GeNeXpreSS

Novel Tools for Detecting Structural Variants: Optical Genome Mapping & More

Saswati Das

Department of Medical Genetics, Nizam's Institute of Medical Sciences, Hyderabad, India

Correspondence to: Saswati Das *Email:* saswati.lef1@gmail.com

The impact of inversions across 33,924 families with rare disease from a national genome sequencing project (Pagnamenta et al., 2024)

Genomic inversions, segments of DNA with reversed orientation compared to the reference genome, are prevalent in human populations and vary in size. Traditional karyotyping can detect these inversions but typically misses those smaller than 10 Mb. Despite advancements in array technologies, copy-neutral structural variants (SVs), such as inversions, remain underexplored in clinical settings. This study aimed to expand the analysis to 33,924 families in the 100,000 Genomes Project (100kGP), examining genes associated with haploinsufficiency (HI). Structural variants were identified using Canvas and Manta, with prioritization of ultra-rare inversions. Ultimately, only 45 families were identified, representing 1-2% of diagnoses across 351 genes. Notable detected inversions included an intragenic *MSH2* founder inversion, a complex maternally inherited structural variant, a de novo inversion in the *HOXD* cluster linked to Kantaputra-type mesomelic dysplasia, and an inversion with a breakpoint in intron 4 of the *APC* gene indicating potential gene disruption. Limitations included a focus on HI genes and potential oversight of inversions in repetitive regions.

A comparison of structural variant calling from short-read and nanopore-based whole-genome sequencing using optical genome mapping as a benchmark (Pei et al., 2024)

This study aimed to assess the clinical efficacy of three genomic technologies: Illumina short-read whole genome sequencing (SR-WGS), Oxford Nanopore Technologies long-read whole genome sequencing (LR-WGS), and Bionano optical genome mapping (OGM) for detecting rare structural variants (SVs) of potential clinical significance. The investigation centered on a model cohort of patients affected by craniosynostosis (CRS), a condition known for its considerable clinical and genetic heterogeneity. As part of the 100,000 Genomes Project (100kGP), 114 CRS families that lacked a genetic diagnosis were recruited and sequenced using Illumina SR-WGS technology. In spite of thorough investigations focused on uncovering causative single nucleotide polymorphisms (SNPs) and structural variants (SVs), 78 families continued to lack a diagnosis. From these families, nine trios were selected. By integrating analyses from LR ONT WGS and Bionano OGM, the uncovered SVs that may have been overlooked by Illumina WGS were considered. A subset of potentially clinically relevant rare SVs was identified through Bionano OGM, which was utilized to create a "truth dataset" for benchmarking the performance of current variant callers from Illumina and ONT to evaluate their clinical utility in rare disease contexts.

VolcanoSV enables accurate and robust structural variant calling in diploid genomes from single-molecule long-read sequencing (Luo et al.,2024)

Advances in long-read sequencing technologies have provided a valuable resource for comprehensive SV detection. However, accurately identifying SV breakpoints and sequences remains challenging. To address this, researchers have developed innovative hybrid SV detection pipelines that utilize both reference genomes and local de novo assembly. One such tool, VolcanoSV, employs phased single nucleotide

polymorphisms (SNPs) and unique k-mer similarity analysis to enable precise haplotype-resolved SV discovery. VolcanoSV constructs comprehensive genetic maps encompassing SNPs, small indels, and all types of SVs, making it well-suited for human genomics studies. Extensive experiments have demonstrated that VolcanoSV surpasses state-of-the-art assembly-based tools in detecting insertion and deletion SVs, exhibiting superior recall, precision, F1 scores, and genotype accuracy across diverse datasets, including low-coverage (10x) datasets. Additionally, VolcanoSV outperforms other tools in identifying complex SVs, such as translocations, duplications, and inversions, in both simulated and real cancer data. The pipeline is also robust to various evaluation parameters and accurately identifies breakpoints and SV sequences.

Optical genome mapping unveils hidden structural variants in neurodevelopmental disorders (Schrauwen et al., 2024)

This study explored the application of optical genome mapping (OGM) in identifying structural variants (SVs) associated with neurodevelopmental disorders (NDDs) that remain undetected by standard exome sequencing. OGM, which surpasses short-read sequencing in capturing complex SVs, was conducted on ultra-high molecular weight DNA from 47 families. OGM analysis of the 47 unsolved families revealed that the majority of identified variants consisted of insertions (67.6%) and deletions (29.2%). Among these families, OGM identified 7 rare variants of interest, including 2 variants of unknown significance and 5 likely pathogenic or pathogenic structural variants (SVs). These likely pathogenic or pathogenic SVs were found in known neurodevelopmental disorder (NDD) genes, such as *BCL11A*, *OPHN1*, *PHF8*, *SON*, and *NFIA*. Additionally, an inversion affecting the *NAALADL2* gene was identified, previously linked to complex rearrangements in NDD cases. The variants missed by exome sequencing primarily included larger insertions (>1 kbp), inversions, and small deletions/duplications (1-4 exons). OGM not only enhances molecular diagnostics for NDDs but also has the potential to uncover novel NDD-related genes harbouring complex SVs often overlooked by conventional sequencing methods.

References

1. Luo C, et al. VolcanoSV enables accurate and robust structural variant calling in diploid genomes from single-molecule long read sequencing. Nat Commun. 2024; 15(1): 6956.
2. Pagnamenta AT, et al. The impact of inversions across 33,924 families with rare disease from a national genome sequencing project. Am J Hum Genet. 2024; 111(6): 1140-1164.
3. Pei Y, et al. A Comparison of Structural Variant Calling from Short-Read and Nanopore-Based Whole-Genome Sequencing Using Optical Genome Mapping as a Benchmark. Genes (Basel). 2024; 15(7): 925.
4. Schrauwen I, et al. Optical genome mapping unveils hidden structural variants in neurodevelopmental disorders. Sci Rep. 2024; 14(1): 11239.