

Extended clusters of transcription factor binding sites in embryonic stem cells

A.V. Tsukanov^{1*}, A.I. Dergilev^{2,3}, Y.L. Orlov^{1,3}

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

² *University Grenoble-Alps, Grenoble, France*

³ *Novosibirsk State University, Novosibirsk, Russia*

* *e-mail: ya.cukanton@yandex.ru*

Key words: genomics, embryonic stem cells, transcription factors, binding sites, ChIP-seq, gene expression regulation

Motivation and Aim: Computer study of transcription regulation presents important problem in molecular biology challenging growing volume of sequencing data. ChIP-Seq allows to detect interactions between DNA and protein factors regulation in genome-wide scale [1]. Additionally, ChIP-seq data with RNA-seq data allows construct transcriptional regulatory network. It can help us better understand the molecular mechanisms of organism development and cellular reprogramming, which is very important for generating various cell types for regenerative therapies [2]. Thus, it is critical to annotate clusters of binding sites of different transcription factors in distal region of genome that may function as enhancers.

Methods and Algorithms: We used own R scripts and some R packages from Bioconductor (<http://www.bioconductor.org/>) for TFBS analysis. The ChIP-seq data of H1 stem cell was downloaded from Cistrome database (<http://cistrome.org/db/#/>).

Results: We have developed set of scripts with using packages from Bioconductor and CRAN (GenomicRanges, AnnotationHub, fastcluster, ggplot2) for transcription factor binding sites analysis. It implements methods for processing, analyzing, and visualizing ChIP-seq data. In particular, next functions are: 1) calculating enrichment of peaks across the genome; 2) calculating distribution density of binding peaks near transcription start site (TSS); 3) generating random peaks; 4) hierarchical clustering of binding peaks. The clusters of transcription factors were obtained. It was shown that clusters contained four transcription factors and more can't be obtained by chance. We have generated the heatmap of co-association between transcription factors, and distribution of binding peaks near gene TSS. We have shown that transcription factors mostly enrich the regions near TSS in embryonic stem cells.

Conclusion: Construction of clusters of transcription factor binding sites with RNA-seq data allowed reconstruct gene regulatory networks. We continue work on integration of ChIP-seq transcription factor binding sites data using available data sources.

Acknowledgements: The research has been supported by RFBR. Computing at Siberian Supercomputer center SB RAS was supported by budget project 0324-2018-0017.

References

1. Chen X., Bhadauria V., Ma B. (2018) ChIP-Seq: A powerful tool for studying protein-DNA interactions in plants. *Curr. Iss. Mol. Biol.* 27:171-180.
2. Bing H., Kai T. (2016) Understanding transcriptional regulatory networks using computational models. *Curr. Opin. Genet. Dev.* 37:101-108.