GeNeViSTA

Next Generation Cytogenetics - Optical Genome Mapping

Sofia Banu, Kanakavalli M K, Chandreswara Raju Kataru, Divya Tej Sowpati, Karthik Bharadwaj Tallapaka

CSIR-Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, Telangana, India

Correspondence to: Dr. T Karthik Bharadwaj *Email:* karthikt@ccmb.res.in

Abstract

Optical genome mapping (OGM) is a state-of-the-art technology that is being increasingly adopted by genomic laboratories across the world to detect structural variations (SVs). Conventional cytogenetic/ molecular genetic technologies used to detect SVs like insertions, deletions, inversions, duplications, and translocations have several limitations and/or are highly dependent on the technical expertise of the personnel. Next-generation sequencing which has revolutionized rare disease diagnosis with its ability to detect small genomic variants (such as single nucleotide variants and small indels) also has several shortcomings in identifying structural variations and certain repeat disorders like facioscapulohumeral dystrophy (FSHD). OGM has a significantly higher resolution than techniques such as karyotyping, fluorescence in situ hybridization (FISH) and chromosomal microarray (CMA) and can detect a wider range of variants in a single assay. With a relatively simple workflow and automated analysis pipelines it is less operator dependent and produces robust and reproducible results with rapid turnaround times. Here, we discuss the technology and elucidate its utility in clinical diagnostic settings.

Keywords: Optical genome mapping, cytogenetics, structural variation, genetic testing

Introduction

Structural variants (SVs) play a pathognomonic role in a wide range of genetic diseases. Conventional cytogenetics methods such as karyotyping, fluorescence in situ hybridization (FISH) and chromosomal microarray (CMA) have been at the forefront of standard of care recommendations (SOC) for detection of these variants (Miller et al., 2010). In case of repeat expansion disorders, Southern blotting or polymerase chain reaction (PCR)-based tests have been used for analysis. With the advent of next-generation sequencing (NGS), exome and genome sequencing have been in the spotlight. While these techniques are widely used in clinical laboratories, they are not without limitations. Karyotyping is the gold standard to detect large visible balanced and unbalanced structural variants in addition to aneuploidy. However, the resolution of detection is ~5 to 10 Mb and SVs below this cutoff will be missed. Additionally, there is high variability in results across samples and laboratories, as it is dependent on the expertise of the operator. CMA has an improved resolution of ~ 50kb but cannot detect balanced rearrangements like translocations. FISH is a targeted approach where only a few loci can be assayed at a time. In light of these shortcomings, there exist lacunae for a single technology to accurately detect a variety of structural variants in a short time.

Optical genome mapping – A one stop solution

Optical genome mapping (OGM) is a next generation cytogenomic technique to comprehensively identify all classes of structural variants, copy number variations (CNVs), repeat expansions and contractions, ring chromosomes, absence of heterozygosity (AOH), aneuploidy and triploidy across the whole genome. Depending on the variant type, it offers 100X - 20,000X more resolution than karyotyping (Smith et al., 2022). For OGM on Bionano Saphyr® (Bionano Genomics, San Diego, California, United States), ultra-high molecular weight (UHMW) DNA (≥150 kbp) is isolated from fresh or frozen blood,

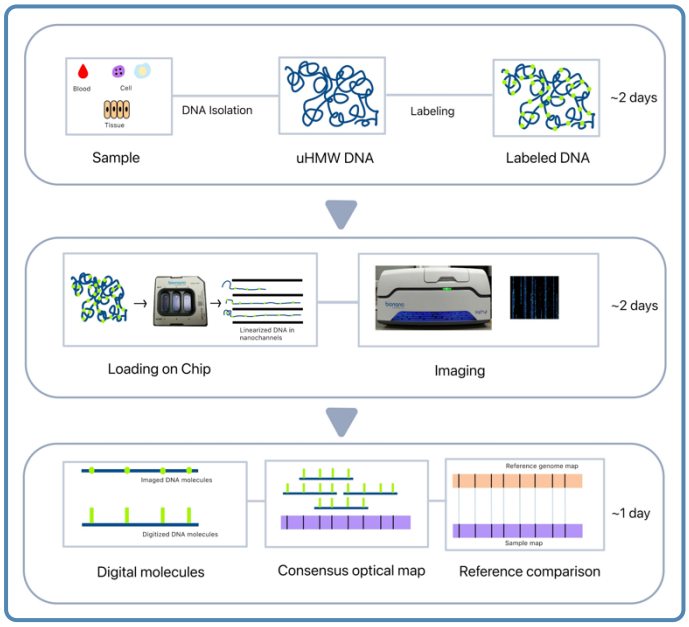

Figure 1 **Optical genome mapping workflow**
Ultra-high molecular weight (uHMW) DNA is isolated from the sample followed by enzymatic labeling with fluorophores. The labeled DNA is loaded onto a Saphyr Chip which linearizes DNA in nanochannels followed by imaging on a Saphyr instrument. The imaged DNA is converted to digitized representations which are then assembled into optical maps. These maps are then compared to the reference genome based on label positions.

bone marrow, tissue, and tumor samples (**Figure 1**). These DNA molecules are then labeled enzymatically with fluorophores using an enzyme DLE-1 at a 6 base-pair sequence motif (CTTAAG) that occurs approximately every 5kb throughout the genome. These labels enable increased resolution and are analogous to bands identified in karyotyping where one band occurs approximately every 5Mb. Three samples can be loaded at once on the Saphyr chip which consists of three flow cells each containing nanochannel arrays within which each single labeled DNA molecule is linearized and imaged on the Saphyr System. Depending on the sample integrity, imaging could take a day or two. The images consist of a unique pattern of bands represented by fluorescent labels which are further converted into digital representations called optical maps and the same can be analyzed bioinformatically. The labels in the optical maps only represent known physical locations of the motif sequence and cannot be used to interrogate regions at single base pair resolution. The optical maps from the sample are mapped to an *in silico* labeled reference genome to detect mismatches in the label patterns. These patterns (**Figure 2**) can then be used to detect structural variants for example, extra labels at a particular locus in the sample when compared to reference could indicate an insertion, missing labels indicate a deletion, and a set of repeated labels could indicate a duplication.

In terms of the bioinformatics workflow, only DNA molecules ≥150 kb are used for analysis. Depending on the SV detection type, there is a choice of four pipelines. For the detection of germline SVs, a *de novo* assembly pipeline can be chosen. In case of somatic detection, the rare variant pipeline can be applied. For targeted approaches encompassing repeat contractions, in case of facioscapulohumeral muscular dystrophy (FSHD) and repeat expansions such as fragile X syndrome, there are specialized EnFocus analysis being offered. OGM can detect SV classes from 500 bp onwards which makes it highly sensitive when compared to other methods. The processing time for the pipelines is very fast when compared to existing methodologies. For our in-house data, the *de novo* assembly pipeline took approximately 12 hours for workflow completion whereas for EnFocus for FSHD it was 8 hours

on our computer servers. We estimate that the end-to-end processing for a clinical sample from DNA extraction to clinical results would take less than a week's time using OGM.

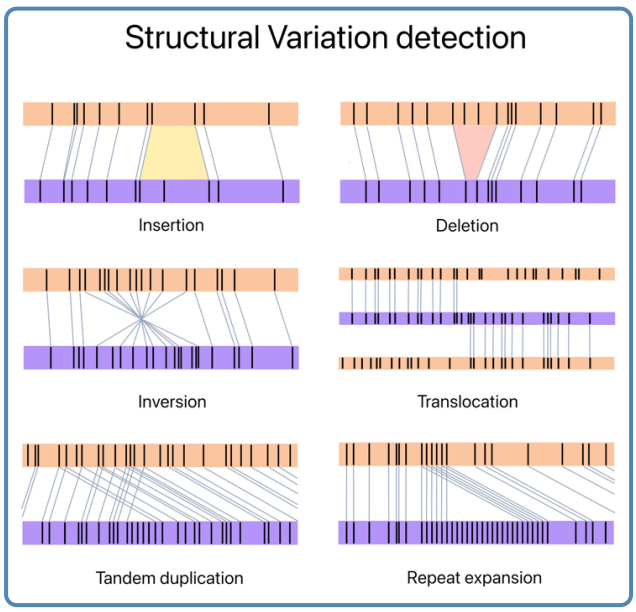

Figure 2 **Structural variant detection**
SVs can be detected by comparing the label patterns between the reference genome (pink) and the sample (purple). In the examples shown, additional labels indicate an insertion, missing labels indicate deletion, inverted label orientation are characteristic of inversions, a single map mapping to distant regions of the genome indicates a translocation and based on the repeating patterns of labels duplications and repeat expansions can be detected. Various other types of complex variants not seen in the above depictions can also be identified by optical genome mapping.

Optical genome mapping versus current standard of care methods

OGM has been assessed in comparison to the current SOC methods in multisite studies involving prospective and retrospective constitutional cohorts, and prenatal and postnatal cohorts. The studies interrogated concordance in technologies across neurodevelopmental problems [autism spectrum disorder (ASD), intellectual disability/ developmental delay (ID/DD), attention-deficit hyperactivity disorder/ oppositional defiant disorder (ADHD/ ODD)], chromosomal aberrations, fragile X syndrome and FSHD (Mantere et al., 2021; Broeckel et al., 2022; Stevenson et al., 2022; Iqbal et al., 2023). Concordance was found ranging from > 98% to 100% depending on the type of anomaly. In addition, several OGM exclusive SVs were also found which were missed out in current SOC methods (Broeckel et al., 2022). The utility of OGM is also being increasingly realized in hematological malignancies where chromosomal abnormalities play a defining role in prognostication of the disease (Neveling et al., 2021; Gerding et al., 2022).

Illustrative clinical scenarios

Patient 1: A 9-year-old male child, born to non-consanguineous parents with severe wasting and weakness of shoulder and arm muscles and abnormal gait was noted to have bilateral facial muscle weakness and scapular winging. There was no significant family history. With a provisional diagnosis of FSHD, his blood sample was analyzed for repeat contraction at D4Z4 locus on chromosome 4 by OGM. After preparation and quality control (QC) checks the sample was run on Bionano Saphyr and the data generated was analyzed using the EnFocus FSHD pipeline, a targeted approach to analyze specific regions on chr4 and chr10 since both contain the D4Z4 arrays. This analysis also enables identification of both permissible and non-permissible haplotypes along with estimation of D4Z4 repeat counts accurately on both the chromosomes. Here, we were able to identify a repeat contraction in the D4Z4 repeat array on 4q35 containing 3 repeats (normal >11) along with the presence of permissive haplotype 4qA **(Figure 3)**. This confirmed the disease diagnosis and the short repeat size detected partly explains the severity of the disease in this child.

Case 2: A 3-year-old female child with developmental delay, seizures and dysmorphism had additional material on p arm of chromosome 12 on karyotyping. On OGM the child was identified to have insertion of duplicated genomic material of ~16Mb from 12q24.21 at 12p13.33 (**Figure 4**). The rest of the CNVs identified were not pathogenic. Thus, OGM was not only able to recognize the chromosomal abnormality but also identified its constitution. This enables accurate genetic diagnosis for better prognostication and

management of such children.

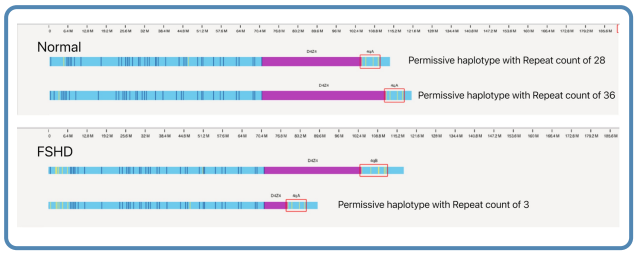

Figure 3 **Identification of D4Z4 repeat contraction in facioscapulohumeral muscular dystrophy**
The first panel shows a sample containing the 4qA permissive haplotype along with a D4Z4 repeat array indicating a normal range. The second panel shows a severe case of FSHD indicated by the permissive haplotype and a repeat contraction indicating only presence of three D4Z4 repeat units.

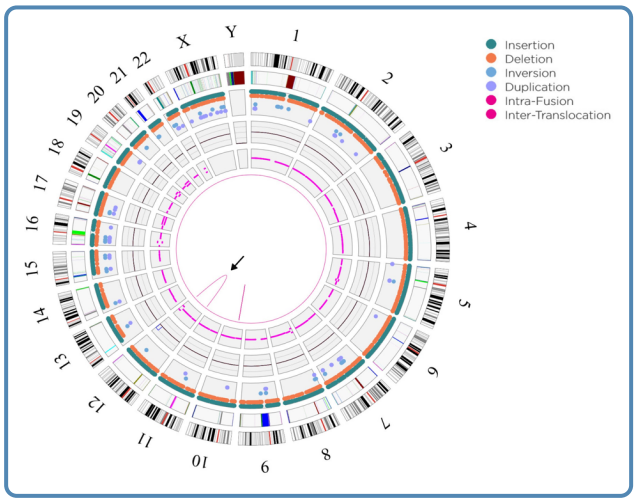

Figure 4 **Circos plot showing translocation and other genome-wide structural variants**
The insertion of material from 12q at 12p is visualized as an arc within chromosome 12 as indicated by an arrow. CNV analysis is suggestive of copy number gain at distal part of 12q (exact loci are evident in a different view not shown in this picture)

Limitations

Extraction of uHMW DNA is an essential requirement necessitating the transport of freshly drawn blood samples to the testing laboratory within 48-72 hrs. This also means that old samples/ stored DNA extracted through conventional methods cannot be used. OGM cannot detect SVs in the centromeric and the telomeric regions. Since it cannot detect acrocentric short arm fusions it cannot distinguish, for example, between a free trisomy or an unbalanced Robertsonian translocation. The technology is relatively expensive, especially in the context of developing countries.

Overall, OGM has been a positive and much needed development in the field of molecular cytogenetics. It has the potential to replace conventional testing methodologies in clinical diagnostics and hopefully, wider adaptation will make the technology more affordable.

References

1. Broeckel U, et al. Multisite Study of Optical Genome Mapping of Retrospective and Prospective Constitutional Disorder Cohorts. medRxiv. 2022: DOI: 10.1101/2022.12.26.22283900.
2. Gerding WM, et al. Optical genome mapping reveals additional prognostic information compared to conventional cytogenetics in AML/MDS patients. Int J Cancer. 2022; 150(12): 1998-2011.
3. Iqbal MA, et al. Multisite Assessment of Optical Genome Mapping for Analysis of Structural Variants in Constitutional Postnatal Cases. J Mol Diagn. 2023; 25(3):175-188.
4. Mantere T, et al. Optical genome mapping enables constitutional chromosomal aberration detection. Am J Hum Genet. 2021; 108(8): 1409-1422.
5. Miller DT, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet. 2010; 86(5): 749-764.
6. Neveling K, et al. Next-generation cytogenetics: Comprehensive assessment of 52 hematological malignancy genomes by optical genome mapping. Am J Hum Genet. 2021; 108(8):1423-1435.

7. Smith AC, et al. Optical genome mapping for structural variation analysis in hematologic malignancies. Am J Hematol. 2022; 97(7): 975-982.

8. Stevenson RE, et al. Multisite evaluation and validation of Optical Genome Mapping for prenatal genetic testing. medRxiv; 2022. DOI: 10.1101/2022.12.19.22283552.