DOI 10.18699/SBB-2022-229

Exo-C is a new Hi-C-based method for all type genome structural variants detection

Gridina M.[1]*, Stepanchuk Y.[1, 2], Nuriddinov M.[1], Mozheiko E.[1], Valeev E.[1, 2], Nazarenko L.P.[3], Lopatkina M.E.[3], Markova Z.G.[4], Yablonskaya M.I.[5], Voinova V.Y.[5], Shilova N.V.[4], Lebedev I.N.[3], Fishman V.[1, 2]

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

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

[3] *Research Institute of Medical Genetics, Tomsk National Research Medical Center, RAS, Tomsk, Russia*

[4] *Research Centre for Medical Genetics, Moscow, Russia*

[5] *Clinical Research Institute of Pediatrics Named After Acad. Y.E. Veltischev, Moscow, Russia*

** gridinam@gmail.com*

Key words: Chromosomal rearrangements, Hi-C, Chromosome conformation capture, Exome sequencing, Structural variation detection

Motivation and Aim: High-throughput methods of genome analysis allow efficient detection of point mutations. However, their sensitivity for detecting chromosomal rearrangements, especially balanced ones, remains rather weak. As a result structural variations are largely understudied, compared to single nucleotide variants (SNV). The Hi-C method is widely used to study the spatial chromatin architecture and to assemble genomes; moreover, it is a powerful tool to detect balanced and unbalanced chromosomal rearrangements [1]. However, Hi-C data are not completely suitable for single nucleotide variant (SNV) detection. First of all, it is limited by a nonuniform genomic coverage, which is result of using restriction endonucleases in Hi-C-protocol. Second, to the reliably detect SNV in exome, the sequencing have to be significantly deeper than for conventional *in situ* Hi-C. Thus, the aim of the research was developing new Hi-C-based method for simultaneous detection of SNV in exome and different type of chromosomal rearrangements at a genome-wide level.

Methods and Algorithms: We developed a new method Exo-C. The first step of Exo-C libraries preparation was DNase Hi-C [2]. The second step was enrichment with exome sequences, which was performed using the KAPA HyperExome kit. Sequencing of obtained libraries was performed on Illumina or MGI platforms, in PE150 paired read mode, with a depth of 30-60 million reads per individual library.

ATo dectect chromsomal rearrangments, we derived statistics from scaling of chromatin 3D contacts with genomic distance. We used joint analysis of coverage and contacts scaling distributions to find significant outliers corresponding to the rearranged loci. To detect SNV, we used stadard approach based on GATK guidelienes.

Results: We modified the existed protocol of DNase Hi-C to robustly obtain the good quality Hi-C data. As a result of using sequence-agnostic exonuclease we obtained more uniform genomic coverage than it would be possible using restriction enzymes. Applying the developed Exo-C protocol to the collection of human blood samples, we obtained three-dimensional contact maps for 45 individuals among which were patients with translocations (13) and inversions (6) previously detected bycytogenetic analysis. As a result, we successfully confirmed all translocations and four inversions by Exo-C method. The rearrangement structure was significantly refined in some cases.

The one of representative example is P47, the patient with multiple exostosis. The patient karyotype was 46XY, t(8;11;21)(q23;q22;q21), however Exo-C data allowed us to determine the chromosome 20 also affected by the rearrangement. This result of Exo-C assay, was further confirmed by Multicolor FISH. Moreover one of the breakpoints was inside the first intron of *EXT1* gene. Mutations in the *EXT1* gene lead to the formation of multiple exostoses; this pathology is inherited in an autosomal dominant manner. Thus, the results of Exo-C analysis allowed us to fully explain the phenotype of the patient and establish the molecular mechanism of the development of the disease.

The SNV detection was the next important task, which we would like resolve. The comparison of the list of SNVs identified by the Exo-C method and classical exome sequencing showed 99 % overlap at a similar sequencing depth. Thus, the Exo-C method makes it possible to find single nucleotide variants with an accuracy comparable to classical exome sequencing.

Conclusion: We developed Exo-C method, which allowed to obtain the information about balanced and unbalanced chromosomal rearrangements as well as clinically relevant SNV from a single experiment and did not require extremely deep sequencing.

Acknowledgements: The study is supported by the Russian Science Foundation (grant No. 21-65-00017).

References

1. Harewood L., Kishore K., Eldridge M.D., Wingett S., Pearson D., Schoenfelder S., Collins V.P., Fraser P. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol.* 2017;27(18(1)):125.
2. Gridina M., Mozheiko E., Valeev E., Nazarenko L.P., Lopatkina M.E., Markova Z.G., Yablonskaya M.I., Voinova V.Y., Shilova N.V., Lebedev I.N., Fishman V. A cookbook for DNase Hi-C. *Epigenetics Chromatin*. 2021;14(1):15.