Whole-exome Sequencing Association Studies On Impaired Spermatogenesis In Different Ethnic Groups In Russia

Semyon K. Kolmykov[1], Gennadiy V.Vasiliev[1], Mikhail P. Ponomarenko[1], Maxim A. Kleshev[1], Aleksandr V. Osadchuk[1], Ludmila V. Osadchuk[1]

1. Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

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.

I. 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 impaired spermatogenesis.

II. Materials And Methods

a. Study sample

The study sample included 62 participants from 3 ethnic groups Slavs, Yakuts, and Buryats (see Tab. 1).

TABLE1: 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

b. 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.

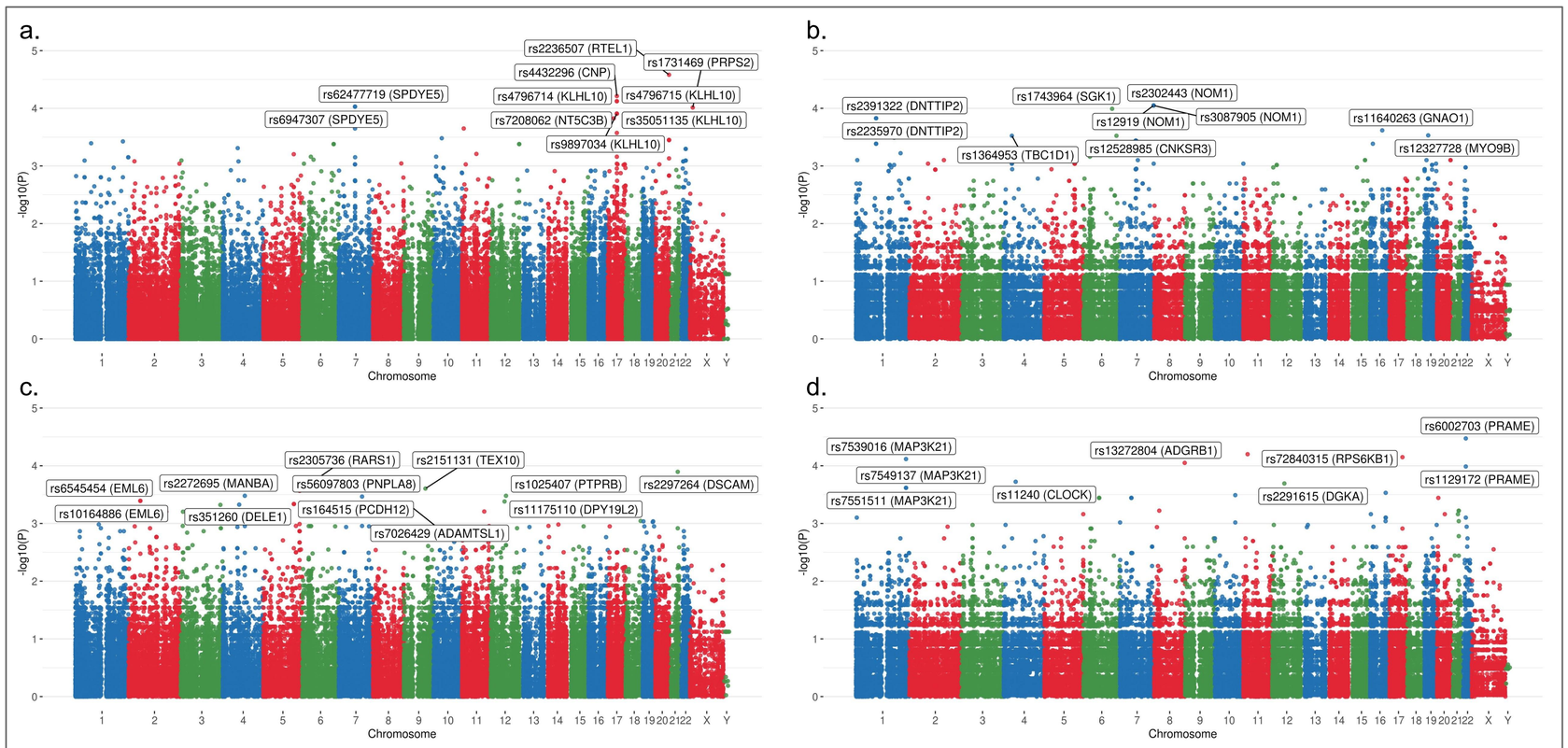

Figure 1. Graphical summary of the results of the association analysis of 62 samples. Plot of –log10(P-values) of chi^2-tests for SNPs that have passed quality control filters in (a) the joint set, (b) the Buryats, (c) Slavs and (d) Yakuts groups.

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 joint data set (62 patients: 32 patients with normospermia and 30 patients with patospermia).

c. Association analysis

An analysis of the association of identified polymorphisms with spermatogenesis defects was performed using χ2-test using PLINK [5].

III. Results

Figure 1 illustrates the results of association analyses. 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 the Slavs, Yakuts, Buryats and joint groups respectively.

The top-20 lists of polymorphisms associated with impaired spermatogenesis were analysed. In particular, the presence of an association of polymorphism in *KLHL10, CNP and PRPS2* genes in a joint group has been shown. An association of other polymorphisms in gene *KLHL10* 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 № 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.