GeNeXpreSS

Utility of Long-read Sequencing in Human Genetic Disorders

Usha R Dutta

Diagnostics Division, Centre for DNA Fingerprinting ad Diagnostics, Hyderabad

Correspondence to: Dr Usha R Dutta *Email:* ushadutta@hotmail.com

Cytogenomic characterization of a novel de novo balanced reciprocal translocation t(1;12) by genome sequencing leading to fusion gene formation of EYA3/EFCAB4b (Dutta et al., 2022)

Disease-associated structural variants (SV) often have breakpoints within the gene or in its vicinity. The accurate detection of breakpoints helps in identifying the molecular mechanism and the risks involved with the disrupted genes. The detection of SVs is mostly done by chromosomal analysis, but karyotyping is a low-resolution technique. In this study, the authors have evaluated a girl with short stature wherein the cytogenetic analysis revealed an additional material on the 12p region and fluorescence *in situ* hybridization (FISH) with whole chromosome paint (WCP) FISH probe for chromosome 12 revealed a translocation of t(1;12)(p36.1;p13.32). The parents did not have the translocation; hence it was most likely of *de novo* origin. Microarray comparative genomic hybridization (CGH) ruled out significant copy number variations. The physical mapping with FISH clones identified the spilt signals. Long-read sequencing by Oxford Nanopore Technologies was done and an efficient pipeline was developed which helped in the identification of 4 unique chimeric reads. Sanger sequencing confirmed the junction fragment, revealing microhomology-mediated break-induced repair (MMBIR). The breakpoint on chromosomes 1 and 12 was found to disrupt *EYA3* and *EFCAB4B* genes. Real time PCR showed an over-expression of both genes due to fusion gene formation. This study highlights the utility of long-read sequencing in delineating breakpoints of balanced reciprocal translocations and easy development of the pipelines.

Identification and characterization of two DMD pedigrees with large inversion mutations based on a long-read sequencing pipeline (Geng et al., 2023)

Duchenne muscular dystrophy (DMD) is an X-linked recessive disorder. The mutation spectrum includes 60-70% of deletion/duplications, 20% of point mutations, and rarely structural variations like translocations and inversions. Most of the deletion/duplications can be detected by multiplex ligation-dependent probe amplification (MLPA) and sequence variants/ point mutations using Sanger sequencing or short-read next-generation sequencing (NGS). In this report, two cases of DMD were studied wherein no deletion/duplication or point mutation was identified. First was a 5-year-old boy with family history of DMD. His serum creatine kinase (CPK) levels were high, and muscle biopsy showed absence of dystrophin. MLPA could not detect any copy number variations (CNVs) and exome sequencing did not detect any sequence variants. In view of the clinical phenotype being strongly suggestive of DMD, a rare mutation was suspected. Chromosomal analysis showed normal karyotype of 46, XY. RNAseq and cDNA array capture sequencing showed absence of 3-55 exons hinting at a rare mutation. Optical genome mapping revealed a 55 Mb pericentric inversion at Xp21 to Xq21 region. Long-read sequencing was performed by Oxford Nanopore sequencing and SMRT technology of PacBio, and the exact breakpoint region was identified. The breakpoint region showed a 16 base pair insertion and repeat elements. Induced pluripotent stem cells iPSC and full cDNA sequence study revealed 4 transcripts, one was small with exons 1 and 2 and the other 3 were fusion transcripts, formed as a result of inversion. Sanger confirmation was done in several family members and carrier analysis of a

pregnant cousin was normal. The second case was a 10-year-old boy with no family history but incidentally found to have elevated serum CPK levels. Immunohistochemistry (IHC) of muscle biopsy showed absence of dystrophin staining. Direct long-read sequencing identified a 96 Mb inversion. Both the breakpoint regions anchored several repeat elements. The authors were able to establish an efficient long-read sequencing pipeline for inversion detection.

Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing (Stevanovski et al., 2022)

Short tandem repeats (STR) are 2-6 base pair repeats of DNA. Usually long expanded STR alleles are pathogenic and more than 40 heritable disorders like Huntington disease, Fragile X syndrome, Myotonic dystrophy, etc. are known to be caused by this mechanism. This study demonstrates the validity and utility of programmable targeted Oxford nanopore sequencing for the genetic diagnosis of STR expansion disorders. Targeted long-read sequencing is one of the techniques used by Oxford Nanopore Technologies (ONTs) using "Read Until" functionality. The ONT device is programmed to recognize and reject/accept specific target selections. It can be used to achieve accurate molecular characterization of all known neuropathogenic STRs in a single assay. The authors customized a panel with all known STRs in a single assay primarily for neurological and neuromuscular diseases. For each gene, the entire locus was targeted including 50 kb of flanking sequences in either direction. Readfish software was used to target sequences of 37 DNA samples from patients and observed a consistent reduction in read length for off-target reads when compared to on-target reads. Although this study could identify the tandem repeat expansion disorders it is still considered to be a research tool.

Detecting cell–of–origin and cancer–specific methylation features of cell–free DNA from Nanopore sequencing (Katsman et al., 2022)

Circulating cell-free DNA (cfDNA) can reveal informative features of its tissue of origin, including somatic genome alterations, DNA modifications, and cell type-specific fragmentation patterns. ONT can call accurate DNA methylation from native DNA and produce single base-pair resolution results highly similar to bisulfite sequencing. This study shows that cell type and cancer-specific methylation can also be detected, as well as cancer-associated fragmentation signatures. This study also highlights that ONT's shallow whole-genome sequencing (WGS) could be a powerful tool for liquid biopsy. In this study, the cell type fractions from cfNano were estimated. The sample size was small but their results suggest that cancer-specific features of DNA methylation, and fragmentation were concordant between cfNano and Illumina-based WGS methods. Also, the results suggest that short di-nucleosomes could be a more robust cancer marker than short mono-nucleosomes, although this needs to be validated in a larger study. This study showed the feasibility of ONT sequencing for circulating tumor DNA detection by comparing methylation and several fragmentation features to matched Illumina samples and comparable Illumina-based datasets.

References

1. Dutta UR, et al. Cytogenomic Characterization of a novel de novo balanced reciprocal translocation t(1;12) by genome sequencing leading to fusion gene formation of EYA3/EFCAB4b. Mol Syndromol. 2022; 13 (5): 370–380.
2. Geng C, et al. Identification and characterization of two DMD pedigrees with large inversion mutations based on a long-read sequencing pipeline. Eur J Hum Genet. 2023; 31(5): 504-511.
3. Katsman E, et al. Detecting cell-of-origin and cancer-specific methylation features of cell-free DNA from Nanopore sequencing. Genome Biol. 2022; 23(1):158.
4. Stevanovski I, et al. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. Sci Adv. 2022; 8(9): eabm5386.