DOI 10.18699/SBB-2022-062

Simulating of 3D genome data with predefined chromosomal rearrangements

Nuriddinov M.[1]*, Mozheiko E.[1], Fishman V.[1, 2]

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

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *nuriddinov@bionet.nsc.ru*

Key words: Hi-C, 3D genomic, chromosome rearrangement, simulation

Motivation and Aim: Many genetic based diseases including development disorders and cancer are caused by structural chromosome abnormalities that affect the genome architecture [1]. This means the detection of structural variants and aberrant positions of regulatory elements are clinically relevant. The Hi-C-like methods are shown to be efficient and precise tool to reveal both known and novel chromosomal rearrangements [2]. Using a machine-guided search to mining the Hi-C-like data seems to be useful; unfortunately, the small sizes of available experimental Hi-C datasets describing rearrangements interferes with benchmarking and an adjusting of searching algorithm parameters. To solve this problem, we developed a flexible and tunable tool to simulate the Hi-C-like contact map with the predetermined chromosomal rearrangements.
Methods and Algorithms: We developed a python-based software to generate Hi-C matrices of chromatin contacts expected for different types of chromosomal rearrangements. Basing on a given reference, the software estimates a dependence of Hi-C contacts from: a distance between contacted loci; a coverage of target locus by Hi-C reads; a contact preference been defined by chromatin architecture features. Accounting that, the software builds the model of contact map for rearranged genome. Additionally, the software can simulate an experimental noise by the binomial or normal distribution and differences in sequencing depth.
Results: We developed the software that simulates contact map for the given chromosomal rearrangement, such as copy number variations, inversion and translocations or their combination. The tool can adjust parameters to simulate different types of Hi-C experiments including capture-Hi-C. The generated models are represented as contact matrices in sparse or Juicebox-preferred .hic-format.
Conclusion: Described software allows to train, to tune and to test computational tools searching for chromosomal rearrangements on different types of Hi-C-like data.
Acknowledgements: The study is supported by Russian Science Foundation (project 21-65-00017).

References

1. Jackson M. et al. The genetic basis of disease. *Essays Biochem*. 2018;62(5):643-723.
2. Harewood L. et al. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol*. 2017;18(1):125.