Секция «Биоинформатика»

Functional annotation of allele-specific binding sites of human transcription factors

Научный руководитель – Кулаковский Иван Владимирович

Быкова Дарья Ильинична
Студент (специалист)
Lomonosov Moscow State University, Факультет биоинженерии и биоинформатики, Москва, Россия
E-mail: db.bykova@gmail.com

Many single nucleotide polymorphisms (SNPs) are located in the regulatory regions of human genome. These variants can affect gene expression and phenotypic traits, particularly, through alterations in transcription factor (TF) binding. Analysis of ChIP-Seq data allows to identify allele-specific binding (ASB) events, the heterozygous sites with TF binding affinity strongly depending on a particular allele. In a joint project of VIGG RAS and EIMB RAS, thousands of ASBs were identified from ChIP-Seq data available in the GTRD database [1]. In this project, we annotated the detected ASB and identified particular cases where functionally relevant TF binding alterations were associated with pathogenic phenotypes.

First, we annotated ASB locations in regards to different genomic segments and observed the enrichment of ASBs in promoters and FANTOM5 enhancers. Second, we compared the ASBs with those already reported in existing databases, including the allele-specific behavior database AlleleDB [2], allele-specific DNA accessibility data [3], ASBs observed in Wenqiang Shi's work [4] and reporter assay quantitative trait loci (raQTLs) data [5] and found a notable overlap, that was increasing with increased ASB significance. Third, we inspected particular ASBs and found several cases with strong links between the respective SNP, TF, and a pathogenic phenotype.

Источники и литература

1) Ivan Yevshin, Ruslan Sharipov, Semyon Kolmykov, Yury Kondrakhin, Fedor Kolpakov. GTRD: a database on gene transcription regulation-2019 update. Nucleic acids research 47 (2018). https://doi.org/10.1093/nar/gky1128
2) Chen, J., Rozowsky, J., Galeev, T. et al. A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. Nat Commun 7, 11101 (2016). https://doi.org/10.1038/ncomms11101
3) Maurano, M., Haugen, E., Sandstrom, R. et al. Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. Nat Genet 47, 1393–1401 (2015). https://doi.org/10.1038/ng.3432
4) Wenqiang Shi, Oriol Fornes, Anthony Mathelier, Wyeth W. Wasserman, Evaluating the impact of single nucleotide variants on transcription factor binding, Nucleic Acids Research 44, 10106–10116 (2016). https://doi.org/10.1093/nar/gkw691
5) van Arensbergen, J., Pagie, L., FitzPatrick, V.D. et al. High-throughput identification of human SNPs affecting regulatory element activity. Nat Genet 51, 1160–1169 (2019). https://doi.org/10.1038/s41588-019-0455-2