ultrathin sections were examined in an electron microscope «JEM 1400» (Japan). Morphometric processing and statistical analysis of the results were performed.

RESULTS

The presence of lymphatic vessels in conjunctiva was verified and it have typical organization, so we used conjunctiva as a control tissue.

With simultaneous immunohistochemical staining of conjunctiva and choroid samples using antibodies to markers of lymphatic vessels (LYVE-1, Prox-1, Pooplanin), in the structure of the conjunctiva revealed positively stained typical lymphatics, and positively stained lymphatic-like structures in the choroid. The choroid was found to contain lymphatic channels in choroidal stroma and layer of choriocapillaris, and lymphatic lacunae in the suprachoroid lamina (the transition zone between choroid and sclera). These channels and lacunae weren't typical lymphatic vessels. The channels were limited by narrow elongated cells, positively stained for markers of lymphatic vessels Prox-1, LYVE-1, Podoplanin and were shown negative CD31 staining; therefore we consider that this channels and lacunae are lymphatic structures.

Using transmission electron microscopy we identified that lymphatic channels did not have a typical vascular structure. We shown that these channels circumscribed by narrow process cells that have ultrastructural differences from

lymphatic endothelial cells: they do not have stropic filaments, contain a greater volume density of the granular endoplasmic reticulum membranes, and a lower volume density of mitochondria and caveolae.

The cells lining the lymphatic channels are similar to fibroblasts, but differ from them in the lower bulk density of the granular endoplasmic reticulum and the lower content of fibrous structures in the microenvironment.

We carried out an immunohistochemical study of the samples using fibroblast markers FGFR-1 and TE-7. During immunohistochemical staining of samples using the TE-7 marker, rare positively colored elongated narrow cells were detected in the choroid of the human eye. As a result, staining of the same cells was noted as when staining for lymphatic endothelial markers. However, not all cells lining the channels studied by us were stained for fibroblast markers TE – 7 and FGFR – 1.

Considering that these cells have ultrastructural differences from fibroblasts and lymphatic endothelial cells, they express lymphatic endothelial markers and do not always have fibroblast markers, i.e. these cells are fibroblast-like cells.

CONCLUSION

Thus, in the structure of the choroid of the human eye, there are non-vascular pathways of aqueous humor outflow, represented by untypical lymphatic channels confined by fibroblast-like cells that have markers of the endothelium of the lymphatic vessels.

REFERENCES

[1] Y. H. Yücel M. G. Johnston, T. Ly, M. Patel, B. Drake, E. Gümüş, S. A. Fraenkl, S. Moore, Tobbia, D. Armstrong, E. Horvath and N. Gupta, “Identification of lymphatics in the ciliary body of the human eye: a novel “uveolymphatic” outflow pathway,” Exp Eye Res., Vol. 89, pp. 810–819, 2009.J.

[2] M. Kim, M. G. Johnston, N. Gupta, S. Moore and Y. H. Yücel, “A model to measure lymphatic drainage from the eye,” Exp Eye Res., Vol. 93, pp. 586–591, 2011..

[3] K. Birke, E. Lütjen–Drecoll, D. Kerjaschki and M. Birke, “Expression of podoplanin and other lymphatic markers in the human anterior eye segment,” Invest Ophthalmol Vis Sci., Vol. 51, pp. 344–354, 2010.

[4] A. Kaser–Eichberger, F. Schrödl, A. Trost, C. Strohmaier, B. Bogner, C. Runge, K. Motloch, D. Bruckner, M. Laimer, S. L. Schlereth, L. M. Heindl and H. A. Reitsamer, “Topography of lymphatic markers in human iris and ciliary body,”. Invest Ophthalmol Vis Sci., Vol. 56, pp. 4943–4953, 2015.

[5] M. E. Koina, L. Baxter, S. J. Adamson, F. Arfuso, P. Hu, M. C. Madigan and T. Chan–Ling, “Evidence for lymphatics in the developing and adult human choroid,” Invest Ophthalmol Vis Sci., № 56, Vol. 2, pp. 1310–1327, 2015.

[6] V.V. Chernykh, Yu.I. Borodin, N.P. Bgatova, S.R. Nogovitsina, I.A. Toporkov, A.V. Eremina, A.N. Trunov and V.I. Konenkov, “Lymphatic structures in the ciliary body and choroid of the human eye,” Sovremennye tekhnologii v oftal'mologii, Vol. 3, pp. 268, 2016. (In Russ.)

DOI 10.18699/BGRS/SB-2020-302

Human retinal photoreceptor cells in glaucoma: destructive changes of mithochondria and mitophagy

Natalia Obanina
Laboratory of ultrastructural research, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences;
Section of cytology and genetics, Department of Natural Sciences, Novosibirsk State University
Novosibirsk, Russia
n.obanina@g.nsu.ru

Nataliya Bgatova
Laboratory of ultrastructural research, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
n_bgatova@ngs.ru

Valerii Chernykh
The academician S.N. Fyodorov Federal State Institution "Intersectoral Research and Technology Complex "Eye Microsurgery", Ministry of Healthcare of Russian Federation, Novosibirsk Branch,
Novosibirsk, Russia
rimma@mntk.nsk.ru

Aleksandr Trunov
The academician S.N. Fyodorov Federal State Institution "Intersectoral Research and Technology Complex "Eye Microsurgery", Ministry of Healthcare of Russian Federation, Novosibirsk Branch,
Novosibirsk, Russia
trunov1963@yandex.ru

Alena Eremina
The academician S.N. Fyodorov Federal State Institution "Intersectoral Research and Technology Complex "Eye Microsurgery", Ministry of Healthcare of Russian Federation, Novosibirsk Branch,
Novosibirsk, Russia
sci@mntk.nsk.ru

Vladimir Konenkov
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
vikonenkov@gmail.com

***Abstract* — The leading place in the development of low vision and blindness worldwide is occupied by glaucoma. This is a disease of non-infectious nature, accompanied by an increase in intraocular pressure and degeneration of retinal neurons, optic nerve, as well as its spread in the central section of the visual analyzer. The aim of the study was to identify features of the ultrastructure of the human retinal photoreceptor cells in glaucoma. The study was carried out using transmission electron microscopy and morphometry. The volume densities (Vv) of mitochondria and autophagosomes containing mitochondria were determined. Disorders of orientation and swelling of rod neurons membrane disks in the outer segments were detected. It is shown that rods are more exposed to destructive changes than cones.**

Keywords* — *glaucoma, photoreceptors, ultrastructure, mitophagy

INTRODUCTION

Glaucoma is a multifactorial disease characterized by progressive damage and death of retinal ganglion cells, axon atrophy, and, as a result, optic neurodegeneration [2, 3]. The process of neurodegeneration distributed outside of the retina and optic nerve to the central section of the visual analyzer [2,4,5,6]. It is known that glaucoma-associated cell death occurs by apoptosis. Apoptosis is caused by oxidative stress through mitochondrial damage, inflammation, endothelial dysregulation and dysfunction, and hypoxia [7]. Autophagy is a complexly organized process that is necessary for any eukaryotic cell for normal life. Dysfunction of autophagy at different stages in neurons leads to the emergence of many neurodegenerative diseases, in particular – glaucoma [8]. The pathogenesis of glaucoma is being studied on various animal models (mice, rats, rabbits, cats, monkeys, etc.). For example, it was shown that in the model of glaucoma, destructive changes in mitochondrial cristae develop in the rats retina. The swelling of the mitochondria is in the cytoplasm of neurons of the inner nuclear and inner plexiform layers (in ganglion cells), and photoreceptor pigment cells [9]. However, data obtained in animal models cannot be fully interpreted for the human eye due to anatomical and physiological differences. There are existing data about damage and ultrastructural changes in retinal ganglion cells, inner nuclear layer, inner plexiform layer [3, 10], but there is no information about ultrastructural changes in the photoreceptors in glaucoma.

METHODS AND ALGORITHMS

The study was conducted in accordance with the principles of the Helsinki Declaration of the World Medical Association "Ethical Principles for Scientific Medical Research with the Participation of Human", Federal Law of the Russian Federation of November 21, 2011 № 323 "About the Basics of Protecting the Health of Citizens in the Russian Federation", and the requirements Federal Law of July 27, 2006 № 152 (as amended on July 21, 2014) "About Personal Data" (as amended and supplemented, entered into force on September 1, 2015). The object of the study were retinal fragments of patients enucleated by medical indications of the eye with terminal stage of glaucoma of the Novosibirsk branch of Academician S.N. Fyodorov Federal State Institution «Intersectoral Research and Technology Complex «Eye Microsurgery», Ministry of Healthcare of Russian Federation. The study was conducted with the permission of the bioethical committee. The written informed consent of the patients to study the biological material was obtained. The collection of material was carried out by employees of the Novosibirsk branch of the IRTC «Eye Microsurgery».

Fragments of the human retina (about 1 mm^3) were fixed in a 4% solution of paraformaldehyde prepared on Hanks medium, fixed for 1 hour in a 1% solution of OsO_4 in phosphate buffer (pH=7.4), dehydrated in increasing concentration of ethanol, and then enclosed in Epon (Serva, Germany). Semi-thin sections with thick 1 μm were prepared

from the obtained blocks on a ultratome Leica UC7/FC7 (Germany / Switzerland), stained with toluidine blue. Then we studied under a light microscope "LEICA DME" and selected the necessary tissue sites for research in an electron microscope. Ultrathin sections thick 70–100 nm were obtained from the selected material, contrasted with a saturated aqueous solution of uranyl acetate and lead citrate, and studied using a electron microscope JEM 1400 (Japan). Microscopic analysis was carried out at the Multiple-access Center for Microscopy of Biological Subjects (Institute of Cytology and Genetics, Novosibirsk, Russia). Morphometry was performed using Image J software (Wayne Rasband, USA) with a closed test system at x25K magnification. The mean (M) and standard deviation (SD) were calculated using Microsoft Excel software. The significance of differences between the parameters was determined by Statistica 6.0 software (StatSoft, Inc.) with Mann-Whitney U-test. The differences were considered significant at $p<0.05$.

RESULTS

The volume densities (V_V) of mitochondria and autophagosomes containing mitochondria in ellipsoids of both rod and cone cells inner segment were counted. In the rod cells outer segments swelling of membrane disks and their disorganization were noted. Disorder of the disk's structure contributes to the destruction of the transmembrane proteins of the visual cascade, which grab the photons. It reduces the photosensitive function of cells. In the rod ellipsoids turgescence of mitochondria was noted, the volume density of which was 4%. Mitochondrial swelling was characterized by a multiple increase of area and a decrease in electron density of the mitochondrial matrix. Volume density of mitochondria with destructive changes of cristae was 20%. Autophagosomes with mitochondria were observed (4.5%). In the outer segments of cone cells the membrane semidisks were located more orderly. Mitochondrial swelling was not observed in their ellipsoids, single autophagosomes were noted (0.6%). Cell death associated with glaucoma occurs through apoptosis [7]. It is initiated by many factors: a) relating to mitochondria directly (oxidative stress); b) affecting the development of apoptosis indirectly (Fas-mediated signaling pathway, the effects of neurotoxic proteins: amyloid beta and tau protein) [7, 11]. During the study, apoptosis was not detected in the human photoreceptors. However, noted the autophagic activity may lead to the death of photoreceptors.

CONCLUSION

In our study mitochondria with destructive changes in cristae were detected both in rod and cone cells ellipsoids of all types of photoreceptors. In rods also present mitochondrial swelling and mitophagy. It was revealed that with glaucoma, rods are subject to greater disturbances in contrast to cones. It indicates their more pronounced reaction on an increase of intraocular pressure. Probably, autophagy can plays a dual role in retinal cells, on the one hand, contribute to the survival of cells by removing damaged structures, and on the other hand, initiate the apoptosis, which requires further research.

ACKNOWLEDGMENT

This work was supported with financing of the Novosibirsk Research Institute of Clinical and Experimental Lymphology as part of a state order, No. 0324-2019-0045.

REFERENCES

[1] V.V. Chernych, N.P. Bgatova, "Lymphatic structures of the human eye and their changes in pimary open-angle glaucoma," Ophthalmology, Moscow, 80 pp., 2019.

[2] V.P. Erichev, V.P. Tumanov, L.A. Panushkina, "Glaucoma and nuerodegererative diseases," Glaucoma, Vol. 1, pp. 62–68, 2012.

[3] M. Almasieh, A.M. Wilson, B. Morquette, J.L. Cueva Vargas and A. Di Polo, "The molecular basis of retinal ganglion cell death in glaucoma," Progress in Retinal and Eye Research, vol. 31(2), pp. 152–181, 2012.

[4] J.A. Ghiso, I. Doudevski, R. Ritch, A.A. Rostagno, "Alzheimer's Disease and Glaucoma: Mechanistic Similarities and Differences," Journal of Glaucoma, Vol. 22, pp. 36–38, June-July 2013.

[5] F.G. Garaci, F. Bolacchi, A. Cerulli, M. Melis, A. Spanò, C. Cedrone, R. Floris, G. Simonetti, C. Nucci, "Optic Nerve and Optic Radiation Neurodegeneration in Patients with Glaucoma: In Vivo Analysis with 3–T Diffusion–Tensor MR Imaging," Radiol, Vol. 252 (2), pp. 496–501, 2009.

[6] M.D. Pinazo-Duran, K. Shoaie-Nia, V. Zanon-Moreno, S.M. Sanz-Gonzalez, J.B. Del Castillo, J.J. Garcia-Medina "Strategies to reduce oxidative stress in glaucoma patients," Curr Neuropharmacol, vol. 16(7), pp. 903–918, 2018.

[7] I.R. Gazizova, I.Yu. Tikhomirova, "The role of mitochondrial dysphunction in glaucoma," Medical bulletin of Bashkortostan, Vol. 10(2), pp. 153–156, 2015.

[8] I.A. Kochergin, M.N. Zakharova, "The Role of Autophagy in Neurodegenerative Diseases," Neurochemistry, vol. 33(1), pp. 12–24, 2016.

[9] H. Celicer, N. Yuksel, S. Solakoglu, L. Karabas, F. Aktar , Y. Caglar, "Neuroprotective Effects of Memantine in the Retina of Glaucomatous Rats: An Electron Microscopic Study," J Ophthalmic Vis Res., vol. 11(2), pp. 174–82, April-June 2016.

[10] T.C. Nag, and S. Wadhwa, "Ultrastructure of the human retina in aging and various pathological states," Micron, vol. 43(7), pp. 759–781, 2012.

[11] E.A. Egorov, V.N. Alekseev, I.R. Gazizova, E.B. Martynova, "Mitochondrial morphological changes of tracular cells in patients with primary open-angle glaucoma," Russian Journal of Clinical Ophthalmology, vol. 16(3), pp. 137-139, March 2016.

DOI 10.18699/BGRS/SB-2020-303

The main phenotypic parameters in the differential diagnosis of monogenic forms of diabetes mellitus and type 2 diabetes in young people

Alla Ovsyannikova
Research Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
aknikolaeva@bk.ru

Oksana Rymar
Research Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
orymar23@gmail.com

Dinara Ivanoshchuk,
Research Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
dinara2084@mail.ru

Elena Shakhtshneider
Research Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia
2117409@mail.ru

***Abstract* — Diagnosis of the correct type of diabetes mellitus (DM) in young patients leads to the appointment of pathogenetic effective therapy, adequate management of pregnancy and the prevention of specific complications. The aim of the study was to investigate main clinical and laboratory criteria for differential diagnosis of type 2 DM (T2DM) and monogenic type (MODY diabetes). We examined 108 people in Novosibirsk city who had debut of DM up to 35 years: 92 patients has been confirmed by molecular - genetic testing MODY, 91 - type 2. All patients had a clinical examination, blood sampling for biochemical and hormone (TSH, C-peptide) analyzes. The main criteria for a differential diagnosis between T2DM and MODY are identified after conducting a regression analysis which can help determine the need for genetic testing. Results: Both groups of patients were comparable in gender, age, and duration of diabetes. Clinical parameters (including heredity) and laboratory parameters in the diagnosis of diabetes were analyzed. The most significant signs ($p < 0.001$) associated with the type of diabetes which were used to create the model of binary logistic regression were determined. The following factors were highly reliable: the presence of relatives with diagnosed diabetes under the age of 35 years (associated with MODY), the presence of relatives with diabetes and overweight, and overweight in the studied patients (associated with type 2 diabetes). These results should be taken into account when conducting differential diagnosis of the type of diabetes mellitus in individuals with diagnosed hyperglycemia up to 35 years.**

Keywords* — *diabetes mellitus, diabetes mellitus type 2, MODY, genetic testing, young patients

Motivation and aim

Most young patients with hyperglycemia are diagnosed with type 1 diabetes mellitus and type 2 diabetes mellitus (T2DM) but up to 10% of all cases of the disease occur in MODY diabetes [1]. Diagnosis of the correct type of diabetes mellitus (DM) leads to the appointment of pathogenetic effective therapy, adequate management of pregnancy and the prevention of specific complications. The aim of the study was to create a model for the differential diagnosis of T2DM and MODY. The aim of the study was to investigate main clinical and laboratory criteria for differential diagnosis of T2DM and monogenic type (MODY diabetes).

Methods

We examined 108 people in Novosibirsk city who had debut of DM up to 35 years: 92 patients has been confirmed by molecular - genetic testing MODY, 91 - type 2. MODY was identified by target sequencing was performed on the genes associated with known subtypes of MODY, 1 to 14. All patients had a clinical examination, blood sampling for biochemical and hormone (TSH, C-peptide) analyzes. A model was constructed to study the most significant factors during the diagnosis of type of diabetes using binary logistic regression.

Results

Both groups of patients were comparable in gender, age, and duration of diabetes. Median of age of the patients at moment of examination was with T2DM - 29,5 [17,7;37,0] years, with MODY - 24,0 [0,0;35,0] years ($p = 0{,}280$). After creating various regression models the most significant factors associated with diabetes were identified: “the presence of relatives with diagnosed diabetes for them under 35 years of age”, “the presence of relatives with impaired carbohydrate metabolism and obesity”, “overweight and obesity”. A final regression model is constructed to determine the significance of these factors. Among the analyzed factors associated with the type of diabetes all three were statistically significant ($p < 0.001$).

Conclusions

1. It is necessary to take into account the presence of a burdened family history of impaired carbohydrate metabolism and overweight in the differential diagnosis of type of diabetes in people with a debut of the disease under 35 years of age.
2. The presence of excess weight in young patients is associated with T2DM.

Acknowledgment

The research work was carried out within the framework of the budget topic No. AAAA-A17-117112850280-2.

References

[1] Lachance C.H. Practical Aspects of Monogenic Diabetes: A Clinical Point of View // Can. J. Diabetes. 2016. Vol. 40. P. 368–375.

DOI 10.18699/BGRS/SB-2020-304

Whole genome of novel Lactobacillus fermentum HFD1 strain producing various antimicrobial metabolites

Georgii Ozhegov
Kazan Federal University,
Kazan, Russia
Perm State Pharmaceutical Academy,
Perm, Russia
georgii_provisor@mail.ru

Monyr Nait Yahia
Kazan Federal University,
Kazan, Russia

Dina Yarullina
Kazan Federal University,
Kazan, Russia

Natalya Gogoleva
Kazan Federal University,
Kazan Institute of Biochemistry and Biophysics, FRC Kazan Scientific Center of RAS, Kazan, Russia

Airat Kaumov
Kazan Federal University,
Kazan, Russia
kairatr@yandex.ru

Alexey Vasilchenko
University of Tyumen,
Tyumen, Russia

***Abstract* — Lactobacilli, generally recognized as safe for humans, are characterized with high antagonistic properties. Among 40 strains of Lactobacilli isolated from the faeces of healthy humans, *Lactobacillus fermentum* HFD1 exhibited significant antimicrobial activity against various nosocomial pathogens. The genome of *L. fermentum* HFD1 was sequenced on Illumina MiSeq and ONT MinION instruments. After reads assembly 2 circular contigs with respective sizes of 2101878 bp and 5386 bp have been obtained. The raw reads alignment with BWA and post-assembly assessment showed that 99.9% of Illumina reads and 99.2% of MinION reads were mapped to assembled genome. The genome was read with 575x coverage for Illumina data and 454x coverage for MinION data. Based on genome alignment *L. fermentum* FTDC8312 was found as closest related strain. The BLAST search identified the short contig as similar to bacterophage phiX174. In the genome 2120 coding sequences (CDS) were predicted. Among them 606 were annotated as "hypothetical" without known biological role. By using CAMPr3, ADAM and AMPA services 5 CDSs with highest summary prediction score were identified as coding for antimicrobial peptides. Further validation of their antimicrobial activity in vitro is required.**

Key words* — *Lactobacilli, de novo assembly, antibacterial peptides.

Introduction

Lactobacilli, generally recognized as safe for humans, are characterized with high antagonistic properties. Besides some low-molecular agents, such as short-chain fatty acids, H_2O_2 and etc., *Lactobacilli* produce various antimicrobial peptides (AMPs) [1]. The resistance development to AMPs generally takes much longer period when compared to traditional antibiotics, that makes AMP promising antimicrobials considering rapid spread of drug-resistant pathogenic bacteria the screening for new antimicrobial agents. Among 40 strains of *Lactobacilli* isolated from the faeces of healthy humans, *Lactobacillus fermentum HFD1* exhibited significant antimicrobial activity against various nosocomial pathogens [2]. From the culture liquid of *L. fermentum* HFD1 several antimicrobial compounds repressing growth of *P. aureginosa* were isolated by using reverse-phase chromatography. The chemical structure and nature of these antimicrobial metabolites remain to be identified and for this purpose the whole genome sequence is strictly required.

Materials and methods

DNA was extracted by phenol-chloform method. Sequencing was performed on Illumina MiSeq and ONT MinION instruments. Hybrid assemblers Unicycler v0.4.8-beta [3] and SPAdes v3.13.0 [4], and long-read assembler Canu 1.8 [5] combined with short-read polishing with Pilon v. 1.23 [6] have been used. Genome has been annotated by using PROKKA 1.14.5 software [7]. Putative genes and genes clusters responsible for antimicrobial peptides synthesis were predicted by using ADAM [8], AMPA [9] and CAMPr3 [10] web-services.

Results

The combined sequencing on Illumina MiSeq and ONT MinION platforms allowed use advantages of both platforms. Illumina sequencing yielded 5 million filtered pair-end reads obtaining 1.2 Gbp data. MinION generated 530 thousand reads with cumulative length of 1Gbp and maximum read length 53364. By comparing different strategies, Unicycler exhibited better performance whereas the assembly quality was the same for all assemblers. After assembly 2 circular contigs with respective sizes of 2101878 bp and 5386 bp has been obtained. The raw reads alignment with BWA and this post-assembly assessment showed that 99.9% of Illumina reads and 99.2% of MinION reads were mapped to assembled genome. Error rate was 8×10^{-4} for Illumina and 0.1 for MinION. Finally, the genome was read with 575× coverage for Illumina data and 454× coverage for MinION data. Based on genome alignment with KmerFinder [11] *L. fermentum FTDC8312* was found closest related strain. The BLAST search identified the short contig as similar to bacterophage phiX174.

Genome annotation was performed by using PROKKA software. In total 2120 coding sequences (CDS) were predicted. Among them 606 were annotated as "hypothetical" without known biological role. Data available in Genbank database by accession number CP050919. By using CAMPr3 (protocols SVM, RF, Ann DAC), ADAM and AMPA services 5 CDSs with highest summary prediction score were identified as coding for antimicrobial peptides with length from 56 to 133 amino acids. Further validation of their antimicrobial activity in vitro is required.

ACKNOWLEDGMENTS

This research was funded by the Russian Foundation for Basic Research (grant number 17-00-00456) and by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, project No. 13556.2019/13.1

REFERENCE

[1] Gavrilova E. et al. (2019). Newly isolated lactic acid bacteria from silage targeting biofilms of foodborne pathogens during milk fermentation. *BMC microbiology*, *19*(1), 248.

[2] Pavlova A. S. et al. (2020). Identification of Antimicrobial Peptides from Novel Lactobacillus fermentum Strain. *The Protein Journal*, 1-12.

[3] Wick R. R., Judd L. M., Gorrie C. L., Holt K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS computational biology*, *13*(6), e1005595.

[4] Bankevich A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, *19*(5), 455-477.

[5] Koren S., Walenz B. P., Berlin K., Miller J. R., Bergman N. H., Phillippy A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research*, *27*(5), 722-736.

[6] Walker B. J. et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one*, *9*(11).

[7] Seemann T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068-2069.

[8] Lee H. T., Lee C. C., Yang J. R., Lai J. Z., Chang K. Y. (2015). A large-scale structural classification of antimicrobial peptides. *BioMed research international*, *2015*.

[9] Torrent M., Nogués V. M., Boix E. (2009). A theoretical approach to spot active regions in antimicrobial proteins. *BMC bioinformatics*, *10*(1), 373.

[10] Waghu F. H., Barai R. S., Gurung P., Idicula-Thomas S. (2016). CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic acids research*, *44*(D1), D1094-D1097.

[11] Larsen, M. V. et al. (2014). Benchmarking of methods for genomic taxonomy. *Journal of clinical microbiology*, *52*(5), 1529-1539.

DOI 10.18699/BGRS/SB-2020-305

MBP-hydrolyzing abzymes as a peripheral markers associated with impaired myelination in schizophrenia

Daria Parshukova
Mental Health Research Institute
Tomsk National Research Medical
Center RAS
Tomsk, Russia
Susl2008@yandex.ru

Liudmila Smirnova
Mental Health Research Institute
Tomsk National Research Medical
Center RAS
Tomsk, Russia
Lpsmirnova@yandex.ru

Ekaterina Dmitrieva
Mental Health Research Institute
Tomsk National Research Medical
Center RAS
Tomsk, Russia
Egdtomsk@mail.ru

Arkady Semke
Mental Health Research Institute
Tomsk National Research Medical
Center RAS
Tomsk, Russia
Asemke@mail.ru

Vasily Yarnykh
University of Washington, Department
of Radiology,
Seattle WA, USA

Svetlana Ivanova
Mental Health Research Institute
Tomsk National Research Medical
Center RAS
Tomsk, Russia
Ivanovaniipz@gmail.com

Abstract — **It is well-known that the pathology of myelin and oligodendrocytes is involved in the pathogenesis of schizophrenia. Anomalies in oligodendrocytes and myelin can be a source of neuronal disruption. The discovery of catalytic antibodies (abzymes) allows us to investigate their pathological role in various conditions. One of the possible ways of inducing proteolytic antibodies is the appearance in the peripheral blood of a substrate in the form of damaged protein fragments. In our previous study, it was shown that the IgG of schizophrenia patients can hydrolyze the myelin basic protein - one of the main components of the central nervous system myelin. The study of IgG proteases in accessible biomaterial (serum) and their association with myelination disturbance in schizophrenia may be potential criteria for monitoring the severity of mental disorders.**

Keywords — myelin; abzymes; schizophrenia; neuroimaging; MBP

MOTIVATION AND AIM

Numerous data from intravital neuroimaging, as well as genetic and morphological studies, indicate impaired myelination and pathology of oligodendrocytes in the brain of patients with schizophrenia. Post-mortem studies in schizophrenia using electron microscopy and histochemistry confirm structural changes in oligodendroglia cells producing the myelin sheath [1], atrophy of axons, abnormalities of the myelin sheath surrounding [2], reduce in myelin basic protein in the anterior frontal cortex of cases of schizophrenia [3] and an increase in the antibody titer to myelin basic protein. Recently we reported about autoantibodies against myelin basic protein (MBP) which had catalytic activity and were capable to hydrolyze myelin basic protein and its peptides [4]. The level of MBP-hydrolyzing activity in patients with schizophrenia significantly exceeded the level of activity of antibodies from healthy donors and depended on the leading symptoms and duration of the disease.

Various neuroimaging techniques reported volume abnormalities in different areas of the brain in schizophrenia. MRI protocols allowing quantification of myelin density have been developed. Macromolecular proton fraction (MPF) mapping is a quantitative MRI method that reconstructs parametric maps of a relative amount of macromolecular protons causing the magnetization transfer (MT) effect and provides a quantitative biomarker of myelin in neural tissues [5]. This study focuses on investigation the level of activity of MBP-hydrolyzing antibodies and myelin density in patients with schizophrenia.

MATERIALS AND METHODS

Patients

In this work, 15 patients with paranoid schizophrenia and 14 healthy donors of match age were recruited to study. The inclusion criteria were the following: or simple schizophrenia according with the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10: F20.0). The exclusion criteria were the following: the presence of acute and chronic infectious, inflammatory, autoimmune, or neurological diseases, other organic mental disorders (e.g., epilepsy, Parkinson's disease) and mental retardation. The patients were recruited from the Mental Health Research Institute TNMRC, Department of Endogenous Disorders (Tomsk, Russia).

Purification of immunoglobulins G from schizophrenia patients' serum

Antibodies from the blood sera of schizophrenia patients and healthy controls were purified and analyzed by earlier developed procedures for purification of electrophoretically and immunologically homogenous IgG preparations from human blood serum. The procedure included affinity chromatography of serum proteins on Protein G-Sepharose followed by high-performance gel filtration on a Superdex-200 HR 10/30 column.

Evaluation of Proteolytic activity of serum IgGs

IgG of schizophrenia patient were incubated with MBP at 37C during 5-14 hours. The catalytic activity of the IgG in the cleavage of MBP was estimated from the decrease in the intensity of Coomassie-stained MBP band after electrophoresis. The activity of the Abs is expressed in units of specific enzyme activity as quantity mg of substrate cleavage by 1 mg Abs per 1 h. Evaluation of IgG-depended proteolytic activity was performed during the period of exacerbation of clinical symptoms.

In vivo determination of myelin density by mapping of the molecular proton fraction using MRI

Images were acquired by using a 1.5-T imager (Magnetom Essenza; Siemens, Germany) with implication of fast three-dimensional whole-brain MPF mapping protocol [5]. Statistical analyses were performed in the program Statistica

10.0. To take into account partial volume effects, four tissue classes were prescribed within the single-channel segmentation procedure. These classes correspond to pure weight matter (WM), pure gray matter (GM), mixed voxels that contained partial volumes from WM and GM (PVWGM), and mixed voxels that contained a partial volume from cerebrospinal fluid - boundary layer (BL).

RESULTS

In this work, we revealed a statistically significant decrease in MPF in global gray matter (p= 0,002), mixed substance (p=0,014), cortical gray matter (p=0,001) in patients with schizophrenia compared with the control. In addition, a statistically significant decrease in MPF was revealed in schizophrenic patients with predominant negative symptoms compared to the control group in global gray matter (p=0,001), white matter (p=0,01), mixed substance (p=0,004) and cortical gray matter (p<0,001). In our previous study, it was shown that the greatest MBP-hydrolyzing activity is associated with negative symptoms.

Since the indicators of proteolytic activity and MPF density in patients with schizophrenia showed similar changes in the analyzed groups, an analysis of the correlation relationships of these indicators was performed to assess the degree of their mutual influence. For this analysis the MRI scan and the investigation of proteolytic activity was performed in same group of patients. The results of the correlation analysis of myelin density and specific activity are shown in Table 30. Correlation analysis revealed a reliable negative correlation of average strength (r = -0.593; p = 0.019) between the level of specific activity of MBP-hydrolyzing IgG and myelin density (MPF) in the white matter of the brain.

The presented results show that catalytic antibodies with MBP-hydrolyzing activity can be used as peripheral markers which reflect a decrease in the density of myelin in the structure of the brain.

ACKNOWLEDGMENT

Support by Grant of RSF No. 18-15-00053 «Search of peripheral markers associated with impaired myelination of the brain and pathogenesis of schizophrenia» 2018-2020.

REFERENCES

[1] Martins-de-Souza D et al. (2010). Proteome and transcriptome analysis suggests oligodendrocyte dysfunction in schizophrenia. J. Psychiatr. Res. 44(3), 149-156.

[2] Uranova, N. A., Vostrikov, V. M., Orlovskaya, D. D., Rachmanova, V. I. (2004). Oligodendroglial density in the prefrontal cortex in schizophrenia and mood disorders: a study from the Stanley Neuropathology Consortium. Schizophr. Res. 67(2), 269-275.

[3] Matthews, P. R., 2012 23. Matthews, P. R., Eastwood, S. L., Harrison, P. J. (2012). Reduced myelin basic protein and actin-related gene expression in visual cortex in schizophrenia. PloS One. 7(6), e38211.

[4] Parshukova D.A. et al. (2019) Autoimmunity and immune system dysregulation in schizophrenia: IgGs from sera of patients hydrolyze myelin basic protein. J Mol Recognit. e2759 doi: 10.1002/jmr.2759

[5] Yarnykh, V. L. et al. (2016). Time‐efficient, high‐resolution, whole brain three‐dimensional macromolecular proton fraction mapping. Magnetic resonance in medicine. 75(5), 2100-2106.

DOI 10.18699/BGRS/SB-2020-306

Mechanisms of ischemic kidney tolerance in young and senescence-accelerated rats exposured to ischemic preconditioning or calorie restriction

Egor Plotnikov
A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University;
V.I. Kulakov National Medical Research Center of Obstetrics, Gynecology and Perinatology,
Moscow, Russia
plotnikov@belozersky.msu.ru

Irina Pevzner
A.N. Belozersky Institute Of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia
irinapevzner@mail.ru

Denis Silachev
A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University;
V.I. Kulakov National Medical Research Center of Obstetrics, Gynecology and Perinatology,
Moscow, Russia
silachev_dn@belozersky.msu.ru

Nadezda Andrianova
Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University,
A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University;Moscow, Russia
andrianova@belozersky.msu.ru

Ljubava Zorova
A.N. Belozersky Institute Of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia
lju_2003@list.ru

Natalia Kolosova
Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (SB RAS), Novosibirsk; Russia,
kolosova@bionet.nsc.ru

Stanislovas Jankauskas
A.N. Belozersky Institute Of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia
jankauskas.ss@gmail.com

Vasily Popkov
A.N. Belozersky Institute Of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia
popkov.vas@gmail.com

Dmitry Zorov
A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University;
V.I. Kulakov National Medical Research Center of Obstetrics, Gynecology and Perinatology,
Moscow, Russia
zorov@belozersky.msu.ru

***Abstract* — Calorie restriction (CR) and ischemic preconditioning (IPC) are the most efficient approaches ameliorating the severity of different pathological conditions. We investigated the protective potential of long-term calorie restriction and short ischemic preconditioning protocol in the model of acute kidney injury (AKI) in young Wistar and OXYS rats. In young rats, ischemic preconditioning, which consists of 4 cycles of ischemia and reperfusion alleviated kidney injury caused by 40-min ischemia. However, 6-month-old OXYS rats having signs of premature aging lost their ability to protect the ischemic kidney by IPC. However, CR of OXYS rats led to a significant decrease in creatinine and blood urea nitrogen levels after kidney ischemia, indicating significant protection.**

Keywords* — *kidney, ischemia, preconditioning, aging, autophagy

Introduction

The phenomenon of improving of tissue ischemic tolerance by the preliminary impact of various stress factors is a universal intrinsic protective mechanism inherent to a large number of organs and tissues, including the kidney. Induced ischemic tolerance of the kidney is of great interest for nephrology, since activation of tissue resistance signaling pathways can be simulated using a number of pharmacological agents. Currently, a number of nephroprotective effectors have been identified, such as IPC and CR, which demonstrate a pronounced therapeutic potential in experimental studies [1,2]. However, their effectiveness has been studied mainly in young animals and not tested on old ones, which makes it difficult to translate these approaches to the clinic, since the main group of patients suffering from renal pathologies are the elderly.

This may be one of the reasons why most clinical trials have failed for drugs that have shown their effectiveness in in vivo experiments. However, working with old animals is associated with a number of objective difficulties, so an important task is to find models that allow us to study the phenomenon of aging in vivo. One of these models is a strain of OXYS rats that show signs of premature aging. A detailed study of OXYS rats revealed as early as during the first 12 months of life an early manifestation of diseases that are usually observed in old age: retinopathy, similar to age-related maculodystrophy [3], osteoporosis [4], arterial hypertension and cardiomyopathy [5], sarcopenia [6], and neurodegenerative diseases, including Alzheimer's type [7]. It can be assumed that accelerated aging also affects the kidney, which makes these rats a unique model for studying the effect of aging on renal diseases.

The aim of this work was to study the effect of premature aging on the mechanisms of ischemic kidney tolerance.

Materials and Methods

The nephroprotective effect of CR and IPC in young and old animals in renal ischemia/reperfusion (I/R) was studied. The kidney was subjected to 40-min ischemia with preliminary IPC (4 cycles of 15 with ischemia and 15 with reperfusion) or CR. CR was performed for 4 weeks by limiting the amount of feed by 35% of the daily norm to 65% of the norm. To assess AKI severity, the concentration of serum

creatinine and blood urea nitrogen (BUN) in the blood was measured, and the level of NGAL protein in the urine was determined.

Results

Kidney I/R led to the development of AKI in young Wistar rats and OXYS rats, 48 hours after ischemia, we observed an increase in the concentration of creatinine and urea in the blood, and the level of NGAL in the urine, while in OXYS rats the severity of renal failure was higher, as indicated by higher increase of serum creatinine and BUN after I/R. Both types of nephroprotective treatments reduced the severity of AKI in young animals, which was manifested in a decrease in serum creatinine and BUN. However, the effects on OXYS rats were rather different. In OXYS rats, ischemia after IPC keep causing a significant increase in AKI markers, meaning that the protective effect of this exposure was not occurred. Conversely, CR of OXYS rats led to a significant decrease in creatinine and BUN levels after kidney ischemia.

To analyze the mechanisms that may be responsible for the loss of efficiency of nephroprotection, the processes of autophagy and mitophagy in the kidneys of young and old animals in norm and after CR or IPC exposure were studied. Autophagy activation was observed in young animals after I/R: the number of lysosomes increased in the proximal tubules (measured by LysoTracker Green staining), and the LC3I/LC3II ratio increased in the whole kidney homogenates. In rats, OXYS I/R did not lead to such changes. Measuring the LC3II/LC3I ratio and staining with a lysosomal probe showed that young animals after CR showed a more pronounced activation of autophagy compared to OXYS rats. The intensity of mitophagy, which was determined by the level of PINK-1 protein, was also higher in Wistar rats.

Discussion

A recent study has shown that the response of OXYS rats to fasting within 24 or 48 hours may differ in terms of expression of a number of genes associated with autophagy (Kozhevnikova et al, 2019). Furthermore, it was found that many effects coincide in direction, although gene expression differs in absolute values of increase/decrease. For example, after 24 hours of fasting, an increase in the expression of the Atg5, Becn1, and p62 genes was observed in both Wistar and OXYS rats, although the magnitude of the effect was different. We can assume that with prolonged dietary restriction, the effects of autophagy activation in OXYS rats may reach the threshold values necessary for adequate kidney protection. That is, despite the initial reduced effectiveness of autophagy and quality control pathways, long-term exposure compensates for this disadvantage. There is some evidence of this in the literature, since it was shown that in older animals with reduced autophagy efficiency, longer periods of CR had a more pronounced protective effect than short ones.

Conclusion

Thus, both protective approaches, CR and IPC, cause induction of ischemic kidney tolerance in young Wistar rats, but not in OXYS rats. Loss of efficiency of nephroprotection may be associated with disturbances of autophagy and mitophagy mechanisms. This ineffectiveness of autophagy may be critical for a short protective pulse (IPC), but prolonged activation stimuli (CR for a month) could induce it anyway.

Acknowledgment

The work was supported by the Russian science foundation grant #18-15-00058.

References

[1] Jankauskas, S.S.; Pevzner, I.B.; Andrianova, N.V.; Zorova, L.D.; Popkov, V.A.; Silachev, D.N.; Kolosova, N.G.; Plotnikov, E.Y.; Zorov, D.B. The age-associated loss of ischemic preconditioning in the kidney is accompanied by mitochondrial dysfunction, increased protein acetylation and decreased autophagy. Sci. Rep. 2017, 7, 44430

[2] Andrianova, N.V.; Jankauskas, S.S.; Zorova, L.D.; Pevzner, I.B.; Popkov, V.A.; Silachev, D.N.; Plotnikov, E.Y.; Zorov, D.B. Mechanisms of Age-Dependent Loss of Dietary Restriction Protective Effects in Acute Kidney Injury. Cells 2018, 7, 178.

[3] Markovets, A. M. et al. Alterations of retinal pigment epithelium cause AMD-like retinopathy in senescent-accelerated OXYS rats. Aging (Albany. NY). 3, 44–54 (2011).

[4] Muraleva, N. A., Ofitserov, E. N., Tikhonov, V. P. & Kolosova, N. G. Efficacy of glucosamine alendronate alone & in combination with dihydroquercetin for treatment of osteoporosis in animal model. Indian J. Med. Res. 135, 221–227 (2012).

[5] Bobko, A. A. et al. 19F NMR measurements of NO production in hypertensive ISIAH and OXYS rats. Biochem. Biophys. Res Commun. 330, 367–370 (2005).

[6] Vays, V. B. et al. Antioxidant SkQ1 delays sarcopenia-associated damage of mitochondrial ultrastructure. Aging (Albany. NY). 6, 140–148 (2014).

[7] Stefanova, N. A., Muraleva, N. A., Skulachev, V. P. & Kolosova, N. G. Alzheimer's disease-like pathology in senescence-accelerated OXYS rats can be partially retarded with mitochondria-targeted antioxidant SkQ1. J. Alzheimer's Dis. 38, 681–694 (2014).

DOI 10.18699/BGRS/SB-2020-307

Opisthorchis felineus extracellular vesicles increase cell proliferation and migration rates of human H69 cholangiocytes

D.V. Ponomarev*, O. Zaparina, M.Y. Pakharukova, V.A. Mordvinov
Laboratory of Molecular Mechanisms of Pathological Processes
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
*E-mail: p.dmitr@outlook.com

***Abstract* — Opisthorchis felineus, is a food-borne liver trematode and the main cause of opisthorchiasis in Russia and Europe. It affects hepatobiliary system of fish-eating mammals, including humans. Opisthorchiasis is associated with chronic inflammation, biliary epithelium proliferation, liver fibrosis, and even might cause cholangiocarcinoma among chronically infected individuals. Secreted proteins and extracellular vesicles of liver flukes might play an important role in the development of pathology. We investigated the response of human H69 cholangiocytes and human hepatoma HepG2 cells to adult liver flukes and to extracellular vesicles released from the flukes. We have demonstrated high mitogenic and cell migration stimulating activity of flukes and EVs vesicles. The activity was specific for cholangiocytes, but not for HepG2 cells. Specific mitogenic effect of liver fluke extracellular vesicles on proliferation and migration of human cholangiocytes in vitro may reflect the mechanisms of development of precancerous biliary intraepithelial neoplasia during opisthorchiasis in vivo.**

Key words* — *Opisthorchis felineus, liver flukes, extracellular vesicles, cholangiocytes, cell proliferation

Motivation and aim

The causative agent of opisthorchiasis *Opisthorchis felineus* causes chronic inflammation, biliary intraepithelial neoplasia and cholangiofibrosis [1]. Chronic opisthorchiasis may cause cholangiocarcinoma, cancer of the biliary tract. The molecular mechanisms of pathology are not studied. One of the potential key events in host-parasite interaction might be fluke secretory proteins, released from the fluke.

The aim was to investigate the effect of adult worms and their extracellular vesicles on cell proliferation and migration rates of human H69 cholangiocytes.

Matherial and methods

Adult *Opisthorchis felineus* liver flukes were extracted from laboratory hamsters previously infected with the metacercariae. Adult flukes were cultured in the media for 7 days. EVs were extracted from the cultivation media and purified, using ultrafiltration approach. Human SV-40 immortalized cholangiocytes (H69) [2] and human hepatoma cells (HepG2) were used to assess the effects of EVs and adult flukes on cell proliferation and cell migration.

To examine the effect of adult flukes on cell proliferation, 1-6 flukes were co-cultured with either H69 or HepG2 cells using a non-contact approach in 12-well Transwell plates with a 1 or 8 μm pore size chambers. The numbers of cells was calculated at days 4, 9 and 14. Cell migration analysis was performed after 14 days of co-cultivation with adult flukes using traditional scratch test.

Results

In contrast to HepG2 cells, proliferation of human cholangiocytes H69 was 4-17 times increased after co-cultivation with the liver flukes. The rate of H69 cell migration, (not HepG2) was also elevated (2 times) after the liver flukes and was accompanied by cell morphological changes.

The effect of fluke extracellular vesicles on proliferation and rate of cell migration was similar to the effect of adult flukes. Unlike H69 cells, the influence on HepG2 cells was not significant. Specific mitogenic effect of liver fluke extracellular vesicles on proliferation and migration of human cholangiocytes *in vitro* may reflect the mechanisms of development of precancerous biliary intraepithelial neoplasia during opisthorchiasis *in vivo*.

Acknowledgment

The study was financially supported by the Russian Science Foundation No. 18-15-00098.

References

[1] Mariya Y. Pakharukova and Viatcheslav A. Mordvinov, "The liver fluke Opisthorchis felineus: biology, epidemiology and carcinogenic potential," Transactions of The Royal Society of Tropical Medicine and Hygiene, Volume 110, Issue 1,, p. 28–36, 06 January 2016., in press

[2] S. A. Grubman, R. D. Perrone, D. W. Lee, S. L. Murray, L. C. Rogers, L. I. Wolkoff et al, "Regulation of intracellular pH by immortalized human intrahepatic biliary epithelial cell lines," American Physiological Society, pp. G1060-G1070, 1 June 1994., in press

DOI 10.18699/BGRS/SB-2020-308

Cytokines as markers of oncogenesis and therapy efficiency in chemically induced breast cancer

Poveshchenko Alexandr
Research Institute of Clinical andExperimental Lymphology - Branch of the Institute of Cytology and Genetics,Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
poveshchenkoa200@mail.ru

Kazakov Oleg
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics,Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
kazakoff_oleg@mail.ru

Orlov Nikolai
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
nbo700@mail.ru

Kabakov Alexey
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
doctor03-85@ngs.ru

Strunkin Dmitry
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
Strunkind@mail.ru

Vladimir Konenkov
Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics,Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
vikonenkov@gmail.com

***Abstract* — Breast cancer was induced by the administration of n-methyl-N-nitrosourea to Wistar rats. Some animals underwent only surgical intervention or only polychemotherapy (cyclophosphamide, methotrexate, 5-ftorouracil). In some animals, both types of therapy were combined, and in a separate group, the administration of the Panagen drug, which was a fragmented DNA molecule, was added to polychemotherapy. To study the concentration of cytokines in lymph, we used the Bio-Plex Pro Rat Cytokoness 24-Plex Assay test system (Bio-Rad, USA). It was shown that in rats with breast cancer the content in the lymph and blood of most of the studied cytokines such as IL-1β, IL-2, IL-4, IL-6, IL-7, IL-12, IL-13, MIP-1α, MIP-3α, RANTES, TNF-α, MCP-1 were statistically significantly higher, and IL-10 and GRO / KC, lower than in intact animals. Surgical removal of the tumor led to a significant decrease in the content of both pro-inflammatory and anti-inflammatory cytokines in the lymph. PCT led to a significant decrease in the content of IL-1β, IL-4, IL-6, IL-7, MIP-1α, MIP-3α, RANTES in the blood serum and lymph of rats with breast cancer.**

Keywords — breast cancer, Wistar rat, cytokines

Introduction

Breast cancer (BC) maintains a leading position among oncological pathologies in women around the world [1,2]. The purpose of the study is to determine the quantitative characteristics and qualitative features of the blood and lymph cytokine profile at various stages of breast cancer oncogenesis, to conduct a comparative study of lymph cytokines and blood serum of Wistar rats after surgery, polychemotherapy, polychemotherapy in combination with the Panagen therapy.

Materials and Methods

The studies were carried out on sexually mature (3 months old) female rats of the Wistar breed weighing 250-300 g in the amount of 9-10 animals in each group: group 1 — intact rats; 2nd group - rats with breast cancer (tumor carriers); 3rd group - rats with breast cancer after polychemotherapy (PCT); Group 4 — rats after removal of breast cancer (operated animals), Group 5 — rats after removal of breast cancer and PCT, group 6 — rats after removal of breast cancer in combination with PCT and administration of the panagen preparation (fragmented DNA). BC was induced by the introduction of N-methyl-N-nitrosourea (MNU) 5 times with an interval of 7 days, subcutaneously in the region of the 2nd breast on the right.,resulting in the formation of breast cancer (adenocarcinoma) after 6 months [1]. Surgical treatment was performed 6 months after the induction of breast cancer. For PCT, the CMF regimen was used (cyclophosphamide - at a dose of 100 mg / m2 from 1 to 14 days; methotrexate - at a dose of 40 mg / m2 on days 1 and 8; 5-ftoruracil - at a dose of 600 mg / m2 in 1 and 8) (Sigma-Aldrich, USA)]. The course of Panagen (fragmented DNA (5 mg / kg) therapy was carried out by intraperitoneal administration once for 14 days 3 hours after cyclophosphamide administration. The experiments used the substance of the Panagen isolated from human placenta. Lymph was taken from the thoracic lymphatic duct [1]. Concentrations of 24 cytokines in lymph were evaluated by flow fluorimetry on a 2-beam laser automated analyzer (Bio-Plex Assay System (Bio-Rad, USA) using a commercial test system (determined dynamic range of 2-32000 pkg / ml) in accordance with manufacturer's instructions Bio-Plex Pro Rat Cytokoness 24-Plex Assay (Bio-Rad, USA).Differences between the groups were considered significant at $p<0.05$. Significance of the differences was calculated by U-criterion the Mann-Whitney.

Results

In rats with breast cancer, the content of most of the studied cytokines such as IL-1β, IL-2, IL-4, IL-6, IL- 7, IL-12, IL-13, IL-17A, IFN-γ, MIP-1α, MIP-3α, RANTES, TNF-α, MCP-1 was statistically significantly higher in serum and lymph than in intact animals. According to many researchers, pro-inflammatory cytokines inside the tumor and its microenvironment are associated with the progression of breast cancer [1,2]. The high content of these biologically active factors in animals with breast cancer allows them to be attributed to tumor growth markers. Induction of breast cancer

led to a decrease in the production of anti-inflammatory cytokine IL-10, which can contribute to immunosuppression and weakening of antitumor defense.

Surgical removal of the tumor resulted in a significant decrease in serum and lymph levels, such as pro-inflammatory cytokines (IL-1β, IL-2, IL-12, IL-18, TNF-α), chemokines (MIP-1α, MIP-3α, RANTES , GRO / KC) and anti-inflammatory cytokines IL-4, IL-13, compared with breast cancer.

A comparative study of the cytokine content after removal of the tumor with intact animals showed that the content of IL-10, IL-18, EPO, GRO / KC, RANTES was significantly higher in the group of control animals. Moreover, the level of production of IL-10 and EPO after surgery was comparable with the level in breast cancer. The main source of IL-10 in the tumor microenvironment is macrophages, which increase the expression of IL-10 in response to signals from dying cancer cells. Removal of the primary tumor site obviously weakened the signaling pathways, and IL-10 production was low. The content of IL-6, TNF-α, and MCP-1 chemokine in the group of operated animals was significantly higher in comparison with the control, which allows us to consider them not only pro-inflammatory markers, but also metastasis markers.

PCT led to a significant decrease in the content of IL-1β, IL-4, IL-6, IL-7, IL-12, IL-18, MIP-1α, MIP-3α, RANTES in the blood serum and lymph of rats with breast cancer. When studying the concentration of cytokines in the blood serum of rats after tumor removal, compared with animals after removal of breast cancer and PCT, no significant differences were found. Nevertheless, the content of IL-2, IL-5, IL-7, IL-13, IL-17A, G-CSF, MCP-1 in the blood serum was significantly higher in the group of operated animals compared with the combined therapy. In this case, there are no changes in the production of metastatic markers such as IL-1β, IL-6, MIP-1α, MIP-3α, VEGF, MCP-1. But at the same time, the level of production of the chemokines GRO / KC and RANTES is higher with a combination of PCT and surgical treatment. Obviously, GRO / KC and RANTES determine not only metastasis, but also chemotaxis of immunocompetent cells.

The concentration of cytokines in the blood serum of the operated animals after PCT and administration of the Panagen significantly increased compared to the group of rats after surgical removal of breast cancer and PCT for most of the studied cytokines: IL-2, IL-4, IL-5, IL6 , IL-7, IL-10, IL-12, IL-17A, IL-18, G-CSF, GM-CSF, GRO / KC, IFNγ, MIP-1α, MIP-3α. A study of the cytokine content in the lymph of the operated animals after polychemotherapy and the administration of the panagen drug also showed that most indicators of the cytokine content such as IL-5, IL-6, IL-7, IL-10, IL-13, IL-17A, IL- 18, GRO / KC, IFNγ, MIP-3α in lymph was also higher after administration of panagen. The stimulating effect of the drug Panagen on lymphoid cells is expressed in an increase in the production of a number of cytokines. This coincides with the notion that cytostatics in combination with Panagen possibly provide an antitumor effect and stimulate lymphoid cells of the immune system [3].

Conclusion

Our comparative study of the cytokine profile using a test system consisting of 24 cytokines of blood serum and lymph of Wistar rats found that the quantitative content of cytokines in breast cancer depends on the type of treatment for the disease. The production of cytokines may change in the process of progression, inhibition of tumor development or metastasis.

Thus, lymph and serum cytokines in experimental breast cancer are markers of tumor growth and metastasis, their levels depend on the type of treatment.

References

[1] A.F. Poveshchenko, O.V. Kazakov, N.B. Orlov, O.V. Poveshchenko , I.I. Kim, N.A. Bondarenko, I.G. Solovieva, D.N.Strunkin, A.V. Kabakov, T.V. Reiter, A.P. Lykov, S.S. Bogachev, E.A. Pokushalov V.I. Konenkov. Lymphatic cytokines as markers of oncogenesis and treatment efficacy in experimental Wistar rat mammary tumor. Pathological physiology and experimental therapy. 2016. 60(3): 68 - 75

[2] Sosnina A.V., Velikaya N.V., Autenshlus A.I. The role of cytokines in the pathogenesis of malignant neoplasms. Novosibirsk: Vector-Best, 2013, 80. (In Russian).

[3] E.A. Alyamkina, O.Y.Leplina, LV.Sakhno, E.R. Chernykh, A.A.Ostanin, Y.R. Efremov, A.G.Shilov, A.S. Proskurina, K.E. Orishchenko, E.V. Dolgova, V.A. Rogachev, V.P. Nikolin, N.A. Popova, S.N. Zagrebelniy, S.S. Bogachev, M.A. Shurdov. Effect of double-stranded DNA on maturation of dendritic cells in vitro. Cellular Immunology. 2010. 266: 46-51.

DOI 10.18699/BGRS/SB-2020-309

Genomics, nutrition, and the microbiome: toward an immunotoxicology of the GI tract

Peter Pressman, MD
The Daedalus Foundation
San Clemente, California, USA
drpressvm2@gmail.com

***Abstract* — Genomic vulnerabilities to environmental challenges can be modulated by nutritionally upregulating or downregulating specific pathways via genomically targeted use of nutrients. The intersection of nutrition, gut microbes and xenobiotics, is reviewed, and possible pathologic and therapeutic pathways are suggested.**

Keywords* — *Nutrition, Genomics, Microbiota, Toxicnts, B-glucan, Lactoferrin, Acemannan

Motivation and Aim

Environmental toxicants have contributed to an array of devastating disorders. Recent pandemics involving HIV, corona viruses, dengue, and Ebola have challenged global resources and our medical fund of knowledge. In the realm of allergy, there has been significant increase in the prevalence of asthma and food allergies. The spectrum of malignancies continues unabated, and the awareness of zoonotic and phytoviral spillover has led to speculation that microbes may be at the root of infectious, allergic, and neoplastic pathology.

The identification of nutritional approaches to upregulating and downregulating specific pathways relative to detoxification via genomically targeted use of foods and supplements appears to be a realistic goal. Enhancement of immunocompetence utilizing hitherto untapped sources of bioactive polysaccharides also presents promising approaches to lowering the risk of pathologies.

Methods

Genomically, we discuss critical pathways relating to phase I and phase 2 detox such as GSTP1, GPX1, GSTT1, deletions, PON1 and some of the CYP 450 system exemplifying genomic vulnerabilities to toxicants. We identify selective nutrient impact on gut immunology, and review the multidimensional interaction between microbial metabolism, gut immunology, and various infectious agents and chemical toxicants.

Results

A nutritional framework for augmentation of immunocompetence and detoxification is suggested and illustrated through an examination of the B-glucans, Lactoferrin and Acemannan.

ACKNOWLEDGMENT

Travel supported by PMI.

DOI 10.18699/BGRS/SB-2020-310

X-chromosome Inactivation in American Mink iPSCs

Inna Pristyazhnyuk
Sector of Genomic Mechanisms of Ontogenesis
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
iprist@mail.ru

Aleksei Menzorov
Sector of Cell Collections
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
menzorov@bionet.nsc.ru

***Abstract* — Induced pluripotent stem cells (iPSCs) were produced from many species. One of the key iPSC characteristics is X-chromosome inactivation status. Both active X-chromosomes mark naïve pluripotency, the presence of inactive X-chromosome suggests primed pluripotency. We had previously shown that American mink embryonic stem cells have both X-chromosomes active. In this study, we produced American mink iPSCs from XX embryonic fibroblasts and studied X-chromosome inactivation. We performed immunocytochemical staining of H3K27me3 chromatin modification (marker of inactive chromatin) and reverse transcription PCR of the *Xist* expression (marker of inactivated X-chromosome) in single cells. We have shown that according to the immunocytochemical analysis of three iPSC lines, from 56.8 to 95.6 % of cells had both X-chromosome homologs active. *Xist* expression data confirmed these results for one iPSC line, but unexpectedly the majority of cells in the other two iPSC lines were positive for *Xist* expression, so one of the X-chromosomes was presumably inactivated. Thus, we analyzed X-chromosome inactivation in XX American mink iPSCs for the first time and revealed the presence of different X-chromosome inactivation patterns.**

Keywords* — *induced pluripotent stem cells, X-chromosome inactivation, American mink, Neovison vison, Xist

Introduction

Induced pluripotent stem cells (iPSCs) share similar properties with embryonic stem cells and allow studying early embryonic development of species whose biological material is difficult to obtain. Currently, iPSCs have been produced for less than 20 species. Among order Carnivora, there are just a few species: snow leopard, dog, cat, and American mink. One of the important pluripotency markers is X-chromosome. Naïve pluripotent stem cells have both X-chromosome homologs active (XaXa) and primed – one inactivated homolog (XaXi). We have previously shown that American mink (*Neovison vison*) embryonic stem cells have both active X-chromosomes [1; 2]. In this study, we analyzed X-chromosome status in American mink iPSCs produced from XX embryonic fibroblasts.

Materials and Methods

American mink iPSC lines iNV1XX1, iNV2XX5, and iNV5XX2 were produced from XX embryonic fibroblasts NV1, NV2, and NV5, respectively, according to previously published protocol [1] with minor modifications. Cell culture medium contained 10 % fetal bovine serum and 10 % KnockOut Serum Replacement (Thermo Fisher Scientific, USA) during iPSC derivation, after third passage 20 % fetal bovine serum was used.

X-chromosome inactivation was estimated by immunocytochemistry and single-cell reverse transcription (RT) PCR.

One of the markers of inactive X-linked chromatin modifications is histone H3 trimethylated at lysine 27 (H3K27me3). Cells were stained for H3K27me3 according to the previously published protocol [1]. H3K27me3 signal was interpreted as the presence of the inactive X-chromosome.

For RT-PCR analysis, we manually collected individual cells into lysis buffer and synthesized cDNA in 5 µl volume using M-MuLV–RH Reverse Transcription Kit (Biolabmix, Russia) according to the manufacturer's recommendations. Then, 2 µl of cDNA solution were used for PCR using BioMaster HS-Taq PCR-Color kit (Biolabmix, Russia) according to the manufacturer's recommendations. Primer pairs for *Xist* and *Rex1* amplification were located in different exons, so *DNAse*I treatment of RNA was not necessary.

Results

iPSCs had normal diploid karyotype (30, XX), expressed pluripotency marker genes *Oct4*, *Rex1*, and as expected had practically no expression of *Nanog*.

We assumed that the majority of iPSCs would be XaXa with *Xist-/Rex1+* pattern, where the absence of *Xist* transcript means lack of Xi, and the presence of *Rex1* serves as a pluripotency marker.

Immunochemical analysis of H3K27me3 was verified on XX and XY embryonic fibroblasts. XX NV1 and NV5 embryonic fibroblasts showed that from 84.9 to 86.7 % of cells are XaXi, as one of the X-chromosomes is inactivated in somatic cells. All cells of XY NV4 fibroblasts lacked Xi, as expected. Analysis of 77-89 cells for each iPSC line revealed that from 56.8 to 95.6 % of cells had both X-chromosome homologs active (Fig. 1).

To verify RT-PCR method we analyzed the expression of *Xist* transcripts in 20 individual embryonic fibroblast cells. 19 XX NV5 cells had *Xist* transcript marking Xi, none XY NV4 cells had it, as expected. Then we analyzed the number of *Xist-/Rex+* cells in iPSCs. It varied from 0 to 30.8 % (Fig. 1). The number of *Xist-/Rex+* cells was 30.8 % for iNV1XX1 which is comparable with 45.7 % of XaXa cells. As for iNV2XX5 and iNV5XX2, there was just one *Xist-/Rex+* cell, the majority of cells were pluripotent with XaXi status (*Xist+/Rex1+*) or differentiated (*Xist+/Rex1-*).

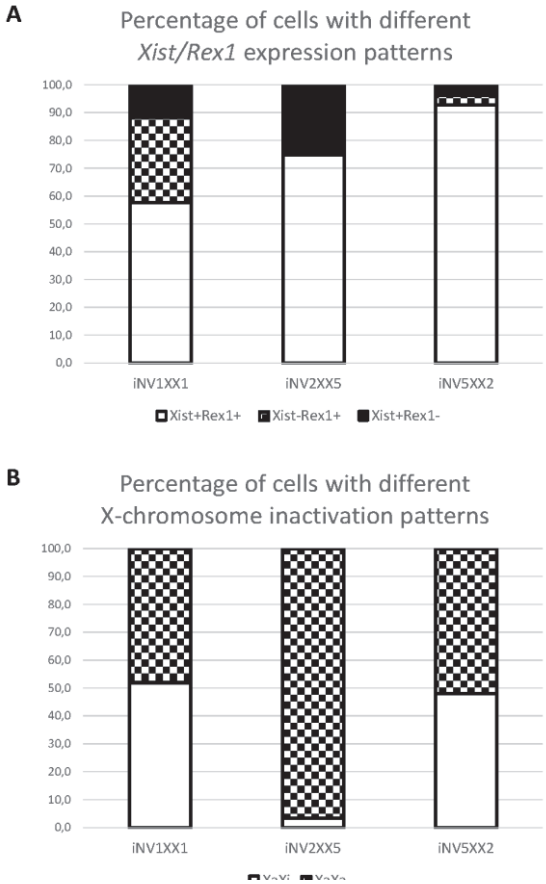


Fig. 1. X-chromosome inactivation patterns in XX American mink iPSCs.

Discussion

Mink embryonic stem cells have both X-chromosomes active [1; 2]. We expected that iPSCs would have XaXa status, as iPSCs are considered practically identical to ES cells. The immunochemical analysis was consistent with that; the majority of cells were XaXa, and some XaXi cells were expected, as there is a certain percentage of differentiated pluripotent stem cells in in vitro culture. RT-PCR analysis of single cells had given different results.

The discordance between immunocytochemical and RT-PCR analyses could have several reasons. First, the iPSCs could have changed XaXa status to XaXi without any changes in morphology and/or gene expression. It is known that cell density, duration of *in vitro* culture, and other factors could promote pluripotent stem cell differentiation. Second, though *Xist* is a well-known X-chromosome inactivation marker, its expression is one of the first steps of the X-chromosome inactivation process. It is possible, that there is an onset of *Xist* expression in XaXa cells that is revealed by RT-PCR but does not coincide with H3K27me3 chromatin modification, as it has not started yet. Third, there might be a certain threshold of *Xist* and other genes expression to trigger X-chromosome inactivation [3]. We were not able to quantify the level of *Xist* expression, so the presence of *Xist* transcript may not be enough to conclude X-chromosome inactivation.

Overall, we produced American mink iPSCs from XX embryonic fibroblasts for the first time. Our data show that iPSCs could have XaXa state of X-chromosomes that is one of the naïve pluripotency markers or could have XaXi state that is either marker of primed pluripotency or sign of differentiation.

Acknowledgment

The reported study was funded by RFBR according to the research project No. 20-04-00369 and by the Ministry of Education and Science of Russia, state project No. 0324-2019-0041-C-01. Cell lines are available at the Collective Center of ICG SB RAS "Collection of Pluripotent Human and Mammalian Cell Cultures for Biological and Biomedical Research" (http://ckp.icgen.ru/cells/; http://www.biores.cytogen.ru/icg_sb_ras_cell/).

References

[1] A. G. Menzorov, N. M. Matveeva, M. N. Markakis, V. S. Fishman, K. Christensen et al., "Comparison of American mink embryonic stem and induced pluripotent stem cell transcriptomes," BMC Genomics, vol. 16(Suppl 13), S6, December 2015.

[2] M. A. Sukoyan, S. Y. Vatolin, A. N. Golubitsa, A. I. Zhelezova, L. A. Semenova, O. L. Serov, "Embryonic stem cells derived from morulae, inner cell mass, and blastocysts of mink: comparisons of their pluripotencies," Mol Reprod Dev., vol. 36(2), pp. 148-158, October 1993.

[3] T. S. Barakat, I. Jonkers, K. Monkhorst, J. Gribnau, "X-changing information on X inactivation," Exp Cell Res., vol. 316(5), pp. 679-687, March 2010.

DOI 10.18699/BGRS/SB-2020-311

3D-models creation based on solid tumor cell lines for assessment of antitumor treatment

Elizaveta Prosekina
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg State University,
St. Petersburg, Russia
elizaveta.prosekina@gmail.com

Natalia Avdonkina
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg, Russia
nataliaavdonkina@gmail.com

Anna Danilova
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg, Russia
anna_danilova@bk.ru

Tatiana Nekhaeva
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg, Russia
nehaevat151274@mail.ru

Irina Baldueva
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg, Russia
biahome@mail.ru

***Abstract* — The creation of three-dimensional cellular models promotes the study of tumor microenvironment. It contributes to a more adequate assessment of the therapeutic effects on malignant cells, which improves the prognostic value of preclinical research and promotes the progress of effective drug therapy for malignant tumors. In this study, 3D models (tumoroids) were obtained from solid tumors cells of patients treated at the N.N. Petrov National Medical Research Center of Oncology. The therapeutic effect of chemotherapy on tumoroids with previous treatment of patients was analyzed.**

Keywords* — *Tumoroids, 3D system, solid tumors, antitumor therapy

Motivation and Aim

Motivation

A malignant tumor is a complex multi-component system, which has individual characteristics for each patient. A model system must adequately reproduce the key characteristics of the tumor in order to study the molecular composition of the tumor and its response to drug therapy.

The most common cell models in vitro for many decades were two-dimensional (monolayer) cultures, in vivo - animal models. However, these models have several disadvantages. In particular, two-dimensional models are well reproducible, have low cost in content, but practically do not retain phenotypic similarity with the tumor material from which they were derived. Animal models perfectly simulate the complex three-dimensional arrangement of cells, the gradient of substances, and features of the tumor microenvironment. But this models are difficult to reproduce and expensive in content.

Adequate model systems should reflect the anatomical and physiological complexity of the malignant tumor that affects the distribution of cellular receptors and various substances, including drug molecules. Tumoroids (3D models) can be used as a micromodel to reflect the main features of the patient's tumor. They are as close in structure and physiological properties to the natural tumor system as possible, well reproducible and inexpensive to maintain [1].

Aim

The aim of the study is experimental 3D models development from solid tumor cell lines for imitation a tumor foci in vitro and for creation a personalized approach in treatment of malignant tumors.

Methods

Three cultures of tumor cells were used in the study: skin melanoma (n=1), myxofibrosarcoma (n=1), embryonic rhabdomyosarcoma (n=1). Cultures were derived from the surgical material of patients and cryoconserved in the biobank of the N.N. Petrov National Medical Research Center of Oncology. Cells were cultivated in monolayer (2D-format) and in the spheroids (3D-format), obtained by the "hanging drop method" under standard conditions (37 °C, 5% CO_2, 100%) with using DMEM/F-12 medium, containing 20% fetal bovine serum and glutamine. Various clinical chemotherapeutic agents and their combinations were used to evaluate the chemosensitivity and invasive properties of tumour cells cultivated in 3D systems. Selection of various chemotherapeutic agents and their combinations, used in the clinic, was carried out retrospectively on the basis of previous patients treatment at the N.N. Petrov National Medical Research Center of Oncology. (Table.1) [2].

TABLE I. SELECTION OF TEST DRUG TO STUDY THE CHEMOSENSITIVITY AND INVASIVE POTENTIAL OF TUMOROIDS IN ACCORDANCE WITH CLINICAL TREATMENT OF PATIENTS

Cell culture	***Patient's previous treatment***	***Model system***
Skin melanoma	Paclitaxel	Paclitaxel
Myxofibrosarcoma	Gemcitabine+Docetaxel; Doxorubicin+ifosfamide.	Doxorubicin+ifosfamide; Paclitaxel
Embryonic rhabdomyosarcoma	Doxorubicin+ifosfamide	Doxorubicin+ifosfamide

MTT-test was used to evaluate chemoresistance of tumor cells. Two concentrations of drug were used for MTT-test staging: therapeutic (corresponding to 10% of the peak

concentration in blood plasma) and ten times higher than therapeutic (corresponding to the peak concentration in blood plasma - 100%). Tumor spheroids and monolayer cultures were placed in the wells of the 96-well plate with the addition of 200 µl of full nutrient medium (DMEM/F-12, 10% FBS, Glu) (control group) and full nutrient medium (DMEM/F-12, 10% FBS, Glu) + drug and incubated during the time corresponding to the period of drug half-life in the organism. Then MTT (0.5 mg/ml) (Roche, Basel, Switzerland) was introduced at the wells where monolayer cell lines and tumor spheroids were cultivated and incubated under CO_2-conditions (+37°C, 5-6.5% CO_2). After 4 hours 100 µl DMSO per 1 hour was added to monolayer cultures and spheroids. Optical density of the obtained solutions was measured on iMark microplate photometer (Bio-Rad Laboratories, USA) at wavelength 595 nm.

To assess the invasive potential tumoroids were placed in a 48-well pre-coated with a polymerized matrix plate with the addition of 700 µl of full nutrient media (DMEM/F-12, 10% FBS, Glu) (control group) and full nutrient media (DMEM/F-12, 10% FBS, Glu) + drug for 96 hours in CellIQ (automated cell culture and analysis system,CM Technologies Oy, Finland). The CellIQ Analyser™ software (CM Technologies Oy, Finland) was used to analyze changes in the area occupied by a tumoroid (Area size).

Results

The chemosensitivity to the combination of doxorubicin+ifosfamide was found in the rhabdomyosarcoma cell culture in 2D and 3D-systems in the MTT-test. Peak concentration of this drugs combination (1 µg/ml + 440 µg/ml) decrease 17% of tumor spheroids viability, while in monolayers the number of live cells after incubation was 59.5%. The invasive potential of the spheroids was almost unchanged when peak concentrations of drugs were added to the medium. This cell response to the combination of doxorubicin+ifosfamide was correlated with the clinical course of the disease in response to the combination of antitumor drugs.

Sensitivity to the combination of doxorubicin+ifosfamide was noted for the monolayer culture of myxofibrosarcoma, while the tumoroids weren't practically sensitive to such therapy. Myxofibrosarcoma tumoroids showed the same response to the combination of doxorubicin+ifosfamide in two concentrations (1 µg/ml + 440 µg/ml; 0.1 µg/ml + 44 µg/ml). The number of viable cells was 83 and 83.5%, respectively. The maximum cytotoxic effect of the combination of doxorubicin+ifosfamide was observed in monolayer cultures. Addition of peak concentration (1 µg/ml + 440 µg/ml) leads to 59.5% of viable cells and for therapeutic dose (0.1 µg/ml + 44 µg/ml) - 79.5. In the clinic background of chemotherapy lines observed progression of tumor disease. At the same time myxofibrosarcomas tumoroids in vitro under the influence of paclitaxel (2,17 µg/ml; (0,217 µg/ml) demonstrated decrease of invasive potential. Chemosensitivity to paclitaxel was also demonstrated in MTT-test.

The changes results in the invasive potential by using the automatic analytical system CellIQ in culturing on the matrix with drug correlated with the results of sensitivity in MTT-test for skin melanoma culture. Tumoroids don't react to the presence in the medium the peak paclitaxel concentration (2.17 µg/ml). Melanoma tumoroids were reacted without their invasive properties changing and proliferative activity, which corresponds to the clinical course of the disease, in which the progression was observed.

An individual comparative analysis of the chemotherapy effectiveness and their combinations in vitro and in vivo on tumor cells with different origins and cultivated in different model systems, in general, demonstrates that tumoroids as a test system for the selection of chemotherapy drugs and individualization of patients treatment with malignant tumors is justified.

Acknowledgment

Supported by the RFBR (18-29-09014).

References

[1] Colella, G., Fazioli, F., Gallo, M., De Chiara, A., Apice, G., Ruosi, C. Cimmino, A., and De Nigris, F. (2018). Sarcoma spheroids and organoids — Promising tools in the era of personalized medicine. Int. J. Mol. Sci. 19(2): 1-15;

[2] Salmon S.E. (1980) Cloning of Human tumor stem cells: background and overview Prog. Clin. Biol. Res. 48 (3): 13.

DOI 10.18699/BGRS/SB-2020-242

Clinical and immunological characteristics of patients sarcomas with different clonogenic potential

Elizaveta Prosekina
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg State University,
St. Petersburg, Russia
elizaveta.prosekina@gmail.com

Tatiana Nekhaeva
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg, Russia
nehaevat151274@mail.ru

Irina Baldueva
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg, Russia
biahome@mail.ru

Anna Danilova
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg, Russia
anna_danilova@bk.ru

Natalia Emelyanova
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg, Russia
emelyana.79@mail.ru

Natalia Avdonkina
N.N. Petrov National Medical Research Center of Oncology, St. Petersburg, Russia
nataliaavdonkina@gmail.com

***Abstract* — The search of new predictor markers for disease outcomes and treatment efficacy is crucial for metastatic sarcoma, where prognosis is unfavourable and the choice of chemotherapy is limited. In our work we present the results of in vitro clonogenic abilities study of tumor cells connection with immunological parameters of peripheral blood and survival rate of patients with metastatic soft tissue sarcomas and bones sarcomas (STBS).**

Keywords* — *sarcoma, clonogenic potential, immunological indicators

Motivation and aim

Today the state of the immune cellular link is widely discussed as a predictor of the response to immunotherapy. It is undoubted interest for sarcoma, where the effectiveness of chemotherapy is low and prognosis is unfavorable, especially in the development of metastatic disease [1]. Genetic, histological and morphological diversity of different clones inside the tumor and peculiarities of their relations with the patient's immune system mainly presented on tumors models with epithelial origin. Tumours with mesenchymal origin, such as STBS, haven't been sufficiently studied in this regard. Cloning with the limiting dilution method demonstrates the presence of self-regulating tumor cells in the population and can serve as an independent predictor of adverse prognosis.

Aim was to study immunological parameters of peripheral blood and to analyze survival parameters of patients with metastatic STBS, whose tumors differed in ability to form clones during in vitro cultivation.

Methods

54 patients with metastatic STBS were included in the study. Cell cultures were isolated from the STBS tumor samples. All patients were treated at the N.N. Petrov National Medical Research Center of Oncology in the 2013 - 2019. Immunological tests were carried out for 21 patients. At the first stage, cloning was performed using the limiting dilution method [2]. Clone formation efficiency (CFE) was assessed as the ratio of the number of formed single colonies to the number of adherent cells, expressed in a percentages. At the second stage the main populations markers of the immunocompetent cells in peripheral blood of patients were evaluated on the BD FACS Canto™ II flow cytomfluorimeter (BD Bioscience, USA). Total and non-relapsing survival rate of patients was analyzed depending on clonogeneous characteristics of the tumor. Relative values for numbers of B-lymphocyte , T-lymphocytes, cytotoxic T-lymphocytes, activated cytotoxic T-lymphocytes, T-helpers, double positive T-lymphocytes, NKT- and NK cells were calculated in relation to the number of lymphocytes; for regulatory T-lymphocytes in relation to the number of T-helpers. The absolute value for all populations was presented in the standard form: number of cells $*10^9$/L.

Results

The cloning procedure was successful in 40.7% of cases (22/54), while in the group with clonogenic tumors the CFE varied from 9.1-93.2% with a median of 47.5%. The total survival rate (TSR) of patients with clonogenic tumors was lower than that of the group with tumors without clonogenic potential ($p<0.05$), while survival rate before progression (SRBP) didn't differ significantly between the groups. Subpopulation structure analysis of immunocompetent cells of peripheral blood in patients with clonogenic sarcomas revealed a reduced content of activated cytotoxic T-lymphocytes with CD3+CD8+HLA-DR+ phenotype and an increased number of NK cells ($p<0.05$) in comparison with nonclonogenic tumors. Differences weren't found in other investigated parameters. Earlier it has been demonstrated by Wedekind M.F. et al. (2018) that initially low levels of cytotoxic T-lymphocytes presence in infiltrates of synovial and desmoplastic sarcoma have been further reduced with disease progression [3]. The work shows the growth in dynamics of NK cell level in these tumor histotypes. We have observed similar changes in peripheral blood parameters in patients with the most aggressive clonogenic sarcomas. Our observations demonstrate the significance of the tumor cells clonogenic potential as an independent predictor of adverse prognosis in patients with STBS and indicate for further research in this area.

Acknowledgment

Supported by the RFBR (18-29-09014).

References

[1] Petitprez F. et al. (2020) B cells are associated with survival and immunotherapy response in sarcoma. Nature. 577(7791):556-560.

[2] Cree I.A. (2011) Principles of Cancer Cell Culture: Methods and Protocols. Second Edition, Methods in Molecular Biology. 731:502.

[3] Wedekind M.F. et al. (2018) Immune profiles of desmoplastic small round cell tumor and synovial sarcoma suggest different immunotherapeutic susceptibility upfront compared to relapse specimens. Pediatr Blood Cancer. 65(11):e27313.

DOI 10.18699/BGRS/SB-2020-312

Circulating microRNAs potentially associated with progression of castration-resistant prostate cancer

Elena Anatolevna Pudova
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
pudova_elena@inbox.ru

Boris Yakovlevich Alekseev
Urological department
FSBI NMRRC
Moscow, Russia
mnioi@mail.ru

Kirill Mikhailovich Nyushko
Urological department
FSBI NMRRC
Moscow, Russia
kirandja@yandex.ru

Anastasiya Andreevna Kobelyatskaya
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
kaa.chel@mail.

Georgy Sergeevich Krasnov
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
gskrasnov@mail.ru

Anna Victorovna Kudryavtseva
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
rhizamoeba@mail.ru

Abstract — **Metastatic castration-resistant prostate cancer (mCRPC) is characterized by a high mortality rate due to its aggressiveness. In the field of modern oncourology, monitoring the activity of a disease for patients of this category during therapy is an important clinical problem, for the solution of which informative markers are needed. Circulating microRNAs can serve as such markers, which demonstrate diagnostic and prognostic potential in various types of cancer, including prostate cancer. The aim of this study is to search for circulating miRNAs of mCRPC patients plasma during therapy that are potentially associated with the progression of the disease, as well as to study the contribution of these miRNAs to the general progression mechanism. Based on high-throughput sequencing, a bioinformatic analysis was carried out with identification of the microRNA profile potentially associated with the progression of the disease (p-value <0.05): miR-221-3p, miR-99b-3p, miR-145-5p, miR-375-3p, miR-20a-5p, miR-21-5p, miR-451a, miR-199a-5p, miR-183-5p, miR-let-7i-5p and miR-let-7f-5p.**

Key words — castration-resistant prostate cancer, progression, microRNAs, exosomes, sequencing, markers

Introduction

Metastatic castration-resistant prostate cancer (mCRPC) is a phase of the disease with increased aggressiveness and overall survival of patients up to three years [4]. The most important clinical problem is the monitoring of disease activity during ongoing therapy, namely, determining the moment of progression (inefficiency of the current therapeutic concept). The classic diagnostic marker, prostate-specific antigen (PSA), at this disease stage is not a reliable biomarker, as patients had visceral metastases without increasing PSA levels [5]. To solve this problem, reliable and informative non-invasive markers are needed. Currently, the search for such markers is taking place in the direction of microRNA profiling. The aim of our study is the profiling of plasma samples circulating miRNAs, in which an increase in PSA levels during therapy was noted, which is associated with treatment failure in this period and, accordingly, with the progression of the disease. The results of this study may be useful for identifying predictive markers with subsequent validation by qPCR, and will also contribute to a fundamental understanding of the progression mechanisms of mCRPC.

Methods

The study included patients diagnosed with mCRPC during therapy. Blood was taken before treatment and every month until the disease progresses. Blood sampling was carried out in vacutainers with EDTA, followed by centrifugation. Isolation of total RNA from exosomes was carried out from 1 ml of blood plasma using a exoRNeasy Serum-Plasma Midi Kit (Qiagen, Germany), according to the manufacturer's protocol. MicroRNA libraries for subsequent high throughput sequencing were prepared using a NEBNext® Small RNA Library Prep Set for Illumina (New England Biolabs) according to the manufacturer's protocol. High throughput sequencing was performed on a NextSeq500 instrument. The quality control of the reads was performed using the FastQC program. To perform the subsequent bioinformatics data analysis, the miRge 2.0 pipeline was used. The analysis of the number of reads of data (counts per million; CPM) was performed using the edgeR package in the statistical environment R. To normalize the data, the TMM (trimmed mean of M-values) method was used. Further, as part of the analysis of differential expression, a quasi-likelihood F-test (QLF) was used to calculate the reliability of the results. The results were considered statistically significant with a QLF p-value <0.05. Next, we tried to identify the pathways most affected by differentially expressed (DE) miRNAs. For this purpose, we used the mirPath 3.0 (DIANA tools) web service and TarBase 7.0 data.

Results

As a result of the study, the following microRNAs were identified that were potentially associated with mCRPC progression: miR-221-3p, miR-99b-3p, miR-145-5p, miR-375-3p, miR-20a-5p, miR-21- 5p, miR-451a, miR-199a-5p, miR-183-5p, miR-let-7i-5p and miR-let-7f-5p. According to various studies, the identified microRNAs demonstrate oncogenic function in various types of cancer, including mCRPC [1,2,3,6]. For identified microRNAs, mirPath has found 49 pathways with $p < 0.01$, Among these pathways, we noted the following most oncologically significant pathways: proteoglycans in cancer, Hippo signaling pathway, cell cycle, pathways in cancer, focal adhesion, transcriptional misregulation in cancer, prostate cancer and TGF-beta signaling pathway.

ACKNOWLEDGMENT

This work was financially supported by the RFBR, grant no. 17-29-06085. This work was performed using the equipment of EIMB RAS "Genome" center (http://www.eimb.ru/ru1/ckp/ccu_genome_c.php).

REFERENCES

[1] Garofalo M, Quintavalle C, Romano G, Croce CM, Condorelli G. miR221/222 in cancer: their role in tumor progression and response to therapy. Curr Mol Med. 2012 Jan;12(1):27-33. doi: 10.2174/156652412798376170.

[2] Huang X, Yuan T, Liang M, Du M, Xia S, Dittmar R, Wang D, See W, Costello BA, Quevedo F, Tan W, Nandy D, Bevan GH, Longenbach S, Sun Z, Lu Y, Wang T, Thibodeau SN, Boardman L, Kohli M, Wang L. Exosomal miR-1290 and miR-375 as prognostic markers in castration-resistant prostate cancer. Eur Urol. 2015 Jan;67(1):33-41. doi: 10.1016/j.eururo.2014.07.035. Epub 2014 Aug 14.

[3] Lin HM, Castillo L, Mahon KL, Chiam K, Lee BY, Nguyen Q, Boyer MJ, Stockler MR, Pavlakis N, Marx G, Mallesara G, Gurney H, Clark SJ, Swarbrick A, Daly RJ, Horvath LG. Circulating microRNAs are associated with docetaxel chemotherapy outcome in castration-resistant prostate cancer. Br J Cancer. 2014 May 13;110(10):2462-71. doi: 10.1038/bjc.2014.181.

[4] Mostaghel EA, Montgomery B, Nelson PS. Castration-resistant prostate cancer: targeting androgen metabolic pathways in recurrent disease. Urol Oncol. 2009 May-Jun;27(3):251-7. doi: 10.1016/j.urolonc.2009.03.016.

[5] Pezaro C, Omlin A, Lorente D., Rodrigues DN, Ferraldeschi R, Bianchini D, Mukherji D, Riisnaes R, Altavilla A, Crespo M, Tunariu N, de Bono J, Attard G. Visceral disease in castration-resistant prostate cancer. Eur Urol. 2014 Feb;65(2):270-273. doi: 10.1016/j.eururo.2013.10.055.

[6] Zhu J, Wang S, Zhang W, Qiu J, Shan Y, Yang D, Shen B. Screening key microRNAs for castration-resistant prostate cancer based on miRNA/mRNA functional synergistic network. Oncotarget. 2015 Dec 22;6(41):43819-30. doi: 10.18632/oncotarget.6102.

DOI 10.18699/BGRS/SB-2020-313

Wingless-type inducible signaling pathway protein-1 inflammation, diabetes and aging

Natalia Nikolaevna Rudovich
Hospital of Bülach, Bülach, Switzerland; Associated Professor of Charité – Medical University Berlin, Germany
natalia.rudovich@spitalbuelach.ch

Abstract — **Obesity is a growing global health issue linked to the development of type 2 diabetes and other aging associated conditions like cancer and neurocognitive disorders. Adipose tissue dysfunction results in altered adipokine secretion, low-grade inflammation and ectopic fat accumulation. Recently, our group identified Wnt1-inducible signaling pathway protein-1 (WISP-1) as a multifunctional factor in inflammatory and fibrotic processes in a cell-context-specific role.**

Keywords — aging, inflammation, obesity, type 2 diabetes

Obesity is a growing global health issue linked to the development of type 2 diabetes and other aging associated conditions like cancer and neurocognitive disorders [1]. Adipose tissue dysfunction results in altered adipokine secretion, low-grade inflammation, and ectopic fat accumulation [2]. Recently, our group identified Wnt1-inducible signalling pathway protein-1 (WISP-1 or CCN4) as inflammatory adipokine, which supress insulin signalling in muscle cells and hepatocytes, contributing to obesity and associated diseases [3-5]. In further studies, a correlation between *WISP-1* and body weights as well as metabolic parameters such as fasting glucose and insulin levels in gestational diabetes and type 2 diabetes was described [5-7]. In addition, recent studies suggest possible effects of circulating WISP-1 on fat cell differentiation in animals [8]. The underlying mechanisms still need clarification and are currently investigated. In humans, the expression of *WISP-1* in the visceral adipose tissue and liver positively correlated with the body mass index [3, 7]. An association of WISP-1 with inflammatory markers and transforming growth factor-β in humans and mice was demonstrated [9]. Furthermore, recently published data links WISP-1 to fibrosis in different tissue [9-13]. In contrast, WISP1 is required for efficient muscle regeneration and loss of WISP1 during aging contributes to age-related regenerative failure of skeletal muscle [14]. Taken together, WISP-1 is a multifunctional factor in inflammatory and fibrotic processes in a cell-context-specific role. Based on recently findings, WISP-1 is a promising target for future research studies in the field of obesity, fibrosis and ageing

References

[1] Mokdad, A. H., E. S. Ford, B. A. Bowman, W. H. Dietz, F. Vinicor, V. S. Bales & J. S. Marks (2003) Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. JAMA, 289, 76-9.

[2] Gregor, M. F. & G. S. Hotamisligil (2011) Inflammatory mechanisms in obesity. Annu Rev Immunol, 29, 415-45.

[3] Murahovschi, V., O. Pivovarova, I. Ilkavets, R. M. Dmitrieva, S. Docke, F. Keyhani-Nejad, O. Gogebakan, M. Osterhoff, M. Kemper, S. Hornemann, M. Markova, N. Kloting, M. Stockmann, M. O. Weickert, V. Lamounier-Zepter, P. Neuhaus, A. Konradi, S. Dooley, C. von Loeffelholz, M. Bluher, A. F. Pfeiffer & N. Rudovich (2015) WISP-1 is a novel adipokine linked to inflammation in obesity. Diabetes, 64, 856-66...

[4] Horbelt, T., C. Tacke, M. Markova, D. Herzfeld de Wiza, F. Van de Velde, M. Bekaert, Y. Van Nieuwenhove, S. Hornemann, M. Rodiger, N. Seebeck, E. Friedl, W. Jonas, G. H. Thoresen, O. Kuss, A. Rosenthal, V. Lange, A. F. H. Pfeiffer, A. Schurmann, B. Lapauw, N. Rudovich, O. Pivovarova & D. M. Ouwens (2018) The novel adipokine WISP-1 associates with insulin resistance and impairs insulin action in human myotubes and mouse hepatocytes. Diabetologia./

[5] Tacke, C., K. Aleksandrova, M. Rehfeldt, V. Murahovschi, M. Markova, M. Kemper, S. Hornemann, U. Kaiser, C. Honig, C. Gerbracht, S. Kabisch, T. Horbelt, D. M. Ouwens, M. O. Weickert, H. Boeing, A. F. H. Pfeiffer, O. Pivovarova & N. Rudovich (2018) Assessment of circulating Wnt1 inducible signalling pathway protein 1 (WISP-1)/CCN4 as a novel biomarker of obesity. J Cell Commun Signal, 12, 539-548.

[6] Klimontov VV, Bulumbaeva DM, Fazullina ON, Lykov AP, Bgatova NP, Orlov NB, Konenkov VI, Pfeiffer AFH, Pivovarova-Ramich O, Rudovich N. Circulating Wnt1-inducible signaling pathway protein-1 (WISP-1/CCN4) is a novel biomarker of adiposity in subjects with type 2 diabetes. J Cell Commun Signal. 2020 Mar;14(1):101-109

[7] Barchetta, I., F. A. Cimini, D. Capoccia, R. De Gioannis, A. Porzia, F. Mainiero, M. Di Martino, L. Bertoccini, M. De Bernardinis, F. Leonetti, M. G. Baroni, A. Lenzi & M. G. Cavallo (2017) WISP-1 Is a Marker of Systemic and Adipose Tissue Inflammation in Dysmetabolic Subjects With or Without Type 2 Diabetes. J Endocr Soc, 1, 660-670..

[8] Ferrand N, Béreziat V, Moldes M, Zaoui M, Larsen AK, Sabbah M. WISP1/CCN4 inhibits adipocyte differentiation through repression of PPARγ activity. Sci Rep. 2017 May 11;7(1):1749.

[9] Berschneider B, Ellwanger DC, Baarsma HA, Thiel C, Shimbori C, White ES, Kolb M, Neth P, Königshoff M. miR-92a regulates TGF-β1-induced WISP1 expression in pulmonary fibrosis. Int J Biochem Cell Biol. 2014 Aug;53:432-41.

[10] Yang X, Wang H, Tu Y, Li Y, Zou Y, Li G, Wang L, Zhong X. WNT1-inducible signaling protein-1 mediates TGF-β1-induced renal fibrosis in tubular epithelial cells and unilateral ureteral obstruction mouse models via autophagy. J Cell Physiol. 2020 Mar;235(3):2009-2022

[11] Zhao X, Hua Y, Chen H, Yang H, Zhang T, Huang G, Fan H, Tan Z, Huang X, Liu B, Zhou Y. Aldehyde dehydrogenase-2 protects against myocardial infarction-related cardiac fibrosis through modulation of the Wnt/β-catenin signaling pathway. Ther Clin Risk Manag. 2015 Sep 11;11:1371-81.

[12] Li X, Chen Y, Ye W, Tao X, Zhu J, Wu S, Lou L. Blockade of CCN4 attenuates CCl4-induced liver fibrosis. Arch Med Sci. 2015 Jun 19;11(3):647-53.

[13] Heise RL, Stober V, Cheluvaraju C, Hollingsworth JW, Garantziotis S. Mechanical stretch induces epithelial-mesenchymal transition in alveolar epithelia via hyaluronan activation of innate immunity. J Biol Chem. 2011 May 20;286(20):17435-44.

[14] Brown BA, Connolly GM, Mill CEJ, Williams H, Angelini GD, Johnson JL, George SJ. Aging differentially modulates the Wnt pro-survival signalling pathways in vascular smooth muscle cells. Aging Cell. 2019 Feb;18(1):e12844

DOI 10.18699/BGRS/SB-2020-314

Melatonin as a key regulator in molecular-genetic network of glucose variability related to circadian rhythm

Olga Saik
Laboratory of Computer Proteomics
Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences (ICG SB RAS)
Novosibirsk, Russia; Laboratory of Endocrinology
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia
saik@bionet.nsc.ru

Pavel Demenkov
Laboratory of Computer Proteomics
Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences (ICG SB RAS)
Novosibirsk, Russia
demps@bionet.nsc.ru

Vladimir Ivanisenko
Laboratory of Computer Proteomics
Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences (ICG SB RAS)
Novosibirsk, Russia
salix@bionet.nsc.ru

Vadim Klimontov
Laboratory of Endocrinology
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics,
Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia
klimontovvv@bionet.nsc.ru

Abstract — **The circadian rhythm is an important factor in the diurnal fluctuations in blood glucose levels, especially in diabetes mellitus. Analysis of molecular-genetic networks performed by ANDSystem showed that melatonin can be considered as one of the key participants in associative networks that describe the relationship between circadian rhythm and glycemic variability. It also can be assumed that melatonin is able to participate in the formation of the dawn phenomenon.**

Keywords — glucose variability, diabetes mellitus, molecular-genetic networks, ANDSystem, melatonin, circadian rhythm

Motivation and Aim

Pathological fluctuations in glucose levels are associated with various aspects of diabetes and especially with the development of its complications. The glycemic variability depends, among other factors, on the circadian rhythm.

The ANDSystem, developed at the Institute of Cytology and Genetics SB RAS, allows to automatically build associative molecular genetic networks associated with complex phenotypic characters and biological processes. Analysis of molecular-genetic networks can improve understanding of the mechanisms of glucose variability.

Matherials and Methods

An associative network of interactions of molecular objects with a circadian rhythm, hyperglycemia and hypoglycemia, as well as a network describing the relationships between these molecular objects, were created using the "Query Wizard" function of ANDSystem. The calculation of the betweenness centrality for network nodes was carried out by the "Statistics" function in the "Analysis" section of ANDSystem. An associative network describing the effect of melatonin on insulin and glucose via intermediary proteins was constructed using the "Pathway Wizard" function of the ANDSystem system.

Results

According to ANDSystem molecular-genetic networks, 49 objects are associated with circadian rhythm, hyperglycemia, and hypoglycemia, including hormones, growth factors, cytokines, chemical substances, and metabolites. Among the participants in the network were leptin, adiponectin, insulin, interleukin-6, melatonin, growth hormones, cortisol, dopamine, serotonin, estrogens, androgens, and others.

Analysis of the network showed that melatonin and its precursor serotonin possessed the highest values of betweenness centrality. Melatonin plays a key role in maintaining sleep-wake cycles. It is interesting that at 4-8 o'clock in the morning there is a sharp decline in the concentration of melatonin and at the same time some patients with diabetes show dawn phenomenon. Analysis of molecular-genetic networks showed that melatonin is able to regulate insulin and glucose levels. This regulation requires participation of a large number of intermediary genes and proteins.

Conclusion

Analysis of associative networks suggests that normal daily glycemic variability is controlled by a big set of biological signaling pathways that have a multidirectional effect. This ensures that the glucose concentration is maintained at a normal level. However, the loss of a number of links of this network due to mutations and/or pathological processes disrupts the system that maintain glucose balance in the physiological range during circadian oscillations.

Acknowledgment

The research was supported by Grant of Russian Science Foundation № 20-15-00057 «Development of personalized approaches to assessing glycemic variability in patients with type 1 diabetes based on mathematical methods and artificial intelligence».

DOI 10.18699/BGRS/SB-2020-315

Toxicity testing in the 21st century: an overview

John Michael Sauer, Ph.D.
Predictive Safety Testing Consortium (PSTC),
Critical Path Institute, Tucson,
Arizona, United States of America

***Abstract* — For the last 40 years the safety testing methods used during the development of pharmaceutical agents have not substantially changed. Furthermore, in many cases the predictive or translational value of the currently used approaches is at best uncertain. This presentation will discuss an integrated systems toxicology approach to safety testing in which traditional in vivo evaluations are combined with human biology-based in vitro models, novel safety biomarkers, computational tools, and non-safety data. The objective of a more modern approach to toxicity testing is to reduce drug attrition through better decision making and ultimately protect public health through the more rapid launch of therapeutics and elimination of unsafe compounds earlier in the drug development cycle. The introduction of contemporary safety testing methods will require an evolution in thinking by both health authorities and drug developers. Although many of the tools that are required for a modernized approach to safety testing are already available, the more difficult challenge is likely to be gaining their acceptance by drug developers and health authorities.**

Keywords* — *Modernization of toxicology, novel safety biomarker, systems biology, in vitro toxicology

Disclosure Statement: The author does not have any conflicts of financial interest to declare.

DOI 10.18699/BGRS/SB-2020-316

Semi-quantitative analysis of serum proteome in patients with bipolar disorder

Seregin Alexander Alexandrovich
Mental Health Research Institute
Tomsk National Research Medical Center of the Russian Academy of Sciences
Tomsk, Russia
email: apocalips1991@mail.ru

Smirnova Lyudmila Pavlovna
Mental Health Research Institute
Tomsk National Research Medical Center of the Russian Academy of Sciences
Tomsk, Russia
email: lpsmirnova@yandex.ru

Simutkin German Gennadievich
Mental Health Research Institute
Tomsk National Research Medical Center of the Russian Academy of Sciences
Tomsk, Russia
email: ggsimutkin@gmail.com

Abstract — **The work analyzed the protein spectrum of blood serum of 45 with bipolar disorder. Protein recruitment in BD was mainly associated with the immune response, regulation of transport processes through the cell membrane and cellular communication, the development of neurons and oligodendrocytes, and cell growth. Cadherin-5 and vascular endothelial growth factor receptor 1 are a marker of endothelial dysfunction. The detection of these proteins suggests the presence of endothelial dysfunction in the pathogenesis of BD. The detection of myelin 1 transcription factor and the literature data on the effect of mutations in the MYT1L gene on the risk of major depressive disorder suggest the MYT1 protein as a potential BD biomarker. The results may help to discover new pathways associated BD.**

Keywords — biomarker, bipolar disorder, schizophrenia, proteome, mass spectrometry, serum

Introduction

Bipolar disorder (BD) leads to a significant decrease in the quality of life of patients and increases the risk of suicidal behavior. The disease often has a similar clinical picture with other mental disorders, and there are no paraclinical criteria for differential diagnosis for these disorders [1][2]. This is due to a lack of understanding of the molecular mechanisms of BD pathogenesis. In recent years, interest in proteomic research in mental disorders has increased and research in this direction has been intensively conducted [2][4][5]. Identification of the marker protein or regulatory proteins involved in the pathogenesis of BD in biomaterial available for diagnostic purposes - blood serum - will bring us as close as possible to understanding the specific pathogenetic mechanisms of this disorder and can serve as the basis for new methods of differential diagnosis and the development of new pathogenetically based medicines.

Methods

In this work, the protein spectrum of the blood serum of the following patient groups was analyzed 45 patients (aged 32 [21, 52] years and a disease duration of 8 [5, 11] years) with bipolar disorder (F31), 12 of them were in-patients of the Department of Affective States of MHRI and 11 were in-patients of the Department of Endogenous Mental Disorders and Affective States of the Mental Health Research Center (MHRC, Moscow). Patients were hospitalized in the acute state and blood serum samples were taken before the start of the study before the start of the course of therapy. The control group consisted of 24 mentally and somatically healthy volunteers matched to the study groups by gender and age. Exclusion criteria for all subjects were the presence of acute and chronic diseases in the acute phase; taking medication psychoactive drugs. Blood serum was taken from the ulnar vein in the morning on an empty stomach before the start of therapy. Further, the serum was purified by affinity chromatography from 6 major proteins (albumin, immunoglobulin G, immunoglobulin A, antitrypsin, transferrin and haptoglobin) on an AKTA pure chromatograph (GE Healthcare). After that, the purified proteins were concentrated to 1 ml using Amicon Ultra-0.5 ultrafilters (MILLIPORE) at 3 kDa. and separated by vertical electrophoresis in a 12% polyacrylamide gel according to the Lemmli method (Laemmli, U.K, 1970). Then, after trypsinolysis and peptide extraction from the gel, the proteins were identified by HPLC / mass spectrometry on an LTQ Orbitrap Velos mass spectrometer (Thermo Scientific). was carried out in IBMC, Moscow (Centre of Collective Usage "Human proteome"). The mass spectrometric data were analyzed with the MaxQuant software (version 1.6.3.4). Default parameters were used unless otherwise specified below. A false discovery rate (FDR) of 0.01 for proteins and peptides and a minimum peptide length of 6 amino acids were required. The Andromeda search engine was used to search the MS/MS spectra against the UniProt human database combined with 262 common contaminants and concatenated with the reversed versions of all sequences. Enzyme specificity was set to trypsin specificity. A maximum of two missed cleavages were allowed. Peptide identification was based on a search with an initial mass deviation of the precursor ion of up to 7 ppm. The fragment mass tolerance was set to 20 ppm on the m/z scale. Only proteins quantified with at least two peptides were considered for quantitation. MS1-intensity based label-free quantitation was used to assess differences in the abundance of proteins between all the studied groups. LFQ intensities for the proteins were log2-transformed and normalized to ensure equal median protein abundance across the samples. A two-tailed unpaired t-test with an FDR value of 0.05 and S0 = 2 was applied to identify proteins for which the abundance was significantly changed between all the studied groups.

Results

As a result of mass spectrometric analysis, between 50 and 350 proteins in each lane were identified and about 1,600 proteins were identified for each person. Comparison of the proteome profiles of different groups revealed 18 proteins specific for BD.

Proteins found in patients with BD, but missing from the control sample, are involved in the regulation of DNA and cell

cycle synthesis, differentiation of progenitor cells, in the development of neurons and oligodendrocytes, such as Myelin transcription factor 1 (MYT1; iBAQ 607,717; LFQ 8,153,738), which regulates genes encoding myelin proteins and other central nervous system proteins [6]. This protein is a zinc finger DNA-binding protein that affects the proliferation, differentiation and transcription of myelin cell genes. MYT1 can regulate the critical transition point in oligodendrocyte lineage development by modulating oligodendrocyte progenitor proliferation relative to terminal differentiation and up-regulation of myelin gene transcription [7]. Ti Wang et al. Have shown that mutations in the MYT1L gene encoding a protein of the myelin transcription factor 1 family can increase the risks of major depressive disorder. However, a connection with pathogenesis has not yet been identified [8]. A significant increase in the expression of the MYT1 gene in the prefrontal cortex of patients with BD is also known. These facts suggest the MYT1 protein as a potential biomarker for BD. Desmoplakin (DSP iBAQ 3,736,801; LFQ 5,864,175) was also found in the serum of patients with BD, which participates in the organization of cadherin-placoglobin desmosome complexes, forming them in separate domains of the plasma membrane. Cadherin 5 (CADH5 iBAQ 4,472,995; LFQ 6,134,045) was also found in a group of patients with BD. The cadherin protein family not only provides mechanical contact between neighboring cells, but also participates in intracellular signal transduction, regulation of proliferation, migration, sorting, differentiation, and cell morphogenesis. In the tissues of adult organisms, cadherins regulate cell renewal, provide a physiological barrier between contacting tissues and selectivity of the transport of soluble substances. Some mediators of the inflammatory response, such as the vascular endothelial growth factor (VEGFR1) (FLT1; iBAQ 4,665,292; LFQ 6,331,822) we discovered, can interfere with the organization of contacts when their receptors are bound, thereby opening the barrier, and plasma proteins can pass through the endothelial barrier. In particular, VEGFR1 is involved in the initiation of autophosphorylation of cadherin 5 signaling cascades, which directly affect the development of endothelial dysfunction [9]. In addition, it is known that cadherins mediate cell sorting, migration, and segregation, morphogenesis, and axon growth in embryogenesis [10]. Van et al. The involvement of genes encoding the CDH9 and CDH10 families have been demonstrated in autism [11]. However, there is no data yet, which indicates the participation of cadherin 5 in the pathogenesis of mental disorders, which requires further study.

Conclusions

Comparison of the proteome profiles of the studied groups revealed 18 proteins specific for BD. The set of proteins in BD was mainly associated with the immune response, regulation of transport processes across the cell membrane, and cell communication. Many of these pathways are involved in the pathogenesis of mental disorders. They are interesting as potential markers. Cadherin-5 and vascular endothelial growth factor receptor 1 are a marker of endothelial dysfunction. The detection of these proteins suggests the presence of endothelial dysfunction in the pathogenesis of BD. Detection of desmoplakin can indirectly indicate a violation of the intercellular contacts of endotheliocytes. The detection of myelin 1 transcription factor and the literature data on the effect of mutations in the MYT1L gene on the risk of major depressive disorder suggest the MYT1 protein as a potential BD biomarker. However, its role in the pathogenesis of BD remains to be seen. Further study of the proteins that we identified that claim to be BD biomarkers may help to uncover unclear aspects of pathogenesis and develop new paraclinical criteria for the differential diagnosis of BD.

ACKNOWLEDGMENT

The study was carried out as part of the government assignment (budgetary funding within the framework of the comprehensive research topic AAAA-A19-119020690013-2 “A comprehensive study of clinical-psychopathological patterns and pathobiological mechanisms of the formation and progression of socially significant mental and behavioral disorders with the development of innovative methods of early diagnosis, personalized treatment and prevention strategies”). Mass spectrometric experiments were performed the equipment of “Human Proteome” Core Facility of the Institute of Biomedical Chemistry (Moscow).

We thank Dr Alexey Chernobrovkin, PhD, Department of Bioinformatics, the Institute of Biomedical Chemistry (IBMC) Moscow, Russia, for his help in work with the MaxQuant software

REFERENCES

[1] Kornetov A. N. “ Ontogenetic aspects of depressive disorders,” vol. 103(8), S.S. Korsakov Journal of Neurology and Psychiatry. 2003, pp. 80-81.

[2] Marusin A.V., Kornetov A.N. and Swarovskaya M.G. “Association of genes for susceptibility to alcoholism, schizophrenia and Alzheimer's disease with psychodiagnostic traits in the Russian population,” vol. 15(5), Bulletin of Siberian Medicine. 2016, pp. 83-96. doi: 10.20538/1682-0363-2016-5-83-96

[3] Smirnova L.P., Loginova L.V., Dmitrieva E.M., Seregin A.A., Semke A.V., Simutkin G.G., Ivanova S.A. “The first results of comparison of blood serum proteomes of patients with schizophrenia and bipolar affective disorder,” vol. 2 (91), Siberian Herald of Psychiatry and Addiction Psychiatry. 2016, pp. 42-47.

[4] Dudley E., Hassler F. and Thome J. “Profiling for novel proteomics biomarkers in neurodevelopmental disorders,” vol. 8(1) Expert Rev Proteomics. 2011, pp. 127–36. doi: 10.1586/epr.10.97.

[5] Föcking M., Dicker P. and English J.A. “Common proteomic changes in the hippocampus in schizophrenia and bipolar disorder and particular evidence for involvement of cornu ammonis regions 2 and 3,” vol. 68(5), Arch Gen Psychiatry. 2011, pp. 477–88. doi: 10.1001/archgenpsychiatry.2011.43.

[6] Singh S. M., Castellani C. and O'Reilly R. “Autism meets schizophrenia via cadherin pathway,” vol. 116, Schizophr Res. 2010, pp. 293–4.

[7] Ryan, M., Lockstone, H. and Huffaker. “Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes,” vol. 11, Mol Psychiatry. 2006, pp. 965–978.

[8] Ti Wang, Zhen Zeng and Tao Li. “Common SNPs in Myelin Transcription Factor 1-Like (MYT1L): Association with Major Depressive Disorder in the Chinese Han Population,” vol. 5(10), Yongyong Shi PLoS One. 2010, doi: 10.1371/journal.pone.0013662

[9] Pang V, Bates DO and Leach L. “Regulation of human feto-placental endothelial barrier integrity by vascular endothelial growth factors: competitive interplay between VEGF-A165a, VEGF-A165b, PIGF and VE-cadherin,” vol. 23;131(23), Clin Sci (Lond). 2017

[10] Perez-Moreno M, Jamora C and Fuchs E. “Sticky business: orchestrating cellular signals at adherens junctions,” vol. 112(4), Cell. 2003, pp. 535–548.

[11] Wang K, Zhang H and Ma D, “Common genetic variants on 5p14.1 associate with autism spectrum disorders,” vol. 28;459(7246), Nature. 2009, pp. 528–533.

DOI 10.18699/BGRS/SB-2020-317

Possible effect of SNP TATA-boxes of human erythropoiesis gene promoters on cognitive disorders

Ekaterina Sharypova
Molecular Genetics Department
Institute of Cytology and Genetics
ICG SB RAS
Novosibirsk, Russia
sharypova@bionet.nsc.ru

Irina Drachkova
Molecular Genetics Department
Institute of Cytology and Genetics
ICG SB RAS
Novosibirsk, Russia
drachkova@bionet.nsc.ru

Ludmila Savinkova
Molecular Genetics Department
Institute of Cytology and Genetics
ICG SB RAS
Novosibirsk, Russia lksav@bionet.nsc.ru

Mikhail Ponomarenko
Systems Biology Department
Institute of Cytology and Genetics
ICG SB RASNovosibirsk, Russia
pon@bionen.nsc.ru

Irina Chadaeva
Systems Biology Department
Institute of Cytology and Genetics ICG SB RAS
Novosibirsk, Russia
ichadaeva@bionet.nsc.ru

***Abstract* — In recent years, there is increasing evidence that various forms of anemia, changes in the quantity and quality of blood cells, may be involved in the pathogenesis of various cognitive disorders accompanying Alzheimer's and Parkinson's diseases, depression of various degrees, etc. In addition to erythroid cells, hemoglobin has been shown to be widely expressed in non-erythroid cells, including neurons of different parts of the brain. We analyzed in silico and in vitro unannotated SNPs in TATA boxes of human genes involved in erythropoiesis. Experimental verification in vitro using the method of electrophoretic mobility shift assay (EMSA) showed the correspondence of prognosis and experimental data. The obtained estimates of the effect of TATA box SNP markers on the formation of TBP/TATA complexes make it possible to consider some SNP markers of erythroid genes as markers of cognitive disorders.**

Keywords* — *TATA-binding protein, TATA-box, TBP/TATA-affinnity, erythropoiesis, single nucleotide polymorphism, cognitive disorders

Background

According to World Health Organization estimates, ~7% of the world population have hereditary aberrations in the synthesis of hemoglobin, meaning that these are the most prevalent monogenic disorders [1]. In recent years, there is increasing evidence that various forms of anemia, changes in the quantity and quality of blood cells, may be involved in the pathogenesis of various cognitive disorders accompanying Alzheimer's and Parkinson's diseases, depression of various degrees, etc. [2]. In addition to erythroid cells, hemoglobin has been shown to be widely expressed in non-erythroid cells, including neurons of the cortex, hippocampus, neurons of the embryos and adult mice brain, and mesencephalic dopaminergic cells of the human, mouse, and rat brain [3]. The morbidity and prevalence of anemia and elevated hemoglobin levels increase with age, which is a heavy burden for society. All this determines the relevance of the problem and the goal of the work, which is to study the effect of SNP TATA boxes of human gene promoters on erythropoiesis and cognitive disorders.

Materials and Methods

The task was to extract from databases of unannotated single nucleotide polymorphisms (SNP) in the TATA elements of human erythroid genes, to analyze them *in silico* using the SNP-TATA_Z-tester Web service developed at the ICG SB RAS (http://beehive.bionet.nsc.com/cgibin/mgs/tatascan_fox/start.pl), the identification of substitutions with potential of functional significance and their experimental verification *in vitro* using the EMSA method. *In vivo*, the task is to identify the differential expression of genes involved in erythropoiesis in various parts of the brain of rats domesticated and aggressive towards humans, in order to identify possible candidate genes for cognitive impairment among them.

Results and Discussions

We analyzed unannotated SNPs in TATA boxes of human globin genes - *HBZ*, *HBB*, *HBD*, *HBG1*, as well as genes encoding enzymes involved in heme biosynthesis - *ALAS1*, *CA1*, *EPOR*, *GYPC* and others involved in erythropoiesis, a total of 161 SNPs 25 human genes. As a result of the analysis using the SNP-TATA_Z-tester Web service, 45 candidate SNP markers were predicted for 15 of 25 genes that can reliably change the affinity of TBP to the TATA boxes of gene promoters participating in erythropoiesis, and an *in silico* prognosis of the effect of SNPs on affinity TBP interactions with TATA / TATA-like sequences. Experimental verification *in vitro* using the method of electrophoretic mobility shift assay (EMSA) showed the correspondence of prognosis and experimental data. As an example, consider the *EPOR* gene - as you know, the glycoprotein hormone EPO is the main regulator of red blood cell production and exhibits protective properties in cerebral ischemia [4]. To perform the function, it binds to its receptor, EpoR, which dimerizes and activates, which leads to the launch of a cascade of genes responsible for the proliferation, survival and differentiation of erythroid progenitor cells. While EpoR in the blood regulates the differentiation of red blood cells, in the brain it protects several types of neuronal cells from death, including A9 dopaminergic neurons (DA) of the Substantia Nigra (SN) and stores oxygen during hypoxia, which is especially important for neurons with an increased need for energy. In addition to the regulation of erythropoiesis in hematopoietic tissues, where *EPOR* mutations are the cause of primary hereditary erythrocytosis and are found in 15% of all cases, erythropoietin is expressed in other tissues, including the nervous system. For functioning, erythropoietin uses a homodimeric receptor (EpoR), which is also widely expressed in the nervous system. The main role of EpoR is to stimulate the proliferation of red blood cell precursor cells for their survival. A 15-fold increase in affinity is caused by the SNP

of -27T> A (rs1006576690) in the TATA box of the *EPOR* gene (K_D = 170 ± 30 nM and 11 ± 3 nM.

Gene		*EPOR*	
Alleles: WT or mut		WT	-27t>A
(5'–3'strands), 26 bp		ctattttgt	ctatAttt
Prediction	-lnK_D	18.24	19.39
	δ		+1.15
Experiments	-lnK_D	15.59	18.33
	δ		+2.74
	K_D, nM	170±30	11±3
	k_a, $M^{-1}s^{-1}*10^3$	(2.8±0,3)	(9.3±0,7)
	k_d, $s^{-1}*10^{-4}$	(4.6±0.6)	(1.0±0.3)
	$t_{1/2}$, min	25±3	115±15

K_D is presented as mean ± standard deviation; δ: the difference between the affinity of hTBP for an ODN with and without the SNP in its TATA box expressed as natural logarithms, δ = -ln (K_D,TATA,Mut) – [-ln (K_D, TATA)]; $t_{1/2}$ = ln2/k_d.

This is accompanied by a 3.3-fold increase in the rate of formation of complexes (k_a) and a 4.6-fold increase in their lifetime (or half-life, $t_{1/2}$). Several studies have shown that cerebral ischemia leads to increased expression of Epo and EpoR to repair damage. Based on this, it can be assumed that substitutions in the TATA-like sequence of the *EPOR* gene promoter are -27T>A, -31C>A and -31C>G, leading to an increase in the affinity of the TBP/TATA interaction by 15.5, 8.1 and 2.1 times, which correlates the level of gene expression may indicate a degree of hereditary ischemic brain damage in carriers of minor alleles and may be candidate markers of this disease.

Conclusions

It is known that hemoglobin genes are expressed in astrocytes of the cortex and hippocampus, in oligodendroglia located in almost all areas of the brain, including the striatum, corpus callosum and medulla oblongata. There are many reports of the association of hemoglobin metabolism disorders with symptoms of mental illness. It is shown that among children with anemia there are significantly more problems with behavior and low intelligence. Many authors believe that cognitive impairment is a causal relationship, and not a consequence of erythropoiesis disorders. Therefore, based on the literature data and the computer-experimental results obtained by us, it can be concluded that the identified candidate SNP markers of erythropoiesis disorders can also be candidate SNP markers of cognitive and mental disorders. Studying the regulatory mechanisms by which detected SNP markers contribute to the development of hemoglobinopathies and associated cognitive impairment will allow doctors not only to take adequate measures to treat hemoglobinopathies in a timely manner, but also to take preventive measures in parallel to prevent or inhibit the development of cognitive and mental disorders. The information received is an additional resource in biomedical research, personalized medicine, diagnostics and drug development.

ACKNOWLEDGMENTS

The work was supported by the budget project No. 0324-2019-0042-c-01

REFERENCES

[1] F.B. Piel, A.P. Patil, R.E. Howes, O.A. Nyangiri, P.W. Gething, M. Dewi, W.H. Temperley, T.N. Williams, D.J. Weatherall, S.L. Hay. "Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates". Lancet, vol. 381, pp. 142-151, 2013.

[2] M..A. Altinoz, B. Ince. "Hemoglobins emerging roles in mental disorders. Metabolical, genetical and immunological aspects". J Dev Neurosci, vol. 61, pp. 73-85, 2017.

[3] M. Biagioli, M. Pinto, D. Cesselli, M. Zaninello, D. Lazarevic, P. Roncaglia, R. Simone, S. Vlachouli, C. Plessy, N. Bertin, A. Beltrami, K. Kobayashi, V. Gallo, C. Santoro, I. Ferrer, S. Rivella, C.A. Beltrami, P. Carninci, E. Pavlova, S. Gustincich. "Unexpected expression of alpha- and betta-globin in mesencephalic dopaminergic neurons and glial cells". Proc Natl Acad Sci USA, vol. 106, pp. 15454-15459, 2009.

DOI 10.18699/BGRS/SB-2020-318

Metabolic response of the Siberian frog *Rana amurensis* to anoxia

Sergei V. Shekhovtsov
ICG SB RAS, Novosibirsk, Russia
shekhovtsov@bionet.nsc.ru

Nina A. Bulakhova
IBPN FEB RAS, Magadan, Russia
sigma44@mail.ru

Yuri P. Tsentalovich
ITC SB RAS, Novosibirsk, Russia
yura.tsentalovich@tomo.nsc.ru

Ekaterina A. Zelentsova
ITC SB RAS, Novosibirsk, Russia
zelentsova.ekaterina@gmail.com

Daniil I. Berman
IBPN FEB RAS, Magadan, Russia
dberman@mail.ru

***Abstract* — The Siberian frog *Rana amurensis* is the only amphibian capable of surviving several months at almost complete anoxia. We performed 1H NMR metabolomics analysis for the liver and heart of the Siberian frog during normoxia and anoxia. We found that anoxia causes significant energetic stress with much less energy molecules available in the studied organs. This is accompanied by the accumulation of several end products, most significantly lactate, alanine, and succinate, as well as most amino acids.**

Keywords* — *anoxia, metabolomics, Rana amurensis, Siberian frog

Motivation and Aim

Hypoxia is a fatal stress for most terrestrial vertebrates. Only a few species are able to survive prolonged hypoxia, most of them are poikiloterms. The recently discovered ability of the Siberian frog *Rana amurensis* to survive complete anoxia for three months [1] makes it the only real anoxia-tolerant amphibian. Noteworthy, these frogs retain the ability to move and react to stimuli under such conditions.

We performed a metabolomics analysis of the heart an liver of Siberian frogs in normoxia and exposed to anoxia in order to (1) quantify the major metabolites present in the frog tissues; (2) elucidate the energetic pathways used under anoxia and their intensity; (3) estimate the degree of cellular stress during anoxia; and (4) to reveal metabolic adaptations to anoxia.

Methods

A set of six experimental animals was exposed to 1 month of anoxia at near-zero temperature; another group of the same size was kept in normoxia at the same temperature. Hearts and liver of the studied sample were extracted as quickly as possible, frozen in liquid nitrogen and used for metabolite extraction. Identification and quantification of metabolites was performed using 1H NMR and mass spectrometry analysis.

Results

Anoxia differed from normoxia in many respects. Glucose, the source of energy for glycolysis, was significantly more abundant during anoxia, as well as end products of glycolysis, lactate and alanine. The Krebs cycle stopped with the accumulation of succinate. Concentrations of most amino acids increased, except for aspartate. Concentrations of AMP and NADH increased severalfold with the concomitant decrease of ATP and NAD+, indicating dramatic decrease in available energetic molecules.

Acknowledgment

Supported by the RFBR (19-04-00312_a) and by Budget Project of FIZ ICiG SB RAS (0324-2019-0040-C-01).

References

[1] Berman, D. I. et al. (2019). The Siberian wood frog survives for months underwater without oxygen. Scientific reports. 9: 13594.

DOI 10.18699/BGRS/SB-2020-319

Prevention of tumor growth by photo- and X-ray activation of tunsgen cluster complex and its conjugate with DNA molecules

Natalia A. Sitnikova
RICEL - Branch of ICG SB RAS
Novosibirsk, Russia

Anastasiya O. Solovieva
RICEL - Branch of ICG SB RAS
Novosibirsk, Russia

Tatiana N. Pozmogova
RICEL - Branch of ICG SB
RASNovosibirsk, Russia

Yuri A. Vorotnikov
RICEL - Branch of ICG SB RAS
NIIC SB RAS
Novosibirsk, Russia

Svetlana M. Miroshnichenko
RICEL - Branch of ICG SB RAS
Novosibirsk, Russia

Michael A. Shestopalov
RICEL - Branch of ICG SB RAS
NIIC SB RAS
Novosibirsk, Russia

Andrey O. Kushnarenko
RICEL - Branch of ICG SB RAS
Novosibirsk, Russia

Abstract — **Photodynamic therapy (PDT) is one of the most promising methods of treating cancer, although it has it's limitations due to applying visible light sources with limited exposure depth. This allows the method to be used only for superficial tumors treatment and can damage healthy cells as well. To increase the efficiency of the method for treatment of deep lying tumors, it is possible to use other radiation sources, for example, x-rays. Thereby search and development of x-ray activated photosensitizers is relevant at the moment. In this work a tungsten (W_6I_8) cluster complex was used as photo- and x-ray- sensitizer and its conjugate with DNA (W_6I_8DNA) as targeting molecules to eliminate Krebs-2 tumor cells and prevent tumor development in vivo.**

Key words — oncology; Krebs-2; photodinamic therapy; X-ray; tunsgen claster complex; DNA moleculs; cancer treatment

Introduction

Oncological diseases are widespread in the world and causing a large number of deaths. Well-known treatment methods of cancer such as surgery, radiation therapy and chemotherapy are efficient in some way, but do not prevent relapse of tumors. Photodynamic therapy (PDT) is new and one of the most promising methods of cancer treatment. It is based on an ability of photosensitizers to be activated by specific wavelength and produce singlet forms of oxygen that are known for its cytotoxic effect. Visible light sources are used for photosensitizer activation, but it can be applied only to superficial tumors. To increase the efficiency of the method, it is possible to use other radiation sources, for example, x-rays that can penetrate deep tissues. Chen and colleges have already created x-ray activated nanoparticles, but it requires high irradiation dose [1]. It was detected that certain tungsten cluster complex has an ability to be activated by light and x-ray and produce reactive oxygen species (ROS) [2], [3]. Such properties allow it to be possible photo- and x-ray sensitizer. In this study, we also used tungsten cluster complex conjugated with DNA molecules for the better targeting of induced cancer cells. It has been shown that certain type of tumor cells absorb DNA molecules more efficiently than other cancer cells and normal cells [3].

The main goal for our work is to exanimate tungsten cluster complex and its conjugate with DNA for photodynamic and x-ray dynamic abilities on in vivo Krebs-2 tumor cells.

Materials and Methods

Tungsten cluster complex $(Bu_4N)_2[\{W_6I_8\}(NO_3)_6]$ was synthesized in Laboratory of Bioactive Inorganic Compounds (Nikolaev Institute of Inorganic Chemistry SB RAS) [4]. Its conjugation with DNA molecules was carried out by intercalating the cluster into double-stranded DNA using the DNA denaturation step.

Krebs-2 mouse carcinoma line cells were used in current study. Cells were cultivated with (W_6I_8) and (W_6I_8DNA) for 2 hours in standard culture conditions (37°C, 5% CO_2). After that, cells were irradiated with 500W halogen lamp for 30 minutes or with x-ray source. The resulting radiation dose was 20 J/cm-2 (light) and 120Gr (x-ray).

CBA/Lac mice 4 month old were vaccinated with irradiated cancer Krebs-2 cells intramuscularly (200 $*10^3$ cells). The next 42 days dynamic of tumor growth was observed. On 42d day all mice with tumors were decapitated and tumors were collected for histological analyze. Control group consisted with mice that received cells with no exposure to light or x-ray. All the animal studies conform to standards and a protocol approved by the Institutional Animal Care and Use Committee.

Statistic analyze was carried out by Statistic 8. Quantitative data were expressed as mean ± s.e.m. A nonparametric Mann-Whitney U-test was used for comparing the treatment groups and the control groups. $P < 0.05$ was considered statistically significant.

Results

Solid tumors Krebs-2 appear palpable on 7-11th day after injection in 100% mice in control group. Other control groups,

where cells were previously irradiated with light or x-ray, show no difference in growth dynamic and final size of tumors.

Preliminary incubation of Krebs-2 cells with tungsten cluster complex $(Bu_4N)_2[\{W_6I_8\}(NO_3)_6]$ also did not affect tumor growth in mice itself (control W_6I_8). However, irradiation of these cells with x-ray led to the latter appearance of tumors and smaller sizes in comparison with both control W_6I_8 and light W_6I_8 (Fig. 1).

Incubation of Krebs-2 cells with tungsten cluster complex $(Bu_4N)_2[\{W_6I_8\}(NO_3)_6]$ conjugate with DNA molecules did not affect tumor growth in mice without activation of cluster. However, irradiation of these cells with x-ray, as in the case with just W_6I_8, led to the latter appearance of tumors and smaller sizes in comparison with both control W_6I_8-DNA and light W_6I_8-DNA. In some mice in this group tumors did not appeared at all (Fig. 1). Incubation of Krebs-2 cells with tungsten cluster complex $(Bu_4N)_2[\{W_6I_8\}(NO_3)_6]$ and with its conjugate with DNA molecules has no differences between sizes of tumors induced by injections of such cells both without irradiation and with light or x-ray irradiation.

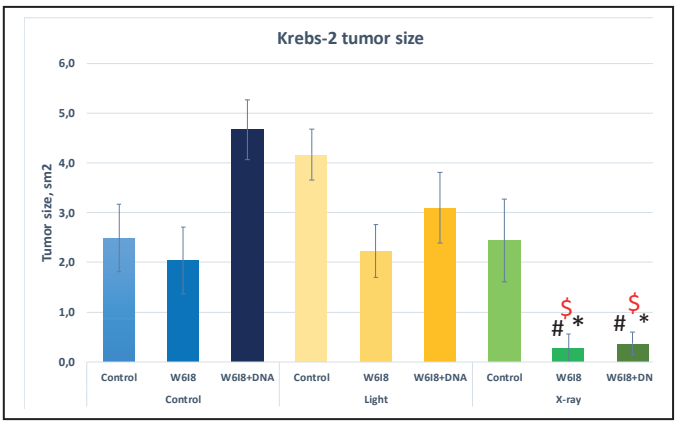


Fig. 1. Krebs-2 tumor size (sm^2) after 42 days of cultivation in mice CBA/Lac 4-6 month old. Krebs-2 cells were previously incubated with (W6I8) and (W6I8DNA) for 2 hours in standard culture conditions (37°C, 5% CO2). After that, cells were irradiated with 500W halogen lamp for 30 minutes or with x-ray source. The resulting radiation dose was 20 J/cm-2 (light) and 120Gr (x-ray). Group with tumors from cells with no exposure to light or x-ray was control group. Quantitative data were expressed as mean ± s.e.m. N=6. A nonparametric Mann-Uitny test was used for comparing the treatment groups and the control groups.

(* - $p < 0.05$) – X-ray compared to light, (# - $p < 0.05$) – x-ray compared to control, ($ - $p < 0.05$) – (W6I8) and (W6I8DNA) compared with control in X-ray group.

Conclusion

Therefore we found out that Tungsten cluster complex $(Bu_4N)_2[\{W_6I_8\}(NO_3)_6]$ and its conjugate with DNA (W_6I_8DNA) can be activated with x-ray which leads to elimination of Krebs-2 cells. Considering, that W_6I_8 itself has an ability to produce ROS, than this part of conjugate is responsible for elimination. Irradiated Krebs-2 cells intramuscular injection reduce tumor growth or leads to its absence. Meanwhile in control group Krebs-2 tumors reached considerable size. As we know that cancer initiating cells absorb DNA molecules more efficiently [3], we can assume that this type of tumor cells was eliminated specifically and tumor cells suspension was not capable to induce tumor growth or the growth was not so fast and efficient.

Furthermore, it was revealed that tungsten cluster complexes $(Bu_4N)_2[\{W_6I_8\}(NO_3)_6]$ have no effect on cells viability, which is one of the essential properties for photo- and x-ray sensitizers. Consequently, X-ray irradiation can be an activator of particular photosensitizers, since there is no toxic effect on living cells on its own. The use of X-rays will reduce the likelihood of death of normal cells and increase the depth of the location of tumors that can be removed by this method. However, to evaluate the effectiveness of X-ray irradiation in photodynamic therapy, further studies are necessary.

Acknowledgment

This research is supported by RFBR № 18-315-00235.

References

[1] H. Chen, G. Wang, Y. Chuang et al. “Nanoscintillator-mediated X-ray inducible photodynamic therapy for in vivo cancer treatment,” Nano Lett., vol. 15, no. 4, pp. 2249–2256, 2015.

[2] E. V. Svezhentseva, Y. A. Vorotnikov et al. “From photoinduced to dark cytotoxicity through an octahedral cluster hydrolysis,” Chem. - A Eur. J., vol. 24, no. 68, pp. 17915–17920, 2018.

[3] D. V. Evtushok, A. R. Melnikov, N. A. Vorotnikova et al. “A comparative study of optical properties and X-ray induced luminescence of octahedral molybdenum and tungsten cluster complexes,” Dalt. Trans., vol. 46, no. 35, pp. 11738–11747, 2017.

[4] E. V. Dolgova, E. A. Potter, A. S. Proskurina et al. “Properties of internalization factors contributing to the uptake of extracellular DNA into tumor-initiating stem cells of mouse Krebs-2 cell line,” Stem Cell Res. Ther., vol. 7, no. 1, pp. 1–16, 2016.

DOI 10.18699/BGRS/SB-2020-320

Cytokines of the IL-6 family in the modulation of carbohydrate and lipid metabolism in patients with metabolic syndrome

Daria Skuratovskaia
IKBFU, Kaliningrad, Russia
DSkuratovskaya@kantiana.ru

Alexandra Komar
IKBFU, Kaliningrad, Russia
alexandkomar@gmail.com

Maria Vulf
IKBFU, Kaliningrad, Russia
mary-jean@yandex.ru

Elena Kirienkova
IKBFU, Kaliningrad, Russia
elenamed@list.ru

Egor Shunkin
IKBFU, Kaliningrad, Russia
egor.shunkin@gmail.com

Larisa Litvinova
IKBFU, Kaliningrad, Russia
larisalitvinova@yandex.ru

***Abstract* — According to the world community, metabolic complications are associated with inflammation in obesity. The cytokines of the interleukins (IL) - 6 family consist of IL-6, IL-11, IL-27 and have similar, but individual effects, such as the regulation of liver damage, the balance between regulatory and effector T cells and metabolic regulation. The anti-inflammatory and regenerative activities of IL-6 are usually mediated by the classical signaling regimen, while the pro-inflammatory responses in many pathological conditions include trans-signaling. The aim of the study was to assess the role of the cytokines of the IL-6 family: IL-6, sIL-6Ra, gp130 / sIL-6Rb, IL-11 in the regulation of carbohydrate (glucose level) and lipid (cholesterol (Hol), high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides (TR)) exchanges in serum / plasma in patients with metabolic syndrome (MetS). This fact may indicate the protective role of sIL-6Ra trans-signaling and protective role of IL-6 signal inhibition due to high levels of sgp130 / sIL-6Rb in relation to metabolic disorders in obesity patients.**

Keywords* — *IL-6, sIL-6Ra, gp130 / sIL-6Rb, IL-11, metabolic syndrome, cholesterol, glucose

Introduction

Obesity and accompanying complications (type 2 diabetes mellitus (type 2 diabetes), diabetic foot, atherosclerosis, non-alcoholic fatty liver disease (NAFLD), steatohepatitis, etc.) occupy a leading position among the causes of mortality [1]. The main signs of metabolic syndrome (MetS) are arterial hypertension, central obesity, hyperglycemia and atherogenic dyslipidemia, and concomitant signs are liver steatosis, polycystic ovary syndrome and hyperuricemia [2]. The main long-term complications of MetS are type 2 diabetes, atherogenesis, and cognitive impairment [2]. According to the world community, metabolic complications are associated with inflammation in obesity. Chronic subclinical inflammation contributes to the lipid and carbohydrate metabolism disorders, and its local foci lead to dysfunction of different tissues (adipose tissue, hepatic, vascular endothelium, etc.).

Initially, it was believed that IL-6 is assigned to pro-inflammatory cytokines and is a predictor of type 2 diabetes, secreted by TNF-a-mediated manner [3; 4]. IL-6 is known to mediates adverse metabolic effects, promotes insulin resistance and worsens glucose homeostasis [4; 5]. However, the role of IL-6 in potentiation of metabolic disorders is still controversial. Different types of cells involved in the regulation of glucose homeostasis respond differently to IL-6. Many experiments have shown that IL-6 can potentiate insulin resistance in hepatocytes, but the results on adipocytes and skeletal muscles are contradictory/inconsistent [3; 6].

The interleukin (IL) - 6 family cytokines is a group of cytokine consisting of IL-6, IL-11, IL-27 and belong to the same family, because the receptor complex of each cytokine contains two (for IL-6 and IL-11) or one molecule (for all other cytokines) of the gp130 signaling receptor subunit [6]. Cytokines of the IL-6 family have similar, but individual effects, such as the regulation of liver damage, the balance between regulatory and effector T cells, and metabolic regulation [6].

IL-6 realizes its effects through the membrane-bound receptor of IL-6R, which can exist in a soluble form (sIL-6R), the latter lacks the transmembrane and cytoplasmic domains. The main sources of sIL-6R are hepatocytes and immune cells. IL-6 binds to IL-6R or sIL-6R and initiates subsequent signaling through interaction with gp130 [3; 6]. The IL-6 / sIL-6R complex acts as an agonist of gp130-mediated IL-6 signaling, which is called trans-signaling. This type of signal transmission expands the spectrum of potential IL-6 targets due to the ubiquitous expression of gp130. However, gp130 can also be in soluble form and inhibit IL-6 trans-signaling without affecting the classical type of signal transmission [6]. Soluble gp130 (sgp130) is found in the bloodstream of healthy people by buffering the systemic response to circulating IL-6 [6; 7].

The anti-inflammatory and regenerative activities of IL-6 are usually mediated by the classical signaling regimen, while the pro-inflammatory responses in many pathological conditions include trans-signaling.

In this regard, the purpose of the study was to assess the role of the IL-6 family cytokines: IL-6, sIL-6Ra, gp130 / sIL-6Rb, IL-11 in the regulation of components carbohydrate and lipid metabolism (glucose level, cholesterol (Hol), high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides (TR)) in serum / plasma in patients with metabolic syndrome (MetS).

Methods and Materials

This study included 118 patients with MetS (53 men and 65 women, 48 (41 – 53) years, BMI = 46.25 (41.48 – 54.68) kg/m^2) who were admitted to the Regional Clinical Hospital.

All patients were diagnosed with abdominal obesity. The presence of MetS was confirmed based on a detailed clinical and instrumental examination using the World Health Organization diagnostic criteria (Swarup S, Zeltser R, 2020). The control group consisted of healthy donors with normal anthropometric and biochemical parameters (23 men and 17 women, 35 (30.5 – 49.5) years, BMI = 23.4 (20.9 – 24.9) kg/m^2, n=40).

All study participants provided informed consent to participate in a research study. The study was carried out in accordance with the World Medical Association (WMA) Declaration of Helsinki (2000) and the Protocol to the Convention on Human Rights and Biomedicine (1999). The study protocol was approved by the Local Ethical Committee of the Innovation Park of the Immanuel Kant Baltic Federal University (Protocol No. 4 from October 23, 2013).

Mediators of carbohydrate (glucose (GL)), and lipid metabolism (cholesterol (Hol), high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides (TR), C-reactive protein (CRP)) were measured on a CA-180 automatic biochemical analyzer (Furuno Electric Co., Ltd, Japan).

Quantitative determination of IL-6, sIL-6Ra, gp130/sIL-6Rb, IL-11 was evaluated by flow fluorimetry on an analyzer «Bio-Plex Protein Assay System» (Bio-Rad, USA) using Bio-Plex Pro™ Human Inflammation Panel 1, 37-Plex (Bio-Rad, USA). Verification of quantitative indicator distribution normality was carried out using the Shapiro-Wilk test. For detecting statistically significant differences between groups, a pairwise analysis was performed using the nonparametric Mann-Whitney criterion for independent groups was used. Correlations between the studied indices were determined using Spearman correlation analysis. Differences were considered significant at the level of $p < 0.05$. All statistical analyses have been performed in GraphPad Prism 8 (USA).

Result and Discussion

Biochemical parameters of carbohydrate and lipid metabolism

Ectopic lipid deposition, chronic inflammation, disorder carbohydrate and lipid metabolism are characteristics of the metabolic syndrome (MetS) [1; 2]. The concept of metabolic syndrome combines the presence of obesity and at least two concomitant signs of metabolic disorders, according to the criteria for determining MetS of the World Health Organization [1]. The study revealed the disorder of carbohydrate and lipid metabolism in patients with MetS. glucose, cholesterol, LDL, triglycerides serum levels in patients with MetS were higher compared to controls ($p < 0.05$), and the HDL levels were lower than control values ($p < 0.05$) (Table 1).

Pro-inflammatory mediators

Obesity inflammation plays an important role in the development of MetS and other complications CRP is an inflammatory biomarker and refers to acute phase proteins. Elevated CRP levels are a companion to metabolic disorders, in particular MetS [8; 9]. In our study, the CRP level in patients with MetS (9.31 (7.37 - 14.15) pg / ml) was higher than the control (3.96 (1.56 - 8.83) pg / ml) ($p < 0.05$), which indicates the presence of systemic inflammation in this category of patients (Table 1).

TABLE 1. BIOCHEMICAL PARAMETERS OF CARBOHYDRATE AND LIPID METABOLISM IN BLOOD SERUM IN THE PATIENT GROUPS

Parameters (mmol/l)	**Control, n=40**	**Patients with MetS, n=118**	**p-value**
Glucose	5.09 (4.66 – 5.36)	6.50 (5.40 – 9.33)	p = 0.0011*
HOL	4.42 (3.80 – 5.17)	5.29 (4.34 – 5.96)	p < 0.0001*
HDL	1.46 (1.05 – 1.71)	1.16 (0.96 – 1.34)	p = 0.0140*
LDL	2.19 (1.65 – 2.47)	2.79 (2.25 – 3.38)	p = 0.0002*
TR	0.86 (0.65 – 1.37)	1.85 (1.27 – 2.44)	p < 0.0001*
CRP	3.96 (1.56 – 8.83)	9.31 (7.37 – 14.15	p = 0.0003*

Note: the significance is determined using the Mann-Whitney criterion for two independent samples (* - p <0.05).

CRP is formed and released from the liver by the action of the cytokines IL-1 and IL-6 [9]. In our IL-6 study, plasma levels in patients with MetS (4.59 (3.99 - 6.59) pg / ml) exceeded the control of 1.45 (0.55 - 6.59) pg / ml ($p < 0.01$) (Figure 1). However, IL-6 levels were negatively correlated with plasma CRP levels in patients with MetS ($r = - 0.522$, $p < 0.05$).

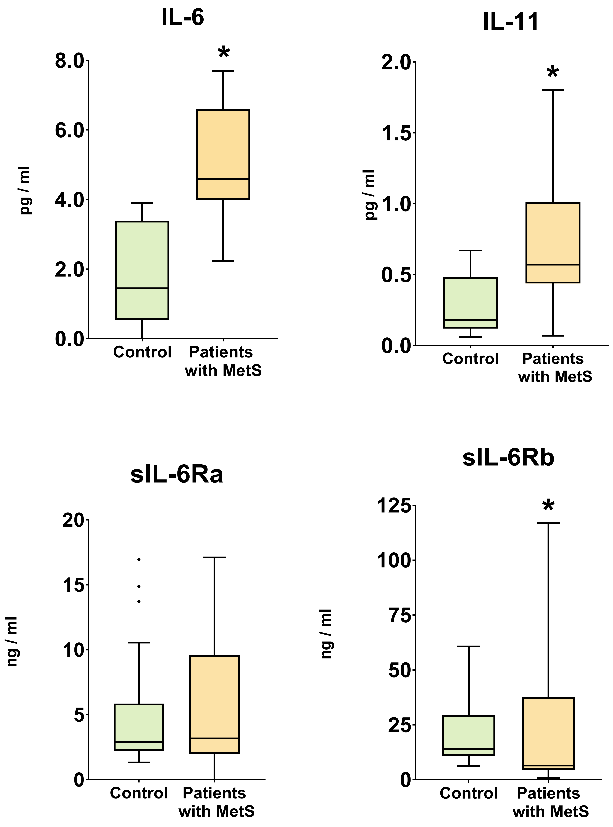


Fig. 1. The content of pro-inflammatory mediators in the blood plasma of patients with metabolic syndrome

IL-6 is usually not expressed constitutively, but its expression is intensively induced by a spectrum of stimuli, such as viral and bacterial infections, pro-inflammatory cytokines, oxidative stress, etc. [10]. IL-6 was initially considered a pro-inflammatory cytokine. There is an opinion that IL-6 and CRP belong to markers of type 2 diabetes [9; 10]. In our study, an increase in IL-6 level was positively correlated with an increase in BMI ($r = 0.56$, $p < 0.05$) and an increase in glucose concentration ($r = 0.50$, $p < 0.05$) in the blood serum of patients with MetS (Fig. 2). The observation that serum IL-6 levels correlate with obesity, impaired carbohydrate and lipid metabolism suggests that IL-6 is associated with the development of metabolic diseases.

However, there is increasing evidence of anti-inflammatory activity of IL-6. Mice with IL-6 deficiency develop obesity in adulthood, liver inflammation, and systemic insulin resistance [3]. In accordance with these observations, elevated cholesterol and triglyceride levels, accompanied by weight gain, were observed in patients who were treated with IL-6 inhibitors [3; 10]. Thus, the association between IL-6 and the progression of metabolic disorders in obesity may simply reflect an attempt to counteract the chronic inflammation caused by other inflammatory mediators.

The explanation may lie in a change in the signaling of IL-6 in conditions of chronic inflammation in obesity. The scientific periodicals describe the properties of IL-6 dependent signaling in the regulation of metabolic parameters in obesity. sIL-6R binds IL-6 and prolongs the half-decay of IL-6 in animal models [11], enhancing regenerative processes in the liver. In addition, it was shown that the complex of IL-6 and sIL-6R internalizes much less efficiently than IL-6, which leads to an increase in the duration of the IL-6 signal during trans-signaling [11]. In our study, sIL-6Ra did not change plasma levels, while gp130 / sIL-6Rb decreased in patients with MetS 6.49 (4.44 - 37.62) ng / ml compared to the control group 13.9 (10.9 - 29.4) ng / ml ($p < 0.05$). The level of soluble sIL-6Ra and gp130 / sIL-6Rb receptors was negatively correlated with BMI ($r = -0.43$, $p < 0.05$) ($r = -0.45$, $p < 0.05$), with LDL level ($r = -0.33$, $p < 0.05$) ($r = -0.30$, $p < 0.05$) and positively correlated with the level of HDL ($r = 0.33$, $p < 0.05$) ($r = 0.29$, $p < 0.05$) in MetS patients (Figure 2). The level of gp130 / sIL-6Rb, but not the level of sIL-6Ra, negatively correlated with the level of glucose ($r = -0.29$, $p < 0.05$) in MetS patients (Figure 2).

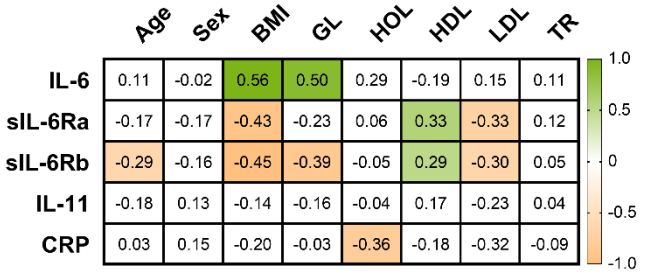


Fig. 2. Correlation matrix of biochemical parameters and pro-inflammatory IL-6 family cytokines in patients with MetS

IL-6 level negatively correlated with sIL-6Ra ($r = -0.59$) and gp130 / sIL-6Rb ($r = -0.45$) levels in MetS patients ($p < 0.05$). The increase in the level of IL-6 in plasma was accompanied by a decrease in concentration gp130 / sIL-6Rb, which is probably associated with impaired signaling of IL-6 and the development of metabolic complications.

IL-11 promotes fibrosis, steatosis, hepatocyte death, inflammation and hyperglycemia in mice with diet-induced steatohepatitis. Inhibition of IL-11 improved the cardiometabolic profile in animal models [12]. In contrast, other authors have shown that IL-11 inhibits NF-kappaB and AP-1 [13]. In this study, we recorded increased IL-11 loss in patients with MetS 0.57 (0.44 – 1.01) pg / ml compared to a control of 0.18 (0.12 – 0.48) pg / ml ($p < 0.05$). However, we did not find any correlation of IL-11 level with other studied parameters, which does not allow us to demonstrate its role in the pathogenesis of MetS.

Conclusion

Thus, a decrease in the concentration of soluble sIL-6Ra and gp130 / sIL-6Rb receptors was accompanied by increased dysregulation of carbohydrate and lipid metabolism.sIL-6Ra did not change, but was negatively correlated with BMI, LDL levels and HDL levels in blood plasma of patients with MetS. A decrease in the concentration of the soluble gp130 / sIL-6Rb receptor was accompanied by metabolic dysregulation: an increase in BMI, glucose levels, LDL levels and decrease in HDL levels in blood plasma of patients with MetS.

This fact may indicate the protective role of sIL-6Ra trans-signaling and protective role of IL-6 signal inhibition due to high levels of sgp130 / sIL-6Rb in relation to metabolic disorders in obesity patients.

Acknowledgment

The research was supported by Russian Foundation for Basic Research [No. 18-015-00084-a to L.L. (patient recruitment and classification), by Russian Foundation for Basic Research and KLG [No.19-415-393004 - r_mol_a to D.S. (biochemical parameters), No.19-44-390005 - r_a to D.S. (flow fluorimetry)].

References

[1] S. Swarup and R. Zeltser, Metabolic syndrome, StatPearls Publishing, 2020.

[2] C.A. Aguilar-Salinas and T. Viveros-Ruiz “Recent advances in managing/understanding the metabolic syndrome.” F1000Res, vol. 8, April 2019.

[3] D. Qu, J. Liu, C.W. Lau and Y. Huang “IL-6 in diabetes and cardiovascular complications.” Br. J. Pharmacol, August 2014, vol. 171(15), pp. 3595-3603.

[4] H.M. Sarbijani, M. Khoshnia and A. Marjani (2016) “The association between Metabolic Syndrome and serum levels of lipid peroxidation and interleukin-6 in Gorgan.” Diabetes. Metab. Syndr, March 2016, vol. 10(1 Suppl 1), pp. 86-89.

[5] J.P. Bastard et al. ”Elevated levels of interleukin 6 are reduced in serum and subcutaneous adipose tissue of obese women after weight loss.” J. Clin. Endocrino.l Metab, September 2000, vol. 85(9), pp. 3338-3342.

[6] S. Rose-John “Interleukin-6 Family Cytokines.” Cold. Spring. Harb. Perspect Biol., February 2018, vol. 10(2).

[7] A.J. Covarrubias and T. Horng “IL-6 strikes a balance in metabolic inflammation.” Cell Metab., June 2014, vol. 19(6), pp. 898-899.

[8] S. Devaraj et al. (2008) “CRP and adiponectin and its oligomers in the metabolic syndrome: evaluation of new laboratory-based biomarkers.” Am. J. Clin. Pathol., May 2008, vol. 129(5), pp. 815-822.

[9] M. Blüher et al. “Association of interleukin-6, C-reactive protein, interleukin-10 and adiponectin plasma concentrations with measures of obesity, insulin sensitivity and glucose metabolism.” Exp. Clin. Endocrinol. Diabetes, October 2005, vol. 113(9), pp. 534-537.

[10] S. Shelbaya et al. “Study of the role of interleukin-6 and highly sensitive C-reactive protein in diabetic nephropathy in type 1 diabetic patients.” Eur. Rev. Med. Pharmacol. Sci., February 2012, vol. 16(2), pp.176-182.

[11] D. Schmidt-Arras and S. Rose-John “IL-6 pathway in the liver: From physiopathology to therapy.” J. Hepatol., June 2016, vol. 64(6), pp. 1403-1415.

[12] A.A. Widjaja et al. (2019) “Inhibiting Interleukin 11 Signaling Reduces Hepatocyte Death and Liver Fibrosis, Inflammation, and Steatosis in Mouse Models of Nonalcoholic Steatohepatitis.” Gastroenterology, September 2019, vol. 157(3), pp. 777-792.

[13] A. Lgssiar et al. “Interleukin-11 inhibits NF-kappaB and AP-1 activation in islets and prevents diabetes induced with streptozotocin in mice.” Exp. Biol. Med. (Maywood)., May 2004, vol. 229(5), pp. 425-436.

DOI 10.18699/BGRS/SB-2020-321

Gene polymorphism IL13 in moderate-to-severe asthmatic Siberian children with different diseases control

Marina Smolnikova
Laboratory of cell-molecular physiology and pathology
Scientific Research Institute of Medical Problems of the North FRC KSC SB RAS
Krasnoyarsk, Russia
smarinv@yandex.ru

Nina Gorbacheva
Clinical department somatic and mental health of children
Scientific Research Institute of Medical Problems of the North FRC KSC SB RAS
Krasnoyarsk, Russia
n-n-gorbacheva@yandex.ru

Marina Malinchik
Laboratory of cell-molecular physiology and pathology
Scientific Research Institute of Medical Problems of the North FRC KSC SB RAS
Krasnoyarsk, Russia
seapearl1995@gmail.com

Sergey Tereshchenko
Clinical department somatic and mental health of children
Scientific Research Institute of Medical Problems of the North FRC KSC SB RAS
Krasnoyarsk, Russia
legise@mail.ru

Abstract — **Multifunctional cytokines, such as IL-13, play an important role in asthma pathogenesis. Several single nucleotide polymorphisms of cytokines are associated with asthma susceptibility in specific populations; however, further replicative studies in other ethnic groups are mandatory, especially with respect to asthma severity in children. Child with moderate-to-severe asthma were divided to patients uncontrolled course of the disease (n = 107) and controlled course of the disease (n = 95). The frequency of the CT genotype IL13 (rs1800925) in patients with controlled moderate-to-severe asthma was significantly higher in comparison with control (46,2% / 36,6%, OR 1,51, p=0,03). Was obtained statistically significant differences in the frequency of the TT genotype between the population sample (6,7%) and the group with uncontrolled asthma (15,5%, OR 1,65, p=0,03). The results have contributed to the data on the role of polymorphisms of IL-13 into the development of asthma in children as exemplified by a European population of East Siberia, Russia.**

Keywords — bronchial asthma, IL13, gene polymorphism, disease severity

Introduction

Bronchial asthma (BA) is a multifactorial disease, since both environmental factors and the person's genetic predisposition contribute to the development of the disease [1, 6]. A number of associative relationships of cytokine gene polymorphism produced by Th2 cells with the development of bronchial asthma have been obtained [1]. Interleukin-13 (IL-13) is the central mediator involved in the pathogenesis of bronchial asthma. In the bronchial tissues in asthmatic patients, there is a cell expansion expressing IL-13 mRNA, with overexpression of IL-13 in the lungs accompanied by the development of allergic inflammation, mucus hypersecretion, airways hyperresponsiveness, subepithelial fibrosis and eotaxin production [2]. Currently, anti-IL-13 (the trade name is "Lebrikizumab") is used in the treatment of bronchial asthma, which is a monoclonal antibody that blocks the activity of IL-13. When applying Lebrikizumab in patients, a significant improvement in the volume of forced expiratory flow for the first second of maneuver (FEV1) was observed [3].

The IL13 gene has a number of polymorphisms associated with the clinical manifestations of BA. Of particular interest is the study of rs1800925 polymorphism due to its location in the promoter region of the IL13 gene, as well as the revealed association between this SNP and BA in both adult and children populations [4]. Carriers of the homozygous variant for the minor allele of this polymorphism developed an association with increased IL-13 production, airway hyperreactivity, and a positive skin test for allergen, and heterozygotes and homozygotes for the minor allele are sure to be associated with an elevated concentration of IgE [4, 5].

The purpose of the study was to evaluate the associations between rs1800925 polymorphism of the IL13 gene and bronchial asthma with different control levels and disease severity in children of Caucasian origin in Eastern Siberia.

Materials and methods

In the studied group of people there were children with bronchial asthma (n = 202, average age was 12.8 ± 1.2 years old) and practically healthy children and adults (n = 135) in Krasnoyarsk. The control group included children (n = 33, average age was 13.6 ± 2.5 years old) and adults with no history of BA and allergies (n = 102, average age 38.3 ± 5.4 years). A comparative analysis of the allele frequency in the control groups of different ages was carried out, no statistically significant differences were found, which allowed one to combine individuals of different ages in one control group. Child patients were divided into groups: moderate-to-severe BA with an uncontrolled course of the disease (UBA, n = 107) and moderate BA with a controlled course of the disease (CBA, n = 95). The diagnosis, severity and the disease progression control was established upon recommendation of the GINA-2018.

DNA from peripheral blood was isolated using salting out method. Genotyping of IL13 (rs1800925) was carried out using the RT-PCR. Comparison of the allele frequency and genotypes between groups was provided using an online calculator, χ-square test (http://gen-exp.ru/calculator_or.php).

Results and discussion

The frequency of the alleles rs1800925 of the IL13 gene in the control group obtained during the study corresponds the distribution in Caucasoid population: the frequency of the C * allele is 75%, the T * allele is 25% (according to http://www.ensembl.org).

A statistically significant difference in the frequency of genotypes between BA patients and the control group was shown by comparison of the rs1800925 genotype frequency. Genotypes with the T* allele are more common in patients as compared to control (OR 1.53 (1.08-2.15), p=0.027), which might indicate that CT and TT genotypes are risk factors for BA development. The frequency of the CT genotype in patients with UBA is statistically significantly higher than in the control group (46.9% and 46.6 % vs. 36.3 %, p=0.008, p=0.033, respectively). This genotype can be assumed to be associated with UBA. Significant differences comparing the TT genotype frequency between the control group and the group with severe BA was also obtained. Thus, in the severe BA group, the homozygous TT genotype is more common (15.49% vs. 6.7 %, p=0.027). Therefore, this genotype is associated with the progression of pathology, i.e. the development of severe bronchial asthma.

In general, our results are consistent with previous studies on the association of bronchial asthma with the rs1800925 polymorphism of the IL13 gene. According to Liu Z. et al., an association was defined between the rs1800925 IL13 polymorphism and the risk of asthma in children, since the CT and TT genotypes compared to the CC genotype, as well as high IL-13 expression in serum, were more common in the patient group. Using the further stratification of the study on ethnicity, a higher risk of the BA development in Caucasians compared to Mongoloids was shown [6]. Meta-analysis performed by Yang H.et al., revealed that rs1800925 polymorphism is associated with the risk of BA development among Caucasians [7].

CONCLUSION

Thus, Th2-lymphocytes play an important role in the pathogenesis of bronchial asthma, and some single-nucleotide variants of the IL13 gene are associated with a risk of the development, course and progression of bronchial asthma in children was confirmed by the obtained data.

REFERENCES

[1] S. Georas, J. Gue, U. Fanis, V. Casolaro, “T-helper cell type-2 regulation in allergic disease”, European respiratory journal, 2005, vol. 26, pp. 933–947.

[2] X.X. Yang, F.X. Li, Y.S. Wu, D. Wu, J.Y. Tan, M. Li, “Association of TGF-beta1, IL-4 and IL-13 gene polymerphisms with asthma in a Chinese population”, Asian Pacific Journal of Allergy and Immunology, 2011, vol. 29(3), pp. 273-277.

[3] J. Corren, “Role of interleukin-13 in asthma. Curr Allergy Asthma Rep.”, 2013, vol. 13(5), pp. 415-420.

[4] L. Cui, J. Jia, C.F. Ma, S.Y. Li, Y.P. Wang, X.M. Guo, Q. Li, H.B. Yu, W.H. Liu, L.B. Gao, “IL-13 polymorphisms contribute to the risk of asthma: a meta-analysis”, Clinical Biochemistry, 2012, vol. 45, pp. 285–288.

[5] B. Beghe, I.P. Hall, S.G. Parker, M.F. Moffatt, A. Wardlaw, M.J. Connolly, L.M. Fabbri, C. Ruse, I. Sayers, “Polymorphisms in IL13 pathway genes in asthma and chronic obstructive pulmonary disease”, Allergy, 2010, vol. 65(4), pp. 474-481.

[6] Z. Liu, P. Li, J. Wang, Q. Fan, P. Yan, X. Zhang, B. Han, “A meta-analysis of IL-13 polymorphisms and pediatric asthma risk”, Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, 2014, vol. 11(20), pp. 2617-2623.

[7] H. Yang, H. Dong, Y. Dai, Y. Zheng, “Association of interleukin-13 C-1112T and G+2044A polymorphisms with asthma: a meta-analysis”, Respirology, 2011, vol. 16(7), pp. 1127-1135.

DOI 10.18699/BGRS/SB-2020-323

Quantitative differences in the proteomic composition of the blood serum of patients with simple and paranoid schizophrenia

Smirnova Liudmila
Laboratory of Molecular Genetics and Biochemistry
Mental Health Research Institute of Tomsk National Research Medical Center
Tomsk, Russia
lpsmirnova@yandex.ru

Zgoda Victor
Laboratory of systems biology
Institute of Biomedical Chemistry
Moscow, Russia
vic@ibmh.msk.su

Dmitrieva Elena
Laboratory of Molecular Genetics and Biochemistry
Mental Health Research Institute of Tomsk National Research Medical Center
Tomsk, Russia
lena-312tom@yandex.ru

Seregin Alexandr
Laboratory of Molecular Genetics and Biochemistry
Mental Health Research Institute of Tomsk National Research Medical Center
Tomsk, Russia
lena-312tom@yandex.ru

Semke Arkadiy
Department of Endogenous Disorders
Mental Health Research Institute of Tomsk National Research Medical Center
Tomsk, Russia
arkan1959@mail.ru

Ivanova Svetlana
Laboratory of Molecular Genetics and Biochemistry
Mental Health Research Institute of Tomsk National Research Medical Center
Tomsk, Russia
ivanovaniipz@gmail.com

Abstract — **In the present study we analyzed the protein spectrum of blood in schizophrenic patients with simple and paranoid form of diagnosis. A set of proteins in schizophrenia were typically associated with processes, which are responsible for protein synthesis and the processes of transduction and translation; immune response, oxidative stress, apoptosis and cell communication. Significant increase in the number of RIPK1, mGluR6 receptors in blood serum between patients and healthy controls was revealed by comparative analysis. Besides, the amount of mGluR6 was significantly higher in patients with simple form of schizophrenia compared with paranoid schizophrenia patients (p = 0.021). A quantitative assessment of the specific minor proteins that we studied, using labeled standard peptides, demonstrated an increase in their number depending on the severity of the disease (a simple form of schizophrenia with leading negative symptoms).**

Keywords — biomarker, schizophrenia, proteome, mass spectrometry, serum

Introduction

The study of mental disorders becomes more relevant in the 21st century. Schizophrenia refers to socially significant diseases, since it is a chorionic disease that begins at a young age. Its etiology and pathogenesis, response to treatment and its consequences are not yet completely understood. As a rule, only anamnestic and clinical psychopathological data are used for diagnostics. Patients, entering a psychiatric clinic for the first time, have particular difficulties in making a diagnosis. At present, the search for peripheral biomarkers, which can be used for differential diagnosis and prognosis of therapy (theranostic), becomes particularly important. Work on proteomic analysis in schizophrenic patients was carried out mainly on post-mortem material. However, so far, the search for markers of mental disorders has not been successful. The identification of peripheral biomarkers is an important step towards personalized medicine [1][2].

Methods

The protein spectrum of the blood serum was analyzed in following groups: 20 patients (9 men, 11 women, aged 32 [26, 41] years and a disease duration of 7 [4, 16] years) with acute paranoid schizophrenia (F20.0); 15 patients (6 men, 9 women, aged 34 [29, 44] years and a disease duration of 10 [8, 20] years) with simple schizophrenia (F20.6) and 10 healthy people, the average age of 32,6 [22; 37] years. The patients were hospitalized in the Department of Endogenous Disorders of the Research Institute of Mental Health (MHRI) of the Tomsk National Research Medical Centre (NRMC, Tomsk). Blood serum samples were obtained from most patients in acute state before the start of the study. This was done in order to get clean samples before starting their course of therapy. The serum was purified by affinity chromatography from six major proteins (albumin, immunoglobulin G, immunoglobulin A, antitrypsin, transferrin and haptoglobin). In the next step the purified proteins were concentrated to 1 ml using Amicon Ultra-0.5 ultrafilters (MILLIPORE) at 3 kDa and separated by vertical electrophoresis in a 12% polyacrylamide gel according to the Lemmli method. Then, after trypsinolysis and peptide extraction from the gel, the proteins were identified by HPLC / mass spectrometry on Q-exactive HF mass spectrometer (Thermo Scientific). Mass spectrometry was performed in Centre of Collective Usage “Human proteome» of IBMC, Moscow.

The mass spectrometric data were analyzed with the MaxQuant software (version 1.6.3.4). The Andromeda search engine was used to search the MS/MS spectra against the UniProt human database combined with 262 common contaminants and concatenated with the reversed versions of all sequences. Only proteins quantified with at least two peptides were considered for quantitation. Quantitative LC-MS-SRM analysis was performed on QQQ TSQ Vantage (Thermo Scientific) equipped with a nano-electrospray ion source. Each sample was analyzed for five times using Dionex UltiMate 3000 RSLCnano System Series (Thermo Scientific).

Statistical analysis was performed using STATISTICA version 10.0 (StatSoft, Tulsa, OK, USA). The Kolmogorov–Smirnov test was used to determine whether the data were normally distributed. Non-parametric Mann-Whitney U-test and Fisher's exact test with Yates' correction was used to check the statistical significance between-group differences. Kruskal-Wallis one-way analysis of variance (ANOVA) and median test were used for comparative data analysis between three groups. Differences and correlations were considered significant at p-value.

Results

In the work, a qualitative and quantitative analysis of the proteomes of the blood serum between patients with two different forms of schizophrenia (simple and paranoid) and healthy individuals was carried out. In patients with a paranoid form of schizophrenia, 5202 proteins were identified. Among them 1097 proteins showed significant differences with the proteins of healthy individuals. Patients with a simple form of schizophrenia identified 4098 proteins. Among them 574 proteins showed significant differences with control. Statistically significant differences in 302 proteins were obtained between patients with simple and paranoid forms of schizophrenia. Most of them from set of proteins in schizophrenia were typically associated with processes, which are responsible for protein synthesis and the processes of transduction and translation; immune response, oxidative stress, apoptosis and cell communication.

For further research by the method of quantitative mass spectrometry using labeled peptide standards, several proteins involved in neurogenesis and proteins of neuronal receptors were selected: NMDAZ1, RIPK1, mGluR6, DCLK1. Kruskal-Wallis test ANOVA by Ranks showed significant differences between three studied groups for protein DCLK1 (p =,0118) and, accordingly, significant pairwise differences between all the studied groups. In the blood serum of patients in the total group of schizophrenia, a significant increase in the number of RIPK1, mGluR6 receptors was revealed in comparison with healthy people. In addition, the amount of mGluR6 was significantly higher in patients with a simple form of schizophrenia compared with paranoid form patients (p = 0.021). In the total group of patients with schizophrenia, divided accordingly by the prevailing symptoms, a more than twofold increase in the number of RIPK1 was revealed in patients with leading negative symptoms. Thus, most of the specific proteins that we studied that has been identified in patients with schizophrenia using screening were significantly increased depending on the clinical symptoms.

Conclusions

Saia-Cereda with co-workers describes a large amount of proteins typical for brain proteome in schizophrenia, as well as for major depressive disorder (MDD) [3]. Through comparative analysis of proteome profiles between schizophrenia, BD, and MDD, only 30 proteins were similarly changed in all three disorders. Hence, small overlapping between changes in protein levels typical for major mental disorders can be a feature maintaining specificity of every disease at proteome level.

A quantitative assessment of the specific minor proteins that we studied using labeled standard peptides showed an increase in their number depending on the severity of the disease (a simple form of schizophrenia, leading negative symptoms). Thus, the identified proteins have pathogenetic and prognostic significance.

Hence, one may suppose that finally the efforts must not be concentrated on the search for a specific protein, but rather a protein set (panel) must be revealed reflecting the main pathogenetic mechanisms and serving as a starting point for diagnosis and prognosis of mental disorder development [4].

ACKNOWLEDGMENT

This work was supported by the RSF grant No.18-15-00053 "Search for peripheral markers associated with impaired myelination of the brain and the pathogenesis of the disease in schizophrenia" 2018-2020.

REFERENCES

[1] Semke A.V., Fedorenko O.Yu., Lobacheva O.A., Rakhmazova L.D., Kornetova E.G., Smirnova L.P., Mikilev F.F., Schigoreva Yu.G. Clinical, epidemiological and biological background of the adaptation of patients with schizophrenia as the basis of a personalized approach to antipsychotic therapy (2015) Siberian Bulletin of Psychiatry and Addiction. 3(88): 19-25.

[2] Loginova L.V., Smirnova L.P., Koval V.V., Fedorova O.S., Semke A.V., Ivanova S.A. Mass-spectrometric analysis of proteins of blood serum of patients with schizophrenia (2011) Bulletin of the Siberian Branch of the Russian Academy of Medical Sciences. 31(6): 63-68.

[3] Saia-Cereda V.M., Cassoli J.S., Martins-de-Souza D., Nascimento J.M. (2017) Psychiatric disorders biochemical pathways unraveled by human brain proteomics. Eur Arch Psychiatry Clin Neurosci. 26: 3-17.

[4] Loginova L.V., Smirnova L.P., Seregin A.A., Dmitrieva E.M., Mazin E.V., Simutkin G.G. To the question of the search for biomarkers in bipolar affective disorder (2014) Bulletin of the Ural Medical Academic Science. 3(49): 139-141.

DOI 10.18699/BGRS/SB-2020-324

The analysis of different longitudinal biomarkers association with the overall survival in non-small cell lung cancer by means of joint modeling

Alina Sofronova
M&S decisions LLC, Moscow, Russia
alina.sofronova@msdecisions.ru

Sergey Gavrilov
M&S decisions LLC, Moscow, Russia
sergey.gavrilov@msdecisions.ru

Oleg Stepanov
M&S decisions LLC, Moscow, Russia
oleg.stepanov@msdecisions.ru

Kirill Peskov
M&S decisions LLC, Moscow, Russia
Computational Oncology Group, I.M. Sechenov First Moscow State Medical University, Moscow, Russia
kirill.peskov@msdecisions.ru

Kirill Zhudenkov
M&S decisions LLC, Moscow, Russia
kirill.zhudenkov@msdecisions.ru

***Abstract* — It is well-known that tumor size is predictive of overall survival for patients with non-small cell lung cancer (NSCLC). Joint modeling is an advanced approach to quantify the association between longitudinal biomarkers and overall survival. The obtained results suggest that assessment of biomarker dynamics improved the accuracy of survival prediction in comparison with consideration of only baseline biomarker values for investigated patients with NSCLC.**

Keywords* — *non-small cell lung cancer, tumor size, survival analysis, joint modeling

Motivation and aim

In non-small cell lung cancer (NSCLC) studies tumor size may be represented by Sum of Longest Diameters of target lesions (SLD). SLD is known as important biomarker for overall survival (OS) prediction [1,2]. Typically, different other baseline and longitudinal biomarkers (representing oxidative stress, blood status, kindey and liver function, etc.) may be available from NSCLC study data. Multiple longitudinal and baseline biomarkers and their association to OS can be analyzed by means of joint modeling (JM) [3]. Current work is aimed to quantify the association of different longitudinal biomarkers and OS in patients with NSCLC and compare the results between two different NSCLC studies with different treatments.

Methods

Data from the comparator arms of two phase 3 clinical studies in second-line metastatic NSCLC were obtained from Project Data Sphere repository, comprising 512 patients treated with target therapy erlotinib (NCT00364351) and 596 patients treated with standard chemotherapy (docetaxel) from ZODIAC study (NCT00312377). Patients in both considered studies were similar in their disease status, demographics and baseline characteristics. Particularly, they had similar baseline values of most biomarkers, in spite of different types of treatment. Cox proportional hazards models of baseline SLD in combination with other baseline biomarkers (lactase dehydrogenase (LDH), neutrophils, white blood cells (WBC), neutrophil-to-lymphocyte ratio (NLR), alkaline phosphatase, aspartate aminotransferase, alanine aminotransferase, creatinine) were tested first. The final Cox model for each dataset was selected based on the lowest Bayesian Information Criterion value and statistical significance of considered covariates. On the basis of the optimal Cox model, multiple joint models were developed using one longitudinal biomarker and other biomarkers as baseline. Joint model parameter estimation was made in JM package in R. Longitudinal biomarker trajectories were modeled with natural splines. We assessed and compared the quality of survival predictions of Cox and JM models by means of Receiver Operating Characteristic Curve AUC (ROC AUC) and Brier Score (BS) [3]. In the case of JM, ROC AUCs and BSs were calculated with 90-day cut-off of longitudinal data.

Results

The optimal Cox model was identified for each study. Cox model structures turned out to be identical for both datasets and included SLD, LDH, WBC and NLR baseline biomarkers. The significant association between each of these identified biomarkers and OS was detected in JM as well, where the biomarkers were considered longitudinal. Additionally, the association between longitudinal SLD, LDH, WBC and the risk of death turned out to be close for both studies with different NSCLC treatments. JM outperformed Cox models in ROC AUC and BS analysis. Accumulation of longitudinal data up to 90 days improved accuracy of survival prediction independently of which biomarker was considered longitudinal. This was confirmed in analysis of both study data. The highest prediction accuracy was received for JM models with assessment of SLD dynamics and baseline values of LDH and WBC.

Conclusions

We identified the statistically significant association of different biomarkers with OS in selected NSCLC data. JM allowed to consider the longitudinal biomarker trends in addition to the baseline values only, that are used in Cox models. Longitudinal biomarkers trajectories assessment allowed to achieve higher patient survival discrimination in JM, compared to Cox models.

REFERENCES

[1] Gavrilov S et al. PAGE Meeting, 2019.

[2] Gavrilov S et al. Longitudinal tumor size and NLR are more predictive of individual survival than their baseline values in patients with non-small cell lung cancer treated with durvalumab. 2019 ASCO Annual Meeting.

[3] Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data. CRC press, 2012, ISBN 9781439872864.

DOI 10.18699/BGRS/SB-2020-325

Novel tools and methods for detecting pathogenic RNA on SARS-Cov-2 model

Grigory Stepanov
ICBFM SB RAS, Novosibirsk, Russia
stepanovga@niboch.nsc.ru

Evgenii Zhuravlev
ICBFM SB RAS, Novosibirsk, Russia
evgenijur@gmail.com

Daria Novopashina
ICBFM SB RAS, Novosibirsk, Russia
danov@niboch.nsc.ru

Vladimir Richter
ICBFM SB RAS, Novosibirsk, Russia
richter@niboch.nsc.ru

Denis Antropov
ICBFM SB RAS, Novosibirsk, Russia
des_ant_nik95@mail.ru

Georgiy Shevelev
ICBFM SB RAS, Novosibirsk, Russia
metatezis@gmail.com

Leonid Kurbatov
IBMC RAS, Moscow, Russia
leonid15@mail.ru

Dmitrii Pyshnyi
ICBFM SB RAS, Novosibirsk, Russia
pyshnyi@niboch.nsc.ru

Igor Oscorbin
ICBFM SB RAS, Novosibirsk, Russia
osc.igor@gmail.com

Maxim Kupryushkin
ICBFM SB RAS, Novosibirsk, Russia
kuprummax@gmail.com

Maksim Filipenko
ICBFM SB RAS, Novosibirsk, Russia
max@niboch.nsc.ru

***Abstract* — The current SARS-Cov-2 pandemic has shown that effective surveillance and the speed of research of a pathogen directly depend on the quality of instruments used for its detection in laboratories. The results obtained suggest the development of the series of instruments allowing to rapidly detect the viral pathogene RNAs.**

Keywords* — *SARS-CoV-2, nucleic acid detection system, NASBA, LAMP, Cas13a-interference

Motivation and aim

Up to date, RT-PCR detection systems are predominantly used for different nucleic acid detection applications. However, such systems require specialized technical equipment and are quite time-consumable. The approaches suggested made it possible to quickly and cheaply amplify pathogenic RNA molecules. The fluorescent imaging techniques used in our studies took a step for the creation of the rapid and convenient visual readout.

Methods

We utilized some modifications of commonly known methods, such as RT-PCR using novel modified phosphoryl guanidine oligonucleotides [1], and naked eye-detection system based on the loop-mediated isothermal amplification (LAMP) [2]. We also used the combination of NASBA (Nucleic Acid Sequence-Based Amplification) for specific amplification of the required RNA and Cas13a-mediated cleavage [3] for its detection via triggering of the collateral hydrolysis of RNA-detector, representing a short (15-nucleotide length) fluorescently-labelled RNA.

Results

To improve the routine real-time PCR-systems, we proposed the use of novel modified phosphoryl guanidine oligonucleotides, which showed an improvement in the quality of PCR results and a decrease in the accumulation of non-specific products (PCT/RU2018/050115). In particular, we observed an increase in the sensitivity level and an increase in the RFU level in Taq-man assay, which is most critical at the lowest possible concentrations of the SARS-Cov-2 RNA.

In this work, enzymatically assembled DNA fragments of SARS-Cov-2 genome were used as templates for synthesis of RNA transcripts as an internal control for PCR test system. A set of oligonucleotides for synthesis of DNA templates using Polymerase Cycling Assembly was selected using Genecut alghorithm (ICBFM SB RAS - Unipro Ugene, Novosibirsk) and synthesized using protocols optimized in the context of error minimization. The approach used made it possible to quickly and cheaply obtain templates for the RNA-control synthesis in vitro. We also developed the CRISPR/Cas13a-based SARS-CoV-2 detection system. We used the combination of NASBA (Nucleic Acid Sequence-Based Amplification) and Cas13a-mediated cleavage allowing to specifically detect the SARS-CoV-2 RNA at the concentrations as low as several copies per reaction. NASBA isothermal amplification increases the target RNA amount by 10^6-10^8-fold, producing the detectable specific RNA concentration for the subsequent applications. The detection mentioned is based on the collateral cleavage activity of (LwaCas13a:crRNA) complex. The signal production is triggered by the fluorescently-labelled RNA-detector cleavage, taking place upon the target RNA recognition and cleavage. The fluorescent imaging of our combined reactions took a step for the creation of the rapid and convenient visual readout. In Laboratory of pharmacogenomics prototype test-system for detecting SARS-CoV2 coronavirus genetic material was developed. The total analysis time starting with a biological sample don't exceed one hour, the results can be detected with the naked eye. The protocol is based on loop-mediated isothermal amplification (LAMP) which, unlike conventional PCR, does not require the use of expensive PCR devices and can be used outside of laboratories. In the prototype test system, a chimeric DNA polymerase previously constructed at the Laboratory of pharmacogenomics, which is more resistant to amplification inhibitors compared to commercial analogues, was used. Thus, the pandemic stimulated the development of a technique for detecting pathogenic nucleic acids, which should soon come into use and become new tools for diagnosing viral infections.

Acknowledgment

Research using phosphoryl guanidine oligonucleotides was supported by the Russian Science Foundation grant (18-14-00357). Studies on Cas13a-based methods was partially supported by State Budget Program (0245-2019-0001).

References

[1] Kupryushkin M.S., Pyshnyi D.V., Stetsenko D.A. (2014). Acta Naturae, 6(4):116118.
[2] Notomi, T. et al. (2000). Nucleic acids research, 28(12), e63.
[3] Abudayyeh, O., Gootenberg, J., Essletzbichler, P. et al. (2017). Nature 550, 280–284.

DOI 10.18699/BGRS/SB-2020-326

The visualization of fluid content during intestinal MRI

Olga A.Subbotina[1], Andrey Yu. Letyagin[1, 2], Mariya V. Rezakova[1]
[1] Department of MRI, State Scientific-Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
o.a.subbotina@yandex.ru
[2] Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
letygin-andrey@yandex.ru

***Abstract* — during a routine abdominal MRI, detailed visualization of the intestine is not possible. Using the MRI technique with peroral hydrocontrast, it becomes possible to analyze the state of the intestine along its entire length due to its liquid content, biphasic contrast and obtaining a balanced MR signal from the intestinal wall with a high T1 / T2 ratio.**

Keywords* — *intestinal MRI, Hydro-MRI, MR-enterography, peroral contrast

INTRODUCTION

In 2016, clinical recommendations were issued of the Russian Gastroenterological Association and the Association of Russian Coloproctologists for the diagnosis and treatment of Crohn's disease, where MR enterography is one of the methods for diagnosing and monitoring the effectiveness of treatment, as well as with suspected presence of stenosis and fistulas. Of course, the diagnostic capabilities of hydro-MRI are much wider and capture a greater number of nosologies, including not only other types of inflammatory bowel diseases, but also a number of infectious diarrhea, tumor pathology. Since the advent of MRI, visualization of the gastrointestinal tract was considered beyond the boundaries of this method, but hydrocontrasting of the intestinal lumen with special peroral contrasts made it possible to visualize and analyze it. The determination of fluid content in MRI required the use of modified signal sequences, which became the basis of the new method.

MATERIALS AND METHODS

When receiving hydrocontrast, different types of peroral contrasts are used: negative, positive and biphasic. Each of them has its own advantages and disadvantages. Negative contrasts have a hypo-intense signal both on the T1-weighted image (T1-WI) and on the T2-weighted image (T2-WI), so the intestinal lumen is always dark. This is an indisputable advantage in assessing the intestinal wall with contrast enhancement and the contrasting pattern (white on dark), but leads to an inadequate assessment of the thickness of the intestinal wall, especially in the absence of its edematous changes (dark on dark). Positive contrasts, on the contrary, have an increased signal at T1-WI and T2-WI. This allows you to perfectly visualize the intestinal wall (dark on white), but makes it difficult to analyze contrast images (white on white). Biphasic contrast, having a low signal at T1-WI and a high signal at T2-WI, is devoid of the previously described disadvantages. An example of biphasic contrast is water, but due to the rapid absorption, it does not give good fullness of the intestinal lumen and is therefore not used. The most widely used biphasic contrast, which we use, is a 2.0–2.5% solution of ®Mannitol, which gives a long and adequate fullness of the intestine for the duration of the scan. In order to qualitatively visualize liquid content, in addition to using the T1-WI and T2-WI methods, additional signal sequences are used. To assess the state of the intestine, we use the pulse sequence FIESTA (Cor, Ax, Sag) - a coherent technique that uses a full balanced gradient signal (Figure 1). The image contrast is ensured by a high T1 / T2 ratio. That is, substances or tissues that have the maximum difference in signal intensity at T1-WI and T2-WI will be most contrasted with this technique. Accordingly, peroral biphasic contrast enhancement will provide excellent visualization of the intestinal lumen. Images are more T2-weighted. The high scanning speed and very low sensitivity to movement when acquiring images make this technique reliable in patients who have difficulty holding their breath.

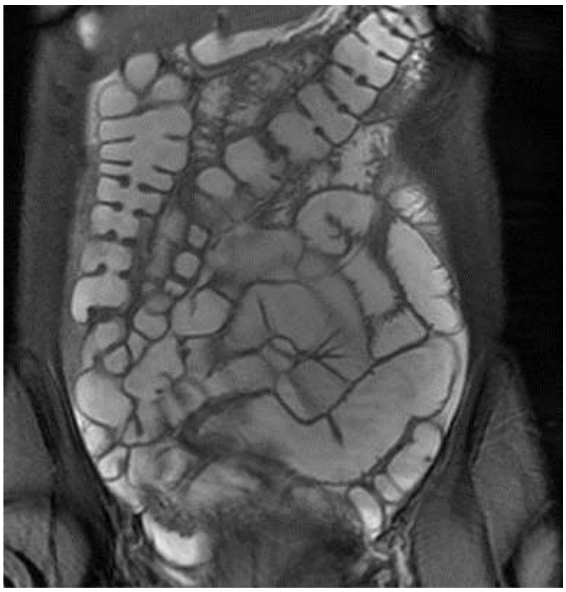


Fig. 1. MR-image of the intestine by FIESTA

Direct methods for determining gastrointestinal activity are electro-gastrography, manometry and the use of electrodes and calibration voltage transducers. The invasiveness of these methods does not allow their frequent use for diagnostic purposes. Conducting dynamic FIESTA provides a unique opportunity to visualize intestinal motility not just in vivo, but also in real time, to evaluate the movement of fluid content in the intestine, to analyze the activity of the ileocecal valve (complete or incomplete closure of its valve flaps, the presence of retrograde current at the level of the small inteFstine transition), slowing down and difficulty in moving the liquid contents in the area of narrowing of the lumen to assess the degree of compensation of stenosis. There is no unified approach to the interpretation of functional studies of the intestine, and the analysis is only quality, but even at this stage of the development of functional techniques, the information received has high diagnostic value.

Since MRCP (HASTE) is based on the turbo-spin echo technique, which is used to sequentially obtain high-resolution T2-WI and allows to obtain a 3D model of the intestine (Figure 2). This allows you to easily select the exact position of the plane to quickly obtain the desired viewing angle and position in space. HASTE uses the technique of short single signals to obtain enough data for the entire image in one sequence repetition period (TR). HASTE imaging uses half-Fourier in phase technology. Only half of the K-space is collected and the remainder is displayed. This technique is essentially an image of exclusively liquid content in the study area - intestinal lumen filled with peroral contrast. Due to the possibility of building a 3D-model, this technique is primarily used to analyze the anatomical pathway of the intestine, verify congenital anomalies, dystopia, dolichocolon and dolichosigma, but also to determine the places of narrowing and expansion of the intestinal lumen. 3D tomography is a modern diagnostic standard, which allowed the quality of diagnosis and treatment to reach a whole new level. The advantage of three-dimensional images is the ability to identify defects that are often overlooked in conventional images.

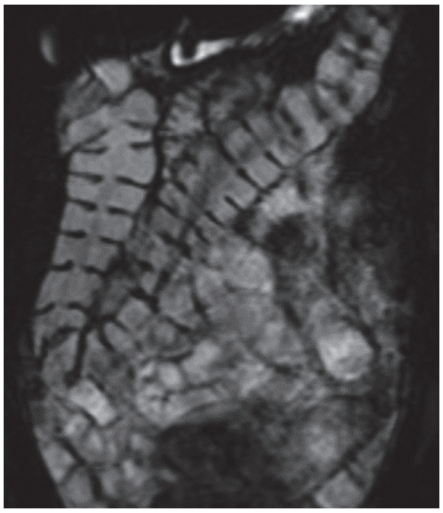


Fig. 2. MR-image of the intestine by MRCP

DISCUSSION

To obtain MR images of the intestine suitable for analysis and interpretation, it is necessary to fill it with special oral liquid contrasts. Using this technique allows you to visualize not only the small, but also the large intestine throughout its entire length. But because of the high diagnostic significance of the information received on the condition of the small intestine, which is less accessible for diagnosis, this technique is often called MR enterography, although it is more correctly called hydro-MRI due to the use of peroral hydrocontrast. The presence of liquid content allows you to straighten the intestinal wall, improve visualization of its course, wall thickness, the location of loops and folds. Thus, when conducting MR enterography, it is the visualization of liquid content that makes it possible to conduct a qualitative analysis of the state of the intestine and visualize its internal structure. In addition, modern MR technologies allow real-time dynamic studies of the intestine and hydrocontrasting makes it possible to judge the presence of functional changes by changing the rate of advancement of the liquid contents. Based on the data of intestinal MRI, it is possible to create a personalized visualization model of the gastrointestinal tract

REFERENCES

[1] Westerland O, Griffin N. Magnetic Resonance Enterography in Crohn ʻs Disease. Semin

Ultrasound CT MR 2016; 37(4):282-2915.

[2] Schukina O.B., Sobko V.Yu. Fecal calprotectin and hydro-MRI in assessing Crohn's disease activity. Bulletin of the North-West State Medical University. I.I. Mechnikov 2013; 5 (1): 78-83.

[3] Trofimova T.N., Dementieva T.V., Karpenko A.K. Possibilities of modern methods of radiation diagnostics in visualization of the colon in the norm and with its pathological changes: CT-colonography // Radiation diagnostics and therapy.— 2010.— No. 2 .- S. 64–68.

[4] Draft clinical guidelines for the diagnosis and treatment of inflammatory bowel disease. Coloproctology 2013; 3 (45): 40

DOI 10.18699/BGRS/SB-2020-327

The *rs12255372 and rs7903146* polymorphisms of the *TCF7L2* gene among Buryats and Russians of Eastern Siberia

Ludmila Tabikhanova,
Novosibirsk State University
Novosibirsk, Russia
tabikhan@bionet.nsc.ru

Ludmila Osipova
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia,
Novosibirsk State University
Novosibirsk, Russia
ludos77@yandex.ru

Tatiana Churkina
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia,
Novosibirsk State University
Novosibirsk, Russia
tan646464@gmail.com

Daria Lichman
Institute of Cytology and Genetics,
SB RAS, Novosibirsk, Russia,
Novosibirsk State University
Novosibirsk, Russia
daria.lichman@gmail.com

Elena Voronina
Institute of Chemical Biology and
Fundamental Medicine, SB RAS,
Novosibirsk, Russia,
Novosibirsk State University
Novosibirsk, Russia
voronina_l@mail.ru

Maxim Filipenko
Institute of Chemical Biology and
Fundamental Medicine, SB RAS,
Novosibirsk, Russia,
Novosibirsk State University
Novosibirsk, Russia
max@niboch.nsc.ru

Abstract — **The objective of the present study was to investigate the polymorphism of metabolism genes in indigenous populations of Siberia. To this end, we have studied the ethnic features of allele frequency distribution for polymorphic variants in gene *TCF7L2 G103894T (rs12255372)* and *C53341T (rs7903146)* in the samples taken from Buryats, and Russians of Eastern Siberia, and compared it with data on world populations. Samples of Eastern (N = 132) and Western (N = 278) Buryats, Russians (N = 122) and persons of mixed Buryat-Russian origin (N = 56) were genotyped by real-time PCR using competitive TaqMan- probes. Compared to the Russians, the Buryats showed a statistically significantly lower incidence of the *TCF7L2 103894T* and *TCF7L2 53341T* alleles associated with metabolic diseases and Type 2 diabetes mellitus (T2DM). Buryats are intermediate between Caucasian and East Asian populations. This agrees with the lower susceptibility of Buryats to metabolic disorders compared to the Caucasian population described in the literature. Intermediate frequencies alleles in the group of mixed origin may indicate a higher risk of the associated metabolic disorders in the descendants of mixed marriages compared to the Buryats.**

Keywords — Buryats; Russians of Eastern Siberia; mixed origin; real-time PCR; metabolism; genetic polymorphism; TCF7L2 (rs12255372); TCF7L2 (rs7903146)

Motivation and Aim

For centuries indigenous peoples of Siberia have been adapting to severe climatic and geographical conditions and predominantly protein-lipid diet, and are now characterized by a distinct type of metabolism with increased protein-lipid and minimized carbohydrate exchange [1]. As the ongoing urbanization impacts the indigenous population, their living and economic conditions change, and the so-called 'civilization diseases' related to metabolic disorders increasingly strike the Siberian peoples [2]. Investigation of gene polymorphism in the metabolic profile of indigenous Siberian populations is critical for understanding the molecular-genetic foundations of adaptive potential they developed by adapting to specific climatic and geographical conditions and via certain nutritional habits, and for identifying genetic reserves of ethnic groups in a rapidly changing world [3,4]. Prior studies indicated that some of the variants of the transcription factor 7-like 2 (TCF7L2) gene such as *rs12255372* and *rs7903146* are associated with cardiovascular and other metabolic diseases and Type 2 diabetes mellitus (T2DM) in various populations and ethnic groups [6-7]. The purpose of this study was to determine the frequencies of *TCF7L2 G103894T (rs12255372)* and *TCF7L2 C53341T (rs7903146)* polymorphisms in healthy Buryats, Russians and descendants of mixed marriages (Metis).

Methods and Algorithms

Persons of Buryat nationality having no ancestors with foreign ethnic backgrounds and living in Alkhanay and Orlovsky settlements of the Agin-Buryat Autonomous District in Zabaykalsky Krai were included in the Eastern Buryat group (N=132). Ethnic Buryats living in the settlements of the Ekhirit-Bulagatsky District of the Ust-Ordyn Buryat Autonomous District in the Irkutsk Region (N=278) were included in the western sample. First- and second-generation descendants from mixed Russian-Buryat marriages were included in the metis sample (N=56). Russians from Eastern Siberia, whose ancestors lived in the settlements of Zabaykalsky Krai and the Irkutsk Region for several generations, were included in the fourth sample (N=122).

Genotyping was performed using real-time PCR with competitive TaqMan allele-specific probes. Population allele frequencies of polymorphic variants were determined based on the observed genotype frequencies. The match between the empirically observed genotype frequency distribution and the expected theoretical distribution in Hardy-Weinberg equilibrium was examined using $\chi 2$ Pearson test (the equilibrium holds at $p>0.05$). Significance of differences in allele frequencies between the studied samples was determined using $\chi 2$ test with Yates's correction for continuity; the results were considered statistically significant at $p<0,05$.

Results and discussion

Genotype distribution for polymorphic loci of the TCF7L2 G103894T (rs12255372) and TCF7L2 C53341T (rs7903146) and allele frequencies in samples of Buryats, their metis, Teleuts, and Russians from Eastern Siberia is presented in Table 1. Genotype distribution matched the Hardy-Weinberg equilibrium for both polymorphic loci.

TABLE II. GENOTYPE DISTRIBUTION FOR *TCF7L2 G103894T (RS12255372)* AND *TCF7L2 C53341T (RS7903146)* AND ALLELE FREQUENCIES IN SAMPLES OF BURYAT, THEIR METIS AND RUSSIANS FROM EASTERN SIBERIA

Population			**Eastern Buryats**	**Western Buryats**	**Metis**	**Russians**
G103894T	Genotype count	*G/G*	116	251	45	73
		G/T	15	24	8	40
		T/T	0	3	2	7
	N, ppl		131	278	55	120
	P (H-W)		0,905	0,668	0,654	0,628
	103894T frequency, %		5,7	5,4	10,9	22,5
C53341T	Genotype count	*C/C*	119	225	49	70
		C/T	13	49	7	44
		T/T	0	2	0	8
	N, ppl		132	276	56	122
	P (H-W)		0,925	0,932	0,926	0,449
	53341T frequency, %		4,9	9,6	6,3	24,6

N is the sample size, P (H-W) is the probability of deviation from the Hardy-Weinberg equilibrium, metis are the descendants of mixed Russian-Buryat marriages

The *TCF7L2 103894T* and *TCF7L2 53341T* alleles associated with metabolic diseases and T2DM [5-7] are widely spread in Caucasian populations - at 20%-40% [8]. The incidence of these alleles is much lower in East Asian, i.e. Chinese, Japanese, and Vietnamese, populations at 1%-3%.

The frequencies of the *TCF7L2 103894T* and *TCF7L2 53341T* alleles in the Russian samples corresponds to the frequency range found in European populations [8]. Frequencies of both alleles in samples of Eastern and Western Buryats were statistically significantly lower than in samples of Russians from Eastern Siberia and Caucasian groups described in the literature [8]. On the other hand, frequencies of the *TCF7L2 103894T* variant in both Buryats samples significantly increased if compared to a number of East Asian groups. Also significant differences in the incidence of the *TCF7L2 53341T* allele were found between Western Buryats and the East Asian sample. It was shown earlier for other gene frequencies that indigenous Siberian populations were intermediate between Caucasians and East Asians [9].

There are literature data showing a reduced risk of metabolic disorders in Buryats [10, 11]. Adaptation to the hypercontinental climate of the Siberian region, livestock-based economy, and high-calorie diet with high protein and fat contents appears to have affected the frequencies of polymorphic variants of metabolic genes in this indigenous Siberian ethnic group. Lower prevalence of lipid and carbohydrate metabolism disorders in Buryats matches decreased incidence of *TCF7L2 103894T* and *TCF7L2 53341T* in the population.

Conclusion

Compared to the Russians, the Buryats showed a statistically significantly lower incidence of the *TCF7L2 103894T* and *TCF7L2 53341T* alleles associated with metabolism-associated diseases and T2DM. This agrees with the lower susceptibility of Buryats to metabolic disorders compared to the Caucasian population described in the literature. Intermediate frequencies of alleles in the metis group may indicate a higher risk of the associated metabolic disorders in the descendants of mixed marriages compared to the Buryats.

ACKNOWLEDGMENT

The work was supported by the Russian Scientific Foundation, project 19-15-00219.

REFERENCES

[1] L.E. Panin, “Energy aspects of adaptation,” Leningrad,1978. (In Russian).

[2] Lyudinina A.Yu, Potolitsyna N.N., Solonin Yu.G., Osadchuk L.V., Gutorova N.V., Petrova P.G., Troev I.P., Ostobunaev V.V., Boyko E.R. Lipid profile in men of Komi Yakut ethnic groups with overweight and obesity. Jekologija cheloveka. 2014;1:13-19. (In Russian).

[3] B. Hallmark, T.M. Karafet, P.H. Hsieh, L.P. Osipova, J.C. Watkins and M.F. Hammer, “Genomic Evidence of Local Adaptation to Climate and Diet in Indigenous Siberians,” Molecular Biology and Evolution, 2018, vol. 36(2), pp. 315-327. DOI: 10.1093/molbev/msy211.

[4] P.H. Hsieh, B. Hallmark, J.C. Watkins, T.M. Karafet, L.P. Osipova, R.N. Gutenkunst and M.F. Hammer, “Exome Sequencing Provides Evidence of Polygenic Adaptation to a Fat-Rich Animal Diet in Indigenous Siberian Populations,” Mol Biol Evol., 2017, vol. 34(11), pp. 2913–2926. DOI:10.1093/molbev/msx226.

[5] D. Bodhini, V. Radha, M. Dhar, N. Narayani and V. Mohan, “The rs12255372(G/T) and rs7903146(C/T) polymorphisms of the TCF7L2 gene are associated with type 2 diabetes mellitus in Asian Indians,” Metabolism Clinical and Experimental, 2007, vol. 56, pp. 1174–1178 DOI:10.1016/j.metabol.2007.04.012

[6] S. Oktavianthi, M. R. Saraswati, K. Suastika, P. Dwipayana, A. Sulfianti, R. F. Hayati, H. Trimarsanto, C. A. Febinia, H. Sudoyo and S. G. Malik, “Transcription factor 7-like 2 single nucleotide polymorphisms are associated with lipid profile in the Balinese” Molecular Biology Reports, 2018, vol. 45, pp. 1135–1143. DOI:10.1007/s11033-018-4265-x

[7] S. Kalantaria, A. Sharafshahb, P. Keshavarzb, A. Davoudic and R. Habibipourb, “Single and multi-locus association study of TCF7L2 gene variants with susceptibility to type 2 diabetes mellitus in an Iranian population,” Gene, 2019, vol. 696, pp. 88–94. DOI:10.1016/j.gene.2019.01.040

[8] The 1000 Genomes Project Consortium, “An integrated map of genetic variation from 1,092 human genomes,” Nature, 2012, vol. 491(7422), pp. 56-65. DOI: 10.1038/nature11632.

[9] L.E. Tabikhanova, L.P. Osipova, E.N. Voronina, A.O. Bragin and M.L. Filipenko, “Polymorphism of lipid exchange genes in some populations of South and east Siberia,” Vavilovskii Zhurnal Genetiki i Selektsii, 2019, vol. 23(8), 1011-1019. (In Russian) DOI: 10.18699/VJ19.578

[10] T.A. Bairova, V.V. Dolgikh, L.I. Kolesnikova and O.A. Pervushina, “Nutriciogenetics and risk factors of cardiovascular disease: associated research in Eastern Siberia populations,” Bjulleten' VSNC SO RAMN, 2013, vol. 4(92), pp. 87-92. (In Russian).

[11] N.V. Semenova, I.M. Madaeva, M.A. Darenskaya, O.A. Gavrilova, R.M. Zhambalova and L.I. Kolesnikova, “Lipid profile in menopausal women of two ethnic groups,” Acta Biomedica Scientifica, 2018, vol. 3(3), pp. 93-98. (In Russian). DOI: 10.29413/ABS.2018-3.3.14.

DOI 10.18699/BGRS/SB-2020-328

The lithium effects on morphology and apoptosis in hepatocellular carcinoma cells

Iuliia Taskaeva
Laboratory of ultrastructural research, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences;
Laboratory of boron-neutron capture therapy, Department of Physics, Novosibirsk State University
Novosibirsk, Russia
inabrite@yandex.ru

Izabella Gogaeva
Laboratory of ultrastructural research, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences;
Department of Natural Sciences, Novosibirsk State University
Novosibirsk, Russia
i.gogaeva@g.nsu.ru

Natalia Obanina
Laboratory of ultrastructural research, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences;
Section of cytology and genetics, Department of Natural Sciences, Novosibirsk State University
Novosibirsk, Russia
n.obanina@g.nsu.ru

Viktoriia Makarova
Laboratory of ultrastructural research, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
shedina_vika@mail.ru

Nataliya Bgatova
Laboratory of ultrastructural research, Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia
n_bgatova@ngs.ru

***Abstract* — Hepatocellular carcinoma (HCC) is characterized dysregulation of cell death mechanisms, and the imbalance of pro- and anti-apoptotic signals. The development of HCC is accompanied by genetic mutations in the signaling pathways involved in the cell proliferation, growth and death. The molecular changes in apoptosis signaling in HCC determine the requirement for targeted chemotherapy to increasing apoptosis in HCC cells. The aim of this study was to assess the ability of lithium to influence on the hepatocellular carcinoma-29 (HCC-29) cells apoptosis in vivo. Light and transmission electron microscopy, and immunofluorescence staining were used to evaluate of apoptosis development in HCC-29 cells after administration of 20 mM lithium carbonate. It was revealed that lithium extremely increased the pro-apoptotic proteins Bad and caspase-3 expression, and decreased the anti-apoptotic protein Bcl-2 expression. These results indicate that lithium carbonate induces apoptosis pathways in HCC-29 cells. Lithium administration can enhance pro-apoptotic chemotherapeutic drugs potential and overcome the resistance of tumor cells to apoptosis in HCC.**

Keywords* — *hepatocellular carcinoma, lithium, morphology, apoptosis, electron microscopy

Introduction

Hepatocellular carcinoma (HCC) is a highly malignant type of liver cancer characterized by the presence of multiple genetic mutations in the Wnt/beta-catenin, p53, PI3K/Ras signaling pathways, as well as in the processes of oxidative and endoplasmic reticulum stress [1]. These pathways are the main intracellular signaling cascades involved in the proliferation, growth and death of tumor cells. Furthermore, HCC is characterized by molecular changes that affect on apoptosis processes: mutations in the P53 tumor suppressor gene (TP53), increased expression of transforming growth factor (TGF-β) and the anti-apoptotic phenotype, resistance to extrinsic apoptosis [2]. The dysregulation of cell death and proliferation mechanisms, as well as the imbalance of pro- and anti-apoptotic signals in HCC require the targeted chemotherapy to enhancing HCC apoptosis [3]. The effects of lithium on carcinogenesis and tumor progression are currently being actively studied [4]. Lithium implements its effect mainly by inhibiting glycogen synthase kinase 3β (GSK-3β), an enzyme that regulates the proliferation, differentiation and apoptosis of tumor cells [5]. We have previously shown that lithium salts decrease viability and stimulate apoptosis in hepatocellular carcinoma-29 cells in vitro [6, 7]. The aim of this study was to investigate the lithium effects on cell morphology and HCC-29 apoptosis in vivo.

Methods and algorithms

HCC-29 was obtained by the researchers of the Institute of Cytology and Genetics, and was kindly provided for our study [8]. Mice were maintained at a constant room temperature (23 °C) with a natural day/night light cycle in a conventional animal colony. Standard laboratory food and water were provided. Male CBA mice of 6–8 weeks of age, with weights of 18–20 g were used in experiment. For tumor induction, $1 \cdot 10^6$ HCC-29 cells were transplanted into the abdominal cavity, after 10 days, ascitic fluid was removed and $2 \cdot 10^6$ HCC-29 cells suspended in 100 μL of PBS were injected into the right thigh muscle. Mice were randomly divided into two experimental groups (five mice in each). Lithium (Li_2CO_3) was delivered in 100 μL volumes given intramuscularly along periphery of the tumor every day. Experimental groups were as follows: the mice with intact tumor (Control) and mice, received 20 mM lithium carbonate (LC). All mice were euthanized by cervical dislocation under overdose of anesthesia on the 23rd day of the experiment. Animal testing was performed in accordance with Directive 2010/63/EU. *Transmission electron microscopy (TEM).* Tumor tissue was fixed with a 4% paraformaldehyde and then incubated with 1% osmium tetroxide (OsO4) at 4 °C for 1h. The tumor tissue

was then incubated with 1% uranyl acetate after which it were further processed at the JEM 1400 electron microscope (JEOL, Japan). Microscopic analysis was carried out at the Multiple-access Center for Microscopy of Biological Subjects (Institute of Cytology and Genetics, Novosibirsk, Russia). *Immunofluorescent staining (IF-F).* Frozen sections of tumor tissue (10 μm) were analyzed by immunofluorescence staining using anti-Bad (ab32445, Abcam, UK), anti-Bcl-2 (ab692, Abcam, UK) and anti-caspase-3 (ab13847, Abcam, UK) primary antibodies. Alexa Fluor 488-conjugated anti-rabbit IgG (ab150077, Abcam, UK) and Alexa Fluor 594-conjugated anti-mouse IgG (ab150108, Abcam, UK) were used to detect the corresponding primary antibodies. Images were analyzed using an Axio Observer Z1 (Zeiss, Germany) fluorescence microscope at ×400 final magnification. Fourteen fields per group were captured (total area was 0.08 mm^2 for each group). The number of Bad-, Bcl-2- and caspase-3 positive cells was counted and analyzed using ImageJ software (National Institutes of Health, Bethesda, MD). *Statistical analysis.* Data are presented as mean (M) ± standard deviation (SD). Mann-Whitney nonparametric tests were used to assess differences by statistical package Statistica 6.0 (StatSoft, USA). Statistically significant differences were considered at $P < 0.05$.

Results

An ultrastructural study of tumor tissue revealed apoptotic changes in HCC-29 cells after administration of 20 mM lithium carbonate. In cancer cells fragmentation of the nucleus and condensation of chromatin, a decrease in cell volume and cytopodia, as well as plasma membrane bleeding were detected. These changes are typical morphological manifestations of apoptosis. After immunofluorescence staining of tumor tissue cryosections, it was found that the introduction of lithium significantly increased the number of Bad-positive and caspase-3-positive cells (Table 1). In addition, the number of Bcl-2 positive cells decreased significantly after lithium administration. On hematoxylin-stained paraffin sections, necrosis zones were calculated for each group. There were no significant differences between the groups: in the control group, the percentage of such zones was 5.41 ± 5.29, and after lithium administration – 5.44 ± 3.3%. Apoptosis is a process of regulated cell death that develops as a result of intracellular changes (intrinsic pathway of apoptosis) or changes in the microenvironment of the cell (extrinsic pathway of apoptosis) [9]. The intrinsic pathway of apoptosis is initiated by disturbance of intracellular homeostasis, leading to permeabilization of the outer mitochondrial membrane, which occurs under the control of the regulating apoptosis Bcl-2 proteins family [10]. The extrinsic apoptosis pathway is triggered mainly by death receptors, which results in the formation of a death-inducing signaling complex (DISC) and activation of caspase-8, -10, -3 and others [10]. The results obtained in this study demonstrate the possible ability of lithium to influence on the mitochondrial membranes permeability and participate in the intrinsic and probably extrinsic apoptotic pathways.

On the various HCC cell lines, lithium has been shown to induce apoptosis [11]. An increase in caspase-3, -8, and p53 levels and TRAIL-induced apoptosis in HCC cells has also been identified [12]. Thus, the results obtained are consistent with the literature data on the lithium effects on HCC cells apoptosis.

TABLE I.

Group	Apoptotic Markers		
	Bad	***Caspase-3***	***Bcl-2***
Control	29.57±10.44	4.71±3.41	38.29±21.42**
Lithium carbonate	60.93±45.9**	21.57±17.89*	16.79±8.65

*P < 0.05; **P < 0.005

Conclusion

GSK-3β is the main intracellular target of lithium involved in a multiple signaling pathways associated with tumor cell growth and proliferation, cell death and survival (PI3K/Akt/mTORC1, Ras/Raf/MEK/ERK, Wnt/beta-catenin, Hedgehog and others). Most likely, lithium affects the development of apoptosis due to inhibition of GSK-3β, nevertheless, further studies of the lithium-related apoptotic changes in HCC cells are required. The molecular features of HCC that contribute to the apoptosis resistance determine the relevance of further research and the development of new strategies for enhancing apoptosis in HCC cells. Lithium administration can enhance pro-apoptotic chemotherapeutic drugs potential and overcome the resistance of tumor cells to apoptosis in HCC.

Acknowledgment

This work was supported with financing of the Novosibirsk Research Institute of Clinical and Experimental Lymphology as part of a state order, No. 0324-2019-0045.

References

[1] W. K. Sung et al., "Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma," Nat. Genet., vol. 44, pp. 765–769, 2012.
[2] I. Fabregat, "Dysregulation of apoptosis in hepatocellular carcinoma cells," World J. Gastroenterol., vol. 15, pp. 513–520, February 2009.
[3] J. Moreno-Càceres and I. Fabregat, "Apoptosis in liver carcinogenesis and chemotherapy," Hepat. Oncol., vol. 2, pp. 381–397, October 2015.
[4] I. S. Taskaeva and N. P. Bgatova, "Lithium salts in experimental oncology: review," Siberian Scientific Medical Journal, vol. 39, pp. 12–18, May 2019.
[5] J. A. Quiroz, T. D. Gould and H. K. Manji, "Molecular effects of lithium," Mol. Interv., vol. 4, pp. 259–272, October 2004.
[6] Y. S. Gavrilova, N. P. Bgatova, A. O. Solov'eva, K. E. Trifonova, A. P. Lykov, Y. I. Borodin and V. I. Konenkov, "Target cells for lithium in different forms within a heterogeneous hepatocarcinoma-29 population," Cell Tiss. Biol., vol. 10, pp. 284–289, April 2016.
[7] N. P. Bgatova, Yu. S. Gavrilova, A. P. Lykov, A. O. Solovieva, V. V. Makarova, Yu. I. Borodin and V. I. Konenkov, "Apoptosis and autophagy in hepatocarcinoma cells induced by different forms of lithium salts," Cell Tiss. Biol., vol. 11, pp. 261–267, April 2017.
[8] V. I. Kaledin, N. A. Zhukova, V. P. Nikolin, N. A. Popova, M. D. Belyaev, N. V. Baginskaya, E. A. Litvinova, T. G. Tolstikova, E. L. Lushnikova and D. E. Semenov, "Hepatocellular carcinoma-29 – metastatic transplantable tumor of mice, causing cachexia," Bull. Exp. Biol. Med., vol. 148, pp. 664–669, December 2009.
[9] L. Galluzzi, et al., "Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018," Cell Death Differ., vol. 25, pp. 486–541, March 2018.
[10] L. Galluzzi, O. Kepp and G. Kroemer, "Mitochondrial regulation of cell death: a phylogenetically conserved control," Microb Cell, vol. 3, pp. 101–108, February 2016.
[11] E. Erdal, N. Ozturk, T. Cagatay, E. Eksioglu-Demiralp and M. Ozturk, "Lithium-mediated downregulation of PKB/Akt and cyclin E with growth inhibition in hepatocellular carcinoma cells," Int. J. Cancer, vol. 115, pp. 903–910, July 2005.
[12] E. Beurel, M. J. Blivet-Van Eggelpoël, M. Kornprobst, S. Moritz, R. Delelo, F. Paye, C. Housset and C. Desbois-Mouthon, "Glycogen synthase kinase-3 inhibitors augment TRAIL-induced apoptotic death in human hepatoma cells," Biochem. Pharmacol., vol. 77, pp. 54–65, January 2009.

DOI 10.18699/BGRS/SB-2020-329

miRNA expression profile in abnormally invasive placenta: accreta, increta and percreta cases

Angelika Timofeeva
FSBI «National Medical Research Center For Obstetrics, Gynecology And Perinatology Named After Academician V.I.Kulakov» Ministry of Healthcare of the Russian Federation, Moscow, Russia
avtimofeeva28@gmail.com

Oksana Vasilchenko
FSBI «National Medical Research Center For Obstetrics, Gynecology And Perinatology Named After Academician V.I.Kulakov» Ministry of Healthcare of the Russian Federation, Moscow, Russia
vasilchenko-on@mail.ru

Mariya Pirogova
FSBI «National Medical Research Center For Obstetrics, Gynecology And Perinatology Named After Academician V.I.Kulakov» Ministry of Healthcare of the Russian Federation, Moscow, Russia
pirogovamariya@gmail.com

Ivan Fedorov
FSBI «National Medical Research Center For Obstetrics, Gynecology And Perinatology Named After Academician V.I.Kulakov» Ministry of Healthcare of the Russian Federation, Moscow, Russia
i_fedorov@oparina4.ru

Vitaliy Chagovets
FSBI «National Medical Research Center For Obstetrics, Gynecology And Perinatology Named After Academician V.I.Kulakov» Ministry of Healthcare of the Russian Federation, Moscow, Russia
vvchagovets@gmail.com

Larisa Ezhova
FSBI «National Medical Research Center For Obstetrics, Gynecology And Perinatology Named After Academician V.I.Kulakov» Ministry of Healthcare of the Russian Federation, Moscow, Russia
l_ezhova@oparina4.ru

Roman Shmakov
FSBI «National Medical Research Center For Obstetrics, Gynecology And Perinatology Named After Academician V.I.Kulakov» Ministry of Healthcare of the Russian Federation, Moscow, Russia
r_shmakov@oparina4.ru

Gennadiy Sukhikh
FSBI «National Medical Research Center For Obstetrics, Gynecology And Perinatology Named After Academician V.I.Kulakov» Ministry of Healthcare of the Russian Federation, Moscow, Russia
g_sukhikh@oparina4.ru

***Abstract* — It is well-known that abnormal placental invasion is a condition characterized by pathological invasion of chorionic villi into the the basal layer of the uterus that is complicated by massive obstetric hemorrhage, blood transfusions, a need for hysterectomy and even mortality. The altered miRNA expression profiles were found by deep sequencing not only directly in the area of pathological invasion, but also in the outside distant placenta areas with more pronounced changes in the former case. An increase in the level of miR-21-5p, miR-25-3p, miR-92a-3p, miR-320a in maternal plasma may indicate abnormally invasive placenta and potentially be used in the early diagnosis and management of this condition.**

Keywords* — *deep sequencing, RT-PCR, miRNA, placenta presentation, accreta, increta, percreta

Motivation and aim

Abnormal placental invasion is a condition characterized by pathological invasion of chorionic villi into the the basal layer of the uterus and as a result, it is complicated by massive obstetric hemorrhage, blood transfusions, a need for hysterectomy and even mortality [1]. Abnormal uterine scars after previous cesarean section especially in combination with placenta previa are thought to serve as a basis for defective decidualization and deep trophoblast invasion [2]. However, the molecular-biological profile of the placenta itself has not yet been studied in this pathology.

The aim of this study was to analyze and compare the miRNA expression profiles directly in the area of abnormal placental invasion (P-area) as well as in placenta areas outside this region (N-area) in pregnant women with placenta accreta, increta and percreta in comparison with that in placenta presentation. Identified placental miRNAs were analyzed in maternal plasma as possible indicators of abnormally invasive placenta for further use as diagnostic and predictive tests.

Methods

69 pregnant women were taken into the study. Samples of a placenta tissue taken immediately after delivery were homogenized in a QIAzol Lysis Reagent, and the total RNA was isolated by an miRNeasy MicroKit and RNeasy MinElute Cleanup Kit (Qiagen, Hilden, Germany). miRNA from maternal blood plasma was isolated by an miRNeasy Serum/Plasma kit (Qiagen). Evaluation of the miRNA expression profile in placental tissue was performed using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (Set11 and Set2, New England Biolab®, Frankfurt am Main, Germany) for cDNA library synthesis and the NextSeq 500 platform (Illumina, San Diego, CA, USA) for sequencing. Differential expression analysis of the miRNA count data was performed with the DESeq2 package. The validation of deep sequencing data was carried out by reverse transcription coupled with polymerase chain reaction (RT-PCR) in real-time in a StepOnePlus™ thermocycler (Applied Biosystems, Foster City, CA, USA).

Results

The altered miRNA expression profiles were detected not only directly in the area of pathological invasion (P-area), but also in the outside distant placenta areas (N-area). In the N-sites, statistically significant changes in the expression level more than twice were found for 73 miRNAs when comparing the «increta group» with the «placenta presentation group» (for example, a 23-fold increase in the expression of miR-25-3p, p=0.0002; a 16.9-fold increase in the expression of miR-92a-3p, p=0.0003); for 67 miRNAs, when comparing the «percreta group» with the «placenta presentation group» (for example, a 15-fold increase in miR-25-3p expression, p=7.8e-5; a 16-fold increase in miR-92a-3p expression, p=0.0017). In the P-sites, statistically significant changes in the expression level more than twice were found for 84 miRNAs when comparing the «increta group» with the «placenta presentation group» (for example, a 19-fold increase in the expression of miR-25-3p, p=0.0002; a 15-fold increase in the expression of miR-92a-3p, p=0.0004; a 2-fold increase in the expression of miR-320a, p=0.02); for 269 miRNAs, when comparing the «percreta group» with the «placenta presentation group» (for example, a 43-fold increase in miR-25-3p expression, p=5.7e-40; a 37-fold increase in miR-92a-3p expression, p=1.7e-44; a 3.5-fold increase in the expression of miR-320a, p=1.78e-15; a 2-fold increase in the expression of miR-21-5p, p=0.029). miRNAs specific for pathologic invasion of the placenta were validated in blood plasma by the quantitative RT-PCR on an independent cohort of pregnant women. There were 8.8- and 10-fold increases in miR-21-5p level (p=0.0003 and p=0.0029, respectively), 3- and 5-fold increases in miR-25-3p level (p=0.05 and p=0.04, respectively), 1.35 - and 1.9-fold increases in miR-92a-3p level (p=0.033 and p=0.001, respectively), 1.9- and 2.2-fold increases in miR-320a (p=0.005 and p=0.001, respectively) in the «increta group» and «accreta group», respectively, relative to the «placenta previa group» without an uterine scar. An increase in the level of miR-21-5p, miR-25-3p, miR-92a-3p, miR-320a in maternal plasma may indicate abnormally invasive placenta.

Acknowledgment

Supported by the State assignment of the Ministry of Healthcare of the Russian Federation.

References

[1] L. Say et al. Global causes of maternal death: A WHO systematic analysis. Lancet Glob. Heal. 2014.Vol. 2, № 6, P. 323–333.

[2] Tantbirojn P, Crum CP, Parast MM. Pathophysiology of placenta creta: The role of decidua and extravillous trophoblast. Placenta 2008; 29:639–645

DOI 10.18699/BGRS/SB-2020-330

Genome mining for novel bioactive peptides

Dmitrii Travin
Center of Life Sciences, Skolkovo Institute od Science and Technology, Moscow, Russia
dmitrii.travin@skoltech.ru

Dmitry Bikmetov
Institute of Molecular Genetics RAS, Moscow, Russia
bikdm12@gmail.com

Konstantin Severinov
Center of Life Sciences, Skolkovo Institute od Science and Technology, Moscow, Russia
k.severinov@skoltech.ru

Abstract — **Genome mining approach was applied to the search of novel bioactive peptide compounds sharing a functionally important post-translational modification – azol(in)e cycles. It revealed the great diversity of uncharacterized biosynthetic gene clusters, some of which likely encode the biosynthetic pathways of novel compounds inhibiting key cellular functions including translation.**

Keywords — genome mining, LAPs, antibiotics, translation inhibitors

Motivation and Aim

The spread of drug resistance to existing antimicrobials among pathogenic bacteria becomes a major threat to modern public healthcare. The urgent need for new antibiotics leads to the development of methods of drug discovery alternative to classical activity-based screenings. Genome mining strategy, which benefits from rapid accumulation of publically available genomic information, is becoming one of the most powerful tools, which has been already successfully applied to the search for novel compounds from diverse groups of microbial natural products.

Earlier we applied BLAST-based search for novel biosynthetic gene clusters (BGCs) of linear azol(in)e modified peptides (LAPs), which allowed the identification and subsequent structural and functional characterization of two compounds inhibiting protein biosynthesis on the bacterial ribosome – klebsazolicin [1] and phazolicin [2]. In this work we aimed to identify novel groups of BGCs of azol(in)e-containing peptide natural products through the comprehensive bioinformatic analysis of all genomic regions containing genes of YcaO-domain containing proteins, which are considered to be a hallmark of azol(in)e modification installation [3].

Methods

First, we retrieved the sequences of YcaO-containing proteins from RefSeq databases using HMMs for YcaO from Pfam and TIGRFAMs databases. Subsequent filtration allowed to get rid of the closely related sequences from the genomes of overrepresented species. Second, genomic regions around the collected genes of YcaO-containing proteins were annotated and used to build a sequence similarity network of BGCs, which was then analyzed manually.

Results

The approach described allowed the identification of approximately 15000 unique YcaO-domain containing BGCs in the publically available genomes. The analysis of the sequence similarity network of clusters revealed both BGCs directing the biosynthesis of already known and novel compounds. We performed manual in-depth bioinformatic analysis for three groups of previously uncharacterized clusters, which allowed to predict possible structure of the encoded bioactive peptides as well as shed the light on several intriguing aspects of the evolution of the modified peptide biosynthesis-encoding clusters.

Acknowledgment

Supported by RSF grant (19-14-00266).

References

[1] M. Metelev *et al.*, “Klebsazolicin inhibits 70S ribosome by obstructing the peptide exit tunnel.,” *Nat. Chem. Biol.*, vol. 13, no. 10, pp. 1129–1136, Oct. 2017.

[2] D. Y. Travin *et al.*, “Structure of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific mode of bacterial translation inhibition,” *Nat. Commun.*, vol. 10, no. 1, p. 4563, 2019.

[3] B. J. Burkhart, C. J. Schwalen, G. Mann, J. H. Naismith, and D. A. Mitchell, “YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function.,” *Chem. Rev.*, vol. 117, no. 8, pp. 5389–5456, Apr. 2017.

DOI 10.18699/BGRS/SB-2020-331

The mTOR pathway activity in ASD and post-infectious neuropsychiatric autoimmune disorders

Ekaterina A. Trifonova
Gene Engineering Laboratory
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
trifonova.k@rambler.ru

Alexandra I. Klimenko
Laboratory of Molecular Genetic Systems
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
klimenko@bionet.nsc.ru

Sviatoslav L. Bezrodny
Laboratiry of Bifidobacteria Biology
Institute of Epidemiology & Microbiology
Moscow, Russia
frebiotik@mail.ru

Zakhar S. Mustafin
Laboratory of Molecular Genetic Systems
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
mustafinZS@bionet.nsc.ru

Sergey A. Lashin
Laboratory of Molecular Genetic Systems
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
lashin@bionet.nsc.ru

Alex V. Kochetov
Gene Engineering Laboratory
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
ak@bionet.nsc.ru

Abstract — **Autism spectrum disorder (ASD) has a strong and complex genetic component with an estimate of more than 1000 genes implicated cataloged in SFARI (Simon's Foundation Autism Research Initiative) Gene database. We conducted gene-set analyses and revealed that 179 out of 281 genes (64%) included in the first three categories of the database ("High confidence", "Strong candidate", and "Suggestive evidence") can be attributed to one of the four groups: 1. FMRP target genes, 2. mTOR signaling network genes, 3. mTOR-modulated genes, 4. Vitamin D3 sensitive genes. The additional gene network analysis revealed 43 new genes and 127 new interactions so in the whole 222 out of 281 (79%) high scored genes from SFARI Gene database are connected with mTOR signaling activity and/or dependent on vitamin D3 availability directly or indirectly. Infections and the resulting immune response to these infections have recently been increasingly recognized as pathogenic mechanism for neuropsychiatric disorders. Sydenham's Chorea (SC), a widely recognized post-streptococcal autoimmune disease, provides a model for studying the disorders. Significant overlap of ASD and SC/PANS/PANDAS symptoms makes complicated both differential diagnosis and search for the common molecular mechanisms underlying the disorders. We analyzed bioinformatically the small intestine microbiome obtained by mass spectrometry of microbial markers in the Russian population of patients with ASD. As a result, quantitative data on the percentage of patients with significantly elevated levels of Streptococcus spp and Clostridium spp in this population were first obtained.**

Keywords *— autism spectrum disorders (ASD), Sydenham's Chorea (SC), PANDAS (pediatric autoimmune neuropsychiatric disorder associated with streptococcal infections), PANS (pediatric autoimmune neuropsychiatric disorder), mTOR (mechanistic target of rapamycin)*

Introduction

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder with complex genetic, environmental, and epigenetic components. The search for genetic factors underlying ASD has led to the identification of more than one thousand genes cataloged in SFARI (Simon's Foundation Autism Research Initiative) Gene database that has scored and ranked genes into one of the seven categories [1]. A significant part of both syndromic and idiopathic autism cases can be attributed to disorders caused by mTOR-dependent translation deregulation. Protein biosynthesis, or translation, is a finely regulated process, with the central role played by mTOR kinase (mechanistic target of rapamycin). A mutational aberration in at least one of the links of the mTOR signaling pathway impairs the synaptic plasticity and behavior. The deregulation of local translation in dendrites is connected with the following monogenic ASDs: neurofibromatosis type 1, Noonan syndrome, Costello syndrome, Cowden syndrome, tuberous sclerosis, fragile X syndrome, and Rett syndrome [2].

Results and Discussion

We conducted gene-set analyses and revealed that 179 out of 281 genes (64%) included in the first three categories of the database ("High confidence", "Strong candidate", and "Suggestive evidence") can be attributed to one of the four groups: 1. FMRP target genes, 2. mTOR signaling network genes, 3. mTOR-modulated genes, 4. Vitamin D3 sensitive genes. The additional gene network analysis revealed 43 new genes and 127 new interactions so in the whole 222 out of 281 (79%) high scored genes from SFARI Gene database are connected with mTOR signaling activity and/or dependent on vitamin D3 availability directly or indirectly. A significant portion of genes that belong to more than one of four categories was of particular interest (Table 1).

TABLE 1. HIGH SCORED GENES IMPLICATED IN AUTISM SUSCEPTIBILITY THAT FALL INTO MORE THAN ONE CATEGORY

Categories of genes	Total number	Elements
FMRP target ∩ Vitamin D sensitive ∩ mTOR signaling pathway ∩ mTOR-modulated	1	PTEN
FMRP target ∩ mTOR signaling pathway ∩ mTOR-modulated	4	SYNGAP1 CUL7 TSC2 CTNNB1

FMRP target ∩ Vitamin D sensitive ∩ mTOR-modulated	4	SMARCC2 ANKRD11 KDM5C CLASP1
Vitamin D sensitive ∩ mTOR signaling pathway ∩ mTOR-modulated	1	CCT4
FMRP target ∩ mTOR-modulated	29	EP400 USP7 UBR5 PHRF1 GGNBP2 UBE3C KMT2E CHD8 AKAP9 ATP2B2 PLXNB1 MED13 LEO1 DPYSL2 KAT6A TRIP12 DYNC1H1 STXBP1 EHMT1 MYH10 AGAP2 CUX1 TAOK2 ANK3 SETD5 SLC12A5 CC2D1A TNRC6B NCKAP1
FMRP target ∩ mTOR signaling pathway	2	MTOR PRKCB
FMRP target ∩ Vitamin D sensitive	15	NUAK1 SMARCA4 MYO9B SHANK3 MED13L ARID1B DLGAP1 SPARCL1 TCF20 WDFY3 KCNQ3 SBF1 TCF4 PRICKLE2 HIVEP3
Vitamin D sensitive ∩ mTOR-modulated	10	RAB2A ELP4 PHB PRICKLE1 BCKDK NAA15 ASTN2 ETFB CAPRIN1 SLC35B1
Vitamin D sensitive ∩ mTOR signaling pathway	2	DYRK1A SLC7A5

We have hypothesized that genetic and/or environment mTOR hyperactivation including provoked by vitamin D deficiency might be a common mechanism controlling expressivity of most autism predisposition genes and even core symptoms of autism.

Infections and the resulting immune response to these infections have recently been increasingly recognized as pathogenic mechanism for neuropsychiatric disorders. Sydenham's Chorea (SC), a widely recognized post-streptococcal autoimmune disease, provides a model for studying the disorders. PANDAS (pediatric autoimmune neuropsychiatric disorder associated with streptococcal infections) has been proposed as a variant of SC, so that they have a common pathogenesis, despite the unique profile of the predominantly psychiatric symptoms of PANDAS [3]. It should be noted that streptococcus spp are not unique in their ability to cause autoimmune neuropsychiatric complications; among alternative inducers, influenza and chickenpox viruses, mycoplasmas, in general, the entire spectrum of such syndromes is called PANS.

Significant overlap of ASD and SC/PANS/PANDAS symptoms makes complicated both differential diagnosis and search for the common molecular mechanisms underlying the disorders. We analyzed bioinformatically the small intestine microbiome obtained by mass spectrometry of microbial markers in the Russian population of patients with ASD. As a result, quantitative data on the percentage of patients with significantly elevated levels of Streptococcus spp and Clostridium spp in this population were first obtained.

It was recently shown that mTORC1 is a central pathway in the pathogenesis of systemic lupus erythematosus and other autoimmune diseases [4]. At the same time, a significant percentage of ASD is associated with activation of the mTORC1 signaling pathway. The causal relationship between mTORC1 activation, an increased level of pathogenic microflora in the intestines of patients with ASD, and autoimmune syndromes often found in ASD should be the subject of future research.

ACKNOWLEDGMENT

This research was supported by the Russian State Budget (project No. 0259-2019-0008).

REFERENCES

[1] B. S. Abrahams, et al SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism. 2013, vol. 4, pp. 36-39.

[2] J.O. Lipton, M. Sahin. The Neurology of mTOR. Neuron. 2014, vol. 84, pp. 275–291.

[3] K.A. Williams, S.E. Swedo. Post-infectious autoimmune disorders: Sydenham's chorea, PANDAS and beyond. Brain Res. 2015 vol. 1617, pp. 144-54.

[4] T. Suto, T. Karonitsch. The immunobiology of mTOR in autoimmunity. J Autoimmun., in press.

DOI 10.18699/BGRS/SB-2020-332

Placental transcriptome co-expression analysis reveals key biomarkers and pathways of preeclampsia

Trifonova E.
TNRMC RAS, Tomsk, Russia
SibMed, Tomsk, Russia
ekaterina.trifonova@medgenetics.ru

Zarubin A.
TNRMC RAS, Tomsk, Russia
aleksei.zarubin@medgenetics.ru

Babovskaya A.
TNRMC RAS, Tomsk, Russia
anastasia.babovskaya@medgenetics.ru

Markov A.
TNRMC RAS, Tomsk, Russia
anton.markov@medgenetics.ru

Stepanov V.
TNRMC RAS, Tomsk, Russia
vadim.stepanov@medgenetics.ru

***Abstract* — Preeclampsia is a complication of pregnancy characterized by new-onset hypertension and proteinuria of gestation, with serious consequences for mother and infant. Although a vast amount of research has been performed on the pathogenesis of preeclampsia, the underlying mechanisms of this multisystemic disease have remained to be fully elucidated. We identified the significant role of disturbance of intercellular interactions and regulation of proteins modification in placental tissue during the development of the PE. Among the genes involved in these key pathways, 9 hub genes and 3 master regulators were identified from the co-expression and upstream analysis networks. The present study may provide a basis for exploring potential novel genes and pathways as therapeutic targets for preeclampsia.**

***Keywords* — preeclampsia, gene expression, placenta, WGCNA, upstream analysis**

Motivation and aim

Preeclampsia (PE) remains a leading cause of maternal/fetal mortality and morbidity associated with gestational hypertension and proteinuria. The underlying mechanism and preventive treatment remain unknown [1]. Due to possible multifactorial causes involved, an increase in "omics" experimental approaches is noted, generating a large amount of information for PE. Therefore, the identification of key genes and pathways is of much importance for clarifying molecular mechanism of PE initiation and progression.

The identification of key genes and pathways of PE by bioinformatics analysis of transcriptomic data.

Methods

Genome-wide expression profiling was performed on placental tissue from preeclamptic and normal (n = 47) pregnancies. The examined patients were from the Russian and Yakut populations. Eight original datasets (GSE25906, GSE30186, GSE35574, GSE44711, GSE60438, GSE6573, GSE73374, GSE94643) from patients with PE and normal pregnancy (n = 129) were downloaded from Gene Expression Omnibus and were further integrated and analyzed with our data.

Statistical software R was used for significance analysis of differentially expressed genes (DEGs) between PE samples and normal samples. Gene Set Enrichment Analysis (GSEA) were applied for the identification of pathways in which DEGs significantly enriched. Cytoscape software was for the construction of protein-protein interaction (PPI) network and module analysis to find the hub genes and key pathways. We applied upstream analysis approach implemented in geneXplain platform (genexplain.com) for detection of upstream transcriptional regulators and their regulatory networks. Finally, weighted correlation network analysis (WGCNA) was conducted to further screen critical gene modules with similar expression pattern and explore their biological significance.

Results

Subsequently, 4939 DEGs between PE patients and healthy women were identified. Using Gene Set Enrichment Analysis we identified the significant role of disturbance of intercellular interactions and regulation of proteins modification in placental tissue during the development of the PE. Weighted genes co-expression network analysis reveals a similar distribution along the modules detected both in normal and preeclampsia conditions. The present co-expression network construction yielded 3987 nodes based on 129 PE and normal pregnancy expression samples. A total of 31 modules were identified as PE related clusters and 34 modules were identified in control group. Using WGCNA, we found 6 clusters containing 86 genes associated only with PE. We identified 9 hub genes (*IFIH1, IFI44L, IFI44, CXCL9, CXCL10, RAD21, YY1, GYPA, GYPB*) in these clusters using cytoHubba (MCC, rank < 5) and STRING (score≥0,7.). We applied upstream analysis approach implemented in geneXplain platform and identified master regulators (MAPK3, TP53 and UBE2D1) that are new therapeutic targets. These key genes may be potential biomarkers of diagnosis, therapy and prognosis for PE.

Acknowledgment

Supported by the RFBR (grant No. 18-29-13045, No. 18-44-700007).

References

[1] Apicella C. et al. (2019) The Role of Epigenetics in Placental Development and the Etiology of Preeclampsia. Int J Mol Sci. 20(11). pii: E2837. doi: 10.3390/ijms20112837.

DOI 10.18699/BGRS/SB-2020-333

Loci and genes involved in chronic musculoskeletal pain identified via analysis of genetically independent pain phenotypes

Yakov A. Tsepilov
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia
tsepilov@bionet.nsc.ru

Sodbo Z. Sharapov
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Lennart C. Karssen
PolyOmica
's-Hertogenbosch, the Netherlands

Yurii S. Aulchenko
Laboratory of Recombination and Segregation Analysis
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Maxim B. Freidin
Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences
King's College London
London, UK

Elizaveta E. Elgaeva
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Pradeep Suri
Division of Rehabilitation Care Services
VA Puget Sound Health Care System
Seattle, USA

Alexandra S. Shadrina
Laboratory of Theoretical and Applied Functional Genomics
Novosibirsk State University
Novosibirsk, Russia

Jan van Zundert
Department of Anesthesiology and Pain Medicine
Maastricht University Medical Centre
Maastricht, The Netherlands

Frances M.K. Williams
Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences
King's College London
London, UK

Abstract — **We have evaluated four genetically independent pain phenotypes of four common chronic musculoskeletal pains (GIPs). We assume that the first GIP represents a biopsychological component of chronic musculoskeletal pain, related to physiological and psychological aspects and possibly reflecting pain perception and processing.**

Keywords — musculoskeletal pains; genome-wide association study; UK Biobank; pleiotropy

Introduction

Pain is an unpleasant sensory experience caused by tissue damage (or by the threat of its occurrence). It is an adaptive response of organism to the pathogenic influence, which activates various protection mechanisms. Pain and especially chronic pain, conventionally defined as a pain that lasts for 3 and more months sequentially, has a negative impact on all spheres of human life. However, nowadays little is known about the mechanisms of chronic pain onset. Somewhat, it is caused by the fact, that genetic studies of pain are rather complicated due to the high complexity and heterogeneity of the phenotype. Nevertheless, recent studies have shown, that chronic pain is a complex genetic trait with heritability up to 60%, based on different estimates. Moreover, it was found, that various types of chronic pain have a number of genetic factors in common. In this research we decided to apply the method of principal component analysis (PCA) to reduce the heterogeneity and studying of commonalities and pathways that are shared by distinct, although related, pain phenotypes.

The aim of this work was to study the main genetic component that underlies the risk of chronic musculoskeletal pain observed at four different sites: back, neck/shoulder, hip, and knee, using data from the UK Biobank.

Materials and methods

This work is based on the results of genome-wide association studies in the UK Biobank [1] data with total sample size of 434 thousand people, split into a discovery set of 256,000 people and a replication set of 178,000 people. The matrix of genetic correlations for four chronic musculoskeletal pains was calculated using LD Score regression [2], then the principal components analysis (PCA) was applied to compute the genetically independent pain phenotypes (GIPs). The meta-analysis was conducted using METAL [3]. The further in silico functional annotation of GIPs was performed using DEPICT [4], COJO [5] and GWAS-MAP platform [6].

Results

We have evaluated four GIPs of four chronic musculoskeletal pains. Six loci were shown to be associated with the first genetic principal component (GIP1), five of them were replicated. Also we have found three loci, associated with the second genetic principal component (GIP2) with one of them being replicated. We identified 13 genes located near the lead SNPs (± 250 kb) in found loci. Most of them were shown by us to be related to musculoskeletal disorders, nervous system and skeletal development. Results of DEPICT analysis for SNP sets associated with GIP1 at $P < 5.0e-08$ and $P < 1.0e-05$ have identified a significant enrichment of terms, related to nervous system and its development and to sensory organ morphogenesis, while the DEPICT analysis of other GIPs did not provide any significant results. The analysis of genetic correlations for GIPs has revealed high correlations between GIP1 and different anthropometric, socio-demographic and psychiatric human traits. The other GIPs

turned to be genetically correlated with various morphological and anthropometric traits.

We speculate that in the framework of the conventional biopsychosocial model of the chronic pain, the GIP1 relates to the physiological ("bio-") and psychological aspects and is likely reflecting pain perception and pain processing.

ACKNOWLEDGMENT

The work of SZS was supported by the Russian Ministry of Education and Science under the 5-100 Excellence Programme. The work of YSA was supported by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genetics (project 0324-2019-0040-C-01 / AAAA-A17-117092070032-4). The work of YAT, ASSh, and EEE was supported by the Russian Foundation for Basic Research (project 19-015-00151). The contribution of LCK was funded by PolyOmica. Dr. Suri was supported by VA Career Development Award # 1IK2RX001515 from the United States (U.S.) Department of Veterans Affairs Rehabilitation Research and Development (RR&D) Service. Dr. Suri is a Staff Physician at the VA Puget Sound Health Care System. The contents of this work do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

REFERENCES

[1] C. Sudlow et al. (2015) UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age, PLOS Medicine, 12: e1001779.

[2] B. Bulik-Sullivan et al. (2015) An atlas of genetic correlations across human diseases and traits, Nature Genetics, 47: 1236–1241.

[3] C.J. Willer et al. (2010) METAL: fast and efficient meta-analysis of genomewide association scans, Bioinformatics, 26: 2190–2191.

[4] T.H. Pers et al. (2015) Biological interpretation of genome-wide association studies using predicted gene functions, Nature Communications., 6: 5890.

[5] J. Yang et al. (2013) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits Genetic Investigation of ANthropometric Traits (GIANT) Consortium 4 , DIAbetes Genetics Replication And Meta-analysis (DIAGRAM), Nature Genetics, 44: 369–372.

[6] Gorev DD, Shashkova TI, Pakhomov E, Torgasheva A, Klaric L, Severinov A, et al. GWAS-MAP: a platform for storage and analysis of the results of thousands of genome-wide association scans. Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2018) The Eleventh International Conference. Novosibirsk: ICG SB RAS; 2018. p. 43. doi:10.18699/BGRSSB-2018-020

DOI 10.18699/BGRS/SB-2020-334

Mitochondrial dysfunction and redox balance alterations in the development of AD-like pathology in OXYS rats

Mikhail Tyumentsev
Molecular mechanisms of aging lab
ICG SB RAS
Novosibirsk, Russia
landselur@bionet.nsc.ru

Natalia Muraleva
Molecular mechanisms of aging lab
ICG SB RAS
Novosibirsk, Russia
Myraleva@bionet.nsc.ru

Yulia Polienko
Laboratory of Nitrogen Compounds
NIOCH SB RAS
Novosibirsk, Russia
polienko@nioch.nsc.ru

Artyom Gorodetsky
Laboratory of magnetic radio spectroscopy
NIOCH SB RAS
Novosibirsk, Russia
gorodaa@nioch.nsc.ru

Elena Bagryanskaya
Laboratory of magnetic radio spectroscopy
NIOCH SB RAS
Novosibirsk, Russia
egbagryanskaya@nioch.nsc.ru

Abstract — **This study focuses on the relationship between mitochondrial dysfunction and redox-status in the context of the development of signs of Alzheimer's disease (AD). Mitochondrial dysfunction is considered the missing link between brain aging and AD [1], the most common type of age-related dementia worldwide, but the exact causal relationship between mitochondrial dysfunction and the transition from healthy aging to AD remains to be fully understood. Oxidative stress is thought to play a significant role in this mitochondrial dysfunction, leading to cellular damage in redox imbalance. However, the extent of these processes and timing of their occurrence within the scope of AD remain hard to study, especially so in the early, pre-clinical stages of the disease. We explored the mechanisms underlying the disruption of mitochondrial function, their impact on the initiation and progression of pathological molecular cascades of AD, and assessed the changes in redox status as one of the main consequences of oxidative stress. This investigation was conducted using senescence-accelerated OXYS rats, which spontaneously develop all major signs of AD and largely reproduce the stages of the disease. We concluded that mitochondrial dysfunction appears to mediate or possibly even initiate AD-like pathology in OXYS rats. Importantly this takes place with no apparent connection to redox imbalance as on both transcriptional and biochemical levels OXYS rats display no significant changes in redox-status and ROS production as compared to controls.**

Keywords— mitochondria, redox, EPR, Alzheimer's disease, OXYS rats

Introduction

Growing evidence suggests that mitochondrial dysfunction is an early event in sporadic Alzheimer's disease (AD), but the timeframe within which mitochondrial dysfunction participates in the transition from healthy aging to AD remains to be studied in depth due to the difficulty of studying early changes in humans and lack of appropriate animal models. Importantly, mitochondria are also the main source of the reactive oxygen species (ROS) and the focal point of many cellular redox reactions. Thus mitochondrial dysfunction is assumed to be directly associated with oxidative stress and redox imbalance which are thought to be tightly connected to aging and thus to age-related diseases.

Using OXYS rats – animal model of sporadic AD exhibiting neuronal and synaptic loss, mitochondrial structural abnormalities, tau hyperphosphorylation, increased Aβ levels, and Aβ deposition [2] – we evaluated the interrelationship between mitochondrial dysfunction and the initiation and progression of AD signs, as well as its connection with redox imbalance in the brain. We assessed the mitochondrial ultrastructure in OXYS rat hippocampal neurons at the age preceding AD-like pathology (from birth to the age of 20 days), age of its manifestation (5 months), and well-pronounced changes (24 months). Ultrastructural alterations were then collated with amounts of proteins of mitochondrial dynamics (mitofusins MFN1 and 2, dynamin-1-like protein DRP1) and activity of respiratory chain complexes I, IV, and V in the mitochondria of prefrontal cortex and hippocampus. To compare the timeline of mitochondrial dysfunction with the timeline of changes in redox balance we have analyzed differential expression of genes related to redox regulation using RNAseq data. Results of this analysis were then verified by measuring the rate of ROS production by brain mitochondria, overall redox status of the brain, NAD/NADH ratio and glutathione content. We have found that signs of mitochondrial dysfunction are present even in the early age preceding the manifestation of AD signs and can exist without pronounced signs of redox dysregulation.

Methods

Wistar and OXYS rats were obtained from the Breeding Experimental Animal Laboratory of the Institute of Cytology and Genetics, the Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia). All animals received food and water ad libitum and were kept at 12/12 light/dark cycle.

In order to assess the extent of mitochondrial dysfunction in different ages, mitochondrial number and morphology was assessed by electron microscopy at the Interinstitutional Shared Center for Microscopic Analysis of Biological Objects (Institute of Cytology and Genetics, Novosibirsk, Russia). Activity of the respiratory chain complexes and the content of the proteins of mitochondrial dynamics were measured by ELISA assays.

Then, to compare the progression of mitochondrial dysfunction against the changes of redox status and redox-

responding systems, a list of genes related to redox regulating systems was composed using WikiPathways database and relevant terms from GO Ontology database. RNAseq data from the cortex of Wistar and OXYS rats aged 20 days, 12 and 18 months was then used to assess differential expression of genes related to redox status of OXYS rats in different stages of the development of AD signs. Redox status of the brain milieu was measured by EPR spectroscopy using redox-sensitive spin probe methoxycarbonyl-PROXYL. ROS production, glutathione content and NAD/NADH were measured using biochemical assays.

Results

Electron microscopy shows that on the morphological level, from 5 months on there is marked depletion of neuronal mitochondria. At the same time it should be noted that ultrastructurally damaged mitochondria with rarefied cristae are more abundant in OXYS rats only at the age of 24 months – at the advanced stage of AD signs. Levels of fusion protein MFN1 is higher in OXYS rats which, in conjunction with neuronal mitochondria showing certain degree of enlargement might be a sign of upregulated fusion. However, with age levels of fusion proteins do not increase in OXYS rats – unlike Wistar. OXYS rats display signs of upregulated fission with age. Activity of the respiratory chain complexes is decreased with 20 days old OXYS rats displaying decreased activity of complex IV – which is a known hallmark of brain aging and neurodegeneration [3]. Interestingly, decreased respiratory chain activity in young OXYS rats is not accompanied by NAD/NADH ratio shifting towards NADH. In fact, there are signs of a slight shift towards NAD+ which might be a sign of compensatory reaction.

Analysis of DEGs related to redox regulation revealed several genes that increase their expression with age both in OXYS and Wistar rats. While there are genes upregulated in OXYS rats compared to Wistar, we have found a number of genes upregulated in Wistar compared in OXYS. Thus, OXYS rats do not present a clear transcriptional profile that would signify a response to oxidative stress or redox imbalance more severe than that in age-matched Wistar rats. Supporting this, we have found no increase in the rate of ROS production in OXYS rats in the ages of manifestation and progression of the disease and only a slight increase while oxidizing a single combination of substrates at the age of 20 days. Importantly, with no increase of ROS production to act upon, the previously shown ability of mitochondrially targeted antioxidant SkQ1 to alleviate signs of the AD [4,5] might be due to changes in signal pathways in response to the treatment. The data pointing at lack of redox imbalance in OXYS rats are further corroborated by the fact that redox balance of the brain milieu has displayed an expected shift to the pro-oxidative state from 20 days to 12 months of age in both Wistar and OXYS rats with no interstrain differences. Additionally, we have observed an increase of glutathione content occurring from the age of 5 months to the age of 18 months, possibly as a part of compensatory response, which was slightly less pronounced in OXYS rats.

Conclusion

Here we present the timeline of the decline of mitochondrial function in OXYS rats. We show that mitochondrial dysfunction precedes other signs of AD and further progresses along with the progression of the disease itself. Interestingly, redox imbalance in the brain of OXYS rats appears to be largely uncoupled from the progression mitochondrial dysfunction and the development of AD signs, and thus might be seen as optional rather than inherent part of these conditions.

Acknowledgment

This study was supported by the Russian Foundation for Basic Research (Grant No. 17-03-01132-a) and Russian Science Foundation (Grant No. 19-13-00235).

References

[1] Grimm, Amandine, Kristina Friedland, and Anne Eckert. "Mitochondrial dysfunction: the missing link between aging and sporadic Alzheimer's disease." Biogerontology, vol. 17, no. 2, pp. 281–296, 2016.

[2] Stefanova, Natalia A., Natalia A. Muraleva, Elena E. Korbolina, Elena Kiseleva, Kseniya Yi Maksimova, and Nataliya G. Kolosova. "Amyloid accumulation is a late event in sporadic Alzheimer's disease-like pathology in nontransgenic rats." Oncotarget, vol. 6, no. 3, pp. 1396–1413, 2015.

[3] Tyumentsev, Mikhail A., Natalia A. Stefanova, Natalia A. Muraleva, Yulia V. Rumyantseva, Elena Kiseleva, Valentin A. Vavilin, and Nataliya G. Kolosova. "Mitochondrial dysfunction as a predictor and driver of Alzheimer's disease-like pathology in OXYS rats." Journal of Alzheimer's Disease, vol 63, no. 3, pp. 1075–1088, 2018

[4] Stefanova, Natalia A., Anzhela Zh Fursova, and Nataliya G. Kolosova. "Behavioral effects induced by mitochondria-targeted antioxidant SkQ1 in Wistar and senescence-accelerated OXYS rats." Journal of Alzheimer's Disease, vol 21, no. 2, pp. 479-491, 2010

[5] G Kolosova, Nataliya, Mikhail A Tyumentsev, Natalia A Muraleva, Elena Kiseleva, Anton O Vitovtov, and Natalia A Stefanova. "Antioxidant SkQ1 alleviates signs of Alzheimer's disease-like pathology in old OXYS rats by reversing mitochondrial deterioration." Current Alzheimer Research, vol 14, no. 12, pp. 1283-1292, 2017

DOI 10.18699/BGRS/SB-2020-335

Syndecan-1 expression in ovarian endometrioid cancer tissue has negative correlation with estrogen status

Alexander V. Volchek
Novosibirsk State Medical University
Novosbirsk, Russia
alexander@volcheck.ru

Yulia S. Timofeeva
Novosibirsk State Medical University
Novosbirsk, Russia
dr.yustimofeeva@gmail.com

Yanina M. Evseeva
Novosibirsk State Medical University
Novosbirsk, Russia
janinaevseeva@yahoo.com

Dmitriy V. Morozov
City Clinical Hospital №1
Novosibirsk, Russia
mdvil07@mail.ru

Igor O. Marinkin
Novosibirsk State Medical University
Novosbirsk, Russia
rectorngmu@yandex.ru

Svetlana V. Aidagulova
Novosibirsk State Medical University
Novosbirsk, Russia
s.aydagulova@gmail.com

Abstract — **Cancer treatment shifts from mono-targeted to combination therapy, considering the tumor microenvironment with heparan sulfates, including Syndecan-1 (SDC1). The aim of this work was to study correlation between core-protein SDC1 and estrogen receptors expression in surgical material of 15 women aged 53,0 ± 10,9 years with endometrioid ovarian cancer (EOC) using immunohistochemistry and automatic image analysis. The staining intensity of SDC1 in tumor tissue was 1304,23 ± 65,85 units, the staining intensity of the estrogen receptors as 1137,88 ± 135,40 units. The significant inverse correlation (R = - 0,512, p = 0,003) between SDC1 and estrogen receptors expression levels were obtained. Most likely, this deals with progressive character of EOC.**

Keywords — ovarian endometrioid cancer, syndecan-1, estrogen receptors, tumor tissue, immunohistochemistry

Motivation and Aim

Ovarian cancer has the 3rd place among all gynecological neoplasms but is one of the leading causes of death in the structure of gynecological oncology worldwide [1]. Status of estrogen receptors was shown to play crucial role for the progression of OEC and being predictor of the response to hormone therapy [2]. The lack of expression of these receptors is associated with low-grade tumors, which are more aggressive and have poor prognosis. Nowadays, it is expected that as cancer treatment shifts from mono-targeted to combination therapy, the management of biosynthesis and degradation of heparan sulfates (of which SDC1 is a representative) are likely to become the main new direction of cancer therapy [3]. The aim of this work was to study correlation between SDC1 and estrogen receptors expression in women with EOC.

Methods and Algorithms

A clinical and morphological study of 15 women with EOC aged 40 to 77 years (mean age 53,0 ± 10,9 years) was carried out. The diagnosis of EOC was based on specific endometrioid histotype of surgical samples with positive expression of immunohistochemical markers: Cytokeratin 7, PAX 8, Vimentin and estrogen and progesterone receptors. Women with pregnancy and severe immunodeficiency were excluded. In all cases the informed consent of the patient to the examination and treatment of EOC was taken in accordance with the directives of the European Community (86/609 / EEC) and the Helsinki Declaration, in compliance with the «Ethical Principles for Scientific Medical Research with Human Participation» and in accordance with the «Rules of Clinical Practice in Russian Federation». Surgical tumor samples were fixed in 10% buffered formalin solution. For immunohistochemistry, 5-mkm serial sections of formalin-fixed, paraffin-embedded tissue sections were deparaffinized and antigens were retrieved. The mouse monoclonal anti-SDC1 and rabbit monoclonal anti-Estrogen Receptor primary antibodies (ThermoScientific, cat. № MS-1793RQ and RM-91-50, Clone SP1, respectively) were used. For negative control, the primary antibodies were replaced by 5% bovine serum. Immunostaining patterns were visualized using UltraVision Quanto Detection System HRP (ThermoFisher Scientific). The sections were counterstained with Hematoxylin and observed by light microscopy using Axio Scope.A1 microscope with the camera AxioCam MRc5 and software ZEN blue. For the quantitative analysis the average intensity of the intra- and extracellular DAB-positive products was estimated, taking into account the area of the staining zone in pixels and the relative gray gradations by the recommended formula [4] minimum of 50 images per antibody with magnification 40 x 10 (Zeiss, Oberkochen, Germany). Statistical processing of the results was performed using the statistical software application package of StatSoft inc. (2011). STATISTICA (data analysis software system) version 10.www.statsoft.com. A value of $p < 0,05$ was considered to indicate a statistically significant difference. Due to small amount of patients nonparametric statistic methods were used. Correlations were evaluated using Spearman rank R. Comparing of intensity values (staining concentrations) of SDC1 and estrogen receptors were done with Wilcoxon test.

Results

The optimal cytoreductive surgery was performed for 10 patients, in the volume of complete cytoreduction in 3 cases and in suboptimal cytoreduction in 2 cases. After surgical staging according to the FIGO classification, 2nd stage of EOC was diagnosed in 4 patients, 3rd stage in 10 patients and 1 woman had stage 4. Microscopically tumor tissue in the surgical material consisted of a large number of fused glands or cribriform proliferations of spindle-shaped cells, oval or tubular pseudoglands, which had a clear luminal edge, lined with stratified, mucin-free epithelium. Some areas have destructive growth with obvious stromal invasion consists of cells or group of cells which infiltrated randomly the desmoplastic stroma. Secretory changes were resemble the

early phase of endometrial secretion; sometimes in the stroma luteinized cells were detected. In EOC samples the automated image analysis revealed the SDC1 core-protein staining intensity as 1304,23 ± 65,85 units (varying from 1281,02 to 1335,71). The estrogen receptors in tumor tissue were expressed in average as 1137,88 ± 135,40 units (varying from 1024,73 to 1239,84 units) (p = 0,003). The significant inverse correlation between SDC1 and estrogen receptors expression levels were obtained (R = - 0,512). Cancer progression is closely related to interactions between cancer cells and tumor microenvironment that are in a dynamic interplay and are regulated by extracellular matrix. Overexpression of SDC1 has been reported in breast, pancreatic, prostate, ovarian, and endometrial cancer [5].

Conclusion

The significant inverse correlation (R = - 0,512, p = 0,003) between SDC1 and estrogen receptors expression levels were obtained. Most likely, this concerns the progressive feature of the EOC, which requires further researches.

ACKNOWLEDGMENT

The study was carried out as part of the State Assignment of the Ministry of Health of Russian Federation for research and development № 056-00141-18-00 (2018 – 2020).

REFERENCES

[1] Torre L.A. et al. (2018) Ovarian cancer statistics. CA Cancer J. Clin. 68: 284–296.

[2] Piperigkou Z., Karamanos N.K. (2019) Estrogen receptor-mediated targeting of the extracellular matrix network in cancer. Semin Cancer Biol. pii: S1044-579X(19)30088-4.

[3] Nagarajan A. et al. (2018) Heparan sulfate and heparan sulfate proteoglycans in cancer initiation and progression. Front Endocrinol (Lausanne). 9: 483.

[4] Bouzin C. et al. (2016) Digital pathology: elementary, rapid and reliable automated image analysis. Histopathology. 68(6):888-896.

[5] Palaiologou M. et al. (2014) CD138 (syndecan-1) expression in health and disease. Histology and Histopathology. 29(2): 177–189.

DOI 10.18699/BGRS/SB-2020-336

Evaluation of biological activity of the conjugates of granulocyte-macrophage colony stimulating factorwith alendronic acid

Volosnikova Ekaterina Aleksandrovna
department of technology development and pilot production of biologicals
IMBT FBRI SRC VB «Vector»,
Rospotrebnadzor, Berdsk, Russia
volosnikova_ea@vector.nsc.ru

Esina Tat'yana Igorevna
department of technology development and pilot production of biologicals
IMBT FBRI SRC VB «Vector»,
Rospotrebnadzor, Berdsk, Russia

Bateneva Alena Vladimirovna
department of biological studies
IMBT FBRI SRC VB «Vector»,
Rospotrebnadzor, Berdsk, Russia

Shimina Galina Grigor'evna
department of biological studies
IMBT FBRI SRC VB «Vector»,
Rospotrebnadzor, Berdsk, Russia

Danilenko Elena Dmitrievna
department of biological studies
IMBT FBRI SRC VB «Vector»,
Rospotrebnadzor, Berdsk, Russia

Abstract — **We have developed a method for the synthesis of rhGM-CSF protein conjugates with alendronic acid by use of water-soluble carbodiimide. The biological activity and accumulation in bone marrow of GM-CSF conjugates with ALN have been confirmed.**

Keywords — Recombinant human granulocyte-macrophage colony stimulating factor, neutropenia, affinity, activity, conjugate

Introduction

Recombinant human granulocyte-macrophage colony stimulating factor (GM-CSF) is used as a hematopoietic stimulant in patients with various diseases accompanied by neutropenia. However, the clinical use of GM-CSF preparations revealed a number of adverse reactions associated with its long-term use (fever, myalgia, bone pain, etc.). The development of targeted therapeutics is one of the ways to reduce the administered dose of the drug, and, as a consequence, its toxicity. Targeted drug delivery to the bone marrow, apparently, can enhance the growth and differentiation of hematopoietic cells due to GM-CSF localization at the site of hematopoiesis. As a vector molecule, alendronic acid (ALN), having affinity for the bone tissue, can be used.

Materials and Methods

For this study we used reagents from AppliChem (Germany), Sigma-Aldrich (USA), Bio-Rad (USA). For the synthesis of conjugates we used the preparation of GM-CSF (substance) manufactured by IMBT FBRI SRC VB "Vector".

The molecular weight of the conjugates was determined by use of vertical gel electrophoresis in 15% polyacrylamide gel under denaturing conditions, with Coomassie R-250 staining. The study of the conjugates` ability to accumulate in bone tissue was performed on an in vitro bone matrix model by use of hydroxyapatite chromatography. The in vitro biological activity of GM-CSF within the conjugates was evaluated by the stimulation level of proliferation of cytokine-dependent human erythroleukemia cells (TF-1). The in vivo hemostimulating activity was determined by use of a model of cytostatic myelosuppression induced by cyclophosphamide administration into mice. The level of distribution and accumulation of GM-CSF preparations in blood, femoral tissue, and bone marrow was performed on CD-1 outbred mice (ICR) with a single intravenous administration at the effective dose of 90 μg / kg. The content of rhGM-CSF in the samples taken 3 minutes, 1, 4 and 24 hours post administration was determined by the enzyme immunoassay with “Human Granulocyte-Macrophage Colony Stimulating Factor (GM-CSF) ELISA Kit” (Cusabio, China).

Results

The synthesis of GM-CSF with ALN was carried out through the carboxyl group of the protein with crosslinking agent 1-ethyl-3- [3-dimethylaminopropyl] carbodiimide (EDC). The conjugation was carried out on a solid phase, for what a chromatographic column with hydroxyapatite (HAP) was used. During the synthesis, the reaction components: GM-CSF, EDC, ALN were subsequently loaded onto a HAP column balanced with 2 mM potassium phosphate buffer, pH 7.0. Conjugates were eluted with 0.2 M potassium phosphate buffer, pH 7.0. The resulting preparation of GM-CSF conjugates was sterilely poured and frozen at - 20 °C.

It has been determined that the molecular weight of the resulting conjugate was 15.4 ± 0.3 kDa, which corresponds to the molecular weight of GM-CSF and, therefore, indicates the monomeric form of the protein within the conjugate. The protein, within the conjugate with ALN, retained the ability to stimulate the proliferation of human erythroleukemia cells TF-1: the ED50 of the conjugate was 0.38 ng / ml and did not differ much from the ED50 of GM-CSF substance (0.28 ng / ml). The in vitro experiment confirmed a high affinity of the conjugate for hydroxyapatite, used as a bone matrix analogue.

On the model of cytostatic myelosuppression, it was shown that GM-CSF conjugate, as well as the initial protein, accelerated the restoration of bone marrow cellularity and the number of segmentonuclear neutrophils in peripheral blood. The stimulating effect of the conjugate on the total number of karyocytes was more pronounced.

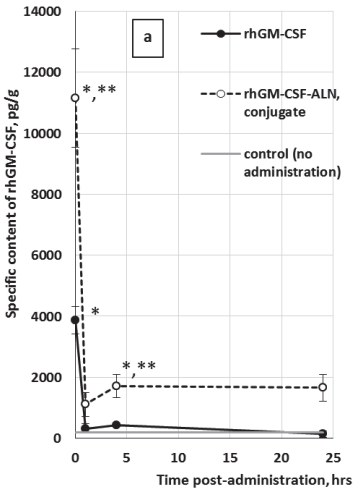
A study of distribution and accumulation of GM-CSF conjugates in the organs and tissues of mice confirmed their increased ability to penetrate into the bone tissue and bone marrow. Thus, the protein content in the femur during 4 hours after the conjugate administration was 2.9-3.9 times higher than the values recorded after the administration of GM-CSF

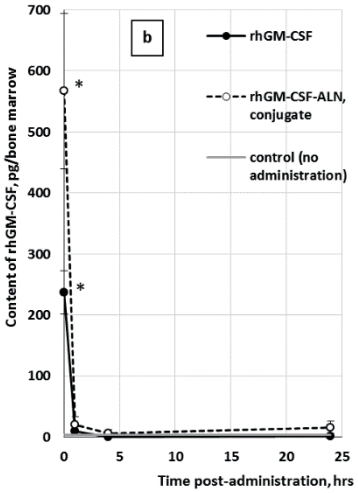
substance (Fig. 1a). The values of the indicator in this group remained higher compared to the control level (intact animals) until the end of the first day. The level of GM-CSF in bone marrow cells 3 minutes after the conjugate administration significantly exceeded the control indicator by 161 times, while after the substance administration, only 67 times (Fig. 1b).

The conjugated GM-CSF circulated longer in the peripheral blood of mice. Thus, 3 minutes post administration, the protein content in the blood was 19.2% of the administered dose for the conjugate and 7.2% for the substance; statistically significant differences between the groups persisted for at least 4 hours (Fig. 1c). It is important to note that GM-CSF conjugate was not detected in the blood by the end of the first day after the administration.

Conclusion

The data obtained indicate that conjugation of GM-CSF with ALN has enhanced the accumulation of GM-CSF in the bone marrow and increased its level of hemostimulating activity. Therefore, the developed approach is a promising platform for the further development of drugs with increased tropism to the bone marrow and hemostimulant effect.





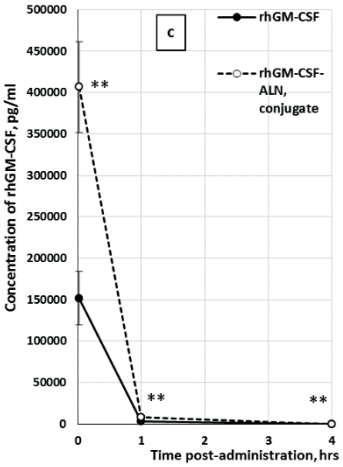


Fig. 1. Dynamics of changes of GM-CSF content in the femurs (a), bone marrow (b) and blood serum (c) of mice after a single intravenous administration of GM-CSF preparations. The significance of intergroup differences was assessed by the nonparametric Mann-Whitney U-test at a significance level of $p < 0.0170$ for tissues, $p < 0.05$ for blood; * - differences are statistically significant compared to the control; ** - differences are statistically significant compared to GM-CSF substance.

ACKNOWLEDGMENT

This work was financially supported by the Russian Foundation for Basic Research, a project in the framework of the contest "My First Grant", application No. 18-315-00236\19.

DOI 10.18699/BGRS/SB-2020-337

Impact of the gut microbiota on chemical risk assessment

A Wallace Hayes,
University of South Florida,
Tampa, FL USA

Abstract — **The multitude of microbes that reside in the human intestine plays an important role in the health and well being of humans. Not only do chemicals affect the composition of the gut microbiota but they affect the uptake and metabolism of ingested chemicals in multiple ways. The interplay among gut microbes, ingested xenobiotics, and toxicological effects is complex and not well understood. Recent studies have shown that the gut microbiota is a key player in the toxicity of drugs and chemicals, which is typically overlooked in current approaches for risk assessment. The microbial community in the gut may mediate or mitigate detrimental effects of chemicals by various mechanisms. Hence, there is a need to include, or at least consider, the gut microbiota as part of a human toxicological risk assessment protocol. Factors such as animal provider, batch/litter differences, and co-caging, among other factors, may significantly influence the outcome of a toxicity evaluation that is based on rodent experiments. A framework for assessment of gut microbiota upon exposure to drugs or chemicals is suggested.**

DOI 10.18699/BGRS/SB-2020-338

DNA damage to nervous tissue due to lead intoxication combined with glucose loading

N.L. Yakimova
Estern-Siberian Institute of Medical and Ecological Research, Angarsk, Russia
ynl-77@list.ru

E.S. Andreeva
Estern-Siberian Institute of Medical and Ecological Research, Angarsk, Russia

E.V. Buinova
Estern-Siberian Institute of Medical and Ecological Research, Angarsk, Russia

Abstract — **The effects of heavy metals on different organs and systems are known. Disorders in diabetes are also well understood. The compounding contribution of glucose loading was manifested in the increase of genotoxic lesions of nerve cells caused by lead intoxication. The DNA damage in the tail of nerve cell comets after exposure to lead and glucose loading increased to 52.86(13.16-73.34) percent compared to 25.28(12.24-54.79) percent of lead-intoxicated individuals. In lead-exposed and glucose-affected animals, cells with high DNA damage in the tail of comets dominated the brain, 62.77(54.68-70.87) percent of cells were apoptosis positive compared to 48.42(38.37-58.47) percent of rats receiving lead acetate without glucose loading.**

Keywords — lead; glucose loading; nervous system; DNA damage

Motivation and aim

According to observations from 2008 to 2017, many localities in the Russian Federation are included in the list with moderately dangerous and dangerous categories of lead contamination of soil, air and water bodies. The effects of heavy metals on various organs and systems, including DNA damage and apoptotic processes, have been studied [1,2]. Studies of diabetes mellitus on experimental models are relevant [3], but its genotoxic effects on the nervous system are not well understood. There is also a lack of genotoxicity data for lead intoxication combined with metabolic disorders. Research on these aspects is important for improving preventive measures and early diagnosis of health problems caused by aggravating factors in the workers and the population.

The aim was to study genotoxic changes in nerve cells in lead intoxication, which was aggravated by glucose loading in rats.

Methods

The experiment was performed on adult male rats. The animals were divided into three groups of 12. The control individuals were in group 1. Rats in group 2 received lead acetate daily at a dose of 50 mg/kg bw (calculated as metal) with drinking water for 30 days. In animals group 3 simulated lead acetate intoxication under similar conditions and simultaneously injected glucose at a dose of 6000 mg/kg bw twice a day at an interval of 6 hours for 30 days. Behavioral reactions were studied by means of the test «open field». The DNA damage in the brain tissue was evaluated using the «DNA comet» method.

Results

Under conditions of heavy metal exposure and glucose injection, there has been a reduction in locomotor and research activity in rats. The DNA content of the comet's tail reflects the extent of DNA damage. In lead exposure and glucose loading in rats of group 3, DNA damage in nerve cells increased compared to animals after lead poisoning "Table 1".

TABLE 1

The level of DNA damage in the brain cells of experimental animals, Me(Q25-Q75)	DNA content in the tail of a comet (%)		
	1 group Control	***2 group Lead acetate***	***3 group Lead acetate+ glucose***
brain	8.72(0-25.69)	25.28(12.24-54.79)^{*1}	52.86(13.16-73.34)$^{*2, *1}$

Footnote: ^{*2} – differences compared to group 2, p<0.001; ^{*1} – differences compared to group 1, p<0.001

In rats, simulated intoxication combined with glucose loading, the amount of DNA in the comet's tail in the brain was significantly higher than in individuals after exposure to lead acetate (p=0.0019) and compared to the control group ($p=3.808*E^{-15}$). The largest amount of intact DNA was observed in the brain of the comet in control group cells, and the smallest – in group 3 rats.

Apoptosis is considered to be those cells in which the tail of comets contains more than 30% of the damaged DNA. Overall, in rats of group 3, there was an increase in the number of apoptosis-positive cells to 62.77(54.68-70.87)%, compared to 48.42(38.37-58.47)%, in individuals of group 2 (p=0.035), and compared to 18.71(12.22-25.19)%, in the control group (p=0.0001). So, in animals with lead intoxication, combined with glucose loading, cells with high DNA damage in the tail of comets dominated.

Acknowledgment

Supported by the state task on the program of search scientific researches.

References

[1] Y. Ahmed, H. Eldebaky, K.Gh.M. Mahmoud, M. Nawito "Effects of lead exposure on DNA damage and apoptosis in reproductive and vital organs in female rabbits," Global Veterinaria, 9, pp. 401-408, 2012.

[2] S. Jadoon, A. Malik, "DNA damage by heavy metals in animals and human beings: an overview," Biochem. Pharmacol., 6, pp. 1000235, 2017.

[3] M. Küçük, R.B. Kalayci, A. Çevik, I. Elmas, M. Kaya, "Effect of aluminum on the blood - brain barrier permeability in acute and chronically hyperglycemic rats," Biol. Trace Elem. Res., 80, pp. 181-189, 2001.

DOI 10.18699/BGRS/SB-2020-339

The biomarkers for genetic predisposition to some connective tissue autoimmune disorders in Belarus

Hanna Yatskiu
Institute of Genetics and Cytology
Minsk, Belarus
a.yatskiv@igc.by

Elizabeth Siniauskaya
Institute of Genetics and Cytology
Minsk, Belarus
e.siniauskaya@igc.by

Nataliya Nikitchenko
Institute of Genetics and Cytology
Minsk, Belarus
N.Nikitchenko@igc.by

Natalia Dostanko
Belarusian State Medical University
Minsk, Belarus
inill2@bsmu.by

Victor Yagur
Belarusian State Medical University
Minsk, Belarus
yagur1@tut.by

Alexandr Sukalo
Belarusian State Medical University
Minsk, Belarus
childill1@bsmu.by

Tatyana Kuzhir
Institute of Genetics and Cytology
Minsk, Belarus
t.kuzhir@igc.by

Roza Goncharova
Institute of Genetics and Cytology
Minsk, Belarus
R.Goncharova@igc.by

***Abstract* — Autoimmune disorders (ADs), such as rheumatoid arthritis (RA), juvenile idiopathic arthritis (JIA), systemic lupus erythematosus (SLE) complicated by lupus nephritis (LN) are the common and socially significant multifactorial diseases. The present study aims to highlight the contribution of certain gene variants involved in the inflammatory and autoimmune responses to predisposition to ADs and identify, if possible, the overlapping biomarkers. The study engaged the patients diagnosed with JIA, RA, SLE and LN as compared to controls genotyped for some SNPs at the *IL6*, *IL6R* and *STAT4* loci using PCR-RFLP or real time PCR. Results have demonstrated the association between the *IL6* rs1800795 SNP and JIA/RA in the Belarusian population, as well as between the *STAT4* rs7574865 and RA/SLE in adults, and SLE + LN in children suggesting its interference in the same pathogenic pathways of different ADs. The results show the association of the *IL6* rs1800795 SNP with JIA and RA in Belarus. Evidence has been obtained that at least the STAT4 gene shares its risk potential between autoimmune diseases such as RA and SLE.**

Keywords* — *gene polymorphisms, IL6, IL6R and STAT4 genes, rheumatoid arthritis, juvenile idiopathic arthritis, systemic lupus erythematosus

Background

Connective tissue autoimmune disorders (AD), such as rheumatoid arthritis (RA), juvenile idiopathic arthritis (JIA) and systemic lupus erythematosus (SLE) belong to common and socially significant diseases with serious consequences for patients in the form of cartilage and bone destruction, joint ankylosis in RA and JIA, and vascular injuries in SLE often complicated by kidney pathology, i.e., lupus nephritis (LN). According to current estimations, their prevalence varies from 5 to 8%, and continues to increase steadily [1]. They arise from innate and adaptive immunity mistakes leading to self-damaging immune response and develop due to interplay between environmental triggers and genetic predisposition [2]; the latter is in the focus of majority of modern investigations. Among genetic factors, HLA locus and non-HLA genes involved in immune and inflammatory responses are carefully studied in order to reveal molecular markers indicating increased susceptibility of their carriers to such diseases. In the past decade using genome-wide association studies, numerous AD-associated alleles (SNPs) has been established, most of the polymorphisms being located in non-coding regions of genome, which regulate gene expression modulating immune responses.

Studies concerning a genetic nature of different ADs face a number of problems, e.g., ethnic disparities in the implementation of the risk alleles [3]. Therefore, there is a need to examine the associations between known AD-associated gene variants and independent autoimmune diseases in the certain ethnogeographic conditions. The other problem is due to a minor contribution of individual SNPs to risk of developing ADs, when odds ratio (OR) does not exceed 1.5. Presumably, risk of developing the disease may be increased by multiple genes. Enhancement of the effects of individual SNPs during gene-gene interaction has been recently discussed using JIA in Belarus as an example [4], as well as reviewed in [5]. Herein, the results of a comparative study of the effects of several SNPs in three autoimmune diseases (RA, JIA and SLE) in the Belarusian population will be presented.

Materials and Methods

1) Study groups were recruited among patients of the 2nd City Children's Clinical Hospital (Minsk) and the Republican Scientific and Practical Center of Transfusiology and Medical Biotechnology (Minsk). Participants of the study (or their relatives) signed informed consent in accordance with international standards for human research. Four groups were formed among children: patients diagnosed with JIA (1), SLE + LN (2); patients with articular syndrome (AS) other than JIA etiology (3), and patients without any inflammatory and autoimmune diseases as a control (4). Three groups were formed among adults: patients diagnosed with RA (1) and SLE (2), and clinically healthy donors as controls (3). DNA was extracted from blood samples (542 from children and 597 from adults) using ordinary phenol-chloroform method and genotyped for the *IL6* rs1800795, *IL6R* rs 4845618 and rs2228145, as well as the *STAT4* rs7574865 loci using PCR-RFLP or real time PCR. The association between these SNPs

and ADs was estimated using odds ratio and 95% confidential interval, i.e., OR [95%CI].

Results and Discussion

The main results obtained in the children's cohort indicate the association of JIA with the polymorphic variant rs1800795 of the *IL6* gene in terms of both the C allele frequency (OR [95% CI] = 1.43 [1.07–1.94]; p = 0.018) and the homozygous CC genotype frequency (OR [95% CI] = 2.39 [1.48–3.86], p = 0.0008). At the same time, minor alleles in the polymorphic loci of the *IL6R* gene were found to be associated with AS of other than JIA origin. The risk potential was manifested by CC homozygotes (OR = 2.07 [1.13–3.78], p = 0.00018) and the C allele (OR = 1.63 [1.20–2.22], p = 0.0018) at the rs2228145 locus, as well as TT homozygotes (OR = 1.91 [1.19–3.05], p = 0.02) and the T allele (OR = 1.41 [1.05–1.91], p = 0.023) at the rs4845618 locus. As to the *STAT4* rs7574865 variant, we failed in detection of statistically significant differences between the JIA, AS and control groups. However, it was associated with developing SLE and LN in girls (OR = 2.54 [1.00–6.42] p = 0.05 for the genotypes GT + TT and OR = 2.14 [1.07-4.27] p = 0.04 for the T allele, respectively).

In the adult cohort, the CC genotype at the *IL-6* rs1800795 locus was associated with RA (OR = 1.52 [1.02–2.27]; p = 0.0456). The studied polymorphic variants of the *IL6R* gene did not affect the sensitivity of the Belarusian population to RA and SLE. Statistically significant differences were found between the frequencies of the TT genotype and the minor T allele at the *STAT4* rs7574865 locus in the RA and SLE patients as compared to controls, as well as in the SLE group versus RA. It was shown that these SNPs can serve as biomarkers of a predisposition to RA (OR = 2.36 [1.04–5.38] p = 0.04 for the TT genotype), but they even more effectively predict the risk of SLE (OR = 7, 61 [2.35–24.64] p = 0.0007 for the TT genotype; OR = 2.83 [1.53–5.22] p <0.0009 for the T allele).

Conclusion

The results show the association of the *IL6* rs1800795 SNP with JIA and RA in Belarus. The *STAT4* rs7574865 SNP increases susceptibility of adults to RA and SLE, and is associated with SLE and LN in children sharing risk effects between different autoimmune diseases.

References

[1] G.S. Cooper, M. L.K. Bynum and E.C. Somers “Recent Insights in the Epidemiology of Autoimmune Diseases: Improved Prevalence Estimates and Understanding of Clustering of Diseases”, J Autoimmun., vol. 33, pp. 197–207, Nov-Dec 2009.

[2] M. D. Rosenblum, K.A. Remedios, and A.K. Abbas “Mechanisms of human autoimmunity”, J Clin Invest., vol. 125, pp. 2228–2233, April 2015.

[3] P.S Ramos, A.M Shedlock, and C.l. D. Langefeld “Genetics of autoimmune diseases: insights from population genetics”, J Hum Genet., vol. 60, pp. 657–664, 2015.

[4] H.A. Yatskiu et al. “Genetic susceptibility to juvenile idiopathic arthritis in the belarusian population: gene-gene interactions analysis”, Ecological genetics, vol. 17, pp. 65–76, 2019 [А.А. Яцкив и др. “Генетическая предрасположенность к ювенильному идиопатическому артриту в белорусской популяции: анализ межгенных взаимодействий”, Экологическая генетика. № 4, С. 63–74, 2019].

[5] T.D. Kuzhir “Polygenic nature of rheumatoid arthritis”, Ecological genetics, vol. 17, pp. 77–90, 2019 [Т.Д. Кужир “Полигенная природа ревматоидного артрита ”, Экологическая генетика. № 4, С. 75–88, 2019].

DOI 10.18699/BGRS/SB-2020-340

Accumulation of oxidative hepatobiliary lesions during experimental opisthorchiasis

Oxana Zaparina
ICG SB RAS, Novosibirsk, Russia
zp.oksana.93@gmail.com

Anna Kovner
ICG SB RAS, Novosibirsk, Russia
kovner@bionet.nsc.ru

Maria Pakharukova
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
pakharukova@bionet.nsc.ru

Viacheslav Mordvinov
ICG SB RAS, Novosibirsk, Russia
mordvin@bionet.nsc.ru

***Abstract* — Opisthorchiasis caused by the liver fluke Opisthorchis felineus is widespread on Russian territory, Eastern Europe and Asia. This disease leads to liver dysfunctions: cholecystitis, cholestasis, periductal fibrosis, chronic inflammation, and precancerous changes in the epithelium of the bile ducts (biliary neoplasia). The mechanisms of these processes are not study. In the result of study, we showed time- dependent oxidative lesions accumulation and hepatoprotective effect of various antioxidants by reducing inflammation and biliary neoplasia.**

Keywords — opisthorchiasis, inflammation, reactive oxygen species, antioxidants, liver fluke

Motivation and Aim

Opisthorchiasis caused by the liver fluke *Opisthorchis felineus* in the hepatobiliary system of humans and fish-eating mammals is widespread on Russian territory, Eastern Europe and Asia. This disease leads to liver dysfunctions: cholecystitis, cholestasis, periductal fibrosis, chronic inflammation, and precancerous changes in the epithelium of the bile ducts (biliary neoplasia). The causative agents of these diseases, liver flukes *O. viverrini* and *C. sinensis*, are officially recognized as Group 1 biological carcinogens and are classified as the main risk factors for cholangiocarcinoma [1]. Recently specific oxysterol-like molecules were found to be secreted from the flukes. These molecules might have genotoxic and pro-oxidative properties, increase production of reactive oxygen species, oxidative damage to DNA and precancerous lesions in host cells [2]. The mechanisms of development of this disease have not been studied; however, increased production of reactive oxygen species probably plays an important role in the pathogenesis of opisthorchiasis [3].

To study the possible role of oxidative stress in the pathogenesis of opisthorchiasis we assessed lipid peroxidation and inflammation markers accumulation, DNA damage in infected hamsters from 1 to 18 months postinfection and the effect of antioxidants SKQ1 (10- (6'-Plastoquinonyl) decyltriphenylphosphonium) and resveratrol.

Methods

Experimental opisthorchiasis in vivo on golden hamsters *M.auratus*. Using immunohistochemistry and immunoassays, we showed accumulation of inflammatory and oxidative damage markers. Pathological changes in the liver, including inflammation, biliary neoplasia, metaplasia, proliferation of the epithelium of the bile ducts, were assessed using histological analysis. Using real time PCR we measured mRNA expresson of the inflammation and fibrogenesis related genes.

Results

We showed time-dependent accumulation of inflammatory, lipid peroxidation (MDA, HNE) and oxidative DNA damage (8-OHdG) markers. Moreover, liver histopathology, including inflammation, biliary neoplasia, periductal fibrosis were assessed.

The histopathology of liver i.e. biliary neoplasia, lipid peroxidation markers and serum biochemistry liver damage markers were significantly attenuated by SKQ1 and resveratrol treatment.

Conclusion

Thus, reactive oxygen species play a pivotal role in the pathogenesis of opisthorchiasis, probably mediating the processes of inflammation and biliary neoplasia. The histopathology of liver i.e. biliary neoplasia and lipid peroxidation markers, were significantly attenuated by SKQ1and resveratrol treatment. Thus, SKQ1 and resveratrol have a hepatoprotective effect.

ACKNOWLEDGMENT

Supported by the RFBR (19-34-90060-A).

REFERENCES

[1] Sripa B., Kaewkes S., Sithithaworn P., Mairiang E., Laha T., Smout M., Loukas A. (2007) Liver fluke induces cholangiocarcinoma. PLoS Med. 4(7): 1148-1155.

[2] Gouveia MJ, Pakharukova MY, Laha T, Sripa B, Maksimova GA., Rinaldi G, Brindley PJ., Mordvinov VA., Amaro T, Santos LL, Correia da Costa JM, Vale N. (2017) Infection with Opisthorchis felineus induces intraepithelial neoplasia of the biliary tract in a rodent model. Carcinogenesis. 38(9): 929–937.

[3] [3] Pakharukova MY, Correia da Costa JM, Mordvinov VA. (2019) The liver fluke Opisthorchis felineus as a group III or group I Carcinogen. 4 Open. 2(23): 1-10.

DOI 10.18699/BGRS/SB-2020-341

Identifying cell subtypes in the carotid atherosclerotic plaques by the deconvolution MiRNA analysis

Aleksei Zarubin
RIMG TNRMC, Tomsk, Russia
aleksei.zarubin@medgenetics.ru

Anton Markov
RIMG TNRMC, Tomsk, Russia
anton.markov@medgenetics.ru

Aleksei Sleptcov
RIMG TNRMC, Tomsk, Russia
alexei.sleptcov@medgenetics.ru

Maria Nazarenko
RIMG TNRMC, Tomsk, Russia
maria.nazarenko@medgenetics.ru

***Abstract* — MiRNAs are small non-coding RNAs that have a specific cell-type-dependent expression and may be influenced by diseases, such as atherosclerosis. In this study, we demonstrate possibility of determining differentially expressed miRNA genes in cell subtypes by using a cell deconvolution approach.**

Keywords — atherosclerosis, miRNA, deconvolution

Motivation and Aim

There has been reported that miRNAs participate in the atherosclerotic process actively. The cell- and tissue-specific manner expression of miRNAs results in a challenging identification of the shifts in expression levels through disease, especially in atherosclerosis. Whether the shifts are caused by a twist of the cell-type ratios or by the intracellular dysregulations are still unclear. We suggest that the cell-type deconvolution algorithms can be helpful to assess the cellular composition of bulk tissue samples [1].

To assess the potential of the deconvolution approaches to simultaneously decomposition onto cell types and detecting expression changes in it on the example miRNA sequencing data of atherosclerosis tissues.

Methods

Both atherosclerotic-affected and intact tissue of carotid artery biopsy specimens obtained from 5 patients simultaneously. Preparation of miRNA sequencing libraries was carried out by the NEBNext Multiplex Small RNA Library Prep Set for Illumina following with sequencing on the HiSeq 1500 system (1x50bp mode). Data processing and evaluation of miRNA expression were performed using miARma-Seq software package. The pre-made database was used as well (https://github.com/alekseizarubin/miRNA_reference_deconvolution_arteries). The deconvolution and statistical analysis was executed in the R software (edgeR, DeconRNASeq, TOAST).

Results

As a result of the deconvolution, a statistically significant difference of the predicted fraction of macrophages (MF), smooth muscle cells and fibroblasts (SMC_Fib), lymphocytes (T_NK and B cells), and neutrophils (NF) was found between atherosclerotic-affected and intact carotid arteries. MiRNAs were differentially expressed in specific cell types (see Table 1). Unfortunately, the identification of expression shifts of miRNAs in endothelial cells (EC) and MF is unfeasible owing to an absence of cell subtypes in the intact tissue specimens.

Thus, we demonstrated possibilities for the utilization of miRNA sequencing data into cell deconvolution analysis to reconstruct cell subtypes proportion and the identification of differentially expressed miRNA genes in them.

TABLE 1. THE PREDICTED FRACTION OF CELL TYPES IN SAMPLES AND DIFFERENTIALLY EXPRESSED MIRNAS

	Cell subtypes					
Sample ID	EC	SMC Fib	MF	T_NK	NF	B
iv24	0.0	0.37	0.0	0.00	0.32	0.31
iv27	0.0	0.52	0.0	0.00	0.30	0.18
iv31	0.0	0.39	0.0	0.00	0.32	0.28
iv33	0.0	0.46	0.0	0.01	0.32	0.21
iv34	0.0	0.43	0.0	0.00	0.35	0.23
ap24	0.0	0.45	0.03	0.03	0.31	0.18
ap27	0.0	0.50	0.05	0.08	0.27	0.10
ap31	0.0	0.47	0.02	0.06	0.32	0.13
ap33	0.0	0.54	0.02	0.11	0.27	0.07
ap34	0.0	0.52	0.01	0.06	0.30	0.11
Differentially expressed miRNAs	NA	miR-132 miR-18a miR-323b miR-410 miR-126 miR-10b	NA	miR-654 miR-410 miR-323b miR-504 miR-24-1 miR-147b miR-10a miR-204	miR-132 miR-199b miR-10b miR-126	miR-199b

iv# - intact vessel; ap# - atherosclerotic plaque; NA – not available.

REFERENCES

[1] Li Z. & Wu H. (2019). TOAST: improving reference-free cell composition estimation by cross-cell type differential analysis. Genome biology, 20 (1), 190.

DOI 10.18699/BGRS/SB-2020-342

The role of microRNA-370 in steroid-resistant focal segmental glomerulosclerosis

Sepideh Zununi Vahed
Kidney Research Center
Tabriz University of Medical Sciences
Tabriz, Iran
sepide.zununi@gmail.com

Seyedeh Mina Hejazian
Kidney Research Center
Tabriz University of Medical Sciences
Tabriz, Iran
smhjz.biotech@gmail.com

Mohammadreza Ardalan
Kidney Research Center
Tabriz University of Medical Sciences
Tabriz, Iran
ardalan34@yahoo.com

***Abstract* — Dysregulated levels of microRNAs may be involved in the pathogenesis and response to steroid therapy in patients with focal segmental glomerulosclerosis (FSGS). An in-silico analysis indicated that miR-370 is contributed to biological adhesion, cation transport, regulation of ion transport, and metal ion transport in FSGS.**

Keywords* — *Nephrotic syndrome, FSGS, MicroRNAs, Steroid response

Motivation and Aim

Focal segmental glomerulosclerosis (FSGS) is the most common type of glomerular disease in adults [1]. Steroids are considered first-line therapeutic opportunities in these patients [2]. Patients with steroid-resistant FSGS without any improvement in their proteinuria are facing a gradual decline of kidney function and are destined to end-stage renal diseases (ESRD) [3]. Numerous lines of evidence suggested that microRNAs play a role in the modification of steroid responsiveness, consequently, they may act as potential molecular markers for steroid response.

An in-silico analysis was performed to understand the signaling pathways and biological procedures that may be regulated by miR-370 in steroid-resistant FSGS.

Methods

Gene expression data sets were downloaded from GEO datasets to find the significant differentially expressed mRNA. The GEO2R online tool was used for determining the differential expression of genes. miRwalk online tool was utilized for predicting target mRNAs of the miRNA. Twelve databases were selected for the predicting (miRWalk, Microt4, miRanda, mirbridge, miRDB, miRMap, miRNAMap, Pictar2, PITA, RNA22, RNAhybrid, and Targetscan). The target genes were chosen which exist in at least 3 aforementioned databases. The overlapping genes between the GEO datasets differentially expressed genes and predictive miRNA target genes were indicated by generating Venn diagram by R v3.5 programming. The shared gene sets were enriched by Gene Ontology (GO) (http://software.broadinstitute.org/gsea/msigdb/annotate.jsp). The GO data was visualized with R programming.

Results

Five newly updated datasets (GSE125779 (n=353), GSE129973 (n=176), GSE108113 (n=1531), GSE104948 (n=106), and GSE104066 (n=693) were downloaded for FSGS. Because of the similarity between the data, GSE108113 and GSE104066 were chosen for further investigations. GO annotation demonstrated that miR-370, mostly contributes to biological adhesion, cation transport, regulation of ion transport, cation transmembrane transport, and metal ion transport.

Acknowledgment

Supported by the National Institute for Medical Research Development (Grant No.: 978622).

References

[1] Ka Hull RP, Goldsmith DJ. (2008) Nephrotic syndrome in adults. Bmj. 2008;336(7654):1185-9

[2] L.A. Greenbaum, R. Benndorf, W.E. Smoyer (2012) Childhood nephrotic syndrome—current and future therapies, Nature Reviews Nephrology, 8, 445.

[3] J.H. Ehrich, C. Geerlings, M. Zivicnjak, D. Franke, H. Geerlings, J. Gellermann (2007) Steroid-resistant idiopathic childhood nephrosis: overdiagnosed and undertreated, Nephrology Dialysis Transplantation, 22 2183-2193.

DOI 10.18699/BGRS/SB-2020-343

Expression of glycogen synthase kinase 3β in nephrotic syndrome

Sepideh Zununi Vahed
Kidney Research Center
Tabriz University of Medical Sciences
Tabriz, Iran
sepide.zununi@gmail.com

Elham Ahmadian
Kidney Research Center
Tabriz University of Medical Sciences
Tabriz, Iran
ahmadian.elham@yahoo.com

Mohammadreza Ardalan
Kidney Research Center
Tabriz University of Medical Sciences
Tabriz, Iran
ardalan34@yahoo.com

***Abstract* — Aberrant expression of glycogen synthase kinase-3 (GSK-3β) in kidney cells has a harmful role in podocyte injury. In this work, we found that dysregulated levels of GSK-3β may be involved in the pathogenesis of nephrotic syndrome, kidney disease with proteinuria, with different etiology.**

Keywords* — *Nephrotic syndrome, Proteinuria, Membranous glomerulonephritis, Focal Segmental Glomerulosclerosis, GSK-3β.

Motivation and Aim

Glycogen synthase kinase-3 (GSK-3) is a conserved multi-functional serine/threonine kinase that regulates various physiological processes, including gene expression, cell signaling [1], cellular proliferation, apoptosis, and intracellular communication. A detailed understanding of the GSK-3β function and its expression in several pathological conditions will help the clinic manage different kidney diseases.

Given the role of GSK-3β in the podocytes injury [2,3], we evaluate its expression levels in PBMCs samples of patients with NS.

Methods

The expression levels of GSK-3β was evaluated in peripheral blood mononuclear cells (PBMCs) of cases with the most common types of nephrotic syndrome (NS); MGN (membranous glomerulonephritis) and FSGS (focal segmental glomerulosclerosis) using real-time PCR. Sixty cases (30 FSGS and 30 MGN) were included based on the strict criteria. The results compared with healthy controls (n=24). The receiver operating characteristic (ROC) curve analysis was used for evaluating the potential of GSK-3β in discriminating cases from controls.

Results

There were statistically significant increases in GSK-3β expression level in NS (P=0.002) and FSGS (P<0.001) groups when compared to controls; however, it was not significant in MGN group (P= 0.137). GSK-3β level was also significantly higher in FSGS group in comparison to MGN group. ROC curve analysis approved a diagnostic power of GSK-3β in discriminating patients from healthy controls (AUC: 0.72, P= 0.002) with high sensitivity and specificity.

Acknowledgment

Supported by the Kidney Research Center at Tabriz University of Medical Sciences (Grant No. 60546 and 62640).

References

[1] Maurer U. et al. (2014) GSK-3 - at the crossroads of cell death and survival. Journal of Cell Science. 127, (Pt 7), 1369-78.

[2] Li C. et al. (2016) The beta isoform of GSK3 mediates podocyte autonomous injury in proteinuric glomerulopathy. The Journal of Pathology. 239, (1), 23-35.

[3] Xu W., Ge Y., Liu, Z. (2014) Glycogen synthase kinase 3beta dictates podocyte motility and focal adhesion turnover by modulating paxillin activity: implications for the protective effect of low-dose lithium in podocytopathy. The American Journal of Pathology. 184, (10), 2742-2756.

Symposium
Systems biology of DNA repair processes and programmed cell death

DOI 10.18699/BGRS/SB-2020-344

Activity of SNP variants of human uracil-DNA glycosylases SMUG1 and MBD4

Irina V. Alekseeva
Institute of Chemical Biology and Fundamental Medicine
Novosibirsk, Russia
Irina.Alekseeva@niboch.nsc.ru

Artemiy S. Bakman
Institute of Chemical Biology and Fundamental Medicine
Novosibirsk, Russia
art-bakman@yandex.ru

Olga S. Fedorova
Institute of Chemical Biology and Fundamental Medicine
Novosibirsk, Russia
fedorova@niboch.nsc.ru

Nikita A. Kuznetsov
Institute of Chemical Biology and Fundamental Medicine
Novosibirsk, Russia
Nikita.Kuznetsov@niboch.nsc.ru

Abstract — **Human uracil-DNA glycosylases SMUG1 and MBD4 are crucial enzymes of the base excision repair pathway, responsible for uracil repair. It has been hypothesized that single-nucleotide polymorphic variants of SMUG1 and MBD4 might be responsible for an increased risk of some types of cancer. In the present work, analysis of SNPs of these uracil-DNA glycosylases was performed to select the set of variants having substitutions of amino acid residues on the surface of the enzyme globule and in the DNA-binding site, thereby affecting protein–protein interactions or the catalytic reaction, respectively. Conformational dynamics and catalytic activities of four SMUG1 and four MBD4 SNP variants were examined. The conformational changes in the molecules of enzymes and in a DNA substrate were recorded as fluorescence changes of Trp and 2-aminopurine residues as well as FRET-signal.**

Keywords — DNA repair, human uracil–DNA glycosylase, stopped-flow enzyme kinetics, fluorescence

Base excision repair (BER) is a key genome maintenance pathway that removes endogenously damaged DNA bases that arise in cells at very high levels on a daily basis. Amino acid substitutions in BER enzymes associated with single-nucleotide polymorphisms (SNPs) are widespread in the human population. The relation between SNP variants of BER proteins and etiology of some human diseases requires further investigation. It is known that some polymorphisms of the human AP-endonuclease or 8-oxoguanine-DNA glycosylase genes lead to decrease in the enzymatic activity. Such variants of BER enzymes in the human population can be associated with an increased risk of diseases [1,2]. The association of SNPs of BER genes with some diseases was recently discussed in a number of reviews [3,4]. It should be noted that a decrease in the functional activity of individual BER enzymes and disruption of coordination between them or with protein factors such as XRCC1 (which acts as a scaffold protein in the BER pathway) can have severe negative consequences for the human body [5–7].

Uracil in DNA can be introduced via two mechanisms, deamination of cytosine and misincorporation of dUMP during replication. Deamination of cytosine has been calculated from measured deamination rates to occur at a rate of 100–500 per human cell/day to yield mutagenic U/G mispairs. Uracil may also appear as a consequence of misincorporation of dUMP instead of dTMP during replication, resulting in a U/A base pair. Human single-strand selective monofunctional uracil–DNA glycosylase SMUG1, thymine–DNA glycosylase (TDG), uracil–DNA glycosylase (UNG), and methyl-CpG-binding domain (MBD4) have high affinity to uracil-containing DNA and catalyze a cleavage of the N-glycosidic bond involving a target base. They also possess individual substrate specificity to a single-stranded DNA substrate, T/G or U/G mismatches, and to some oxidized bases such as 5-hydroxymethyluracil, 5-formyluracil, and 5-hydroxyuracil. Interestingly to note, UNG, TDG, and SMUG1 belong to different families (I, II, and III, respectively) of the large uracil–DNA glycosylase structural superfamily [8], whereas MBD4 belongs to structural superfamily HhH (helix–hairpin–helix) [9] of DNA-binding proteins. Given that human has four uracil–DNA glycosylases, indicating importance of uracil repair, we have conducted a systematic analysis of single-nucleotide polymorphism (SNP) variants of two uracil-DNA glycosylases SMUG1 and MBD4. Our aim was to determine the influence of known SNPs on the conformation and activity of human SMUG1 and MBD4. Using the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/), exonic polymorphic variants of the human SMUG1 and MBD4 were selected. Subsequent analysis of potential importance of an SNP was based on a principle of a maximal change in the chemical nature of the side group of the amino acid encoded by a corresponding nucleotide triplet. Spatial location of the amino acid residue close to the DNA-binding site and/or active site of the enzyme was also taken into account. Four SMUG1 (G90C, P240H, N244S and N248Y) and four MBD4 (S470L, G507S, R512W and H557D) SNP variants were chosen for stopped-flow analyses of conformational dynamics and kinetics of interaction with DNA substrates. The activity of SNP variants was studied by direct PAGE analysis of the kinetics of accumulation of products. The conformational transitions in the molecules of WT enzymes and its SNP variants in the course of interaction with DNA-substrate were monitored as changes in the intrinsic Trp fluorescence. To determine the effect of amino acid substitutions on the conformational changes in the DNA substrate, the changes in fluorescence intensity of aPu residues in the model substrate were monitored as well. A FAM/BHQ1-labeled substrate was also subjected to FRET analysis of DNA distortion processes and of the cleavage reaction. Obtained results made it possible to determine the kinetic mechanism underlying the interactions of the SNP variants with DNA substrates, to calculate the rate constants of the elementary stages, and to identify the stages of the process affected by mutation.

Acknowledgment

This study was supported partially by the Russian Ministry of Science and Education (project #AAAA-A17-117020210022-4). A pre-steady-state kinetic analysis of the enzyme interaction with DNA substrates was supported by a grant of the Russian Science No. 16-14-10038.

REFERENCES

[1] Karahalil, B.; Bohr, V. A.; Wilson, D. M. Impact of DNA polymorphisms in key DNA base excision repair proteins on cancer risk. Hum. Exp. Toxicol. 2012, doi:10.1177/0960327112444476.

[2] Sweasy, J. B.; Lang, T. M.; DiMaio, D. Is base excision repair a tumor suppressor mechanism? Cell Cycle 2006, 5, 250–259, doi:Doi 10.4161/Cc.5.3.2414.

[3] Wallace, S. S.; Murphy, D. L.; Sweasy, J. B. Base excision repair and cancer. Cancer Lett. 2012.

[4] Marsden, C. G.; Dragon, J. A.; Wallace, S. S.; Sweasy, J. B. Base Excision Repair Variants in Cancer. Methods Enzym. 2017, 591, 119–157, doi:10.1016/bs.mie.2017.03.003.

[5] Abbotts, R.; Wilson, D. M. Coordination of DNA single strand break repair. Free Radic. Biol. Med. 2017, 107, 228–244, doi:10.1016/j.freeradbiomed.2016.11.039.

[6] Kwiatkowski, D.; Czarny, P.; Galecki, P.; Bachurska, A.; Talarowska, M.; Orzechowska, A.; Bobinska, K.; Bielecka-Kowalska, A.; Pietras, T.; Szemraj, J.; Maes, M.; Sliwinski, T. Variants of Base Excision Repair Genes MUTYH , PARP1 and XRCC1 in Alzheimer's Disease Risk. Neuropsychobiology 2015, 71, 176–186, doi:10.1159/000381985.

[7] Czarny, P.; Kwiatkowski, D.; Toma, M.; Kubiak, J.; Sliwinska, A.; Talarowska, M.; Szemraj, J.; Maes, M.; Galecki, P.; Sliwinski, T. Impact of Single Nucleotide Polymorphisms of Base Excision Repair Genes on DNA Damage and Efficiency of DNA Repair in Recurrent Depression Disorder. Mol Neurobiol 2017, 54, 4150–4159, doi:10.1007/s12035-016-9971-6.

[8] Schormann, N.; Ricciardi, R.; Chattopadhyay, D. Uracil-DNA glycosylases-Structural and functional perspectives on an essential family of DNA repair enzymes. Protein Sci. 2014, 23, 1667–1685, doi:10.1002/pro.2554.

[9] Zhang, W.; Liu, Z.; Crombet, L.; Amaya, M. F.; Liu, Y.; Zhang, X.; Kuang, W.; Ma, P.; Niu, L.; Qi, C. Crystal structure of the mismatch-specific thymine glycosylase domain of human methyl-CpG-binding protein MBD4. Biochem. Biophys. Res. Commun. 2011, 412, 425–428, doi:10.1016/j.bbrc.2011.07.091.

DOI 10.18699/BGRS/SB-2020-345

Activity of human AP-endonuclease APE1 on DNA- and RNA-substrates forming non-canonical structures

Anastasiia T. Davletgildeeva
Institute of Chemical Biology and Fundamental Medicine
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
davleta94@gmail.com

Alexandra A. Kuznetsova
Institute of Chemical Biology and Fundamental Medicine
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
sandra-k@niboch.nsc.ru

Olga S. Fedorova
Institute of Chemical Biology and Fundamental Medicine
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
fedorova@niboch.nsc.ru

Nikita A. Kuznetsov
Institute of Chemical Biology and Fundamental Medicine
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
Nikita.Kuznetsov@niboch.nsc.ru

***Abstract* — Human apurinic/apyrimidinic endonuclease APE1 is responsible for detecting and initiating the elimination of AP-sites from DNA. APE1 makes an incision of the phosphodiester bond on the 5′-side of some damaged nucleotides and possesses 3′-5′ exonuclease, 3′-phosphodiesterase, 3′-phosphatase, and RNase H activities. It remains unclear how such structurally different substrates can be recognized by the only active site of the enzyme. In the present work the mechanism of APE1 catalyzed target nucleotide cleavage was studied. As model substrates we used DNA- and RNA-oligonucleotides forming non-canonical structures such as quadruplexes, loops and hairpins. Obtained data demonstrate the ability of APE1 to excise AP-site containing DNA as well as undamaged RNA in different non-canonical structures on the basis of substrate topology.**

Keywords* — *AP endonuclease, non-canonical structures, fluorescence

Human apurinic/apyrimidinic (AP) endonuclease APE1 catalyses the hydrolysis of phosphodiester bonds on the 5´-side of an AP-site and of some damaged nucleotides such as etheno-derivatives of DNA bases [1,2], bulky photoproducts [3], benzene-derived DNA adducts [4], α-anomers of 2´-deoxynucleosides [5], oxidatively damaged pyrimidines [6] and 2´-deoxyuridine [7]. APE1 also possesses other activities, such as 3´-5´ exonuclease [8,9], 3´-phosphodiesterase, 3´-phosphatase [10], and RNase H [11]. In spite of numerous studies on APE1, the molecular origin of its broad substrate specificity and the mechanism of discrimination between modified bases and/or nucleotides as well as undamaged DNA and RNA bases are not yet clear. Despite the successful characterization of crystal structures of human APE1 bound to a DNA substrate containing an F-site (a stable analogue of a natural AP-site, lacking the hydroxyl group on the C1-atom of ribose) or a cleaved DNA product, at present, there is no published structure of an APE1 complex with other DNA or RNA substrates.

The substrate specificity to different bases may be associated with different distortions of local DNA or RNA structure around the target base, influencing the efficiency of forming the catalytically active conformation. Another basis for substrate recognition may be related to differences in the efficiency of target base eversion from substrate into the active site of enzyme, which probably is dependent on the nature of the base and/or on stability of the target base pair in DNA or RNA.

The main purpose of our study was to elucidate the key steps of the mechanism behind protein–substrate interaction that ensure specific recognition of target nucleotide in various DNA- and RNA-substrates forming non-canonical structures such as quadruplexes, and various loop- and hairpin-containing structures, which can facilitate nucleotide eversion stage. The structures of these substrates were proved by CD-spectroscopy. Analysis of product accumulation using gel-electrophoresis has shown that APE1 excises all of the DNA substrates used, i.e. DNA quadruplexes, containing the single F-site in core or loop regions, bulged DNA duplexes, containing the F-site, as well as undamaged RNA hairpin-structures. Obtained data demonstrate the ability of APE1 to excise AP-site containing DNA as well as undamaged RNA in different non-canonical structures. Taken together, our data support the idea that the mechanism of substrate specificity of APE1 is based on the ability of the target nucleotide to flip out of the DNA duplex into the active site owing to an enzyme-induced substrate distortion and bending.

ACKNOWLEDGMENT

This work was supported by the Russian Science Foundation No. 19-74-10034.

REFERENCES

[1] P. Prorok, C. Saint-Pierre, D. Gasparutto, O.S. Fedorova, A.A. Ishchenko, H. Leh, M. Buckle, B. Tudek, M. Saparbaev, Highly mutagenic exocyclic DNA adducts are substrates for the human nucleotide incision repair pathway, PLoS One. 7 (2012) e51776. doi:10.1371/journal.pone.0051776.

[2] P.P. Christov, S. Banerjee, M.P. Stone, C.J. Rizzo, Selective Incision of the alpha-N-Methyl-Formamidopyrimidine Anomer by Escherichia coli Endonuclease IV, J Nucleic Acids. 2010 (2010) 850234. doi:10.4061/2010/850234.

[3] M.G. Vrouwe, A. Pines, R.M. Overmeer, K. Hanada, L.H. Mullenders, UV-induced photolesions elicit ATR-kinase-dependent signaling in non-cycling cells through nucleotide excision repair-dependent and -independent pathways, J Cell Sci. 124 (2011) 435–446. doi:10.1242/jcs.075325.

[4] A.B. Guliaev, B. Hang, B. Singer, Structural insights by molecular dynamics simulations into specificity of the major human AP endonuclease toward the benzene-derived DNA adduct, pBQ-C, Nucleic Acids Res. 32 (2004) 2844–2852. doi:10.1093/nar/gkh594.

[5] L. Gros, A.A. Ishchenko, H. Ide, R.H. Elder, M.K. Saparbaev, The major human AP endonuclease (Ape1) is involved in the nucleotide incision repair pathway, Nucleic Acids Res. 32 (2004) 73–81. doi:10.1093/nar/gkh165.

[6] S. Daviet, S. Couve-Privat, L. Gros, K. Shinozuka, H. Ide, M. Saparbaev, A.A. Ishchenko, Major oxidative products of cytosine are substrates for the nucleotide incision repair pathway, DNA Repair. 6 (2007) 8–18. doi:10.1016/j.dnarep.2006.08.001.

[7] P. Prorok, D. Alili, C. Saint-Pierre, D. Gasparutto, D.O. Zharkov, A.A. Ishchenko, B. Tudek, M.K. Saparbaev, Uracil in duplex DNA is a substrate for the nucleotide incision repair pathway in human cells, Proc. Natl. Acad. Sci. U S A. 110 (2013) E3695–E3703. doi:10.1073/pnas.1305624110.

[8] K.-M. Chou, Y.-C. Cheng, The exonuclease activity of human apurinic/apyrimidinic endonuclease (APE1). Biochemical properties and inhibition by the natural dinucleotide Gp4G., J. Biol. Chem. 278 (2003) 18289–96. doi:10.1074/jbc.M212143200.

[9] A. Kuznetsova, O. Fedorova, N. Kuznetsov, Kinetic Features of 3′-5′ Exonuclease Activity of Human AP-Endonuclease APE1, Molecules. 23 (2018) 2101. doi:10.3390/molecules23092101.

[10] D.S. Chen, T. Herman, B. Demple, Two distinct human DNA diesterases that hydrolyze 3'-blocking deoxyribose fragments from oxidized DNA, Nucleic Acids Res. 19 (1991) 5907–5914. http://www.ncbi.nlm.nih.gov/pubmed/1719484.

[11] G. Barzilay, I.D. Hickson, Structure and Function of Apurinic/Apyrimidinic Endonucleases, Bioessays. 17 (1995) 713–719. doi:DOI 10.1002/bies.950170808.

DOI 10.18699/BGRS/SB-2020-346

Activity of DNA glycosylases on non-canonical DNA substrates

Evgeniia Diatlova
Novosibirsk State University, Novosibirsk, Russia
ICBFM SB RAS, Novosibirsk, Russia
jannie.lapt@gmail.com

Dmitry Zharkov
Novosibirsk State University, Novosibirsk, Russia
ICBFM SB RAS, Novosibirsk, Russia
dzharkov@niboch.nsc.ru

***Abstract* — Until recently, DNA repair was almost exclusively considered in the B-DNA context, since non-canonical DNA was not believed to be a significant target for damage in the bulk genome. A paradigm shift that followed the discoveries of functionally important roles of non-canonical DNA and the cases of apparent connection of regulation with DNA damage put DNA repair in non-canonical structures into the focus of active studies. In this work we investigated base excision repair in non-canonical DNA structures involving hairpins, quadruplexes and DNA/RNA heteroduplexes.**

Keywords* — *DNA glycosylase, non-canonical DNA, base excision repair.

Introduction

B-form is the main conformation adopted by double-stranded DNA under physiological salt, pH, and temperature conditions. Although genomic DNA in living cells mostly exists in the B-form, it became clear over the past decades that DNA exhibits significant conformational polymorphism and that many functionally important genome elements can assume alternative structures in vivo, either permanently or temporarily. Such non-B-DNA structures (thereafter referred to as "noncanonical DNA") include cruciform DNA, hairpins, triplexes, quadruplexes, intercalated motifs (i-motifs), single-stranded DNA, Z-DNA, RNA/DNA heteroduplexes, and displacement loops (D-loops) containing either RNA (R-loops) or DNA invading strands. These elements are often involved in gene activity regulation and genome stabilization, or, on the contrary, could be intermediates of deleterious processes causing genome instability.

As a chemically active molecule, DNA is in a constant flux of spontaneous and directed modification. Notably, most noncanonical DNA forms is more sensitive to DNA damage than B-DNA, Recently, some non-canonical DNA, such as quadruplexes, were shown to serve as oxidation-dependent transcriptional regulatory elements in human cells [1-3]. On the other hand, non-canonical structures, including DNA hairpins and Z-DNA, turned out to be associated with severe human diseases due to their mutation-prone nature and gene activity misregulation [4-5].

Methods

In this work we investigated the ability of mammalian OGG1, NEIL1 and NEIL2 DNA glycosylases; human and E. coli uracil-DNA glycosylase; E. coli and Lactococcus lactis formamidopyrimodine-DNA glycosylase; and E. coli MutY, Nei and Nth DNA glycosylases to process substratres other than double-stranded DNA. Oligonucleotide substrates containing one of damaged bases (oxoguanine, 5-hydroxyuracil, 5,6-dihydrouracil, uracil, AP site) were annealed with complementary DNA (forming one base bulge substrate), or RNA strand (forming DNA/RNA heteroduplexes with each of four nucleotides in the opposite position). For uracil-glycosylases we also prepared oligonucleotides containing uracil and G-quadruplex.

RESULTS

We performed a screen of activity for all enzyme–substrate pairs. Both human and E. coli uracil-DNA-glycosylases removed uracil from structures containing G-quadruplex, one base bulges and DNA/RNA heteroduplexes regardless of the opposite nucleotide. There was no activity of MutY. Enzymes of the helix–two-turn–helix (H2TH) structural superfamily (Fpg from E. coli and Lactococcus lactis, Nei E. coli and mouse NEIL1 and NEIL2) all cleaved DNA containing oxoguanine, 5-hydroxyuracil, 5,6-dihydrouracil, AP site in any context, but Nei and NEILs removed oxoguanine relatively slight. Nei also showed small activity on the substrate with a bulged uracil. Nth removed all damage except uracil and hydroxyuracil. OGG1 processed substrates with oxoguanine from bulge and when C or U were opposite in RNA.

We also obtained kinetic data of Fpg, OGG1, human and E.coli UNG on substrates with one base bulges, and also on DNA containing G-quadruplex for uracil glycosylases.

ACKNOWLEDGMENT

This work was supported by Russian Science Foundation (No. 17-14-01190).

REFERENCES

[1] Zarakowska E., Gackowski D., Foksinski M., Olinski R. (2014) Are 8-oxoguanine (8-oxoGua) and 5-hydroxymethyluracil (5-hmUra) oxidatively damaged DNA bases or transcription (epigenetic) marks? Mutat. Res., 764-765, p. 58-63.

[2] Fleming A.M., Burrows C.J. (2017) 8-Oxo-7,8-dihydroguanine, friend and foe: Epigenetic-like regulator versus initiator of mutagenesis. DNA Repair, 56, p. 75-83.

[3] Seifermann M., Epe B. (2017) Oxidatively generated base modifications in DNA: Not only carcinogenic risk factor but also regulatory mark? Free Radic. Biol. Med., 107, p. 258-265.

[4] (McMurray C.T. (2010) Mechanisms of trinucleotide repeat instability during human development. Nat. Rev. Genet., 11, p. 786-799.

[5] Zhao X., Krishnamurthy N., Burrows C.J., David S.S. (2010a) Mutation versus repair: NEIL1 removal of hydantoin lesions in single-stranded, bulge, bubble, and duplex DNA contexts. Biochemistry, 49, p. 1658-1666.

DOI 10.18699/BGRS/SB-2020-347

Sensitization mechanism of cells with TDP1 inhibitors to the action of topotecan

Nadezhda S. Dyrkheeva
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

Anastasiya A. Malakhova
Federal Research Centre, Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Konstantin P. Volcho
N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry SB RAS
Novosibirsk, Russia

IrinaV. Il'ina
N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry SB RAS
Novosibirsk, Russia

Sergey P. Medvedev
Federal Research Centre, Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Nariman F. Salakhutdinov
N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry SB RAS
Novosibirsk, Russia

Nikolay S. Li-Zhulanov
N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry SB RAS
Novosibirsk, Russia

Suren M. Zakian
Federal Research Centre, Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia

Olga I. Lavrik
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

***Abstract* — A novel structural type of monoterpene-derived tyrosyl-DNA phosphodiesterase 1 (TDP1) inhibitors have been discovered that was synthesized through preliminary isomerization of (+)-3-carene to (+)-2-carene followed by reaction with heteroaromatic aldehydes. All compounds inhibit the TDP1 enzyme at micro- and submicromolar levels with the most potent compound having an IC50 value of 0.65 μM. TDP1 is an important DNA repair enzyme and promising inhibition target for the development of new chemosensitizing agents. Panel of isogenic clones of the HEK293FT cell line knockout for TDP1 gene was created using the CRISPR-Cas9 system. Cytotoxic effect of topotecan (Tpc) and non-cytotoxic compounds of the novel structural types were investigated separately and jointly on the TDP1 gene knockout cells. For two TDP1 inhibitors synergistic effect was observed with Tpc on HEK293FT parental wild-type cells but was not found on TDP1-/- cells. Thus, it is likely that the synergistic effect with Tpc for these compounds on HEK293FT cell growth is caused by TDP1 activity inhibition. The synergy was also found for these compounds on the cancer cell lines. Such an approach in chemotherapy, with a sensitizing effect when adding a non-cytotoxic drug, can enhance the efficacy of currently used pharmaceuticals and concomitantly reduce the potential for toxic side effects.**

Keywords* — *tyrosyl-DNA phosphodiesterase 1; TDP1 gene knockout cells; synergy; topotecan; inhibitor

Introduction

There are a number of problems arising in the course of chemotherapeutic treatment for oncological diseases, namely, low efficiency of chemotherapy, the resistance of malignant tumors to drugs, numerous side effects and high toxic load on the body. The cytotoxic effect of chemotherapy is caused by DNA damage, and the ability of cells to recognize and repair DNA lesions results in resistance. The design of new compounds that inhibit DNA repair enzymes is a promising strategy for potentiating the cytotoxicity of DNA damaging agents in clinical use. The enzymes involved in DNA repair, for example, tyrosyl-DNA-phosphodiesterase 1 (TDP1), are promising therapeutic targets for the treatment of cancer [1-4]. Being one of the DNA repair enzymes, TDP1 plays a key role in the removal of DNA lesions, including those resulting from the action of DNA topoisomerase I (TOP1) inhibitors. TOP1 is an essential enzyme that regulates DNA topology by reducing DNA supercoiling. TOP1 introduces a transient single-strand break in DNA, enabling the broken strand to rotate around the TOP1-bound DNA during fundamental cellular events such as replication, transcription, and repair [5]. TOP1 inhibitors, for example, topotecan (Tpc), are important chemotherapeutic agents that stabilize the cleavage complex TOP1–DNA, thereby inducing significant DNA damage leading to cell death [6]. However, TDP1 removes lesions caused by TOP1 inhibitors thus reducing their efficacy and leading to resistance [7-9]. A promising approach to increase the selectivity and potency of TOP1 inhibitors to cancer cells is to combine them with TDP1 inhibitors. In recent years, TDP1 inhibitors of various types have been discovered [1-4] including natural product derivatives. Monoterpenes, which have a unique diverse structure and are inexpensive, available and often enantiomerically pure is an attractive renewable raw material for the development of physiologically active agents. Some of the derivatives exhibit analgesic, antiviral, neuroprotective and antitumor properties [10]. Monoterpene (+)-3-carene is one of the main components of turpentine. In the current work we have studied anti-TDP1 activity of (+)-3-carene derivatives for the first time. For this purpose, two sets of (+)-3-carene derivatives of new structural types were synthesized and tested as potent TDP1 inhibitors.

To clarify the inhibitors' action on the cellular level a panel of HEK293FT TDP1 knockout isogenic clones was created using the CRISPR-Cas9 approach. The cytotoxic effect of Tpc and the new TDP1 inhibitors was measured separately and jointly.

Results

We tested 15 compounds synthesized starting from (+)-3-carene for their TDP1 inhibitory properties by measuring their IC_{50} (half maximal inhibitory concentration) values using a real-time fluorescent oligonucleotide biosensor [11]. This biosensor demonstrated high sensitivity and specificity for detection of TDP1. The IC_{50} values for substances with heterocyclic substituents varied from 0.65 μM to 28 μM.

To examine the potential impact of TDP1 knockout on cell survival under the treatment by the synthesized compounds, we generated HEK293FT TDP1 deficient

(TDP1-/-) cell line. Using paired gRNA CRISPR-Cas9 strategy and polymerase chain reaction (PCR) screening of cell clones we obtained clones containing deletions in the first protein coding exon of TDP1 gene. After sequencing and first treating of TDP1 proficient (wild-type) and TDP1 deficient (TDP1-/-) cells with increasing concentrations of Tpc we selected one clone for subsequent work as it has homozigous deletion causing open reading frame shift in both alleles of the TDP1 gene. It was subsequently screened by biochemical assay for 3'-phosphotyrosyl cleavage activity to identify detectable TDP1 activity in the cell extract. There was no established cleavage activity in the TDP1-/- cell extract in contrast to control WT cell extract and purified TDP1. Than we analyzed, using colorimetric test, the relative amount of viable cells after treating HEK293FT WT and TDP1 deficient (TDP1-/-) cells with increasing concentrations of Tpc for 72 hours. TDP1 knockout reduced cells viability after Tpc treatment, thus HEK293FT TDP1-/- cells where more sensitive to Tpc, thus reflecting the contribution of TDP1 on cell survival.

An analysis of the intrinsic cytotoxicity for the synthesized compounds was performed on HEK293FT WT and TDP1 deficient (TDP1-/-) cell lines by colorimetric test. We tested several compounds with the highest TDP1 inhibitory activity. Cytotoxicity is absent or insignificant in the range of studied concentrations (0.08-100 μM) for all the compounds for both cell lines, which is of great importance in terms of the absence of additional side effects of cancer therapy.

TDP1 can work against the action of TOP1 poison such as Tpc due to the ability to cleave the TOP1-DNA stalled complex. Thus, TDP1 inhibition could increase the efficiency of TOP1 poisons. Next, we checked the influence of the most effective TDP1 inhibitors on the cytotoxic effect of Tpc on the HEK293FT WT and mutant cell lines. We tested HEK293FT cells viability after treatment with 11h and 12k alone and with Tpc. There is suppressed cell growth in the joint presence of TDP1 inhibitor and Tpc on WT cells and no effect of Tpc presence in TDP1 deficient cells (Fig. 1). Thus, the synergistic action of Tpc in conjunction with 11h and 12k on HEK293FT WT cells is likely due to the TDP1 inhibition making this enzyme the main target of action.

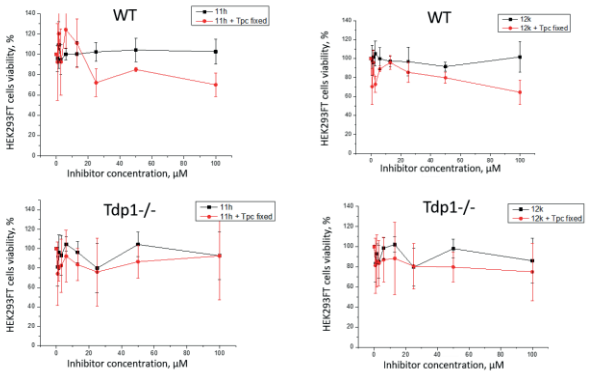


Fig. 1. TDP1 inhibitors' (11h and 12k) cytotoxicity with topotecan (Tpc) on HEK293FT WT and TDP1-/- cells.

Conclusions

We found a new class of TDP1 inhibitors with activity in the low micromolar or submicromolar concentrations, it was synthesized based on readily available monoterpene 3-carene. All the tested derivatives exhibited low intrinsic cytotoxicity when tested on various cell lines (data not shown). We created a panel of isogenic clones of HEK293FT cell line knockout for TDP1 using the CRISPR-Cas9 system and selected one clone for the experiments. We investigated the cytotoxic effect of Tpc and two TDP1 inhibitors separately and jointly on TDP1 gene knockout cells. HEK293FT TDP1-/- cells were more sensitive to Tpc compared to WT cells. For both tested compounds we observed cell growth suppression in the presence of Tpc only for WT cells but not for the TDP1 knockout cells. Thus, it is likely that the synergistic effect with Tpc on HEK293FT cell growth is only caused by TDP1 activity inhibition by these two new compounds minimizing any off-target effects. Two most promising compounds with the best combination of TDP1 inhibition exhibited IC_{50} values of 0.75 μM and 1.20 μM and demonstrate low cytotoxicity at concentrations up to 100 μM and enhancement of the Tpc efficacy on the cells. Thus, these new inhibitors are promising candidates for adjuvant therapy, mainly due to the absence of additional toxic load.

Acknowledgment

The work was supported by RFBR (project no. 19-415-540002).

References

[1] Huang, S.N.; Pommier Y., Marchand C. Tyrosyl-DNA Phosphodiesterase 1 (Tdp1) inhibitors. Expert Opin. Ther. Pat. 2011, 21, 1285–1292.

[2] Laev, S.S.; Salakhutdinov, N.F.; Lavrik O.I. Tyrosyl-DNA phosphodiesterase inhibitors: Progress and potential. Bioorg. Med. Chem. 2016, 24, 5017-5027.

[3] Zakharenko, A.; Dyrkheeva, N.; Lavrik, O. Dual DNA topoisomerase 1 and tyrosyl-DNA phosphodiesterase 1 inhibition for improved anticancer activity. Med Res Rev.; 2019, 39, 1427-1441.

[4] Kawale, A.S.; Povirk, L.F. Tyrosyl-DNA phosphodiesterases: rescuing the genome from the risks of relaxation. Nucleic Acids Res. 2018, 46, 520-537.

[5] Interthal, H.; Pouliot, J.J.; Champoux, J.J. The tyrosyl-DNA phosphodiesterase Tdp1 is a member of the phospholipase D superfamily. PNAS 2001, 98, 12009-12014.

[6] Pommier, Y. Topoisomerase I inhibitors: camptothecins and beyond. Nat. Rev. Cancer 2006, 6, 789–802.

[7] Dexheimer, T.S.; Antony, S;. Marchand, C.; Pommier, Y. Tyrosyl-DNA phosphodiesterase as a target for anticancer therapy. Anti-Cancer Agent ME 2008, 8, 381-389.

[8] Beretta, G.L.; Cossa, G.; Gatti, L.; Zunino, F.; Perego P. Tyrosyl-DNA Phosphodiesterase 1 targeting for modulation of camptothecin-based treatment. Curr. Med. Chem. 2010, 17, 1500-1508.

[9] Ledesma, F.C.; El Khamisy, Sh.F.; Zuma, M. C.; Osborn, K.; Caldecott, K.W. A human 5′-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. Nature 2009, 461, 674–678.

[10] Salakhutdinov, N.; Volcho, K.; Yarovaya, O. Monoterpenes as a renewable source of biologically active compounds. Pure Appl. Chem. 2017, 89, 1105-1117.

[11] Zakharenko, A.L.; Khomenko, T.M.; Zhukova, S.V.; Koval, O.A.; Zakharova, O.D.; Anarbaev, R.O.; Lebedeva, N.A.; Korchagina, D.V.; Komarova, N.I.; Vasiliev, V.G.; Reynisson, J.; Volcho, K.P.; Salakhutdinov, N.F.; Lavrik, O.I. Synthesis and biological evaluation of novel tyrosyl-DNA phosphodiesterase 1 inhibitors with a benzopentathiepine moiety. Bioorg. Med. Chem. 2015; 23, 2044-2052.

DOI 10.18699/BGRS/SB-2020-348

Loss of *Drosophila Hyperplastic disc* promotes massive cell death and germline atrophy in oogenesis

Iuliia Aleksandrovna Galimova
IMCB SB RAS, Novosibirsk, Russia
galimova@mcb.nsc.ru

Natalia Vladimirovna Dorogova
ICG SB RAS, Novosibirsk, Russia
dorogova@bionet.nsc.ru

Elena Ustinovna Bolobolova
ICG SB RAS, Novosibirsk, Russia
elbol@bionet.nsc.ru

Svetlana Aleksandrovna Fedorova
ICG SB RAS, Novosibirsk, Russia
fsveta@bionet.nsc.ru

***Abstract* — Drosophila *hyd* is ortholog of the mammalian tumor suppressor *EDD* that implicated in a wide variety of cellular processes and its regulation is impaired in various types of tumors. It is a member of highly conservative HECT family of E3 ubiquitin ligases that directly transfers the ubiquitin to targeted substrates. Hyd is required for the regulation of cell proliferation during development, mutations in the *hyd* gene resulting in developmental abnormalities including adult sterility. Reduction of *hyd* expression results in large-scale germ cells death at different stages of oogenesis. Germ cells lacking Hyd demonstrated a slowdown in the growth and development during mid-oogenesis, leading to the total cell death due to metabolic impairment.**

Keywords* — *hyperplastic disc gene, oogenesis, cell death, E3 ubiquitin ligase, germ cells, drosophila

Motivation and Aim

One of the important aspects of carcinogenesis investigation deals with tumor suppressor genes that control the cellular constancy of the body and prevent the formation of tumors. *Drosophila melanogaster hyperplastic disc* gene (*hyd*) is ortholog of the mammalian tumor suppressor *EDD* (E3 identified by differential display, EDD/UBR5/hHyd, hereafter called EDD). EDD implicated in a wide variety of cellular processes and its regulation is impaired in various types of tumors. Hyd, as well as EDD, has E3 ubiquitin ligase activity and is critical for cell proliferation and differentiation [1,2]. Mutations in *hyd* gene display a range of developmental phenotypes, including imaginal disc abnormalities [3], larval lethality, sterility [4], premature photoreceptor differentiation [5], failed oogenesis [6], and defective spermatogenesis [7]. The clone analysis identified that Hyd negatively regulates expression of *Hedgehog (Hh)* and *Decapentaplegic (Dpp)* in eye and wing imaginal discs [5]. Here, we study the Hyd function in *Drosophila* oogenesis. We showed that the reduction of *hyd* expression by mutations or RNA interference results in large-scale germ cells death at different stages of oogenesis. This cell death does not depend on Dpp or Hh signaling.

Methods

We used Bloomington *Drosophila* stocks: kni hyd15 e1 /TM3, Sb1 (3718) and yw; PBac{3HPy+}hydC017/TM3, Sb1 Ser1 (16256) as the source of mutant *hyd* alleles. To suppress *hyd* expression in different ovarian cells we used the *hyd* RNA interference. To induce the RNA interference we combined UAS-*hyd*-RNAi transgene with a set of different ovarian-specific GAL4 drivers. We performed a detailed cytological analysis of oogenesis using fluorescent and electronic microscopy.

Results and Discussion

For the cytological analysis of the mutant effect in oogenesis, viable females carrying the interallelic combination *hydC017/hyd15* were used. *hydC017/hyd15* females are sterile and lay only single eggs from which embryos do not develop. The gonads of such females mainly contain partially or completely reduced ovaries: either with a small number of egg chambers, or without them at all.

Antibodies staining of germline cells marker, Vasa protein, showed that more than 90% of the ovaries consisted of ovarioles, either containing no germline cells or with single egg chambers. Although gonads form normally in embryos and larvae, but their internal structure degrades at later stages, only part of the cells gives rise to cysts that form a normal egg chamber. However formed egg chambers also degrade during oogenesis, only 1% of them develop into an egg. Thus, *hyd* mutations affect the viability of cells involved in oogenesis and egg chamber forming. Stem cells also die, and, as a result, ovarioles without germ cells are formed. Also cells die in germarium and egg chambers at different stages of development.

Oogenesis defects detected in *hydC017/hyd15* are associated with a decrease in *hyd* gene expression. Quantitative analysis of mRNA using real-time PCR showed that the level of gene expression is statistically different and about 2 times lower in the mutant than in the control (wild type).

The association of cell death with a decrease of the *hyd* activity is confirmed by our data obtained using *hyd* RNA interference (hyd-RNAi). For this experiment, we used the *UAS/GAL4* system. In order to ectopically suppress the *hyd* gene expression in different ovary cells, tissue-specific drivers were selected: *nanos*-GAL4, *oskar*-GAL4 - for germline cells, 7023-GAL4, 7024-GAL4, 36287-GAL4 - for somatic cells. Only combination *UAS-hyd-RNAi* with germline specific drivers, *nanos-GAL4* and *oskar-GAL4*, resulted in massive cell death and ovarian atrophy, which corresponded to the phenotypic manifestation of mutations.

As the *nanos-GAL4* starts to express in primordial germline cells from embryogenesis, in the *nanos-GAL4/UAS-hyd-RNAi* females, germ cell death occurred before reaching the adult stage. Electron microscopy revealed signs of cell death everywhere in the germ line of such female at the end of 3rd larval stage. Normally, at this stage, germ cells actively proliferate and increase in their number without subsequent differentiation. Cell death at this stage may be associated with misregulation of factors that control the cell cycle, cell death pathways, nuclear envelope and cytoskeleton remodulating.

The *oskar-GAL4* is activated at stages 3-4 of oogenesis, and in *oskarGAL4/UAS-hyd-RNAi* females, germ cell death occurred in mid-oogenesis. In this case, after stage 4-5, germ cells slowed down in the development. Although they slightly increased in size, the oocyte does not grow normally and does not increase in volume relative to the nurse cells area, compared wild type. The developmental arrest ended with the total (100%) death of the egg chambers. Electron microscopic analysis showed that the Hyd lack did not lead to the anomalies in the cell ultrastructure. Germ cells degradation occurred at the same stage of oogenesis as in the wild type and did not differ morphologically. Massive germ cells death and egg chambers in mid oogenesis were also observed when the expression of inhibitors of apoptosis Bruce and Diap was reduced [8]. However, in the Hyd loss background, the germ cells death occurred after a slowdown in the development and growth, which indicates a gradual extinction of cellular metabolism and a decrease in viability. We believe there is an effect similar to a protein deficiency or environmental stress. It is known that starvation and lack of proteins activate an oogenesis checkpoint triggering the mechanism of cell death [9]. Probably, the lack of E3 ubiquitin ligase Hyd leads to a disruption in the ubiquitination mechanism of the protein substrates responsible for the metabolism and synthetic activity of germ cells, causing cell death due to deficit of energy resources.

ACKNOWLEDGMENT

Supported by the RFBR (20-04-00496), and by ICG SB RAS budget Project (0324-2019-0042-C-01).

REFERENCES

[1] Clancy J.L. et al. (2003) *EDD*, the human orthologue of the hyperplastic discs tumour suppressor gene, is amplified and overexpressed in cancer. Oncogene. 22: 5070–5081.

[2] *Shearer R.F. et al. (2015)* Functional Roles of the E3 Ubiquitin Ligase UBR5 in Cancer. *Mol Cancer Res. 13(12):1523-1532.*

[3] Martin P. et al. (1977) Studies of l(3)c43hs1 a polyphasic, temperature-sensitive mutant of *Drosophila melanogaster* with a variety of imaginal disc defects. Dev Biol. 55:213–232.

[4] Mansfield E. et al. (1994) Genetic and molecular analysis of *hyperplastic discs*, a gene whose product is required for regulation of cell proliferation in *Drosophila melanogaster* imaginal discs and germ cells. Dev Biol. 165:507–526.

[5] Lee J.D. et al. (2002) The ubiquitin ligase Hyperplastic discs negatively regulates *hedgehog* and *decapentaplegic* expression by independent mechanisms. Development. 129:5697–5706.

[6] Szabad J. et al. (1991) Requirement for cell-proliferation control genes in Drosophila oogenesis. Genetics.127:525–533.

[7] Pertceva J.A. et al. (2010) The role of *Drosophila hyperplastic discs* gene in spermatogenesis. Cell Biol. Int.34:991–996.

[8] *Nezis I.P. et al. (2010)* Autophagic degradation of dBruce controls DNA fragmentation in nurse cells during late *Drosophila melanogaster* oogenesis. *.J Cell Biol. 190(4):523-531.*

[9] Jenkins V.K. et al. (2013) Diversity of cell death pathways: insight from the fly ovary.Trends Cell Biol. 23(11):567-574.

DOI 10.18699/BGRS/SB-2020-349

Computational insights into molecular mechanisms of CD95 programmed cell death activation

Nikita V. Ivanisenko
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
n.ivanisenko@gmail.com

Vladimir A. Ivanisenko
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia

Laura K. Hillert
Translational Inflammation Research,
Medical Faculty, Otto von Guericke
University Magdeburg, Magdeburg,
39106, Germany

Corinna König
Translational Inflammation Research,
Medical Faculty, Otto von Guericke
University Magdeburg, Magdeburg,
39106, Germany

Inna N. Lavrik
Translational Inflammation Research,
Medical Faculty, Otto von Guericke
University Magdeburg, Magdeburg,
39106, Germany
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia

Abstract — **The assembly of the death-inducing signaling complex (DISC) and death effector domain (DED) filaments at CD95/Fas results in activation of extrinsic apoptosis. This process is deregulated in the variety of tumors and neurodegenerative diseases. In this work computational modeling was carried out to describe the molecular mechanism of DISC assembly and to develop small molecules facilitating cell death induction.**

Keywords — CD95, cell death, apoptosis, c-FLIP, virtual screening, docking.

Introduction

There are two types of apoptosis induction: intrinsic-mediated via mitochondria and extrinsic-mediated via death receptor (DR) activation. Currently six DRs are characterized: CD95/Fas, TNF-R1, TRAILR1/2, DR3 and DR6, while CD95/Fas is one of the most studied members of the DR family. The induction of apoptosis via CD95 is largely controlled by the Death-Inducing Signaling Complex (DISC), which is formed upon CD95 stimulation. CD95 DISC comprises oligomerized, CD95, the adaptor protein FADD, procaspases-8/10 and cellular FLICE inhibitory proteins (c-FLIP). Deregulation of the CD95 pathway accompanies a variety of tumors and neurodegenerative diseases. Currently a limited number of small-molecule agents targeting this pathway is available. Molecular modeling of the components of DISC complex and development of compounds targeting them is of great interest and the goal of the current work.

Methods

Virtual screening was conducted using GLIDE software (Schrodinger Inc.). Molecular modeling was carried out using Rosetta Software.

Results

Using virtual screening techniques of large databases of chemical compounds we could identify first in class small molecule targeting heterodimer caspase-8/c-FLIPL [1]. Chemical probe was designed to target the heterodimerization interface leading to allosteric stabilization of the complex. Kinetic mathematical model was further developed to analyze the observed effects of FLIPinBγ on DISC activation. Based on the modeling results we could predict that the stabilized FLIPinBγ/ caspase-8/c-FLIPL complex plays a major role at the very initial stages of the DED chain assembly and procaspase-8 processing. Furthermore conducted structural analysis of the DISC complex [2-3] suggests high therapeutic potential of c-FLIP targeting compounds to enhance cell death in cancer cell lines that are characterized by high c-FLIP levels.

Acknowledgment

The study has been financially supported by RFBR grant (No.19-54-45015).

References

[1] Hillert, L.K., Ivanisenko, N.V., Busse, D., Espe, J., König, C., Peltek, S.E., Kolchanov, N.A., Ivanisenko, V.A. and Lavrik, I.N., 2020. Dissecting DISC regulation via pharmacological targeting of caspase-8/c-FLIP L heterodimer. Cell Death & Differentiation, pp.1-14.

[2] Hillert, L.K., Ivanisenko, N.V., Espe, J., König, C., Ivanisenko, V.A., Kähne, T. and Lavrik, I.N., 2020. Long and short isoforms of c-FLIP act as control checkpoints of DED filament assembly. Oncogene, 39(8), pp.1756-1772.

[3] Ivanisenko, N.V., Buchbinder, J.H., Espe, J., Richter, M., Bollmann, M., Hillert, L.K., Ivanisenko, V.A. and Lavrik, I.N., 2019. Delineating the role of c-FLIP/NEMO interaction in the CD95 network via rational design of molecular probes. BMC genomics, 20(3), p.293.

DOI 10.18699/BGRS/SB-2020-350

Single-nucleotide polymorphisms of hNEIL2 gene: from protein structure to functions in base excision DNA repair

Zarina Kakhkharova
ICBFM SB RAS, Novosibirsk, Russia
Novosibirsk State University,
Novosibirsk, Russia
zarinkaapels@gmail.com

Petrova Daria
ICBFM SB RAS, Novosibirsk, Russia
Novosibirsk State University,
Novosibirsk, Russia
dashapetr21@gmail.com

Inga Grin
ICBFM SB RAS, Novosibirsk, Russia
Novosibirsk State University,
Novosibirsk, Russia
grin@niboch.nsu.ru

Abstract — **Base excision repair (BER) is one of most universal mechanisms of DNA repair in living cells. BER initiated by specific enzymes – DNA-glycosylases recognizing damaged nucleotide and can influence on efficiency of all DNA repair pathway. Nucleotide polymorphisms (SNPs) of the NEIL2 gene could be associated with the risk of cancer develope in actively transcribed organs. We found two SNPs located in conservative regions responsible for catalysis and DNA binding and purified corresponding recombinant protein variants hNEIL2(R103W) and hNEIL2(P304T), respectively. A comparison of the enzymatic characteristics with the wild-type hNEIL2 protein revealed reducing enzyme activities for hNEIL2(P304T).**

Keywords — NEIL2, base excision repair, DNA-glycosylase, single nucleotide polymorphism

Motivation and Aim

Every day in our cells the DNA molecule is chemically modified by numerous exogenous and endogenous factors. To correct these damages there are same protective mechanisms named DNA repair and base excision repair (BER) is one of most universal. BER is initiated by specific enzymes DNA-glycosylases capable to recognize and excise the modified DNA bases with followed by some enzymatic stages for recovering intact structure of DNA. Gene sequence variation like a single-nucleotide polymorphism (SNP) can lead to another polypeptide sequence and enzymatic activity changes. Therefore, SNPs in DNA-glycosylase genes can decrease efficiency of BER and promote DNA damages accumulation causing the oncological transformation of cells.

Mammalian DNA-glycosylase NEIL2 is a bifunctional enzyme with AP-lyase and glycosylase activity [1]. Additionally NEIL2 has unique ability to recognize oxidative damages in “bubble” structures [2] forming during transcription. On the other hand, it is known that some SNPs of hNEIL2 gene are associated with risk of cancer developing in actively transcribed organs particular, one of the most dangerous – lung cancer. This project aims to study the effect of SNPs in the conserved regions of the hNEIL2 gene on the enzymatic functions of protein variants. This understanding can provide a real opportunity to predict an increased cancer risk in carriers of the corresponding alleles.

Methods

We analyzed the available SNP databases hNEIL2 gene and the structure of the hNEIL2 protein using bioinformatics methods (since XRC data are not available) and found that some SNP lead to large changes in the protein structure. We cloned the wild type gene and SNP corresponding genes using standard mutagenesis methods, expressed and purified recombinant proteins by affinity chromatography. Then, the substrate specificity and enzyme activities were studied for hNEIL2 recombinant proteins.

Results

We found that two SNPs are located in the coding region of the gene and alter the class of amino acids in conservative regions responsible for catalysis and DNA binding, hNEIL2(R103W) and hNEIL2(P304T) respectively. According to the 1000genome and ExAC projects the frequencies of these alleles were equal to MAF (Minor Allele Frequency) = 0.0148 for hNEIL2(R103W) and MAF = 0.0088 for hNEIL2(P304T). All proteins exhibit increased DNA glycosylase activity observed in the following chain: double-stranded DNA, single-stranded DNA and DNA with “bubble” structure. The hNEIL2(P304T) polymorphic variant dramatically reduces both AP-lyase and DNA glycosylase activities.

Acknowledgment

This study has been supported by RFBR according to the research project No. 18-44-540029.

References

[1] Grin I., Zharkov D.O. (2010) Inactivation of NEIL2 DNA glycosylase by pyridoxal phosphate reveals a loop important for substrate binding. Biochemical and Biophysical Research Communications Res. 394(1): 100-105.

[2] Makasheva K.A., Zharkov D.O. Requirements for DNA bubble structure for efficient cleavage by helix-two-turn-helix DNA glycosylases. Mutagenesis 35(1): 119-128.

DOI 10.18699/BGRS/SB-2020-351

The role of DNA repair in active DNA demethylation is studed by the construct based on the CRISPR/Cas9 system

Zarina Kakhkharova
ICBFM SB RAS, Novosibirsk, Russia
Novosibirsk State University,
Novosibirsk, Russia
zarinkaapels@gmail.com

Darya Khantakova
ICBFM SB RAS, Novosibirsk, Russia
Novosibirsk State University,
Novosibirsk, Russia
d.khantakova@alumni.nsu.ru

Inga Grin
ICBFM SB RAS, Novosibirsk, Russia
Novosibirsk State University,
Novosibirsk, Russia
grin@niboch.nsu.ru

Abstract — **Recently published data have demonstrated the possibility of using CRISPR / Cas9-based systems that direct intracellular DNA repair systems to consolidate changes in the genome and epigenome. We develop an approach for assessing the contribution of various DNA repair processes for integrity of DNA, and we proposed the mechanisms of active epigenetic DNA demethylation includes the participation of base excision repair and non-canonical mismatch repair system in this process.**

Keywords — CRISPR/Cas9, base excision repair, non-canonical mismatch repair, epigenetics, active DNA demethylation.

Motivation and Aim

Motivation

In mammalian cells the cytosine methylation of the CpG-rich regulatory regions of the individual genes forms the epigenetic landscape of the cell. During differentiation processes the molecular mechanism of local methylation status changes still remains a mystery. We develop the system based on CRISPR/Cas9 for investigation the molecular mechanism of the active DNA demethylation through DNA repair pathways: base excision repair and non-canonical mismatch repair.

Aim

Active DNA demethylation occurs via hydroxylation and/or deamination of 5-methylcytosine (5mC). The resulting 5mC derivatives are removed through the base excision DNA repair pathway [1] and cell needs the special mechanism for eliminate closely spaced toxic intermediates as single- and double-strand breaks. Recently, we demonstrated cooperative mechanism between the base excision repair and non-canonical mismatch repair pathways for local active DNA demethylation. [2]. This project aims to fundamental questions about the contribution of different DNA repair pathways into epigenetic DNA demethylation and also the development of technics for directed epigenome changes.

Methods

The phagemid DNA substrate containing defined mismatches and methylation mark was treated by mouse embryonic fibroblasts (MEF) cell-free extracts differing in the level of DNA repair activity followed by restriction analysis. For estimate of the contribution base excision repair and non-canonical mismatch repair systems in active DNA demethylation we initiated nick-formation by CRISPR/Cas9 system.

Results

The presence of multiple mispairs and nick-formation by CRISPR/Cas9 system promote proficient ncMMR resulting DNA demethylation in the internal region. In addition, we found that MSH2 protein is not required for the BER/Cas9-dependent removal of tandem oxidative mismatches.

ACKNOWLEDGMENT

This study has been supported by RFBR according to the research project No. 18-29-07059.

REFERENCES

[1] Moréra S., Grin I., Vigouroux A., Couvé S., Henriot V., Saparbaev M., Ishchenko A.A. (2012) Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyuracil-containing DNA. Nucleic Acids Res. 40(19): 9917–9926.

[2] Grin I., Ishchenko A.A. (2016) An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. Nucleic Acids Res. 44(8): 3713–3727.

DOI 10.18699/BGRS/SB-2020-352

Conformational dynamics in methylated DNA repair by human Fe(II)/alpha-ketoglutarate dependent dioxygenases ALKBH2 and ALKBH3

Lyubov Yu. Kanazhevskaya
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
Lyubov.Kanazhevskaya@niboch.nsc.ru

Denis A. Smyshlyaev
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
Department of Natural Sciences, Novosibirsk State University
Novosibirsk, Russia

Olga S. Fedorova
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
fedorova@niboch.nsc.ru

Abstract — **Alpha-ketoglutarate dependent dioxygenases of AlkB family are non-heme Fe(II) containing enzymes which protect cells against alkyl DNA lesions. Human homologs of AlkB protein ALKBH2 and ALKBH3 catalyse direct repair of 1-methyladenine, 3-methylcytosine and ethenoadenine in DNA. Conversion of the substrate proceeds in several steps that involve oxidation of the alpha-ketoglutarate cofactor, activation of molecular oxygen, generation of the reactive Fe(IV)=O intermediate and hydroxylation of the target alkyl moiety. Here we applied a pre-steady-state kinetic approach to assess the conformational dynamics of the enzyme-substrate complex in the course of catalysis. An intrinsic Trp fluorescence of enzymes as well as fluorescent base analogue 2-aminopurine and a dye–quencher pair FAM-BHQ1 inserted into model oligonucleotide substrates were used for detection of kinetics by a stopped-flow method. The data obtained are indicative of conformational flexibility of ALKBH-DNA complexes. The structure of methylated substrate is substantially changed during binding to the enzyme active site. The kinetic mechanism and rate constant describing specific conformational changes were determined by a quantitative analysis of fluorescence progression curves.**

Keywords — DNA dioxygenases, DNA repair, conformational dynamics, protein–DNA interactions, stopped-flow method, fluorescent probes, FRET analysis.

Repair of alkylation damages in *E. coli* is fulfilled by proteins of the "adaptive response", particularly, by Fe(II)/α-ketoglutarate dependent dioxygenase AlkB. Several AlkB homologs reconstitute this function in mammalian cells [1]. The AlkB family members ALKBH2 and ALKBH3 are able to remove 1-methyladenine (1-meA), 3-methylcytosine (3-meC), and ethenoadenine (εA) from DNA and RNA in human cells through an oxidative demethylation mechanism. The process begins with binding of the molecular oxygen to the Fe(II) cofactor in the enzyme active site. The subsequent oxidation of α-KG substrate to succinate results in formation of a reactive Fe(IV)=O intermediate that hydroxylates an alkyl moiety of DNA or RNA substrates [2]. The substrate specificity of ALKBH2 and ALKBH3 dioxygenases differs [3], and that determines their biological functions. The ALKBH2 protein preferably acts on double-stranded DNA, providing the integrity of genetic information. ALKBH3 demethylates single-stranded DNA and RNA, indicating its potential role in gene expression control. It was shown that expression level of these enzymes is associated with human cancers and genetic diseases [4]. Efforts are being made for developing specific inhibitors of oxidative demethylases to improve the efficacy of chemotherapy.

Recently, we have studied the interactions between bacterial dioxygenase AlkB and methylated DNA by a pre-steady-state kinetic approach [5]. To elucidate a detailed mechanism of the DNA substrate demethylation by ALKBH2 and ALKBH3 proteins we have studied a conformational dynamics within the enzyme-substrate complex in the course of catalysis. A stopped-flow method combined with fluorescence monitoring was applied for detection of conformational changes in the interacting molecules under pre-steady-state conditions. A 15 nt single- or double-stranded oligodeoxynucleotides containing 1-meA or 3-meC lesions were used as model substrates. Conformational transitions of the enzyme molecule during interactions with methylated DNA were recorded as changes in the intrinsic fluorescence of its Trp residues. To determine the overall dynamics of the DNA structure, an 1-meA-containing oligodeoxynucleotide was modified by the fluorescence emitter/quencher pair, FAM/BHQ1, and Förster resonance energy transfer (FRET) was measured. The kinetic curves obtained after rapid mixing of methylated DNA with the ALKBH2 (or ALKBH3) protein revealed the conformational flexibility of the dioxygenases and their substrates. Two global conformational transitions of the enzyme-substrate complex were detected in the time intervals 1–5 ms and 10–100 s. The kinetic mechanisms were evaluated in the fitting procedures using non-linear least-squares and the individual rate constants of each step were determined.

Acknowledgment

This work was supported by grant from the Russian Science Foundation (Grant No. 16-14-10038).

References

[1] B.I. Fedeles, V. Singh, J.C. Delaney, D. Li, J.M. Essigmann, The AlkB Family of Fe(II)/alpha-Ketoglutarate-dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond, J Biol Chem, 290, 2015, pp. 20734–20742.

[2] S. Martinez, R.P. Hausinger, Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases, J Biol Chem, 290, 2015, pp. 20702–20711.

[3] P.O. Falnes, M. Bjoras, P.A. Aas, O. Sundheim, E. Seeberg, Substrate specificities of bacterial and human AlkB proteins, Nucleic Acids Res, 32, 2004, pp. 3456–3461.

[4] S. Dango, N. Mosammaparast, M.E. Sowa, L.J. Xiong, F. Wu, K. Park, M. Rubin, S. Gygi, J.W. Harper, Y. Shi, DNA unwinding by ASCC3 helicase is coupled to ALKBH3-dependent DNA alkylation repair and cancer cell proliferation, Mol Cell, 44, 2011, pp. 373–384.

[5] L.Yu. Kanazhevskaya, I.V. Alekseeva, O.S. Fedorova, A Single-Turnover Kinetic Study of DNA Demethylation Catalyzed by Fe(II)/alpha-Ketoglutarate-Dependent Dioxygenase AlkB, Molecules, 24, 2019, pii. E4576.

DOI 10.18699/BGRS/SB-2020-353

The influence of ligand structure of ruthenium nitrosyl complexes on their biological activity

Darya Khantakova
ICBFM SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
d.khantakova@alumni.nsu.ru

Inga Grin
ICBFM SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
grin@niboch.nsu.ru

Abstract — **The study is focused on the novel photoactive ruthenium nitrosyl complex. In previous studies, we identified new pyrimidine complex that showed the best profile of cytotoxicity. We compared the effect of chemical nature and structure of different ligands on the cytotoxic potential of complexes.**

Keywords — ruthenium nitrosyl complex, nitric oxide, DNA repair

Motivation and Aim

Metal-based anticancer drugs have been used successfully in clinical therapy worldwide. The success of platinum compounds (such as cisplatin, carboplatin and oxaliplatin) in chemotherapy is undeniable. However, there are still many limitations, such as high mutagenic potential and cytotoxic activity against normal cells, resistance to the treatment of certain types of tumors. An active search for solutions to these problems has revealed a new class of compounds – ruthenium complexes (L_6Ru, L = ligand). They have several major advantages: ability to affect many types of cancer, lower toxicity and effectiveness against platinum-resistant types of cancer. Moreover, ruthenium complexes have shown selective biological activity as they can be administered as prodrugs and later activated within the reducing microenvironment of tumor. It is known that the radical intermediate particle $L_5Ru•$ selectively accumulates in cancer tumors along through transferrin-dependent pathway, leading to the appearance of reactive oxygen species (ROS), DNA damage and initiation of apoptosis mechanisms. The oxidative damages of DNA bases could be repaired by one the most common DNA repair pathway – base excision repair (BER) initiated by specific enzymes – DNA glycosylases and it maintains the stability of the genome in healthy cells, but accelerates the tolerance to anticancer therapy. However, the molecular mechanism of action of ruthenium complexes remains unclear. On the other hand, the ligands in ruthenium complex can significantly change the effect. Nitroso-ruthenium complexes are particularly attractive candidates because of their ability to release active radicals NO• and $L_5Ru•$ through photoactivation *in vitro* or redox reactions in cells. In previous studies, we demonstrated possibility of using novel nitroso-ruthenium complex $[Ru(NO)(bPic)_2(NO_2)_2OH]$ ({RuNO}) for investigation of the inhibition effect on DNA-glycosylases *in vitro* [1] and measurement the cytotoxicity potential *in vivo* [2]. This project aims to compare the influence of chemical nature and structure of different ligands of nitroso-ruthenium complexes on the cytotoxic potential for human cell line 239FT.

Methods

Because of collaboration with the Nikolaev Institute of Inorganic Chemistry SB RAS, we have great opportunity to select compounds with certain ligands to adjust the necessary properties: cytotoxicity and solubility in water. We measured the cytotoxicity potential of novel nitrosyl complexes with different ligands on a cancer cell line by MTT analysis.

Results

For ruthenium nitroso complexes, an increase in cytotoxicity was shown in the series: water-soluble NH_3-containing complexes (IC_{50}> 0,9mM), NO_2-containing complexes, chlorine-containing complexes (IC_{50}< 20 μM). Among chlorine-containing complexes, the complex $[Ru(NO)(C_8H_9NO_2)_2Cl_3]$ exerted the highest cytotoxicity (IC_{50}=1,5 ± 0,4 μM). The shift of the ester moiety in the pyridine ligand from the fourth to the third position decreased cytotoxicity by a factor of five (IC_{50}=7,8 ± 0,8 μM), and the replacement of the ethyl substituent in the ester moiety with methyl one, decreased cytotoxicity by a factor of two.

Acknowledgment

This work was partially supported by Russian State funded budget project # AAAA-A17-117020210023-1 to ICBFM SB RAS.

References

[1] Mikhailov A. A., Khantakova D. V., Nichiporenko V. A., Glebov E. M., Grivin V. P., Plyusnin V. F., Yanshole V. V., Petrova D. V., Kostin G. A., Grin I. R. (2019) Photoinduced inhibition of DNA repair enzymes and the possible mechanism of photochemical transformations of the ruthenium nitrosyl complex [RuNO(b-Pic)2(NO2)2OH]. Metallomics, 11 (12): 1999-2009.

[2] Kostin G.A., Mikhailov A.A., Kuratieva N.V., Pischur D.P., Zharkov D.O., Grin I.R. (2017) Influence of pyridine-like ligands on the structure, photochemical and biological properties of nitro-nitrosyl ruthenium complexes. New J. Chem., 41: 7758-7765.

DOI 10.18699/BGRS/SB-2020-354

The effect of protein-protein interactions on the activity of APE1 SNP forms

Olga A. Kladova
Laboratory of biopolymer modification
Institute of chemical biology and fundamental medicine
Novosibirsk, Russia
kladova@niboch.nsc.ru

Irina V. Alekseeva
Laboratory of biopolymer modification
Institute of chemical biology and fundamental medicine
Novosibirsk, Russia
Irina.Alekseeva@niboch.nsc.ru

Olga S. Fedorova
Laboratory of biopolymer modification
Institute of chemical biology and fundamental medicine
Novosibirsk, Russia
fedorova@niboch.nsc.ru

Nikita A. Kuznetsov
Laboratory of biopolymer modification
Institute of chemical biology and fundamental medicine
Novosibirsk, Russia
Nikita.Kuznetsov@niboch.nsc.ru

Abstract — **Human apurinic/apyrimidinic endonuclease 1 (APE1) is known to be an important participant in base excision repair (BER) and activator of several transcription factors. It was previously shown that APE1 can stimulate the activity of some DNA glycosylases, which indicates possible protein-protein interactions between APE1 and other BER enzymes. In this study, we tested several natural single-nucleotide polymorphic variants of APE1 (R221C, N222H, R237A, G241R, M270T, R274Q, P311S) on its ability to influence on protein-protein interactions with different human DNA repair enzymes such as OGG1, UNG, AAG, XRCC1, PCNA, and Pol β. Obtained data indicate differences in the BER enzyme effect on some APE1 variants supporting the important role of protein-protein interactions in the coordination of repair enzymes.**

Keywords *— base excision repair, protein-protein interactions, fluorescence*

Introduction

One of the major pathways to remove DNA lesions is the base excision repair (BER). BER pathway is a multi-step process, and can be reconstituted with a limited number of proteins such as damage-specific DNA glycosylases, AP endonuclease, DNA polymerase, DNA ligase and additional coordination proteins XRCC1 and PCNA. Coordination of the enzymatic activities of DNA glycosylases and AP endonuclease is essential to ensure a complete repair of the damaged bases. A "passing-the-baton" model proposed for BER is consistent with the findings that DNA glycosylases coordinate with other proteins to process the damaged DNA substrate. Indeed, numerous studies have revealed that APE1 promotes dissociation of the DNA glycosylase–product complex, and this situation in turn increases multiple turnover rates of some DNA glycosylases. Two mechanisms were proposed to explain the stimulation of DNA glycosylases by APE1: a passive mechanism, when an AP endonuclease cleaves a free AP-site to prevent its rebinding with the DNA glycosylase during reaction and an active mechanism that consists of direct displacement of DNA glycosylase from the AP site either *via* specific protein–protein interactions or *via* distortion of the duplex DNA structure to disrupt the DNA glycosylase–AP site DNA complex. Nevertheless, the detailed molecular mechanism of DNA glycosylase stimulation by APE1 remains unknown.

In this study, we focused on the issue of damaged DNA transfer from the complex with various human DNA glycosylases (OGG1, UNG, AAG), DNA polymerase β, XRCC1 or PCNA to AP endonuclease APE1. We tested several single-nucleotide polymorphic (SNP) variants of APE1 (R221C, N222H, R237A, G241R, M270T, R274Q, P311S) on its ability to influence on protein-protein interactions with different human DNA repair enzymes. SNP variants have substitutions of amino acid residues on the surface of the enzyme globule and in the DNA-binding site, thereby affecting protein-protein interactions or the catalytic reaction, respectively (Table 1).

Materials and Methods

The real time cleavage of the DNA duplex containing stable AP-site analog (F-site) was detected by stopped-flow method with fluorescence detection. A FAM/BHQ1-labeled substrate 5'-FAM-GCTCAFGTACAGAGCTG-3'/5'-CAGCTCTGTACGTGAGC-BHQ1-3' was used to FRET analysis of the DNA cleavage reaction. The F-site cleavage activity of WT APE1 and SNP forms was estimated by calculating observed catalytic rate constant k_{obs}. The activity of SNP variants in the presence of other BER enzymes was also studied by direct PAGE analysis of the product accumulation. Typical experiment was conducted at 37˚C in buffer containing 50 mM Tris-HCl, 50 mM KCl, 1 mM (Na_2EDTA), 1 mM DTT, 9% glycerol, 5 mM $MgCl_2$, pH = 7.5. The reaction mixture contained 10 nM WT APE1 or SNP forms, 1 μM F-containing DNA duplex, and 1 μM effector protein (OGG1, UNG, AAG, Pol β, XRCC1 or PCNA).

Results and Discussion

The analysis of the effect of different BER proteins on the activity of WT APE1 and SNP variants revealed several differences in protein-protein interactions. It was shown that OGG1, AAG, Pol β and XRCC1 stimulate WT, R237A and P311S APE1, but revealed negligible effect on other SNP variants (R221C, N222H, G241R, M270T, R274Q), indicating that these amino acids can be involved in the protein-protein interactions during damage repair. Interestingly to note that UNG and PCNA slightly stimulate WT APE1, but at the same time decrease the activity of R221C, M270T and R274Q, and have no effect on N222H and G241R. Previously it was shown that APE1 G241R has

little activity on nucleosomal DNA in comparison to WT APE1. Obtained data support that enzymes involved in the BER pathway have mutual effect on the activity of each other and that some amino acid substitution caused by SNPs lead to change of this effect.

TABLE 1 – AMINO ACID SUBSTITUTIONS ANALYZED IN THIS WORK

Amino acid substitution	**Location**
R221C	DNA-binding site (distance 8.0 Å between N η 1 of Arg221 and O atom of 3′ -phosphate group of second nucleotide on 3′ side of F-site).
N222H	DNA-binding site (distances 4.5 and 5.0 Å between N δ 2 atom of Ans222 and two O atoms of 5′ - and 3′ -phosphate groups of second nucleotide on 3′ side of F-site).
R237A	Internal coordination of α-helixes (distances 3.9 and 7.5 Å between Nη1 of Arg237 and Oε1 atom of Glu217 and Oδ1 atom of Asp219, respectively).
G241R	Exterior of protein globule.
M270T	Active site. Met270 is embedded into DNA minor groove, thereby displacing base opposite to AP site.
R274Q	Internal coordination near active site (distances 2.6 and 3.0 Å between Nη1/Nη2 of Arg274 and carbonyl O atoms of Ala304 and Ser307, respectively).
P311S	Structural element at end of β-sheet, bordering catalytic loop.

ACKNOWLEDGMENT

This work was supported by a grant MD-3775.2019.4.

DOI 10.18699/BGRS/SB-2020-355

The effect of empagliflozin and its combinationwith linagliptin on the renal autophagy and apoptosis regulators in *db/db* diabetic mice

Anton I. Korbut
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia
ORCID: 0000-0003-3502-5892

Natalia A. Muraleva
Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia
ORCID: 0000-0002-0665-1723

Iuliia S. Taskaeva
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia
ORCID: 0000-0002-2812-2574

Nataliya P. Bgatova
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia
ORCID: 0000-0002-4507-093X

Maksim V. Dashkin
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia

ORCID: 0000-0002-5099-5144

Vadim V. Klimontov
Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (RICEL – Branch of IC&G SB RAS)
Novosibirsk, Russia
ORCID: 0000-0002-5407-872

***Abstract* — Recent data indicated the emerging role of glomerular autophagy in diabetic kidney disease. We aimed to assess the effect of SGLT2 inhibitor empagliflozin, DPP4 inhibitor linagliptin, and their combination, on glomerular autophagy in a model of type 2 diabetes. Eight-week-old male db/db mice were randomly assigned to treatment with vehicle, empagliflozin or combonation of empagliflozin and linagliptin for 8 weeks. Age-matched non-diabetic *db/+* mice were acted as control. To estimate glomerular autophagy, immunohistochemistry for beclin-1 and LAMP-1 was performed. Podocyte autophagy was assessed by counting the volume density (Vv) of autophagosomes, lysosomes and autolysosomes by transmission electron microscopy. LC3B and LAMP-1, autophagy markers, caspase-3 and Bcl-2, apoptotic markers, were evaluated in renal cortex by western blot. Vehicle-treated *db/db* mice had weak glomerular staining for beclin-1 and LAMP-1 and reduced Vv of autophagosomes, autolysosomes and lysosomes in podocytes. Empagliflozin and the combination empagliflozin and linagliptin, enhanced the areas of glomerular staining for beclin-1 and LAMP-1 and increased Vv of autophagosomes and autolysosomes in podocytes. Renal LC3B and Bcl-2 was restored in actively-treated animals. LAMP-1 expression was enhanced in empagliflozin group; caspase-3 expression decreased in the empagliflozin-linagliptin group only. Mesangial expansion, podocyte foot process effacement and urinary albumin excretion were mitigated by both agents. The data provide further explanation for the mechanism of renoprotective effect of SGLT2 inhibitors and DPP4 inhibitors in diabetes.**

***Keywords* — *type 2 diabetes, chronic kidney disease, podocutes, autophagy, apoptosis, empagliflozin, linagliptin*.**

Introduction

Diabetic nephropathy is the most common cause for end-stage renal disease worldwide [1]. Synthetic dysfunction and violations in the life cycle of glomerular cells play an important role in the development of diabetic kidney disease [2]. Recent data indicate emerging role of autophagy and apoptosis dysregulation in the process [3-5]. Moreover, autophagy and apoptosis are considered as the therapeutic targets in diabetes [4-6]. Inhibitors of sodium-glucose cotransporter-2 (SGLT2) [7,8] and inhibitors of dipeptidyl peptidase-4 (DPP4) are considered as promising therapeutic agents in diabetic nephropathy [9,10], but little is known about the exact mechanisms of their protective activity.

Thus, we aim of our study to assess the effects of SGLT2 inhibitor empagliflozin, either alone or in combination with DPP4 inhibitor linagliptin, on the renal expression of autophagy and apoptosis regulators in a model of type 2 diabetes.

Matherials and Methods

Eight-week-old male *db/db* diabetic mice (BKS.Cg-Dock7^{m}+/+Lepr^{db}/J) were treated by vehicle, empagliflozin (10 mg/kg) or combination of empagliflozin and linagliptin (10 mg/kg of each agent) for 8 weeks. Non-diabetic heterozygous *db/+* mice were acted as control. Plasma levels of glucose, fructosamine, glycated albumin, creatinine and urinary albumin to creatinine ratio (UACR) were measured at Week 0 and 8 of experiment. Renal structural changes were analyzed quantitatively from the light and electron microscopic images. The levels of LC3B and LAMP-1, the markers for autophagy, caspase-3 and Bcl-2, the markers for apoptosis, were assessed in renal cortex by Western blot. To estimate autophagy in glomeruli, staining areas for beclin-1 and LAMP-1 were estimated by immunohistochemistry. Autophagy in podocytes was assessed by counting the volume density (Vv) of autophagosomes, lysosomes and autolysosomes in the electron microscopy images.

Results

The *db/db* mice were obese and hyperglycemic prior to the start of experiment. The blood glucose, fructosamine and glycated albumin levels, as well as UACR, remained elevated throughout the experiment in the vehicle group; these parameters decreased significantly in empagliflozin and empagliflozin-linagliptin-treated animals (all $p<0.05$). Empagliflozin, either alone or in combination with linagliptin, attenuated mesangial expansion, the thickening of glomerular basement membrane and effacement of podocyte foot processes (all $p<0.05$ *vs.* vehicle group).

Vehicle-treated *db/db* mice demonstrated higher renal expression of LC3B and caspase-3 as compared to *db/+* mice; while expression of LAMP-1 and Bcl-2 was decreased (all $p<0.01$). The weaker glomerular staining for beclin-1 and LAMP-1, and lower Vv of autophagosome, autolysosomes and lysosomes in podocytes was revealed (all $p<0.05$ *vs*. *db/+* mice).

There was increase in Vv of podocyte autophagosomes, autolysosomes and lysosomes and the areas of glomerular staining for beclin-1 and LAMP-1 under empagliflozin and empagliflozin-linagliptin treatment (all $p<0.05$). The renal expression of LC3B and Bcl-2 was restored in actively treated animals (all $p<0.05$ *vs.* vehicle group). Besides, LAMP-1 expression was enhanced in empagliflozin group ($p=0.03$ *vs.* vehicle group). The protein content of caspase-3 was decreased significantly in the combination group only ($p=0.008$).

Conclusion

The data demonstrate that empagliflozin, either alone or in combination with linagliptin, promotes autophagy and suppresses apoptosis in the kidneys in a model of type 2 diabetic nephropathy. These effects could contribute to preservation of glomerular structure and mitigation of podocyte injury. The data provide further explanation for the mechanism of renal protective effect of SGLT-2 and DPP-4 inhibitors in diabetes.

REFERENCES

[1] A.C. Webster, E.V. Nagler, R.L. Morton, P. Masson. "Chronic kidney disease", Lancet 2017, vol. 389(10075), pp. 1238-1252. doi: 10.1016/S0140-6736(16)32064-5.

[2] I.A. Bondar, V.V. Klimontov, E.M. Parfentyeva. "A role of glomerular cell dyfunction in the development of diabetic nephropathy", Probl Endokrinol. (Mosk.) 2006, vol. 52(4), pp. 45-49. doi: 10.14341/probl200652445-49.

[3] T.A. Lin, V.C. Wu, C.Y. Wang. "Autophagy in chronic kidney diseases", Cells 2019, vol. 8(1), p. 61. doi:10.3390/cells8010061.

[4] D. Yang, M.J. Livingston, Z. Liu, G. Dong, M. Zhang, J.K. Chen, Z. Dong. "Autophagy in diabetic kidney disease: regulation, pathological role and therapeutic potential", Cell Mol Life Sci 2018, vol. 75(4), pp. 669-688. doi: 10.1007/s00018-017-2639-1.

[5] H. Dai, Q. Liu, B. Liu. "Research progress on mechanism of podocyte depletion in diabetic nephropathy", J Diabetes Res 2017, vol. 2017, p. 2615286. doi:10.1155/2017/2615286.

[6] M. Kitada, Y. Ogura, I. Monno, D. Koya. "Regulating autophagy as a therapeutic target for diabetic nephropathy", Curr Diab Rep 2017, vol. 17(7), p. 53. doi:10.1007/s11892-017-0879-y.

[7] J. Nespoux, V. Vallon. "SGLT2 inhibition and kidney protection", Clin Sci (Lond) 2018, vol. 132(12), pp. 1329–1339. doi:10.1042/CS20171298.

[8] A.I. Korbut, V.V. Klimontov. "Empagliflozin: a new strategy for nephroprotection in diabetes", Diabetes Mellitus 2017, vol. 20(1), pp. 75-84. doi: 10.14341/DM8005.

[9] A.J. Scheen, P. Delanaye. "Renal outcomes with dipeptidyl peptidase-4 inhibitors", Diabetes Metab 2018, vol. 44(2), pp. 101–111. doi:10.1016/j.diabet.2017.07.011.

[10] A.I. Korbut, V.V. Klimontov. "Incretin-based therapy: renal effects", Diabetes Mellitus 2016, vol. 19(1), pp. 53-63. doi: 10.14341/DM7727.

DOI 10.18699/BGRS/SB-2020-356

PARP1 and PARP2 affinity to the lessions in the context of nucleosomes

Tatiana Andreevna Kurgina
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
t.a.kurgina@gmail.com

Ekaterina Anatolyevna Belousova
LBCE
Institute of Chemical Biology and Fundamental Medicin SB RAS
Novosibirsk, Russia
rina@niboch.nsc.ru

Michail Michailovich Kutuzov
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
kutuzov.mm@mail.ru

Rashid Oktamovich Anarbaev
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
ranarbaev@gmail.com

Olga Ivanovana Lavrik
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
lavrik@niboch.nsc.ru

Svetlana Nikolaevna Khodyreva
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
svetakh@niboch.nsc.ru

Members of Poly(ADP-ribose)polymerase family are key enzymes regulating the DNA repair system. For example, PARP1 binds damaged DNA resulting catalytically active complex. Using NAD+ as a substrate PARP1 catalyzes the synthesis of poly(ADP-ribose) covalently attached to many target proteins, including itself. This enzyme is a promising target for creating anti-cancer drugs. However, these compounds often display high cytotoxicity. These problems require new inhibitors as well as testing approaches. The inhibitors verification is additionally obstructed by complex chromatin organization of DNA in cells. The use of short DNA makes it difficult to extrapolate the results obtained in vitro to in vivo conditions. In the current work we present a test system that allows real time analyzing of PARP1 DNA binding in nucleosome context.

Keywords – poly(ADP-ribosyl)ation, nucleosome, DNA repair.

Introduction

Genome stability in higher eukaryotes is supported by the activity of several DNA repair pathways, which depend on the type of damage. Poly(ADP-ribose)polymerases 1 and 2 (PARP1 and PARP2) are some of the key enzymes regulating DNA repair system. This enzymes functions as a "sensor" of DNA damage. The interaction with damaged DNA stimulates the NAD+-dependent poly(ADP-ribosyl)ation PARPs activity. The polymer of poly(ADP-ribose) has a negative charge and its covalent attachment to PARPs leads to the dissociation of poly(ADP-ribosyl)ated PARPs from DNA due to electrostatic repulsion. The poly(ADP-ribose) is a signal molecule for the enzymes and cellular systems concerning DNA damage.

Recently, we created a method that can be used in real time measurements [1]. PARP1 was activated with DNA duplexes containing damage in one strand and fluorophore at the 3'-end of the other strand. Fluorescent label is necessary for detection of the fluorescence anisotropy. Fluorescence anisotropy was defined as the ratio of the polarized component to the total intensity: $A=(I1 - I2)/(I1 + 2I2)$, where I1 and I2 are the intensities of the light emitted by a fluorophore along different axes of polarization. The anisotropy level was used to estimate the size of the complex containing fluorescent DNA.

In this work, we applied this method to the nucleosome core particle (NCP). This system is more relative to *in vivo* conditions than naked DNA. In this system, anisotropy increases when PARP proteins bind to the nucleosome in the immediate vicinity of the FAM-labeled DNA ends.

Results

In this study, we used 3 different nucleosome structures and 3 different DNA structures. First, DNA and NCP without anything lesions (except blunt DNA-ends) were used. Further, we use two lesion-intermediates of base excision repair: AP-site and gap. We evaluated dissociation constants (Kd) for PARP1 and PARP2 on this structures.

We did not detect any difference between affinity to native and damage nucleosomes. It can be explained by strong affinity of PARP1 to blunt DNA ends: at the all cases, PARP1 bind DNA ends and direct influence on the fluorophore mobility.

The dramatic decreasing of Kd for the PARP2 and gaped-nucleosome is shown. This data are consistent with previous studies [2] of PARP2 interaction with DNA: Kd for single strand break is lower than Kd for blunt DNA end.

Conclusions

We have shown the applicability of our spectrophotometry method to detect PARP1 and PARP2 interaction with NCP. We detected binding of these enzymes to different lesions with different affinity in the context of nucleosome.

Acknowledgment

The work was supported by the grant from RFFS 17-00-00097.

References

[1] Kurgina T.A., Anarbaev R.O., Sukhanova M.V. and Lavrik O.I. "A rapid fluorescent method for the real-time measurement of poly(ADP-ribose) polymerase 1 activity", Anal Biochem., vol. 15, pp. 91-97, March 2018.

[2] Sukhanova M.V., Abrakhi S., Joshi V., Pastre D., Kutuzov M.M., Anarbaev R.O., Curmi P.A., Hamon L. and Lavrik O.I. "Single molecule detection of PARP1 and PARP2 interaction with DNA strand breaks and their poly(ADP-ribosyl)ation using high-resolution AFM imaging.", Nucleic Acids Res, vol. 7, April 2016.

DOI 10.18699/BGRS/SB-2020-357

Lesion recognition and cleavage of damage-containing G-quadruplexes by DNA glycosylases

Aleksandra A. Kuznetsova
Institute of Chemical Biology and Fundamental Medicine of the SB RAS
Novosibirsk, Russia
sandra-k@niboch.nsc.ru

Olga S. Fedorova
Institute of Chemical Biology and Fundamental Medicine of the SB RAS
Novosibirsk, Russia
fedorova@niboch.nsc.ru

Nikita A. Kuznetsov
Institute of Chemical Biology and Fundamental Medicine of the SB RAS
Novosibirsk, Russia
Nikita.Kuznetsov@niboch.nsc.ru

Abstract — **Human telomeres contain repeated TTAGGG elements, in which the 3′ exposed strand may adopt a G-quadruplex (Q4) structure. In addition, more than 40% of human genes have been found to contain at least one sequence near the promoter regions, which potentially could form a quadruplex structure. The guanine-rich regions of telomeres are hotspots for oxidation forming 8-oxoguanine, thymine glycol as well as abasic sites (AP-sites), the lesions that are handled by the base excision repair (BER) pathway. The first step in BER is initiated by DNA glycosylases, which locate damaged DNA bases in the large amount of normal DNA. DNA glycosylases catalyze excision of the abnormal base and produce an AP-site. However, the features of DNA repair processes in non-canonical structures, including quadruplexes, are still poorly understood. Therefore, the purpose of this work was a comparative analysis of the efficiency of the removal of damaged nucleotides from G-quadruplexes by human BER enzymes. We analyzed activity and substrate specificity of 8-oxoguanine-DNA glycosylase (OGG1), endonuclease VIII-like 1 (NEIL1), endonuclease III (NTH1) to their specific damaged nucleotides. As the control prokaryotic formamidopyrimidine-DNA glycosylase (Fpg) and endonuclease VIII (Nei) were used. Direct detection of the product formation by PAGE allowed to estimate the efficiency of enzymatic conversion of damaged G-quadruplexes under the same experimental conditions. The stages of formation of enzyme-substrate complexes and catalysis were analyzed by stopped-flow technique, which allows to register conformational rearrangements of model DNA substrates.**

Keywords *— **Base excision repair, DNA glycosylases, G-quadruplex***

Introduction

Endogenous and exogenous agents such as highly reactive cell metabolites, external environmental compounds, UV or ionizing irradiation continually damage cellular DNA. The major sources of endogenous DNA damage are reactive oxygen species (ROS). ROS generate a variety of DNA base lesions, including 7,8-dihydro-8-oxoguanine (8-oxoguanine, oxoG), thymine glycol (Tg), 2,6-diamno-4-hydroxy-5-formamidopyrimidine (Fapy-G), 4,6-diamno-5-formamidopyrimidine (Fapy-A) and many others [1].

Most of this endogenous burden is handled by the base excision DNA repair (BER) pathway. BER involves several DNA *N*-glycosylases specific for removing a wide array of oxidized, alkylated or deaminated bases, and in some cases mispaired normal bases, which yields abasic (AP) sites [2].

Beyond the canonical duplex form, DNA can populate a wide range of states, from single stranded conformations to four-stranded arrangements. Guanine-rich nucleic acids can fold into the non-B DNA or RNA structures called G-quadruplexes (Q4). G-quadruplexes are four-stranded secondary structures formed by particular G-rich nucleic acid sequences. They result from the stacking of multiple stable "G-quartets", planar arrangements of four guanines held together by Hoogsteen-type hydrogen bonding and further stabilized by monovalent cations (generally K^+ or Na^+). The topologies of G-quadruplexes can be classified into several types: antiparallel, parallel, and hybrid. Usually, the syn/anti-glycosidic conformation of guanines is considered to be an important factor in the G-quadruplex structure folding. Structure and function of G-quadruplexes have become an area of great interest. Recently, they have been associated with some human genetic neurodegenerative diseases [3]. G-quadruplexes are also believed to be important for telomere maintenance [4], DNA replication, genome rearrangements, DNA damage response, chromatin structure, RNA processing and transcriptional or translational regulation.

Despite the important roles that G-quadruplexes play in telomere biology and gene transcription, the features of DNA repair processes in non-canonical structures, including quadruplexes, are still poorly understood. Therefore, the purpose of this work was a comparative analysis of the efficiency of the removal of damaged nucleotides from G-quadruplexes by human BER enzymes. We analyzed activity and substrate specificity of 8-oxoguanine-DNA glycosylase (OGG1), endonuclease VIII-like 1 (NEIL1), endonuclease III (NTH1) to their specific damaged nucleotides. As the control prokaryotic formamidopyrimidine-DNA glycosylase (Fpg) and endonuclease VIII (Nei) were used. Direct detection of the product formation by PAGE allowed to estimate the efficiency of enzymatic conversion of damaged G-quadruplexes under the same experimental conditions. The stages of formation of enzyme-substrate complexes and catalysis were analyzed by stopped-flow technique, which allowed to register conformational rearrangements of model DNA substrates.

Results and Discussion

Several model DNA substrates were used. The model G-quadruplexes contained Tg residue at the loop region (Q4-Tg) or oxoG residue at the middle of GGG (Q4-oxoG). Also control G-quadruplex without damage (Q4) and control duplexes with Tg or oxoG (Tg/A, oxoG/C) were used for comparison ofg cleavage data.

Structural analysis of G4 sequences by circular dichroism (CD) spectroscopy

Firstly, we analyzed the ODN structures by CD spectroscopy under several buffer conditions including those required for the enzymatic assays. The DNA sequences were folded in the quadruplex buffer containing K^+, in which the ODN without a lesion had a hybrid quadruplex structure. The CD spectrum of the unmodified quadruplex Q4 has shown two pronounced maxima at 290 nm and 265 nm. It should be noted, that the control G-quadruplex and quadruplexes containing Tg at the loop region gave similar CD spectra. The thermal melting of these substrates did not gave significantly

reduced Tm values: 66°C for control Q4 and 58–62°C for modified Q4-Tg.

In contrast, the ODN with oxoG in the center G-quartet showed a dramatic change in the CD spectra compared to the native hybrid fold, namely, a maximum peak at 263 nm and a minimum at 245 nm, that are characteristic of parallel G-quadruplex structures adopted by telomeric sequences in molecular crowding conditions. Nevertheless, the Tm values of quadruplexes containing oxoG were 70–75°C. These data suggest the coexistence of parallel and other types of G-quadruplex conformations for Q4-Tg and Q4-oxoG, with a clear prevalence of the parallel conformation in the case of oxoG containing substrates.

Glycosylase activity

The model 17-nt duplexes containing a specific damage oxoG or Tg were used in order to compare the glycosylase activity towards G-quadruplex targets. Fpg, Nei and NEIL1 led to formation of the reaction product corresponding to the β,γ-elimination, while OGG1 and NTH1 led to formation of the reaction product corresponding to the β-elimination.

The interaction of Fpg with Q4-oxoG in 140 mM KCl led to the formation of several cleavage products, whereas in 50 mM KCl the cleavage of DNA backbone occurred preferably at the oxoG residue. It should be noted that interaction of OGG1 with Q4-oxoG did not lead to the substrate cleavage at all. The interaction of Nei and NEIL1 Q4-Tg led to formation of the cleavage products at 140 and 50 mM KCl, but the cleavage was more efficient in 50 mM KCl. It should be noted that NTH1 cleaved Q4-Tg only at 50 mM KCl.

Kinetic analysis of OGG1 and Fpg interaction with oxoG-substrates

One of the most important goals in the study of DNA repair is the elucidation of the enzymatic mechanism that provides highly precise recognition of the damaged base and its removing from DNA. Time-dependent stopped-flow experiments were carried out using 17-nt duplex substrates containing a single lesion at the sixth position from the 5'-end and G-quadruplexes. DNA substrates were labeled by aPu or FAM/BHQ1. It should be noted that we used two type of 17-nt duplex substrates ^{F}oxoG/C^{B} and ^{F}oxoG/^{B}C. The first one contained FAM and BHQ1 residues on the opposite side of the duplex, whereas in the second one the FAM and BHQ1 residues were located at the one side of the duplex. These substrates allow to register not only the substrate cleavage but also the bending of DNA in the course of catalytic complex formation.

The interaction of Fpg with ^{F}oxoG/C^{B} led to the strong arising of FAM fluorescence up to 100 second, whereas the interaction of Fpg with ^{F}oxoG/^{B}C resulted in FAM fluorescence decrease up to 5 second (duplex bending) and then to FAM fluorescence increase up to 100 second (substrate cleavage). The same FAM fluorescence changes were observed for interaction of OGG1 with ^{F}oxoG/C^{B} and ^{F}oxoG/^{B}C substrates.

The interaction of Fpg with Q4-oxoG$_{17}^{F/B}$ led to slow increase of FAM fluorescence probably indicating the slow cleavage of G4 quadruplex. Also we examined the G-quadruplex with oxoG and aPu located from the 3'- or 5'-side of damaged nucleotide. The intensity of emission of aPu is influenced by its environment; in particular, it is strongly quenched when aPu is stacked in a DNA duplex. The interaction of Fpg with aPu containing quadruplexes led to slow increase of aPu fluorescence intensity at the same manner indicating a shift of aPu to a more hydrophilic environment, revealing the destabilization of the G-quadruplex upon formation of complex with enzyme.

The interaction of OGG1 with Q4-oxoG did not cause effect on fluorescence of FAM. Whereas the OGG1 interaction with Q4-oxoG quadruplex with aPu located from the 5'-side from damage led to the slight increase of aPu fluorescence probably indicating the binding process.

Kinetic analysis of Nei, NEIL1 and NTH1 interaction with Tg-substrates

The interaction of Nei with ^{F}Tg/A^{B} led to the strong arising of FAM fluorescence intensity up to 300 second, whereas the interaction of Nei with ^{F}Tg/^{B}A resulted in FAM fluorescence decrease up to 0.1 second (duplex bending) and then to the increase of FAM fluorescence up to 200 second (substrate cleavage). The same FAM fluorescence changes were observed for interaction of NEIL1 with ^{F}Tg/A^{B} and ^{F}Tg/^{B}A substrates. The interaction of NTH1 with these substrates led to increase followed by decrease of FAM fluorescence in both cases.

The interaction of Nei and NEIL1 with Q4-Tg led to slow increase followed by fast decrease of FAM fluorescence probably indicating the slow cleavage of quadruplex. At the same time the interaction of NTH1 with Q4-Tg only led to decrease of FAM fluorescence up to 100 seconds, indicating the binding process. For all enzymes Nei, NEIL1 and NTH1 the interaction with aPu-labeled quadruplexes did not led to significant change of aPu fluorescence, suggesting that complex formation has no effect on aPu fluorescence intensity when aPu residue located in the loop region of quadruplex.

ACKNOWLEDGMENT

This work was supported partially by Russian Foundation of Basic Research grant No 19-04-00012 and Russian State funded budget project of ICBFM SB RAS # AAAA-A17-117020210022-4.

REFERENCES

[1] M.D. Evans, M. Dizdaroglu, M.S. Cooke, “Oxidative DNA damage and disease: induction, repair and significance”, Mutat. Res., 2004, vol. 567, pp.1–61.

[2] D. Svilar, E.M. Goellner, K.H. Almeida, R.W. Sobol, “Base excision repair and lesion dependent subpathways for repair of oxidative DNA damage”, 2011, Antioxid. Redox Signal., vol. 14, pp. 2491–2507,https://doi.org/10.1089/ars.2010.3466.

[3] A.R. Haeusler, C.J. Donnelly, G. Periz, E.A. Simko, P.G. Shaw, M.S. Kim, N.J. Maragakis, J.C. Troncoso, A. Pandey, R. Sattler, J.D. Rothstein, J. Wang, “C9orf72 nucleotide repeat structures initiate molecular cascades of disease”, 2014, Nature, vol. 507, pp. 195–200.

[4] Y. Xu, “Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA”, Chem. Soc. Rev., 2011, vol. 40, pp. 2719–2740.

DOI 10.18699/BGRS/SB-2020-358

Initial steps of base excision repair on DNA-substrates with non-canonical structures

Alexandra A. Kuznetsova
Institute of Chemical Biology and Fundamental Medicine
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
sandra-k@niboch.nsc.ru

Anastasiia T. Davletgildeeva
Institute of Chemical Biology and Fundamental Medicine
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
davleta94@gmail.com

Olga S. Fedorova
Institute of Chemical Biology and Fundamental Medicine
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
fedorova@niboch.nsc.ru

Nikita A. Kuznetsov
Institute of Chemical Biology and Fundamental Medicine
Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
Nikita.Kuznetsov@niboch.nsc.ru

Abstract — **Although canonical B-form is thought to be prevalent in human genome, DNA is able to adopt many non-canonical forms. Such DNA sites are often localized in genome regions having important biological significance. For example, human telomeres contain repeated TTAGGG elements, in which the 3′-exposed strand may adopt a G-quadruplex structure. In addition, more than 40% of human genes have been found to contain at least one sequence near the promoter regions, which potentially could form a quadruplex structure. Some regions of such nature have an increased susceptibility to oxidative damage. Indeed the guanine-rich regions of telomeres are hotspots for oxidation forming 8-oxoguanine, thymine glycol as well as abasic sites (AP-sites), the lesions that are handled by the base excision repair (BER) pathway. However, the features of DNA repair processes in non-canonical structures, including quadruplexes, are still poorly understood. Therefore, the purpose of this work was a comparative analysis of the efficiency of the removal of damaged nucleotides from non-canonical DNA structures by human BER enzymes. We designed a set of DNA substrates such as DNA bubbles, DNA loops and G-quadruplexes. We analyzed activity of DNA glycosylases and AP endonuclease on model substrates. Direct detection of the product formation by PAGE allowed to estimate the efficiency of enzymatic hydrolysis of damaged G-quadruplexes under the same experimental conditions. The stages of formation of enzyme-substrate complexes and catalysis were analyzed by stopped-flow technique, which allows to register conformational rearrangements of model DNA substrates.**

Keywords *— Base excision repair, DNA glycosylases, AP endonuclease, DNA non-canonical structures*

Human genome DNA adopts canonical B-form right-handed double-helical structure. However, many alternative DNA conformations are known to exist, and their roles in affecting biological processes have just begun to be understood. In particular, tandem DNA repeats for example, GGG-rich sequences, which are found in telomeric DNA, known to play critical role in aging and disease development. In addition, more than 40% of human genes have been found to contain at least one sequence near the promoter regions, which potentially could form a quadruplex structure. Moreover, complex nucleic acid structures can be formed during the processes of DNA replication and transcription, and other cellular events.

These domains of the genome could be susceptible to oxidative attack and other endogenous and exogenous agents such as highly reactive cell metabolites, external environmental compounds, UV or ionizing irradiation. Abasic sites are estimated to arise spontaneously ~10,000 times per mammalian genome equivalent per day. However, the features of DNA repair processes in non-canonical structures, including DNA bubbles, DNA loops or G-quadruplexes, are still poorly understood.

It is known that recognition and removal of non-bulky damage of nitrogenous bases proceed *via* the base excision repair (BER) pathway. In general, the BER pathway includes sequential actions of two enzymes for DNA incision: a DNA glycosylase and an AP endonuclease. DNA glycosylases generate AP-sites and blocking 3′-end groups, which should be removed in subsequent steps of BER in order to initiate the DNA repair synthesis and ligation. An AP endonuclease hydrolyses the phosphodiester bond located 5′ to the AP-site and introduces a break into the deoxyribophosphate backbone, in addition, it removes the remaining 3′-blocking groups: either the 3′-phospho-α,β-unsaturated aldehyde or the 3′-terminal phosphate group.

Therefore, the purpose of this work was a comparative analysis of the efficiency of the removal of damaged nucleotides from DNA non-canonical structures by human BER enzymes. We analyzed activity and substrate specificity of 8-oxoguanine-DNA glycosylase (OGG1), endonuclease VIII-like 1 (NEIL1), endonuclease III (NTH1) and AP endonuclease (APE1) to their specific damaged nucleotides. As the control prokaryotic formamidopyrimidine-DNA glycosylase (Fpg) and endonuclease VIII (Nei) were used. The model G-quadruplexes contained 8-oxoguanine (oxoG), thymine glycol (Tg) or stable AP-site analog (F-site) at the loop and core region were used. A set of DNA duplexes containing in the central part 1–5 nt loops with damaged nucleotide served as bulge-substrates. Also control G-quadruplex without damage and control duplexes with Tg, oxoG and F-site were used for comparison of cleavage activity. In order to obtain structural insights, we first analyzed the ODNs by CD spectroscopy, under several buffer conditions including those required for the enzymatic assays.

Direct detection of the product formation by PAGE allowed to estimate the efficiency of enzymatic hydrolysis of damaged DNA under the same experimental conditions. The model 17-nt duplexes containing a oxoG, Tg or F-site were used in order to compare DNA glycosylases and AP endonuclease activity relative to damaged G-quadruplexes and bulge-substrates. In the control DNA duplexes APE1

makes an incision of the phosphodiester bond on the 5′-side of F-site, Fpg, Nei and NEIL1 lead to formation of the reaction product corresponding to the β,γ-elimination, while OGG1 and NTH1 lead to formation of the reaction product corresponding to the β-elimination. Analysis of product accumulation using gel-electrophoresis has shown that APE1 excises all of the DNA substrates used, i.e. DNA quadruplexes, containing the single F-site in core or loop regions and bulged DNA duplexes, containing the F-site. The interaction of Fpg with Q4-oxoG and Nei and NEIL1 with Q4-Tg led the formation of cleavage product. It should be noted that cleavage of damaged quadruplex Q4-Tg by NTH1 was inefficient. Interaction of OGG1 with Q4-oxoG did not lead to the substrate cleavage at all.

The stages of formation of enzyme-substrate complexes and catalysis were analyzed by stopped-flow technique, which allowed to register conformational rearrangements of model DNA substrates. DNA substrates were labeled by aPu or FAM/BHQ1. It should be noted that we used two type of 17-nt duplex substrates: the first one contained FAM and BHQ1 residues on the opposite side of the duplex, whereas in the second one the FAM and BHQ1 residues were situated at the one side of the duplex. These substrates allow to register not only the substrate cleavage but also DNA bending and local melting in the course of catalytic complex formation.

Taking together, a set of DNA substrates with non-canonical structures was tested to evaluate the structural aberrations in non-canonical structures required for the cleavage of damaged DNA by human DNA glycosylases and AP endonuclease.

ACKNOWLEDGMENT

The part of this work involving analysis of DNA glycosylases was specifically funded by Russian Foundation of Basic Research grant 19-04-00012. The part of this work involving analysis of AP endonuclease was specifically funded by Russian Science Foundation grant 18-14-00135.

DOI 10.18699/BGRS/SB-2020-359

PARP1 activation directs RNA binding proteins to DNA damages to form PARG reversible compartments enriched in damaged DNA

Olga Lavrik
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

Anastasia Singatulina
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

Loic Hamon
SABNP
INSERM U1204
Université Paris-Saclay
Evry, France

Maria Sukhanova
LBCE
Institute of Chemical Biology and Fundamental Medicin SB RAS
Novosibirsk, Russia

Konstantin Naumenko
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

David Pastré
SABNPINSERM U1204
Université Paris-Saclay
Evry, France

Abstract — **Poly(ADP-ribose) polymerase 1 (PARP1) synthesizes poly(ADP-ribose) (PAR) at DNA damage sites to recruit DNA repair factors. PAR plays a key role in regulation of multiple processes in high eukaryotic cells and interacts with DNA and RNA binding proteins. Among proteins relocated on damaged DNA, the nuclear RNA-binding protein FUS is one of the most abundant, raising the issue about its involvement in DNA repair. We reconstituted the PARP-1/PAR/damaged DNA and analyzed this system by atomic force microscopy in the presence of FUS. We demonstrated the dissociation of FUS from mRNA in the presence of PAR, its recruitment at DNA damage sites via its binding to PAR, and observed assembly of dynamic compartments in which damaged DNA is concentrated. FUS contains disordered prion-like domains and its interaction with PAR is a driving force in formation of membrane less compartments by liquid-liquid phase separation. The hydrolysis of PAR by poly(ADP-ribose) glycohydrolase (PARG) results in dissociation of damaged DNA-rich compartments and initiates the nucleocytoplasmic shuttling of FUS in cells. We anticipate that FUS facilitates DNA repair through the transient compartmentalization of DNA damage sites and concentration in these compartments of DNA repair proteins. This reversible process permits to select damaged DNA from undamaged one and accelerates DNA repair. Other RNA binding protein YB-1 interacts with PAR and stimulates PARP1 activity. Multifunctional protein YB-1 is located in cytoplasm but also shows its nuclear localization. The role of YB-1 in the formation of liquid compartments formed at DNA damages under PARP1 activation was investigated. Therefore, this study shed light on the role of PARP1 and RNA binding proteins containing prion-like domains in the formation of dynamic liquid compartments, which can facilitate DNA repair by concentrating in these structures damaged DNA and repair proteins.**

Keywords — DNA repair; RNA-binding proteins; atomic force microscopy, cancer, liquid-liquid phase separation, neurodegenerative disease, poly(ADP-ribose), poly(ADP-ribose) polymerase 1

Introduction

In mammalian cells, DNA single- and double-strand breaks trigger a complex cascade of events in which the members of the poly(ADP-ribose)polymerase family such PARP1 recognize damaged DNA and synthesize poly(ADP-ribose) (PAR) polymer covalently attached to themselves or to other acceptor proteins [1]. The central enzyme for PAR production in cells and the main target of poly(ADP-ribosyl)ation (PARylation) during DNA damage is PARP1 [2]. PARP1-mediated PARylation of proteins and PAR polymer synthesis is one of the earliest events in DNA damage response. It facilitates the initial recruitment of proteins to DNA lesions and orchestrates a wave of chromatin remodelling events during the DNA repair. The protein PARylation is subjected to tight control by the enzymes that degrade PAR. The key ADP-ribose-degrading enzyme is poly(ADP-ribose)glycohydrolase (PARG), which makes protein PARylation a reversible post-translational modification [2]. The synthesis of PAR polymers in response to genotoxic stress provides a landing platform for a plethora of PAR-binding proteins among them number of RNA-binding proteins (RBPs) [3]. Both high PARylation level among RBPs and their abundance in nuclear regions damaged by short laser beam exposures raise issues about the putative role of the proteins in DNA damage response [4]. Here, we focus our attention on RBPs such as FUS (FUS/TLS, Fused in Sarcoma) and YB-1 (Y-box-binding protein 1). FUS is a multifunctional DNA/RNA-binding protein that is involved in the regulation of transcription, pre-mRNA splicing, mRNA transport and local mRNA translation, storage and can be implicated in DNA repair response. FUS is a member of the FET family and one of the most abundant and highly PARylated nuclear RNA-binding proteins. Similar to FUS, YB-1 carries out its functions in RNA methabolism and likely in DNA repair [3]. YB-1, as a transcription factor, controls the expression of stress-induced genes and the genes involved in DNA repair and as an RNA-binding protein, YB-1 mediates pre-mRNA splicing, is one of the major proteins constituting RNP granules in the cytoplasm, and modulates mRNA translation. It has been recently shown that PAR can nucleate the intracellular phase transitions of such RBPs such as FUS at microlaser-generated sites of DNA lesions [4]. Intracellular compartmentalization initiated by PAR-dependent phase separation can underlie the mechanisms by which PAR is involved in DNA- and RNA-dependent cellular events: for example, the formation of stress-granules, nucleoli, spliceosomes, and transcriptosomes [3]. The molecular mechanisms responsible for the formation and the possible functions of these compartments are, therefore, difficult to

The work was supported by RSF (project No. 19-04-00882), RFBR (project No. 18-04-00882) and Genopole Evry.

address in a cellular context. Further advances in the understanding of the function of FUS and YB-1 in DNA repair, notably when in interaction with PAR, are critically required. In the present work, we investigate *in vitro* and *in vivo* the FUS and YB-1 interaction with PAR to elucidate the mechanism involvement of these proteins in DNA repair through PARP1 activation.

Results

We developed an original approach based on a single molecule analysis by atomic force microscopy (AFM) to reconstitute the molecular system that serves to recruit FUS at DNA damage sites, including factors such as intact and damaged DNA, PARP1, and mRNA [5]. We also have applied the new real-time technique to explore YB-1-PARP1 interplay during the poly(ADP-ribosyl)ation process [6]. We have shown the local recruitment of FUS to PAR synthesized by PARP1 at damaged DNA sites and found that FUS then triggers the formation of large compartments in which damaged DNA is enriched. We have demonstrated the reversible nature of these compartments as the hydrolysis of PAR by PARG is sufficient to dissociate damaged DNA compartments formed by FUS. We also shown that YB-1 can stimulate PARP1 in the absence of magnesium, and that YB-1-PARP1 interplay can be mediated and regulated not only by the DNA at the initial stage of poly(ADP-ribosyl)ation, but also by poly(ADP-ribose) during elongation stage of the polymer synthesis.

Using AFM-based single molecule visualization, we have analyzed FUS-PAR interactions in the reconstituted molecular system including mRNA to mimic nuclear mRNA targets of FUS, damaged DNA, PARP-1 (to recognize DNA damages sites), NAD^+ (to trigger the synthesis of PAR by PARP1), and PARG (to hydrolyze PAR). We have demonstrated not only the strong affinity of FUS for PAR, but the recruitment of FUS to PAR synthesized by PARP1 at DNA damage sites followed by the formation of large aggregates or compartments in which damaged DNA is concentrated (Fig. 1a). Howevev, YB-1 or two other mRNA-binding proteins (HuR, G3BP1) that also have low complexity domain and also bind to PAR but fail to trigger the formation of large DNA-rich compartments (data not shown). In the case of YB-1, we found that YB-1 can form heteromeric complex with PARP1 on damaged DNA, serving as preferable PAR acceptor at the initiation stage. During elongation unmodified YB-1 molecules appear to non-selectively bind growing polymers of poly(ADP-ribose) rather than DNA at the PARP1 boarding site (data not shown).

Conclusions

Our data suggest that transient molecular assemblies triggered by FUS upon PARP-1 activation may facilitate DNA repair through compartmentalization of DNA damage sites. The hydrolysis of these compartments with PARG provides reversibility of the whole process by dissociation of these compartments (Fig. 1b).

In the case of YB-1, non-covalent binding of unstructured positively charged C-tail of YB-1 to PAR polymers during auto-modification of PARP1 stabilizes the catalytically active PARP1-DNA complex and stimulates PAR elongation.

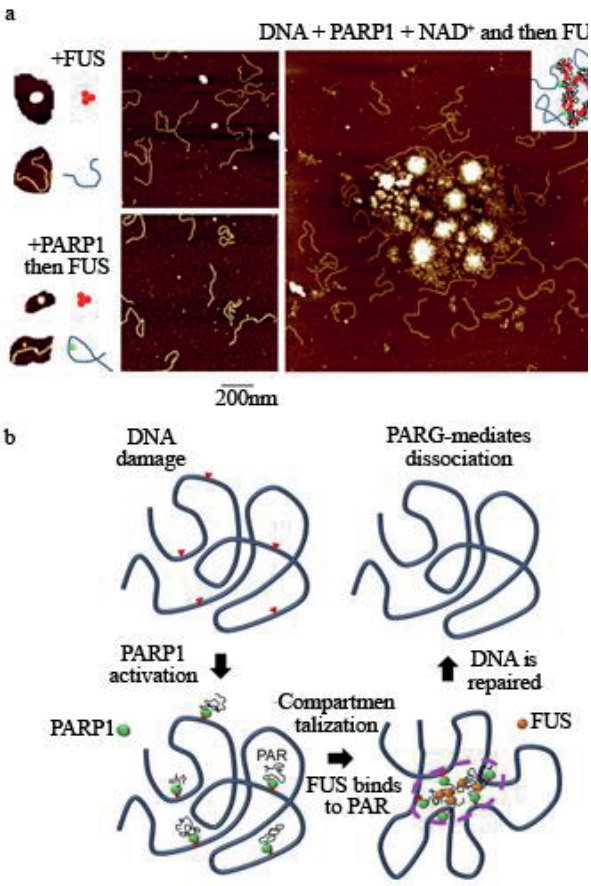


Fig. 1. FUS forms DNA-rich compartments upon PARP1 activation.a. AFM images of DNA after incubation with PARP1 in the presence of NAD^+ to trigger the synthesis of PAR, followed by the addition of FUS. b. Schematic view of the different steps of damaged DNA compartmentalization by FUS.

It is possible that regions of low complexity or prione-like domain in FUS (or YB-1) plays primary role in modulation of interaction of both protein with PAR.

References

[1] E. E. Alemasova, and O. I. Lavrik, "Poly (ADP-ribosyl) ation by PARP1: reaction mechanism and regulatory proteins" Nucleic Acids Res., vol. 47, pp. 3811-3827, May 2019.

[2] D. d'Amours, S. Desnoyers, I. d'Silva, and G.G. Poirier, "Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions", Biochem. J., vol. 342, pp. 249-268, September 1999.

[3] E. E. Alemasova, and O. I. Lavrik, "At the interface of three nucleic acids: the role of RNA-binding proteins and poly (ADP-ribose) in DNA repair", Acta Naturae, vol. 9, pp. 4-16, April 2017.

[4] S. L. Rulten, A. Rotheray, R. L. Green, G. J. Grundy, D. A. Moore, F. Gomez-Herreros, and K. W. Caldecott, "PARP-1 dependent recruitment of the amyotrophic lateral sclerosis-associated protein FUS/TLS to sites of oxidative DNA damage", Nucleic Acids Res, vol. 42, pp. 307-314, January 2014.

[5] A. S. Singatulina, L. Hamon, M. V. Sukhanova, B. Desforges, V. Joshi, A. Bouhss, O. I. Lavrik, and D. Pastré, "PARP-1 activation directs FUS to DNA damage sites to form PARG-reversible compartments enriched in damaged DNA" Cell Rep., vol. 27, pp. 1809-1821, May 2019.

[6] E. E. Alemasova, K. N. Naumenko, T. A. Kurgina, R. O. Anarbaev, and O. I. Lavrik, "The multifunctional protein YB-1 potentiates PARP1 activity and decreases the efficiency of PARP1 inhibitors", Oncotarget, vol. 9, pp. 23349-23365, May 2018.

DOI 10.18699/BGRS/SB-2020-360

The interplay between NHEJ and BER in NHEJ deficient cells

Polina Loshchenova
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
polilos@bionet.nsc.ru

Svetlana Sergeeva
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
sergeeva-sv@bionet.nsc.ru

Grigory Dianov
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
Oxford Institute for Radiation Oncology, University of Oxford, UK
grigory.dianov@oncology.ox.ac.uk

***Abstract* — Normally, in human cells, both DNA single-strand breaks (SSBs) and double-strand breaks (DSBs) arising due to exogenous and endogenous DNA damaging agents, are efficiently repaired by base excision repair (BER) and several DSB repair pathways, correspondingly. However, mutations in genes involved in DNA repair or extensive DNA damage caused by exogenous or endogenous mutagens may lead to the accumulation of unrepaired DNA strand breaks (SBs). Accumulation of the persistent SBs results in genomic instability and may lead to many human diseases including cancer. Unlike BER deficiency, that causes downregulation of NHEJ genes and initiates cell death, the knockdown of *XRCC4*, key NHEJ gene has a less dramatic effect on the expression of BER genes.**

Keywords — DNA damage, Base excision repair (BER), Non-homologous End Joining (NHEJ), Transcription factor Sp1

Motivation and Aim

Motivation

We earlier investigated how BER deficiency affects non-homologous end joining (NHEJ) repair pathway in normal human cells. We found that in normal human fibroblast in response to persistent DNA strand breaks, ATM protein kinase (ataxia-telangiectasia mutated), initiates proteasomal degradation of transcription factor Sp1 and leads to downregulation of BER genes expression, further accumulation of DNA single-strand breaks (SSBs) and renders cells susceptible to elimination via apoptosis (1). Furthermore, we also demonstrated that in cells, that can proliferate in presence of the DNA strand breaks (SB), knockdown of key BER gene *XRCC1* does not cause degradation of Sp1 but leads to downregulation of Lig4/XRCC4 and Ku70/80 at the transcription and protein levels by the Sp1-independent backup mechanism (2).

Aim

The interaction between NHEJ and BER deficiency is not thoroughly investigated. This work aimed to study the effect of NHEJ deficiency on BER in normal human fibroblast TIG-1.

Methods

The normal human fetal lung fibroblast cell line TIG-1 (RRID:CVCL_0560) was used in this study. Knockdown genes was carried out using siRNA transfections with lipofectamin. Protein expression levels were estimated by western blot analysis.

Results

In this study, we knocked down the *XRCC4* gene to model NHEJ deficiency in TIG-1 cells. We found that *XRCC4* knockdown results in minor downregulation of XRCC1, LigIII and APE1 proteins essential for BER pathway. Thus, in normal human fibroblasts deficiency of NHEJ has a less dramatic effect on the expression of BER genes, suggesting that the frequency of DSB is very low in normal human cells.

Acknowledgement

Supported by FASO (0324-2019-0040) and by RFBR (19-04-00067).

References

[1] Fletcher S.C. et al. (2018) Sp1 phosphorylation by ATM downregulates BER and promotes cell elimination in response to persistent DNA damage. Nucleic Acids Res. 46 (4): 1834-1846

[2] Loshchenova P. S. et al. (2020) Sp1-independent downregulation of NHEJ in response to BER deficiency. DNA REPAIR. 2020 Feb; 86:102740. Epub 2019 Nov 11. DOI: 10.1016/j.dnarep.2019.102740

DOI 10.18699/BGRS/SB-2020-361

Processing of clustered DNA damages by Nucleotide Excision Repair pathway

Natalia Lukyanchikova
ICBFM SB RAS, Novosibirsk, Russia
lunata9@yandex.ru

Irina Petruseva
ICBFM SB RAS, Novosibirsk, Russia
irapetru@niboch.nsc.ru

Olga Lavrik
ICBFM SB RAS, Novosibirsk, Russia
lavrik@niboch.nsc.ru

Alexander Lomzov
ICBFM SB RAS, Novosibirsk, Russia
lomzov@niboch.nsc.ru

***Abstract* — Nucleotide excision repair (NER) is a multistep process that eliminates a wide range of damages in DNA, including UV photoproducts and base modifications by many carcinogenic and chemotherapeutic agents [1]. Clustered DNA damages are defined as two or more damages situated within one to two helical turns of dsDNA [2]. In this study, we examined how the presence of AP site analog within a cluster with bulky damage in DNA impacts on functioning of Nucleotide Excision Repair machinery of mammalian cell.**

Keywords — DNA repair, Nucleotide Excision Repair.

Motivation and Aim

Motivation

We aim to analyze systematically the properties of the model DNAs bearing clustered damages. The main goal was to identify how the efficiency of recognition and removal of damages forming a cluster depends on both the structure and the distance between damages.

Aim

Earlier we have developed a mathematical model of cardiomyocyte electro-mechanical function [1] that predicted a significant role of the intra- and extracellular mechanical factors in arrhythmogenesys. Model prediction was verified in experiments on papillary muscles from the right ventricle of guinea pigs overloaded with calcium [2].

Methods

In order to evaluate the efficiency of NER we utilized the approach based on reproducing the repair reaction by mixing protein extracts from mammalian cells with modellinear DNAs, bearing clustered damages. We havesynthesized the set of long linear DNAs (137 bp), containing nFlu, the nonnucleoside bulky lesion, recognized and processed by NER system, and DEG, AP site analog). We also used the fluorescence anisotropy measurements for evaluation XPC affinity to the cluster-containing DNAs. Using photoaffinity labeling we have analyzed the interactions between XPD protein and the DNA probes bearing clustered structures. Molecular dynamics simulations have revealed structural for the differences observed.

Results

Results of fluorescence anisotropy measurements have shown that XPC has a significantly increased affinity to the cluster-containing DNAs. We have demonstrated that the NER-catalyzed excision of the DNA fragments that contain nFlu is fully abrogated in the presence of DEG insert in the certain positions of the complementary strand. The excision efficiency from the DNAs, containing the bulky lesion and the AP-site analog in one strand was not impeded. The experimental data together with Molecular Dynamic simulation results contribute to an understanding of the repair mechanisms of bulky adducts-containing clusters of different topologies in mammalian cell.

Acknowledgment

Supported by RFBR grant No. 19-04-00018.

References

[1] Orlando D. Schärer O. (2013) Nucleotide Excision Repair in Eukaryotes. Cold Spring Harb Perspect Biol. 5(10): a012609.

[2] Sage E., Harrison L. (2011) Clustered DNA lesion repair in eukaryotes: relevance to mutagenesis and cell survival. Mutat Res. 711(1-2): 123–133.

DOI 10.18699/BGRS/SB-2020-362

Human apurinic/apyrimidinic endonuclease 1 is modified by poly(ADP-ribose) polymerase 1 via the DNA structure-controlled mechanism

Nina Moor
Institute of Chemical Biology and Fundamental Medicine (ICBFM), Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia

Inna Vasil'eva
Institute of Chemical Biology and Fundamental Medicine (ICBFM), Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia

Nikita Kuznetsov
Institute of Chemical Biology and Fundamental Medicine (ICBFM), Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia

Olga Lavrik
Institute of Chemical Biology and Fundamental Medicine (ICBFM), Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russia

Abstract — **Apurinic/apyrimidinic endonuclease 1 (APE1) is an essential multifunctional protein in mammals involved in base excision DNA repair (BER) and other processes. Poly(ADP-ribose) polymerase 1 (PARP1) modifies itself and target proteins with poly(ADP-ribose), contributing to regulation of many processes. To understand molecular basis of cooperation between APE1 and PARP1 in BER, we examined poly(ADP-ribose)-binding activity and ADP-ribosylation of human APE1 in comparison with known targets of PARP1, using full-length, N-terminally truncated and catalytically inactive APE1 forms. The protein binds preferentially large ADP-ribose polymers, being similar to DNA polymerase β (Polβ) but contrasting with the scaffold XRCC1 protein. The interaction with poly(ADP-ribose) involves the universally conserved portion and the eukaryote-specific extension of APE1. The ADP-ribosylation of APE1 depends on the structure of PARP1-activating DNA, contrasting APE1 with Polβ and XRCC1. Relative levels of APE1 modification in the presence of different DNAs were found to correlate with affinities of the DNAs for APE1 and substrate activities in the enzymatic incision, suggesting the ADP-ribosylation to occur within the DNA-mediated complex. This conclusion was confirmed by importance of the length of DNA region 3' to the AP site for the modification. Deletion of the N-terminal extension of APE1 produced no significant influence on the ADP-ribosylation and hydrolytic stability of the modified protein, suggesting localization of target residues in the catalytic core. The most efficient ADP-ribosylation of the inactive APE1 reduced the level of PARP1 automodification, suggesting role of APE1 in modulating PARP1 activity. Our data provide new insights into mechanisms of protein targeting for ADP-ribosylation.**

Keywords *— apurinic/apyrimidinic endonuclease 1, poly(ADP-ribose) polymerase 1, protein ADP-ribosylation, base excision repair, posttranslational modification*

Introduction

APE1 is an essential protein in mammals with multiple functions in BER, regulation of gene expression and RNA metabolism [1]. BER is a molecular pathway devoted to correction of apurinic/apyrimidinic (AP) sites, modified bases and single-strand breaks (SSBs) [2, 3]. The major enzymatic function of APE1 in BER is incision of AP sites. In addition, APE1 can remove terminal blocking groups of BER DNA intermediates. The multiple activities of APE1 are modulated by protein-protein interactions and post-translational modifications [4, 5]. PARP1 is among the proteins modulating the APE1 activities in BER [6, 7]. This abundant nuclear protein acts as a rapid DNA damage sensor activated by binding to SSBs or double-strand breaks [2]. The subsequent PARP1 automodification with poly(ADP-ribose) (PAR) is a key mechanism for coordination of BER, via the PAR-dependent interaction with other BER proteins, primarily with a scaffold X-ray repair cross-complementing protein 1 (XRCC1), which in turn interacts with multiple BER enzymes. PARP1-catalysed modification of proteins modulates their properties and functions [2, 5]. A regulatory role of PARP1 in BER is also based on direct interactions with enzymes and competition for the interaction with DNA intermediates [6–8]. We have shown direct interaction of APE1 with PARP1 and its DNA-dependent modulation [8]. To further extend our knowledge of molecular mechanisms underlying the APE1-PARP1 interaction, we examined the non-covalent binding of APE1 to PAR and PARP1-catalysed ADP-ribosylation of APE1. The study was performed with the full-length human APE1, its N-terminally truncated forms deprived of the entire eukaryote-specific extension (APE1NΔ61) or its disordered fragment (APE1NΔ35) and an inactive APE1 D210N mutant to localize the structural regions involved in the non-covalent and covalent interaction with PAR. The ADP-ribosylation of APE1 was explored in detail, using different synthetic DNA intermediates of BER and their structural variants.

Results and Discussion

Characterization of PAR-binding activity of APE1

Proteins identified as targets of poly(ADP-ribosyl)ation interact non-covalently with PAR [9]. Therefore we first explored the non-covalent interaction of APE1 with PAR in comparison with Polβ and XRCC1, using partially fractionated polymer. APE1 and Polβ bind preferentially large linear and branched polymers (>20-mers) with similar parameters, in contrast to high-affinity interaction of XRCC1 with small ADP-ribose polymers. Using the full-length and truncated forms of APE1 it was shown that both the N-terminal eukaryote-specific extension and the C-terminal universally conserved catalytic portion of APE1 are involved in PAR binding, with the first of them contributing for the most part to the interaction with linear polymers [10]. The positively charged part of the extension conserved in mammals is critically important for the interaction with small polymers (6–10-mers). Overlapping of PAR- and DNA-binding domains in both APE1 and Polβ was shown in the competitive binding experiments performed in the presence of DNA intermediates.

Covalent ADP-ribosylation of APE1 catalysed by PARP1

The ADP-ribosylation of APE1 was explored in parallel with known targets of PARP1, Polβ and XRCC1, using various DNAs for PARP1 activation. To exclude impact of DNA incision at AP site, the experiments were performed with the

wild-type APE1 and catalytically inactive D210N mutant in the presence of calcium ions as PARP1 cofactor. The ADP-ribosylation of Polβ and XRCC1 is independent on the presence and type of damage (intact/incised AP site or gap) in the DNA duplex. The most efficient modification of APE1 was detected in the presence of 32-mer DNA duplex containing a medially located synthetic AP site and a purine base in the opposite position (F/Pur-d32). The relative levels of APE1 ADP-ribosylation in the presence of DNA variants correlate with their affinities for APE1, indicating that the modification is controlled by the strength of APE1-DNA interaction and occurs within the ternary complex of proteins with DNA. The alternative PAR-mediated mode of APE1 interaction with PARP1 is excluded by the fact that deletion of the N-terminal extension of APE1 contributing to the interaction with PAR revealed no negative influence on the ADP-ribosylation efficiency. The primary role of DNA in proper positioning of APE1 within the ternary complex is further evidenced by results obtained in experiments with sequence variants of F/A-d32 DNA. The effects produced by the sequence variations on the initial rates of the APE1-catalysed AP-DNA incision in multiple-turnover conditions and on the ADP-ribosylation efficiency of APE1 (both wild-type and truncated/mutant forms) well correlate with each other. These results indicate that, in addition to the affinity of APE1 for DNA, the conformational dynamics of APE1-DNA contacts mediating substrate binding and product release determines the efficiency of PARP1-catalysed modification of APE1, thus suggesting localization of the acceptor residue(s) in the catalytic core of APE1.

The experiments performed with truncated and elongated variants of F/A-d32 DNA have shown that the efficiency of PARP1-catalysed APE1 modification is determined by the length of base-paired region 3' to the AP site. PARP1 is activated predominantly via interaction with the blunt ends of DNA duplex [11]. In the X-ray structure of its complex with DNA duplex, each end of the duplex is independently bound to the protein via interactions of Zn1 and WGR domains with 3' and 5' terminated DNA strands respectively [12]. Therefore, the catalytic domains of PARP1 bound to the DNA ends in the proposed ternary complex are oriented to opposite directions relative to the AP site and as a consequence relative to the bound APE1 (Fig. 1). Our results show that PARP1 bound to the 3' end of AP site-containing strand is capable of catalysing the modification of APE1. The optimal length between the binding sites of APE1 and PARP1 in the DNA was estimated being about 20 base pairs.

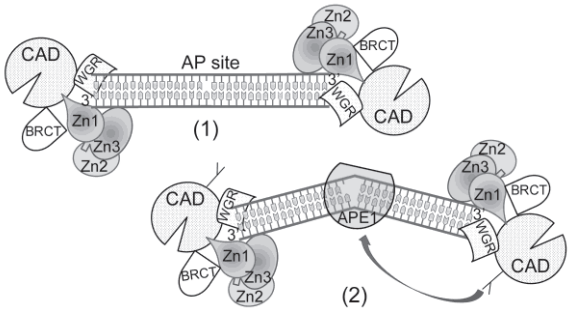


Fig. 1. A mechanism of PARP1-catalysed ADP-ribosylation of APE1 controlled by the DNA structure. State (1) shows PARP1 binding to AP site-containing DNA. PARP1 contains three Zn-finger domains, WGR, BRCT and the catalytic (CAD) domains positioned in accordance with the PARP1-DNA structure [12]; the 3' and 5' ends of each DNA strand are bound to the Zn1 and WGR domains respectively. Specific interaction of APE1 with AP site (state 2) results in kinking the DNA. PARP1 bound 3' to the AP site at a distance of about two helical turns is shown to be capable of catalysing the modification of APE1.

Our studies provide evidence that diverse mechanisms may be responsible for the functional cooperation between APE1 and PARP1. The ADP-ribosylation of APE1 and direct interaction between APE1 and PARP1 modulated by the structure of DNA intermediates are evidently related to assistance between the proteins during DNA repair. The unusual mechanism of protein ADP-ribosylation provides new insights into mechanisms of protein targeting for the post-translational modification with ADP-ribose. DNA acting as the activator of PARP1 can bring the catalytic domain of PARP1 in close proximity to the acceptor protein and thereby play a direct role in the substrate targeting.

ACKNOWLEDGMENT

This work was supported by the Russian Science Foundation 19-14-00107. Protein purification was supported by grant of Russian Fund for Basic Research 17-04-00925.

REFERENCES

[1] G. Antoniali, M. C. Malfatti, and G. Tell, Unveiling the non-repair face of the Base Excision Repair pathway in RNA processing: A missing link between DNA repair and gene expression? DNA Repair, 56 (2017) 65–74.

[2] R. Abbotts, and D. M. Wilson 3rd, Coordination of DNA single strand break repair, Free Radic. Biol. Med., 107 (2017) 228–244.

[3] A. M. Whitaker, M. A. Schaich, M. R. Smith, T. S. Flynn, and B. D. Freudenthal, Base excision repair of oxidative DNA damage: from mechanism to disease, Front. Biosci. (Landmark Ed.), 22 (2017) 1493–1522.

[4] S. Thakur, M. Dhiman, G. Tell, and A. K. Mantha, A review on protein-protein interaction network of APE1/Ref-1 and its associated biological functions, Cell Biochem. Funct., 33 (2015) 101–112.

[5] K. L. Limpose, A. H. Corbett, and P. W. Doetsch, BERing the burden of damage: pathway crosstalk and posttranslational modification of base excision repair proteins regulate DNA damage management, DNA Repair, 56 (2017) 51–64.

[6] M. V. Sukhanova, S. N. Khodyreva, N. A. Lebedeva, R. Prasad, S. H. Wilson, and O. I. Lavrik, Human base excision repair enzymes apurinic/apyrimidinic endonuclease1 (APE1), DNA polymerase beta and poly(ADP-ribose) polymerase 1: interplay between strand-displacement DNA synthesis and proofreading exonuclease activity, Nucleic Acids Res., 33 (2005) 1222–1229.

[7] S. N. Khodyreva, R. Prasad, E. S. Ilina, M. V. Sukhanova, M. M. Kutuzov, Y. Liu, E. W. Hou, S. H. Wilson, and O. I. Lavrik, Apurinic/apyrimidinic (AP) site recognition by the 5'-dRP/AP lyase in poly(ADP-ribose) polymerase-1 (PARP-1), Proc. Natl. Acad. Sci. U.S.A., 107 (2010) 22090–22095.

[8] N. A. Moor, I. A. Vasil'eva, R. O. Anarbaev, A. A. Antson, and O. I. Lavrik, Quantitative characterization of protein-protein complexes involved in base excision DNA repair, Nucleic Acids Res., 43 (2015) 6009–6022.

[9] J. P. Gagné, M. Isabelle, K. S. Lo, S. Bourassa, M. J. Hendzel, V. L. Dawson, T. M. Dawson, and G. G. Poirier, Proteome-wide identification of poly(ADP-ribose) binding proteins and poly(ADP-ribose)-associated protein complexes, Nucleic Acids Res., 36 (2008) 6959–6976.

[10] N. A. Moor, I. A. Vasil'eva, N. A. Kuznetsov, and O. I. Lavrik, Human apurinic/apyrimidinic endonuclease 1 is modified *in vitro* by poly(ADP-ribose) polymerase 1 under control of the structure of damaged DNA, Biochimie, 168 (2020) 144–155.

[11] W. Lilyestrom, M. J. van der Woerd, N. Clark, and K. Luger, Structural and biophysical studies of human PARP-1 in complex with damaged DNA, J. Mol. Biol., 395 (2010) 983–994.

[12] M. F. Langelier, J. L. Planck, S. Roy, and J. M. Pascal, Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1, Science, 336 (2012) 728–732.

DOI 10.18699/BGRS/SB-2020-363

YB-1 as modulator of PARP1 activity

K. N. Naumenko
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

E. E. Alemasova
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

M. M. Kutuzov
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

M. V. Sukhanova
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

T. A. Kurgina
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

O. I. Lavrik
LBCE
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia

Abstract — **Base excision repair (BER) is one of the main DNA repair pathways is aimed at repairing the most common DNA damages such as single-strand breaks *apurinic*/*apirimidinic* (AP) sites or modified bases resulting from its oxidation, deamination or alkylation. The poly (ADP-ribose) polymerase 1 (PARP1) is considered as one of the key regulators of BER process. Despite the fact that several key proteins are sufficient for reconstruction *in vitro* BER reactions, it is currently believed that additional regulatory proteins can be involved in the BER in cells. One of these proteins is the multifunctional Y-box binding protein (YB-1). In the current work, using model DNAs carrying a various types of damage, we quantified the level of poly (ADP-ribose) (PAR) synthesized by PARP1 in the presence or absence of the regulatory protein YB-1. We found that effect of YB-1 on the activity of PARP1 is independent on the structure of the DNA substrates. Using a model nucleosome structure, we shown that PARP1 and YB-1 form a heteromeric complex and YB-1 is PARylated by PARP1, suggesting that YB-1 can be an acceptor of ADP-ribose in the context of chromatin. Using nuclear extracts from HeLa cells, we showed that the addition of YB-1 to the extract leads to increased level of poly(ADP-ribose) synthesis and prevented the PAR degradation. We suggest that YB-1 is able to mediate stress response by increasing the total yield of poly(ADP-ribose).**

Key words* — *poly(ADP-ribosyl)ation, stress response, PARP1, YB-1

Introduction

Human DNA is constantly exposed to damaging exogenous and endogenous agents [1]. Preservation of the integrity of genetic information is provided by DNA repair systems. BER is one of the main repair paths aimed at repairing the most common DNA damage – AP-sites and bases that underwent modification (oxidation, deamination, alkylation). One of the key regulators of the BER process are poly(ADP-ribose)polymerases (PARPs) [2]. PARP1 is the most active enzyme in this family. PARP1 is activated by damaged DNA and synthesizes up to 90% of total cellular poly(ADP-ribose) covalently attached to acceptor proteins including PARP1 itself. Poly(ADP-ribosyl)ation is type of protein post-translational modification leading to more efficient dissociation of proteins from the complex with DNA [3] and modulates the cellular localization of proteins [4]. In addition, poly(ADP-ribose) performs a signaling function and attracts DNA-, RNA-, and PAR-binding proteins to the site of DNA damage [5]. Despite the long history of studying the role of PARP1 in the cellular response to damage, the exact mechanism of stimulation and regulation of PARP1 activity by other proteins has not been fully described. Currently, several proteins modulating PARP1 activity have been discovered: Sam68 [6], p53 [7], RPA [8]. Y-box binding protein 1 (YB-1) is one of the potential candidates for the role of regulatory protein of PARP1. Mass spectrometry proteome-wide identification of poly (ADP-ribose) binding proteins revealed YB-1 as protein which associates with poly (ADP-ribose) in cell [9]. It was shown that YB-1 poly(ADP-ribosyl)ated by PARP1 and PARP2 enzymes *in vitro* [10]. Recently, we have found that YB-1 stimulates the synthesis of poly (ADP-ribose) and reduces efficacy of various PARP1 inhibitors [11].

In the present study, we report new findings concerning functional interactions of PARP1 and YB-1. It was shown that YB-1 forms a heteromeric complex with PARP1 and damaged DNA, acting as the primary target for covalent modification by a poly(ADP-ribose). Using a set of model DNAs, the effect of the structure of a model duplex on PARP1 activity was studied. It was found that in the presence of YB-1, the level of ADP-ribose synthesis does not depend on the structure of the DNA duplex. The effect of YB-1 on PARP1 activity in the chromatin context was studied in the presence of mononucleosome substrate. It has been established that YB-1 is a predominant target for poly (ADP-ribosyl)ation in such system. In addition, it was shown that YB-1 increases the level of synthesis of poly(ADP-ribose) in nuclear extracts of HeLa cells.

Results

We hypothesized that the interaction between YB-1and PARP1 and observed stimulation of PARP1 activity may depend on the type of damaged DNA. The level of poly(ADP-ribose) synthesized by PARP1 was tested in the presence of YB-1 using different DNA structures (single- or double-strand breaks, mismatched nucleotides, bubble-type duplexes, single-stranded DNA and dumbbell DNAs with Nick or Gap). Altogether, obtained results demonstrate, that PARP1 activity is dependent on DNA-structure. In particular, PARP1 activation with dumbbell DNAs less intense than it was observed for blunt-ended DNA duplexes (Fig. 1 A, B). The effect of PARP1 stimulation with YB-1 was more pronounced in the case of dumbbell DNAs, where PARP1 activity was strongly increased. We demonstrated that C-terminal domain of YB-1 is essential for the ability of YB-1 to stimulate PARP1 (Fig. 1. A, B). Also, we tested whether YB-1 modulates PARP1 activity when mononucleosomes were used as substrates. For these experiments, we used mononucleosomes containing

undamaged 147-bp DNA or damaged 147-bp DNA with one-nucleotide gap, and free undamaged 147-bp, or damaged 147-bp DNAs with one-nucleotide gap were used as control. We observed that YB-1 stimulates of PARP1 activity in the case of both mononucleosomes and free 147-bp DNAs (Fig. 1 C).

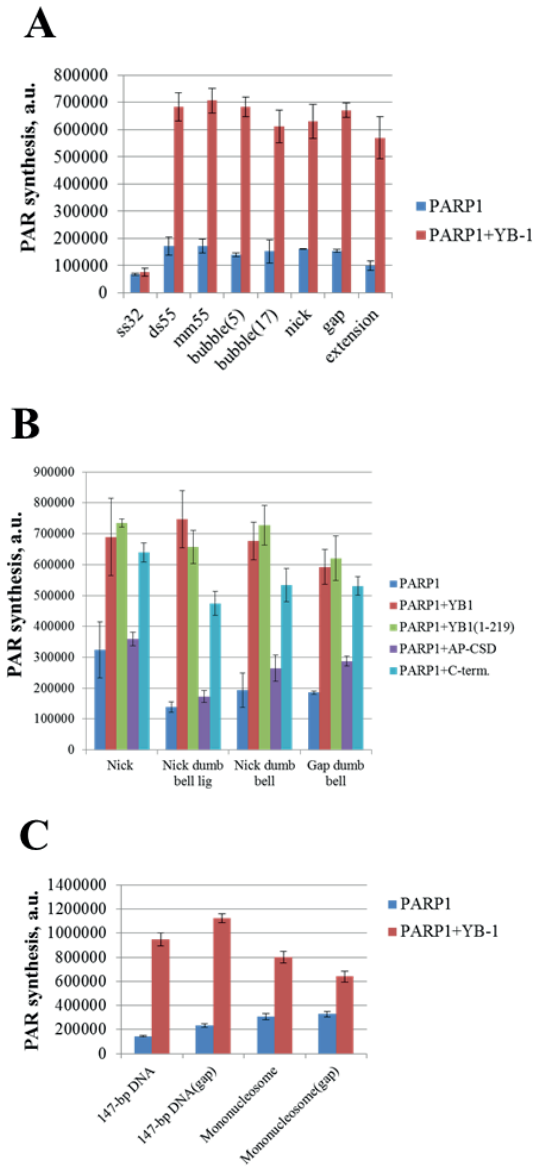


Fig. 1. A) Quantification of PARP1 activation. 100 nM PARP1 was incubated with 100 nM DNA substrate in the presence of 1600 nM YB-1 (where indicated), ^{32}P-PAR-modified proteins were TCA-precipitated and counted. B) Quantification of PARP1 activation. 100 nM PARP1 was incubated with 100 nM DNA substrate in the presence of 1600 nM YB-1 or its mutant (where indicated), ^{32}P-PAR-modified proteins were TCA-precipitated and counted. C) Quantification of PARP1 activation. 100 nM PARP1 was incubated with 100 nM substrate in the presence of 1600 nM YB-1 (where indicated), ^{32}P-PAR-modified proteins were TCA-precipitated and counted.

Conclusions

Thus, YB-1 might function as regulator of DNA damage-dependent PARP1 activation, acting on broad spectrum of DNA structures increasing the overall yield of poly(ADP-ribosyl)ation reaction.

Acknowledgment

The work was supported by RSF (project No. 19-14-00107) and RFBR (project No. 17-00-00097).

References

[1] T. Lindahl. Instability and decay of the primary structure of DNA. Nature, vol. 362, pp. 709–715, 1993.

[2] R. Chaudhuri, A. Nussenzweig. The multifaceted roles of PARP1 in DNA repair and chromatin remodeling. Nat. Rev. Mol. Cell Biol., vol. 18(10), pp. 610-621, 2017.

[3] P. Hassa, S. Haenni, M. Elser, M. Hottiger. Nuclear ADP-ribosylation reactions in mammalian cells: Where are wetoday and where are we going? Microbiol. Mol.Biol. Rev., vol. 70, pp. 789-829, 2006.

[4] Y. Duan et al. PARylation regulates stress granule dynamics, phase separation, and neurotoxicity of disease-related RNA-binding proteins. Cell Res., vol. 29(3), pp. 233-247, 2019.

[5] E. Alemasova, O. Lavrik. Poly(ADP-ribosyl)ation by PARP1: reaction mechanism and regulatory proteins. Nucleic Acids Res., vol. 47(8), pp. 3811-3827, 2019.

[6] X. Sun. Sam68 Is Required for DNA Damage Responses via Regulating Poly(ADP-ribosyl)ation. PLoS Biol., vol. 14(9), 2016.

[7] A. Fischbach, A. Mangerich et al. The C-terminal domain of p53 orchestrates the interplay between non-covalent and covalent poly(ADP-ribosyl)ation of p53 by PARP1. Nucleic Acids Res., vol. 46(2), pp. 804-822, 2018.

[8] E. Maltseva, O. Lavrik. Replication protein A as a modulator of the poly(ADP-ribose)polymerase 1 activity. DNA Repair (Amst), vol. 72, pp. 28-38, 2018.

[9] J. Gagne, G. Poirier et al. Proteome-wide identification of poly(ADP-ribose) binding proteins and poly(ADP-ribose)-associated protein complexes. Nucleic Acids Res., vol. 36(22), pp. 6959-6976, 2008.

[10] E. Alemasova, O. Lavrik et al. Poly(ADP-ribosyl)ation as a new posttranslational modification of YB-1. Biochimie, vol. 119, pp. 36-44, 2015.

[11] E. Alemasova, O. Lavrik et al. The multifunctional protein YB-1 potentiates PARP1 activity and decreases the efficiency of PARP1 inhibitors. Oncotarget, vol. 9(34), pp. 23349-23365, 2018.

DOI 10.18699/BGRS/SB-2020-364

Helicase XPD *Chaetomium thermophilum* as a functional analogue of human XPD

Irina Petruseva
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
irapetru@niboch.nsc.ru

Olga Lavrik
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
lavrik@niboch.nsc.ru

Janette Kappenberger
Rudolf Virchow Center for Experimental Medicine, University
Wurzburg, Germany
jeannette.kappenberger@uniwuerzburg.de

Natalia Lukyanchikova
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
lunata9@yandex.ru

Jochen Kuper
Rudolf Virchow Center for Experimental Medicine, University
Wurzburg, Germany
jochen.kuper@virchow.uniwuerzburg.de

Rashid Anarbaev
Institute of Chemical Biology and Fundamental Medicine SB RAS
Novosibirsk, Russia
anarbaev@niboch.nsc.ru

Caroline Kisker
Rudolf Virchow Center for Experimental Medicine, University
Wurzburg, Germany
caroline.kisker@virchow.uniwuerzburg.de

***Abstract* — XPD helicase performs the verification step of damage recognition during NER. The detailed investigation of this step will help answer the question of interrelation of the chemical and spatial structure of the damage and its NER removal efficiency. The solution of this problem is also fundamentally important in the context of the induced cytotoxic DNA damages safekeeping and NER inhibitors search. Structurallyfunctional basis of the hXPD operatio as a damage verifier in NER n in NER is not quite clear. *Chaetomium thermophilum* (ct) proteins have recently been proposed as more accessible and stable analogues of human ones. This study is a systematic analysis of the interaction of ctXPD and the more complete ctXPD-p44 system with DNA, which contain structurally different bulky lesions with previously estimated NER repair efficiencies. The data obtained indicate the similarity of the mechanisms for verifying damage by helicases ctXPD and hXPD and prove the possibility of further use of ctXPD in extensive biochemical and structural studies of the mechanism for verifying damage.**

Keywords* — *DNA repair, XPD helicase, bulky damages recognition/verification

Motivation and Aim

Motivation

The efficient functioning of DNA repair systems is a critical cell function, necessary to maintain genome stability. The efficiency of bulky damages elimination with nucleotide excision repair system (NER) is largely determined by the damage recognition efficiency. The key step of recognition damage verification with XPD and interrelation with the efficiency of damage eliminations is not sufficiently understood up to date.

Aim

The understanding of the mechanism and structure of hXPD is based on the results of biochemical studies, comparative mutagenesis, and studies of crystals of archaeal helicases which contain all domains necessary for XPD activities, however, operate as isolated monomers. A close Chaetomium thermophilum (ct)XPD functions as part of TFIIH available for isolation [1]. To prove the legitimacy of ctXPD as functional model of hXPD we have used the series of model DNA early characterized as NER substrates [2, 3].

Methods

XPD and auxillary ct protein p44 were expressed in bacterial system and isolated mainly as in [1]. To characterize ctXPD - damaged DNA interaction we have applied fluorescence spectroscopy and photoaffinity modification. CtXPD–damaged DNA affinity evaluation was based on fluorescence anisotropy measurements. The helicase activity was measured using the designed for like

DNA substrates: unmodified one and bearing exo-N-{2-[N(4-azido-2,5-difluoro-3-chloropyridin-6-yl)-3aminopropionyl]aminoethyl}-deoxycytidinine (FapdC); exoN-[4-(4-azido-2,3,5,6,-tetrafluorobenzylidenehydrazinocarbonyl)-butylcarbamoyl]- 2'-deoxycytidine (FabgdC); N-[6-(9-antracenylcarbomoyl)hexanoyl]-3-amino-1,2-propandiol (nAnt), N-[6-(5(6)-fluoresceinylcarbomoyl)hexanoyl]-3amino-1,2-propandiol (nFlu) in translocated strand. When BHQ1 quencher was introduced in 3'-end of translocated strand, FAM group was at 5-end of the leaving strand; when at 3'-end of translocated strand was Cy3 group, at 5'-end of leaving stand was Dabcyl group. DNA probes for photocrosslinking were designed using photoactivable damages FapdC and FabgdC.

Results

The analysis of interaction of ctXPD with the set of the model DNAs revealed interconnection of the efficiency of NER performed elimination of structurally diverse bulky damages and their recognition with ctXPD. The increased ctXPD affinity for DNA bearing efficiently repairable damages (Fig.1) in row with the results of helicase activity measurements and XPD photocrosslinking results indicate the similarity of the mechanisms of damage verification by ctXPD and hXPD helicases.

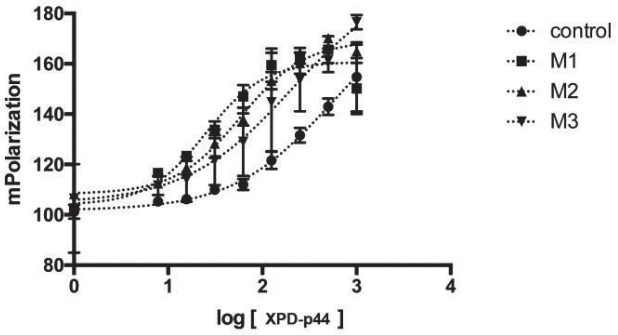


Fig. 1. XPD-p44 binding with DNA. Control – non modified DNA, M1 – nFlu-DNA; M2 – nAnt DNA; M3 - FapdC-DNA.

The data obtained in the given study prove legitimacy of ctXPD as the model for further investigations of bulky DNA damages verification during NER in higher eucaryotes and indicate correlation among efficiency of bulky damage elimination and helicase activity.

ACKNOWLEDGMENT

Supported by RFBR grant No. 19-04-00018.

REFERENCES

[1] Kuper J., Braun C., Elias A. et al. // pLoS Biol. 2014. V. 12 No. 9. p. 1-13.predictions and experiments. Progress in Biophysics and Molecular Biology. 107(1): 81-89*)*

[2] Evdokimov A., Petruseva I., Tsidulko A. et al. // Nucleic Acids Res. 2013. V. 41. No. 12. p. 1-10.Lashin S.A., Matushkin Yu.G. (2012) Haploid evolutionary constructor: new features and further challenges. In Silico. Biol. 11(3): 125-135.

[3] Evdokimov A., Tsidulko A., Popov A. et al. // DNA repair. 2018. V. 41. No. 61. p. 86-98.

DOI 10.18699/BGRS/SB-2020-365

Regulated cell death in Hetherocephalus glaber

Aleksei Popov
ICBFM SB RAS, Novosibirsk, Russia
depolice@mail.ru

Aleksei Evdokimov
ICBFM SB RAS, Novosibirsk, Russia
an_evdokimov@mail.ru

Irina Petruseva
ICBFM SB RAS, Novosibirsk, Russia
irapetru@niboch.nsc.ru

Svetlana Romanenko
IMCB SB RAS, Novosibirsk, Russia
rosa@mcb.nsc.ru

Vladimir Trifonov
IMCB SB RAS, Novosibirsk, Russia
vlad@mcb.nsc.ru

Olga Koval
ICBFM SB RAS, Novosibirsk, Russia
o.koval@niboch.nsc.ru

Inna Lavrik
Institute of Experimental Internal Medicine, Magdeburg, Germany
inna.lavrik@med.ovgu.de

Elena Ryabchikova
ICBFM SB RAS, Novosibirsk, Russia
lenryab@niboch.nsc.ru

Olga Lavrik
ICBFM SB RAS, Novosibirsk, Russia
lavrik@niboch.nsc.ru

***Abstract — Heterocephalus glaber*, or the naked mole rat, is the longest-living rodent species that is extraordinary resistant to cancer and aging-related diseases. The molecular basis for the unique phenotypic traits of naked mole-rat has been extensively studied recently, but the role of the programmed cell death in longevity and cancer resistance of naked mole rat is still insufficiently understood. Regulated cell death is a mechanism restricting the proliferation of damaged or premalignant cells which counteracts aging and oncotransformation. In this study, the systematic analysis of naked mole rat fibroblasts cells to undergo DNA-damage-induced apoptosis has been carried out using conventional methods of apoptosis detection. Skin fibroblasts from the naked mole rat were shown to be more resistant to several types of stress effects compared with fibroblasts from the *Mus musculus*. Naked mole rat cells also exhibit a limited apoptotic response and seem to undergo necrotic cell death under severe stress.**

Keywords — aging, longevity, naked mole rat, stress resistance, apoptosis

Motivation and Aim

Motivation.

It was previously demonstrated that vascular endothelial cells in naked mole rat (NMR) are highly resistant to apoptotic stimuli such as H2O2 treatment or heat exposure compared with mouse cells [1], the ability of NMR cells to undergo apoptosis in response to stress has been studied insufficiently, the details of the processes leading to apoptosis in NMR cells remain unclear. Also, no comprehensive studies of the regulated cell death (RCD) in NMR and its role in cancer development have been conducted yet.

Aim

To determine the features of the regulated cell death in Heterocephalus glaber, we have carried out a systematic comparison of DNA-damage-induced apoptosis in skin fibroblasts from Heterocephalus glaber and Mus musculus.

Methods

Experiments were provided using NMR skin cell line and 3T3 mouse cell line. Three compounds that are well known to induce three different types of DNA damage agents were used as genotoxic stress agents: methyl methanesulfonate (MMS) [2], 5-fluorouracil (5FU) [3], and etoposide [4]. The resistance of NMR and mouse fibroblasts to stress induced by selected reagents was evaluated using cell metabolic activity assay. FITC Annexin V Apoptosis detection kit and FACSCanto II (BD Biosciences, San Jose, USA) flow cytometer were used to detect cell death. To perform differential study of the development of apoptotic cascade dynamics, caspase activities were measured using the Caspase-Glo 3/7 assay kit according to the manufacturer's instructions (Promega, Madison, WI). Luminescence of the samples was measured using an Infinite M200 plate reader (Tecan, Research Triangle Park, USA). Electron microscopy of ultrathin sections was used to delineate between the types of cell death in NMR and mouse cells. Statistical analysis was performed using Statistica 10 software.

Results

The comparative studies have revealed that naked mole rat cells are highly resistant to all types of exposures used. The decline in cellular metabolic activity and cell death take place at much higher concentrations of cytotoxic agents. Unlike mouse cells, naked mole rat cells are especially resistant to some proapoptotic reagents (e.g., 5-fluorouracil). The data obtained using various approaches (flow cytometry, electron microscopy, and measuring caspase activity during incubation in the presence of reagents) give grounds for assuming that efficiency of apoptosis activation in the naked mole rat is lower compared to that in mice. When exposed to high-dose toxic agents, the naked mole rat cells preferentially undergo necrotic rather than apoptotic death (unlike mouse cells exposed to the same agents).

Acknowledgment

This study was supported by Russian Science Foundation project No. 19-74-10056.

References

[1] Labinskyy N., Csiszar A., Orosz Z., et al. Comparison of endothelial function, O2• and H2O2 production, and vascular oxidative stress resistance between the longest-living rodent, the naked mole rat, and mice. Am. J. Physiol. Heart. Circ. Physiol. 2006; 291: 2698-2704.

[2] Wyatt M.D., Pittman D.L. Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. Chemical Research in Toxicology. 2006;19(12):1580–94.

[3] Longley D.B., Harkin D.P., Johnston P.G. 5-fluorouracil: mechanisms of action and clinical strategies. Nat. Rev. Cancer. 2003; 3:330–338.

[4] Hande K.R. Etoposide: four decades of development of a topoisomerase II inhibitor. Eur. J. Cancer. 1998; 34 (10): 1514–21.

DOI 10.18699/BGRS/SB-2020-366

Nucleotide excision repair proteins and PARP1/PAR interplay regulats protein assembly on damaged DNA

Nadejda Rechkunova
Lab of Bioorganic Chemistry of Enzymes
ICBFM SB RAS
Novosibirsk, Russia
nadyarec@niboch.nsc.ru

Ekaterina Maltseva
Lab of Bioorganic Chemistry of Enzymes
ICBFM SB RAS
Novosibirsk, Russia

Yuliya Krasikova
Lab of Bioorganic Chemistry of Enzymes
ICBFM SB RAS
Novosibirsk, Russia

Maria Sukhanova
Lab of Bioorganic Chemistry of Enzymes
ICBFM SB RAS
Novosibirsk, Russia

Olga Lavrik
Lab of Bioorganic Chemistry of Enzymes
ICBFM SB RAS
Novosibirsk, Russia

Abstract — **Nucleotide excision repair (NER) is one of the major mechanisms to prevent genomic DNA instability. This process removes a wide range of lesions distorting the double helix and bulky chemical adducts resulting from environmental factors or chemotherapeutic agents. The coordination of the assembly of the NER complexes and the sequential individual reactions is achieved through multiple DNA-protein and protein-protein interactions. The interactions of key protein factors of the NER process, xeroderma pigmentosum protein C in the complex with RAD23B (XPC-RAD23B), xeroderma pigmentosum protein A (XPA), and replication protein A (RPA) with DNA mimicking NER intermediates and their modulation by the activity of poly(ADP-ribose)polymerase 1 (PARP1) and poly(ADP-ribose) have been analyzed. Using several biochemical approaches we have analyzed the influence of PARP1 and PAR synthesis on the interaction of the NER factors with damaged DNA. XPC-RAD23B, RPA and XPA are the targets of poly (ADP-ribosyl)ation catalyzed by PARP1, bind free poly(ADP-ribose) with an affinity depending on the PAR strand length and influence on PAR synthesis. Both XPC-RAD23B and XPA stimulate PAR synthesis in all experimental conditions whereas effect of RPA depends on the structure of DNA used for PARP1 activation. RPA inhibits PAR synthesis in the presence of ssDNA and stimulates it in the presence of DNA duplex, in particular containing a nick or a gap. RPA stimulation effect can be due to acceleration of the replacement of auto-(ADP-ribosyl)ated PARP1 in the complex with DNA by unmodified one. The data obtained suggest participation of RPA in regulation of PARP1 activity, which, in its turn, could play an important role in regulation of not only NER, but also replication-coupled repair and other DNA repair pathways. Our study clearly shows that PARP1 can be regarded as the universal regulator in DNA repair processes.**

Keywords *— **nucleotide excision repair, poly(ADP-ribose)polymerase 1, DNA-protein interactions, protein-protein interactions***

Introduction

Genetic stability of living organisms is substantially maintained by the action of the DNA repair systems. Regulation of the activity of the DNA repair systems is required to efficiently protect genomic DNA under genotoxic stress. One of the key mechanisms modulating the activity of DNA repair systems is the synthesis of poly(ADP-ribose) attached covalently to some repair proteins. Cellular DNA damage response by PAR synthesis is mainly mediated by PARP1, which is the most abundant and well-studied member of the PARP family [1, 2]. Poly(ADP-ribosyl)ation targets are certain nuclear proteins including PARP1 itself. Poly(ADP-ribosyl)ation of the proteins can modulate their interactions with DNA due to the linkage to the negatively charged PAR [3] and can also be regarded as a signal of DNA damage. The PARylation reaction is reversible: PAR is cleaved by the poly(ADP-ribosyl)glycohydrolase (PARG) that additionally regulates the cellular level of protein poly(ADP-ribosyl)ation and PAR synthesis. Here we analyzed interactions of key protein factors of the NER process, xeroderma pigmentosum protein C in the complex with RAD23B (XPC-RAD23B), xeroderma pigmentosum protein A (XPA), and replication protein A (RPA) with DNA mimicking NER intermediates and their modulation by the activity of PARP1 and poly(ADP-ribose).

Results and Discussion

PARP1 modulates XPC-RAD23B interaction with DNA via poly(ADP-ribosyl)ation)

The XPC-RAD23B heterodimer is the major factor responsible for the primary damage recognition in DNA and initiating assembly of the NER complex. We found that electrophoretic mobility of XPC-RAD23B-DNA complex decreases in the condition of PAR synthesis and it is restored by the PARG treatment. It might correspond to formation of the covalent or noncovalent complexes of XPC-RAD23B with PAR [4].

Using ^{32}P-labeled NAD$^+$ and immunoblotting analyses we have shown that both subunits of XPC-RAD23B heterodimer are PARylated by PARP1 in the presence of damaged DNA duplex. Direct interaction of these proteins was demonstrated in experiments with the fluorescence labeled PARP1. To demonstrate XPC-RAD23B PARylation relevance to the NER process we analyzed PARP1 activation by UV-induced DNA damages. We demonstrated that XPC-RAD23B is poly(ADP-ribosyl)ated by PARP1 in response to UV irradiation of plasmid DNA in dose dependent manner. These findings are in line with observations that UV irradiation triggers both stimulation of PAR synthesis and association of PARP1 with UV photolesions in chromatin [5]. Free or covalently bound to other proteins (including PARP1) PAR might serve as target for the attraction of XPC-RAD23B to lesions in DNA. PARP1 also influences XPA binding to DNA but in

The work was supported by RSF (project No. 19-14-00107) and RFBR (project No. 18-04-00596).

different manner. PARP1 displaces XPA from the complex with DNA and PAR synthesis inhibits XPA binding to DNA.

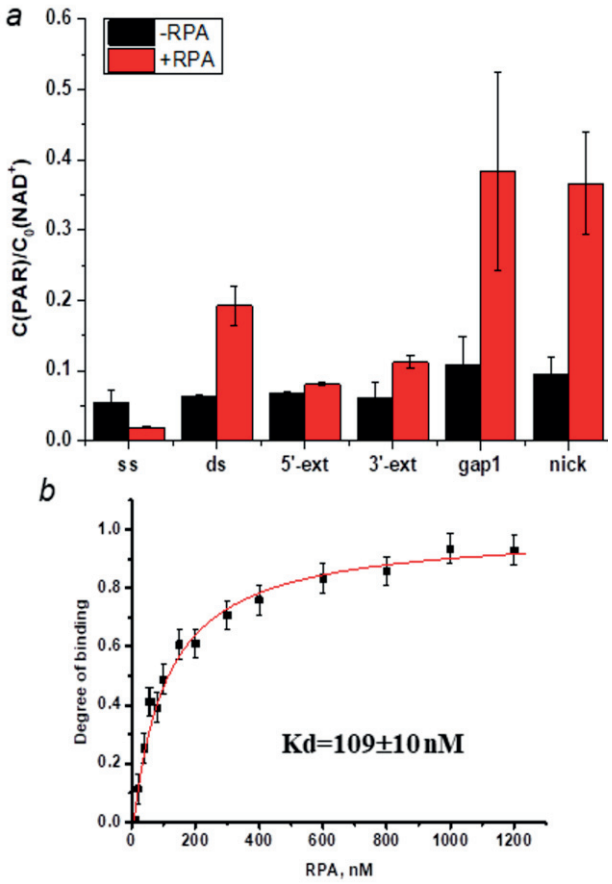


Fig. 1. RPA influences the PAR synthesis dependently on DNA structure (a) and physically interacts with PARP1.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

RPA modulates PARP1 activity in DNA dependent manner

RPA is a major eukaryotic single-stranded DNA binding (SSB) protein, which is indispensable in DNA replication, repair, and recombination. RPA is heterotrimer consisting of p70, p32, and p14 subunits. We examined the efficiency of RPA PARylation by PARP1 in the presence of different DNA structures: ssDNA, double-stranded DNA (dsDNA), DNA duplexes with a 3′- or 5′-single-stranded extension (3′-ext-DNA and 5′-ext-DNA, respectively), DNA duplex with a single-nucleotide gap (gap1-DNA), and DNA duplex with a single-strand break (nick-DNA).

For all DNA structures, the major reaction product corresponded to the product of PARP1 automodification. Modification of RPA, presumably the large p70 subunit, was also detected with the efficiency depended on the DNA structure. The maximal level of RPA modification was observed for the 3′-ext-DNA. For this structure, PARylation of p32 subunit was also displayed. In the case of ssDNA, efficiency of the proteins modification was decreased significantly in the presence of RPA.

RPA is not only a target PARylation but it also changes the efficiency of PARP1 automodification. Therefore, RPA can influence PARP1 activity. To check this possibility, we tested the effect of RPA on the PAR synthesis catalyzed by PARP1 in the presence of different DNA. RPA was shown to significantly increase the efficiency of PAR synthesis when the DNA structures with blunt ends were used. This effect was most visible with nick-DNA and gap1-DNA. For DNA structures with protruding single-stranded parts, RPA either had no significant effect (5′-ext-DNA, and 3′-ext-DNA) or inhibited PAR synthesis (ssDNA) (Fig. 1 *a*). RPA-PARP1 interaction in the absence of DNA was detected by fluorescent titration (Fig. 1 *b*).

RPA also influences leaving of PARP1 from DNA during PARylation. DNA-bound RPA can facilitate interaction of unmodified PARP1 with DNA. As a result, the replacement of modified PARP1 molecules by the unmodified ones on DNA accelerated, and the length of synthesized PAR was shortened [6]. The recent data on PAR synthesis at sites of DNA replication in normal S phase cells allowed indicating PARP1 is a sensor of unligated Okazaki fragments [7]. This observation is in line with our result that PARP1 interacts with RPA on nick/gap1-DNA.

Conclusions

The data obtained in this study clearly show that PARP1 interacts with and influences DNA binding of key NER factors which modulate its activity. Therefore, PARP1 can be regarded as the universal regulator in DNA repair processes.

Acknowledgment

The authors are grateful to Dr. Rashid Anarbaev for the contribution in preparation of the fluorescence labeled PARP1 and RPA and fluorescence anisotropy measurements.

References

[1] N. Ogata, K. Ueda, M. Kawaichi, and O.Hayaishi, "Poly(ADP-ribose) synthetase, a main acceptor of poly(ADP-ribose) in isolated nuclei", J. Biol. Chem., vol. 256, pp. 4135-4137, May 1981.

[2] D. D'Amours, S. Desnoyers, I. D'Silva, and G.G. Poirier, "Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions", Biochem. J., vol. 342, pp. 249-268, September 1999.

[3] J.M. de Murcia, C. Niedergang, C. Trucco, M. Ricoul, B. Dutrillaux, M. Mark, F. J. Oliver, M. Masson, A. Dierich, M. LeMeur, C. Walztinger, P. Chambon, and G. de Murcia, "Requirement of poly(ADP-ribose) polymerase in recovery from DNA damage in mice and in cells", Proc. Natl. Acad. Sci. U S A, vol. 94, pp. 7303–7307, July 1997.

[4] E.A. Maltseva, N.I. Rechkunova, M.V. Sukhanova, and O.I. Lavrik, "Poly(ADP-ribose) Polymerase 1 Modulates Interaction of the Nucleotide Excision Repair Factor XPC-RAD23B with DNA via Poly(ADP-ribosyl)ation", J. Biol. Chem., vol. 290, pp. 21811-21820, September 2015.

[5] M.D. Vodenicharov, M.M. Ghodgaonkar, S.S. Halappanavar, R.G. Shah, and G.M. Shah, "Mechanism of early biphasic activation of poly(ADP-ribose) polymerase-1 in response to ultraviolet B radiation" J. Cell. Sci., vol. 118, pp. 589–599, February 2005.

[6] E.A. Maltseva, Y.S. Krasikova, M.V. Sukhanova, N.I. Rechkunova, and O.I. Lavrik, "Replication protein A as a modulator of the poly(ADP-ribose)polymerase 1 activity", DNA Repair, vol. 72, pp. 28-38, December 2018.

[7] H. Hanzlikova, I. Kalasova, A.A. Demin, L.E. Pennicott, Z. Cihlarova, K.W. Caldecott. "The importance of poly(ADP-ribose) polymerase as a sensor of unligated Okazaki fragments during DNA replication", Mol. Cell., vol. 71, pp. 319-331, 2018.

DOI 10.18699/BGRS/SB-2020-367

PARP1 activation promotes FUS translocation to cytoplasm and incorporation into stress granules

Anastasia Shavkatovna Singatulina
LBCE, Institute of Chemical Biology
and Fundamental Medicine SB RAS
Novosibirsk, Russia

Pastré David
SABNP, INSERM U1204
Université Paris-Saclay
Evry, France

Ahmed Bouhss
SABNP, INSERM U1204
Université Paris-Saclay
Evry, France

Bénédicte Desforges
SABNP, INSERM U1204
Université Paris-Saclay
Evry, France

Maria Vladislavovna Sukhanova
LBCE, Institute of Chemical Biology
and Fundamental Medicin SB RAS
Novosibirsk, Russia

Loic Hamon, SABNP
INSERM U1204
Université Paris-Saclay
Evry, France

Olga Ivanovana Lavrik
LBCE, Institute of Chemical Biology
and Fundamental Medicine SB RAS
Novosibirsk, Russia

***Abstract* — Synthesis of poly(ADP-ribose) catalyzed by nuclear poly(ADP-ribose) polymerases (PARPs) is one of the earliest events in the cellular response to DNA damages in higher eukaryotes. Over the past few years, a huge body of evidence has been accumulated about the involvement of PARPs in the regulation of not only DNA-dependent processes in the nucleus but of RNA-dependent processes in the cell cytoplasm. Many RNA-binding proteins have been described to interact with poly(ADP-ribose) and (or) undergo post-translational modification through poly(ADP-ribosyl)ation. Among them, fused in sarcoma (FUS) protein was found to function in DNA damage response through PARP-dependent mechanism. In the present study, we show that PARP1 activation leads to the shuttling of FUS from the nucleus to the cytoplasm in HeLa cells after hydrogen peroxide -induced oxidative stress. We have observed that stress granule assembly depends on PARP1 activity and found a correlation between the nucleocytoplasmic shuttling of FUS and its accumulation in stress granules, which may in turn promote stress granule assembly. Thus, FUS appears to mediate stress granules formation downstream of PARP1 activation in response to oxidative DNA damage.**

Keywords* — *fused in sarcoma, poly(ADP-ribosyl)ation, stress granules

Introduction

The cellular response to genotoxic stress occurs through a series of regulatory processes that not only control DNA repair but also can cause changes in the level of gene expression [1]. In response to genotoxic impact, DNA damage sensors activate complex signaling networks that affect a number of cellular processes [2]. Nuclear PARPs are sensors of DNA strand breaks and play a key role in the regulation of DNA damage response [3]. Upon binding to DNA breaks, PARPs catalyze the synthesis of poly(ADP-ribose) (PAR) covalently attached to PARPs itself or an acceptor protein using NAD+ as substrate for the post-translation modification [4]. PAR covalently attached to acceptor proteins and can be subsequently hydrolysed to form free mon(ADP-ribose) or oligo(ADP-ribose) by PAR glycohydrolase (PARG) [5]. Recently, PAR has been identified as stress granules (SG) component, and PAR-dependent regulation of nuclear-cytoplasmic transport of set of RNA-binding proteins (RBPs) has been shown to involve in the formation of SG [6,7]. SGs are ribonucleoprotein complexes containing untranslated mRNA and RBPs that appear in the cytoplasm of mammalian cells upon the dissociation of polysomes, which occurs generally after various stresses including osmotic stress, oxidative stress, viral infection, hypoxia and hyperthermia [8]. SG dynamics is thought to play a major role in timely mRNA-related processes such as mRNA storage, sorting and protection from degradation. It was shown that PAR produced by nuclear PARPs such PARP1 and PARP2 can function as signaling molecules, driving the translocation of stress granule components from specific subcellular compartments toward the nascent stress granule [9]. In particular, FUS (Fused in Sarcoma) has focused our attention since it is one of the most abundant nuclear RBPs which can be PARylated, interact with PAR and be the component of SGs [10]. Recently, we have found that oxidative stress induces FUS translocation which is dependent on PARP1 and PARG activities [11]. Here, we have analyzed PARP1-dependent formation and composition of stress granules in cells under oxidative stress. We have shown that in hydrogen peroxide -treated cells stress granule assembly was dependent on PARP1 activation despite the absence of PAR in these granules. We have found a correlation between the nucleocytoplasmic shuttling of FUS and its accumulation in stress granules, which may in turn promote stress granule assembly. Thus, the relocation of FUS in the cytoplasm may participate to an adapted translational response to DNA damages but its accumulation in stress granules may be an intermediary step between PARP1 activation and the aggregation of RNA-binding proteins in neurodegenerative diseases.

Results

To test whether PARP1 activation in nuclear affect the assembly of SGs, we examined the PAR synthesis and SG formation in HeLa cells under oxidative stress induced by arsenite or hydrogen peroxide and puromycin (Fig. 1a). Arsenite is widely used to induce oxidative stress and SGs formation in cells, however, arsenite exposure does not or weakly activates PARP1 in contrast to hydrogen peroxide (Fig. 1a).

In arsenite-treated cells, chemical inhibition of PARP1 by Olaparib does not influence the formation of SGs (Fig. 1b). Nevertheless, SG formation in cells treated with hydrogen peroxide/puromicine is prevented by olaparib (Fig. 1b). We then considered the possibility that PARP1 activation and PAR synthesis could promote stress granule assembly

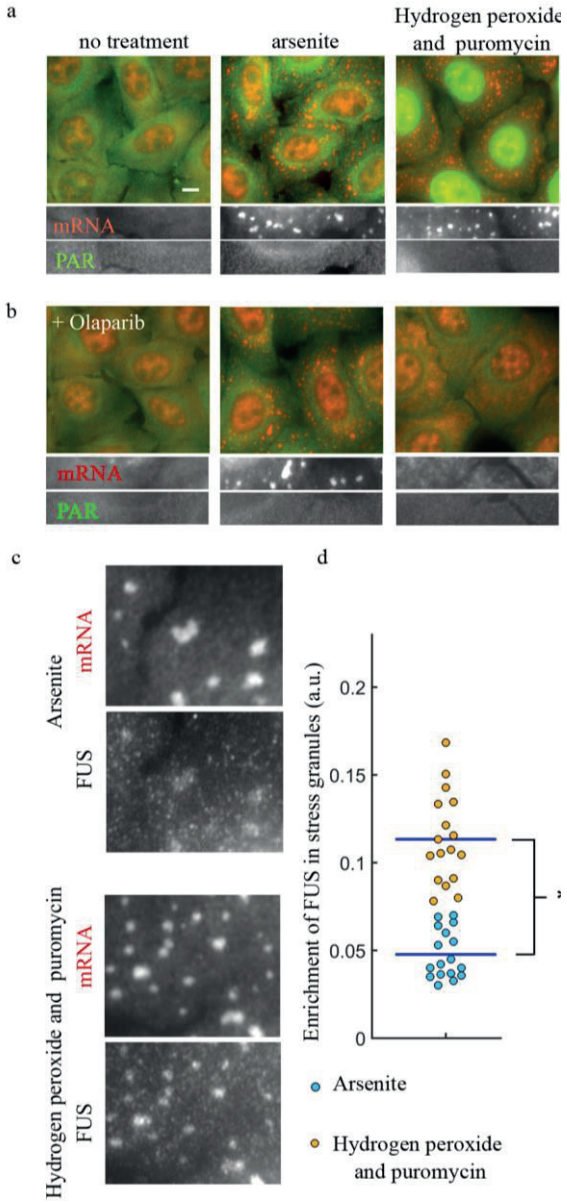


Fig. 1. PARP1 activation is required for the assembly of FUS-rich stress granules after hydrogen peroxide and puromycin treatment.
a and b. Hela cells were pretreated without (a) or with (b) olaparib for 15 min, then exposed to either arsenite (300 µM, 60 min) or hydrogen peroxide and puromycin (300 µM and 2.5 µg/mL, respectively, 90 min). mRNA was detected by using cy3-labeled poly(T)-probe (red). anti–PAR antibodies fluorescence (green). Scale bar: 10 µm.
c. Fluorescence images of stress granules in HeLa cells obtained under indicated conditions.
d. Ratio of FUS to mRNA fluorescence integrated intensity in stress granules under indicated conditions. Anti FUS antibodies were used. $p < 0.05$, **; paired t-test, n=17.

indirectly through the translocation of FUS. To explore this idea, we measured the relative enrichment of FUS in SGs in cells exposed to either arsenite or hydrogen peroxide in combination with puromycin (Fig. 1c and d). We observed that the accumulation of FUS in SGs is more pronounced in hydrogen peroxide/puromycin – than in arsenite-treated cells (Fig. 1d).

Conclusions

We hypothesize that FUS could mediate stress granules formation downstream of PARP1 activation in response to oxidative DNA damage. The PARP-dependent relocation of FUS in the cytoplasm may participate in an adapted translational response to DNA damages but FUS accumulation in stress granules may be also an intermediary step between PARP1 activation and the aggregation of RNA-binding proteins such FUS that occurs in neurodegenerative diseases.

Acknowledgment

The work was supported by RFBR (project no. 18-04-00882) and Genopole Evry.

References

[1] M. Dutertre, S. Lambert, A. Carreira, M. Amor-Guéret, and S. Vagner, “DNA damage: RNA-binding proteins protect from near and far”, Trends Biochem. Sci., vol. 39, pp. 141-149, March 2014.

[2] L. von Stechow, and J. V. Olsen, “Proteomics insights into DNA damage response and translating this knowledge to clinical strategies”, Proteomics, vol. 17, pp. 1600018, December 2017.

[3] M. L. Meyer-Ficca, R.G. Meyer, E.L. Jacobson, and M. K. Jacobson, “Poly (ADP-ribose) polymerases: managing genome stability”, Int. J. Biochem. Cell Biol., vol. 37, pp. 920-926, May 2005.

[4] D. d'Amours, S. Desnoyers, I. d'Silva, and G.G. Poirier, “Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions”, Biochem. J., vol. 342, pp. 249-268, September 1999.

[5] M. E. Bonicalzi, J. F. Haince, A. Droit, and G.G. Poirier, “Regulation of poly(ADP-ribose) metabolism by poly (ADP-ribose) glycohydrolase: where and when?”, Cell. Mol. Life Sci., vol. 62, pp. 739-750, April 2005.

[6] A. K. Leung, S. Vyas, J. E. Rood, A. Bhutkar, P. A. Sharp, and P. Chang, “Poly(ADP-ribose) regulates stress responses and microRNA activity in the cytoplasm”, Mol. Cell, vol. 42, pp. 489-499, May 2011.

[7] M. Isabelle, J. P. Gagné, I. E. Gallouzi, and G. G. Poirier, “Quantitative proteomics and dynamic imaging reveal that G3BP-mediated stress granule assembly is poly(ADP-ribose)-dependent following exposure to MNNG-induced DNA alkylation”, J. Cell Sci., vol. 125, pp. 4555-4566, October 2012.

[8] J. R. Buchan, and R. Parker, “Eukaryotic stress granules: the ins and outs of translation”, Mol. Cell, vol. 36, pp. 932-941, December 2009.

[9] A. K. Leung, “Poly(ADP-ribose): an organizer of cellular architecture”, J. Cell Biol., vol. 205, pp. 613-619, June 2014.

[10] R. R. Sama, C. L. Ward, and D. A. Bosco, “Functions of FUS/TLS from DNA repair to stress response: implications for ALS”, ASN Neuro, vol. 6, pp. 1759091414544472, August 2014.

[11] A. S. Singatulina, L. Hamon, M. V. Sukhanova, B. Desforges, V. Joshi, A. Bouhss, O. I. Lavrik, and D. Pastré, “PARP-1 activation directs FUS to DNA damage sites to form PARG-reversible compartments enriched in damaged DNA” Cell Rep., vol. 27, pp. 1809-1821, May 2019.

DOI 10.18699/BGRS/SB-2020-368

Nucleosome assembling: quick-time reconstitution protocol

Alexander Ukraintsev
ICBFM SB RAS, Novosibirsk, Russia
a.ukraintsev@g.nsu.ru

Michael Kutuzov
ICBFM SB RAS, Novosibirsk, Russia
kutuzov.mm@mail.ru

Tatyana Kurgina
ICBFM SB RAS, Novosibirsk, Russia
t.a.kurgina@gmail.com

Ekaterina Belousova
ICBFM SB RAS, Novosibirsk, Russia
rina@niboch.nsc.ru

Svetlana Khodyreva
ICBFM SB RAS, Novosibirsk, Russia
svetakh@niboch.nsc.ru

Olga Lavrik
ICBFM SB RAS, Novosibirsk, Russia
lavrik@niboch.nsc.ru

***Abstract* — In cells, a DNA molecule is compacted in chromatin, and the first level of compaction is nucleosomal. Nucleosome core particles, NCPs, consists of eight core proteins – histones, around which 147 nucleotide pairs of DNA are wound (McGinty R. K. et al 2015, Chemical Reviews). Traditionally, for studies used to nucleosomes reconstituted *in vitro* from histones and DNA structures with a specific sequences that provide a clear positioning of the DNA molecule relative to the histone octamer. Such DNA structures were obtained and characterized by J. Widom and colleagues (Lowary P.T. et al 1998, Journal of molecular biology). The same group of scientists developed a method for reconstitution of NCP *in vitro* in the NaCl gradient using dialysis cells (Thåström A. et al 2004, Methods). However, the method have a several drawbacks: firstly, it requires large consumption of chemical reagents; secondly, the used dialysis buffers are applied repeatedly; thirdly, during the dialysis process high osmotic pressure is created, which can lead to damage to the dialysis cell with subsequent loss of reaction products. That's why the main objective of the presented research was to create a universal method of nucleosome assembly *in vitro*.**

Keywords* — *nucleosome core particle, assembly, chromatin, core histones

Motivation and Aim

Motivation

The *in vitro* nucleosome reconstitution method used by J. Widom and colleagues [1] requires a large number of biochemical reagents. It's well known that to NCP reconstitution the equimolar ratio of DNA to histone octamers is needed. However, different degrees of samples purity can affect the accuracy of determining the concentration of DNA and histones. In turn, this can lead to the formation of high molecular weight complexes during the reconstitution process or to a residual amount of DNA in the final reaction mixture. To resolve this problem a method should be developed that would allow determining the molar ratio of DNA to histone octamers before carrying out the nucleosome reconstitution process. In addition, the method should require relatively low reagent consumption for the buffer systems. It is necessary to improve the method for carrying out nucleosome reconstitution in the gradient of NaCl.

Aim

The aim of this study was to optimize the procedure for reconstitution of NCP from purified histone octamers and various types of synthesized model DNA.

Methods

Nucleosome assembling: quick-time reconstitution. In this approach either radioactive- or fluorescent-labelled DNA should be used. The DNA in final concentration of 0.1 μM was mixed with histones in different concentrations in low-salted buffer (0.2 mM EDTA, 5 mM beta mercaptoethanol, 0.1 % NP-40, 10 mM Tris-HCl pH 7.5 with 10 mM NaCl), incubated for 15 min at 37°C and analyzed by 4 % PAGE under non-denaturing conditions. The minimal DNA:histone ratio corresponding to the mixture with the absence of naked DNA should be used for preparative reconstitution.

Preparative reconstitution. The DNA and histones should be taken in the concentrations determined under previous procedure, mixed with high-salted buffer with 2 M NaCl, and dialyzed against the buffer with gradient of NaCl from 2 M to 250 mM during 6 hours at 4°C with gentle stirring. Gradient dilution of NaCl was carried out by means of a peristaltic pump with an addition of a non-salted buffer. Then the probes were dialyzed against low-salted buffer with 10 mM NaCl overnight at 4°C with gentle stirring.

Analysis of nucleosome reconstitution using Eva-Green dye. The samples of 10 μl containing 10 mM NaCl, 10 mM DTT, 50 mM Tris-HCl pH 8.0, 50 nM nucleosome (the concentration was defined by nucleic acid component) and 0.5x Eva-Green (Biotium) were placed to the 384-well round bottom microplate to estimate the fluorescence intensity. The measurement was performed in microplate reader CLARIOstar (BMG Labtech) at 520 nm. The average value of five independent measurements was normalized to fluorescent level of naked dsDNA and analyzed using MARS Data Analysis Software (BMG Labtech).

Results

The design of DNA substrates was based on nucleotide sequences named clone 603 that were constructed and published by Lowary and Widom [2]. The length of this DNA duplex is 147 nucleotides, which is exactly corresponds to the “core” part of nucleosomes *in vivo*. The feature of this sequence consists in the precise positioning of the histone octamer on DNA during the formation of NCP.

The plasmid DNA encoding the 603 sequence was amplified by PCR. The PCR primers were chemically synthesized to obtain the final DNA containing or not the discriminant mark in the 5'-end of the DNA duplex. Such approach permits to produce the target DNA constructs with the specific damage at the defined position or special label at the 5'-end of the resulting DNA.

First of all, to assemble the nucleosomes correctly it should be pay attention on the molar ratio of the DNA and histone octamer samples. Due to highly basic nature of the histones it is impossible to determine the precise protein concentration in the samples using standard Bradford

protocol. Therefore, the absorbance measurements at 230 nm are commonly used. However, the most part of small molecules absorb at this wavelength. Thus, it is difficult to detect of the protein concentration in the mixture for each sample. We proposed the fast assembling of NCPs in low-salted nucleosome buffer under titration of the DNA-histone ratio. The analysis in non-denaturing 4 % PAAG allows selection of the nucleic acid to protein proportion providing the correct assembling of NCP. Then, the octamers were initially mixed with DNA in the buffer containing 2 M NaCl and subjected to a gradient dialysis at +4°C for about 6 hours to decrease the salt concentration to 250 mM. Finally, the NCP solution was dialyzed overnight against the buffer with 10 mM NaCl at +4°C. The efficiency of nucleosome assembling was analyzed by classical method with an electrophoretic separation in 4 % PAAG under non-denaturing conditions.

This approach has an obvious disadvantage as far as it doesn't allow the estimation of percent of out-nucleosome DNA directly in the solution under equilibrium conditions. This problem could be solved by utilizing an external fluorescent dye in the solution after the assembling procedure. Accordingly, we suggested an alternative approach for analyzing the efficiency of nucleosome assembling in solution using the Eva-Green dye [3]. It should be mentioned that this method are in full agreement with the data obtained by electrophoretic analysis under non-denaturing conditions.

Using this method and the method described in the article [1] we obtained NCPs with DNA containing uracil in out-warded conformation. Uracil has been introduced in DNA during amplification with plasmid DNA "603" by PCR with U-containing primer. The efficiency of the enzymes UDG and APE1 on both types of substrates didn't differ to each other and corresponde to published data.

ACKNOWLEDGMENT

Supported by the RFBR (20-04-00674).

REFERENCES

[1] Thåström A., Widom J. (2004) Measurement of histone – DNA interaction free energy in nucleosome. Methods. 33(1): 33-44.

[2] Lowary P.T., Widom J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. Journal of Molecular Biology. 276(1): 19-42.

[3] Kutuzov M. M., Kurgina T. A., Belousova E. A., Khodyreva S. N., Lavrik O. I. (2019) Optimization of nucleosome assembling from histones and model DNAs and estimation of the reconstitution efficiency. Biopolymers and Cells. 35(2): 91-98.

DOI 10.18699/BGRS/SB-2020-369

Platinum Polyoxoniobates have a potential as an anticancer agent

Anna V. Yudkina
ICBFM SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
yudkinaanya@gmail.com

Ivan P. Vokhtantsev
ICBFM SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
ivanvohtancev@gmail.com

Maxim N. Sokolov
NIIC SB RAS, Novosibirsk, Russia
caesar@niic.nsc.ru

Pavel A. Abramov
NIIC SB RAS, Novosibirsk, Russia
abramov@niic.nsc.ru

Inga R. Grin
ICBFM SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
grin@niboch.nsc.ru

Dmitry O. Zharkov
ICBFM SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
dzharkov@niboch.nsc.ru

***Abstract* — It is well-known that platinum coordination complexes are widely used anticancer agents. Here we have investigated the biological properties of platinum polyoxometalates — inorganic coordination compounds with a Pt, as a potential antitumor agent.**

Keywords* — *polyoxometalates, DNA polymerases, DNA damage

Motivation and Aim

Motivation

Platinum complexes are among the most commonly prescribed drugs in cancer therapy. However, they have serious restrictions such as significant toxicity to healthy tissues and cell resistance. One strategy to overcome these limitations is the development of new, improved platinum drugs. One of such promising group of compounds could be platinum polyoxometalates — platinum complexes liganded with cluster anions consisting of oxygen atoms and transition metals. Their properties related to a potential use as anticancer drugs have never been studied.

Aim

In this work, we aim to investigate the cytotoxic effect of platinum (IV) polyoxoniobate of the $[Nb_6O_{19}\{Pt(OH)_2\}]_2$ structure containing two platinum centers and two polynuclear Lindqvist type anions [1].

Methods

First, we have tested the effect of platinum (IV) polyoxoniobate on the activity of a number of DNA polymerases belonging to different families (Klenow fragment of *Escherichia coli* DNA polymerase I, bacteriophage RB69 DNA polymerase, human DNA polymerases β and κ, DNA polymerase IV from *Sulfolobus solfataricus*) to investigate *in vitro* possible mechanisms of its cytotoxicity. Moreover, the cytotoxic effect of platinum (IV) polyoxoniobate was investigated in several cell lines. The nature of platinum (IV) polyoxoniobate adduct with DNA was confirmed by MALDI-mass spectrometry.

Results

It has been shown that platinum (IV) polyoxoniobate of the structure $[Nb_6O_{19}\{Pt(OH)_2\}]_2$ demonstrates properties similar to that of cisplatin: it was capable of forming covalent bulky adducts with DNA, which interfere with the primer elongation by the studied DNA-polymerases, and mostly contribute to studied compound cytotoxicity. At the same time, platinum (IV) polyoxoniobate did not act as an inhibitor of DNA polymerases.

Acknowledgment

The authors thank Facilities of the Joint Center for Genomic, Proteomic and Metabolomics Studies Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia.

This study was supported by the Russian Foundation for Basic Research (19-44-543011-r-mol-a) and Novosibirsk region.

References

[1] Abramov P. A. et al. (2015) Platinum polyoxoniobates. Chem. Commun. 51(19): 4021-4023

DOI 10.18699/BGRS/SB-2020-370

Computer-assisted analysis of caspases molecular evolution

Alexey Zamaraev
MSU, Moscow, Russia
a-zamaraev@yandex.ru

Konstantin Gunbin
ICG SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
genkvg@gmail.com

Boris Zhivotovsky
MSU, Moscow, Russia
Karolinska Institutet, Stockholm, Swede
boris.zhivotovsky@ki.se

Gelina Kopeina
MSU, Moscow, Russia
lirroster@gmail.com

***Abstract* — The functions of any protein are driven by their chemical and physical properties, which, in turn, are determined by steric and physico-chemical folding requirements. Therefore, it is expected that replacement of the amino acid tightly interacting with a large number of other amino acids is related to changes in the context of interactions in the protein globule. Recent studies of the protein evolution revealed various signatures of substitution asymmetry and heterotachy. Here, based on reconstruction of ancestral libraries we analyze the substitution asymmetry in the molecular evolution of caspases protein family characterized by huge number of molecular functions.**

Keywords* — *caspases, molecular evolution, substitution asymmetry, ancestral libraries

Motivation and Aim

Up to now, the vast majority of available procedures of reconstruction of ancestral sequences are based on the symmetric single (applied to all protein sites) matrix of amino acid substitution rates. It seems obvious that new software tools for ancestral protein reconstruction, taking into account substitution limitations from the 3D protein structure and from the stability of its folding (for example, ProtASR), should be most useful. However, unfortunately, the experimentally resolved 3D protein structures are still lacking. Another way for substitution asymmetry accounting in the ancestral protein reconstruction is the construction of ancestral libraries [1]. To make the ancestor libraries accurate enough, it has recently been proposed to use the AltAll reconstruction approach. This approach combines all possible alternative states introduced into one protein, and then characterizes this protein with a set of these states [2, 3]. It has been shown that this approach significantly improves the imperfection of individual ancestral sequences reconstructed by Bayesian approach.

Methods

Multiple alignment was done using PROMALS. The best reversible and symmetric models of amino acid substitution rates were selected by IQTree v. 1.5.4 (in our case it is C20+G4). We adjusted initial phylogenetic tree topology using Metazoa species tree from TimeTree DB by the TreeFix v. 1.1.10 software. After this, the branch lengths were re-optimized using IQTree v. 1.5.4 and the best reversible model of amino acid substitution rates. Bayesian sampling of ancestral sequences in each internal tree node was carried out using PhyloBayes v. 4.1, the CAT model [4], and 6 rate categories of sites. In order to construct complete and truncated (using our modified approach, called 'AltAll * N') libraries of ancestral sequences we used AltAll * N' procedure. Our procedure 'AltAll * N' is an iterative rewriting of all probable (with a posteriori probability> 0.1) alternative states in a consensus ancestral sequence of given tree internal node. For example, if there are 3 alternative states in site A and 4 alternative states in site B of ancestral sequence (node) X, then we must rewrite this ancestral sequence 4 times to get 4 alternative ancestors: a) a sequence consisting of the best states in A- and B-sites, b) a sequence with the second most probable states of sites A and B, c) a sequence with the third probable states of sites A and B, and d) a sequence with the third probable state in sites A and with the fourth probable in site B.

Substitution asymmetry during protein evolution was detected by analyzing deviations in the protein evolution rates from the reversible and symmetric model of amino acid substitution rates on each of the branches of the protein tree (1) and by a comparative analysis of the lengths of the internal tree branches taking into account the structure of the protein (2). (1) In order to solve first task, we: a) reconstructed a protein-specific, time-reversible model of amino acid substitution rates using Modelestimator software; b) for each possible amino acid replacement on each internal node of the tree, we calculated d=PPa*PPb*2*NC, where PPa and PPb are the posterior probabilities of amino acids a and b, a $\neq$ b, NC=1/(1+e^(200*RFab)), RFab is the relative rate of ab substitution in the time-reversible model of amino acid substitution rates; c) summed d values along all sites in each internal tree node and calculated the natural logarithms of these sums ($\ln(\sum d)$); d) in order to identify branches with a maximum $\ln(\sum d)$, we conducted nonparametric comparison of these values over the entire tree. (2) In order to solve second task, we used truncated 'AltAll * N' libraries, particularly we: a) for each alternative ancestral sequence in each internal tree node, using the RaptorX_Property Fast software system, we assigned secondary structures to each amino acids, derived amino acid solvent accessibilities, and predicted disorder for each amino acid residue; b) calculated the frequency of changes of these measures between all alternative ancestral sequences of neighboring nodes residing on each internal tree branch; c) in order to identify branches with maximum structural changes, conducted nonparametric comparison of the abovementioned frequencies of changes across the entire tree.

Results

We extracted and verified 1565 Deuterostomia caspase proteins from NCBI GenBank. After that, we processed this protein set as described above.

One of the most promising observation we made is the finding of strict substitution asymmetry in the diversification of lower rodents, characterized by higher longevity comparing to higher rodents (Fig 1). The same evolutionary event is characteristic for caspase 7 in Xenopus (Fig 2).

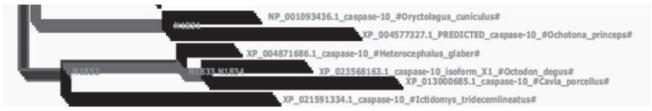


Fig. 1. Selected part of caspases phylogenetic tree, containing caspase 10 of lower rodents. Each branch composed of 5 lines, corresponding to (from down to up) changes in amino acid disorder, protein secondary structures (3 and 8 types of secondary structure), amino acid solvent accessibilities, and deviations from the reversible and symmetric model of amino acid substitution. Black lines show terminal branches, light grey indicates evolution without significant deviations, while dark grey indicates significant deviations from expectations.

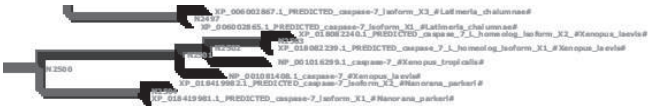


Fig. 2. Selected part of caspases phylogenetic tree, containing caspase 7 of Xenopus. Same designations as in Fig. 1.

It is of interest, that such evolutionary events are characteristic for the divergences of large protein clades tightly related with origination of large taxonomical groups, for example, avian caspase 6, mammalian caspase 13, etc.

ACKNOWLEDGMENT

This work supported by the grant from the RSF # 19-1500125. BZ is supported by the Swedish (190345) and Stockholm (181301) Cancer Societies.

REFERENCES

[1] Gumulya Y, Gillam EM. (2017) Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the
[2] 'retro'approach to protein engineering. Biochem J. 474(1):1–9.
[3] Anderson DP, et al. (2016) Evolution of an ancient protein function involved in organized multicellularity in animals. elife. 5:e10147.
[4] Eick GN, et al. (2017) Robustness of reconstructed ancestral protein functions to statistical uncertainty. Mol Biol Evol. 34(2):247–61.
[5] Quang le S, Gascuel O, Lartillot N. (2008) Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics. 24(20):2317–2323.

DOI 10.18699/BGRS/SB-2020-371

Actors of the base excision repair play: how well do we know the credits?

Dmitry O. Zharkov
Novosibirsk State University,
SB RAS Institute of Chemical Biology
and Fundamental Medicine,
Novosibirsk, Russia
dzharkov@niboch.nsc.ru

Anton V. Endutkin
ICBFM, Novosibirsk, Russia
aend@niboch.nsc.ru

Evgeniia A. Diatlova
ICBFM, Novosibirsk, Russia
jannie.lapt@gmail.com

Anna V. Yudkina
Novosibirsk State University,
ICBFM, Novosibirsk, Russia
ayudkina@niboch.nsc.ru

Alexander V. Popov
ICBFM, Novosibirsk, Russia
avpopov@niboch.nsc.ru

***Abstract* — Base excision repair (BER) is a pathway responsible for removal of small non-bulky DNA lesions from genomes of all living organisms. BER is initiated by one of several DNA glycosylases that excise the damaged base and is continued by apurinic/apyrimidinic endonucleases, DNA polymerases and DNA ligases. Although the whole BER pathway in both human and *E. coli* cells operates with 20–30 polypeptides, recent finding from other organisms suggest the existence of BER proteins that are not found in well-characterized model organisms. Even the homologs of known BER proteins may form families clearly separate from the biochemically characterized ones, and have different functions. Moreover, BER proteins often consist of a universally conserved catalytic core and accessory domains specific for a certain clade. Here, several examples of newly discovered bacterial DNA glycosylases are presented. Evolution of substrate and reaction specificity in BER enzymes is analyzed. The functions of non-catalytic domains or tails of BER proteins are discussed.**

***Keywords* — DNA damage, DNA repair, base excision repair, DNA glycosylases, protein structure, protein evolution.**

Motivation and Aim

Motivation

Base excision repair (BER) is a pathway responsible for removal of small non-bulky DNA lesions from genomes of all living organisms. BER is initiated by one of several DNA glycosylases that excise the damaged base and is continued by apurinic/apyrimidinic endonucleases, DNA polymerases and DNA ligases. Although the whole BER pathway in both human and *E. coli* cells operates with 20–30 polypeptides, recent finding from other organisms suggest the existence of BER proteins that are not found in well-characterized model organisms.

Aim

We aimed to mine sequence databases for homologs of known BER proteins in order to find polypeptides and/or their domains forming families clearly separate from the biochemically characterized ones, and to characterize their functions biochemically.

Methods

We have used Conserved Domain Database [1] to obtain consensus sequences of all BER proteins and used them as queries in BLAST [2] to recover all homologous protein sequences from the NCBI RefSeq database [3]. The sequences were aligned and distance trees were built using Clustal Omega [4]. Coding sequences for Flp1, Flp2, Flp3a, Flp3b, Flp4, *Sau*Fpg1, *Sau*Fpg2, *Sau*MutY, Ulp2, and Ulp3 proteins were made by total gene synthesis and cloned into pET series plasmids for expression. The proteins were overproduced and purified by a combination of metal affinity chromatography, ion exchange chromatography, and gel filtration. Substrate specificity of the enzymes was studied using modified oligonucleotides essentially as described for other DNA glycosylases [5,6]. X-ray data were collected at BNL National Synchrotron Light Source II. The protein structures were built by molecular replacement using *Phaser* [7], *Coot* [8], Phenix [9], Refmac [10], and BUSTER [11]. Molecular dynamic studies were done in Amber 11 [12] employing ff99SB force field with parmbsc0 corrections [13, 14].

Results

All sequences bearing homology to the known members of helix–two-turn–helix (H2TH) DNA glycosylases (Fpg, Nei, NEIL1, NEIL2, NEIL3, MMH) could be grouped in twelve families, five of which had none biochemically characterized members. To establish substrate specificity and enzymatic mechanisms of these Flp proteins (Fpg-Like Proteins), we have synthesized coding sequences of Flp1 from *Streptomyces coelicolor*, Flp2 from *Xanthomonas campestris*, Flp3a and Flp3b from *Bacteroides thetaiotaomicron*, and Flp4 from *Salinibacter ruber*, and purified the proteins. The enzymes were mostly Nei/NEIL-like, removing oxidized pyrimidines from DNA and generating a significant fraction of β-elimination product, unlike Fpg that predominantly yields products of β,δ-elimination. Flp3a and Flp3b are different from most H2TH glycosylases in having N-terminal nucleophilic Ile and Lys, respectively, instead of Pro. The crystal structure of Flp3b was determined and revealed a closed protein conformation (like Fpg and NEIL1 but unlike Nei), a lesion-binding loop significantly different from other H2TH superfamily members, and a partial electron density corresponding to an unprocessed initiator Met residue. Molecular dynamics was used to obtain the structure of the fully processed Flp3b and analyze its interactions with various damaged bases. Additionally, *E. coli* Fpg with a series of N-terminal substitutions was analyzed biochemically and revealed that Pro1 is indeed required for activity in H2TH glycosylases out of the Flp3 family.

All sequences bearing homology to the known members of uracil–DNA glycosylase superfamily (Ung/UNG, Mug/TDG, SMUG1) could be grouped in nine families, three of which had none biochemically characterized members. To establish substrate specificity and enzymatic mechanisms of these Ulp proteins (Ung-Like Proteins), we have synthesized coding sequences of Ulp2 from *Helicobacter hepaticus* and Ulp3 from *Pseudomonas syringae*, and purified the proteins. Both were monofunctional DNA glycosylases, as all known superfamily members. Unexpectedly, Ulp3 turned out to lack uracil-removing activity but efficiently processed DNA with oxidized pyrimidines such as 5-hydroxyuracil and 5,6-dihydrouracil.

Some species contain more than ortholog from a given family of DNA glycosylases. For instance, the reference genome of *Staphylococcus aureus* (strain NCTC 8325), an important human opportunistic pathogen, possesses one homolog for Fpg, yet many isolates of *S. aureus* and other *Staphylococcus* species contains another homolog, more closely related to Fpg than to Nei. We have produced and characterized both enzymes, and found them to be essentially Fpg-like. However, another DNA glycosylase, MutY, which in *E. coli* and humans counteracts the mutagenic action of 8-oxoguanine (oxoG), was different in *S. aureus*: it excised guanine much better than oxoG.

E. coli MutY and its human homolog, MUTYH, possess a C-terminal domain important for the preferential excision of oxoG. This domain is similar to NUDIX superfamily proteins, which hydrolyze a number of mononucleotide substrate; in particular, *E. coli* MutT and human MTH1 are pyrophosphatases converting oxodGTP to oxodGMP and preventing its incorporation into DNA. Interestingly, in the static crystal structure of MutY, oxoG interacts with the NUDIX domain in an entirely different manner than in MutT and MTH1, and the oxoG interactions with MutT and MTH1 are also different [15], while kinetic data suggest that the early stages of oxoG recognition by MutY occur in a MutT/MTH1-like conformation. All NUDIX sequences could be grouped in 54 families. Comparison of sequences and computer simulation of structures and their interactions with oxoG showed that MutY C-terminal domain is closer to human MTH1; a putative early recognition pocket with a set of specific enzyme–oxoG interactions was revealed in MutY.

Apurinic/apyrimidinic (AP) endonucleases are important enzymes catalyzing the next step of BER after base excision. They cleave DNA at AP sites, and also have a 3′→5′-exonuclease activity, 3′-phosphodiesterase, and 3′-phosphatase activity. The major human AP endonuclease, APEX1, belongs to the exonuclease–endonuclease–phosphatase superfamily, and the homologs of APEX1 can be grouped in five families, some of which include enzymes with preferential 3′-end processing function. An analysis of sequence conservation and structure allowed us to single out the residues that are responsible for 3′-processing rather than AP endonuclease activity. Evolution of non-catalytic domains of APEX1-like enzymes shows that this group of polypeptides is very modular, combining the DNA repair function with redox sensing, recombination, and likely other, still undefined functions. The N-terminal tail of human APEX1, in addition to its redox function, was needed to form oligomers on DNA, which might be required for APEX1 interaction with the rest DNA repair machinery and with transcription factors.

ACKNOWLEDGMENT

Supported by RFBR (20-04-00554-a) and RSF (17-14-01190).

REFERENCES

[1] Marchler-Bauer A. et al. (2015) CDD: NCBI's conserved domain database. Nucleic Acids Res. 43(D1):D222-D226.

[2] Altschul S.F. et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25(17):3389-3402.

[3] O'Leary N.A. et al. (2016) Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44(D1):D733-D745.

[4] Sievers F. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7:539.

[5] Tchou J. et al. (1994) Substrate specificity of Fpg protein: Recognition and cleavage of oxidatively damaged DNA. J. Biol. Chem. 269(21):15318-15324.

[6] Zharkov D.O. et al. (2000) Substrate specificity and reaction mechanism of murine 8-oxoguanine-DNA glycosylase. J. Biol. Chem. 275(37):28607-28617.

[7] McCoy A.J. et al. (2007) *Phaser* crystallographic software. J. Appl. Crystallogr. 40(4):658-674.

[8] Emsley P. et al. (2010) Features and development of *Coot*. Acta Crystallogr. D Biol. Crystallogr. 66(4):486-501.

[9] Zwart P.H. et al. (2008) Automated structure solution with the PHENIX suite. Methods Mol. Biol. 426:419-435.

[10] Murshudov G.N. et al. (2011) *REFMAC5* for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr., 67(4):355-367.

[11] Smart O.S. et al. (2012) Exploiting structure similarity in refinement: Automated NCS and target-structure restraints in *BUSTER*. Acta Crystallogr. D Biol. Crystallogr., 68(4):368-380.

[12] Case D.A. et al. (2010) AMBER 11. University of California, San Francisco.

[13] Hornak V. et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 65(3):712-725.

[14] Pérez A. et al. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. Biophys. J. 92(11):3817-3829.

[15] Yudkina A.V. et al. (2019) Reading and misreading 8-oxoguanine, a paradigmatic ambiguous nucleobase. Crystals, 9(5):269.

Section
Structural computational biology

DOI 10.18699/BGRS/SB-2020-372

Learning the changes of barnase mutants thermostability from structural fluctuations obtained using anisotropic network modeling

Nikolay Alemasov
Kurchatov Genomics Center,
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
ORCID iD: 0000-0002-7511-5385

Nikita Ivanisenko
Kurchatov Genomics Center,
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
ivanisenko@bionet.nsc.ru

Vladimir Ivanisenko
Kurchatov Genomics Center,
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
salix@bionet.nsc.ru

Abstract — **In biotechnology applications, rational design of new proteins with improved physico-chemical properties includes a number of important tasks. One of the greatest practical and fundamental challenges is the design of highly thermostable protein enzymes that maintain catalytic activity at high temperatures. This problem may be solved by introducing mutations into the wild-type enzyme protein. In this work, to predict the impact of such mutations in barnase protein we applied the anisotropic network modeling approach, revealing atomic fluctuations in structural regions that are changed in mutants compared to the wild-type protein. A regression model was constructed based on these structural features that can allow one to predict the thermal stability of new barnase mutants. Moreover, the analysis of regression model provides a mechanistic explanation of how the structural features can contribute to the thermal stability of barnase mutants.**

Keywords — barnase, thermostability, free energy, prediction, machine learning

Introduction

The thermal stability of proteins reflects the ability to preserve the unique spatial structure of the polypeptide chain under high temperature. Thermostability is determined numerically through the differences between the enthalpy and entropy of the protein in the native and denatured states [1] and can be determined from the thermodynamic properties of the amino acids that make up the protein [2]. There are several experimental methods for improving the thermal stability of a protein: stabilizing its native form, destabilizing its denatured form, or by combining both approaches [3].

The present work is aimed at developing an approach to construct protein structure ”profiles” indicating the positions of amino acid residues that are most sensitive in terms of protein thermal stability to mutations. Such approaches were previously developed to study the survival time of patients with mutations in the SOD1 protein. Here, the feasibility of this approach to assess thermal stability was shown for the first time using barnase protein as an example [4].

Methods

The experimental thermal stability of 134 barnase mutants as assessed by Gibbs free energy (ddG, kJmol^{-1}) was taken from the work of Seeliger and de Groot [5].

The initial structure for the study was taken from the PDB (1BNI). Mutant structures were obtained by introducing an appropriate mutation using the PDBFixer tool (https://github. com/pandegroup/pdbfixer). The dynamics of the barnase mutants and their wild types were simulated using anisotropic network models (ANM) implemented in the ProDy [6]. The simulation was performed with the following parameters: $m \in \{16,48,80,159\}$, $n = 10$, $A \in \{0.5,1,2,3\}$A, where˚ m is the number of first vibrational modes, n is the number of conformations obtained, A is the desired root-meansquared deviation from the original structure. Each of the n conformations obtained for a given barnase mutant was subjected to geometry optimization using the RDKit package (http://www.rdkit.org/).

To obtain the properties of the protein structure, after the simulation the MDTraj was used [7]. The following measures were calculated using MDTraj for each conformation: secondary structure (DSSP), gyration radius (Rg), root-meansquared fluctuations of protein residues (RMSF), and solvent accessible surface area (SASA) of each residue. The random forest method implemented using the scikit-learn was used to construct the regression models [8]. Stability selection was used as a feature selection algorithm [9].

The training set included 90 % of the mutants while the test set contained 10 % of the mutants. At each bootstrap step for the mutants from the test sample, the root-meansquared prediction error (RMSE) of the Gibbs free energy was estimated.

The treeinterpreter (https://github.com/andosa/treeinterpreter) method of interpreting the results of predictions obtained using the random forest was used to study the mechanism influencing the mutations that affect the thermal stability of barnase.

Results

As a result of the ANM simulation of the wild-type protein, 318 modes with non-zero eigenvalues were obtained. In total, 136 regression models were constructed (on 16 single parameters and 120 various pairs of simulation parameters) using the random forest method, linking the calculated parameters of the structure of the barnase mutants and the thermal stability of these mutants.

The most accurate regression model was a model based on a pair of variants (A = 0.5A,˚ m = 80 and A = 2A,˚ m = 159) with the accuracies RMSE = 5.06±1.27 kJmol^{-1}and AUE = 3.87±0.06 kJmol^{-1}. Hereinafter, this model will be denoted by its serial number 114 among all 136 models.

TABLE I – TOP TEN IMPORTANT FOR PREDICTING THE THERMAL STABILITY OF BARNASE MUTANTS USING REGRESSION MODEL NO. 114 FEATURES. THE NAME OF THE FEATURE CONTAINS A CALCULATED PARAMETER (DSSP, RMSF OR SASA), AN INDICATION OF THE AVERAGE VALUE OF THE PARAMETER (BAR) OR STANDARD DEVIATION FROM THE MEAN (Σ), AS WELL AS THE NUMBER OF AMINO ACID RESIDUES IN THE SEQUENCE INDICATED AFTER "R"

#	A, A°	m	Feature name	Normalized feature importance, a.u.
1	0.5	80	RMSF, r21	0.05±0.02
2	2.0	159	$\frac{\quad}{\text{SASA, r3}}$	0.05±0.03
3	0.5	80	$\frac{\quad}{\text{SASA, r88}}$	0.02±0.02
4	0.5	80	RMSF, r47	0.02±0.01
5	0.5	80	RMSF, r20	0.02±0.02
6	2.0	159	RMSF, r41	0.02±0.01
7	2.0	159	σ(SASA), r23	0.02±0.02
8	0.5	80	σ(SASA), r74	0.02±0.01
9	2.0	159	RMSF, r34	0.02±0.01
10	0.5	80	$\frac{\quad}{\text{SASA, r74}}$	0.02±0.01

While constructing the regression model No. 114 using the stable selection method, 29 features important for predictions were selected. The first ten features with the highest importance are presented in table I. Among the top ten important features were those associated with amino acid residues at positions 3, 20–21, 23, 34, 41, 47, 74 and 88. An analysis of the contributions of each of the features selected as important for predicting the thermal stability of the barnase mutants revealed that deviations in the values of five features (# 1, # 2, # 6, # 8, # 10, see table I) account for more than 80 % of the variance in the contributions of all features.

Discussion

Previously, three hydrophobic cores were found in barnase [10]. In the present work, features associated with residues at the positions Val3, Leu20, Pro21, Asn23, Gly34, Asn41, Pro47, Ala74, and Ile88 were found to be among the most important ones for predicting thermostability. Of these amino acid residues, Ile88 is located in the hydrophobic core 1. The residues at positions 20, 21, and 23 are located on the same loop as the Ile25 residue from the core 2, and the Asn41 residue is located in the same alpha helix as Leu42 from core 2. The Ala74 is located in the same beta-strand as the Ile76 from the hydrophobic core 1. It can be assumed that the barnase mutants studied here can promote structural changes affecting significant elements, such as hydrophobic core 1 and core 2.

The computational analysis of the barnase mutants conducted in the current work leads to the following conclusion: it is possible to determine the kind (stabilization or destabilization) of the mutational effect on the structure of the protein based on the remaining three features: the average surface area accessible to the solvent in the region of the amino acid residues Val3 and Ala74; the deviation of this area in the region of the residue Ala74 from the average; the mean-square fluctuations in the region of the residues Pro21 and Asn41 in mutants. Such an opportunity can allow us to propose new mutations in this protein that will stabilize its structure compared to the wild-type protein.

Acknowledgment

The authors thank the Siberian Branch of the Russian Academy of Sciences for the budget project № 0324-20190040-C-01 and the Russian Foundation for Basic Research together with the Government of the Novosibirsk Region for the Project № 19-44-543002.

References

[1] S. Talluri, "Advances in engineering of proteins for thermal stability", International Journal of Advanced Biotechnology and Research, vol. 2 (1), pp. 190–200, 2011.

[2] C. Tanford, "Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins", Journal of the American Chemical Society, vol. 84 (22), pp. 4240–4247, 1962.

[3] B. W. Matthews, "Structural and genetic analysis of protein stability", Annual review of biochemistry, vol. 62, pp. 139–60, 1993.

[4] N. Alemasov, N. Ivanisenko and V. Ivanisenko, "Learning the changes of barnase mutants thermostability from structural fluctuations obtained using anisotropic network modeling", Journal of Molecular Graphics and Modelling, vol. 97 (107572), pp. 1–9, 2020.

[5] D. Seeliger and B. L. de Groot, "Protein thermostability calculations using alchemical free energy simulations", Biophysical journal, vol. 98 (10), pp. 2309–2316, 2010.

[6] A. Bakan, L. M. Meireles, and I. Bahar, "ProDy: Protein Dynamics Inferred from Theory and Experiments", in Bioinformatics, vol. 27 (11), pp. 1575–1577, 2011.

[7] R. T. McGibbon et al., "MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories", Biophysical Journal, vol. 109 (8), pp. 1528–1532, 2015.

[8] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python", Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[9] N. Meinshausen and P. Buhlmann, "Stability selection", Journal of the¨ Royal Statistical Society. Series B: Statistical Methodology, vol. 72 (4), pp. 417–473, 2010.

[10] A. R. Fersht, "Protein folding and stability: the pathway of folding of barnase", FEBS Letters, vol. 325 (1-2), pp. 5–16, 1993.

DOI 10.18699/BGRS/SB-2020-373

Interpretation of the features of a linear regression model for predicting the survival time of the amyotrophic lateral sclerosis patients with mutated SOD1

Nikolay Alemasov
The Federal Research Center
Institute of Cytology and Genetics
The Siberian Branch of the Russian
Academy of Sciences
Novosibirsk, Russia
ORCID iD: 0000-0002-7511-5385

Alexandr Shcherbakov
Novosibirsk State Technical University
Novosibirsk, Russia
aleksandr.serbakov@yandex.ru

Vladimir Timofeev
Novosibirsk State Technical University
Novosibirsk, Russia
v.timofeev@corp.nstu.ru

Vladimir Ivanisenko
The Federal Research Center
Institute of Cytology and Genetics
The Siberian Branch of the Russian
Academy of Sciences
Novosibirsk, Russia
salix@bionet.nsc.ru

***Abstract* — Amyotrophic lateral sclerosis (ALS) is an incurable neurodegenerative disease characterised by the inevitable degeneration of central and peripheral motor neurons. Aggregation of mutant SOD1 is one of the molecular mechanisms underlying the onset of the disease. Previously, we proposed regression models linking the change in the stability of hydrogen bonds in mutant SOD1 calculated using molecular dynamics with patients' survival time. In this work, we developed an approach to the interpretation of features of linear regression models with the aim of a deeper understanding of the structural effects of mutations in the SOD1 protein. The approach suggested was based on the principal component analysis and over-representation of features including the sequence position of amino acid residues forming the hydrogen bonds detected within a range of the structural elements of the SOD1. This study can help one to deduce important structural regions in the protein which could be further targeted by small chemical compounds.**

Index Terms* — *ALS, SOD1, hydrogen bonds, regression models, PCA, over-representation

Introduction

Amyotrophic lateral sclerosis (ALS) is an incurable neurodegenerative disease [1]. It is known that the second most common and widely studied cause of the familial form of ALS are mutations in the *SOD1* gene, which encodes superoxide dismutase-1 [2]. One of the hypotheses regarding the mechanism at the molecular level of the disease is an aggregation of the SOD1 protein induced by its incorrect folding as a result of mutations [2].

Previously, we proposed a regression models for predicting the survival time of patients with ALS on the basis of the analysis of the change in the stability of hydrogen bonds in SOD1 mutants compared to the wild-type protein [3]. The stability of hydrogen bonds was estimated using the analysis of molecular dynamics (MD) trajectories.

In this work, we aimed at interpretation of features included into linear regression models to deeper understand the structural effects of mutations in the SOD1 protein. For this to achieve an approach was suggested based on the principal component analysis and over-representation of positions of multiple features such as the hydrogen bonds detected previously.

Methods

The stability of the hydrogen bonds between protein atoms and atoms of water molecules and the stability of water bridges from MD trajectories were taken from our previous paper [3]. For the MD simulation AMBER 12 software suite was applied there [4] and hydrogen bonds were detected with the cpptraj utility for each conformation via AmberTools 13 [5].

The observed survival time of patients with different mutations in the SOD1 protein was obtained from the literature [6]. These data was used to construct multiple regression model. To isolate latent factors, the principal component analysis was used, followed by varimax rotation. Later, latent factors were used as independent variables in constructing regression models, the response of which was the survival time of ALS patients. In addition to the classical least-squares method, robust M-estimates were used based on the Huber loss function to construct regression models. There were 100 bootstrap steps with each estimating the standard error (RMSE) as the mean square difference between the predicted and published survival times.

An analysis of principal components was also applied to unravel the internal structure of hydrogen bonds stability formed in the different SOD1 mutants. Over-representation of amino acid residues within different structural regions was calculated with scipy [7].

The several structural regions were considered in the overrepresentation analysis (see e.g. reference [8]).

Results

As a result of the simulation of the wild-type SOD1 and its mutants 6381 hydrogen bonds were detected in the MD trajectories. However the best RMSE (5.44 years) was for the regression model with only 1545 hydrogen bonds left after filtering those H-bonds with stability less than 0.1.

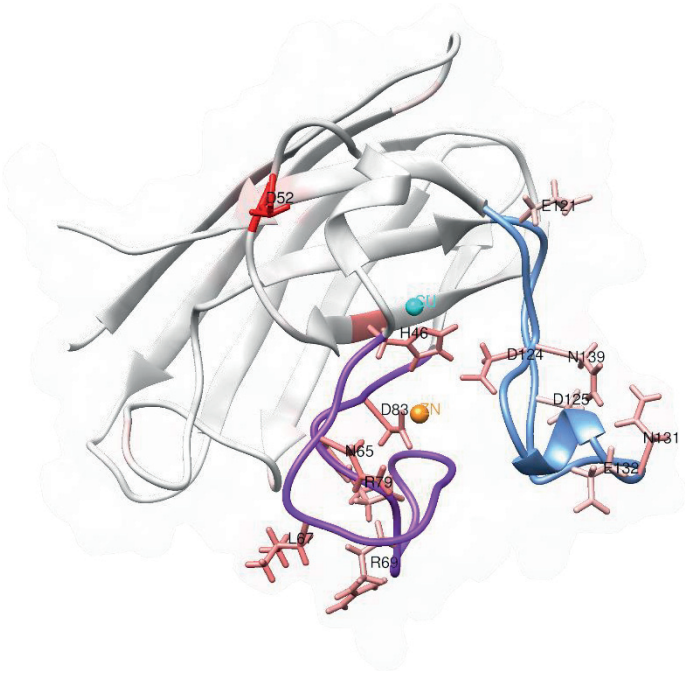


Fig. 1. Spatial structure of the SOD1 monomer with over-represented aminoacid residues shown

The results of evaluating the over-representation in figure 1. The best regression model using the robust M-estimates was logistic one resulting in the RMSE of 3.59 years (see table I).

TABLE I – RMSE OF THE REGRESSION MODELS BASED ON M-ESTIMATOR AND LEAST-SQUARE APPROACHES

Regression model	RMSE, years
Logistic based on M-estimator	3.588
Linear based on M-estimator	4.19
Cubic based on M-estimator	5.6
Cubic least-square	5.75

Discussion

The literature describes amino acid residues that can participate in the formation of aggregates of SOD1 protein mutants, which are one of the potential factors in the development of the ALS disease [8]. The authors identified regions which involved in the formation of aggregates: the electrostatic loop (residues 121 to 144), the disulfide loop (residues 49-62) and the loop containing the zinc ion-binding site (residues 63-84). According to our results, it was seen that positions 65, 67, 69, 79, 83 from the zinc-binding loop and 121, 124–125, 131– 132, 139 from the electrostatic loop known to be involved in aggregation were also involved in the formation of hydrogen bonds from the PC1. Thus, a change in hydrogen bond stability in these important positions can simultaneously be associated with survival time and affect the aggregation of mutant SOD1.

ACKNOWLEDGMENT

The authors thank the Siberian Branch of the Russian Academy of Sciences for the budget project No. 0324-20190040-C-01 and the Indian Council of Medical Research and Russian Foundation for Basic Research for the joint Project No. 17-54-49004.

REFERENCES

[1] M. C. Kiernan et al., “Amyotrophic lateral sclerosis”, The Lancet, vol. 377 (9769), pp. 942–955, 2011.

[2] E. Tokuda and Y. Furukawa, “Copper Homeostasis as a Therapeutic Target in Amyotrophic Lateral Sclerosis with SOD1 Mutations”, International Journal of Molecular Sciences, vol. 17 (5), P. 636, 2016.

[3] N. A. Alemasov et al., “Dynamic properties of SOD1 mutants can predict survival time of patients carrying familial amyotrophic lateral sclerosis”, Journal of Biomolecular Structure and Dynamics, vol. 35 (3), pp. 645–656, 2017.

[4] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, “An overview of the Amber biomolecular simulation package”, Wiley Interdisciplinary Reviews: Computational Molecular Science, vol. 3 (2), pp. 198–210, 2013.

[5] D. R. Roe, T. E. Cheatham, and T. E. Cheatham III, “PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data”, Journal of Chemical Theory and Computation, vol. 9 (7), pp. 3084–3095, 2013.

[6] Q. Wang et al., “Protein aggregation and protein instability govern familial amyotrophic lateral sclerosis patient survival”, PLoS biology, vol. 6 (7), P. e170, 2008.

[7] P. Virtanen, et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”, Nature Methods, vol. 17, pp. 261–272, 2020

[8] S. Antonyuk et al., “Structural consequences of the familial amyotrophic lateral sclerosis SOD1 mutant His46Arg”, in Protein science, vol. 14 (5), pp. 1201–1213, 2005.

DOI 10.18699/BGRS/SB-2020-374

Modelling of Nef interaction with ABCA1 revealed potential binding sites for inhibitor compounds

Anastasia A. Anashkina
EIMB RAS, Moscow, Russia
nastya@eimb.ru

Yaroslav V. Tkachev
EIMB RAS, Moscow, Russia
yat@eimb.ru

Alexei A. Adzhubei
EIMB RAS, Moscow, Russia
alexei.adzhubei@eimb.ru

***Abstract* — The effect of HIV-1 Nef protein on the demyelination of central nervous system cells is mediated by its effect on the cholesterol transporter protein ABCA1. To determine the possible interactions of Nef-ABCA1, an expert model of the cytoplasmic fragment ABCA1 was constructed and modelling of the reciprocal binding sites in the Nef and ABCA1 structures was carried out. This made it possible to determine the interface for the interaction of proteins and localize binding sites in them.**

Keywords — Nef, ABCA1, demyelination, model, binding site

Motivation and aim

Motivation

It is known that the Nef inhibits the activity of ABCA1, disrupts its maturation in the endoplasmic reticulum, and also induces the internalization and degradation of the membrane protein ABCA1. Chaperon calnexin, with which Nef forms a complex, plays an important role in this process.

Aim

Earlier we had shown that Nef interacts with the structural epitope ABCA1, which consists of sites in the two adjacent cytoplasmic domains [1]. While the site in the C-terminal domain ABCA1 was identified by us earlier, other possible interaction sites in ABCA1, as well as sites in Nef responsible for the interaction, remain unknown. This information is of major importance for the rational design of inhibitors.

Methods

To perform the task of searching for other binding sites of the ABCA1 cytoplasmic domain with Nef, an expert model of the cytoplasmic plus transmembrane domains of ABCA1, equilibrated by molecular dynamics, was built. Next, modeling of binding sites in the Nef and ABCA1 structures was performed using global docking, and employing our previously developed approach and server [2].

The structure of ABCA1 cytoplasmic domains D1 (847-1324) and D2 (1870-2261) was modeled by the servers Phyre2, iTasser, RaptorX, M4T, SwissModel. In addition, a model of the full-sized protein ABCA1 was built using the SwissModel server, based on the incomplete PDB 5XJY structure (with multiple missing fragments) obtained by electron microscopy. Further construction of the model structure was carried out by expert modeling.

For subsequent equilibration by molecular dynamics, a model was prepared, which contained the transmembrane and cytoplasmic parts of ABCA1 and a fragment of the lipid membrane. The system was minimized in 5,000 steps, then balanced in 1 fs steps for 75,000 steps and 2 fs steps for 150,000 steps using temperature and pressure coupling. For a 60 ns trajectory, the NPT ensemble was modeled at a temperature during the 310K simulation. All simulations were performed using the full atomic force field CHARMM36 and GROMACS software.

The previously constructed model of the structure of Nef protein [3] was also equilibrated by molecular dynamics method. Given the mobility of the Nef protein loops, 5 different conformation clusters were selected that most fully represent the space of conformational states. Representative structures of these clusters were used for docking.

The model of the cytoplasmic part of ABCA1, obtained after molecular dynamics, was used for docking with models of the Nef structure. Docking was carried out between the Nef structure models and the structure model of the cytoplasmic part of ABCA1 without a membrane, using ZDock, SwarmDock, GrammX and ClusPro servers. A total of 80 docking models of the interaction of cytoplasmic part of ABCA1 with Nef were obtained.

After excluding docking models in which Nef overlapped the structural region occupied by the membrane, eighteen ABCA1 - Nef models were selected.

Results

Considering that the N-terminus of Nef protein is lipidated at the Gly2 residue with the myristic acid residue, which serves to anchor Nef in the membrane, there remains only one ABCA1-Nef docking model, satisfying such spatial criteria. Using the QASDOM server [2], we determined the ABCA1 and Nef protein regions that form the interaction interface. The binding sites are centered on residues for which the largest number of intermolecular atomic interactions is shown. For ABCA1, these are Ser1023, Leu1026, Leu1972, Thr1261, Pro1267, Leu1282, Pro1284, Thr1305. For Nef: Asp32, Asp40, Asn51, Cys59, Trp61, Tyr85, Asp127. These putative binding sites will be used for structurally oriented virtual screening.

Acknowledgment

This work was supported by a grant from RFBR 17-54-30021 NIZ_a.

References

[1] Jacob D. et al. The ABCA1 domain responsible for interaction with HIV-1 Nef is conformational and not linear // Biochem Biophys Res Commun 2014. Vol. 444. P. 19.

[2] Anashkina AA. et al. Meta-server for automatic analysis, scoring and ranking of docking models // Bioinformatics 2018. Vol. 34. P. 297.

[3] Hunegnaw R. et al. Interaction Between HIV-1 Nef and CalnexinHighlights // Arteriosclerosis, thrombosis, and vascular biology 2016. Vol. 36. P. 1758.

DOI 10.18699/BGRS/SB-2020-375

Beta-bends as an example of conformationally predetermined segments of Protein. Conditions of stabilization of the structure and role of context

Anastasia A. Anashkina
EIMB RAS, Moscow, Russia
nastya@eimb.ru

Ivan Yu. Torshin
FIC IU RAS, Moscow, Russia
tiy135@yahoo.com

Natalia G. Esipova
EIMB RAS, Moscow, Russia
nge@eimb.ru

Vladimir O. Chekhov
EIMB RAS, Moscow, Russia
bock@eimb.ru

Leonid A. Uroshlev
IGG RAS, Moscow, Russia
leoniduroshlev@gmail.com

Vladimir G. Tumanyan
EIMB RAS, Moscow, Russia
tuman@eimb.ru

Abstract — **Conformations of the local protein structures is determined by the balance of their own energy characteristics and the context (flanking regions). The conformation of a given fragment becomes predetermined if the contribution of the context is clearly defined. Beta bends were chosen as the object of the study due to availability of formalised classification of this structure, small number of variables that determine the conformation of a segment, as well as the beta bends high prevalence in the three-dimensional structures of proteins. Using the PDB data, a complete cluster analysis of the conformationally predetermined beta-bend segments of the four main types was performed. Statistical data on the conformational parameters, amino acid composition and the amino acid sequences in beta-bends and their flanking regions were summarized. Factors have been established that stabilize the beta bend (additional hydrogen bonds between the residues involved in the bend) and the beta hairpin, flanking the bend (presence of a conformationally stable peptide). The existence of such stabilizing hydrogen bonds is confirmed by quantum chemical calculations of the energy of the fragment including bending. Stabilization of the hairpin neck is achieved by a conformationally stable hexapeptide in the conformation of the beta strand that is included in the structure, which is also confirmed by the method of denaturing molecular dynamics.**

Keywords — beta-bends, conformational analysis, predetermined conformations, quantum chemistry, molecular dynamics

Motivation and aim

Motivation

The solution of the problem of protein structure can be reached it seems, by pursuing the studies of multi-scale components of the native protein structure. And though the purely structural aspects of these units are described sufficiently well and exhaustively, the issues of their stabilization and conditions for the formation of local structures are only waiting to be resolved. To approach these issues we employ the example of beta bends, the local structures whose conformations are characterized by a small number of the degrees of freedom.

Aim

Geometric aspects of the formation of beta-bends were explained by us earlier on the basis of the idea of a topological lock [1, 2], where closing a pseudocycle through a hydrogen bond reduces the number of significant independent parameters determining the conformation to four. Thus the number of theoretically possible structures decreases to two. Moreover, one solution corresponds to an experimentally determined structure, and the second solution cannot be realized due to the unacceptable atom-atom contacts [2]. The purpose of this study is to establish the energy aspects of the formation of predetermined conformations, including the contribution of flanking regions.

Methods

To calculate the geometry of molecules and optimize energy, we used the methods of clustering, Voronoi-Delaunay decomposition, enumeration of conformations, distance geometry, AM1, as well as the ab initio method in the approximation of 6-31 ++ G * with the MP2 correlation.

Results

Using the data of PDB database (including structures solved by the neutron diffraction method), a cluster analysis was performed of the four main types of beta bends identified by us as conformationally predetermined. Statistical data were collected on the conformational parameters, amino acid composition and amino acid sequences (including assessment of the residue conservation) in the beta-bends and flanking regions. A quantum chemical calculation of the change in total energy depending on the orientation in space of the NH-bond of the second residue in the beta-bend, made it possible to draw a conclusion that an additional H bond is formed between the carbonyl oxygen of the residue i in the beta-bend and the amine nitrogen of the residue i + 2. Another source of stabilization of the local structure as a whole is the confomationally stable (adopting the beta-strand conformation) peptide WKVEVND that secures the neck of the beta hairpin.

ACKNOWLEDGMENT

Supported by the RFBR grant 20-04-01085.

REFERENCES

[1] Uroshlev, L.A., Torshin, I.Y., Batyanovskii, A.V., Esipova, N.G. and Tumanyan, V.G., 2015. Disallowed conformations of a polypeptide chain as exemplified by the β-turn of the β-hairpin in the α-spectrin SH3 domain. Biophysics. 60(1): 1-9.

[2] Uroshlev, L.A., Torshin, I.Y., Batyanovskii, A.V., Esipova, N.G. and Tumanyan, V.G., 2019. Predetermined Conformations in Bends of Polypeptide Chains: A Geometric Analysis. Biophysics, 64(2): 195-202.

DOI 10.18699/BGRS/SB-2020-376

Modeling of single-molecule FRET-experiments on protein folding: from coarse-grained to all-atom simulations

Vladimir A. Andryushchenko
IT SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
vladimir.andryushchenko@gmail.com

Sergei F. Chekmarev
IT SB RAS, Novosibirsk, Russia
NSU, Novosibirsk, Russia
chekmarev@itp.nsc.ru

***Abstract* — A key question in the application of the single-molecule Förster resonance energy transfer (smFRET) to study protein folding is how the dyes can affect the process of folding. Understanding of these effects is particularly important for small proteins, for which the dyes can be comparable in size with the protein. Also, the smFRET-experiments require a support from simulations in order to interpret the results of the measurements.**

Keywords* — *FRET-experiments protein folding molecular dynamics

Motivation

A key question in the application of the single-molecule Förster resonance energy transfer (smFRET) to study protein folding is how the dyes can affect the process of folding. Understanding of these effects is particularly important for small proteins, for which the dyes can be comparable in size with the protein. Also, the smFRET-experiments require a support from simulations in order to interpret the results of the measurements.

Results

In this work, we modeled smFRET-experiments on folding of BBL domain, the protein that consists of 45 amino acid residues [1,2]. The FRET protein construct is a system that is presented by the protein and two dyes at the protein termini, as it is in the experiment.

First, to gain a general insight into the problem, a coarse-grained representation of the FRET-construct was used [3]. Specifically, the amino acid residues were replaced with monomers at the positions of C-alpha atoms, and the dyes were represented by chains of monomers mimicking the fluorophores and linkers. Using the method of molecular dynamics (MD), the simulation of the folding of the protein itself (as a "baseline" system) and of its FRET-construct was performed. It was found that the presence of dyes and, in particular, the appearance of the excluded volume in the spatial protein conformations due to the dyes, does not prevent the protein from folding into its native state. Also, due to the loading of the FRET-constructs with massive dyes, the FRET-construct folds considerably slower than the protein itself, which leads to a significant increase in the average folding time.

This preliminary study was followed by the most realistic MD simulation available up-to-date – the protein and dyes were described at atomic level of resolution, an explicit solvent was used, and a brute-force modeling of folding trajectories from an extended state of the protein was performed. As the dyes in the FRET-construct, the Cy3 and Cy5 fluorophores were used, which were attached to the protein termini. The protein and 25,000 water molecules (TIP3P) modeling the solvent were placed in a cube with a 95 A edge and periodic boundary conditions. The simulations were performed by the MD method in the framework of the CHARMM36m program [4]. It was found that at T = 300K, the native states of the protein and the FRET-construct are stable. Fifteen MD folding trajectories for both the protein and FRET-construct were generated. It has been found that the presence of dyes does not change the overall picture of folding except that the FRET-construct folds considerably slower than the original protein. The FRET-efficiency histograms constructed on the basis of simulated MD-trajectories have a tendency to converge and are found in agreement the experimental histograms.

Acknowledgment

This work was supported by a grant from the Russian Foundation for Basic Research (18-04-00013).

References

[1] F. Huang, L. Ying, A.R. Fersht, Direct observation of barrier-limited folding of BBL by single-molecule fluorescence resonance energy transfer, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 16239–16244.

[2] J. Liu, L.A. Campos, M. Cerminara, X. Wang, R. Ramanathan, D.S. English, V. Muñoz, Exploring one-state downhill protein folding in single molecules, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 179–184.

[3] S. F. Chekmarev. How the dyes affect folding of small proteins in single-molecule FRET experiments: A simulation study. Biophys. Chem. 254 (2019) 106243.

[4] Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J. Comput. Chem. 30 (2009)1545-1614.

DOI 10.18699/BGRS/SB-2020-377

Extraction of spectral series of ions from mass spectra of peptides by methods of integral transforms and machine learning

Eduard Fomin
Systems Biology Department,
Kurchatov Genomics Center, ICG SB RAS
Novosibirsk, Russia
fomin@bionet.nsc.ru

Nikolay Alemasov
Systems Biology Department,
Kurchatov Genomics Center, ICG SB RAS
Novosibirsk, Russia

Dmitriy Afonnikov
Systems Biology Department,
Kurchatov Genomics Center, ICG SB RAS
Novosibirsk, Russia

Abstract — **Two methods of preprocessing of peptide mass spectra aimed at identification of lines belonging to the same ion series (a, b, c, x, y, z) have been developed. The first method is based on the application of integral autocorrelation and convolution operations. For noise model spectra of peptides, in which m/z values are determined with accuracy up to 10^{-4}, this method allows accurate identification of lines belonging to different series. However, for the spectra obtained experimentally, the accuracy of this method was low. To identify the series in the experimentally obtained spectra, an approach based on the random forest method was developed, which allows each line to be attributed to one of the series in the peptide spectrum, taking into account the multiplicity of the ion charge. When cross-validation with a stratified six-fold sample, the accuracy of this method on data from the PeptideAtlas database (more than 40 spectra and more than 15 thousand lines) averaged 0.85, the value of the AUC parameter was 0.79.**

Keywords — linear peptides, mass spectra, integral transformation, machine learning

Introduction

An important fraction of the peptidoma of bacteria and fungi is non-ribosomal peptides (NRP), representing a class of secondary peptide metabolites, and having an extremely wide range of biological activity and pharmacological properties. According to the Norine database [1], NRPs manifest themselves as antibiotics (61%), toxins (17%), surfactants (16%), siderophores (11%), antitumor agents (4%), immunomodulators (4%), and ~ 25% of the NRPs included in the database show multiple activity. In the vast majority of cases (73%), NRPs have a complex structure: 64% are cyclic (partially or completely), 1% have branches, 8% have overlapping cycles and branches, while only 27% have a simple linear structure. The monomers that make up NRPs have a wide variety of types (~ 500) and include not only 20 proteinogenic amino acids, but also non-proteinogenic amino acids (Aib, Hpg, etc.) and modified forms (methylated, glycolized, D-forms) [2]. In connection with their biosynthesis from the non-rybosomal path [3], the identification of NRPs by classical methods of bioinformatics and genomics is impossible, and is carried out only on the basis of mass spectrometry.

Mathematically, the problem of sequencing peptides from mass spectra is reduced to the problem of reconstructing a set of points X from the set of their pairwise distances ΔX. Increasing the dimension of the point space and complicating its topology greatly complicates the task of reconstructing the set X. Even the relatively simple cases, to which the problems of sequencing linear and cyclic peptides are reduced, and known as turnpike and beltway problems were enrolled in the list of the main problems of computational geometry [4].

We previously showed [5, 6] that the potential for effective reconstruction of the sequence of amino acid residues of a peptide from its mass spectrum directly depends on the accuracy of the available data. We have provided algorithms that effectively $O(N^2)$ reconstructs sequences from the theoretical mass of the spectrum, which includes only a series of b + and y + ions given with an accuracy of 10^{-3} without noise [6] and with noises of various nature [7].

Since the experimental mass spectra include many different series of ions with different charges and are characterized by a large number of gaps, to start the sequencing process with the previously proposed algorithms, it is necessary to extract a given ions series and restore the lines missing in it.

Results

Integral transform method

We have developed an algorithm aimed at identifying a series of spectral lines that correspond to breaks of bonds of a certain type (a,b,c,x,y,z) [4]. The identification of the series occurs through the sequential application of autocorrelation and autoconvolution to the peptide spectra and subsequent search for lines, the distance between which is determined by a constant specific for each of the series (a,x), (b,y), (c,z). The 'integral transform' algorithm can be used to process spectra of peptides of any type, both cyclic and linear. Test calculations using noisy theoretical spectra of linear peptides showed the possibility of extracting any given series without relying on knowledge of the amino acid sequence, which is a new result.

Our experiments showed that the quality of line prediction of a given series substantially depends on the accuracy of specifying the lines of the spectrum, confirming the conclusion made in [6]: the possibility of reconstructing the sequence of amino acid residues of a peptide from its mass spectrum directly depends on the accuracy of the available data. Since the accuracy of the available experimental spectra cannot reach the required values 10^{-3}, the extraction of series from them is accompanied by an increase in the number of errors. In particular, this is shown by the example of the experimental spectrum of the IGITASALTR peptide (see Figure 1a) taken from the Peptide Atlas database [9]. For this spectrum we determined the series of b^+ and y^+ ions using the integral transformation algorithm described above. It turned out that when only lines with high intensity (> 0.1 max) are taken into account, the integral transform method finds 7 out of 18 lines of the

given series (38%). These lines are shown by blue lines in Figure 1a, and the exact lines of these series are indicated by dashed lines. It should be noted that taking into account the complete set of spectrum lines, including those with low intensity, allows us to determine all 18 lines of these series, however, resulting in rise in false positive rate.

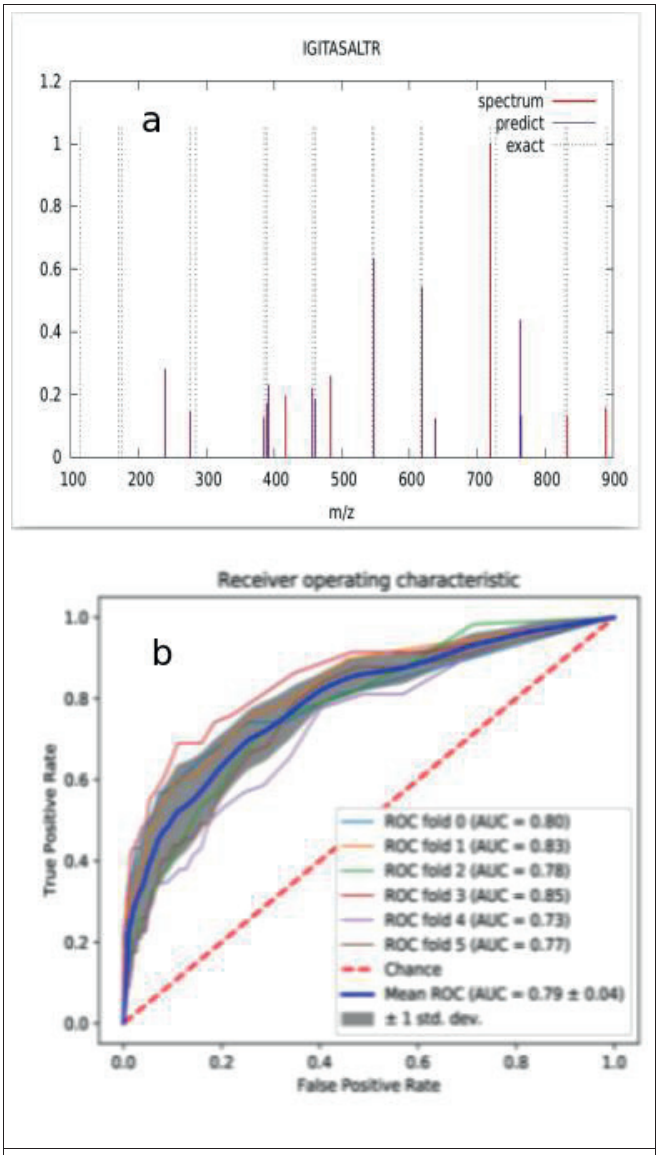


Fig. 1. (a) the result of the prediction of the b^+ and y^+ series in the experimental mass spectrum of the IGITASALTR linear peptide; (b) a set of ROC curves obtained by cross-validation with a stratified six-fold sample. The ROC curve obtained by averaging over the entire set of curves is highlighted in blue. The area under the curve (AUC) is indicated in parentheses.

Thus, the integral transformation algorithm turned out to be sensitive to errors in the spectrum of peptides that are inevitable in the analysis of experimental data. Therefore, we decided to use machine learning algorithms to identify a series of spectral lines.

Machine learning techniques

We have developed a method for classifying all spectrum lines according to the series (a, b, c, x, y, z). In this algorithm, the input is the spectrum of the peptide; at the output, the algorithm classifies each line of the spectrum according to a certain series (a, b, c, x, y, z). The parameters of the algorithm are theoretical displacements of distances for the m/z ratio between ions of a given mass in different series with different charges z (for a charge ion +1, 6 m/z values were calculated for each type of ion a, b, c, x, y, z; similarly, 6 values were calculated for charge +2 and charge +3). After that, if lines with given values of the m/z parameter were identified in the spectrum, then for this value the parameter became equal to 1, otherwise the parameter was assigned to be 0. The m/z ion and m/z precursor were taken as additional features.

As a data source, we used the Peptide Atlas database which contains the spectra of peptides with spectral lines marked by type of ions. We used data on 40 spectra (15,250 lines) for training and data on 8 spectra for control (2998 lines). The training was performed using the random forest method [10]. The results of evaluating the accuracy of the method predicting the series type for spectrum lines are shown in Figure 1b.

The following accuracy and sensitivity values were obtained for the classifier: precision = 0.85 ± 0.12, recall = 0.55 ± 0.04, MCC = 0.25 ± 0.15, AUC = 0.79 ± 0.04, where MCC is the Matthews correlation coefficient, and for assessing accuracy and sensitivity for different classes macro averaging was used. When constructing the classifiers, 10 features were used that provided the maximum correlation coefficient MCC and provided a class-average accuracy greater than 0.9 at the same time.

ACKNOWLEDGMENT

Russian Foundation for Basic Research, project No. 17-00-00462.

REFERENCES

[1] S. Caboche, M. Pupin, V. Leclère, A. Fontaine, P. Jacques, and G. Kucherov. 2008. Norine: a database of nonribosomal peptides. Nucleic Acids Res. 36:D326–D331.

[2] S. Caboche, V. Leclère, M. Pupin, G. Kucherov and P. Jacques. Diversity of Monomers in Nonribosomal Peptides: towards the Prediction of Origin and Biological Activity. Journal of bacteriology, 2010, p. 5143–5150 Vol. 192, No. 19.

[3] Marahiel, M., Nakano, M., and Zuber, P. (1993). Regulation of peptide antibiotic production in Bacillus. Mol Microbiol, 7(5):631–636.

[4] Shamos, M. (1977). Problems in computational geometry. CMU, Pittsburgh, PA.

[5] Fomin, E. (2016). A Simple Approach to the Reconstruction of a Set of Points from the Multiset of n 2 Pairwise Distances in n 2 Steps for the Sequencing Problem: I. Theory. Journal of Computational Biology, 23(9), 769-775.

[6] Fomin, E. (2016). A Simple Approach to the Reconstruction of a Set of Points from the Multiset of n 2 Pairwise Distances in n 2 Steps for the Sequencing Problem: II. Algorithm. J. Comput. Biol. 23, 934–942.

[7] Fomin ES (2019) A Simple Approach to the Reconstruction of a Set of Points from the Multiset of Pairwise Distances in n 2 Steps for the Sequencing Problem: III. Noise Inputs for the Beltway Case. Journal Of Computational Biology, 26(1), 68-75.

[8] Roepstorff, P., & Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomedical mass spectrometry, 11(11), 601-601.

[9] F Desiere, EW. Deutsch, NL. King, AI. Nesvizhskii, P Mallick, J Eng, S Chen, J Eddes, SN. Loevenich, R Aebersold. The PeptideAtlas Project. Nucleic Acids Research, 2006, 34, D655-D658

[10] Ho TK (1998). The Random Subspace Method for Constructing Decision Forests. IEEE Transactions on Pattern Analysis and Machine Intelligence. 20 (8): 832–844.

DOI 10.18699/BGRS/SB-2020-378

Model for stacking monomers in filamentous actin

Anna Glyakina
IMPB RAS, Pushchino, Russia
Institute of Protein Research RAS, Pushchino, Russia
quark777a@gmail.com

Alexey Surin
Institute of Protein Research RAS, Pushchino, Russia
Pushchino Branch, Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Pushchino, Russia
alan@vega.protres.ru

Oxana Galzitskaya
Institute of Protein Research RAS, Pushchino, Russia
ogalzit@vega.protres.ru

***Abstract* — To date, a lot of scientific evidences have accumulated (limited proteolysis, mass spectrometric analysis, electron microscopy) that is not consistent with the generally accepted double-helical model of actin filament (F-actin) organization in eukaryotic cells. This entails an ambiguous understanding of many key cellular processes in which actin filament is involved. For a detailed understanding of the mechanism of actin polymerization and the interaction of actin with its partners, it is necessary to have a correct model of the structural organization of the actin filament at the molecular level. In this work, 155 structures of actin (from wild rabbit) obtained by X-ray diffraction and electron microscopy technique were analyzed. From the pairwise alignment and the calculation of the root mean square deviation for these structures, it follows that they are very similar. Our preliminary data on limited proteolysis and mass spectrometry analysis show that areas inaccessible for acting of proteases are completely open in the existing models of F-actin. Thus, in this work a new model for filamentous actin is proposed.**

Keywords — actin, monomer, filaments, proteolysis, accessible surface area

Motivation and aim

Motivation

Actin is the most abundant protein in eukaryotes. Analogs of actin protein have also been found in bacteria [1-3] and archaea [4,5]. In a cell, actin can exist in two forms: globular (G-actin) and fibrillar (F-actin). The length of globular actin is about 375 amino acid residues, and the molecular weight is about 42 kDa. In vertebrates, depending on the isoelectric point, three isoforms of actin α, β and γ are distinguished. α-actins are mainly characteristic of muscle cells, while β and γ-actins are characteristic of non-muscle cells. α-actins, in turn, are divided into three types: smooth muscle α-actin, skeletal muscle α-actin and cardiac muscle α-actin.

The actin transition between the globular and fibrillar forms is carried out under the influence of ions and actin-binding proteins [6-9]. Actin is involved in many protein-protein interactions [10-13], which allows it to play an important role in cell motility and maintaining their shape. The interaction of fibrillar actin with myosin is the basis for muscle contraction.

Aim

Since the 1950s, intensive studies of the structure of monomeric and filamentous actin have been conducted. According to the X-ray data analysis of polymeric actin, it was suggested that actin filaments can be helical [14]. Based on the EM analysis and data from paper [14], for the first time was stated that actin filaments are a double helix [15]. Since then, the idea of the double helix organization of filamentous actin is generally accepted. The atomic structure of filamentous actin in the double-helical form was proposed by Holmes in 1990 [16] based on fitting the crystal structure of monomeric G-actin in X-ray fiber diffraction data obtained for oriented gels of filamentous actin (F-actin) [16]. The experimental data (limited proteolysis, mass spectrometric analysis, electron microscopy) available today are in disagreement with this assumption. Thus, it is necessary to reconsider the existing data set in a new way and propose a new model that would satisfy the current experimental data.

Methods

Database Collection and Structural Characteristics of Proteins

The list of actin protein structures available in the Protein Data Bank (https://www.rcsb.org) was taken from the UniProtKB database (https://www.uniprot.org), record number P68135, gene ACTA1, wild rabbit species (Oryctolagus cuniculus).

There were 155 structures of actin. From them there were 72 monomer and 83 oligomer structures. 100 structures were obtained by X-ray diffraction analysis with a resolution from 1.29 to 7.88 Å, 52 structures were obtained by electron microscopy with a resolution from 3.6 to 70 Å, and one structure was obtained by fiber diffraction with a resolution 3.3 Å, and two model structures (1ALM and 1UY5).

The spatial alignment of 155 actin structures, the calculation of the root mean square deviation (RMSD) of the Cα atoms for each pair of superimposed structures, and the calculation of the accessible surface area (ASA) for each amino acid residue in the actin structures were performed using the YASARA program [17]. If there were several actin structures in the pdb file, then only one (first) structure was taken for spatial alignment and RMSD calculation. ASA was defined as the surface area that is rolled by a ball of water with radius 1.4 Å.

Mass Spectrometry Analysis

For actin polymerization, the sample was supplemented with KCl solution to a final concentration of 0.1 M and incubated at 37°C for a day. Then, the sample containing the fibrillar form of actin was centrifuged for 20 minutes at 10000 g. The pellet was washed twice with 100 mM NH4HCO3 (pH 7.5). Thus precipitated fibrils were dissolved in 100 mM NH4HCO3 (pH 7.5) to a concentration of 1 mg/ml and incubated during 20 hours with the proteinase K in the actin-to-protease ratio of 25/1). Then, the mixture was supplemented with CaCl2 solution to a concentration of 5 mM to ensure efficient functioning of proteinase K. After incubation with the mixture of proteases, the solution was centrifuged at 10000 g during 20 min and washed twice with 100 mM NH4HCO3 (pH 7.5) and then with 0.1% trifluoroacetic acid. The sample was then dried in a vacuum concentrator (Eppendorf, Germany). Mass spectrometric

analysis of fractions was performed on a high resolution mass spectrometer of the OrbiTrap Elite ETD orbital trap-type high-resolution mass spectrometer (Thermo Scientific, Germany). Fragmentation of ions was carried out by the methods of dissociative activation by collision (DAC) and electron transfer (ET). The mass of ions and ion fragments was recorded with a resolution of 240000 and 60000, respectively. The obtained fragmentation spectra were processed using the PEAKS Studio 7.5 software (Bioinformatics Solution Inc.). Peptides for which the ion current signal intensity was greater than 105 were regarded significant.

Results

The pairwise spatial alignment of 155 actin structures showed that the RMSDs calculated from the Cα atoms for these pairs do not exceed 3 Å. It follows that the structures of the actin proteins of thewild rabbit (Oryctolagus cuniculus) from the UniProtKB record P68135 are very similar.

Since the core of an actin filament is resistant to proteases due to the compact packing of the polypeptide chain, it can be assumed that the rest of the chain will be susceptible to proteolysis. Thus, it is possible to determine the regions of protein that make up the core of actin filament. As a result of proteolysis, the following peptides accumulated: 23-35, 97-107, 130-149, 164-195 and 331-339. From these data, it can be assumed that actin filaments should be formed so that the regions that make up its core are not accessible to the solvent. The accessible surface areas (ASA) of experimentally determined regions (23-35, 97-107, 130-149, 164-195 and 331-339) for filamentous (2W49, 1M8Q, 3G37, 3J8K, 6BNP) and monomeric (2ZWH) structure of actin from our database were calculated. It is worth noting that, in average, accessible surface area for amino acid residue is about 200 Å². An analysis of the data showed that in the filamentous structures only the region 164-195 is less accessible to the solvent in comparison with the monomeric actin structure. The differences in the ASA for the regions (23-35, 97-107, 130-149, and 331-339) between filamentous and monomeric actin structures is not revealed. Thus the existing filamentous structures of actin do not satisfy the experimental data on the treatment of F-actin with proteinase K.

Then, two models of filamentous actin were made, so that the regions that make up the core of the filaments were as little as possible accessible to the solvent. These models were designated as “circle” and “helix” (Fig. 1).

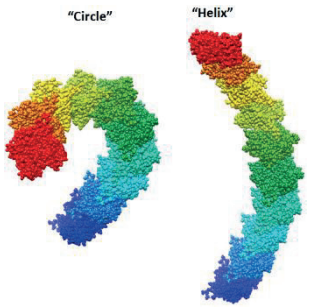


Fig. 1. The models of filamentous actin “circle” and “helix”.

From the calculations it follows that the “circle” model is more consistent with the experimental data on the treatment of F-actin with proteinase K than the “helix” model. Moreover, “circle” model type contradicts the experimental data on electron microscopy (not published), and the “helix” model is consistent with these data.

Thus, in this work, an attempt to construct a model of filamentous actin based on the available new experimental data was made.

Acknowledgment

Supported by the Russian Science Foundation (19-74-10051).

References

[1] Popp D., Narita A., Oda T., Fujisawa T., Matsuo H., Nitanai Y., Iwasa M., Maeda K., Onishi H., Maéda Y. (2008) Molecular structure of the ParM polymer and the mechanism leading to its nucleotide-driven dynamic instability. EMBO J. 27(3): 570-579.

[2] Galkin V.E., Orlova A., Rivera C., Mullins R.D., Egelman E.H. (2009) Structural polymorphism of the ParM filament and dynamic instability. Structure. 17(9): 1253-1264.

[3] Popp D., Narita A., Ghoshdastider U., Maeda K., Maéda Y., Oda T., Fujisawa T., Onishi H., Ito K., Robinson R.C. (2010) Polymeric structures and dynamic properties of the bacterial actin AlfA. J. Mol. Biol. 397(4): 1031-1041.

[4] Izoré T., Duman R., Kureisaite-Ciziene D., Löwe J. (2014) Crenactin from Pyrobaculum calidifontis is closely related to actin in structure and forms steep helical filaments. FEBS Lett. 588(5): 776-782.

[5] Braun T., Orlova A., Valegård K., Lindås A.C., Schröder G.F., Egelman E.H. (2015) Archaeal actin from a hyperthermophile forms a single-stranded filament. PNAS USA. 112(30): 9340-9345.

[6] Oda T., Iwasa M., Aihara T., Maéda Y., Narita A. (2009) The nature of the globular- to fibrous-actin transition. Nature. 457(7228): 441-445.

[7] Thomasson M.S., Macnaughtan M.A. (2013) Microscopy basics and the study of actin-actin-binding protein interactions. Anal. Biochem. 443(2): 156-165.

[8] Dominguez R., Holmes K.C. (2011) Actin structure and function. Annu. Rev. Biophys. 40: 169-186.

[9] Page R., Lindberg U., Schutt C.E. (1998) Domain motions in actin. J. Mol. Biol. 280(3): 463-474.

[10] Carlier M.F., Laurent V., Santolini J., Melki R., Didry D., Xia G.X., Hong Y., Chua N.H., Pantaloni D. (1997) Actin depolymerizing factor (ADF/cofilin) enhances the rate of filament turnover: implication in actin-based motility. J. Cell. Biol. 136(6): 1307-1322.

[11] Silacci P., Mazzolai L., Gauci C., Stergiopulos N., Yin H.L., Hayoz D. (2004) Gelsolin superfamily proteins: key regulators of cellular functions. Cell Mol. Life Sci. 61(19-20): 2614-2623.

[12] Kovar D.R., Harris E.S., Mahaffy R., Higgs H.N., Pollard T.D. (2006) Control of the assembly of ATP- and ADP-actin by formins and profiling. Cell. 124(2): 423-435.

[13] Ferron F., Rebowski G., Lee S.H., Dominguez R. (2007) Structural basis for the recruitment of profilin-actin complexes during filament elongation by Ena/VASP. EMBO J. 26(21): 4597-4606.

[14] Selby C.C., Bear R.S. (1956) The structure of actin-rich filaments of muscles according to x-ray diffraction. J. Biophysic. and biochem. cytol. 2(1): 71-85.

[15] Hanson J., Lowy J. (1963) The structure of F-actin and of actin filaments isolated from muscle. J. Mol. Biol. 6: 46-60.

[16] Holmes K.C., Popp D., Gebhard W., Kabsch W. (1990) Atomic model of the actin filament. Nature. 347(6288): 44-49.

Krieger E., Koraimann G., Vriend G. (2002) Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. Proteins. 47: 393-402.

DOI 10.18699/BGRS/SB-2020-379

Accuracy of disorder predictors results – comparison on DisProt DB data

Nenad Mitić
Faculty of Mathematics
University of Belgrade
Belgrade, Serbia
nenad@matf.bg.ac.rs

Gordana Pavlović-Lažetić
Faculty of Mathematics
University of Belgrade
Belgrade, Serbia
gordana@matf.bg.ac.rs

Saša Malkov
Faculty of Mathematics
University of Belgrade
Belgrade, Serbia
smalkov@matf.bg.ac.rs

Miloš Beljanski
Insistute for General and Physical Chemistry
University of Belgrade
Belgrade, Serbia
mbel@matf.bg.ac.rs

Mirjana Maljković
Faculty of Mathematics
University of Belgrade
Belgrade, Serbia
mirjana@matf.bg.ac.rs

Ana Jelović
Faculty of Transport and Traffic Engineering
University of Belgrade
Belgrade, Serbia
a.jelovic@sf.bg.ac.rs

Abstract — **Some results of accuracy and inconsistencies of disordered predictors prediction compared on proteins regions from DisProt database are presented. Results are compared for more than 20 predictors applied on different versions of DisProt database.**

Keywords — disorder predictors, accuracy, DisProt database

Introduction

It is well known that Intrinsically Disordered Proteins (IDPs) play significant role in key biological processes including signaling, recognition, regulation and cell cycle control [1]. Consequently, they are associated with various human pathologies, such as cancer, diabetes, cardiovascular and neurodegenerative diseases [2] and therefore are in the focus of many researches.

The process of experimentally determining disordered regions in proteins is slow and expensive, resulting in relatively small number of experimentally confirmed disordered regions. Information about such region were collected from the literature (manually curated) in DisProt database of proteins with experimentally detected disorder regions [3,4,5]. Number of collected regions vary from 152 proteins with 176 regions in DisProt 1.0 (released in 2003 year) to 1713 proteins with 7902 regions in DisProt V8.01. (released in 2019 year). The number of proteins and regions are even smaller because database include information about ambiguous and obsolete regions. Because importance of IDPs and their roles in many processes, as a consequence of slow and inefficient process of experimentally detecting, large number of computer programs - disorder predictors - has been developed for prediction of disordered regions in proteins.

As the reasons and mechanism of appearing disordered regions in proteins are not theoretically formally described, different methods and bases are used for prediction algorithms in disorder predictors (for example, see [6]). Some predictors behave better in one class of proteins and opposite. But researchers do not have any guarantee which predictors acts better and more accurate on completely new and previously unknown (unclassified) protein. In some cases, applying two or more predictors can lead to contradictory and confusing results.

The aim of this work was to analyze predictors results over experimentally verified disorder regions. This was motivation to apply predictors on proteins collected in DisProt database, and compare obtained results with expected i.e. experimentally confirmed as potentially "stable point".

Matherial and Methods

Material used in research includes all proteins from DisProt version 1.0 to DisProt Version 8.01. There are a lot of changes between DisProt consecutive versions. The changes include adding and removing proteins that were declared as disordered, obsolete regions, different AA contents of disorder regions and change in fractional differences, change in protein sequence, removing information about ordered regions which exist up to version 6.02, etc. Nevertheless, because DisProt is the one and only available collection of experimentally detected disordered regions it is used as benchmark.

Analysis is done using over 20 most commonly used disorder predictors divided into two group based on the method access possibilities. The first group includes predictors available for download and used in local environment as standalone versions, while the second one includes predictors that can be used over corresponding web sites.

The analysis includes a comparison of predicted and experimentally found disordered regions, as well as different predictors aspects including precision and accuracy at the points on the timeline associated with the detection of the regions and the construction of the predictors.

In addition to use as criteria direct hit (match, intersection) with the positions of the experimentally confirmed regions, other different criteria was used like approximate guessing, i.e. did the predictor predict the position in the environment or nearby the positions of regions in DisProt, method used to identify disordered regions, percentage of matched regions known at the time of predictors were announced and percentage of matched regions known after predictors were announced.

Results

Analysis produced very different results. Some predictors had very high level of correct predictions, but, as expected, there is no predictor that behaves as "the best" in all cases. Moreover, in some cases when consensus of some predictors was taken, these results did not match positions of

experimentally detected regions. Some cause of the lower quality of the results of some predictors are possible consequences of change between consecutive DisProt version that was used as benchmark. This is especially evident for difference between percentage of matched regions known before and after predictor announced. Obtained results suggest that for some predictors additional training which takes latest versions of DisProt DB can be useful.

REFERENCES

[1] V.N. Uversky: "Introduction to Intrinsically Disordered Proteins (IDPs)", Chem Rev. 2014;114(13):6557–60.

[2] V.N. Uversky, C.J. Oldfield,and A.K. Dunker: "Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept", Annual Review of Biophysics (2008), Vol. 37:215-246.

[3] S. Vucetic et al. "DisProt: a database of protein disorder", Bioinformatics. 2005 Jan 1;21(1):137-40. Epub 2004 Aug 13.

[4] A. Hatos et al. "DisProt: intrinsic protein disorder annotation in 2020", Nucleic Acids Research, Volume 48, Issue D1, 08 January 2020, Pages D269–D276.

[5] D. Piovesan et al. "DisProt 7.0: a major update of the database of disordered proteins", Nucleic Acids Research, Volume 45, Issue D1, January 2017, pages D219–D227.

[6] F. Meng, V.N. Uversky, L. Kurgan: "Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions", Cellular and Molecular Life Sciences volume 74, pages 3069–3090(2017).

.

DOI 10.18699/BGRS/SB-2020-380

Status of methylation and transcriptional activity of ecdysone receptor *EcR* gene in the Colorado beetle *Leptinotarsa decemlineata* say after application of S-adenosylmethionine

Yuri Nikonorov
Institute of Biochemistry and Genetics
Ufa Federal research centre of RAS
Russia
griffinriders@mail.ru

Tansulpan Akhmetkireeva
Institute of Biochemistry and Genetics
Ufa Federal research centre of RAS
Russia
tansulpan.ufa@gmail.com

Galina Benkovskaya
Institute of Biochemistry and Genetics
Ufa Federal research centre of RAS
Russia
bengal2@yandex.ru

Keywords — DNA methylation, Colorado potato beetle, ecdysone receptor gene

Motivation and Aim

Gene of ecdysone receptor coding the protein EcR belongs to superfamily of nuclear steroid receptors, which has an evident domain structure and is of great importance for insect development and reproducing [1]. In cells of different tissues protein EcR presents as a combination of its isoforms coded by some variants of mRNA generated consequently by alternative splicing and change in isoforms balance has phenotypical exhibition [2]. In the Colorado beetle *Leptinotarsa decemlineata* Say two variants of mRNA coding A and B1 isoforms with tissue-specific character of their number variations are described [3]. Our aim is the definition of DNA methylation character in separate loci of gene *EcR* coding region in the Colorado beetle, investigation of environmental effects and detection of correlation between DNA methylation and transcriptional activity of *EcR* gene.

Methods and Algorithms

Adult Colorado beetles sampled in potato plantation were used as an object. The source of methyl groups (Heptral – S-adenosylmethionine, SAM, Abbott Laboratories GmbH, Germany) was applied as suspension (80 mg/10 ml of water) for treatment of fresh potato leaves intended for beetles feeding. DNA and RNA were extracted from tissues of thorax and gonads of adults by phenol-detergent method. The presence of 5-methylcytosine in Colorado beetles DNA was determined by MSRE-PCR (methyl-sensitive PCR). Transcriptional activity of genes of ecdysone receptor *EcR,* DNA-methylase *Dnmt1* and protein transporter of juvenile hormone *Juv* was evaluated by qRT-PCR with reference gene *RP18*.

Results: The presence of methylated cytosine in DNA coding the regions of *EcR* gene in the *L. decemlineata* Say is of a specific character: DNA site for HpaII restrictase in exon distinctive for B1 isoform degraded completely under hydrolysis in all cases, i.e. it is constantly demethylated. Considerable part of DNA in exons specific for A isoform and constant for both isoforms is methylated permanently in a varying degree. In two months the addition of methyl group source (SAM) to food caused decrease of methylation level in coding region of *EcR* gene in adults of both sexes. Content of mRNA of *EcR* gene in cells of thorax increased 6,1–fold in males and 2,2-fold in females; in gonads it increased 7,7-fold and 9,1- fold respectively. The ratio of mRNA variants coding A and B1 isoforms changed to prevalence of A-isoform in muscles as well as in gonads in males and females. At the same time mRNA content of *Dmnt1*gene reduced to a certain degree and the content of mRNA of *Juv* gene increased dramatically, especially in female gonads. Similar but more expressed changes in transcriptional activity of these genes were observed right after feeding beetles by nontreated by SAM potato leaves: in an hour after feeding we noted more than two-fold decrease of mRNA of Dnmt1 gene content and more than 10-fold increase of mRNA of *Juv* gene content. However in this case transcriptional activity of *EcR* gene decreased 10-fold.

Conclusion

Effect of SAM addition to food is of long-term character exhibited as change of methylation status and expression of *EcR* gene in the Colorado beetle. We suggest that methylation of DNA takes part in regulation of mRNA splicing of *EcR* gene. A considerable part of methyl group donors most probably includes in juvenile hormone metabolism. Initiation of this process activated by SAM treatment leads to deficit of endogenic donors of methyl groups the consequence of which is the phenomenon of DNA demethylation.

Acknowledgments

Investigations was partially funded by RFBR No. 17-44-020347-r_a.

References

[1] L.M. Riddiford, P. Cherbas, J.W. Truman, “Ecdysone receptors and their biological actions”, Vitam. Horm., 2000, vol. 60, pp. 1-73.

[2] W.S. Talbot, E.A. Swyryd, D.S. Hogness, “*Drosophila* tissues with different metamorphic responses to ecdysone express different ecdysone receptor isoforms”, Cell, 1993, vol. 93, pp. 1323–1337.

[3] T. Ogura, C. Minakuchi, Y. Nakagawa, G. Smagghe, and H. Miyagawa, “Molecular cloning, expression analysis and functional confirmation of ecdysone receptor and ultraspiracle from the Colorado potato beetle *Leptinotarsa decemlineata*”, FEBS J., 2005, vol. 272 (16), pp. 4114-4128

DOI 10.18699/BGRS/SB-2020-381

Study of mutagenic properties of PEG-internferon-λ1

Larisa A. Oleynik
Department of Experimental Pharmacology
Research Institute of Clinical and Experimental Lymphology - a branch of the Institute of Cytology and Genetics of Siberian Branch of Russian Academy of Sciences
Novosibirsk, Russia
larisaoleynik81@gmail.com

Nikolai Kikhtenko
Department of pharmacology, clinical pharmacology and evidence based medicine Novosibirsk State Medical University
Novosibirsk, Russia
dr_kikhtenko@gmail.com

Pavel G. Madonov
Department of experimental pharmacology Research Institute of Clinical and Experimental Lymphology - Branch of the Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russian Federation
Department of pharmacology, clinical pharmacology and evidence based medicine Novosibirsk State Medical University
Novosibirsk, Russia
pmadonov@yandex.ru

Abstract — **Immune regulation at the maternal-fetal interface is complex due to conflicting immunological objectives: protection of the fetus from maternal pathogens and prevention of immune-mediated rejection of the semiallogeneic fetus and placenta. Interferon (IFN) signaling plays an important role in restricting congenital infections as well as in the physiology of healthy pregnancies. While type III IFNs stimulate antiviral protection, which can protect the fetus and placenta from certain viral infections, type I IFNs have the opposite effect, damaging the structure and function of the placenta.**

The relevance of this project is due to the ability to create a domestic highly effective, safe and inexpensive drug based on pegylated interferon lambda (PEG-interferon-λ1). Therefore, it is very important to conduct a preclinical study of the mutagenic properties of the drug based on PEG-interferon-λ1 in experimental animals and in the somatic mosaic test system

Keywords — finished dosage form based on PEG-interferon-λ1, mutagenicity

Introduction

Barrier surfaces such as the epithelium lining the respiratory and gastrointestinal (GI) tracts, the endothelium comprising the blood–brain barrier (BBB), and placental trophoblasts provide key physical and immunological protection against viruses. These barriers utilize nonredundant mechanisms to suppress viral infections including the production of interferons (IFNs), which induce a strong antiviral state following receptor binding. However, whereas type I IFNs control infection systemically, type III IFNs (IFN-ls) control infection locally at barrier surfaces and are often preferentially induced by these cells. In addition, it is widely known that epithelial cells predominantly produce IFN-λ rather than type I IFN; and lambda interferon receptor 1 (IFNLR1), a specific receptor for IFN-λ, is highly expressed in cells of epithelial origin.

It should be noted that, in contrast to other types of barrier cells that have been studied fairly well and are widely described in the literature, evidence that type III IFNs play an important role in protecting human placenta from viral infections has appeared relatively recently.

Interestingly, type III IFNs are constitutively released from human trophoblasts in the absence of any viral infections [4-6]. Accordingly, the medium isolated from uninfected primary donor human trophoblastic cells or from chorionic villi isolated from placenta of average human gestation can exhibit strong antiviral activity against RNA and DNA viruses, including teratogenic viruses such as Zika virus (ZIKV), rubella (RuV), human cytomegalovirus (hCMV), chickenpox virus (VZV) and herpes virus (HSV-1)[4-7].

In contrast, when pregnant uterus was infected later during pregnancy (after complete development of the placenta by ~ E9.5), only placenta lacking functional type III IFN signaling showed high vertical ZIKV transmission, which correlated with high viral loads on the fetus. fetal death, and also led to congenital malformations. Similarly, the treatment of pregnant uterus carrying wild-type fetuses with recombinant IFN-2 reduced infection with ZIKV by a factor of 2500 compared with untreated pregnant uterus [8]. This mouse study showed that IFN type III can protect against vertical transmission of ZIKV. Another in vivo study also showed that recombinant treatment of IFN-λ2 of pregnant uterus limits the vertical transmission of ZIKV.

While type III IFNs stimulate antiviral protection, which can protect the fetus and placenta from certain viral infections, type I IFNs have the opposite effect, damaging the structure and function of the placenta. Accordingly, human chorionic villi isolated in the second trimester of a human pregnancy treated with recombinant type I rather than type III, IFN-β, showed a large number of syncytial nodes that are associated with damage to the placenta and decreased production of the necessary hormones of pregnancy [9]. This study suggests that type I IFNs can damage the placenta, while type III IFNs may have a protective effect [9].

In order for these drugs to be used in obstetrics and gynecology, they must be checked for mutagenic activity.

At the moment, a prototype of the drug interferon lambda is created, which is a recombinant interferon lambda-1 conjugated with polyethylene glycol by the method of radiation synthesis. This drug is a lyophilized substance for parenteral and enteral administration.

For its use in gynecological practice, it is necessary to make sure that it does not have a mutagenic effect.

The aim of the study was to evaluate the mutagenic properties of the preparation based on PEG-interferon-λ1

in CBA / CaLac mice and in somatic Drosophila melanogaster cells.

Research Objectives:

1. To study the mutagenic effect of the drug (INF-l1) on the appearance of cytogenetic abnormalities in bone marrow cells (hereinafter BMC) during single and course intragastric administration of CBA / CaLac mice by chromosomal aberration accounting.

2. Investigate the mutagenic properties of the drug based on PEG-interferon-λ1 in the test system of somatic mosaicism.

The object of the study was a finished dosage form based on PEG-interferon-λ1, CBA / CaLac mouse and flies (Drosophila melanogaster).

Materials and Methods

The finished dosage form based on PEG-interferon-λ1 is an encapsulated form of a lyophilisate immobilized using ionizing radiation on polyethylene oxide with a molecular weight of 1.5 kDa of recombinant interferon-λ1 isolated from the producer strain Escherichia coli SG20050 / pTINF lamda. The content of PEG-interferon-λ1 per capsule is not more than 480 μg and not less than 320 μg. As fillers, maltodextrin and dextran-40 are present in the preparation.

According to the results of preliminary tests of acute toxicity of the drug based on PEG-interferon-λ1 for the treatment of hepatitis C with intragastric administration to white mice, LD50 was not observed. In a single dose administration experiment, mice were given a dose of 500 μg / kg (as the maximum volume) and 50 μg / kg (1/10 of the maximum). During the course of administration of the drug based on PEG-interferon-λ1 was administered to mice intragastrically for 5 days at a dose of 5 μg / kg, which corresponds to the intended therapeutic dose of ED50.indicated.

Results And Discussion

In mice, after a single and course administration of the drug in the studied doses, no clinical changes in appearance were found. The results of the studies showed that the increase in the mass of male mice at the 1st stage of the study after a single injection of the drug at doses of 500 μg / kg and 50 μg / kg did not change compared to the control and the group of mice that were injected with CF once. At the 2nd stage of the study, in the 6m group, which was administered the drug at a dose of 5 μg / kg, there was a tendency to increase the weight gain of male mice, and the weight of the females did not change.

The study found that the number of damaged cells in the group of mice receiving a single dose of PEG-interferon-λ1 at a dose of 500 μg / kg and 50 μg / kg at 24 hours exposure remained at the level of structural disturbances in the negative control and significantly differed in terms of cytogenetic activity when compared with positive control.

Therefore, a single intragastric administration of the drug PEG-interferon-λ1 at these doses does not increase the level of chromosomal aberrations in the bone marrow cells of mice.

It can be concluded that five-time course administration of PEG-interferon-λ1 at a dose of 5 μg / kg in males and females does not change the proportion of damaged bone marrow cells, does not increase the number of cells with chromosome gaps as compared to the control.

Thus, a single and course intragastric administration of the drug does not increase the proportion of damaged metaphases in bone marrow cells of mice**.**

Conclusion

According to the results of the study, it can be concluded that a single intragastric administration of the studied drug based on PEG-interferon-λ1 at a dose of 500 μg / kg and 50 μg / kg, as well as its course administration (at a dose of 5 μg / kg x 5) do not increases the level of cytogenetic disorders in bone marrow cells of CBA / CaLac mice. Using the yellow and singed markers in the somatic recombination test system, no increase in the number of mutant bristles and spots on the body and head was found in Drosophila melanogaster.

Thus, we can conclude that the finished dosage form based on PEG-interferon-λ1 does not have mutagenic properties in these tests.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

[1] M.D. Robek , B.S.Boyd, F.V. Chisari “Lambda Interferon Inhibits Hepatitis B and C Virus Replication”, *J. Virol*. vol.79, no. 6, pp. 3851-3854, March 2005. doi:10.1128/JVI.79.6.3851–3854.2005

[2] N.E. Pagliaccetti, E.N. Chu, C.R. Bolen, S.H.Kleinstein, M.D.Robek, “Lambda and alpha interferons inhibit hepatitis B virus replication through a common molecular mechanism but with different in vivo activities”, *Virology*, vol. 401, no. 2, pp. 197-206, June 2010. doi:10.1016/j.virol.2010.02.022

[3] Y.S. Wang, S. Youngster, M. Grace J. Bausch, R. Bordens, and D. F. Wyss, “Structural and biological characterisation of pegylated recombinant interferon a-2b and its therapeutic implications”, *Adv Drug Deliv Rev*, vol.54, no. 4, pp. 547-570, 2002. doi:10.1016/S0169-409X(02)00027-3

[4] A. Bayer et al., “Type III interferons produced by human placental trophoblasts confer protection against Zika virus infection”, *Cell Host Microbe* no. 19, pp. 705–712, 2016.

[5] J. Corry et al., “Organotypic models of type III interferon-mediated protection from Zika virus infections at the maternal–fetal interface”, *Proc. Natl. Acad. Sci. U. S. A.*, no. 114, pp. 9433–9438, 2017.

[6] A. Bayer et al, “Human trophoblasts confer resistance to viruses implicated in perinatal infection”, *Am. J. Obstet. Gynecol*., no. 212, pp. 71e1–71e8, 2015.

[7] E. Delorme-Axford et al. “Human placental trophoblasts confer viral resistance to recipient cell”, *Proc. Natl. Acad. Sci. U. S. A*. no. 110, pp. 12048–12053, 2013.

[8] B.W. Jagger et al., “Gestational stage and IFN-l signaling regulate ZIKV infection in utero”, *Cell Host Microbe* no. 22, pp. 366–376, 2017

[9] L.J. Yockey, et al. “Type I interferons instigate fetal demise after Zika virus infection”, *Sci. Immunol*., no 3, eaao1680, 2018

DOI 10.18699/BGRS/SB-2020-382

Differentially expressed genes associated with TMPRSS2-ERG molecular subtype of prostate cancer

Anastasiya Andreevna Kobelyatskaya
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
kaa.chel@mail.ru

Elena Anatolevna Pudova
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
pudova_elena@inbox.ru

George Sergeevich Krasnov
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
gskrasnov@mail.ru

Anna Victorovna Kudryavtseva
Laboratory of Postgenomic Research
EIMB RAS
Moscow, Russia
rhizamoeba@mail.ru

Kirill Mikhailovich Nyushko
Urological department
FSBI NMRRC
Moscow, Russia
kirandja@yandex.ru

Boris Yakovlevich Alekseev
Urological department
FSBI NMRRC
Moscow, Russia
mnioi@mail.ru

Abstract — **Prostate cancer (PC) is one of the most common and socially significant oncological diseases in men. This study examined the transcriptome profile of the most common molecular genetic subtype of prostate cancer, TMPRSS2-ERG. As a result of bioinformatics analysis conducted on the basis of The Cancer Genome Atlas project data, 115 differentially expressed genes were identified for this study group, and in particular, the most over-expressing genes were identified: *ALOX15*, *CACNA1D*, *EML6*, *HLA-DMB*, *NKAIN1*, *OGDHL*, *PLA1A*, *SYT13*. Enrichment pathways analysis showed that these genes are participants in important oncologically significant pathways, which emphasizes the association of this molecular subtype with an unfavorable prognosis for prostate cancer.**

Keywords — TMPRSS2-ERG, TCGA, prostate cancer, sequencing

Introduction

Prostate cancer (PC) is one of the most common and socially significant oncological diseases in men. At the molecular genetic level, prostate cancer can be divided into seven subtypes, the most common among them being the subtype characterized by the presence of the chimeric transcript TMPRSS2-ERG (up to 45% of cases) [1, 2]. This transcript is the result of translocation between exon 1 of *TMPRSS2* gene and exon 4 of *ERG* gene [2, 3]. According to the results of numerous studies, the presence of the TMPRSS2-ERG transcript is most often considered an important predictor of unfavorable prognosis [1, 2, 4]. The aim of our study is to investigate the characteristics of the transcriptome profile of this molecular subtype, which can serve as a basis for understanding the mechanisms of progression in prostate cancer and help in the search for informative prognostic markers.

Methods

The study included PC samples RNA-Seq data of The Cancer Genome Atlas project. The cohort was divided into two groups: tumors with TMPRSS2-ERG fusion transcript (88 cases) and TMPRSS2-ERG-free tumors (117 cases). Samples have appropriate *ERG* expression signature. Differential expression analysis was performed in statistical environment R using the EdgeR package was used. The Mann-Whitney test, Exact test and quasi-likelihood method (QLF) were used for statistical analysis. To exclude false positive results FDR method was used. Enrichment pathways analysis was performed using clusterProfiler and ReactomePA packages. The results were considered statistically significant with a QLF p-value <0.05.

Results

We found out at list 115 differentially expressed genes between the studied groups (fig.1 – Differentially expressed genes TMPRSS2-ERG molecular subtype PC). Among those next genes are of most interest with fold-expression level more than 4 for TMPRSS2-ERG-positive group: *ALOX15*, *CACNA1D*, *EML6*, *HLA-DMB*, *NKAIN1*, *OGDHL*, *PLA1A*, *SYT13*. The following genes showing overexpression in TMPRSS2-ERG-positive group of samples are marked: *ALOX15*, *CACNA1D*, *EML6*, *HLA-DMB*, *NKAIN1*, *OGDHL*, *PLA1A*, *SYT13*. According to the results of enrichment pathways analysis, these genes are participants in the following cancer-significant pathways in the KEGG database associated with the progression: Arachidonic acid metabolism (hsa00590), Focal adhesion (hsa04510), Mucin type O-glycan biosynthesis (hsa00512), Notch signaling pathway (hsa04330), PI3K-Akt signaling pathway (hsa04151), Prostate cancer (hsa05215), Sphingolipid metabolism (hsa00600). Thus, the results underscore the potential association with an unfavorable prognosis in the group of samples belonging to the molecular subtype TMPRSS2-ERG.

Acknowledgment

This work was funded by the Russian Science Foundation grant No. 18-75-10127. This work was performed using the equipment of EIMB RAS “Genome” center (http://www.eimb.ru/ru1/ckp/ccu_genome_c.php).

References

[1] Arora K., Barbieri C.E. (2018) Molecular Subtypes of Prostate Cancer. Curr Oncol Rep. 20(8):58. doi: 10.1007/s11912-018-0707-9.

[2] Cancer Genome Atlas Research Network. (2015) The molecular taxonomy of primary prostate cancer. Cell. 163(4):1011-25. doi: 10.1016/j.cell.2015.10.02.

[3] Krumbholz M., Agaimy A., Stoehr R., Burger M., Wach S., Taubert H., Wullich B., Hartmann A., Metzler M. (2019) Molecular Composition of Genomic TMPRSS2-ERG Rearrangements in Prostate Cancer. Dis Markers. 2019:5085373. doi: 10.1155/2019/5085373.

[4] Wang Z., Wang Y., Zhang J., Hu Q., Zhi F., Zhang S., Mao D., Zhang Y., Liang H. (2017) Significance of the TMPRSS2:ERG gene fusion in prostate cancer. Mol Med Rep. 16(4):5450-5458. doi: 10.3892/mmr.2017.7281.

DOI 10.18699/BGRS/SB-2020-383

Nuclear envelope rupture in Drosophila D11 cells inhibits mitosis

Snezhanna Sergeevna Saydakova
Sector of Cell Structural Biology
ICG SB RAS
Faculty of Natural Sciences
NSU
Novosibirsk, Russia
custodian.of.midnight@gmail.com

Ksenia Nikolaevna Morozova
Sector of Cell Structural Biology
ICG SB RAS
Novosibirsk, Russia
morozko@bionet.nsc.ru

Gera Alekseevna Pavlova
Laboratory of Cell Division
IMCB SB RAS
Novosibirsk, Russia
gera.pavlova@mcb.nsc.ru

Elena Vladimirovna Kiseleva
Sector of Cell Structural Biology
ICG SB RAS
Novosibirsk, Russia
elka@bionet.nsc.ru

***Abstract* — It is already known that lesions in the organization of the nuclear envelope (NE) negatively affects the stability of the genome functioning, transport of molecules between the nucleus and the cytoplasm, as well as the process of cell division. In previous studies we demonstrated the atypical interaction of ER membranes with the outer nuclear membrane (ONM) at the prometaphase stage, leading to the formation of 4-layered membranes and a disassembly delay of the NE [1]. Our current study is focused on the ultrastructural organization of *Drosophila* D11 cells, which shows a significant decrease in the number of dividing cells. Our TEM analysis demonstrated that the nuclei of many cells contain numerous microtubules in the nucleoplasm, slightly compacted chromosomes and an atypical NE with extended folds and outgrowths shaped like 4layered membranes. These structures represent membrane complexes consisting of inner nuclear membranes (INM) layers close attached to each other. The formation of multilayer fragments from adherent outgrowths of the NE was previously observed when the lamina protein Lam or the insect-specific protein KUGELKERN was overexpressed [2]. According to these data we suggest that the low mitotic index of *Drosophila* D11 cells division is associated with misregulation of the mitosis process, possibly due to imbalance in the expression of genes encoding lamina proteins, which stimulates the synthesis and elongation of the nuclear membrane without disassembling it. Excessive NE membranes form loops and folds protruding to the cytoplasm and sticking together due to mutual fusion of the INMs forming the quadruple structures.**

Keywords* — *nuclear envelope, endoplasmic reticulum, MLDmD11 cell line, nuclear lamina, Drosophila, TEM

Introduction

The NE consists of the outer and inner membranes (ONM and INM, respectively) clamped together with nuclear pore complexes attached to the lamina underlying the INM. These are dynamic structures that transform throughout cell cycle stages. The proper organization of the NE depends on the expression level genes encoding the protein components of the ER, ONM, INM and lamina [2]. Defects in the NE organization with quadruple membranes formation was observed in Hela and *C. elegans* cells as the consequence of LIPIN and TMEM170 protein level variability, which are ER and the NE proteins, respectively [3,4]. In 2018 we observed similar quadruples formed in about 70% of all membranes in S2 cells during prometaphase stage [1]. We suggested that these abnormalities could be due to LIPIN and TMEM170 levels misbalance.

The atypical NE organization was shown when lamina proteins were overexpressed in intestinal stem cells and enterocytes of *D. melanogaster* [2]. In particular, overexpression of Lam led to extra NE membranes formation with small fragments of quadruple membranes formed by the INM contacts with itself.

The aim of the current study was to perform TEM analysis of the NE organization in D11 cells which derive from the eyeantennal disk of *D. melanogaster*. These cells showed very low mitotic index suggesting the presence of defects in the NE dynamics.

Materials and methods

Cells ML-DmD11 were maintained at 25º C in Shields and Sang M3 Insect medium supplemented with 10% Fetal Bovine Serum (FBS) and 10 μg/ml insulin [6]. Fixation for the TEM analysis performed according to previously described method [5]. Ultrathin sections were made with the diamond knife (Diatome, Switzerland) on the ultramicrotome (Leica ultracut UCT, Vienna, Austria). The sections were observed under the JEOL1400 electron microscope (Jeol, Japan) in the Interinstitutional center for microscopic analysis of biological objects (ICG SB RAS, Novosibirsk, Russia).

Results and conclusions

Detailed TEM analysis revealed the absence of quadruple membranes formed by the NE and ER stacking on each other similar to those observed in S2 cells. At the same time, we found long quadruple membranes formed by NE folds and loop-like structures (Fig.1). Nuclei with abnormal NE observed in 40% of all examined cells. As a result, these NE fragments form long protrusion into the cytoplasm suggesting the extra NE growth. The presence of numerous microtubules in the nucleoplasm of these cells as well as slightly compacted chromosomes may indicate that they are paused at the beginning of mitosis. The cytoplasm is characterized by high ribosome density and the presence of some swellings of ER that also implies that these cells are at the beginning of mitosis. It is possible that the nuclei with an excess of NE could havedivision problems probably due to lamina defects. It was previously demonstrated that similar structures can be the result of the lamina protein

LAM or the insect KUGELKERN overexpression [2]. Presumably, some lamina proteins are overexpressed in D11 cells causing extra NE membranes growth and as a consequence folds of the NE additional fragments due to fusion of INM on itself. To verify this hypothesis, we are going to measure the expression level of genes encoding lamina proteins in D11 cells.

REFERENCES

[1] A. Strunov et al, "Ultrastructural analysis of mitotic Drosophila S2 cells identifies distinctive microtubule and intracellular membrane behaviors", BMC Biology, vol. 16, №1, p. 68, May 2018.

[2] R. Petrovsky, G. Krohne, J. Großhans, "Overexpression of the lamina proteins Lamin and Kugelkern induces specific ultrastructural alterations in the morphology of the nuclear envelope of intestinal stem cells and enterocytes", European journal of cell biology, vol. 97, №2, pp. 102-113, March 2018.

[3] A. Christodoulou et al, "Transmembrane protein TMEM170A is a newly discovered regulator of ER and nuclear envelope morphogenesis in human cells", J Cell Sci., vol. 129, №. 8, pp. 1552-1565, April 2016.

[4] A Golden., J. Liu, O. Cohen-Fix, "Inactivation of the C. elegans lipin homolog leads to ER disorganization and to defects in the breakdown and reassembly of the nuclear envelope", Journal of cell science,. vol. 122, №. 12, pp. 1970-1978, June 2009.

[5] A. Strunov et al, "A simple and effective method for ultrastructural analysis of mitosis in Drosophila S2 cells", MethodsX, vol. 3,.pp. 551559, October 2016.

[6] https://dgrc.bio.indiana.edu/product/View?product=94

ACKNOWLEDGMENT

The work was supported by budjet project 0324-20190042-C-01.

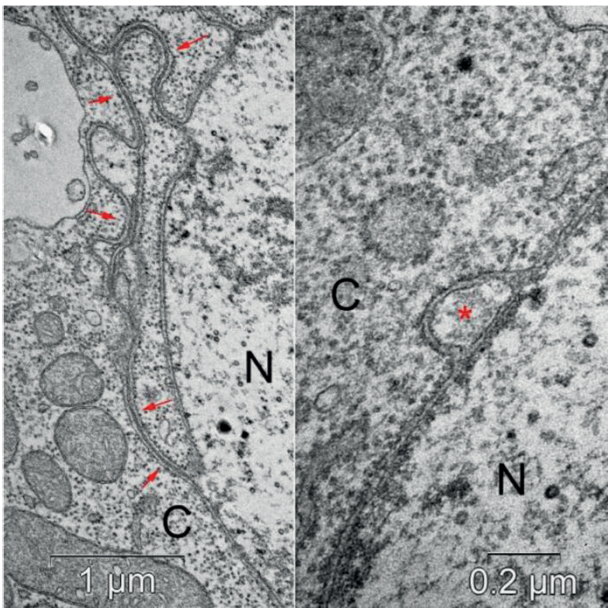


Fig. 1. Abnormal NE in D11 cells. Arrows shows the NE fragments stacked on each other; the loop-like structure is shown with asterisk. N – nucleus, C – cytoplasm. specific protein

DOI 10.18699/BGRS/SB-2020-384

Properties of the C-terminal domain of HlyIICTD suggest that *B. cereus* HlyII is a representative potential member of trimeric autotransporter adhesins among gram-positive bacteria

Siunov A.V.
G. K. Skryabin Institute of Biochemistry and physiology of microorganisms RAS, Pushchino, Russia

Nagel A.S.
G. K. Skryabin Institute of Biochemistry and physiology of microorganisms RAS, Pushchino, Russia

Andreeva-Kovalevskaya Z.I.
G. K. Skryabin Institute of Biochemistry and physiology of microorganisms RAS, Pushchino, Russia

Zamyatina A.V.
Pushchino State Institute of Natural Sciences, Pushchino, Russia
Pushchino Branch, Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Pushchino, Moscow, Russia

Rudenko N.V.
Pushchino State Institute of Natural Sciences, Pushchino, Russia
Pushchino Branch, Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Pushchino, Moscow, Russia

Karatovskaya A.P.
Pushchino Branch, Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Pushchino, Moscow, Russia

Borisova M.P.
Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Russia

Brovko F.A
Pushchino State Institute of Natural Sciences, Pushchino, Russia
Pushchino Branch, Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Pushchino, Moscow, Russia

Salyamov V.I.
G. K. Skryabin Institute of Biochemistry and physiology of microorganisms RAS, Pushchino, Russia

Solonin A.S.
G. K. Skryabin Institute of Biochemistry and physiology of microorganisms RAS, Pushchino, Russia
Pushchino State Institute of Natural Sciences, Pushchino, Russia

Key-words — hemolysin, autotransporter proteins, in pore formation

Hemolysin II of *B. cereus sensu lato* is synthesized in a bacterial cell in the form of a water-soluble secreted monomer and penetrates into eukaryotic membranes forms weakly anion-selective pores, the radius of which varies from 0.6 to 0.8 nm in both natural and model erythrocyte membranes [1]. In a mature state, this toxin is the closest homologue of *S. aureus* alpha toxin (38% amino acids identity). The HlyII protein has a C-terminal extension not previously described in this class of toxins, including 94 amino acid residues designated as HlyIICTD (C-terminal domain) [2]. Removal of HlyIICTD from HlyII does not allow transfer of the deletion variant to *E. coli* cells, possibly due to the attack of the bacterial membrane by the deletion toxin during its accumulation in the periplasm. Additional deletion of the signal peptide, which excludes the penetration of the protein into the periplasm, provides for the cloning of this gene with two deletions in E. coli cells. The deletion variant of hemolysin II has a reduced hemolytic activity. Using monoclonal antibodies against recombinant HlyIICTD, HlyIICTD showed the possibility to bind effectively to red blood cells [3]. Purified HlyIICTD preparations are capable to trimerization in the presence of 4M urea and, forming a possibly beta-barrel-like structure, integrate into the artificial bilayer membrane with the formation of pores. The paper presents the current-voltage characteristic of these pores. Such structures are characteristic of trimeric autotransporter proteins. In this case, the secreted full-sized monomeric form of hemolysin II acts as a passenger, and HlyIICTD acts as an element involved in adhesion. The materials presented in this paper demonstrate the possible functional role of HlyIICTD hemolysin II both in pore formation and in the secretion of toxin from a bacterial cell, suggesting that hemolysin II may belong to trimeric autotransporter proteins – the first case of the description of this family of molecules among gram-positive microorganisms.

References

[1] Andreeva Z.I., Nesterenko V.F., Fomkina M.G., TernovskyV.I., Suzina N.E., Bakulina A.Y., Solonin A.S., Sineva E.V. // Biochim. Biophys. Acta. 2007. V.1768. P.253–263.

[2] Baida G., Budarina Z. I., Kuzmin N.P., Solonin A.S. // FEMS Microbiol. Lett. 1999. V.180. P.7–14.

[3] Rudenko N. V., Karatovskaya A. P., Zamyatina A. V., Siunov A. V., Andreeva-Kovalevskaya Zh. I., Nagel A. S., Brovko F. A., and A. S. Solonin // Russian Journal of Bioorganic Chemistry, 2020, Vol. 46, No. 3, pp. 321–326.

DOI 10.18699/BGRS/SB-2020-385

Mustguseal: versatile bioinformatic platform for knowledge-based protein design and modulation

Dmitry Suplatov
Belozersky Institute of Physicochemical Biology, Lomonosov Moscow State University, Moscow, Russia
d.a.suplatov@belozersky.msu.ru

Yana Sharapova
Belozersky Institute of Physicochemical Biology, Lomonosov Moscow State University, Moscow, Russia
sharapova@belozersky.msu.ru

Vytas Švedas
Belozersky Institute of Physicochemical Biology and Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia
vytas@belozersky.msu.ru

***Abstract* — Systematic bioinformatic analysis of the growing sequence and 3D-structural data representing evolutionarily distantly related proteins with diverse properties has a potential of becoming a particularly important tool in protein engineering and drug discovery, but represents a significant methodological challenge. We have developed the Mustguseal platform – a versatile on-line resource for knowledge-based protein design and modulation to facilitate experimental research, freely available at https://biokinet.belozersky.msu.ru/m-platform.**

Keywords* — *biotechnology, bioinformatic analysis, methods on-line, protein engineering, drug discovery

Motivation and aim

Motivation

Comparative analysis of proteins implementing diverse functional properties within shared 3D-structure of the superfamily is the cornerstone of computational biology to study the relationship between structure and function and solve practical problems of medicine and biotechnology. As number of available protein sequences and 3D-structures has increased exponentially in recent years, their systematic analysis presents growing methodological and computational challenges.

Aim

The common problem of available bioinformatic tools is that they require profound skills in computational biology, and were designed to handle relatively small datasets. This impedes practical use by many investigators in a daily laboratory practice and offers limited productivity at large-scale analysis of families/superfamilies. Our aim was to develop easy-to-use tools featuring content-rich interactive on-line output that will provide advanced computational methods directly to wet-lab scientists.

Results

We have developed the Mustguseal platform – a collection of easy-to-use on-line methods that provide access to comprehensive bioinformatic analysis in a daily laboratory practice. The key web-server Mustguseal can automatically collect and superimpose thousands of sequences and 3D-structures of homologous proteins with high structure, but low sequence similarity to the selected query protein [1]. This data represent all currently known sequence variability within a common fold of a functionally diverse superfamily and can be further studied using the sister web-methods. The Zebra2 web-server can be used to identify subfamily-specific positions in protein structures, i.e. the variable amino acid residues responsible for functional diversity within the common fold, as well as the conserved positions highlighting key catalytic and structural residues. This output can be used to understand how protein/enzyme performs its inherent function, and to select hotspots for rational design or directed evolution experiments [2]. The pocketZebra web-server is an add-on to Zebra2 that maps subfamily-specific positions onto the protein surface to identify and rank binding sites by functional significance, annotate novel regulatory centers, and select particular positions in the structure that are important for selective binding of substrates/inhibitors/effectors [3]. The Yosshi web-server can systematically classify and study disulfide bonds in protein families, and assist the selection of hot-spots for disulfide engineering [4]. The visualCMAT web-server can help to select and interpret correlated mutations/co-evolving residues in protein structures to reveal functionally coupled residues between topologically independent sites, study allosteric communication pathways, and select compensatory mutations for protein engineering [5]. The integration of these web-based bioinformatic tools provides an open-access out-of-the-box solution, first of its kind, to systematically analyze all the available sequence and structural data describing evolutionarily close and distantly related proteins.

We have used the developed tools to assist the experimental research in the fields of protein engineering and drug discovery – to study structure-function relationship in several protein families and design novel enzymes with improved properties (Ntn-hydrolases, penicillin-binding proteins, sialidases, α/β-hydrolases, PLP-dependent enzymes), to identify previously unknown binding sites for modulating ligands and assist the design of complementary selective inhibitors [6-11].

We expect that our bioinformatic tools will further bridge the gap between methods of advanced computational biology and experimental studies, thus promoting the value of bioinformatics in protein engineering and drug discovery.

Acknowledgment

The study was funded by Russian Foundation for Basic Research according to the research project (18-29-13060). The use of HPC computing resources at the Lomonosov Moscow State University supported by the project RFMEFI62117X0011 is acknowledged [12].

References

[1] Suplatov D., Kopylov K., Popova N., Voevodin Vl., & Švedas V. (2018) Mustguseal: a server for multiple structure-guided sequence alignment of protein families. *Bioinformatics*, 34(9): 1583-1585.

[2] Suplatov D., Sharapova Y., Švedas V. (2020) Zebra2: advanced and easy-to-use web-server for bioinformatic analysis of subfamily-specific and conserved positions in diverse protein superfamilies. *Nucleic acids research*. doi:10.1093/nar/gkz385

[3] Suplatov D., Kirilin E., Arbatsky M., Takhaveev V., Švedas V. (2014) pocketZebra: a web-server for automated selection and classification of subfamily-specific binding sites by bioinformatic analysis of diverse protein families. *Nucleic acids research.* 42(W1): 344-349.

[4] Suplatov D., Timonina D., Sharapova Y., Švedas V. (2019) Yosshi: a web-server for disulfide engineering by bioinformatic analysis of diverse protein families. *Nucleic acids research*. 47(W1): 308-314.

[5] Suplatov D., Sharapova Y., Timonina D., Kopylov K., Švedas V. (2018) The visualCMAT: A web-server to select and interpret correlated mutations/co-evolving residues in protein families. *J. Bioinf. Comput. Biol.* 16(02): 1840005.

[6] Suplatov D., Besenmatter W., Švedas V., Svendsen A. (2012) Bioinformatic analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities. *Protein Eng. Des. Sel.* 25(11): 689-697.

[7] Suplatov D., Panin N., Kirilin E., Shcherbakova T., Kudryavtsev P., Švedas V. (2014) Computational design of a pH stable enzyme: understanding molecular mechanism of penicillin acylase's adaptation to alkaline conditions. *PloS one.* 9(6).

[8] Suplatov D., Voevodin V., Švedas V. (2015) Robust enzyme design: Bioinformatic tools for improved protein stability. *Biotechnology J.* 10(3): 344-355.

[9] Fesko K., Suplatov D., Švedas V. (2018) Bioinformatic analysis of the fold type I PLP-dependent enzymes reveals determinants of reaction specificity in l-threonine aldolase from Aeromonas jandaei. *FEBS open bio.* 8(6): 1013-1028.

[10] Sharapova Y., Suplatov D., Švedas V. (2018) Neuraminidase A from Streptococcus pneumoniae has a modular organization of catalytic and lectin domains separated by a flexible linker. *FEBS J.* 285(13):2428-2445.

[11] Suplatov D., Shmalgauzen E., Muronets V., Švedas V. (2018) Selective inhibitors of glyceraldehyde-3-phosphate dehydrogenase of Mycobacteria. Patent #2661151RU.

[12] Sadovnichy V., Tikhonravov A., Voevodin V., Opanasenko V. (2013) Lomonosov: supercomputing at moscow state university. In Vetter J. (ed.) Contemporary High Performance Computing: From Petascale Toward Exascale (Chapman & Hall/CRC Computational Science). Boca Raton, CRC Press, 283–307.

DOI 10.18699/BGRS/SB-2020-386

easyAmber: a step away from inefficient "static" approaches towards a deeper understanding of protein dynamics

Dmitry Suplatov
Belozersky Institute of Physicochemical Biology, Lomonosov Moscow State University, Moscow, Russia
d.a.suplatov@belozersky.msu.ru

Yana Sharapova
Belozersky Institute of Physicochemical Biology, Lomonosov Moscow State University, Moscow, Russia
sharapova@belozersky.msu.ru

Vytas Švedas
Belozersky Institute of Physicochemical Biology and Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia
vytas@belozersky.msu.ru

***Abstract* — Molecular dynamics (MD) can improve the success rate of *in silico* tools for drug discovery and protein engineering by accounting for protein structure flexibility/plasticity and predicting the time-dependent behavior of a molecular system, but requires specialized training and skills, what impedes practical use by many investigators. We have developed easyAmber – a comprehensive toolbox to automate the MD routines implemented in the Amber package, freely available at https://biokinet.belozersky.msu.ru/easyAmber.**

Keywords* — *structural biology, molecular dynamics, protein structure flexibility, time-dependent behavior, protein-ligand interaction, computational drug discovery

Motivation and aim

Motivation

Computational biology is overwhelmed by "static" approaches that take into account only a single PDB entry of the protein of interest, e.g., to select hot-spot positions to improve its biocatalytic function or design selective inhibitors as prototypes of novel drugs. Despite some positive progress, the overall success rates of these approaches remain disappointingly low, probably because protein structures are flexible and the corresponding ensemble of conformations has to be explicitly taken into account to understand the structure-function relationship. Molecular dynamics (MD) can address the challenge of accounting for protein flexibility, but presents a very complex method that requires specialized training and advanced computer experience, thus significantly limiting its everyday use.

Aim

The aim was to develop comprehensive and easy-to-use assist software to automate the simulation workflow and thus improve the user's experience with MD.

Results

We have developed the easyAmber – a set of wrapper scripts as a supplement to the Amber suite to assist the MD simulation starting from the 3D-model of a protein/protein-ligand complex. The **autoligand.pl** master-script assists the preparation of the Amber parameter files for a custom low-molecular-weight ligand. The **automodel.pl** master-script assists the preparation of a full-atom molecular system in water. The **autorun.pl** master-script accommodates the seven-step MD simulation pipeline: the initial optimization of the molecular system (i.e., three-step energy minimization with water relaxation), heating (in the NVT ensemble), equilibration (in the NPT ensemble), followed by the classical/conventional MD (in the NVT ensemble), and optionally concluded by the accelerated MD simulation (in the NVT ensemble). Thus, two major MD methods are supported – the classical and accelerated MD – that together can help to assess protein structure flexibility on a wide range of timescales.

The easyAmber implements advanced MD protocols, but is easy-to-operate on a personal desktop station equipped with a compatible gaming GPU-accelerator, as well as a powerful supercomputer, thus significantly increasing productivity. The new toolkit was used in recent studies of conformational plasticity in NanA from *S.pneumoniae* and human p38α MAPK, both featuring well-known targets for drug design, to operate hundreds of independent trajectories at various settings worth tens of microseconds, using various HPC resources [1-3]. A step-by-step practical tutorial to the easyAmber was published in the corresponding paper [4], with supplementary technical documentation available on-line at https://biokinet.belozersky.msu.ru/easyAmber#techdoc. We hope that easyAmber will contribute to a more widespread use of MD in a daily laboratory routine, supporting a recent trend away from inefficient "static" approaches in protein sciences towards a deeper understanding of structural dynamics, leading to a more efficient selection of hot-spots for protein engineering and complementary ligands/inhibitors by companion *in silico* tools.

Acknowledgment

The study was funded by Russian Foundation for Basic Research according to the research project (19-04-01297). The use of HPC resources at the Lomonosov Moscow State University supported by the project RFMEFI62117X0011 is acknowledged [5].

References

[1] Sharapova Y., Suplatov D., Švedas V. (2018) Neuraminidase A from Streptococcus pneumoniae has a modular organization of catalytic and lectin domains separated by a flexible linker. *FEBS J.* 285(13):2428-2445.

[2] Suplatov D., Kopylov K., Sharapova Y., Švedas V. (2019) Human p38α Mitogen-Activated Protein Kinase in the Asp168-Phe169-Gly170-in (DFG-in) state can bind allosteric inhibitor Doramapimod. *J. Biomol Struct. Dyn.* 37(8): 2049-2060.

[3] Sharapova Y., Suplatov D., Švedas V. (2018) Simulating the Long-timescale Structural Behavior of Bacterial and Influenza Neuraminidases with Different HPC Resources. *Supercomput. Front. Innov.* 5(3): 30-33.

[4] Suplatov D., Sharapova Y., Švedas V. (2020) easyAmber: a comprehensive toolbox to automate the molecular dynamics simulation of proteins. *J. Bioinf. Comput. Biol.* in press

[5] Sadovnichy V., Tikhonravov A., Voevodin V., Opanasenko V. (2013) Lomonosov: supercomputing at moscow state university. In Vetter J. (ed.) Contemporary High Performance Computing: From Petascale Toward Exascale (Chapman & Hall/CRC Computational Science). Boca Raton, CRC Press, 283–307.

DOI 10.18699/BGRS/SB-2020-387

An effective molecular blockers of ion channel of M2 protein as anti-influenza A drug

Yury Nikolaevich Vorobjev
Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of Russian Academy of Sciences, Novosibirsk 630090, Russia
ynvorob@niboch.nsc.ru

Key words — Influenza A; M2 protein; ion channel blockers

Motivation:

Design of a drug molecule that can effectively bind to the M2 ion channel and block a diffusion of ions H^+ through and inhibit influenza A virus replication is an important task.

Methods

A new class of positively charged, +2, molecules is proposed to block diffusion of H^+ ion through the M2 channel. Several drug candidates, derivatives of a lead compound (diazabicyclooctane), is proposed and investigated. Molecular dynamics of thermal fluctuations of M2 protein structure and ionization-conformation coupling of all the ionizable residues were simulated at physiological pH via original methods [1]. The influence of the most probable mutations of key drug-binding amino acid residues in the M2 ion channel are investigated too.

Results

It is shown that the suggested blocker drug molecule has high binding affinity for the native and mutant M2 ion channel. There are two in-channel binding sites of high affinity for the native M2 protein, the first one demonstrates formation of two H-bonds with two of four serine residues Ser-31A (B) or Ser-31C(D), and the second one has H-bonds with two of four histidine residues His-37A (B), or His-37C(D). Six types of the most probable mutations of residues Ser-31A(B,C,D) are analyzed by the same computation protocol, as for the native M2 protein, and it is shown that the binding site with His-37A(B,C,D) residues is highly conservative with high binding affinity. Probability of double mutants, namely Ser-32 and His-37 is quite low and does not exceed 10^{-5}.The main advantages of the new drug molecule is the positive charge, +2, which creates a positive electrostatic potential barrier (in addition to a steric one) for a transfer of H^+ ion through M2 ion channel and may serve as an effective anti-influenza A virus drug.

ACKNOWLEDGEMENTS

This research was supported by the Russian Foundation for Basic Research (№ 18-04-00005-a) and a Russian-Government–funded budget project for the Institute of Chemical Biology and Fundamental Medicine, the Siberian Branch of the Russian Academy of Sciences: № AAAA-A17-117020210022-4.

REFERENCES

[1] Vorobjev, Y.N. Modeling of Electrostatic Effectsin Macromolecules, 163-202. Computational Methods to Study the Structure and Dynamics of Biomolecules. Springer Nature, 2019. ISSN 978-3-319-95843-9; https://doi.org/10.1007/978-3-319-95843-9_6

Section
Systems biology of aging

DOI 10.18699/BGRS/SB-2020-388

Cluster analysis of age-related trends of the expression of metabolically relevant genes in humans PBMCs

Aleksey Alekseev
MSU, Moscow, Russia
alekseev@physics.msu.ru

***Abstract* — New methodology of cluster analysis of gene expression trends with averaging using a new moving average method was developed and applied to metabolically important genes in human PBMC cells. Gene expression trends were obtained from 5 independent datasets (total number of people was 2503, including 1551 males and 952 females), and an original algorithm was developed and implemented to harmonize them. An algorithm for detecting communities on an undirected graph was used, where genes were the nodes, and the edges represented strong direct and inverse correlation of age-related gene expression trends. Optimal values of the hyperparameters for the clustering procedure were found, and the averaging procedure that developed for the division of the genes into clusters. Non-monotonic behavior with local minima and maxima was shown for the trends of expression of some genes, which may correspond to the transition processes between some stages of human aging.**

Keywords* — *bioinformatics, human aging, RNA-seq data harmonization, PBMC, trend analysis, graph clustering

Context and the goal

Introduction

System biology research in the aging, especially human aging, is an urgent, but still underdeveloped area of modern bioscience. Currently, the primary omics data on biomarkers of aging and age-associated genes is insufficient for constructing a generalized theory of human aging, which would explain the facts on the pathophysiology of aging. So far, the aging can be defined as a set of underlying pathological interconnected processes at various levels of organization of human body [1].

A number of genes is known to be associated with human aging and the development of aging-related diseases [2]. However, the relationships between these processes and changes in the levels of gene expression has not been sufficiently studied. Even for genes and proteins with known association with lifespan, which signaling pathways are well studied (such as AMPK, the regulator of cell metabolism, or transcription factor NF-κB, etc.), we do not see unambiguous relationships between gene expression change and the major pathological processes of aging, such as mitochondrial dysfunction, systemic insulin resistance, etc.). There are not many published papers on gene expression changes in various types of human cells or tissues during aging. This is due to the deficiency of open datasets on gene expression in human tissues or cells with large samples and age and sex indication.

Data analysis is also complicated by inter-individual, within-individual and intersex variability, which leads to "fuzzy" distribution of the data on the graphs representing the changes of gene expression during aging. At the same time, average dynamics (trends) of expression of individual genes is of considerable interest, since it can allow us to associate the directions and patterns of such trend curves with the pathological processes of aging, changes in the tissue cellular composition, and the features of the extracellular environment (concentration of various metabolites, signaling molecules, etc.). Methods for clustering age-related trends for proteomic and metabolomic data have been published recently [3, 4]. This report is one of the first analyses of age-related gene expression trends.

Research goals and tasks

The goal of this study is to identify the trends of expression of the key metabolic genes during aging, as well as the gene clusters based on the dynamics of expression of such genes. This requires solving several tasks:

- development of an approach to juxtapositioning data from various datasets (so-called "data harmonization");
- development of a method of averaging the merged data and obtaining gene expression trends for the set of metabolically important genes available in all datasets;
- development of a method for assessing the trends of correlation between the expression of each pair of genes;
- development and application of the methods of clusterization of gene expression trends (finding communities of correlated trends in the graph) to allow identification of the groups of gene with similar dynamics during aging;
- analyzing the clusters of the trends in expression of metabolically significant genes in human peripheral blood mononuclear cells (PBMCs).

Methods

337 genes from human PBMCs were analyzed (separately for men and women). The groups of genes of interest are related with the use or the production of NAD+ and NADPH in cells, the Krebs cycle, glycolysis, pentose phosphate and glutathione metabolic pathways, as well as AMPK. The following method of identifying the gene expression trends and their clusterization was developed.

1. The used pipeline (a script on R language) allowed obtaining gene expression trends in PBMCs of healthy people using datasets from the GEO database [5], including GSE75511, GSE30483, GSE47353, GSE68759, and GSE65907, which contained information on sex and age of people. The dataset GSE65907 with a large sample and a wide age range was chosen as the reference. The data from men and women were merged for normalization. The normalization was performed for each gene from each dataset, except for the referential one (GSE65907), by comparing distribution density of the gene expression levels with that of the same genes in the reference dataset. The distribution turned out to be close to normal. The points for aligning the distribution densities were selected taking into account the differences in the distribution of people by age in each of considered and the reference datasets. Since age distribution was different in the sample, the samples of 28 randomly chosen individuals were selected for each dataset including the reference one for gene alignment.The samples of the selected individuals where only of these age ranges for which age distribution intersected on both sides of the distribution

interval boundaries. A linear function was developed to calculate the normalized expression values for each gene ELnorm (gene) = EL (gene) * k + b, where the pairs of the parameters (k, b) were obtained by iterative minimization. Optim function of the standard set of the R language was used, and random pairs of numbers in a wide range of values were used as initial parameters values. The minimization target function was defined by the proximity between 10% and 90% quantiles in the reference dataset (in the age-limited samples, as described above) and 10% and 90% quantiles of normalized sample from a considered dataset. The sets of best pairs of parameters (k, b) for each dataset and for each gene were obtained separately.

2. The averaging procedure was performed via a new, a specially developed moving average method with varying window width (R code) to obtain a trend for each gene for the combined sample of the data points. The algorithm changes the moving averaging window so that the points from a certain age range fall into it, which was necessary because the distribution of different age groups varied from dataset to dataset. Moreover, the selection of the best pairs of parameters from each non-reference datasets and the reference dataset (k, b) was carried out randomly 20 times, and a set of the points normalized in different ways was formed. Thus, 20 sets of the pairs were obtained and processed by the moving average procedure. Each set was smoothed via cubic spline (standard R language smooth.spline function) and then, each ultimate gene expression trend was calculated as the average for the 20 smoothed trends.

3. The procedure of iterative calculation of the correlation coefficients in a moving window with a certain step was used to cluster the trends for each pair of genes. As a result, a set of correlation coefficients was obtained (ranging from -1 to 1), showing how the trend curves of the two genes were correlated in different age intervals. The distribution density of the correlation coefficients was calculated to identify the strongest correlations within the entire age range. Thus, strongly correlated pairs of the genes were identified, including direct and inverse correlations.

At this stage, the best number of clusters was estimated via hierarchical clustering and clustering of k-means using the Calinsky-Harabasz index [6]. The connection graph was constructed where the connections imply strong correlation of the gene expression trends. I searched for the communities in the graphs obtained from the data from men and women calculating the distribution of genes in the clusters based on the trends of their expression during aging as well as the modularity of graphs using several procedures of the iGraph library of the R language. Functions from d3Network R package were used for plotting the interactive graphs. Figure 1 (Supplementary materials link) shows the gene community graph which was produced based on the patterns of correlations between the genes, using community search methods.

Results

Age-related gene expression trends were calculated for the considered genes. Interestingly, the number of gene expression trends have distinct non-monotonic shape. Thus, the critical age points obtained for men were 34, 43, and 50. The genes with minimum or maximum gene expression at the age of 43 may be associated with metabolic transition in men at this age point, associated with reduced ability to increase NAD+ levels in cells [7] (e.g., IDH3A, SQLE, SPARC, PPBP). Trends in expression of a group of genes display clear non-monotonous behavior near the age of 50 years (AASDHPPT, AGPAT2, DHCR7, MGST2, NDUFB1, NDUFA8, POR), that can be connected to significant increase in AMP concentration in the cells in 50 years [8]. However, the expression of the genes of AMPK subunits (PRKAA1, PRKAB1, PRKAB2, PRKAG1, PRKAG2) is not nonlinear at this age point. Strong correlation between the trends of a priori coexpressed genes during aging separately in men and women validates the new approach and the biological meaning of the results, as well as their similarity to the previously published data on gene expression trends for some of the genes.

Discussion

A large number of gene expression trends was obtained in this data analysis. Further detailed investigations and comparison of gene expression trends to the dynamics of the pathophysiological processes of aging, as well as to the metabolic and proteomic data. It is also necessary to validate the findings of this study with similar analysis of data from other tissues (muscles, liver, adipose tissue, etc.) to separate tissue-specific patterns of gene expression and the effects of aging. It would be also important to take into account the influence of cellular composition of PBMCs on age-related gene expression changes, since some gene expression patterns may be cell-specific. The development of such methods for gene expression analysis in various tissues in different phenotypic groups (different sexes, BMI levels, etc.) can open ways to create a unified omics databases on aging trends with different types of primary data, which would be a resource for developing system biology models of specific aging-related pathological processes and human aging as a whole.

Acknowledgment

The author is grateful to Daria Khaltourina for fruitful scientific discussions, as well as support in editing and text design.

Supplementary materials

The Figures and the tables are available online: https://yadi.sk/d/vdPAXhZ8c1C9Mg

References

[1] Khaltourina, D., et al (2020) Aging Fits the Disease Criteria of the International Classification of Diseases. Mechanisms of Ageing and Developmen: 111230.

[2] Tacutu, R. et al (2018) Human Ageing Genomic Resources: new and updated databases. Nucleic Acids Research. 46(D1): D1083-D1090.

[3] Lehallier, B. et al (2019) Undulating changes in human plasma proteome profiles across the lifespan. Nature Medicine. 25(12): 1843-1850.

[4] Bunning, B. J. et al (2020) Global metabolic profiling to model biological processes of aging in twins. Aging cell. 19(1): e13073.

[5] Gene Expression Omnibus - https://www.ncbi.nlm.nih.gov/geo/

[6] Caliński, T., & Harabasz, J. (1974) A dendrite method for cluster analysis. Communications in Statistics-theory and Methods. 3(1): 1-27.

[7] Clement, J. et al (2019) The plasma NAD+ metabolome is dysregulated in “normal” aging. Rejuvenation research, 22(2): 121-130.

[8] Ravera, S. et al (2019) Discrete changes in glucose metabolism define aging. Scientific reports. 9(1): 1-8.

DOI 10.18699/BGRS/SB-2020-389

Spatial learning as activator of hippocampal neurogenesis during aging and development of Alzheimer's disease-like pathology

Alena Burnyasheva
Molecular mechanisms of aging
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
burnyasheva@bionet.nsc.ru

Tatiana Kozlova
Molecular mechanisms
of aging
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
kozlova@bionet.nsc.ru

Ekaterina Rudnitskaya
Molecular mechanisms of aging
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
rudnickaya@bionet.nsc.ru

Natalia Stefanova
Molecular mechanisms of aging Institute
of Cytology and Genetics SB RAS,
Novosibirsk, Russia
stefanovan@bionet.nsc.ru

Abstract — **Adult neurogenesis in dentate gyrus (DG) is one of the key mechanisms of neuronal plasticity in hippocampus and plays an important role in cognitive function. However, the consequences of its alteration during healthy aging as well as development of neurodegeneration including Alzheimer's disease (AD) remain unclear. It was shown that factors which can activate neurogenesis – such as physical exercises and learning – are able to improve cognitive function. Animal models are useful to clarify the connection between adult neurogenesis and cognitive function during development of AD signs, and OXYS rats are a suitable model for the most common sporadic form of AD. Here we examined effects of spatial learning on neurogenesis in DG of OXYS rats prior to and during manifestation of AD signs. We showed altered reference memory of OXYS rats already at the period prior to neurodegeneration. At the period of active manifestation of AD signs OXYS rats demonstrated altered spatial learning and reversal learning, whereas reference memory was altered only a little. At the period of active amyloid-β accumulation in the brain only reference memory of OXYS rats was altered. Spatial learning resulted in accelerated maturation of immature cells of neuronal and astrocytic cell lineages in DG of OXYS and Wistar rats and decrease of amyloid-β content in aged animals.**

Keywords — neurogenesis, spatial learning, Alzheimer's disease, OXYS rats

Introduction

Alzheimer's disease (AD) is detrimental multifactorial disorder developing asymptomatically for many years prior to its manifestation [1]. There is no effective cure for AD to date, however interventions prior to manifestation of the disease symptoms are able to improve cognitive function in patients with mild cognitive impairment and thus to slow down or even prevent development of AD [2]. Indeed, several studies revealed positive associations of physical activity, cognitive training or both with cognition in elderly people and patients with mild cognitive impairment [3-5]. These cognitive improvements may be achieved because of activation of neuroplasticity: it is well known, that physical exercises and cognitive training result in activation of neurogenesis in hippocampal DG of adult animals [6]. However, the precise mechanisms underlying cognitive improvement in aged animals and, moreover, elderly people are still unknown, and its examination requires the use of suitable animal models for human neurodegenerative disorders. Previously we have shown that senescence-accelerated OXYS rats may be considered as an adequate model of the late-onset sporadic AD because of disease signs are developed spontaneously without mutations in *App*, *Psen1* and *Psen2* genes. First neurodegenerative changes occur in OXYS rats at 3 months of age, and active amyloid-β accumulation in the brain observed at 12 months of age [7].

Materials and methods

To evaluate spatial learning, reversal learning and reference memory of OXYS and control Wistar rats at 1.5, 3 and 12 months of age (n = 8 animals per strain and age) we used Morris water maze. To analyze changes in hippocampal neurogenesis of OXYS and Wistar rats we evaluated density of progenitors and cells from neuronal and astrocytic lineages (n = 3 to 6 per group, strain and age) in DG and the content of amyloid-β by immunohistochemistry using antibodies specific for molecular markers of various cell types as well as for amyloid-β.

Results

Decrease of reference memory in OXYS rats was observed already at 1.5 months of age; however, spatial learning and reversal learning did not differ compared to Wistar rats. At the period of active manifestation of AD meaning 3 months of age OXYS rats demonstrated totally altered spatial learning and reversal learning, whereas reference memory was altered only a little. We observed decrease of reference memory in OXYS rats at stage of amyloid-β accumulation in the brain (12 months of age); however, we did not show any differences in learning and reversal learning abilities between OXYS and Wistar rats at this age. Learning in the Morris water maze from 1.5 months of age resulted in accelerated maturation of immature cells of neuronal and astrocytic cell lineages in Wistar rats and only immature cells of neuronal cell lineage in OXYS rats. Learning from 3 months of age accelerated maturation of immature cells of neuronal lineage in Wistar rats as well as activated astrocytogenesis in both OXYS and Wistar rats. Learning from 12 months of age did not affect cell densities in DG of Wistar rats and resulted in accelerated maturation of immature cells of astrocytic lineage in hippocampal neurogenic niche in OXYS rats. Learning from 12 months of age also affected amyloid-β content in DG: indeed, the parameter was lower in animals trained in Morris water maze.

This work was supported by the Russian Foundation for Basic Research (# 18-015-00320).

Conclusion

We demonstrated that spatial learning affected neurogenesis, however the impact depended on the age at which the animals were trained and the stage of AD-like pathology. Naturally, this positive effect was stronger for young adult rats and less clear when signs of AD in OXYS rats were already developed. Learning also lowered the content of amyloid-β in the DG of aged animals. Thus, hippocampal-dependent learning may be considered as a perspective strategy to intensify neuronal plasticity and thereby improve cognitive function even in aged animals and probably in elder people.

ACKNOWLEDGMENT

Senescence-accelerated OXYS rats and Wistar rats were obtained from the Breeding Experimental Animal Laboratory of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia. The microscopy was conducted at the Multi-Access Center for Microscopy of Biological Objects (Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia).

REFERENCES

[1] M. Crous-Bou, C. Minguillón, N. Gramunt, J. L. Molinuevo, "Alzheimer's disease prevention: from risk factors to early intervention," Alzheimers Research Therapy, vol. 9, 2017, No. 71.

[2] G. K. Bhatti, A. P. Reddy, P. H. Reddy, J. S. Bhatti, "Lifestyle Modifications and Nutritional Interventions in Aging-Associated Cognitive Decline and Alzheimer's Disease," Frontiers in Aging Neuroscience, vol. 11, 2020, pp. 1-15.

[3] M. Roig, S. Nordbrandt, S. S. Geertsen, J. B. Nielsen, "The effects of cardiovascular exercise on human memory: A review with meta-analysis," Elsevier, vol. 37, 2013, pp. 1645-1666.

[4] A. Lampit, H. Hallock, M. Valenzuela, "Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers," PLoS Med, vol.11, 2014, pp. 1-18.

[5] T. Ngandu, J. Lehtisalo, A. Solomon, E. Levalahti, S. Ahtiluoto, R. Antikainen et al., "A 2 year multidomain intervention of diet, exercise, cognitive training and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial," Lancet, vol. 385, 2015, pp. 2255–2263.

[6] P. Baptista and J. P. Andrade, "Adult Hippocampal Neurogenesis: Regulation and Possible Functional and Clinical Correlates,"Fronties in Neuroanatomy, vol.12, 2018, pp. 1- 23.

[7] N. A. Stefanova, O. S. Kozhevnikova, A. O. Vitovtov, K. Y. Maksimova, S. V. Logvinov, E. A. Rudnitskaya, E. E. Korbolina, N. A. Muraleva, N. G. Kolosova, "Senescence-accelerated OXYS rats: A model of age-related cognitive decline with relevance to abnormalities in Alzheimer disease," Cell Cycle, vol. 13, 2014, pp. 898–909.

DOI 10.18699/BGRS/SB-2020-390

Search for single nucleotide polymorphisms (SNPs) associated with hypertension in the genome of senescence-accelerated OXYS rats

Vasiliy A. Devyatkin
Laboratory of molecular mechanisms of aging
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
devyatkin@bionet.nsc.ru

Natalia A. Muraleva
Laboratory of molecular mechanisms of aging
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
myraleva@bionet.nsc.ru

Olga E. Redina
Institute of Cytology and Genetics
SB RAS, Novosibirsk, Russia
oredina@bionet.nsc.ru

Nataliya Kolosova
Laboratory of molecular mechanisms of aging
Institute of Cytology and Genetics
SB RAS Novosibirsk, Russia
kolosova@bionet.nsc.ru

Abstract — **Aging is a risk factor for many diseases, but the the chance of development with age also depends on genetic factors, environmental conditions, lifestyle and the presence of other pathologies. The OXYS rat strain (ICG SB RAS) is a unique model for studying the mechanisms of aging, since already at an early age these animals develop a whole complex of age-dependent diseases, including cataract, retinopathy, osteoporosis, hypertension and Alzheimer's Disease-like pathology. While hypertension has typical for an age-related disease risk factors, it itself is a risk factor for many other pathologies. However, a complex senile phenotype does not appear in other hypertensive models even with a higher blood pressure. The aim of this study was, based on the results of RNA-Seq, to search for single nucleotide polymorphisms that could contribute to the development of hypertension in senescence-accelerated OXYS rats. We found that OXYS rats are genetically distant from other strains and presumably have their own basis for the development of hypertension, which may determine the absence of senile phenotype of OXYS rats in the hypertensive rat strains.**

Keywords — hypertension, SNPs, aging, OXYS rats

Introduction

Aging is a risk factor for many diseases, but the the chance of development with age also depends on genetic factors, environmental conditions, lifestyle and the presence of other pathologies. Currently, studies on biological models, the genetic uniformity and standard conditions of detention of which increase the reproducibility of the results and reduce the influence of external factors, are a productive approach to elucidating the genetic overlap between age-related diseases. The OXYS rat strain, developed in ICG SB RAS, is a unique model of accelerated aging, which already at early age is characterized by a phenotype similar to geriatric diseases in humans, including cataracts, cardiomyopathy, hypertension, retinopathy and neurodegenerative brain pathology with the signs of Alzheimer's disease [1].

The destructive effect of hypertension on blood vessels, leading to hypoperfusion and metabolic stress, is considered in many studies as a risk factor for cataracts, cardiovascular diseases, osteoporosis and neurodegenerative changes in the brain.

At the age of 3 months in OXYS rats was detected the moderate increase in blood pressure. At the same time the first manifestations of other pathologies in OXYS rats appear. In hypertensive rat strains or induced hypertension models many age-dependent diseases do not occur or manifest in a milder form, in contrast to the complex manifestation in OXYS rats. Its may indicate the importance of the genetic component in the development of age-related diseases.

We previously showed that the genome of the OXYS rat contains a large number of single nucleotide polymorphisms (SNPs) not found in other rat strains. The purpose of this study was to compare the data of transcriptome of OXYS rats with the genotypes of hypertensive rat strains for found the SNPs that could contribute to the development of hypertension in senescence-accelerated OXYS rats.

Methods

Male OXYS rats were used at ages 20 days and 3, 5, and 18 months, while age-matched male Wistar rats served as controls (3–5 per group). The animals were kept at the Center for Genetic Resources of Laboratory Animals at the ISG SB RAS under standard laboratory conditions (22 ± 2°C, 60% relative humidity, 12 h light/12 h dark cycle, lights on at 9 a.m.). The food and water were available ad libitum. The protocol of the animal experiment was approved by the Bioethical Committee of the Institute of Cytology and Genetics.

We analyzed RNA-Seq data obtained in early studies for each sample of retinal, prefrontal-cortex, and hippocampal RNA by Illumina nonstranded sequencing (on an Illumina GA IIx at Genoanalytica) in accordance with standard Illumina protocols (mRNA-Seq Sample Prep Kit). The sequencing data were prepared and SNPs positions were identified as described previously [2].

The list of SNPs of OXYS rats was compared with RGSC 6.0 data for genome sequences of 45 rat strains and substrains: 11 of these strains/substrains are commonly employed as a normotensive control (FHL/EurMcwi, LN/MavRrrc, LL/MavRrrc, MNS/Gib, SBN/Ygl, SR/Jr, WKY/N, WKY/Gla, WKY/NCrl, WKY/NHsd and WAG/GSto-Icgn), 22 rat strains/substrains serve as control or experimental animals in the studies on various pathological conditions that have no relation to hypertension or aging (ACI/N, ACI/EurMcwi, BBDP/Wor,

BN-Lx/Cub, BN-Lx/CubPrin, BN/SsN, BUF/N, DA/BklArbNsi, F334/N, F344/NHsd, F344/NCrl, SUO_F344, GK/Ox, LE/Stm (SOLiD), LEW/Crl, LEW/NCrlBR, LE/Stm (Illumina), M520/N, MR/N, WAG/Rij, WN/N and Wistar) and 12 strains and substrains manifestated signs of hypertension (FHH/EurMcwi, LH/MavRrrc, MHS/Gib, SBH/Ygl, SHR/OlaIpcv, SHRSP/Gla, SHR/NCrlPrin, SHR/NHsd, SHR/OlaIpcvPrin, SS/Jr, SS/JrHsdMcwi and ISIAH/Icgn).

Functional enrichment analyses for the list of genes with SNPs was performed with DAVID tool (https://www.david-d.ncifcrf.gov/). List of the quantitative trait loci (QTLs) overlapped with the SNPs position was received using the RGD (https://rgd.mcw.edu/).

The identity by state (IBS) analysis of the alleles performed by SNPRelate in the R Software Environment was used to measure the distances between objects for dendrogram construction and for demonstrating the results of principal coordinates.

Results

A comparison of transcriptome sequences of the prefrontal cortex, hippocampus, and retina of OXYS rats with the reference genome of BN/NHsdMcwi rats revealed 42,478 SNPs in 9,903 genes. Analysis of the genetic similarity based on the identified SNPs showed that OXYS/Icgn rats are genetically closer to the ISIAH/Icgn rat strain, which was also developed from Wistar/Icgn rats, and two strains of Wistar Albino Glaxo rats (WAG/GSto-Icgn and WAG/Rij) than to 41 strains and substrains taken in the analysis.

Among the 42,478 SNPs found in OXYS rats, 40,373 SNPs in 9,699 genes have also been detected in the genomes of normotensive rat strains or strains used in studies on various pathological conditions that have no relation to hypertension or age-related diseases. Group of the remaining 2,105 SNPs includes 725 polymorphisms, which are found in both OXYS rats and in one or more hypertensive rat strains/substrains taken in analysis, as well as 1,380 polymorphisms that are specific to OXYS rats.

Using multidimensional scaling, a similarity between the genotypes of OXYS rats and 12 strains/substrains of hypertensive rats was evaluated. The data obtained show the similarity of the genotypes of OXYS and ISIAH/Icgn rats and their significant differences from the genotypes of strains and substrains of other hypertensive animals. Indeed, among the 725 SNPs that were detected in the hypertensive strains/substrains and OXYS rats, 663 SNPs were common to OXYS/Icgn rats and ISIAH/Icgn rats and were not found in other hypertensive rats. No common SNP was found for all hypertensive rat strains and substrains taken in the analysis. The maximum frequency of occurrence of SNPs detected in hypertensive strains/substrains was eight out of thirteen, with 4 out of this 8 strains represented by the substrains of SHR rats. This result indicates the differences in the genetic basis of different forms of hypertension.

A functional annotation of genes with SNPs common to OXYS rats and one or more hypertensive rat strains revealed enrichment in such functional groups as GTP/ATP binding, various signaling systems and cell division. Genes with SNPs specific for OXYS rats are associated with synapses/intercellular contacts, kinases, ATP binding, various signaling systems, MHC, infection response and calcium binding. Many of those categories in both group involved in the development of hypertension according to various studies.

17 and 23 SNPs from those of 725 and 1380, respectively, presumably have significant effect on mRNA structure or protein function, that we described in detail in our previous work [2]. Among them the *Ephx1* gene annotated in RGD as associated with hypertension. In addition, *Ephx1*, as well as several other genes (*Pla2r1*, *Zmym6*, *Trappc9*, *Nqo2*) are associated with neurodegenerative diseases and/or mental disorders. Harmful polymorphisms are located in 313 and 451 QTLs. The most common parameters associated with these QTLs are blood pressure, bone structure and strength, bone mineral density, body weight, cardiac mass, kidney mass, diabetes mellitus, renal function, serum cholesterol and corticosterone levels, anxiety response.

Conclusion

Transcriptome of OXYS rats contains 725 common SNPs with one or more hypertensive rat strains and substrains. Analysis of genes with SNPs showed that only one gene *Ephx1* annotated in RGD as associated with hypertension. In addition, *Ephx1*, as well as several other genes (*Pla2r1*, *Zmym6*, *Trappc9*, *Nqo2*) are associated with neurodegenerative diseases and/or mental disorders. Multidimensional scaling analysis showed that OXYS rats are genetically distant from other strains and presumably have their own basis for the development of hypertension, which may determine the absence of senile phenotype of OXYS rats in the hypertensive rat strains.

Acknowledgment

This work was supported by the Budget Project (# 0259-2019-0002)

References

[1] Kolosova, N.G., Stefanova, N.A., Korbolina, E.E., Fursova, A.Zh., Kozhevnikova, O.S. The Senescence-Accelerated Oxys Rats--A Genetic Model of Premature Aging and Age-Dependent Degenerative Diseases. Adv Gerontol. 2014, 27(2), 336–340.

[2] Devyatkin, V.A., Redina, O.E., Kolosova, N.G., Muraleva, N.A. Single-Nucleotide Polymorphisms Associated with the Senescence-Accelerated Phenotype of OXYS Rats: A Focus on Alzheimer's Disease-Like and Age-Related-Macular-Degeneration-Like Pathologies. J Alzheimers Dis. 2020, 73(3), 1167–1183. https://doi.org/10.3233/JAD-190956.

DOI 10.18699/BGRS/SB-2020-391

Way to longevity: role of antioxidant defense gene polymorphisms in successful adaptation

Vera Erdman
IBG UFRC RAS, Ufa, Russia
danivera@mail.ru

Timur Nasibullin
IBG UFRC RAS, Ufa, Russia
nasibullintr@yandex.ru

Ilsia Tuktarova
IBG UFRC RAS, Ufa, Russia
iltuk@mail.ru

Ksenia Danilko
BSMU, Ufa, Russia
kse-danilko@yandex.ru

Olga Mustafina
IBG UFRC RAS, Ufa, Russia
anmareg@mail.ru

Tatiana Viktorova
BSMU, Ufa, Russia
t_vict@mail.ru

Alisa Matua
SRI EPT ASA, Sukhum, Abkhazia
azmatua@mail.ru

***Abstract* — We carried out the analysis of associations between polymorphic loci of antioxidant defense genes with ethnicity and longevity. We found the interethnic differences in the distribution of allele frequencies of *SOD1*, *SOD2*, *CAT*, *NQO1* genes. For reaching longevity *SOD1*, *SOD2*, *NQO1*, *GPX1* genes were significant among Russians, *SOD2*, *CAT* genes – among Tatars, *MSRA*, *CAT* genes – among Bashkirs.**

Keywords* — *human longevity, adaptation, antioxidant defense, associative analysis, genetic polymorphism

Motivation and aim

Motivation

Longevity is a complex phenomenon of surviving to an age significantly exceeding the average species lifespan. Among possible causes of aging and longevity about 25% are genetic and 25% are external factors, while the majority - about 50% - is the way of carry out the interaction of exogenous and endogenous factors. To date, the question remains as to which particular combinations of factors of hereditary and environmental nature contribute to achieving the age of longevity. An organism is an open system, therefore it is precisely those external agents (molecules, substances, ions) that primarily come into contact with it deserve special attention. First of all, it is oxygen participating in the energy exchange inside the cell. Getting into the body, it enters a chain of chemical transformations. And already the metabolic products of those structures are involved in the adaptation process and regulate (modify) homeostasis [1]. The enzyme activity level is determined by the genetic variation in the structure of their genes. Thus, individual genotypic features determine the variability of the enzymatic antioxidant system, and, therefore, the plasticity of chemical and physiological reactions that determine the range of adaptive capabilities of the body.

Aim

The purpose of the study was the analysis of the polymorphic markers of some genes-candidate of aging and longevity, which relate to the body's defense system against oxidative stress, considering ethnicity, age gradation and gender differentiation.

Methods

Total group (3664 people) included individuals living in the Republic of Bashkortostan and belonging to three ethnic groups – Russians, Bashkirs and Tatars and in age from 1 to 109 years old. The biological material was DNA isolated from 8 mL of whole venous blood by standard phenol-chloroform extraction. Allelic variants of the genes were identified by RT-PCR using TaqMan probes. For statistical analysis of the results of the study, computer programs SPSS (v.13.0), GENEPOP, and Arlequin (v.3.0) were used.

Results

We found interethnic differences in the distribution of allele frequencies of superoxidedismutases 1 and 2 (Mn, Cu-SOD and Mn-SOD), catalase (CAT), NAD(P)H Quinone dehydrogenase 1 (NQO1) genes. To reach the age of longevity, genotypes *SOD1**A/A, *SOD1**A/G, *SOD2**A/A, *NQO1**C/T, *NQO1**C/C and *GPX1**L/L were significant among Russians, genotypes *SOD2**A/A, *SOD2**V/V, *SOD2**V/A, *CAT**C/T, *CAT**C/C were significant among Tatars, genotypes *MSRA**C/C, *CAT**C/C were significant among Bashkirs. Based on modern ideas about the genes of aging and longevity, antioxidant defense genes related to "frailty genes" [2]. A number of associative studies have shown the participation of antioxidant defense genes in the development of multifactorial and age-associated diseases that limit the lifespan [3-5]. However, genetically determined functioning of the antioxidant defense enzyme system can become the key to the molecular base for the formation of an individual phenotype of longevity.

Acknowledgment

The study was carried out according the state task (State Registration No. AAAA-A16-116020350032-1) in part and funded by RFBR and ASA according to the research project # 19-54-40007.

References

[1] Novikov V.E., Levchenkova O.S., Pozhilova E.V. (2014) The role of reactive oxygen species in the physiology and pathology of the cell and their pharmacological regulation. Reviews on clinical pharmacology and drug therapy. 12(4): 13-21.

[2] Ukraintseva S. et al. (2016) Puzzling role of genetic risk factors in human longevity:"risk alleles" as pro-longevity variants. Biogerontology. 17(1): 109-127.

[3] Vorobyova E.N., Vorobyev R.I. (2005) The role of free radical oxidation in the pathogenesis of diseases of the circulatory system. Siberian Scientific Medical Journal. 4: 24-30.

[4] Soerensen M. et al. (2012) Human longevity and variation in GH/IGF-1/insulin signaling, DNA damage signaling and repair and pro/antioxidant pathway genes: cross sectional and longitudinal studies. Experimental gerontology. 47(5): 379-387.

[5] Levy D., Reichert C.O., Bydlowski S.P. (2019) Paraoxonases activities and polymorphisms in elderly and old-age diseases: An overview. Antioxidants. 8(5): 118-142.

DOI 10.18699/BGRS/SB-2020-392

Lymph nodes morphology as predictor natural and premature aging

Olga Gorchakova
Research Institute of a clinical and experimental Lymphology – branch of Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Science
Novosibirsk, Russia
ORCID 0000-0001-7732-7587

Vladimir Gorchakov
Research Institute of a clinical and experimental Lymphology – branch of Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Science,
Novosibirsk State Universite
Novosibirsk, Russia
ORCID 0000-0001-8135-7842
vgorchak@yandex.ru

Georgy Demchenko
Institute of Physiology of Human and Animals of Committee of Science of the Ministry of Education and Science of the Republic of Kazakhstan
Almaty, Kazakhstan
ORCID 0000-0001-9906-2700

***Abstract* — The research purpose is to estimate the structural organization of lymph nodes of different localization at natural (physiological) and premature aging respectively at OXYS and Wistar rats. We used a morphological method of a research. The age-induced changes of lymphoid tissue of Wistar and OXYS rats differ with type of the immune response and morphological variant of lymph nodes structure. The immune response on humoral type is formed in mesenteric and inguinal lymph nodes, and the immune response on cellular type is formed in a tracheobronchial lymph node. We noted reduction of structures of cortical substance, especially lymphoid follicles and a paracortex, and expansion of medullary substance in a lymph node of OXYS rats. Lymph nodes are subject big morphological changes at rats of OXYS, than at old rats of Wistar. Observed changes of structure of lymph nodes of OXYS rats are a morphological equivalent of premature aging and confirms early decrease of a drainage and immune function of lymph nodes.**

Keywords* — *lymph node, OXYS and Wistar rats, premature aging, gerontology

Introduction

Clarification of the reasons of aging of lymphoid tissue has special relevance because of increase in number of old people in the country and need of specification of pathogenesis and search of method of prevention. Lymph nodes are key elements of different lymphatic regions, and lymph nodes define a regional immunological homeostasis [1, 2]. There is an urgent need of a morph of functional assessment of lymph nodes depending on localization taking into account an age factor. There is no uniform theory of aging [3]. Researchers combine often the mechanism of aging and pathogenesis of immune insufficiency. It is impossible to consider separately age evolution and aging of lymphoid tissue without structural reorganization of lymph nodes. Emergence of model of senilism of the line of rats of OXYS [4] is an argument for studying morphological equivalents of aging of lymph nodes. Morphology of lymph nodes at OXYS rats are not provided..

The purpose of work is a research of age changes of lymph nodes of different localization at OXYS rats with a syndrome of premature aging and Wistar rats with the normal rate of aging.

Material and methods

Work is performed on rats males of the OXYS and Wistar lines according to "Rules of work with use of experimental animals" (86/609/EEC). Rats of the Wistar and OXYS lines are received from Center of collective use "Gene pools of laboratory animals" of Institute of cytology and genetics of the SB RAS. N.G. Kolosova is a holder of the license for OXYS rats [4]. These rats have genetically caused defect, it is considered as an accelerated aging syndrome.

The experiment was executed on young and old white rats. The age of rats is 3 months and 1.5 years. We investigated by a morphological method of mesenteric, inguinal and tracheobronchial lymph nodes. Lymph nodes fixed in 10% neutral formalin. After fixing we adhered to the classical scheme of washing, dehydration, imbibition with a xylol, paraffin and preparation histologic sections on the microtome. Histologic sections of lymph nodes painted hematoxylin and eosine, azure-II-eosine, trichromatic paint on Masson.

The morphometric analysis of structures of a lymph node was carried out by means of a morphometric grid and the Image-Pro Plus 4.1 program. Statistical data processing was performed with licensed statistical software package StatPlus Pro 2009, AnalystSoff Inc. A P-value < 0.05 was considered statistically significant.

Results and Discussion

At natural (physiological) aging of involution of lymphoid tissue the stage of the maximum development of compartments of a lymph node precedes that corresponds three-month-old of young rats of Wistar [1, 5]. The indicator of a ratio of T- and B-zones makes more than unit at three months age of young rats, building a progressive row of lymph nodes: tracheobronchial – inguinal – mesenteric. Each of these lymph nodes differs with extent of development of structural and functional zones. Involution of adenoid tissue leads to structural destabilization of regional lymph nodes at old rats of Wistar.

The coefficient of a ratio of T- and B-zones almost changes in an inguinal lymph node at natural aging. The coefficient considerably decreases in a mesenteric lymph node and increases in a tracheobronchial lymph node. So, immunoactive zones have the following changes by 1.5 years at rats of Wistar.

The inguinal lymph node. There is an expansion of the area of the cortical plateau (in 1.2 times) a medullary sine and medullary cords (in 1.4-1.5 times) and reduction of the area of paracortical area (in 1.2 times), lymphoid follicles (in 1.9 times) in the inguinal lymph node.

The mesenteric lymph node. There is a reduction of the area of the cortical plateau (in 1.9 times), a paracortex (in 1.8 times), lymphoid follicles with the germinative center (in 4 times), a

medullary sine (in 1.5 times) and expansion medullary cords (in 1.8 times) in the mesenteric lymph node.

The tracheobronchial lymph node. There is a reduction by 1.2-1.3 times of the area of the cortical plateau (in 1.2 times), a paracortex (in 1.3 times), lymphoid follicles (in 4.5 times), medullary cords (in 1.2 times) with an invariable width of a medullary sine in the tracheobronchial lymph node.

It is clear, there is a redistribution of the immunoactive compartments which are responsible for the immune response in lymph nodes of different lymphatic regions with age.

Other morphological picture is observed in lymph nodes of premature aging of OXYS rats. Changes of lymphoid tissue are characterized more rapidly development without achievement of the maximum development of compartments. The greatest changes are noted T- and B-zones of lymph nodes. These zones are responsible for forming of the immune response on cellular and humoral type. The coefficient of a ratio of T- and B-zones less than unit also demonstrates prevalence in structure of lymph nodes a thymus-independent area at of OXYS rats at three-months age. Indicators of structural and functional zones of lymph nodes are various at three-months age of OXYS and Wistar rats. We noted the following changes in lymph nodes of young rats of OXYS.

The inguinal lymph node. There is an increase in the area of the cortical plateau (in 1.9 times), a medullary sine (in 1.2 times) and reduction of the area of a paracortex (in 1.4 times), lymphoid follicles (in 3.4 times), medullary cords (in 1.6 times) in an inguinal lymph node

The mesenteric lymph node. There is a reduction of the sizes of the cortical plateau (in 1.8 times), a paracortex (in 2.8 times), lymphoid follicles (in 3.3 times), and increase in a medullary sine (in 1.5 times) in a mesenteric lymph node.

The tracheobronchial lymph node. There is a reduction of the area occupied by the cortical plateau (in 1.2 times), a paracortex (in 1.3 times), lymphoid follicles (in 3.4 times) and increase in the area occupied by medullary cords (in 1.3 times) at minor change of a medullary sine in a tracheobronchial lymph node.

The coefficient of a ratio of T- and B-zones remains low size in inguinal and mesenteric lymph nodes and increases in a tracheobronchial lymph node when aging by 1.5 years. It is connected with regress of the main structural and functional zones (a paracortex, lymphoid follicles) at expansion of medullary substance in lymph nodes of different lymphatic regions. Nevertheless, the T-zone dominates in a tracheobronchial lymph node that defines the immune response of cellular type. Changes of the structural organization of lymph nodes are the morphological sign of decrease immune protection progressing in process of aging at OXYS rats.

The structure of lymph nodes of OXYS old rats does not provide a drainage and the immune status of the lymphatic region. It increases risk of developing pathology [3]. It should be considered by search of means of correction of senile changes of lymphoid tissue for increase of nonspecific resistance [5].

Conclusion

Morphodynamics of compartments of lymph nodes of different localization has distinctions at natural and premature aging. The imbalance of a ratio of T- and B-zones characterizes depression of function of lymph nodes depending on localization and age. The general is minimization of structures of cortical substance and increase medullary substance of lymph nodes. Changes of compartments sizes lead to forming of the immune response of humoral type for a mesenteric and inguinal lymph nodes and the immune response of cellular type for a tracheobronchial lymph node. At the same time at rats of OXYS the aging happens the accelerated rates and regressive changes of lymphoid tissue are available already for young animals. The age-caused imbalance of compartments of a lymph node can be considered as a predictor of premature aging and as risk of developing pathology. Features of morphology of lymph nodes give the grounds to consider of OXYS rats as universal model of immune insufficiency for studying pathogenesis and development of correction methods.

Acknowledgment

The state task to research work # 0324-2019-0045-C-02, AAAA-A19-119031590017-7.

References

[1] Ju.I. Borodin, O.V. Gorchakova, A.V. Suhovershin and V.N. Gorchakov The concept of lymphatic region in preventive lymphology. LAP LAMBERT Academic Publishing, 2018, 74 p.

[2] H. Suami Lymphosome concept: Anatomical study of the lymphatic system / H. Suami // Journal of Surgical Oncology, 2017, vol. 115, no 1, pp.1–5. DOI: 10.1002/jso.24332.

[3] E.G. Zotkin, I.S. Dydykin, Lila A.M. The inflammatory theory of aging, age-associated diseases and osteoarthritis // Russian medical journal, 2020, no. 7, pp.33-38.

[4] N.G. Kolosova, N.A. Stefanova, E.E. Korbolina, A.Zh. Fursova and O.S. Kozhevnikova Senescence-Accelerated OXYS Rats: A Genetic Model of Premature Aging and Age-Related Diseases // ADVANCES IN GERONTOLOGY, 2014, vol. 27, no. 2, pp. 336-340.

[5] O.V. Gorchakova and V.N. Gorchakov, Increase in drainage and immune functions of a lymph node as a factor of endoecological wellbeing at advanced and senile age // ADVANCES IN GERONTOLOGY, 2015, vol. 28, no. 3, pp. 521-526.

DOI 10.18699/BGRS/SB-2020-393

Is there a fecundity/longevity trade-off under heat stress?

Nataly Gruntenko
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
ORCID 0000-0003-3272-1518
nataly@bionet.nsc.ru

Evgenia K. Karpova
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
karpova@bionet.nsc.ru

Elena V. Burdina
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
bella79@list.ru

Natalya V. Adonyeva
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
nadon@bionet.nsc.ru

Petr N. Menshanov
Laboratory of Functional Neurogenomics
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
eternity@bionet.nsc.ru

Inga Yu. Rauschenbach
Department of Insect Genetics
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
iraushen@bionet.nsc.ru

Abstract — **The idea of this study was to discover remote consequences of the long-term effects of regular stress exposure. We investigated the effect of short-term heat stress (38°C, 1 h) of varying frequency on longevity, fecundity, fat content, dopamine metabolism and resistance to acute heat stress (38°C, 4 h) in *Drosophila melanogaster*. We showed that stress once a day causes a significant decrease in both longevity and fecundity, as well as in the fat content, and increases dopamine metabolism and survival under acute stress. We believe that this decrease could possibly contribute to adaptation, allowing insects to save energy. We found that weekly stress in the first two weeks does not affect the longevity, dopamine metabolism and resistance to acute stress, but causes a significant increase in the total level of fertility, despite sharp fertility drops on the exact days of stressing. Thus, two stress modes were found to have opposite effects on the reproductive function of *D. melanogaster* (negative or positive). However, the data obtained allow us to assume that there is no fecundity/longevity trade-off under heat stress. The trade-off we can see here is between reproduction and stress resistance. The flies use their energy resources either on the immediate individual survival (under acute stress) to the detriment of fertility and longevity or on the fecundity increase (under mild stress).**

Keywords *— Drosophila melanogaster, Longevity, Fecundity, Heat stress, Dopamine, Stress resistance*

Introduction

The ability to develop an adequate response to unfavourable conditions of different natures is of crucial importance for the successful adaptation of an organism to ecological challenges. It is known that stress can cause both a decrease and an increase (due to the hormetic effect) in fitness [1]. Many genetic mechanisms, both ensuring resistance to unfavourable factors and underlying the processes of reproduction and ageing are highly conservative, and therefore model objects such as *Drosophila melanogaster* (Diptera: Drosophilidae) can be used for their study.

Results and discussion

Here we studied the effect of repeated episodes of mild heat stress (38°C, 1 h) of varying frequency (once a day, once a week, twice in the first two weeks) on the longevity, fecundity, fat content, dopamine metabolism and resistance to acute heat stress (38 °C, 4 h) in *D. melanogaster*. We found that mild heat stress repeated weekly in the first two weeks did not affect the longevity, resistance to acute heat stress and dopamine metabolism, but caused a decrease in the fat content and increase in the total level of fertility, in spite of a sharp fertility drop on the exact days of stressing. Stress once a week throughout the entire period of reproduction did not affect the total level of fertility, although it resulted in a significant fertility decrease on the exact days of stressing. The mild stress repeated daily caused a significant longevity and fecundity decrease throughout the entire reproduction period. We believe that this decrease could possibly contribute to the adaptation, allowing flies to save energy. This suggestion is supported by the data on dopamine metabolism and fat content in the flies exposed to the mild stress once a day. Dopamine is known to be involved in the control of many cellular and physiological processes in insects, the neuroendocrine stress response and energy metabolism among them [2-4]. The fat body is the most important energy store of insect organism and a key organ in the metabolism of lipids and carbohydrates [5]. We found out that the mild heat stress repeated daily within two weeks resulted in an increased activity of the dopamine metabolism enzymes, dopamine-dependent arylalkylamine N-acetyltransferase and alkaline phosphatase, which, together with an increased survival rate of these flies under acute heat stress, indicates a decrease in dopamine levels [6-8]. These data agree well with the results of transcriptome analysis of the fat body, which demonstrated significant changes in expression levels of genes involved in catecholamine and carbohydrate metabolic processes following the daily mild heat stress. The revealed decrease in the fat content following the mild heat stress repeated daily also supported the idea of energy consumption of adaptation process.

Conclusions

Thus, the data obtained allow us to assume that there is no fecundity/longevity trade-off under heat stress. The weekly stress episodes during first two weeks have a hormetic effect on reproduction but do not decrease longevity, whereas the daily stress episodes result in a decrease in both fecundity and longevity. The trade-off we can see here is between long- and short-term strategies of adaptation. The flies may compromise the reproduction and stress resistance. They could use their energy resources either on immediate individual survival (under more severe stress) to the detriment of fertility and longevity or on the fecundity increase (under less severe stress).

ACKNOWLEDGMENT

This research was supported by the Russian Foundation for Basic Research (Project # 19-04-00458).

REFERENCES

[1] É. Le Bourg, "Using *Drosophila melanogaster* to study the positive effects of mild stress on aging", Experimental Gerontology, 2011, V. 46, P. 345–348.

[2] C. A. Martin, D. E. Krantz, "*Drosophila melanogaster* as a genetic model system to study neurotransmitter transporters", Neurochem Int., 2014, V. 73, P. 71–88.

[3] T. Ueno, J. Tomita, S. Kume, K. Kume, "Dopamine modulates metabolic rate and temperature sensitivity in *Drosophila melanogaster*", PLoS ONE, 2012, V. 7, e31513.

[4] M. E. Hanna, A. Bednářová, K. Rakshit, A. Chaudhuri, J. M. O'Donnell, N. Krishnan, "Perturbations in dopamine synthesis lead to discrete physiological effects and impact oxidative stress response in *Drosophila"*. J Insect Physiol., 2015, V. 73, P. 11–19.

[5] Z. Liu, X. Huang, "Lipid metabolism in *Drosophila*: development and disease", Acta Biochim Biophys Sin (Shanghai), 2013, V. 45, P. 44-50.

[6] E. V. Bogomolova, I. Yu. Rauschenbach, N. V. Adonyeva, A. A. Alekseev, N. V. Faddeeva, N. E. Gruntenko, "Dopamine down-regulates activity of alkaline phosphatase in *Drosophila*: the role of D2-Like receptors", J Insect Physiol., 2010, V. 56, P. 1155–1159.

[7] I. Yu. Rauschenbach, E. K. Karpova, N. V. Adonyeva, O. V. Andreenkova, N. V. Faddeeva, A. A. Alekseev, P. N. Menshanov, N. E Gruntenko, "Disruption of insulin signalling affects the neuroendocrine stress reaction in *Drosophila* females", JEB, 2014, V. 217, P. 3733-3741.

[8] E. V. Burdina, N. V. Adonyeva, E. K. Karpova, I. Yu. Rauschenbach, P. N. Menshanov, N. E. Gruntenko, "The effect of a mild heat stress of different frequencies on the adaptability of *Drosophila melanogaster* females", Archives of Insect Biochemistry and Physiology, 2019, V. 102, e21619.

DOI 10.18699/BGRS/SB-2020-394

Nanobodies design for treatment of age-related diseases

Mohammad Mehdi Heidari
Department of Biology, Faculty of science,Yazd University,
Yazd, Iran
heidarimm@yazd.ac.ir

Yuriy L. Orlov
Institute of Digital Medicine
I.M. Sechenov First Moscow State Medical University
Moscow, Russia
Novosibirsk State University
Novosibirsk, Russia
orlov@d-health.institute

***Abstract* — The problem of the treatment of age related diseases such as Alzheimer disease demands development of new drug design strategies. Reagents that specifically recognize oligomeric morphologies of Aβ have potential diagnostic and therapeutic value. Nanobodies (Nbs) or Single-domain antibodies are the smallest antigen-binding fragments derived from heavy-chain-only antibodies. The E1 nanobody selectively recognizes naturally occurring Aβ aggregates produced in human AD brain tissue We discuss a method for the generation and binding optimization of VHHs that involves the grafting of the complementarity determining regions (CDRs) from already existing, non-camelid antibodies to VHH frameworks, followed by affinity maturation and target binding improvement using *in silico* site-directed mutagenesis.**

Keywords* — *bioinformatics, nanobodies, protein design, Alzheimer disease, age-related diseases

Motivation and aim

Motivation

The prevalence of dementia is increasing in the aging populations at an alarming rate last years. The problem of the treatment of age related diseases including Alzheimer disease demands development of new drug design strategies. Reagents that specifically recognize oligomeric morphologies of Aβ have potential diagnostic and therapeutic value [1].

Aim

Idea is to design *in silico* nanobodies (Single-domain antibodies) for diagnostics and potential therapy. Nanobodies are special derivatives of antibodies, which consist of only a single chain. The E1 nanobody selectively recognizes naturally occurring Aβ aggregates produced in human AD brain tissue indicating that a variety of morphologically distinct Aβ aggregate forms occur naturally.

Methods

Nanobodies (Nbs) or Single-domain antibodies are the smallest (12–15 kDa) antigen-binding fragments derived from heavy-chain-only antibodies (VHH) [2]. Nanobodies are characterized by high thermal stability and solubility, recognizing uncommon or hidden epitopes of protein targets, unique refolding properties. They exhibit affinities comparable to conventional antibodies and superior tissue penetration.

Nanobodies can also recognize epitopes that remain undetected by conventional antibodies. In addition, nanobodies are shown to be non-immunogenic in humans and are easy to manufacture. Because of these favorable characteristics, nanobodies are suitable candidates for the development of imaging probes, therapeutic agents with neutralizing or receptor-ligand antagonizing functions, and for targeted drug therapy.

Results and Discussion

We discuss a method for the generation and binding optimization of VHHs that involves the grafting of the complementarity determining regions (CDRs) from already existing, non-camelid antibodies to VHH frameworks, followed by affinity maturation and target binding improvement using *in silico* site-directed mutagenesis.

To date, many antibodies and fragments have been generated against various antigens. This existing assortment represents an ideal CDR donor repertoire for our approach. CDR grafting is a powerful technique for transferring binding specificities to other antibody frameworks with desired properties. This is typically done to stabilize or humanize antibodies intended for medical use. To start the mutation process, the CDR, non-CDR loops and the locations on the CDR loops which allow insertions or deletions must be identified. This can be done by aligning nanobody sequences to known structures belonging to the same species. There are publicly available CDR numbering tools that can be applied to a certain sequence. After the identification of the orientation or the sequence of the nanobody, amino acids can be selected for mutation to increase the binding affinity for the antigen.

The rational design of nanobodies based on CDRs derived from conventional, pre-existing antibodies would be of high value to avoid animal immunizations and directly generate binders in the desired framework.

ACKNOWLEDGMENT

The research was supported by the RSF grant (19-15-00219) and RFBR grant application.

REFERENCES

[1] Kasturirangan S., Li L., Emadi S., Boddapati S., Schulz P., Sierks M.R. (2012) Nanobody specific for oligomeric ?-amyloid stabilizes nontoxic form. Neurobiol Aging.;33(7):1320-8.

[2] Messer A., Butler D.C. (2020) Optimizing intracellular antibodies (intrabodies/nanobodies) to treat neurodegenerative disorders. Neurobiol Dis. 134:104619.

DOI 10.18699/BGRS/SB-2020-395

Serum Polypeptide Alpha-Fetoprotein (AFP) as a possible powerful geroprotector

Alexander Khalyavkin
Institute of Biochemical Physics of RAS
Federal Research Center"Computer Science and Control" of RAS
Moscow, Russia
antisenesc@mail.ru

Vyacheslav Krut'ko
Federal Research Center "Computer Science and Control" of RAS
Sechenov First Moscow State Medical University
Moscow, Russia
krutkovn@mail.ru

Vitaly Dontsov
Federal Research Center "Computer Science and Control" of RAS
Moscow, Russia
dontsovvi@mail.ru

***Abstract* — AFP is primarily known as an oncomarker for tumors of the liver and some other organs. But then it was identified immunochemically as a normal fetal serum antigen. Now it is one of the best-known embryo specific proteins with pronounced immunotropic and detoxifying effects and has a very small background serum level in adults. Liver reparative regeneration is accompanied by temporal and malignant growth by constitutive AFP synthesis in adult liver. It is also known that rat sexual activity increased 30 min after administration of 1 μg/kg AFP. AFP increased cerebral flow and had pronounced antihypoxic activity. For this reason we studied bioactivating effects of AFP in old mice (females Balb/c aged 18 months) and their survivability. During the two weeks of the experiment, the natural death of animals in the control group was 47%, while AFP administration reduced the mortality up to 17% in the experimental group. The physical activity of the animals, their habitus and immune status were also higher than in the control group.**

Keywords* — *alpha-fetoprotein, geroprotector, survivability, reducing mortality, anti-aging effects

Introduction

AFP is primarily known as an oncomarker for tumors of the liver and some other organs. But then it was identified immunochemically as a normal fetal serum antigen. Now it is one of the best-known embryo specific proteins with pronounced immunotropic and detoxifying effects and has a very small background serum level in adults. Liver reparative regeneration is accompanied by temporal and malignant growth by constitutive AFP synthesis in adult liver.

It is also known that rat sexual activity increased 30 min after administration of 1 μg/kg AFP. AFP increased cerebral flow and had pronounced antihypoxic activity. For this reason we studied bioactivating effects of AFP in old mice, keeping in mind their immunoregulatory features, too.

Materials and Methods

AFP (produced by "Institute of New Medical Technologies", Perm, Russia) was administered to 12 experimental animals (females Balb/c aged 18 months, obtained from kennel Stolbovaya) for 2 weeks, in the morning, intraperitoneally, in 0.5 ml saline, 10 μg per kg of body weight. In 15 control animals, the same saline volume was administered without the drug. In the course of the experiments, we evaluated some parameters of aging as well as survival of animals in the control and experimental groups.

We evaluated the appearance of animals based on the state of their hair – its color and gloss, baldness, and the severity of age-hump. The physical state of the animals was evaluated by the time during which the mice could stay on a string; we took into account the weight of the animals and the relative mass of internal organs (mg of organ weight per gram of animal body weight). After the end of the experiments, the remaining animals were sacrificed and a number of parameters important for assessing the level of aging of the animals were measured. To assess immune status, the following parameters were taken into account: relative weight of the immune organs (spleen and thymus) and degree of age related autoaggression based on the level of autoimmune complexes after the addition of polyethylene glycol (PEG6000) to blood serum expressed in units of optical density under spectrophotometric study. The content of intracellular water in tissues was evaluated based on the extent of weight loss (for kidneys) in hypertonic solution (40% sucrose) after 4 h of incubation.

We traced animal survivability during the studied period (percentage of the original number) and evaluated the mean values of the test parameters and standard deviations for control and experimental values; differences were compared by Student's test, which characterizes the normal distribution of data, and the Mann–Whitney U criterion, which characterizes nonparametric distribution. In the experiments, we used a group of mice at late stages of aging, which was reflected in the mortality level in the population: half of the mice in the vivarium had died during the previous 2 months.

Results and Conclusions

During the experiment, the natural death of animals in the control group was 47%, while AFP administration reduced the mortality to 17% in the experimental group. We studied the general state of the surviving animals (8 in the control and 10 in the experiment), then the animals were sacrificed, and a number of parameters were examined.

A clear trend towards improvement of the appearance was observed in all the animals treated with AFP: hair loss and bald areas were reduced, hair glistened, and it lost yellowness and patchiness characteristic of old age. Habitus evaluated according to these characteristics was clearly better for all the animals of the experimental group. The physical activity of the animals and their immune status were also higher than in the control group.

Thus, AFP administration drastically increased the survival rate of the oldest mice and had a pronounced bioactivating effect in our experiments.

DOI 10.18699/BGRS/SB-2020-396

Gerontology and scientometrics ("Citogerontology")

Alexander N. Khokhlov
Evolutionary Cytogerontology Sector, School of Biology, Lomonosov Moscow State University, Moscow, Russian Federation
khokhlov@mail.bio.msu.ru

***Abstract* — Based on his experience in working as a member of Editorial Boards of some scientific journals and Elsevier/Scopus Expert Content Selection and Advisory Committee (Russian Federation), as well as Associate Editor-in-Chief of Moscow University Biological Sciences Bulletin, the author analyzes current situation with publications in the field of experimental and theoretical gerontology. Special attention is paid to scientometric indices and ranking of the papers, relevant journals, and authors. Some predatory journals in the field are considered too. Various approaches to improving scientometric indices of authors involved in gerontological research are proposed.**

Keywords* — *gerontology, aging, publications, scientific journals, indices, metrics, ranking

In recent years, scientometric indicators (number of citations, impact factor, Hirsch index, SJR index, etc.) of their work efficiency have become increasingly important for scientists of various specialties. Unfortunately, it is precisely such indicators, and not the essence of the research conducted by the authors, that are now often decisive for reviewers from various foundations providing research grants. And gerontology/biology of aging in this regard are no exception. In order to get a big enough grant to study the mechanisms of aging or to develop model systems for the search for geroprotectors, it is now not enough for specialists in the field of biology of aging to submit an application to the science foundation with a description of their advanced ideas or developed research methods. In the first place, reviewers of applications pay attention to the "quality" of works already published by applicants, as well as to the relevant ratings of the authors. And by "quality" and "ratings" they mean the scientometric indicators just mentioned above (of both researchers and the journals they used to publish their results). There is a vicious circle: in order to get high-quality scientific results, you need money, and you can get it only by publishing a significant number of "cool" articles [1].

The paradox is that in many cases the journals from Q1 (25% of journals with the highest scientometric indicators; the publications in them provide scientists with the high ratings necessary for receiving grants, as well as for re-certification or competition for prestigious posts) are not necessarily really serious about reviewing and editing manuscripts submitted to the editorial office. Moreover, they may even be the so-called "predatory journals," the appearance of which is associated with an increasingly widespread recent distribution of articles, publication of which is fully paid by the authors [2, 3].

A few years ago, I was already considering the situation with one small foreign publishing house specializing in gerontological journals mainly [1]. It exists for a little over 10 years and currently publishes 4 scientific journals. The impact factors of these publications reach 5–6, i.e. they can easily be classified as "highly ranked" (Q1). All of them operate within the framework of an open access model paid by the authors. The cost of publishing an article (APC – article processing charge) varies from 3 to 4 thousand US dollars. Moreover, the number of published articles is huge. For example, in the most popular journal of the publisher in 2016 about 1,500(!) articles were published, distributed among 52(!) issues. In addition, publication time is from two weeks(!), which practically excludes the possibility of a normal reviewing of manuscripts. It should be said that this journal was mentioned by the famous fighter for the purity of scientific publications, Jeffrey Beall [2, 3], on his website as a very likely "predator," but this did not prevent the journal from being indexed in international systems of global citation.

It should be borne in mind that journal ranking can be done in quite different ways. In particular, impact factor (Web of Science) and CiteScore (SCOPUS) indices are very similar; both are based on the number of citations in a given year of articles from a given journal published in the previous two (Web of Science) or three (SCOPUS) years. Therefore, the Q1-Q4 quartile distribution in these systems is quite similar. However, in the very popular SCImago Journal Rank system, based on SCOPUS data, journals are ranked according to the SJR index; this is the same citation index of the journal, but already normalized to the rating of serials in which articles are cited. If the rating of citing editions is very high, then SJR may exceed CiteScore, if it is low enough, then SJR will be lower than CiteScore – in some cases at times lower. In many scientific institutions, it is the rating based on the SJR indicator that is used in the certification of teachers and researchers.

It should be emphasized that in each area and category of knowledge there are ranks and quartiles that may not be connected in any way with these indicators in other fields of science. Say, in the subject category "Gerontology," even SJR 0.9 allows the journal to get into Q1, but in the subject category "Aging," SJR 1.7 is already needed for this. The highest ranked gerontological journal *Ageing Research Reviews* has SJR of 4.125, while its CiteScore has now exceeded 11. However, it is very far from the "coolest" oncological journal *Ca-A Cancer Journal for Clinicians* with SJR 72.576 and CiteScore around 180.

In conclusion, I would like to note that my many years of experience as a member of the editorial board of the journal *Advances in Gerontology*, as well as a member of the Elsevier/Scopus Expert Content Selection and Advisory Committee (Russian Federation) and Associate Editor-in-Chief of *Moscow University Biological Sciences Bulletin* [4], which publishes, by the way, a significant number of articles on the biology of aging, has allowed me to assume that a highly ranked scientific journal is difficult to quickly create [5], if you do not ignore the formal requirements of international systems of global citation, implying a rigorous blind peer review, correct preparation of the manuscript, illustrations and bibliographies, as well as a serious attitude of the authors to the statistical processing of the data obtained. However, in the presence of significant financial injections, this process can be accelerated many times, but, as far as I know, the vast majority of gerontological journals that are free of charge for authors, sadly, cannot boast of this kind of financial support.

ACKNOWLEDGMENT

Supported in terms of the state assignment of Lomonosov Moscow State University, part 2 (basic research, no. AAAA-A16-116021660098-8).

REFERENCES

[1] Khokhlov A.N., Klebanov A.A., Morgunova G.V. (2017) How very bad articles are published in very good scientific journals. In: World-Class Scientific Publication – 2017: Best Practices in Preparation and Promotion of Publications: Proc. 6th Int. Sci. & Pract. Conf., April 18–21, 2017. Ekaterinburg: Ural University Press, 150–156.

[2] Beall J. (2012) Predatory publishers are corrupting open access. Nature. 489(7415): 179.

[3] Beall J. (2013) Predatory publishing is just one of the consequences of gold open access. Learn. Publ. 26(2): 79–84.

[4] Kirpichnikov M.P., Morgunova G.V., Khokhlov A.N. (2020) Our journal – 2020: what and how we publish. Moscow Univ. Biol. Sci. Bull. 75(1): 1-6.

[5] Khokhlov A.N. (2019) The worse, the better or how to quickly create a high-ranking scientific journal. In: World-Class Scientific Publication – 2019: Strategy and Tactics of Management and Development: Proc. 8th Int. Sci. & Pract. Conf., Moscow, April 23–26, 2019. Ekaterinburg: Ural University Press, 118–126.

DOI 10.18699/BGRS/SB-2020-397

Cellular senescence in age-related macular degeneration: impact of changes in autophagy and neurotrophic supplementation

Oyuna S. Kozhevnikova
Laboratory of molecular mechanisms of aging, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
oidopova@bionet.nsc.ru

Darya V. Telegina
Laboratory of molecular mechanisms of aging, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
telegina@bionet.nsc.ru

Mikhail A Tyumentsev
Laboratory of molecular mechanisms of aging, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
landselur@bionet.nsc.ru

Nataliya G Kolosova
Laboratory of molecular mechanisms of aging, Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
kolosova@bionet.nsc.ru

The senescence-accelerated OXYS rats spontaneously reproduce the major signs of age-related macular degeneration (AMD): dystrophic alterations of the RPE, thinning of the neuroretina, and impairment of choroidal microcirculation. We found increased NGF staining in Muller cells in OXYS rats with progressive stage of retinopathy and differences in the cell type–specific localization of mBDNF between OXYS and Wistar rats. During the development of AMD-like retinopathy, proBDNF dominated over mBDNF. It was suggested that neurons of OXYS rats have a limited ability to process proBDNF, thereby leading to increased cell loss in retina. We showed that the development of AMD-like retinopathy in OXYS rats is accompanied by retinal transcriptome changes affecting genes involved in autophagy. The impaired reactivity of autophagy was confirmed by a decreased number of autophagosomes under the conditions of blocked autophagosome-lysosomal fusion according to immunohistochemical analysis and transmission electron microscopy. Using OXYS rats as a model of AMD, we demonstrated that decreased capacity for upregulation of autophagic flux in response to metabolic stress accompanies the development of AMD and may reflect an age-related decline in the adaptability of retinal cells. Our results indicate disturbances in the neurotrophic support in the retina of OXYS rats. Maintaining sufficient reactivity of autophagy in the retina and the balance of neurotrophic factors may be considered as a strategy to slow down AMD.

Keywords — autophagy, neurotrophins, aging, retina, AMD, OXYS rats

Introduction

Aging is the major risk factor for age-related diseases, including age-related macular degeneration (AMD). AMD is a progressive retinal disorder causing the severe vision impairment in the elderly but the pathophysiology of this disease is still largely unknown. There is increasing evidence that defective proteostasis due to impaired clearance might be the key process in AMD. Autophagy is a conserved cellular degradation pathway for the breakdown of cytoplasmic components: damaged proteins and organelles. Besides a housekeeping function, autophagy is crucial for response to the stress. Defects in the autophagy are linked to aging and disease pathology. Research on the molecular mechanisms underlying the age-related dysregulation of autophagy at the early stage of AMD and before its development can give clues to the most relevant molecular events triggering the entry into the irreversible stage [1].

Molecular studies on AMD are hampered by the inaccessibility of live retinal tissue from AMD patients, especially in the early stages. There is evidence that a suitable experimental model of AMD is senescence-accelerated OXYS rats, which spontaneously develop a phenotype similar to human age-related disorders including AMD-like retinopathy. Retinopathy that develops in OXYS rats already at a young age corresponds (in terms of clinical manifestations and morphological characteristics) to the dry atrophic form of AMD in humans. Nonetheless, neovascularization develops in some (~ 10–20%) of these rats with age. The clinical signs of AMD-like retinopathy appear by the age of 3 months against the background of a reduction in the transverse area of the RPE, impairment of choroidal microcirculation, and retinal thinning. The progression of these abnormalities in OXYS rats is accompanied by a significant reduction in thickness of the photoreceptor cell layer and a decrease in the number of photoreceptor cell nuclei of the outer nuclear layer. Significant pathological changes in the RPE manifest themselves as excessive accumulation of lipofuscin and amyloid in the RPE regions, disturbances in the morphology of the RPE sheet, including an increase in the proportion of multinucleated cells, hypertrophy, distortion of cell shape, and reactive gliosis.

Here, we analyzed high-throughput RNA sequencing (RNA-Seq) data to identify the specific molecular processes and pathways that take part in the alterations of autophagy during the development and progression of AMD-like retinopathy in OXYS rats. On the basis of previous results revealing an important role of autophagy, we explored the *in vivo* effects of autophagy activation and inhibition by fasting and chloroquine (CQ) treatment, respectively, on the expression of autophagy markers in the retina of OXYS and control Wistar rats. Also we assess the age-related changes in expression of mNGF, mBDNF, proBDNF and their receptors in the retina of OXYS and Wistar rats.

Methods

OXYS and Wistar rats at the age of 4 and 16 months were randomly distributed into treatment and control groups (n = 6 of each genotype per group). The rats consumed feed ad libitum, fasted for 12, 24, or 48 h, or fasted during four daily intraperitoneal injections of CQ (CQ, 50 mg/kg, Sigma-Aldrich, St. Louis, MO, USA) with the last injection administered 3 h before euthanasia [1].

To identify the pathways and biological functions involved in the alteration of autophagy in OXYS rats, we analyzed RNA-Seq data obtained previously. To identify the GO terms and pathways associated with autophagy over-represented in a DEG list, the detected DEGs were subjected to functional enrichment analyses by means of the DAVID tool with a Benjamini p-value cutoff at <0.05. The gene interaction networks related to autophagy were identified on the GeneMANIA web server (http://www.genemania.org/) with default parameters.

Standard techniques of qPCR, immunohistochemistry and electron microscopic examination were used to analyze autophagy modulation effects on retina. Neurotrophins immunofluorescent staining was performed by a standard method [2].

Results

Here we showed that the development of AMD-like retinopathy in OXYS rats is accompanied by retinal transcriptome changes affecting genes involved in autophagy. These genes are associated with kinase activity, immune processes, and FoxO, mTOR, PI3K-AKT, MAPK, AMPK, and neurotrophin pathways at preclinical and manifestation stages, as well as vesicle transport and processes in lysosomes at the progression stage. Dysregulation of the autophagy transcriptome network in retinal cells takes place already at early stages of AMD-like retinopathy. We demonstrated a reduced response to autophagy modulation (inhibition or induction) in the retina of old OXYS rats: expression of genes *Atg5, Atg7, Becn1, Nbr1, Map1lc3b, p62,* and *Gabarapl1* differed between OXYS and Wistar (control) rats. The impaired reactivity of autophagy was confirmed by a decreased number of autophagosomes under the conditions of blocked autophagosome-lysosomal fusion according to immunohistochemical analysis and transmission electron microscopy.

We found increased NGF staining in Muller cells in OXYS rats with progressive stage of retinopathy. In contrast, we observed only subtle changes in the labeling of mature BDNF (mBDNF) and TrkB during the development of AMD-like retinopathy in OXYS rats. Using colocalization with vimentin and NeuN, we detected a difference in the cell type–specific localization of mBDNF between OXYS and Wistar rats. We showed that the mBDNF protein was located in Muller cells in OXYS rats, whereas in the Wistar retina, mBDNF immunoreactivity was detected in Muller cells and ganglion cells. During the development of AMD-like retinopathy, proBDNF dominated over mBDNF. It was suggested that neurons of OXYS rats have a limited ability to process proBDNF, thereby leading to increased cell loss in retina..

Currently there is no consensus on whether autophagic activity increases or decreases with age and disease. Apparently, this phenomenon depends on such factors as cell type, disease, or a specific stage of the disease. We have reported that the basal level of autophagy is elevated at the early stage of retinopathy and declines at progressive stages. We demonstrated that retinal autophagy flux is lower in OXYS rats, thereby reducing the capacity of retinal cells to cope with the elevated proteolytic stress in progressive stages of disease. As one of the reasons for the decreased reactivity of autophagy, we propose the accumulation of senescent RPE cells in the retina of OXYS rats. These cells lead to cellular dysfunction and promote the senescence of neighboring cells by secreting the senescence-associated secretory phenotype. The converse is also possible: autophagy deregulation contributes to senescent cells accumulation. We have observed altered RPE cell morphology in OXYS rats: enlargement, flattening, a loss of the hexagonal shape, accumulation of lipofuscin granules, and multinucleation, which are employed as morphological markers of senescent cells [3].

Conclusion

Thus, using OXYS rats as a model of AMD, we demonstrated that decreased capacity for upregulation of autophagic flux in response to metabolic stress in the retina accompanies the development of AMD and may reflect an age-related decline in the adaptability of retinal cells. Alterations of neurotrophin signaling pathway were found at the advanced stage of AMD-like retinopathy development in OXYS rats. Consequently, maintaining sufficient reactivity of autophagy in the retina and the balance of neurotrophic factors may be considered as a strategy to slow down AMD.

ACKNOWLEDGMENT

This study was supported by the Russian Science Foundation (18-75-00031) and by the Russian Foundation for Basic Research (18-015-00336).

REFERENCES

[1] Kozhevnikova, O.S.; Telegina, D.V.; Tyumentsev, M.A.; Kolosova, N.G. Disruptions of Autophagy in the Rat Retina with Age During the Development of Age-Related-Macular-Degeneration-like Retinopathy. Int. J. Mol. Sci. 2019, 20, 4804. https://doi.org/10.3390/ijms20194804

[2] Telegina, D.V., Kolosova, N.G. & Kozhevnikova, O.S. Immunohistochemical localization of NGF, BDNF, and their receptors in a normal and AMD-like rat retina. BMC Med Genomics 12, 48 (2019). https://doi.org/10.1186/s12920-019-0493-8

[3] Telegina, D., Kozhevnikova, O., Bayborodin, S. et al. Contributions of age-related alterations of the retinal pigment epithelium and of glia to the AMD-like pathology in OXYS rats. Sci Rep 7, 41533 (2017). https://doi.org/10.1038/srep41533

DOI 10.18699/BGRS/SB-2020-398

Evolution of proteins involved in response to ROS

Vassily Lyubetsky
IITP RAS
Moscow, Russia
lyubetsk@iitp.ru

Gregory Shilovsky
IITP RAS; MSU
Moscow, Russia
gregory_sh@list.ru

Alexandr Seliverstov
IITP RAS
Moscow, Russia
slvstv@iitp.ru

Oleg Zverkov
IITP RAS
Moscow, Russia
zverkov@iitp.ru

Lev Rubanov
IITP RAS
Moscow, Russia
rubanov@iitp.ru

***Abstract* — Original software was used to specify the evolution of transcription factors Nrf2 and Bach1 in Deuterostomia. The transcription factors are antagonistically involved in the response to reactive oxygen species (ROS). The original algorithm shows that *Bach* emerged by duplication of *Nfe2*, an ortholog of *Nfe2l2*, in the chordate ancestor. At the N-terminus, the copy was provided with the BTB domain from a gene orthologous to the European lancelet *Branchiostoma lanceolatum* gene *BL03038_cuf1* followed by a domain typical for the Zinc finger C2H2 superfamily.**

Keywords* — *Nrf2, anti-ageing program, anti-ROS; Bach1, ageing program, pro-ROS; ROS

Introduction

In many vertebrates, Nrf2 (Nfe2l2), Bach proteins (Bach1 and Bach2), as well as Keap1, β-TrCP, c-Myc, and GSK3b are components of the regulatory network (e.g., in *Mus musculus*) that is, among other functions, associated with regulation of the level of reactive oxygen species (ROS). In particular, the relationship between Nrf2 and Bach1 is critical for the regulation of heme oxygenase 1 expression, which provides for heme degradation. The functioning of this network is related to *species-specific lifespan* as well as to *many human diseases*. Nrf2 activates about 200 genes; Nrf2 and Bach1 are antagonists in controlling ROS levels, which is also due to the competition for ARE DNA-binding sites. These proteins usually function by complexing with each other as well as with other proteins (e.g., Maf) and DNA. The BTB (POZ) domain is common in zinc-finger transcription factors. Bach proteins are the only BTB proteins with the bZIP domain. Bach has the BTB domain at the N-terminus and the bZIP domain at the C-terminus. Nrf2 also has the bZIP domain at the C-terminus. Phosphorylation of Tyr-486 in the mouse Bach1 inactivates the protein. The heme-binding sites have been identified for Bach as well as the sites of Bach homodimerization. In Bach, the heme-binding regions are not similar, and, in particular, differ by the number of regulatory cysteine-proline (CP) motifs. The following evolutionary assumptions have been proposed: the bZIP family descends from a single eukaryotic gene; the common ancestor of the *Bach* genes existed in chordates before the divergence of vertebrates since it exists in *Ciona* spp.; two *Bach* genes emerged in gnathostomes, apparently, after their separation from lampreys. Invertebrates have a single ortholog to four vertebrates genes (*Nfe2*, *Nfe2l1*, *Nfe2l2*, and *Nfe2l3*). Our work contains results about the evolution of *Nrf2* and *Bach* genes and corresponding proteins. Here we present the results of the computer-aided search for *Nrf2* and *Bach* in deuterostomes using the conditions specified below.

Results

1) Here we consider all deuterostomes represented in Ensembl and GenBank. The *Bach* gene has been identified in tetrapods, cartilaginous and bony fishes, cyclostomes, and ascidians using a local alignment with the human BACH1 (NP_001177.1), the presence of the BTB and bZIP-Maf domains, satisfactory alignment of the secondary structure, and high similarity of the 3D structure at the N- and C-terminal regions with the human BACH1. According to *these criteria*, the *Bach* gene is missing in lancelets, hemichordates, and echinoderms. Similarly, the *Nrf2* gene has been identified in deuterostomes using local alignment with the human NFE2L2 (NP_006155.2), the presence of the bZIP-Maf domain, a satisfactory alignment of the secondary structure, the absence of some other domains (kelch-type beta-propeller for kelch-like proteins; BTB for Bach; bZIP-Jun for Jun and similarly for other bZIP proteins), and high 3D structure similarity with NFE2L2 at the C-terminus coupled with a significant difference from other human bZIP proteins at the N-terminus. *Another* important reason for dividing all found proteins into the Bach and Nrf2 groups as well as other groups is the splitting of the unrooted tree into the clades of Bach and Nrf2 separated with a nearly 100% support. Rooting the tree with echinoderms makes it incongruent with the species tree only for Nrf2 in *Oikopleura dioica* and *Ciona* spp., and these proteins are questionable in other respects. Similarly, we recognize the Bach and Nrf2 clades in the bigger tree of bZIP proteins.

Analysis of the search results. Among invertebrate deuterostomes, Bach proteins have been found only in *Ciona intestinalis* and *C. savignyi* (tunicates diverged early from other chordates). At least one Bach protein has been found in all vertebrates. It looks like *Bach* emerged in the common ancestor of vertebrates and tunicates (and was possibly lost in appendicularians), after which it diverged into *Bach1* and *Bach2* in the common ancestor of cartilaginous and bony fishes. In fishes, the *Bach* genes were many times duplicated and lost; their genomes commonly have many paralogs that are orthologous to *Bach*. E.g., the huchen *Hucho hucho* has four genes orthologous to *Bach2*. The *Bach1* gene is represented by a single copy in all tetrapods. The alignment of Bach proteins indicates considerable conservation of each of them in most mammals. Within Euarchontoglires, Bach1 demonstrates only minor changes in most rodents including the Damaraland mole-rat (DMR) *Fukomys damarensis*. The only exception is the C-terminal region of all Bach1 isoforms in the naked mole-rat (NMR) *Heterocephalus glaber*. One Bach1 isoform has an extended deletion, although a short region upstream of the C-terminus is conserved. Another isoform has a long insertion in the same region. The full-length Bach1 protein of the NMR usually shares the dispensable amino acids with the DMR but not the mouse, which agrees with their taxonomical position. No significant differences in Bach1 have been revealed in primates. Beyond Euarchontoglires, the Bach1 sequences of the sloth *Choloepus hoffmanni*, tenrec *Echinops telfairi*, shrew *Sorex araneus*, dolphin *Tursiops truncatus*, flying fox *Pteropus vampyrus* are similar to those in primates and most rodents. Conversely, other

representatives of Laurasiatheria as well as the nine-banded armadillo *Dasypus novemcinctus* have insertions or deletions in Bach1 in the corresponding region of the NMR. The Bach1 of the hedgehog *Erinaceus europaeus*, elephant *Loxodonta africana*, and hyrax *Procavia capensis* in Bach1 has a very long C-terminal deletion covering both conserved and variable regions. On the other hand, one of the Bach2 isoforms in primates *Macaca mulatta*, *M. nemestrina*, and *Pan troglodytes* has an N-terminal extension not observed in humans.

2) Here we consider all species represented in RefSeq and Ensembl supplemented by those in GenBank. The Bach1 protein of cartilaginous fishes lacks the Tyr-486. The Bach1 in tetrapods has conserved the functionally significant phosphorylated tyrosine and the neighboring amino acids remained largely unchanged, although D→S substitutions and proline loss in the CP motifs are observed in long-lived rodent DMR and NMR. In the Australian ghostshark *Callorhinchus milii*, cloudy catshark *Scyliorhinus torazame*, whale shark *Rhincodon typus*, and brownbanded bamboo shark *Chiloscyllium punctatum*, this tyrosine is replaced with phenylalanine. The Bach in *Ciona* spp. has preserved this tyrosine unlike some of the neighboring amino acids. The predicted heme-binding sites in Bach1 of tetrapods insignificantly differ from those in mouse except the species described below. These sites include the 223-LCPKYR-228 (C→G in marsupials, while in the platypus *Ornithorhynchus anatinus* it is the same as in the mouse; or C→Y in the common wall lizard *Podarcis muralis*); 300-QCPAEQ-305, which considerably changed or disappeared in most mammals; 435-ECPWLG-340 (conserved in all tetrapods); 463-NCPFIS-468 (the cysteine is conserved in tetrapods, and I→M in placentals including DMR, NMR and the common degu *Octodon degus*); 494-PCPYAC-499 (conserved in all tetrapods except the bearded dragon *Pogona vitticeps* and platypus); and 648-DCPLSF-653 (conserved in almost all tetrapods). Bach in *Ciona* spp. has only two conserved CP sites, and their positions differ from those involved in the heme-dependent regulation in human and mouse. A similar pattern is observed for Bach2. The 368-ACPFNK-373 heme-binding site is present in most Myomorpha but not in the Upper Galilee Mountains blind mole rat (GMR) *Nannospalax galili*; the ACPFDK (N→D) site are found in other rodents (the North American beaver *Castor canadensis*, guinea pig *Cavia porcellus*, the GMR, and other mole-rats, DMR and NMR) and nearly all tetrapods; the ACPSDK (F→S) site is found in the Philippine tarsier *Carlito syrichta*; ACSFDK (P→S), in flying foxes; VCPFDK (A→V), in the barbed agama; ACPLDK (F→L), in marsupials and the three-toed box turtle *Terrapene carolina triunguis* (with a negligible aging rate); ACPFEK, in birds, saltwater crocodile *Crocodylus porosus*, and Chinese alligator *Alligator sinensis*; ACPVEK, in the western clawed frog *Xenopus tropicalis*; ACPLNR, in the West Indian Ocean coelacanth *Latimeria chalumnae*; no this site was found in the platypus. Other fishes have the P→S substitution and some other modifications of the site. Overall, Bach2 in flying foxes differ from that (commonly ACPFDK) in other chiropterans. The 498-SCPVPI-503 site is strictly conserved in all species starting from cartilaginous fishes to mammals. The 505-VCPRSP-510 site is also strictly conserved within the same taxonomic range excluding the *L. chalumnae* with the C→Y substitution. The 602-SCPVQD-607 site is conserved in the *L. chalumnae* and all tetrapods. This site is missing in the European cattle *Bos taurus*, while *Bos mutus* shares it with all other tetrapods. This site is missing in the bluespotted mudskipper *Boleophthalmus pectinirostris* as well as in other fishes. The sites notably differ in the *C. milii*, and whale shark *Rhincodon typus* from those in tetrapods. The 728-YCPVLI-733 site is only found in all rodents but not in other taxa. However, most vertebrates including the European rabbit *Oryctolagus cuniculus* and *C. milii* have a different site, YCPVLR (I→R). The Chinese softshell turtle *Pelodiscus sinensis* demonstrate singular modifications in the YFPVLR (C→F) site; the *X. tropicalis*, in YCPVLQ (R→Q); the *R. typus*, in FCPVFR (Y→F, L→F). The evolution of the insertion in the N-hook downstream of the MSLSE motif at the N-terminus in Bach1 has been studied. The evolution of response to mitochondrial ROS will be discussed [1].

Discussion

The genomic rearrangement that gave rise to the *Bach* remains an open problem. The Nfe2 is most similar to the Bach in early-diverging deuterostomes. One can think that *Bach* emerged by duplication of *Nfe2*, an ortholog of *Nfe2l2*, in the chordate ancestor. The BTB domain essential in Bach is missing in Nfe2; however, it exists in dozens of ancient proteins (e.g., the BTB-ZF family) that could provide it for the ancestral Bach. Possible BTB sources include a gene orthologous to the European lancelet *Branchiostoma lanceolatum* gene *BL03038_cuf1* with the domain at the N-terminus followed by a domain typical for the Zinc finger C2H2 superfamily (IPR036236). The bZIP domain specific for the C-terminal regions of Bach proteins is not found in BL03038_cuf1. Thus, *Bach* could emerge as a chimeric gene. Anyhow, such an evolutionary scenario is optimal for the reconstruction of the genomic structures carried out by the original algorithm. The absence of Bach proteins in *Branchiostoma* spp. agrees with the proposed closer phylogenetic similarity between vertebrates and tunicates rather than between vertebrates and lancelets. The absence of Bach in *Oikopleura dioica*, which is relatively close to *Ciona* spp., can be attributed to its neoteny. Indeed, like other appendicularians, adult *O. dioica* has a discrete body and tail and preserves the notochord throughout its life, while the body structure of *Ciona* spp. substantially changes in development. The loss or substantial change of *Bach2* genes in the tuatara *Sphenodon punctatus* and the tortoise *Chelonoidis abingdonii* is accompanied by a long species-specific lifespan. The changes in the NMR Bach1 untypical for rodents can also be related to the unusually long lifespan.

Acknowledgment

We are very grateful to V. P. Skulachev for the original idea, as well as help, advice, and comments on this work. The study was funded by RFBR according to project no. 18-29-13037.

References

[1] M. Y. Vyssokikh, S. Holtze, O. A. Averina, K. G. Lyamzaev, A. A. Panteleeva, M. V. Marey et al., "Mild depolarization of the inner mitochondrial membrane is a crucial component of an anti-aging program," Proc. Natl. Acad. Sci. U.S.A., Mar 2020, 201916414.

DOI 10.18699/BGRS/SB-2020-399

Calorie restriction in gerontological experiments on cell cultures

Galina V. Morgunova
Evolutionary Cytogerontology Sector, School of Biology, Lomonosov Moscow State University, Moscow, Russian Federation
morgunova@mail.bio.msu.ru

Alexander N. Khokhlov
Evolutionary Cytogerontology Sector, School of Biology, Lomonosov Moscow State University, Moscow, Russian Federation
khokhlov@mail.bio.msu.ru

***Abstract* — Lifespan can be increased by the calorie restriction. However, it is not entirely clear whether this effect is manifested at the cellular level. Recently, it has been shown that calorie restriction extends the lifespan of yeast. We conduct similar experiments on mammalian cell cultures.**

Keywords* — *calorie restriction, cell aging, glucose metabolism, survival curves, RT-PCR

Motivation and aim

Motivation

The lifespan of yeast undergoing chronological aging can be extended by calorie restriction (CR) [1, 2]. CR in this case means reducing the amount of glucose or replacing glucose with non-fermentable carbon sources. In addition, it was also possible to increase the lifespan of yeast with the help of some drugs that mimic this effect (for example, rapamycin) [3].

Aim

We studied how CR affects the lifespan of mammalian cell culture in the "stationary phase aging" (similar to chronological aging) model, as well as the expression of genes associated with autophagy and glucose metabolism.

Methods

Transformed Chinese hamster cells were grown at 37°C in Dulbecco's modified Eagle's medium (or Minimum Essential Medium, or Medium 199) supplemented with 10% bovine serum. The influence of various ways of dilution of medium and addition of CR mimetics on viability of the culture, its growth and subsequent dying out in the stationary phase as well as gene-expression profile was analyzed. The resulting cell survival curves were approximated using the Gompertz equation. Differences were considered statistically significant at $p<0.05$. Mathematical calculations and statistical data processing were performed using SigmaPlot 12.0 software (Systat Software Inc., United States).

Results

We found no dependence of lifespan of "stationary phase aging" mammalian cell culture on glucose content; in some cases, the cells lived longer on media with high glucose content. Drugs that could presumably cause CR either did not affect (2,4-dinitrophenol) or reduced (metformin) the lifespan of the culture. Dilution of the culture medium with the isotonic Quinton Marine Plasma increased lifespan, however, this effect can be associated not so much with CR, but with a more successful combination of components in the new medium. Despite the lack of the same encouraging results as in experiments with yeast, it cannot be denied that the effect of CR is realized at the cellular level. Inhibition of autophagy, one of the main mechanisms through which CR is realized, reduced the lifespan of a cell culture. Moreover, the expression of some genes responsible for the synthesis of glucose metabolism enzymes and autophagy-related proteins (LDHa, AMPK, LC3, LAMP1) differs in "young" and "old" cells.

Conclusion

The results of experiments on yeast, at first glance, confirm that CR increases cell lifespan. However, the interpretation of such experiments seems to us not always correct [4, 5]. Yeast are grown in a nutrient-rich medium, and then transferred to a medium with a lack of nutrient components or with CR mimetics. In such experiments, it is necessary to determine the metabolic activity of "aging" yeast; it is likely that they become metabolically inactive. This way of increasing lifespan is not suitable for actively functioning postmitotic cells such as neurons and myocytes. Based on our experiments, we can conclude that the "old" cells are metabolically active (according to the results of changing the expression level of some genes and the kinetics of cell culture dying out), but we do not always get the same results as in experiments with yeast.

Acknowledgment

The study was funded by RFBR according to the research project # 18-34-00813.

References

[1] Burtner C.R., Murakami C.J., Kennedy B.K., Kaeberlein M. (2009) A molecular mechanism of chronological aging in yeast. Cell Cycle. 8(8): 1256-1270.

[2] Murakami C.J., Wall V., Basisty N., Kaeberlein M. (2011) Composition and acidification of the culture medium influences chronological aging similarly in vineyard and laboratory yeast. PloS one. 6(9): e24530.

[3] Alvers A.L., Wood M.S., Hu D., Kaywell A.C., Dunn, Jr. W.A., Aris J.P. (2009) Autophagy is required for extension of yeast chronological life span by rapamycin. Autophagy. 5(6): 847-849.

[4] Morgunova G.V., Klebanov A.A., Marotta F., Khokhlov A.N. (2017) Culture medium pH and stationary phase/chronological aging of different cells. Moscow Univ. Biol. Sci. Bull. 72(2): 47-51.

[5] Khokhlov A.N., Klebanov A.A., Morgunova G.V. (2018) On choosing control objects in experimental gerontological research. Moscow Univ. Biol. Sci. Bull. 73(2): 59-62.

DOI 10.18699/BGRS/SB-2020-400

MAPK pathways and alphaB-crystallin phosphorylation in brain: a focus on aging and Alzheimer's disease

Natalia Muraleva
Molecular mechanisms of aging
Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
Myraleva@bionet.nsc.ru

Abstract — **Accumulation of intracellular damage and protein aggregates is an universal hallmark of aging and accompanies the development of some age-related diseases include Alzheimer's disease (AD). Alpha-B-Crystallin (CryaB) as the molecular chaperone contributes maintenance of proteostasis by prevention of aggregation of proteins (e.g. amyloid beta) and enables their correct refolding. CryaB activity is regulated by MAPK signaling pathway (MAPKsp) through its phosphorylation. Nevertheless, the link between changes in MAPK–dependent CryaB phosphorylation with age and the development of AD remains unclear. Here, we examined p38 MAPK– and ERK-dependent phosphorylation of CryaB in the brain of Wistar rats with normal aging and senescence-accelerated OXYS rats at the different stages of the development of AD-like pathology, including the presymptomatic stage. The most significant changes identified in the p38 MAPK–dependent CryaB phosphorylation. The level of p-Ser59-CryaB in the brain of Wistar rats increased with the age on the background of p38-MAPKsp activation. Similar but more significant changes accompanied the development of AD-like pathology in OXYS rats. The activation of ERK1/2-dependent CryaB phosphorylation (pSer45-CryaB) was detected at the early age and at the late stages of AD-like pathology in OXYS, while changes in the ERK1/2 signaling pathway were detected in Wistar rats with age. Thus, alteration of MAPK-dependent phosphorylation CryaB occurs with the normal aging. Manifestation and progression of the signs of the AD occurs against the background of activation of p38MAPK-dependent phosphorylation of CryaB. Activation of EPK-dependent CryaB phosphorylation is characteristic of the preclinical and progressive stage of the AD-like pathology.**

Keywords — aging, phosphorylation, MAPK pathways, alphaB-crystallin, Alzheimer's disease, OXYS rats

Introduction

Accumulation of intracellular damage and protein aggregates is an universal hallmark of aging. Maintenance of protein proteostasis is attained through precisely coordinated systems that must rapidly correct unwanted proteomic changes. The points of cross talk between the unfolded protein response and MAPK signaling pathways that may contribute to our understanding of the mechanisms of alteration of proteostasis processes with aging and with development of age-related diseases including Alzheimer's disease (AD). MAPK signaling network regulates cell survival and death responses following a variety of stresses including misfolded protein response stress. Two subfamilies of MAPKs (p38MAPK and ERK1/2) participate in this by regulating the activity of the alphaB-crystallin (CryaB) through its phosphorylation. CryaB as a molecular chaperone prevents aggregation of proteins (e.g. amyloid beta) and enables their correct refolding. Nevertheless, the link between changes in MAPK–dependent CryaB phosphorylation with the normal aging and the development of AD remains unclear. Here, we examined CryaB, phospho- (p-) Ser59-CryaB and p-Ser45-CryaB protein amounts in the brain of Wistar rats with normal aging and senescence-accelerated OXYS rats at different stages of the development of AD-like pathology. We compared this result with the changes in expression of genes involved in the p38 MAPK and ERK 1/2 signaling pathways and the content of key proteins of these pathways in brain of Wistar and OXYS rats.

Materials and Methods

The work was carried out on male OXYS rats and Wistar rats (control) aged 20 days, 5, and 18 months based in the Center of Genetic Resources of Laboratory Animals at the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (Novosibirsk, Russia).

To detect changes in the expression of genes involved in MAPK- and ERK1/2 signaling pathways in Wistar and OXYS rats, we analyzed RNA-Seq data obtained previously [1].

Standard techniques of western blott analysis and immunohistochemistry were used to examine of content of CryaB, p-Ser45-CryaB, p-Ser59-CryaB and key proteins of MAPK- and ERK1/2 signaling in the prefrontal cortex and hippocampus of Wistar and OXYS rats. The data were subjected to two-way analysis of variance in the Statistica 8.0 software.

Results

Here, we showed that in the prefrontal cortex of rats OXYS and Wistar p-Ser59-CryaB was detected only in the detergent-insoluble fraction. Its level did not differ between OXYS and Wistar rats at the age of 20 days. Increase of p-Ser59-CryaB content occurred with age in rats of both strains. The manifestation and progression of signs of AD in OXYS rats were accompanied by an increase in the level of p-Ser59-CryaB relative to Wistar rats in the prefrontal cortex and hippocampus. It should be noted that the content of p-Ser59-CryaB in the detergent-insoluble fraction was significantly higher in OXYS rats, which indicates the formation of stronger links of p-Ser59-CryaB with the target molecule. Indeed, immunohistochemical staining of brain sections revealed in the brain of 18-month-old OXYS rats the joint localization of p-Ser59-CryaB with toxic beta amyloid (Aβ1-42). CryAB phosphorylation (S59) is mediated through activation of p38 protein kinase. A significant increase in the p38 MAPK protein content in the detergent-insoluble fraction was found in the cortex of OXYS rats aged 5 and 18 months, while its level did not change in Wistar rats. In these age groups, OXYS rats showed an increase in the level of phosphorylation of p38 MAPK protein in both protein fractions. A similar study of the contents of p38MAPK and p-p38MAPK in the hippocampus of OXYS and Wistar rats during the same age periods revealed similar results with the level of these proteins in the prefrontal cortex.

According to the analysis of transcriptomes of the prefrontal cortex and hippocampus, the development of AD signs in OXYS rats is accompanied by an increase in the number of differentially expressed genes (DEGs) involved in the p38 MAPK signaling pathway. Changes in the mRNA level of genes in OXYS rats are associated with such categories of gene ontologies as the cascade of protein kinases, phosphorylation, calcium signaling, and the protein kinase activating signaling pathway.

Next, we showed that in the prefrontal cortex and hippocampus of OXYS and Wistar rats p-Ser45-CryaB was contained in detergent-soluble and detergent-insoluble protein fractions. The level of p-Ser45-CryaB in OXYS rats was higher than in the control Wistar rats at the age of 20 days and 18 months. It should be noted that the accumulation of p-Ser45-CryaB in the detergent-insoluble fraction increased already at the preclinical stage of AD-like pathology in OXYS rats. CryaB phosphorylation (S45) is mediated through activation of the ERK1/2 signaling pathway controlled by ERK1/2 protein kinases. A significant increase in the protein content of ERK1/2 was found in the prefrontal cortex and hippocampus of OXYS rats at all studied ages. According to the analysis of transcriptomes of the prefrontal cortex and hippocampus, the progression of signs of AD in OXYS rats was accompanied by an increase in the number of DEG involved in the ERK1 / 2 signaling pathway.

CONCLUSION

Thus, alteration of MAPK-dependent phosphorylation of CryaB is happening with the normal aging. Activation of EPK1/2-dependent CryaB phosphorylation occurs with age and is characteristic for the preclinical and progressive stages of the AD-like pathology. Manifestation and progression of the signs of the AD occur against the background of activation of p38MAPK-dependent phosphorylation of CryaB and changes in gene expression of this signaling pathway. An increase in the level of p-Ser59-CryaB and its joint localization with Aβ1-42 can be considered as a response to the accumulation of toxic protein aggregates in the brain, which is an important part of the endogenous mechanism of AD development.

ACKNOWLEDGMENT

Microscopy was performed at the Microscopy Center of the Institute of Cytology and Genetics, SB RAS, Russia. This work was supported by grant from the RFBR (#18-015-00336).

REFERENCES

[1] Stefanova, N.A., Maksimova, K.Y., Rudnitskaya, E.A., Muraleva, N.A., Kolosova, N.G. Association of cerebrovascular dysfunction with the development of Alzheimer's disease-like pathology in OXYS rats. BMC Genomics. 2018, 9, 19(Suppl 3):75. https://doi.org/10.1186/s12864-018-4480-9.

DOI 10.18699/BGRS/SB-2020-401

Resting-state alpha rhythm modulation after divergent problem solving in aged adults

Evgeniya Privodnova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
privodnovaeu@physiol.ru

Nina Volf
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia volf@physiol.ru

Ekaterina Merkulova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia
merkaterine@gmail.com

Victoriya Bilik
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russia v.bilik@mail.ru

Abstract — Given that repetitive cognitive activity leaves post-task traces after task performance and divergent problem solving is accompanied by alpha power enhancement, one may suppose that divergent tasks performance may induce resting-state alpha rhythm modulation. The results obtained in the current pilot study indicated that alpha power increased after 30-40 minute divergent problem solving in parietal brain region in older adults.

Keywords — alpha rhythm modulation, divergent thinking, aging

Motivation and aim

Motivation

Resting-state EEG alpha power is known to dramatically decrease as people ages [1]. At the same time, due to the association between alpha power and inhibitory abilities [2, 3], alpha increase may be useful to preserve top-down inhibitory capacity in older age. Given that repetitive cognitive activity leaves post-task traces after task performance [4] and that divergent problem solving is accompanied by alpha power enhancement [5], one may suppose that divergent tasks performance may results in resting-state alpha rhythm modulation.

Aim

The current study aimed to examine if alpha power increases after performance of verbal and visual divergent tasks. To this end, we explored the changes in resting-state alpha power that followed 30-40 minute divergent problem solving session, which consists of repetitive tasks of verbal and visual domains.

Methods

We recorded EEG in 30 right handed older adults (55-75 years old, mean age=64.2) before, during and after divergent thinking session. All subjects continued their professional activity; exclusion criteria was self-reported history of psychiatric, major medical, and neurological diseases. Divergent thinking session lasted for 30-40 minutes and included 30 verbal tasks (“Alternate Uses Task”, [6]) and 30 visual tasks (“Incomplete figures”, [6]). Originality (uniqueness) of creative ideas in each creativity domain was assessed using database of the response rate in these tests developed by authors previously. The EEG data were registered using 52 Ag–AgCl electrodes placed according to the modified version of the international 10–20 system via “Neuroscan 4.4” (USA). Fronto-central electrode was used as the ground, and electronically linked mastoid electrodes as reference. Electrode impedances did not exceed 5 kΩ. The EEG was amplified using Neuroscan amplifiers with a gain of 250 and a bandpass of 0–50 Hz. Artifacts rejection was made using independent component analysis implemented in EEGLAB toolbox (https://sccn.ucsd.edu/eeglab/). Bandwidth for the upper alpha band was defined as [(individual alpha peak frequency) to (individual alpha peak frequency +2)]. To examine alpha activity, we used three-minute recordings in eyes-open resting-state condition from pre-test and post-test intervals as well as EEG recorded during successfully resolved verbal and visual tasks. In all cases, EEG data were segmented into 2 s epochs and submitted to further analysis. We calculated current source density (CSD) estimates via standardized Low Resolution Brain Electromagnetic Tomography (sLORETA) [13]. For each interval under analysis, we yielded the CSD distribution across all 6239 voxels in a realistic head model, which is based on the MNI 152 template, with spatial resolution of 5x5x5 mm. Statistical contrasts of CSD estimates was performed using statistical nonparametric mapping (SnPM) with 5000 randomizations in the LORETA package. Pearson correlations were calculated in STATISTICA10. MNI coordinates of voxel with maximum difference between pre-test CSD and post-test CSD was used as a “seed” in follow-up region of interest (ROI) analysis. ROI was identified as cluster of voxels within the circle centered at seed, 15 mm in radius. For ROI analysis, eLORETA transformation matrix was used. ROIs were extracted from differential maps of post-task/pre-task changes, visual/pre-task changes, verbal/pre-task changes.

Results

Alpha CSD increased from pre-task to post-task resting-state interval (p=0.019) in the Superior Parietal lobule (Brodmann area 7) (see Figure, A). CSD estimates during visual task were higher in comparison with CSD estimates in pre-task interval with localization of maximal difference (p=0.0003) in the Superior Parietal lobule (Brodmann area 7). However, CSD estimates during verbal task did not differ from CSD estimates in pre-task interval. The results suggest that alpha increase from pre-task to post-task interval is more likely to be formed by activity during performance of visual task. Correlation analysis of CSD estimates within ROIs revealed positive association between visual task and post-task intervals (r=0.46, p=0.013) (see Figure, B), while no significant correlations were found between verbal task and post-task intervals, providing additional evidence for the prominent role of visual domain. ROI estimates extracted from visual/pre-task changes were positively correlated with visual originality (r=0.48, p=0.008) (see Figure, C). This finding suggests that CSD estimates of visual/pre-task changes within ROI reflected specific to visual creativity aspects of task performance. Correlations between efficacy indicator and CSD estimates in ROI, derived from post-task/pre-task changes, failed to reach significance level. Therefore, the traces

in post-task EEG may be associated with extensive use of this region during visual task implementation, independently of efficiency of task performance.

ACKNOWLEDGMENT

Supported by budgetary funding for basic scientific research (theme No. AAAA-A16-116021010228-0, EEG data collection) and by RFBR and Government of the Novosibirsk region according to the research project No. 19-415-543009 (EEG data processing and analysis, conceptualization, paper writing and preparation).

REFERENCES

[1] R. J. Vaden, N. L. Hutcheson, L. A. McCollum, J. Kentros, and K. M. Visscher, "Older adults, unlike younger adults, do not modulate alpha power to suppress irrelevant information", Neuroimage, vol. 63(3), pp. 1127–1133, Aug. 2012.

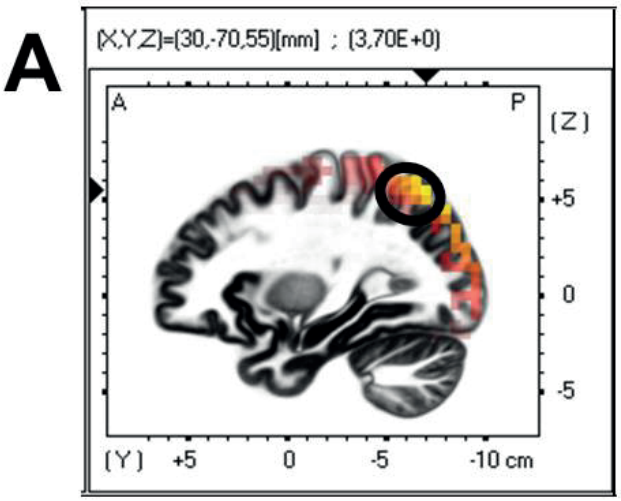
[2] O. Jensen and A. Mazaheri, "Shaping functional architecture by oscillatory alpha activity: gating by inhibition", Front. Hum. Neurosci., vol. 4:e186, Nov. 2010.

[3] G. Borghini, M. Candini, C. Filannino, M. Hussain, V. Walsh, V. Rome, N. Zokaei, and M. Cappelletti, "Alpha Oscillations Are Causally Linked to Inhibitory Abilities in Ageing" J. Neurosci. vol. 38 (18), pp. 4418-4429, May 2018.

[4] C. Moisello, H.B. Meziane, S. Kelly, B. Perfetti, S. Kvint, N. Voutsinas, D. Blanco, A. Quartarone, G. Tononi and M.F. Ghilardi, "Neural activations during visual sequence learning leave a trace in post-training spontaneous EEG", PloS one, vol. 8(6):e65882, Jun 2013.

[5] Fink and M. Benedek, "EEG alpha power and creative ideation", Neurosci. Biobehav. Rev., vol. 44, pp. 111-123, Jul. 2014.

[6] E.P. Torrance, "Torrance tests of creative thinking norms-technical manual : verbal tests, forms A and B : figural tests, forms A and B", Princeton, New Jersey Personal Press, 1966.



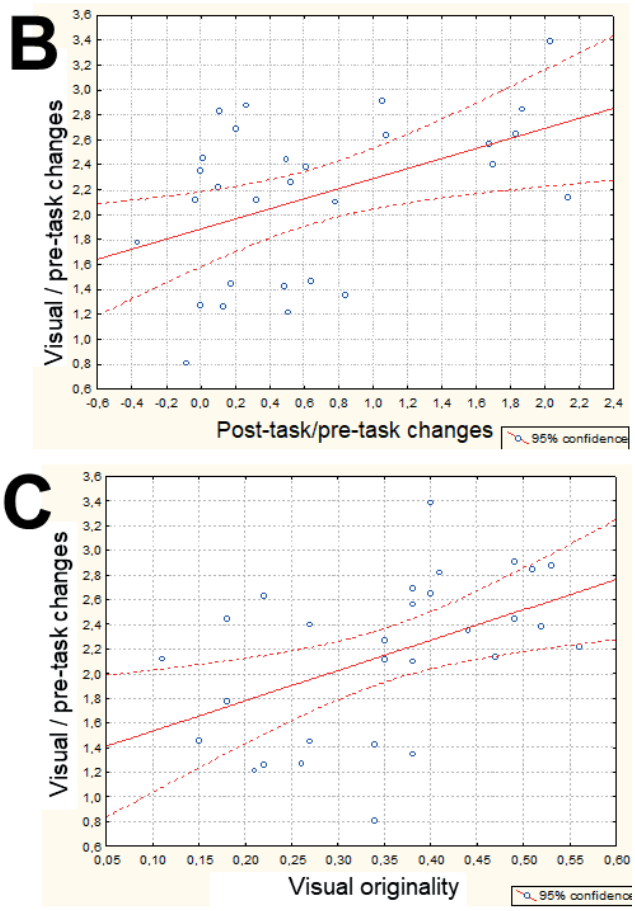


Fig. 1. Associations of CSD estimates within ROI in upper-alpha frequency band. A) Increase in CSD estimates from pre-task to post-task interval was statistically significant according to sLORETA SnPM in right Superior Parietal Lobule (Brodmann area 7). B) Correlations between ROI estimates of post-task/pre-task changes and ROI estimates of visual/pre-task changes. C) Correlations between ROI estimates of visual/pre-task changes with originality in visual task. Significant differences ($p<0.05$) are colored. ROI is marked by circle

DOI 10.18699/BGRS/SB-2020-402

Difference between younger and older adults in post-task traces after creative task performance matches PASA model

Evgeniya Privodnova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russian Federation
privodnovaeu@physiol.ru

Nina Volf
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russian Federation
volf@physiol.ru

Dariya Bazovkina
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russian Federation
daryabazovkina@gmail.com

Ekaterina Merculova
Scientific Research Institute of Physiology and Basic Medicine
Novosibirsk, Russian Federation
merkaterine@gmail.com

***Abstract* — Brain activity during during execution of the experimental task in older adult differ from that in younger ones. HAROLD model describes hemispheric asymmetry reduction in older adults, while PASA model postulates posterior-anterior shift in aging. Up to date, age-related differences in use-dependent neuroplasticity are under-studied. Obtained results showed that in both pre-task and post-task intervals alpha power anterior-posterior gradient (posterior>anterior) was higher in younger group in comparison with older group. This pattern of age-related differences is consistent with PASA model.**

Keywords — alpha rhythm, post-task traces, PASA, aging

Motivation and aim

Motivation

Repetitive cognitive activity induces neuroplasticity prosesses [1], which may manifests itself in electroencephalography (EEG) by post-task traces after task performance. It is well-known that brain activity during during execution of the experimental task in older adult differ from that in younger ones. Empirical data were summarized in several patterns of age-related changes probably reflecting neuro-compensatory processes. HAROLD model describes hemispheric asymmetry reduction in older adults [2], while PASA model postulates posterior-anterior shift in aging [3]. Up to date, age-related differences in use-dependent neuroplasticity are under-studied.

Aim

The current study aimed to explore age-related differences in post-task traces after creative task performance and to examine whether these differences match the functional aging brain models (PASA, HAROLD).

Methods

We recorded EEG in 31 younger (19-33 years) and 30 older adults (55-75 years) before, during and after divergent thinking session. All subjects were students or continued their professional activity; exclusion criteria was self-reported history of psychiatric, major medical, and neurological diseases. Divergent thinking session lasted for 30-40 minutes and included 30 verbal tasks (“Alternate Uses Task”, [4]) and 30 visual tasks (“Incomplete figures”, [4]). The EEG data were registered using 52 Ag–AgCl electrodes placed according to the modified version of the international 10–20 system via “Neuroscan 4.4” (USA). Fronto-central electrode was used as the ground, and electronically linked mastoid electrodes as reference. Electrode impedances did not exceed 5 kΩ. The EEG was amplified using Neuroscan amplifiers with a gain of 250 and a bandpass of 0–50 Hz. Artifacts rejection was made using independent component analysis implemented in EEGLAB toolbox (https://sccn.ucsd.edu/eeglab/). Bandwidth for the upper alpha band was defined as [(individual alpha peak frequency) to (individual alpha peak frequency +2)]. To examine use-dependent changes alpha activity, we used three-minute recordings in eyes-open resting-state condition in pre-test and post-test intervals. EEG data were segmented into 2 s epochs and submitted to further analysis.For each derivation, mean alpha power was calculated using Fourier transform via EEGLAB toolbox. Then derivations were combined in the frontal, central, temporal- central, and parietal-occipital left regions and the same right regions. For statictical analysis we used repeated measures analysis of variance (ANOVA) implemented in STATISTICA10.

Results

We found significant interaction of factors “region” and “time” ($F(3, 177)= 46$, $p<0.0001$). For both age groups, alpha power in post-task interval was higher than at pre-task interval in all regions except of frontal (anterior-posterior gradient, posterior>anterior). ANOVA yielded significant interaction of factors “region” and “age” ($F(3, 177)=3$, $p<0.02$). In younger group, alpha power in central and parietal-occipital region was higher than in frontal and temporal-central regions. In older group, alpha power at central brain region was higher in comparison with all the rest regions, however there was no difference between frontal and parietal-occipital regions. In both pre-task and post-task intervals alpha power anterior-posterior gradient (posterior>anterior) was higher in younger group in comparison with older group ($p<0.05$). This pattern of age-related differences is consistent with PASA model [3].

Acknowledgment

Supported by budgetary funding for basic scientific research (theme No. AAAA-A16-116021010228-0, EEG data collection) and by RFBR and Government of the Novosibirsk region according to the research project No. 19-415-543009 (EEG data processing and analysis, conceptualization, paper writing and preparation).

REFERENCES

[1] P.Voss, M.E. Thomas, J.M. Cisneros-Franco, and E. de Villers-Sidani, "Dynamic Brains and the Changing Rules of Neuroplasticity: Implications for Learning and Recovery", Front. Psychol., vol. 8:e1657, Oct. 2017.

[2] B.M. Duda, M.M. Owens, E.S. Hallowell, and L.H. Sweet, "Neurocompensatory Effects of the Default Network in Older Adults", Front. Aging Neurosci., vol. 11:e111, Jun. 2019.

[3] S.W. Davis, N.A. Dennis, S.M. Daselaar , M.S. Fleck, and R. Cabeza, "Que PASA? The posterior-anterior shift in aging", Cereb. Cortex, vol. 18(5), pp. 1201-1209, May 2008.

[4] E.P. Torrance, "Torrance tests of creative thinking norms-technical manual : verbal tests, forms A and B : figural tests, forms A and B", Princeton, New Jersey Personal Press, 1966

DOI 10.18699/BGRS/SB-2020-403

Delay of early postnatal development as a risk factor for accelerated aging and Alzheimer's disease

Ekaterina Rudnitskaya
Molecular mechanisms of aging
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
rudnickaya@bionet.nsc.ru

Tatiana Kozlova
Molecular mechanisms of aging
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
kozlova@bionet.nsc.ru

Alena Burnyasheva
Molecular mechanisms of aging
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
burnyasheva@bionet.nsc.ru

Natalia Stefanova
Molecular mechanisms of aging
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
stefanovan@bionet.nsc.ru

Nataliya Kolosova
Molecular mechanisms of aging
Institute of Cytology and Genetics
SB RAS
Novosibirsk, Russia
kolosova@bionet.nsc.ru

Abstract — **Alzheimer's disease (AD) is the most common cause of dementia worldwide. In the case of sporadic form of AD (~95% of cases), its mechanisms are unknown. Accumulating data indicate middle age as a critical period for the relevant pathological processes; however, it has been reported only recently that in the early postnatal period — when brain development is completing—preconditions for a decrease in cognitive abilities and for accelerated aging can form. Here, we hypothesized that specific features of early postnatal brain development may be considered some of the prerequisites of AD development at an advanced age. To test this hypothesis, we used OXYS rats, which are a suitable model of sporadic AD. The duration of gestation was lower in OXYS rats compared to control Wistar rats which may result in developmental retardation. Indeed, we noted retardation of emergence of postural and locomotor skills as well as decreased locomotor activity and increased anxiety in OXYS rats already at a young age: possible signs of altered brain development. We demonstrated retardation of the peak of postnatal neurogenesis in the hippocampal dentate gyrus (DG) of OXYS rats. Delayed neuronal maturation led to alterations of mossy-fiber formation and was accompanied by altered astrocytic migration from DG. We assume that the observed retardation in the development of the brain may be predictor of AD signs development in OXYS rats and, possibly, these diseases in humans.**

Keywords — neurogenesis, brain development, Alzheimer's disease, OXYS rats

Introduction

Aging is the major risk factor for development of wide range of psychiatric disorders including AD. The most common (~95% of cases) sporadic AD develops asymptomatically for many years prior to its manifestation [1] and accumulating data indicate middle age as a critical period for the relevant pathological processes [2, 3]; however, the question of when AD starts to develop remains open. It has been reported only recently that in the early postnatal period—when brain development is completing—preconditions for a decrease in cognitive abilities and for accelerated aging can form [4-6]. Evidently, this hypothesis cannot be verified without using adequate animal models. Previously we have confirmed that senescence-accelerated OXYS rats are a suitable model of sporadic AD [7-10]. Thus, we suggested that the study of the specific features of early postnatal development using OXYS rats may shed a light on long-lasting effects of alterations of brain development and its connection to manifestation of AD late in life.

This work was supported by the Russian Science Foundation (# 19-15-00044).

Materials and methods

For evaluation the duration of gestation we used 3-months-old virgin female OXYS and control Wistar rats and age-matched males (n = 25 per strain and gender). For evaluation of body and brain weight we used offspring from female OXYS and Wistar rats (n = 4 female rats for one time point: day of birth, postnatal day 1 (PND1), PND3, PND5, PND7, PND10, PND14, PND20, PND45). For analysis of physical development (auricle detachment, emergence of pelage and incisors, eye opening, dropping of the testes in male and vaginal opening in female rats) and emergence of postural and locomotor skills (elevation of head and shoulders, grasp reflex for fore and hind limbs, righting on surface and in midair, negative geotaxis, placing, jumping and jumping across cliff) we used offspring from 4 female OXYS and Wistar rats per 2-3 parameters. To evaluate locomotor activity we used open field test for animals at PND10, PND14, PND20 and PND45; to evaluate anxiety level we used elevated plus maze test for animals at PND45 (n = 15 per strain and age). To analyze hippocampal development of OXYS and Wistar rats we evaluated the density of neuronal progenitors and proliferating cells in the DG at the day of birth, PND1, PND3, PND5, PND7, PND10, PND14, PND20 and PND45 as well as the density of neuronal and astroglial cell lines at PND10, PND14, PND20 and PND45 by immunochistochemistry; intensity of apoptosis at PND10, PND14, PND20 and PND45 by TUNEL method (n = 3-6 per strain and age). To evaluate morphological characteristics of mossy fibers, hippocampal slices of OXYS and Wistar rats at PND10, PND14, PND21 and PND45 underwent DiI injection into the DG area (n = 6-8 per strain and age) with following microscopic analysis. To clarify the connection between features of early development and AD-like pathology we evaluated the density of progenitors and cells from neuronal and astroglial cell lines in the DG of OXYS rats at period of active manifestation (age 3 months) and progression (age 18 months) of the disease signs. The data were subjected to two-way analysis of variance in the Statistica 8.0 software.

Results

We showed that duration of gestation was shorter in OXYS rats compared to Wistar rats, and OXYS pups were born with lower body as well as brain weight. At PND1 the

body weight did not differ between OXYS and Wistar pups, whereas brain weight remained lower in OXYS pups. However, to PND3 the differences in brain weight between rat strains become insignificant. Shortened duration of gestation in OXYS rats may result in developmental retardation, such as delayed auricle detachment, emergence of pelage and incisors, eye opening, dropping of the testes in male rats and vaginal opening in female rats. Moreover, we observed retardation of emergence of postural and locomotor skills in OXYS rats: elevation of head and shoulders, grasp reflex for hind limbs, righting on surface and in midair, negative geotaxis, placing (when chin elicited) and jumping across cliff. Also we demonstrated decreased locomotor activity in OXYS rats at PND20 and PND45 and increased anxiety at PND45. These changes may reflect deterioration of brain development. Indeed, we demonstrated that OXYS pups were born with decreased density of amplifying neuronal progenitors in DG; however, to PND1 the parameter increased reaching the level of Wistar pups. The density of amplifying neuronal progenitors naturally decreased to PND10 in both rat strains; however, in OXYS pups the decrease was more pronounced: at PND10 the parameter was lower compared to Wistar pups; to PND14 the difference disappeared. The density of immature neurons was higher in DG of OXYS rats at PND10 which may indicate delay of postnatal peak of neurogenesis. Moreover, the peak of postnatal apoptosis was observed at PND20 in OXYS rats, which indicate its retardation. As for neurite development, the following features of mossy-fiber formation in OXYS rats were revealed by our colleagues [11]: a smaller suprapyramidal bundle and lager infrapyramidal bundle, less pronounced fasciculation of granule cell axons, and smaller size and an irregular shape of nuclei in the CA3 pyramidal layer. However, to PND45 these features of hippocampal morphology in OXYS rats disappeared, and the parameters became the same as in Wistar rats. Besides we found altered astrocytic migration from DG in OXYS rats. It's important to emphasize that there were long-lasting consequences of these abnormalities of postnatal hippocampal development in OXYS rats such as insufficiency of astrocytic support of hippocampal neurogenic niche at 3 months of age and lower intensity of neurogenesis during lifespan: indeed, the density of stem quiescence neuronal progenitors was higher in DG of OXYS rats at 18 months of age.

Conclusion

To conclude, in present work we observed shortened duration of gestation and, as a consequence, delayed postnatal development of OXYS rats. Behavioral abnormalities may reflect altered brain development. Indeed, peaks of postnatal neurogenesis and apoptosis in DG of OXYS rats fell later compared to Wistar rats. Moreover, formation of hippocampal mossy fibers was altered in OXYS rats. We suppose that the observed features of early hippocampal development have long-lasting effects on hippocampal plasticity and, thus, are the one of predictors of AD-like pathology in OXYS rats and may be one of risk factors for development of sporadic AD in human.

Acknowledgment

Senescence-accelerated OXYS rats and Wistar rats were obtained from the Breeding Experimental Animal Laboratory of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia. The microscopy was conducted at the Multi-Access Center for Microscopy of Biological Objects (Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia).

References

[1] M. Hersi, B. Irvine, P. Gupta, J. Gomes, N. Birkett, D. Krewski, "Risk factors associated with the onset and progression of Alzheimer's disease: A systematic review of the evidence," Neurotoxicology, vol. 61, 2017, pp. 143–187.

[2] R. J. Bateman, C. Xiong, T. L. Benzinger, A. M. Fagan, A. Goate, N. C. Fox, et al., "Clinical and biomarker changes in dominantly inherited Alzheimer's disease," New England Journal Medicine, vol. 367, 2012, pp. 795–804.

[3] M. Crous-Bou, C. Minguillón, N. Gramunt, J. L. Molinuevo, "Alzheimer's disease prevention: from risk factors to early intervention," Alzheimers Research Therapy, vol. 9, 2017, No. 71.

[4] K. Heinonen, J. G. Eriksson, J. Lahti, E. Kajantie, A. K. Pesonen, S. Tuovinen, et al., "Late preterm birth and neurocognitive performance in late adulthood: a birth cohort study," Pediatrics, vol. 135, 2015, pp. 818–825.

[5] N. N. Nalivaeva, A. J. Turner, I. A. Zhuravin, "Role of Prenatal Hypoxia in Brain Development, Cognitive Functions, and Neurodegeneration," Frontiers Neuroscience, vol. 12, 2015, No. 825.

[6] L. K. Axelrud, J. R. Sato, M. L. Santoro, F. Talarico, D. S. Pine, L. A. Rohde, et al., "Genetic risk for Alzheimer's disease and functional brain connectivity in children and adolescents," Neurobiology Aging, vol. 82, 2019, pp. 10–17.

[7] N. A. Stefanova, O. S. Kozhevnikova, A. O. Vitovtov, K. Y. Maksimova, S. V. Logvinov, E. A. Rudnitskaya, E. E. Korbolina, N. A. Muraleva, N. G. Kolosova, "Senescence-accelerated OXYS rats: A model of age-related cognitive decline with relevance to abnormalities in Alzheimer disease," Cell Cycle, vol. 13, 2014, pp. 898–909.

[8] N. A. Stefanova, K. Y. Maksimova, E. A. Rudnitskaya, N. A. Muraleva, N. G. Kolosova, "Association of cerebrovascular dysfunction with the development of Alzheimer's disease-like pathology in OXYS rats," BMC Genomics, vol. 19, 2018, No. 75.

[9] N. A. Stefanova, N. I. Ershov, K. Y. Maksimova, N. A. Muraleva, M. A. Tyumentsev, N. G. Kolosova, "The Rat Prefrontal-Cortex Transcriptome: Effects of Aging and Sporadic Alzheimer's Disease-Like Pathology," Journals Gerontology: Series A Biological Sciences Medical Sciences, vol. 74, 2019, pp. 33–43.

[10] E. A. Rudnitskaya, T. A. Kozlova, A. O. Burnyasheva, N. G. Kolosova, N. A. Stefanova, "Alterations of hippocampal neurogenesis during development of Alzheimer's disease-like pathology in OXYS rats," Experimental Gerontology, vol. 115, 2019, pp. 32–45.

[11] E. A. Rudnitskaya, T. A. Kozlova, A. O. Burnyasheva, A. E. Tarasova, T.M. Pankova, M. V. Starostina, N. A. Stefanova, N. G. Kolosova, "Alteration of postnatal hippocampal development as a potential risk factor for Alzheimer's disease-like pathology in OXYS rats," Frontiers Neuroscience, 2020, in press.

DOI 10.18699/BGRS/SB-2020-404

Cholinergic deficit in olfactory bulbectomized animals as a model of neurodegenerative diseases

Mikhail Stepanichev
IHNA&NPh RAS, Moscow, Russia
mikhail_stepanichev@yahoo.com

Olga Nedogreeva
IHNA&NPh RAS, Moscow, Russia
nedogreewaolga@gmail.com

Natalia Lazareva
IHNA&NPh RAS, Moscow, Russia
nalaza@rambler.ru

Anna Manolova
IHNA&NPh RAS, Moscow, Russia
ashit@mail.ru

Natalia Gulyaeva
IHNA&NPh RAS, Moscow, Russia
nata_gul@mail.ru

***Abstract* — Cholinergic deficit is a feature of age-related memory decline and age-associated dementia. In the present study we show that olfactory bulbectomy in mice resulted in the development of cognitive impairments and a decrease in the number of cholinergic markers in the medial septal area while the number of neurons remained being stable. The involvement of oxidative stress in the development of cholinergic abnormalities was evident. Thus, olfactory bulbectomy may be used as a model of early stages of neurodegenerative diseases.**

Keywords* — *acetylcholine, olfactory bulbectomy, Alzheimer's disease, oxidative stress

Motivation and aim

Motivation

Acetylcholine (ACh) is a neurotransmitter playing an important role in higher brain functions, including attention, learning and memory. Cognitive decline found in aging is associated with the development of cholinergic deficit probably because of degeneration of cholinergic neurons. The impairments of the cholinergic system often manifests in patients with dementia, including Alzheimer's disease. However, the number of cholinergic neurons in the basal nucleus of Meynert in aged individuals without cognitive impairments or in patients with mild cognitive impairments was similar to that observed in age-matched controls. Similarly, the development of cholinergic deficit in the absence of neuronal loss could be observed in some animal models of neurodegenerative diseases.

Aim

Here, we studied the time course of the development of behavioral impairments and features of cholinergic deficit in olfactory bulbectomized animals.

Methods

The experiments were performed in adult Wistar female rats, CD1 female mice and C57Bl/6 male mice. The animals were subjected to olfactory bulbectomy (OBX) by aspiration under chloral hydrate anesthesia as described previously.

OBX animals were subjected to series of behavioral tests, including tests for estimation of emotionality and cognitive functions. Immunohistochemical and biochemical methods were used to study the indices of cholinergic system and oxidative stress.

Results

OBX in female Wistar rats resulted in the appearance of some depressive-like behavioral features, such as the decreased sucrose consumption, hyperactivity, anxiety-like behavioral features impaired short-term memory and an enhancement of avoidance behavior one month after surgery. However, these behavioral abnormalities could be also associated with some disturbances in hippocampal function, which is supported by the presence of cellular changes in this brain structure. We could not find any effects of OBX on the number of cholinergic neurons in the medial septum-diagonal band of Broca as well as the ACh content and acetylcholinesterase activity in the septum, hippocampus, and neocortex. On the contrary, in female CD1 mice OBX impaired spontaneous alternation behavior and decreased the number of cholinergic neurons in the medial septum-diagonal band of Broca. These data demonstrate that OBX does not significantly affect the cholinergic system in rats. We also conclude that rats and mice differentially respond to OBX.

To futher examine the development of cholinergic deficit we studied the effects of OBX in male C57Bl/6 mice. We revealed that behavioral abnormalities developed in OBX mice starting at lesat two weeks after surgey and remained to be detected until 50 days later on. We revealed a decrease in the number of cholinergic cells in the medial septum-diagonal band of Broca. However, the number of neurons did not change in the OBX mice compared to the control group.

We studied the level of oxidative modifications in nucleic acids using anti-8OHG DNA/RNA antibodies. We did not find any staining in the neuronal nuclei whereas cytoplasm stained intensely. We found that the levels of damaged DNA/RNA staining in the cytoplasm of choline acetyltransferase (ChAT) positive or ChAT negative cells in the medial septum were higher compared to sham-operated mice. Furthermore, the content of oxidatively modified proteins was higher in the hippocampus and medial septum of OBX mice. These data show that oxidative stress may be involved in the development of cholinergic deficit found in OBX animals.

Acknowledgment

Supported by the RFBR (20-015-00226).

DOI 10.18699/BGRS/SB-2020-405

Neuronal transcription factors in lifespan control

Alexander Symonenko
IMG RAS, Moscow, Russia
symonenko@gmail.com

Natalia Roshina
IMG RAS, Moscow, Russia
VIGG RAS, Mosocw, Russia
nwumr@yandex.ru

Anna Krementsova
IMG RAS, Moscow, Russia
IBCP RAS, Moscow, Russia
akrementsova@mail.ru

Elena Pasyukova
IMG RAS, Moscow, Russia egpas@rambler.ru

***Abstract* — In *D. melanogaster*, neuronal transcription factors were shown to participate in lifespan control. To identify common pathways for regulating neuronal properties and longevity, we examined molecular targets for transcription factor Stc. Among the targets, genes (proteins) involved in the control of neurogenesis, chromatin structure, and energy metabolism were identified. Our results allowed us to suggest several possible molecular mechanisms by which Stc is involved in lifespan control.**

Keywords — lifespan, transcription, factors, neurons, Drosophila

Motivation and aim

Motivation

Transcription factors play an important role in the systemic control of gene expression. The role of transcriptional cascades in the regulation of neurogenesis in *D. melanogaster* is well known [1]. Earlier, we showed that a number of genes (*stc*, *Lim3*, *esg*, and others) encoding neuronal transcription factors are involved in controlling lifespan [2–4].

Aim

In order to understand which systemic molecular and genetic mechanisms determine the effect of the transcription factor Stc on lifespan and how they are associated with the control of the neuronal development and function, we examined Stc target genes and regulatory cascades in which it is involved.

Methods

Regular RNA-seq and Real-time RT-qPCR analyses were used to experimentally assess *stc* targets. To build *stc* interaction networks, we used data from the open repositories: Flybase (http://flybase.org/), Flymine

(https://www.flymine.org/) and BioGRID (https://thebiogrid.org/). To build the networks, the esyN toolkit (easy networks, http://www.esyn.org/) was used.

Results

The *stc* gene encodes a transcription factor of RNA polymerase II homologous to human transcription factor NFX1 [5]. We found that Stc might impact the transcription levels of the *E(bx)* gene, which encodes the factor of the nucleosome remodeling complex, and interacts with the Ada2b protein, a component of the multi-subunit SAGA complex, which affects the level of histone acetylation, and thus the chromatin structure. This result reinforces our earlier assumption about the possibility of epigenetic inheritance of properties caused by changes in *stc* expression at the embryonic stage over a number of cell generations. Stc also interacted with the genes encoding the SERCA and THADA proteins that control the body's temperature homeostasis, both in human and drosophila models, as well as with other genes that control energy metabolism. The functional variability of SERCA and THADA may have important adaptive value in populations [6]. As expected, Stc interacted with genes involved in neurogenesis. The most interesting interactions were confirmed by data on the effects of *stc* (Stc) on energy metabolism and neuronal structure and function.

Our results allowed us to identify several possible molecular mechanisms by which *stc* (Stc) is involved in lifespan control. However, data obtained so far did not indicate common regulatory pathways by which *stc* (Stc) might control neuronal functions and lifespan.

Acknowledgment

Supported by the RFBR (15-04-05797a, 18-04-01127-a) and by RF Budget (AAAA-A19-119022590053-3).

References

[1] Jothi R. et al. (2009) Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. Mol. Syst. Biol. 5:294.

[2] Roshina N.V. et al. (2014) Embryonic expression of *shuttle craft*, a *Drosophila* gene involved in neuron development, is associated with adult lifespan. Aging US 6:1076-1093.

[3] Rybina O.Y. et al. (2017) Tissue-specific transcription of the neuronal gene *Lim3* affects *Drosophila melanogaster* lifespan and locomotion. Biogerontology 18:739-757.

[4] Symonenko A.V. et al. (2018) Reduced neuronal transcription of *escargot*, the *Drosophila* gene encoding a Snail-type transcription factor, promotes longevity. Front. Genet. 9:151.

[5] Stroumbakis N.D. et al. (1996) A homolog of human transcription factor NF-X1 encoded by the *Drosophila shuttle craft* gene is required in the embryonic central nervous system. Mol. Cell. Biol. 16:192-201.

[6] Moraru A. et al. (2017) THADA regulates the organismal balance between energy storage and heat production. Dev. Cell 41:72-81.

DOI 10.18699/BGRS/SB-2020-406

Effects of melatonin and SkQ1 long-term treatment during aging and development AMD-like retinopathy

Darya V. Telegina
ICG SB RAS, Novosibirsk, Russia
telegina@bionet.nsc.ru

Oyuna S. Kozhevnikova
ICG SB RAS, Novosibirsk, Russia
oidopova @bionet.nsc.ru

Anzhela Z. Fursova
ICG SB RAS, Novosibirsk, Russia
anzhellafursova@yandex.ru

Abstract — Melatonin and antioxidant SkQ1 act like mitochondria-targeted antioxidants, which concentrate in mitochondria at relatively high levels and they may prevent mitochondrial damage during retinal aging and development of age-related retinal disease such as age-related macular degeneration (AMD). However, detailed effects of melatonin and SkQ1 on the biochemical mechanisms underlying therapeutic effect of these drugs during retinal aging and AMD progression remain unclear. Using Wistar rats with normal aging process and senescence-accelerated OXYS rats, which spontaneously develop a phenotype similar to human age-related disorders including AMD-like retinopathy, we found that treatment of SkQ1 and melatonin decreased the incidence and severity of retinopathy in OXYS rats. In Wistar rats, which do not naturally develop retinopathy, ophthalmoscopic inspections did not reveal pathological alterations in the retina of melatonin and SkQ1-treated rats. SkQ1 decreased p62/SQSTM1 protein but not mRNA levels in both OXYS and Wistar rat's retinas as compared of control rats. We observed reduced level of VDAC1 and increased level of glutaminase by long-term treatment of melatonin and SkQ1 in retina of Wistar rats but not OXYS rats. Taken together, our data indicated that long-term treatment of melatonin and mitochondria-targeted antioxidant SkQ1 may retard an age-related decline in the adaptability of retinal cells and may be considered as a strategy to slow down AMD. At the same time effects of melatonin and SkQ1 on molecular events may be different depending on genotype and disease.

Keywords — melatonin, SkQ1, aging, retina, AMD, OXYS rats, Wistar rats

Introduction

Today the percent of older people are on a dramatic increase result in greater elderly human with age-related diseases such as age-related macular degeneration (AMD). AMD is the predominant cause of visual loss in the macula in old people and is characterized by degeneration of retinal pigment epithelium and photoreceptors, impaired autophagy, DNA damage, mitochondrial dysfunction, increased levels of ROS and impaired of blood vessels. It is important that AMD is characterized by retina alterations similar to normal retinal aging. This fact complicates the study of AMD pathogenesis and the search for new therapy drugs.

Accumulating evidence indicates that the melatonin, N-acetyl-5-methoxytryptamine, and mitochondria-targeted antioxidant SkQ1 have a positive effects on AMD treatment [1]. Both SkQ1 and melatonin act like mitochondria-targeted antioxidants, which concentrate in mitochondria at relatively high levels and they may prevent mitochondrial damage in AMD. Melatonin scavenged free radicals generated in the mitochondria, to reduce electron leakage from the respiratory complexes and to improve ATP synthesis, maintains reduced glutathione levels within the mitochondria thereby enhancing the antioxidative potential. Moreover, melatonin was shown to influence autophagy, to stimulate the activity of telomerase, restoration of inner blood-retina barrier integrity, reduced VEGF and NO levels in the aged retina [2]. SkQ1 (10-(6-plastoquinonyl) decyltriphenyl-phosphonium) is plastoquinol derivative modified by a lipophilic cation then accumulate in the mitochondrial matrix and modulate of mitochondrial superoxide formation at specific sites of Complex I and III [3]. There are evidence that SkQ1 to slow down aging and to retard, arrest, and in some cases even reverse the development of many age-related eye diseases [4]. However, detailed effects of melatonin and SkQ1 on the biochemical mechanisms underlying therapeutic effect of these drug during retinal aging and AMD progression remain unclear.

Here, we analyzed the effects of melatonin and SkQ1 long-term treatment during aging and development AMD using by animal model. We used Wistar rats with normal aging process and senescence-accelerated OXYS rats, which spontaneously develop a phenotype similar to human age-related disorders including AMD-like retinopathy. The clinical signs of AMD-like retinopathy appear by the age of 3 months. Significant pathological changes detected at the age of 12 months and manifest themselves as excessive accumulation of lipofuscin and amyloid in the RPE regions, disturbances in the morphology of the RPE sheet, including an increase in the proportion of multinucleated cells, hypertrophy, distortion of cell shape, and reactive gliosis [5]. We explored the *in vivo* effects of melatonin and SkQ1 long-term treatment (6 months) in the retina of Wistar and OXYS rats during aging and AMD progression on the expression of autophagy-associated genes (*Nbr, Atg7, Becn1, Gabarapl1, LC3B, p62/SQSTM1*), mitochondria-associated genes (VDAC1 and mtTFA) and glutaminase that generates glutamate from glutamine.

Methods

To comparison of the effects of melatonin and SkQ1 long-term treatment on aged retina, 12-month-old OXYS and Wistar rats were randomly assigned to one of three groups (n = 10). The first group consumed a control diet, the second group supplemented with 0.04 mg of melatonin (Melaxen; Unipharm, New York, NY, USA) per kg of body and the third group supplemented with SkQ1 250 nmol per kg of body per day from the age of 12 to 18 months. The animals were provided with standard rodent feed and water *ad libitum*. SkQ1 was production by Institute of Mitoengineering of Moscow State University (Moscow, Russia).

We used standard techniques of Western-blot and qPCR to estimate effects of melatonin and SkQ1 on retina [5]. All the rats were examined by an ophthalmologist two times: before supplementation at the age of 12 months, and during the treatment at ages 18 months as decribed in [6]. Five days after the last eye examination, the rats were euthanized by CO_2 asphyxiation and decapitated. All animal procedures were

in compliance with the Association for Research in Vision and Ophthalmology statement for the Use of Animals in Ophthalmic and Vision Research and the European Communities Council Directive 86/609/EES.

Results

Here, we observed similar effects of melatonin and SkQ1 in the retina of Wistar rats. Preliminary examination showed that there was no difference between 12-month-old OXYS rats assigned to experimental and control groups. By the age of 22 months, retinopathy in control OXYS rats progressed. Both SkQ1 and melatonin decreased the incidence and severity of retinopathy in OXYS rats. In Wistar rats, which do not naturally develop retinopathy, repeated inspection did not reveal pathological alterations in the retina of melatonin and SkQ1-treated rats.

Protein level of VDAC1 was greater in retina of Wistar rats as compared OXYS rats. Treatment of melatonin and SkQ1 decreased VDAC1 level in Wistar rats and no influence on OXYS rats. VDAC1, also known as mitochondrial porin, associate with NADH oxidation and thus plays a role in cellular redox mechanisms. There are evidence that VDAC1 increases with age and may have negative effects by interrupting mitochondrial pore opening and closure. Also VDAC1 may interacts with Aβ, and phosphorylated tau may in turn block mitochondrial pores, leading to mitochondrial dysfunction [7]. Based on current study observations, we propose that reduced levels of VDAC1 by long-term treatment of melatonin and SkQ1 leading to normal mitochondrial function in Wistar rats.

The protein level of glutaminase was lower in Wistar rats as compared OXYS rats. Treatment of melatonin and SkQ1 increased protein level of glutaminase in Wistar rat's retina. In mammalian cells, glutamine is converted to glutamate by glutaminase (GLS), which functions as the rate-limiting enzyme, and then to α-ketoglutarate (α-KG). α-KG enters the tricarboxylic acid cycle, not only for the generation of ATP via oxidative phosphorylation, but also for the production of acetyl-coA as a critical precursor for the synthesis of lipids and nucleotides [8]. On other hand, glutamate is the main excitatory neurotransmitter in the retina, but it is neurotoxic when present in excessive amounts. On the basis of these results, it might be presumed that rat retinal glutaminase level is regulated by antioxidants.

We observed that treatment of SkQ1 dramatically decreased p62/SQSTM1 protein but not mRNA levels in both OXYS and Wistar rats as compared of control rats. p62/SQSTM1 is a scaffold protein for many signaling pathways, such as autophagy, apoptosis and inflammatory. Analyze of mRNA level of autophagy-dependent genes (*Nbr, Atg7, Becn1, Gabarapl1, LC3B*) demonstrated increasing mRNA level of *Beclin1* in OXYS rats after SkQ1 and melatonin treatment. The Beclin1 protein involved in initiation of autophagy. The treatment of SkQ1 is not change the protein level of Atg7 in the retina of OXYS and Wistar rats.

Summary, our data indicated that long-term treatment of melatonin and mitochondria-targeted antioxidant SkQ1 may be the different effects on molecular events resulting from different in genotype and environment.

Conclusion

Thus, using Wistar rats with normal aging process and senescence-accelerated OXYS rats, which spontaneously develop AMD-like retinopathy, we demonstrated that long-term treatment of melatonin and mitochondria-targeted antioxidant SkQ1 may retard an age-related decline in the adaptability of retinal cells and may be considered as a strategy to slow down AMD.

Acknowledgment

This study was supported by the Russian Foundation for Basic Research (18-015-00320) and by a Russian Scientific Foundation Grant (19-15-00044).

References

[1] Blasiak, J., Reiter, R. J., Kaarniranta, K. (2016). Melatonin in retinal physiology and pathology: the case of age-related macular degeneration. Oxidative medicine and cellular longevity, 2016. https://doi.org/10.1155/2016/6819736

[2] Rastmanesh, R. (2011). Potential of melatonin to treat or prevent age-related macular degeneration through stimulation of telomerase activity. Medical hypotheses, 76(1), 79-85. https://doi.org/10.1016/j.mehy.2010.08.036

[3] Ježek, J., Engstová, H., & Ježek, P. (2017). Antioxidant mechanism of mitochondria-targeted plastoquinone SkQ1 is suppressed in aglycemic HepG2 cells dependent on oxidative phosphorylation. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1858(9), 750-762. https://doi.org/10.1016/j.bbabio.2017.05.005

[4] Anisimov V. N. et al. (2011) Effects of the mitochondria-targeted antioxidant SkQ1 on lifespan of rodents. Aging (Albany NY), 3(11), 1110. https://doi.org/10.18632/aging.100404

[5] Kozhevnikova, O. S., Telegina, D. V., Tyumentsev, M. A., & Kolosova, N. G. (2019). Disruptions of Autophagy in the Rat Retina with Age During the Development of Age-Related-Macular-Degeneration-like Retinopathy. International Journal of Molecular Sciences, 20(19), 4804. https://doi.org/10.3390/ijms20194804

[6] Kolosova, N. G., et al. (2018). p62/SQSTM1 coding plasmid prevents age related macular degeneration in a rat model. Aging (Albany NY), 10(8), 2136.

[7] Manczak, M., Reddy, P. H. (2012). Abnormal interaction of VDAC1 with amyloid beta and phosphorylated tau causes mitochondrial dysfunction in Alzheimer's disease. Human molecular genetics, 21(23), 5131-5146. https://doi.org/10.1093/hmg/dds360

[8] Cai, W. F., Zhang, C., Wu, Y. Q., Zhuang, G., Ye, Z., Zhang, C. S., Lin, S. C. (2018). Glutaminase GLS1 senses glutamine availability in a non-enzymatic manner triggering mitochondrial fusion. Cell research, 28(8), 865-867. https://doi.org/10.1038/s41422-018-0057-z

DOI 10.18699/BGRS/SB-2020-407

Cellular response to UVA-B light depends on cellular age and chromatin structure

Bela Vasileva
Laboratory of Molecular Genetics
Institute of Molecular Biology "Acad. R. Tsanev", Bulgarian Academy of Sciences
Sofia, Bulgaria
belavas@outlook.com

Dessislava Staneva
Laboratory of Molecular Genetics
Institute of Molecular Biology "Acad. R. Tsanev", Bulgarian Academy of Sciences
Sofia, Bulgaria
dessysta@gmail.com

Plamen Zagorchev
Department of Physics and Biophysics, Faculty of Pharmacy, Medical University – Plovdiv, Bulgaria
Plovdiv, Bulgaria
plamenz@gbg.bg

Natalia Krasteva
Electroinduced and Adhesive Properties
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences
Sofia, Bulgaria
nataly@bio21.bas.bg

George Miloshev
Laboratory of Molecular Genetics
Institute of Molecular Biology "Acad. R. Tsanev", Bulgarian Academy of Sciences
Sofia, Bulgaria
karamolbiol@gmail.com

Milena Georgieva
Laboratory of Molecular Genetics
Institute of Molecular Biology "Acad. R. Tsanev", Bulgarian Academy of Sciences
Sofia, Bulgaria
milenageorgy@gmail.com

Abstract — **Complex interactions among DNA and nuclear proteins maintain genome organization and direct its stability. Both chromatin organization and stability are crucial for the way DNA is organized, compacted and preserved but also bring about intricate processes for its dynamic reorganization. As the fifth class of histone proteins, the linker histones interact with DNA and other histones and thus take major part in maintaining genome stability and nuclear organization. Our recent results have proved that *Saccharomyces cerevisiae* linker histone – Hho1p, physically interacts with the actin-related protein 4 (Arp4). The abrogation of this interaction through the deletion of the gene for the linker histone – *HHO1* in *arp4* mutant cells leads to global changes in chromatin compaction. Results show that the healthy interaction between the yeast linker histone and Arp4p is critical for maintaining genome stability and for controlling cellular sensitivity to different types of stress, particularly UVA/B stress. The double mutant yeast cells exhibit cellular characteristics of prematurely aged cells. They have completely abolished higher-order chromatin organization, die young and are subtler to different types of stress. This unambiguously proves the role of linker histones and chromatin remodelling in ageing by their cooperation which allows us to hypothesize that the linker histones are essential for maintaining higher-order chromatin compaction and thus preserving genome stability during ageing.**

Keywords — chromatin, ageing, linker histones, chromatin remodelling, Arp4, UVA/B stress, genome stability

Introduction

The molecule of DNA is compacted in the constrained nuclear compartment through binding with histone proteins which allows it to make an intricate nucleoprotein complex, known as chromatin. Chromatin is organized in nucleosomes which are composed of four pairs of the core histones – H2A, H2B, H3 and H4, around which 146 bp of DNA are wrapped (Luger et al., 1997). Outside of the nucleosome though binds a fifth class of histone proteins, recognized also as the family of linker histones or H1. The role of the linker histones is to stabilize the nucleosome by transforming it in a larger unit called chromatosome. A more important function of H1 histones is their engagement in the building and maintenance of the higher-order chromatin structures like the putative 30 nm fibres and structures above it corresponding to differently sized chromatin loops (Georgieva et al., 2012). As a reflection of its complexity chromatin is involved in processes like DNA replication, chromosome segregation, DNA repair and others. To fulfil these functions, it goes through various types of epigenetic modifications like DNA methylation, histone modifications and non-coding RNA binding (Allis et al., 2007; Georgieva et al., 2016). Chromatin secondary structures have great impact on cellular fate. Structural changes in the way the genome is organized at these higher levels of chromatin compaction govern and moderate gene expression and by this modulate cellular response to intrinsic and extrinsic signals. This determines the way cellular programs are performed and also controls cellular normality. Many diseases and other pathologies like cancer, ageing, age-associated diseases and others have been linked with alterations in the structural organization of chromatin and its epigenetic make-up. Although these changes have been linked with aberrant cellular programs little yet is known about the molecular mechanisms underlying cellular adaptability to different types of stress. Here, we show our recent results which demonstrate the role of two chromatin players in the process of cellular ageing and stress resilience.

Experimental

Yeast strains

Wild type (WT) - MATa his4-912 ade2 his4-912 lys2-128 can1 trp1 ura3 ACT3

hho1Δ (in the text appears as hho1delta) - MATa his4-912 ade2 his4-912 lys2-128 can1 trp1 ura3 act3 ypl127C::K.L.URA3

arp4ts26 (designated in the text as arp4) - MATa his4-912 ade2 lys2-128 can1 leu2 trp1 ura3 act3-ts26

arp4ts26Δhho1 (in text denoted as arp4 hho1delta) - MATa his4-912 ade2 lys2-128 can1 leu2 trp1 ura3 act3-ts26 ypl127C::K.L.URA3

CLS of S. cerevisiae

The CLS of *S. cerevisiae* was examined according to (Longo and Fabrizio 2012, Uzunova et al. 2013).

UVA/B irradiation of cells and Colony Forming Units (CFU) assays

Yeast cells were cultivated in minimal medium and probes of them were taken at the 4th, 24th, 72nd hours and 9th day of cultivation. One hundred cells were spread on each Petri dish. These were then irradiated with UVA/B light for 3 or 30 minutes. The used source of UVA/B was a 15W Cleo lamp, with the help of which an energy dose of 2.56 mW/cm2 was obtained. The irradiation was followed by a 2-day recovery period at 30°C on a rich YPD plate. After this time the number of formed colonies was counted and the percentage of viable cells was estimated with a respect to the non-irradiated control.

Fluorescence Activated Cell Sorting (FACS) analysis

FACS analysis was performed according to (Georgieva, et al. 2012). Several changes were made: cells were fixed with 100% ethanol; sonication lasted for 20 seconds at 50% power; cells were incubated with RNase A (0.25 mg/ml), after which they were centrifuged for 4min/5000g. This was followed by a pre-staining with propidium iodide and probing at a BD FACSCanto apparatus. Quantification of the results was made by a FlowJoV10 software.

Gene expression analyses

Aliquots were taken and total RNA was isolated. Briefly, cells were washed with 1 ml of ice-cold RNase-free water and collected by centrifugation at 1500 g. Cell pellet was resuspended in 400 µl of TES solution (10 mM Tris, pH 7.5, 10 mM EDTA, pH 7.5 and 0.5% SDS) followed by addition of 400 µl of pre-warmed at 65°C acid phenol and incubation for 60 min at 65°C with 10 sec vortexing at every 15 min. After centrifugation for 15 min at 10 000 g the purification of nucleic acids continued with three subsequent extractions with 400 µl of hot acid phenol each and a final extraction with 400 µl chloroform. Nucleic acids were precipitated at –20 °C ON with three volumes of cold 100% ethanol and 1/10 volumes of 3M sodium acetate, pH 5.2. Total RNA was then washed with 70% ethanol, dried and dissolved in RNase-free water. The concentration and purity of the isolated total RNA was determined using NanoDrop® ND-1000 spectrophotometer.

The level of the *ScACT1*, *ScCDC28* and *ScRAD9* mRNAs was analyzed by real - time reverse transcriptase PCR (RT-qPCR). 1 µg of DNase I-treated total RNA was reverse transcribed into cDNA using oligo-dT18 primer and RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas) according to manufacturer's instructions.

Gene expression analyses were performed with Maxima® SYBR Green qPCR Master Mix (Fermentas) on Rotor Gene 6000 (Corbett LifeScience) following manufacturers' protocols. All primer pairs were designed through SGD Web Primer design tool. qPCR was performed in duplicate, using 1/40th (25 ng) of the cDNA template and 0.2 µM each of the forward and the reverse gene-specific primer in a 20 µL reaction mixture. Cycling conditions used for all primer sets consisted of an initial denaturation at 95 °C for 10 min and 45 cycles of: 95 °C for 15 s, 60 °C for 60 sec. The specificity of the resulting amplicons was confirmed by melting curve analysis and agarose gel electrophoresis.

The expression of each gene was normalized to the reference gene ScACT1. Relative quantification expression levels were calculated by comparative CT method via Rotor-Gene Q Series Software 1.7.

Results and Discussion

The proliferative potential of the yeast mutant cells was assessed by FACS analysis and was compared to the control wild type cells. Cell aliquots were taken at the four time points: day 1, day 2, day 3 and day 9. Percentage of cells in different phases of the cell cycle has been calculated by means of specialized software FlowJo V10. The results from these experiments showed that *arp4 hho1delta* double mutant cells are short-lived with impaired proliferation potential. Their vacuolar morphology was changed and resembled the morphology of prematurely aged cells. Cells from control cultures and cells exposed to replicative, cold temperature and oxidative stress were plated on solid rich media and were cultured at 30°C for three days. CFUs (colony forming units) were counted. Results showed that the double mutant cells could not survive the applied stress conditions, thus suggesting increased stress sensitivity and lack of resilience.

conclusions

Hho1p interaction with Arp4p is crucial for yeast chronological lifespan. Abrogation of the interaction between the yeast linker histone Hho1p and Arp4p leads to premature ageing phenotypes in the mutant cells. The cells without the linker histone and bearing a mutation in *ARP4* die young. These chromatin mutant cells possess changed vacuolar morphology at early stages of their lifespan. The lack of proper chromatin structure in the mutant cells make them more prone to different types of stress.

ACKNOWLEDGMENT

The work is sponsored by the Bulgarian Science Fund Grant number: DN 11/15 and NATO Science for Peace and Security programme [grant number: NATO SPS MYP G5266].

REFERENCES

[1] Longo and Fabrizio 2012
[2] Uzunova, K., et al. 2013
[3] Georgieva, M., et al. 2012
[4] Luger, K., et al., 1997
[5] Allis, D., et al., 2007
[6] Georgieva, M. et al., 2016

DOI 10.18699/BGRS/SB-2020-416

Are younger people sleepier than older people after missing bedtime and night sleep? It depends…

Arcady Putilov
Research Group for Math-Modeling of Biomedical Systems
Research Institute for Molecular Biology and Biophysics of the Federal Research Centre for Fundamental and Translational Medicine
Novosibirsk, Russia
0000-0003-2779-9046

Olga Donskaya
Research Group for Math-Modeling of Biomedical Systems
Research Institute for Molecular Biology and Biophysics of the Federal Research Centre for Fundamental and Translational Medicine
Novosibirsk, Russia
dolly@niimbb.ru

***Abstract* — Feeling sleepy might evolve for motivating humans and apes to initiate sleep-preparatory behaviors prior to their habitual bedtime. Sleep deprivation experiments consistently revealed that younger rather than older people are more vulnerable to sleep loss. If so, why don't they go to sleep earlier, become sleepier during the day and nap more frequently than older people? To address such questions, we obtained subjective and objective sleepiness measurements during prolongation of wakefulness of 130 and 48 volunteers to the next morning and to the next two days after today, respectively (ages ranged from 15 to 67 years). The results suggested that the answer to a question of whether sleep loss makes younger people sleepier than older people depends upon the method of sleepiness measurement, either objective (e.g., alpha rhythm attenuation in eyes closed condition) or subjective (e.g., alertness-sleepiness self-scoring). We concluded that the motivational function of sleepiness remains preserved in older adults despite a lowered responsiveness of their electroencephalogram to extension of wakefulness beyond its normal duration.**

Keywords* — *EEG, alpha attenuation, aging, sleep loss, drowsiness, alertness

INTRODUCTION

Human aging is associated with notable changes in sleep–wake behaviors. Older adults are often rising early, feeling sleepy during the day, taking a nap, going to bed early, having difficulty falling asleep, and suffering from frequent night sleep interruptions and early morning awakenings. However, studies of age-associated differences in response to night sleep deprivation consistently revealed that older adults are more resistant to impairment of alertness and performance than younger adults (reviewed with suggesting an explanation in [1, 2]). Subjective sensation of sleepiness prior to habitual bedtime might evolve for motivating humans and apes to switch to sleep-preparatory behaviors from any other current daytime activities [3]. If younger adults become sleepier than older adults during prolongation of wakefulness, why don't they go to sleep earlier, become sleepier during the day and nap more frequently?

We addressed such questions by extending a period of wakefulness of people aged between 15 and 67 years to examine the effect of such an extension on subjective and objective measures of their sleepiness. The following two hypotheses were tested:

- An analysis of objective measure of sleepiness would provide further evidence for a reduced vulnerability of older adults to sleep deprivation;
- in contrast, an analysis of subjective measure of sleepiness would reveal age independence of the strength of feeling sleepy due to sleep loss.

METHODS

Stady Participants and Study Protocols

The main electroencephalographic (EEG) dataset was collected in two-day sleep deprivation experiments with 48 study participants (27 females) aged between 15 and 67 years (mean age ± standard deviation were 36.6 ± 13.1 years). The participants were grouped into three ages of 15-18 participants each (younger, 15-26, intermediate, 30-40, and older, 46-67 years). They arrived in the research unit on Friday before 18:30 to leave it on Sunday at 19:30 with an expectation to complete 25 resting EEG recording sessions separated by 2-h intervals (2 and 5 min with eyes open and closed, respectively).

The additional EEG dataset was collected in a shorter (one-night) sleep deprivation experiments with 130 study participants (76 females) 15-67 year old (mean age ± standard deviation were 29.4 ± 12.2 years). The participants were grouped in four ages 30-37 participants each (youngest, 15-21, younger, 22-24, intermediate, 25-39, and older, 40-67 years). They arrived in the research unit between 8:00 and 8:30 to leave it at 9:30 next day after completing 9 sessions of resting EEG recordings separated by 3-hr intervals (one min with eyes open and one min with eyes closed).

The experimental sleep deprivation studies were performed in accordance with the ethical standards laid down in the Declaration of Helsinki. Their protocols were approved by the Ethics Committee of the Institute of Molecular Biology and Biophysics. Informed written consent was obtained from each participant studied as paid volunteer.

Assessments of Sleepiness

Frontal and occipital electrodes were used for all EEG recordings in accord with the International 10-20 system of electrode placement. The EEG signals were recorded via a 16-channel electroencephalograph (Neuron-Spectrum-2, Neurosoft, Ivanovo, Russia), conditioned by the high-pass, low-pass and notch filters (0.5 Hz, 35 Hz and 50 Hz, respectively), sampled and stored on a hard disc with a frequency of 200 Hz.

The EEG records were visually inspected on 2-s epochs for removing from further analysis all epochs containing artifacts. The FFTW (Fastest Fourier Transform in the West) package was applied to compute power spectra for the artifact-free epochs [4]. Rectangular window taper was used on 1-s epochs without overlap for calculating absolute spectral power densities (μV^2) for each of 16 first single-Hz frequency bandwidth. These single-Hz power densities were averaged on each of one-min intervals of EEG record and ln-transformed. These powers were further averaged over derivations, min with eyes open and closed, and, additionally, over 4-Hz frequency ranges. Each value for 13-25 sessions of each of 48 participants of the two-day deprivation study was subtracted from the first (evening) value (at 19:00). In additional dataset,

single-Hz powers were similarly averaged (over derivations and 4-Hz frequency ranges). Each value for the last 5 sessions of each of 130 participants (from 21:00 to 9:00) was subtracted from an evening value (21:00). Similarly, KSS score self-reported during each EEG recording session by using the 9-step Karolinska Sleepiness Scale [5] (a measure of subjective sleepiness) was related to a score self-reported at either 19:00 or 21:00 by either 48 or 130 participants of either two-day or one-night sleep deprivation study, respectively.

Statistical Analyses

Calculations were performed using the $SPSS_{23.0}$ statistical software package (IBM, Armonk, NY, USA). To correlate time courses of objective and subjective sleepiness on the interval of 25 EEG recording sessions and to relate age of study participant to subjective and objective sleepiness measures (Table I), Pearson's coefficient of correlation (*r*) was applied.

TABLE I. ALPHA POWER AS A CORRELATE OF AGE

Clock hour	Day 1		Day 2		Clock hour	Night 1	
	n	*r*	*n*	*r*		*n*	*r*
21	48	.365*	42	.489***	-	-	-
23	48	.280^	42	.564***	-	-	-
1	48	.338*	41	.419**	-	-	-
3	48	.400**	39	.428**	24	130	.028
5	48	.505***	37	.347*	3	130	.204*
7	48	.387**	36	.493**	6	130	.245**
9	48	.326*	33	.361*	9	130	.316***
11	48	.360*	29	.500**	-	-	-
13	48	.380**	27	.490**	-	-	-
15	48	.321*	27	.390*	-	-	-
17	48	.312*	25	.217	-	-	-
19	48	.438**	25	.478*	-	-	-

Notes. Objective sleepiness measure was alpha power related to either 19:00 or 21:00 in either two-day or one-night sleep deprivation study; n: Number of study participants (it is gradually decreasing throughout Day 2 due to dropouts); r: Pearson coefficient of correlation between age of study participant and alpha power; ^ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Moreover, two-way repeated measure ANOVA (rANOVA) was run to test significance of main effect of independent factor "Age" (Table II). The repeated measure was clock hour of the day 1 and day 2 (between 21:00 and 19:00) and for the night 1 (from 24:00 to 9:00).

TABLE II. MAIN EFFECT OF INDEPENDENT FACTOR "AGE"

Sleepiness index	Day 1		Day 2		Night 1	
	df	*F*	*df*	*F*	*df*	*F*
Alpha power	2/45	5.319**	2/22	4.960*	3/126	3.145*
KSS score	2/45	0.479	2/22	0.639	3/126	0.769

Notes. Objective and subjective sleepiness measures were Alpha power and KSS score related to either 19:00 or 21:00 in either two-day or one-night sleep deprivation study; df: Degree of freedom (it is smaller on Day 2 than on Day 1 due to, in total, 23 dropouts); F: F-ratio from two-way rANOVAs (the repeated measure was clock hour) for the independent factor "Age" (either three or four ages – see Methods); * $p < 0.05$, ** $p < 0.01$.

RESULTS

The strongest of the correlations between the time course of KSS score and time courses of 16 single-Hz spectral powers was found for 10Hz power at the 2nd min with eyes closed. It attained a value of <-0.980. Similarly, such strong correlations were obtained after averaging over the first two or five min with eyes closed and over alpha frequency range (9Hz-12Hz). Therefore, the alpha power for the first two and the 1st min within eyes closed section of the record was chosen for relating objective sleepiness with age of study participant in the two-day and one-night study, respectively. KSS score did not show a consistent pattern of correlation with age, while alpha power was found to be a significant correlate of age in almost every EEG recording session (Table I). This relationship suggesting a weaker attenuation of alpha rhythm in older participants was confirmed by two-way rANOVAs yielding significant main effect of independent factor "Age" (Table II).

DISCUSSION

Despite observable aging-related changes in the sleep-wake pattern, such as going to bed early, napping and feeling sleepy during the day, etc., the experimental studies of age-associated differences in the response of alertness and performance measures to sleep loss consistently pointed at younger rather than older people as being highly susceptible to impairment of their alertness and performance levels during extension of wakefulness beyond its normal duration. Such contradictive facts motivated us to compare younger and older people on objective and subjective sleepiness measures in sleep deprivation experiments. The results provided support for each of two hypotheses. 1. Older adults had a reduced vulnerability to sleep loss as indicated by the results of analysis of objective measures of sleepiness. 2. In contrast, the results of analysis of the subjective sleepiness assessments failed to reveal significant difference between ages in the strength of feeling sleepy. Therefore, the answer to such a question as are younger adults become sleepier than older adults due to sleep loss depends upon whether the sleepiness was evaluated with objective or subjective method. The result supporting the second hypothesis pointed on the preserving motivational function of sleepiness in older adults despite the age-associated decline of their EEG responsiveness to sleep deprivation.

ACKNOWLEDGMENT

The experimental studies were partially funded by the Russian Foundation for Basic Research through a research grant number 19-013-00424.

REFERENCES

[1] H. P. Landolt, J. V. Rétey and M. Adam, "Reduced neurobehavioral impairment from sleep deprivation in older adults: contribution of adenosinergic mechanisms," Front Neurol., vol. 3:62, April 2012.

[2] A. A. Putilov and O. G. Donskaya, "Evidence for age-associated disinhibition of the wake drive provided by scoring principal components of the resting EEG spectrum in sleep provoking conditions," Chronobiol. Int., vol. 33, pp. 995-1008, 2016.

[3] J. Axelsson, M. Ingre, G. Kecklund, M. Lekander, K. P. Wright and T. Sundelin, "Sleepiness as motivation: a potential mechanism for how sleep deprivation affects behavior", Sleep [online, 2019 Nov 29].

[4] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3", Proc IEEE 93, pp. 216–231, 2005.

[5] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," Int. J. Neurosci., Vol. 52, pp. 29-37, 1990.

DOI 10.18699/BGRS/SB-2020-415

Author index

Contents

Электронное научное издание

BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE/SYSTEMS BIOLOGY (BGRS/SB-2020)

The Twelfth International Multiconference Abstracts

Printed without editing

БИОИНФОРМАТИКА РЕГУЛЯЦИИ И СТРУКТУРЫ ГЕНОМА/СИСТЕМНАЯ БИОЛОГИЯ (BGRS/SB-2020)

Двенадцатая международная мультиконференция

Тезисы докладов

Публикуется в авторской редакции

Выпуск подготовлен информационно-издательским отделом ИЦиГ СО РАН

Подписано в печать 01.07.2020. Формат 60×84 1/8. Усл. печ. л. 79.3

Федеральный исследовательский центр
«Институт цитологии и генетики Сибирского отделения Российской академии наук»
630090, Новосибирск, проспект Академика Лаврентьева, 10