DOI 10.18699/WIBSB-2018-48

Bioinformatics tools for 3D chromosome contacts analysis

O. Thierry[1, 2], A.G. Galieva[2], A.V. Tsukanov[2, 3], Y.L. Orlov[2, 3]*
[1] *University of Bordeaux, France*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** e-mail: orlov@bionet.nsc.ru*

Key words: genomics, sequencing technologies, gene expression regulation, 3D genome structure, Hi-C, ChIA-PET, database

Motivation and Aim: Chromatin interactions play a critical role for gene expression regulation. Series of post-genome technologies have been developed to study the binding of transcription factors for transcription regulation, such as chromatin immunoprecipitation arrays (ChIP-Seq) [1]. Correspondingly, set of software tool for processing of such data has been developed. Another challenge is to detect functional contacts of 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 [2]. 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 [3].
Methods and Algorithms: The data have been obtained via available data sources containing experimental information from ChIP-seq, Hi-C, ChIA-PET tests using different sequencing platforms. Gene annotation was obtained from UCSC Genome Browser (http://genome.ucsc.edu). We reviewed existing software and created a database prototype of bioinformatics tools for 3D genome structure analysis.
Results and Conclusion: We tested program for analysis of ChIA-PET experimental data. 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. 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 has to be solved by our software development.
Acknowledgements: OT internship at Novosibirsk University was supported by French educational exchange program. The research has been supported in part by RFBR and by the budget project No. 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. Fullwood M.J. et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. Nature. 462(7269):58-64.
3. Li G. et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 148(1-2):84-98.