DOI 10.18699/SBB-2022-021

Efficient detection of structural variations using capture-Hi-C data

Mozheiko E.A., Babenko V.N.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** mozheiko@bionet.nsc.ru*

The diagnosis and treatment of patients with hereditary diseases require the creation of efficient methods for the study of individual genomes. The existing approaches either are aimed at searching for a narrow set of genomic variants or are too expensive to use in routine practice. In this way, it is an actual issue to investigate efficient methods, which could couple detection point mutation in exons and searching for novel translocations and copy number alterations (CNA). We proposed a novel method based on sequencing of enriched Hi-C libraries to detect such variants. This method is based on the assumption that Hi-C technology is more informative for the search for chromosomal rearrangements than whole genome data with the same sequencing depth. This is achieved due to the fact that translocations and CNA significantly change the profile of the chromatin spatial contacts.

In this study, we prepare 10 enriched Hi-C-libraries from 8 patients' peripheral blood samples, and 2 cancer cell lines: k562 and A549 [1]. We used these cell lines as a reference of highly rearranged genomes to test our bioinformatics algorithm for detecting chromosomal rearrangements [2].

We have developed a bioinformatics algorithm for detecting interchromosomal translocations, which takes into account key features of altered profile of the chromatin spatial contacts. In this algorithm, we also took into account that chromosomal rearrangement can influence each other.

For CNA detection, we develop a coverage normalization algorithm, based on intrinsic properties of the chromatin. This type of normalization has never been used in other studies, and highly improves accuracy of CNA detection.

Acknowledgements: DNAse I Hi-C library preparation and sequencing were supported by RFBR Grant 18-29-13021. Data analysis performed by E.M. was supported by RFBR Grant 20-34-90110. High-throughput computations were performed on the nodes of the computational cluster of the Institute of Cytology and Genetics, SB RAS (budget project No. 0259-2021-0016).

References

1. Gridina M., Mozheiko E., Valeev E., Nazarenko L.P., Lopatkina M.E., Markova Z.G., Yablonskaya M.I., Voinova V.Yu., Shilova N.V., Lebedev I.N., Fishman V. A cookbook of DNase Hi-C. *Epigenetics Chromatin*. 2021;14(1):15. doi: 10.1186/s13072-021-00389-5.
2. Mozheiko E.A., Fishman V.S. Detection of point mutations and chromosomal translocations based on massive parallel sequencing of enriched 3C libraries. *Russ J Genet*. 2019;55:1273-1281. doi: 10.1134/S1022795419100089.