

Bioinformatics tools for 3D chromosome contacts analysis

A.V. Tsukanov¹, A.O. Bragin¹, O. Thierry^{2,3}, Y.L. Orlov^{1,3*}

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

²*University of Bordeaux, Paris, France*

³*Novosibirsk State University, Novosibirsk, Russia*

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

Key words: genomics, 3D genome structure, gene expression regulation, transcription factor binding sites, Hi-C, ChIA-PET

Motivation and Aim: Transcription regulation in eukaryotes is a complex process, in which chromatin interactions play a critical role for gene expression regulation as well as to further influence other cellular activities. Series of post-genome technologies have been developed to study the binding of transcription factors (TF) for transcription regulation, such as chromatin immunoprecipitation arrays (ChIP-chip), ChIP-Seq [1]. Another challenge is to define whether such binding sites distal from gene regions are functional, i. e. physically contact target gene promoters via chromosome loops or attracting RNA polymerase II complex for gene transcription. Identification of genome-wide distal chromatin interactions that lead the regulatory elements to their target genes may provide novel insights into the study of transcription regulation. Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) method for such analysis requires development of specialized software that will be reviewed in our presentation [2].

Methods and Algorithms: The aim of the work was to review existing tools for 3D genome structure develop a computer program for statistical data analysis and test it on CTCF binding sites, genes and spatial topological domains. These data have been obtained experimentally by using methods ChIP-seq, Hi-C, ChIA-PET. Gene annotation was obtained from UCSC Genome Browser (<http://genome.ucsc.edu>).

Results: The result of the program is a distribution of CTCF transcription factor binding sites on domains on the human chromosomes. The distributions of human genes relative CTCF binding sites and a randomly generated list of such sites as the program output were used to estimate statistical significance of the associations found.

Conclusion: With the rapidly increasing resolution of Hi-C datasets, the size of the chromatin contact map will soon exceed the memory capacity of general computers. The same problem related to ChIA-PET and subsequent data integration to be solved by our software development.

Acknowledgements: The authors are grateful to I.V. Chadaeva for science discussion. The research has been supported by RFBR (grant 18-34-00496) and Computing was done at Siberian Supercomputer center SB RAS (SSCC) with support by ICG SB RAS budget project (0324-2018-0017).

References

1. Li G. et al. (2014) Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. BMC Genomics. 15(Suppl. 12):S11.
2. Orlov Y.L. et al. (2017) Computer methods of analysis of chromosome contacts in the cell nucleus based on sequencing technology data. Biomed Khim. 63(5):418-422 (in Russian).