

Sequencing technologies for 3D chromosomes contacts analysis and computer tools

O. Thierry¹, S.S. Kovalev², A.V. Tsukanov², Y.L. Orlov^{2,3*}

¹ *University of Bordeaux, France*

² *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

³ *A.O. Kovalevsky Institute of Marine Biology Research RAS, Sevastopol, Russia*

* *e-mail: orlov@bionet.nsc.ru*

Key words: genomics, bacteria, microbiology, sequencing, environments, GC content, bioinformatics

Motivation and Aim: The aim of this work was to review existing computer tools for 3D genome structure data analysis and spatial topological domains, especially for the Hi-C method. Chromatin interactions play a critical role for gene expression regulation. Series of post-genome technologies have been developed to study the total binding of transcription factors in genome, such as chromatin immunoprecipitation arrays (ChIP-Seq) and Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) [1]. Several methods have been developed in order to study the spacial proximity between genomic regions, based on microscopy or high throughput sequencing methods (HTS) as Chromosome Conformation Capture (3C) and chromatin immunoprecipitation. 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 [2]. Microscopy methods involve fluorescence in situ hybridization (FISH), a single-cell assay with optically labeled probes that hybridize to complementary regions of chromosomes, allowing direct measuring physical distance between them. We reviewed computer tools for 3D genome organization analysis.

Methods and Algorithms: As all the HTS methods, the Hi-C generates a huge amount of data which need to be analyzed, targeting the whole genome, leading to the development of various software and tools with different effectiveness and accuracy. We used to review different categories of software developed for 3D genomics experiments support – data processing, construction of contact maps, statistical estimates, visualization. Note that the visualization can be done through four main process: heatmap, 3-dimensional representation, circular representation and network.

Results and conclusion: We tested program for analysis of ChIA-PET experimental data. 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 software development.

Acknowledgements: The research has been supported by the Ministry of Education and Science of the Russian Federation grant #14.W03.31.0015. Computing was done at Siberian Supercomputer center SB RAS.

References

1. M.J. Fullwood et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome, *Nature*, 462(7269):58-64.
2. G. Li et al. (2014) Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application, *BMC Genomics*, 15(Suppl 12):S11.