Functional study of potential regulatory SNPs (rs590352, rs11542583, rs3829202, rs78317230, rs2072580, rs4796672)

A.Degtyareva, E.Leberfarb, I.Brusentsov, T.Kuzina, T.Merkulova

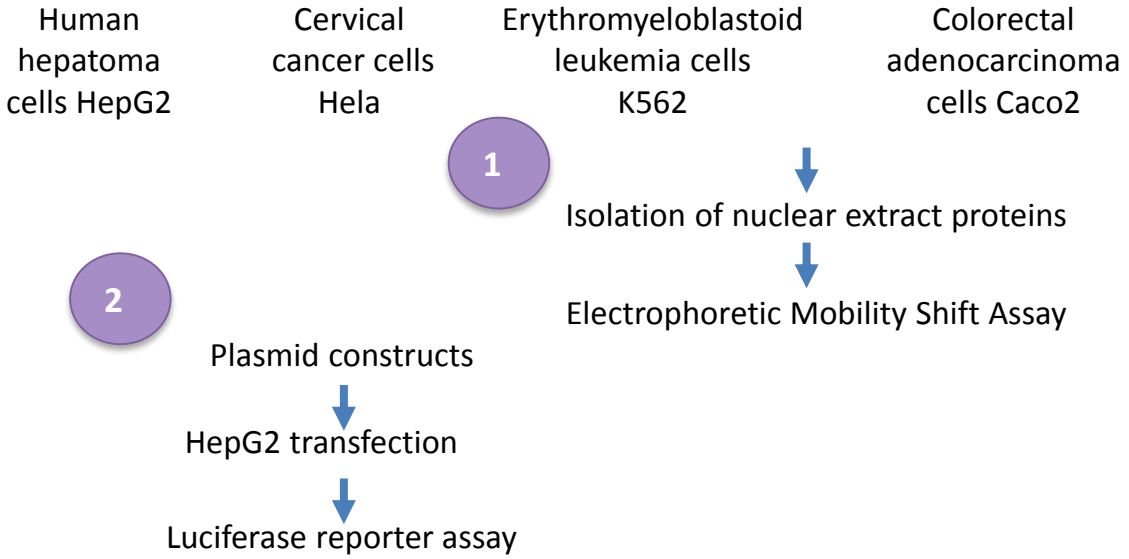
Introduction

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, GWAS does not provide information about the functionality of these variants.

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.

The aim 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 Electrophoretic Mobility Shift Assay (EMSA) and luciferase reporter system.

Materials and methods



Results

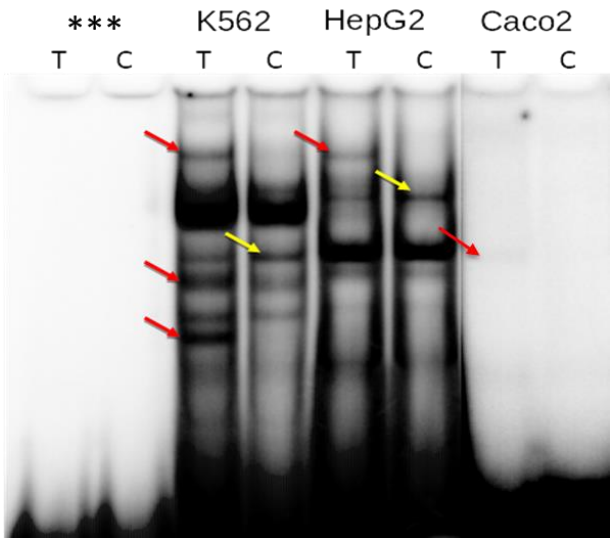
Using EMSA, it was shown that all nucleotide substitutions G>C (rs590352), A>G (rs11542583), (T>C) rs3829202, T>C (rs78317230), A>T (rs2072580), C>T (rs4796672) change the pattern of nuclear extract proteins binding.

EMSA example for rs78317230

A) 5′-ATGGCATGCGTTCCG**A**CACAACTTCGGACAG-3′
3′-TACCGTACGCAAGGC**T**GTGTTGAAGCCTGTC-5′

B) 5′-GCCAGGGGGCAGTAA**C**GCCAGATCATCTGTC-3′
3′-CGGTCCCCCGTCATT**G**CGGTCTAGTAGACAG-5′

C)



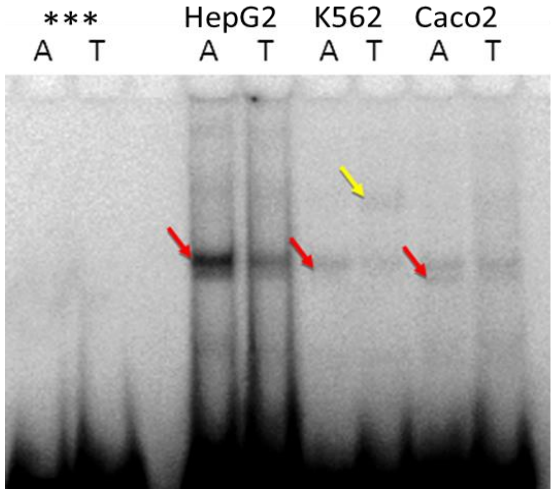
A) nucleotide sequence of DNA-probe containing allele T rs78317230;
B) nucleotide sequence of DNA-probe containing allele C rs78317230;
C) EMSA results: *** – free DNA-probes; red arrows – intensification/ appearance of bands in the case of T allele; yellow arrows – intensification of bands in the presence of C allele

EMSA example for rs2072580

A) 5′-GCCAGGGGGCAGTAA**T**GCCAGATCATCTGTC-3′
3′-CGGTCCCCCGTCATT**A**CGGTCTAGTAGACAG-5′

B) 5′-ATGGCATGCGTTCCG**T**CACAACTTCGGACAG-3′
3′-TACCGTACGCAAGGC**A**GTGTTGAAGCCTGTC-5′

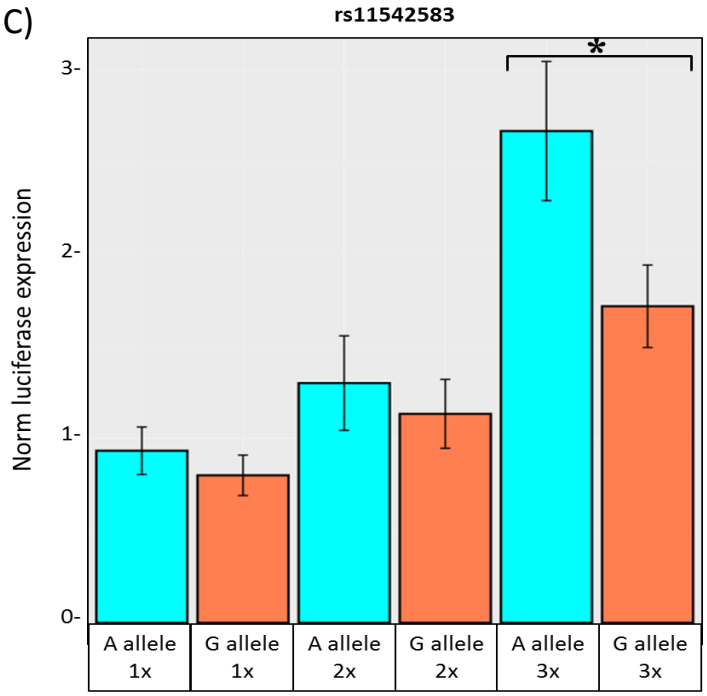
C)



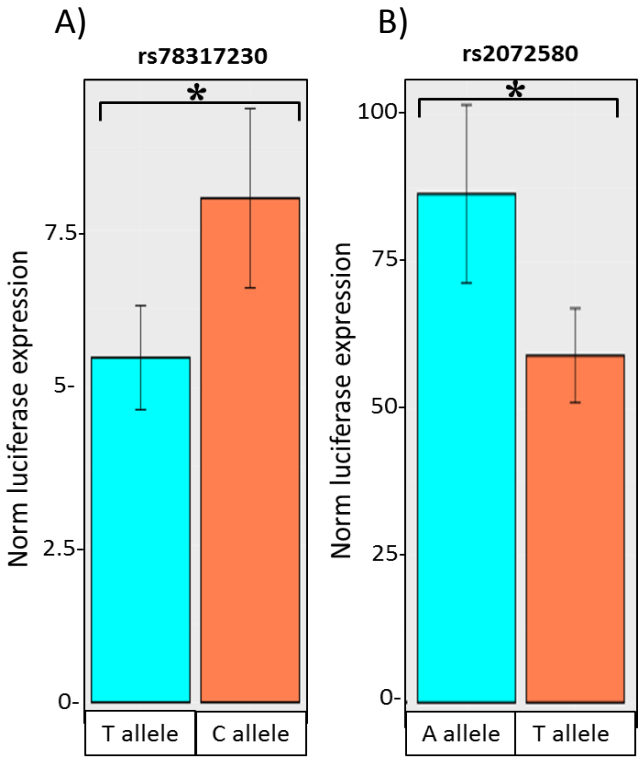
A) nucleotide sequence of DNA-probe containing allele A rs2072580;
B) nucleotide sequence of DNA-probe containing allele T rs2072580;
C) EMSA results: *** – free DNA-probes; red arrows – intensification/ appearance of bands in the case of A allele; yellow arrows – intensification of bands in the presence 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.



Examples



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. According to literature, *ATXN7L3* (rs590352), *COG8* (rs11542583), *KLF6* (rs3829202), *KRT15 (*rs4796672), *U2AF2* (rs78317230), *SART3* (rs2072580) encode proteins that perform important functions, therefore expression change of these genes can lead to serious consequences.