DOI 10.18699/SBB-2022-260

Genome-Wide detection of human regulatory SNPs based on analysis of ChIP-seq and RNA-seq complex data

Damarov I.[1]*, Merkulova T.[1, 2], Bryzgalov L.[1], Ustrokhanova D.[2]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *damarovis@bionet.nsc.ru*

Key words: regulatory SNP, ChIP-seq, GWAS

Motivation and Aim: According to modern concepts, non-coding SNPs, including regulatory (rSNPs), involved in the control of gene expression, make the greatest contribution to the formation of the phenotype, accounting for more than 90 % of the SNPs identified in GWAS [1]. However, GWAS does not provide information about the functionality of the found variants, thus preventing progress in understanding the mechanisms underlying the development of certain diseases. Therefore, functional genomics data are widely used for the functional interpretation of the found variants. These same approaches are applied independently, regardless of the identified regulatory variants in the GWAS. The aim of this study is a genome-wide search for human regulatory SNPs based on the analysis of ChIP-seq and RNA-seq data.
Methods: To search for rSNPs, we used an approach based on the identification of allele-specific events in experimental data to determine the profile of histone modification of open chromatin H3K4me3 (ChIP-seq) and transcriptome (RNA-seq) obtained for peripheral blood lymphocytes from 20 donors. To determine the functional significance of potentially regulatory SNPs using GTEx (The Genotype-Tissue Expression) data [2], a random forest algorithm was developed, which, based on the specified parameters, is able to effectively classify the SNP according to whether it is eQTL / is not eQTL.
Results: ChIP-seq and RNA-seq libraries were prepared and sequenced, averaging 20 and 73 million pair-end reads per sample respectively. As a result of mapping data to the reference genome, a list of rSNPs was formed, demonstrating an allelic imbalance in the data on the representation of H3K4me3 in the promoter regions and the expression of the same genes. For each rSNP from this list, the probability of belonging to eQTL was predicted. Several rSNPs classified as eQTLs significantly change the sequence of transcription factor binding sites, which confirms their regulatory function. Connecting GWAS data provided the association of several identified eQTLs with phenotypic traits.
Conclusion: As a result of the analysis of the obtained experimental data, a panel of about 2 thousand rSNPs was formed, which with a high probability can be eQTL.

References

1. GWAS Catalog [Electronic resource]. URL: https://www.ebi.ac.uk/gwas/home. Accessed 18.03.2022.
2. Genotype-Tissue Expression Project (GTEx) [Electronic resource]. URL: https://www.genome.gov/Funded-Programs-Projects/Genotype-Tissue-Expression-Project. Accessed 18.03.2022.