Bionano Genome Imaging Identifies Large Structural Variants in Genetic Disorders

All types of large structural events are detected as heterozygous or homozygous variants with unrivaled sensitivity and specificity.

Structural variants (SVs) make up the majority of human genomic variation and encompass insertion/deletions, duplications, translocations, inversions, and tandem repeat expansions. These events drive the genetic diversity among individuals and populations. It is estimated that every individual harbors a median of 18 Mbp of structural variants per diploid genome.[1] Although single nucleotide polymorphisms (SNPs) are more frequent in the genome, SVs account for more nucleotide changes and it is suggested that they have a disproportionate impact on gene expression relative to their number. Despite their importance, detection of SVs is still challenging for sequencing-based technologies. SVs are largely missing from genome-wide association studies (GWAS), and their impact on gene expression is greatly underestimated.[1]

For many known syndromes, clinically relevant large SVs are well characterized. These SVs include deletions in Prader-Willi syndrome, DiGeorge syndrome and Williams-Beuren syndrome. They also include duplications as in Charcot-Marie-Tooth disease and inversions such as in Haemophilia A.[2] More recently, large SVs have been found to play a role in neurological diseases, including early-onset neuropsychiatric disorders,[3] Tourette syndrome,[4] and Parkinson's disease.[5] SVs are also known to influence coronary heart disease[6] and congenital heart disease,[7] obesity,[8] and pharmacogenetics.[9]

Two thirds of the human genome consists of repetitive sequences, making it impossible for sequencing to detect SVs comprehensively. The most common repetitive sequences in the genome are long interspersed nuclease elements (LINEs), short interspersed nuclease elements (SINEs), retrotransposons, and segmental duplications (Figure 1). Next-Generation Sequencing (NGS) provides primarily short-read sequences that map with poor accuracy to these repeats. Alignment algorithms typically fail to identify the exact genomic location for these short-reads. When they do align, the limited 100–150 bp read length and spacing of paired-end reads does not allow for a correct positioning of larger repeats. In addition, NGS does not reliably detect balanced SVs such as inversions and translocations. Non-allelic homologous recombination of repetitive sequences is thought to be a predominant mechanism for the origin of many large SVs. The non-unique sequences flanking these SVs often make them invisible to sequencing-based detection methods.

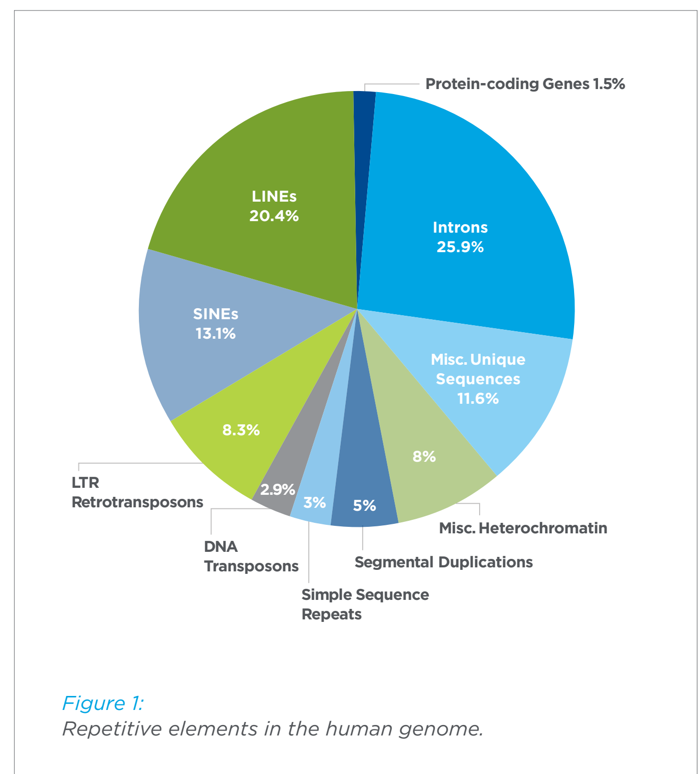

Figure 1:
Repetitive elements in the human genome.

While long-read sequencing has improved significantly over the years and occasionally reads reach hundreds of kilobase pairs, median read lengths are typically in the 10 to 30 kbp range. This range is not sufficient to span longer repetitive areas of the genome or elucidate large, complex events. Sequencing-based approaches lack the sensitivity to reliably detect heterozygous structural variants,[10] making the detection of rare variants in clonal populations impossible. In their current state, long-read sequencing technologies require too great a sacrifice in cost, time, and coverage

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2020, Bionano Genomics Inc.

depth to be suitable for clinical research and diagnosis in cancer and other genetic diseases.

Clinical exome sequencing solves only about 30% of rare diseases.[11] Whole genome or whole exome sequencing can reliably identify single nucleotide variants and small insertions and deletions, but they fail to identify most large insertions, deletions, and copy-number variations in repetitive regions of the genome. This limitation leaves a large number of potentially disease causing variants undetected.

In a multi-platform study, the Human Genome Structural Variation Consortium combined different technologies to comprehensively capture the full spectrum of genetic variations in three family trios.[12] A combination of Illumina short-read sequencing, PacBio long-read sequencing and optical mapping with Bionano Genome Imaging were used to overcome the barriers that limit sequencing algorithms. This initiative led to the discovery of approximately seven-fold more variations than what was uncovered by short-read sequencing alone. Among insertions and deletions detected, the majority of events larger than 1 kb were discovered by both optical mapping and sequencing (either short-read or long-read), but interestingly a vast proportion of the very large variants were systematically missed by sequencing and captured only by optical mapping with Bionano Genome Imaging (Figure 2). This result illustrates the caveats encountered by the 13 algorithms used by sequencing to accurately and comprehensively detect SVs, and highlights the need for a complementary approach that is not sequencing based.

The limitations of sequencing clearly show why direct visualization of DNA is the most reliable approach for the identification of structural variants to date. The way structural variants are detected in samples has barely evolved over the past decades. Cytogenetic methods such as karyotyping, fluorescent *in situ* hybridization (FISH) and array Comparative Genomic Hybridization (aCGH) are still the gold standard for clinical SV detection. Unfortunately, none of those methods alone

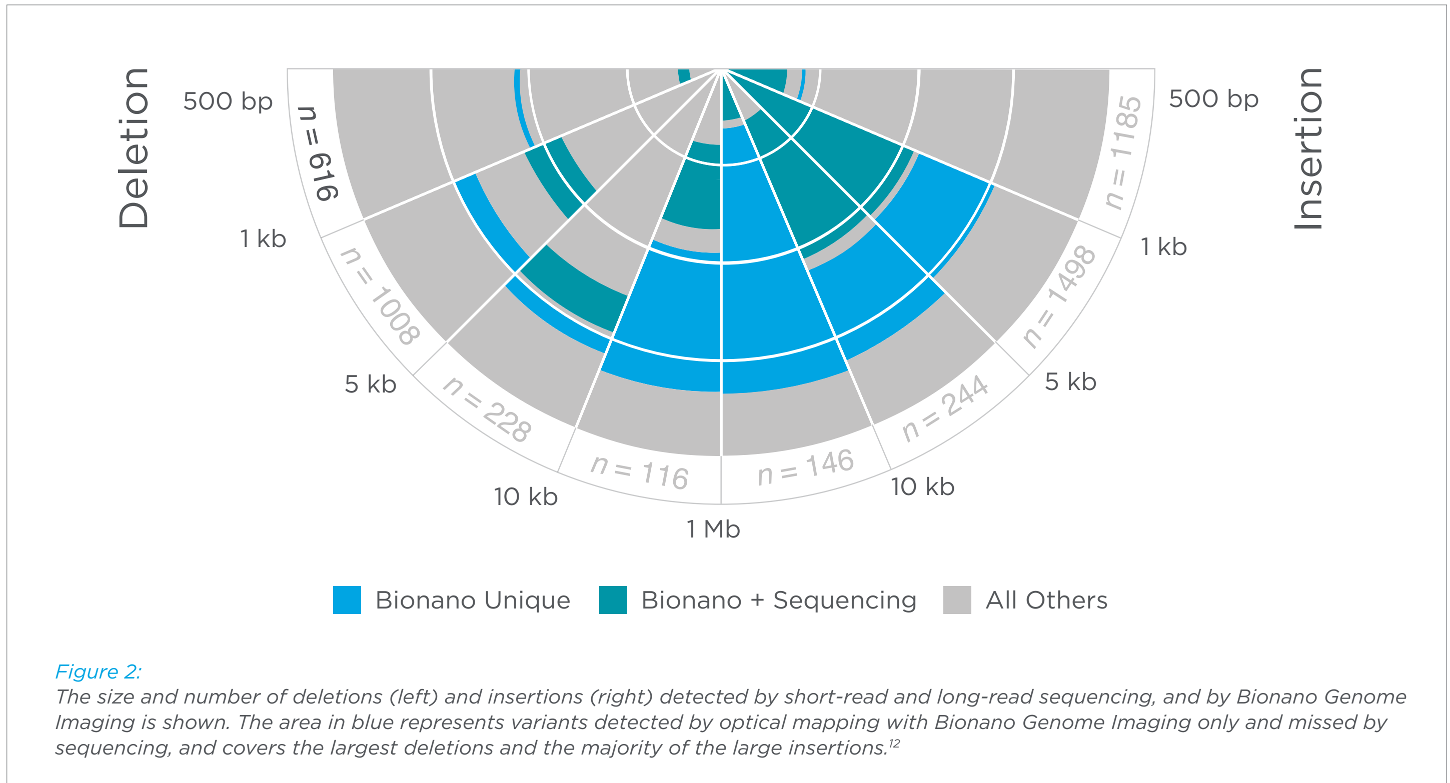

Figure 2:
The size and number of deletions (left) and insertions (right) detected by short-read and long-read sequencing, and by Bionano Genome Imaging is shown. The area in blue represents variants detected by optical mapping with Bionano Genome Imaging only and missed by sequencing, and covers the largest deletions and the majority of the large insertions.[12]

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2020, Bionano Genomics Inc.

can address complex cases due to their technical limitations (Table 1) and must be combined in order to be effective.

Method	Resolution	Limitation
Karyotyping	5-10 Mbp	Extensive training required for interpretation. Cell culture; Cannot detect mosaicism <10%
FISH	100 kbp	Extensive training required for interpretation. Targeted approach; Cannot detect mosaicism <10%
Array based techniques	20-200 kbp	Cannot detect balanced rearrangement, mosaicism <10%; agnostic to the nature of a structural aberration.

Table 1:
Methods used in cytogenomic analysis and their resolution.[13]

Optical mapping with Bionano Genome Imaging is the only technology that detects all types of homozygous and heterozygous SVs, from 500 bp up to millions of bp. Ultra high molecular weight genomic DNA is extracted, labeled, linearized, and uniformly stretched in high density NanoChannel arrays on the Saphyr chip which is then imaged by the Saphyr® genome imager. The Saphyr Genome Imager generates up to 5 Tbp per flowcell, or 400x raw coverage of molecules between 150 kbp and 3 Mbp. A maximum of three samples can be loaded on the Saphyr chip. Bionano's labeling chemistry, Direct Label and Stain (DLS), uses a single direct-labeling enzymatic reaction to attach a fluorophore to the DNA at a specific 6-basepair sequence motif, yielding approximately 15 labels per 100 kbp in the human genome. The label patterns allow the long molecules to be uniquely identified and aligned. Using pairwise alignment of the single molecules, consensus genome maps are constructed, refined, extended, and merged. Molecules are then clustered into two alleles, and a diploid assembly is created to allow for heterozygous SV detection. Genome maps typically span entire chromosome arms in single, contiguous maps. This means that complex genomic rearrangements are visualized in the context of the entire chromosome. While WGS and chromosomal microarray can often detect large deletions and duplications, only Bionano's extremely long molecule data can determine the order and orientation of complex events, and provide direct visualization.

Bionano genome maps are built completely *de novo*, without any reference guidance or bias. Short-read sequences obtained from NGS are typically aligned to a reference. This alignment often fails to detect true structural variants by forcing the short-reads to map to an incorrect or too divergent reference, or by excluding mismatched reads from the alignment. Only *de novo* constructed genomes, such as those generated by optical mapping with Bionano Genome Imaging, allow for a completely unbiased, accurate assembly.

Bionano's SVs are observed, and not inferred as with NGS. When short-read NGS sequences are aligned to the reference genome, algorithms piece together sequence fragments in an attempt to rebuild the actual structure of the genome. SVs are inferred from the fragmented data, resulting in mixed success. Bionano Genome Imaging relies on direct visualization of megabase-size native DNA, just like the traditional cytogenetic methods of FISH and karyotyping. With Bionano Genome Imaging, most large SVs or their breakpoints can be observed directly in the label pattern on the molecules.

Bionano algorithms call SVs by comparing label patterns. To identify a structural variation, a *de novo* genome map assembly can be aligned to a reference genome, or samples can be aligned to each other directly (e.g. trio analysis). When aligning a genome map to a reference assembly, Bionano software identifies the location of the same recognition sequence used to label the DNA molecules in the reference genome and aligns matching label patterns in the sample and reference. This alignment provides all the annotation of the reference to the *de novo* assembled genome. By observing changes in label spacing and comparisons of order, position, and orientation of label patterns, Bionano's automated structural variation calling algorithms detect all major structural variation types (Figure 3), and these structural variants are sized with extremely high accuracy. Compared to the insertions and deletions detected by NGS in a human genome, Bionano SV calls differ by a median of only 60 bp.

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2020, Bionano Genomics Inc.

The performance of SV calling by Bionano Solve, a suite of tools for analyzing data and generating fully annotated SVs, was established from datasets obtained from well described cells lines with known alterations as well as from simulated data in order to manipulate a larger number of events. To that end, SVs were simulated and randomly introduced in hg19 human reference before detection by Bionano Solve. An exhaustive list of performances and detail of the procedure is available in Bionano's Theory of Operation: Structural Variant Calling.[18]

A recent UC San Francisco study by Dr. Pui-Yan Kwok used optical mapping with the Bionano Genome Imaging to analyze structural variation in a record 154 humans previously sequenced as part of the 1000 genome project.[14] Bionano technology identified 8.5 times more large insertions in the same samples than previously reported by the 1000 Genomes Project using short-read whole genome sequencing, and 35% more large deletions. Many of the large structural variations were flanked by repetitive elements, which short-read sequencing fails to accurately resolve, rendering the structural variants undetectable. The high accuracy of insertion/deletion calls was confirmed by proving Mendelian inheritance for a subset of samples assembled in trios and by cross-referencing calls from linked-read sequencing.

Large copy number variants (CNV) are detected at low allele fraction by an independent molecule-based tool. A per-label fractional copy number analysis algorithm normalizes the raw coverage profile on each sample to provide copy number calls for detection of aneuploidy, loss of chromosome arms, and large duplications and deletions above 250 kbp. It handles complex genomes with complex state changes. The copy number tool has a sensitivity of over 90% for large duplications and deletions in as low as 10% allele fraction, and sensitivity of 95-100% for aneuploidy in as low as 10% allele fraction.

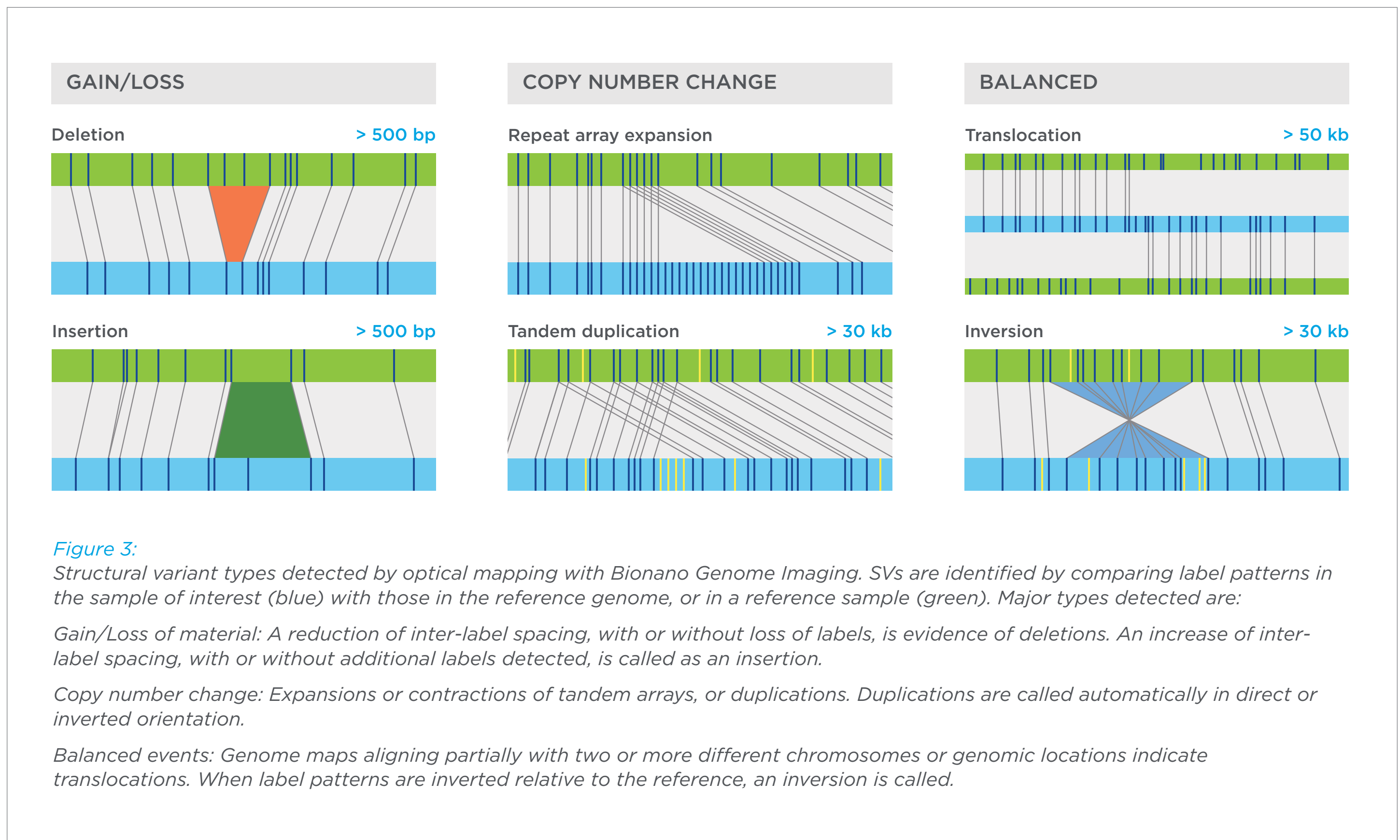

Figure 3:
Structural variant types detected by optical mapping with Bionano Genome Imaging. SVs are identified by comparing label patterns in the sample of interest (blue) with those in the reference genome, or in a reference sample (green). Major types detected are:

Gain/Loss of material: A reduction of inter-label spacing, with or without loss of labels, is evidence of deletions. An increase of inter-label spacing, with or without additional labels detected, is called as an insertion.

Copy number change: Expansions or contractions of tandem arrays, or duplications. Duplications are called automatically in direct or inverted orientation.

Balanced events: Genome maps aligning partially with two or more different chromosomes or genomic locations indicate translocations. When label patterns are inverted relative to the reference, an inversion is called.

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2020, Bionano Genomics Inc.

Bionano detects extremely rare variants in mosaic samples. While the *de novo* assembled consensus maps enable the detection of SVs in homogeneous, diploid genomes, a separate pipeline was developed to analyze mosaic samples, or heterogeneous cancer samples. This tool is based on the alignment of single molecules to the reference and calls SVs starting at 5 kbp that are detected in just a few molecules, which corresponds to 90% sensitivity at a Variant Allele Fraction (VAF) of just 1%. The Rare Variant Pipeline (RVP) also computes variant allele fraction estimates, by calculating the number of relevant single molecules supporting a specific SV call divided by the total number of molecules aligned. As such, 1% allele fraction corresponds to a heterozygous variant present in 2% of diploid cells.

An example of mosaicsm was reported in a paper by a group of scientists from GrandOmics and universities in Beijing and Philadelphia.[15] The study discussed the use of optical mapping with Bionano Genome Imaging to genotype the repeat array that's involved in FSHD.

Facioscapulohumeral Muscular Dystrophy (FSHD) is a common muscular dystrophy affecting the muscles of the face, shoulder blades and upper arms most. Most FSHD cases are caused by a contraction of a tandem repeat array subtelomeric on chromosome 4q. FSHD is currently diagnosed by pulsed-field gel electrophoresis and Southern blot, which is labor-intensive, time consuming and, most importantly, inaccurate. Bionano EnFocus™ FSHD Analysis enables a direct count or quantitative measurement of repeat numbers, differentiates between the pathogenic 4qA allele and the non-pathogenic allele, and between the chromosome 4 repeat and a largely identical repeat on chromosome 10 that's not involved in the disease. Additionally, it can quantitate the degree of mosaicism. The study discusses several patients who show mosaicism for this locus (Figure 4).

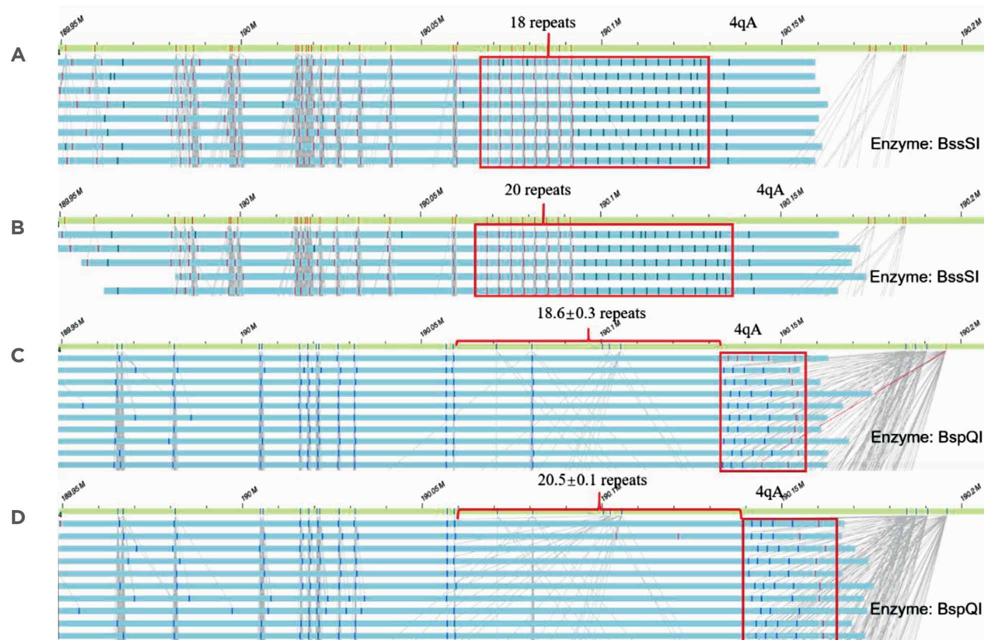

Figure 4:
Detection of somatic mosaicism in a late-onset FSHD patient with 2, 11 and 17 repeats in the D4Z4 repeat region, all with 4qA configuration. The BssSI labeling enzyme places a single label in each repeat unit, making the detection of the repeat copy number as simple as counting the number of labels in the repeat area. Other labeling enzymes, like DLE-1, place labels flanking the repeat, allowing the total repeat length to be accurately measured and the repeat number to be calculated from that measurement.[15]

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2020, Bionano Genomics Inc.

Bionano's Variant Annotation Pipeline reduces thousands of structural variants to those of clinical relevance. The Variant Annotation Pipeline (VAP), part of the Bionano Access® software, streamlines genetic studies. Using VAP, structural variation calls from multiple samples can be analyzed as a group to detect *de novo* mutations when comparing calls from proband with parents. By using a control database of common variants, VAP filters the thousands of identified variants down to hundreds that are rare, or to a handful of *de novo* variants. The built-in control database contains the frequency of each variant in an ethnically diverse sample set of over 200 healthy individuals. It also identifies the genes they overlap with or are closest to in the genome. The VAP is part of Bionano Access, which provides an interface for setting up experiments on on the Saphyr System, starting and monitoring instrument runs, launching *de novo* assemblies and SV calling, visualizing SVs, and annotating variants with the VAP. The results can be exported as a dbVar compliant VCF file, for easy integration with variants identified with NGS or other methods. Within Bionano Access, all structural variants can be visualized as an interactive circos plot, combining SV calls with copy number calls and translocations in a single view. Variants can be filtered to show only those that are somatic, rare, or overlap with a provided list of genes associated with a particular pathology, or with any custom gene list or with any custom gene list as a .BED file.

Bionano Genome Imaging correctly identifies genetic disorders. In a publication in Genome Medicine,[16] professor Eric Vilain of Children's National Medical Center, Washington, DC, presents molecular diagnoses using optical mapping with Bionano Genome Imaging of patients with Duchenne Muscular Dystrophy (DMD). His team successfully mapped deletions, a duplication, and an inversion affecting the X-linked dystrophin gene, identifying deletions 45–250 kbp in size and an insertion of 13 kbp. The Bionano genome maps refined the location of deletion break points within introns. They detected heterozygous SVs in carrier mothers of DMD patients as well, demonstrating the ability of Bionano Genome Imaging to ascertain carrier status for large SVs. Vilain's team identified a 5.1 Mbp inversion involving the DMD gene, previously only identified by RNA sequencing of a muscle biopsy sample but missed by standard clinical methods (Figure 5).

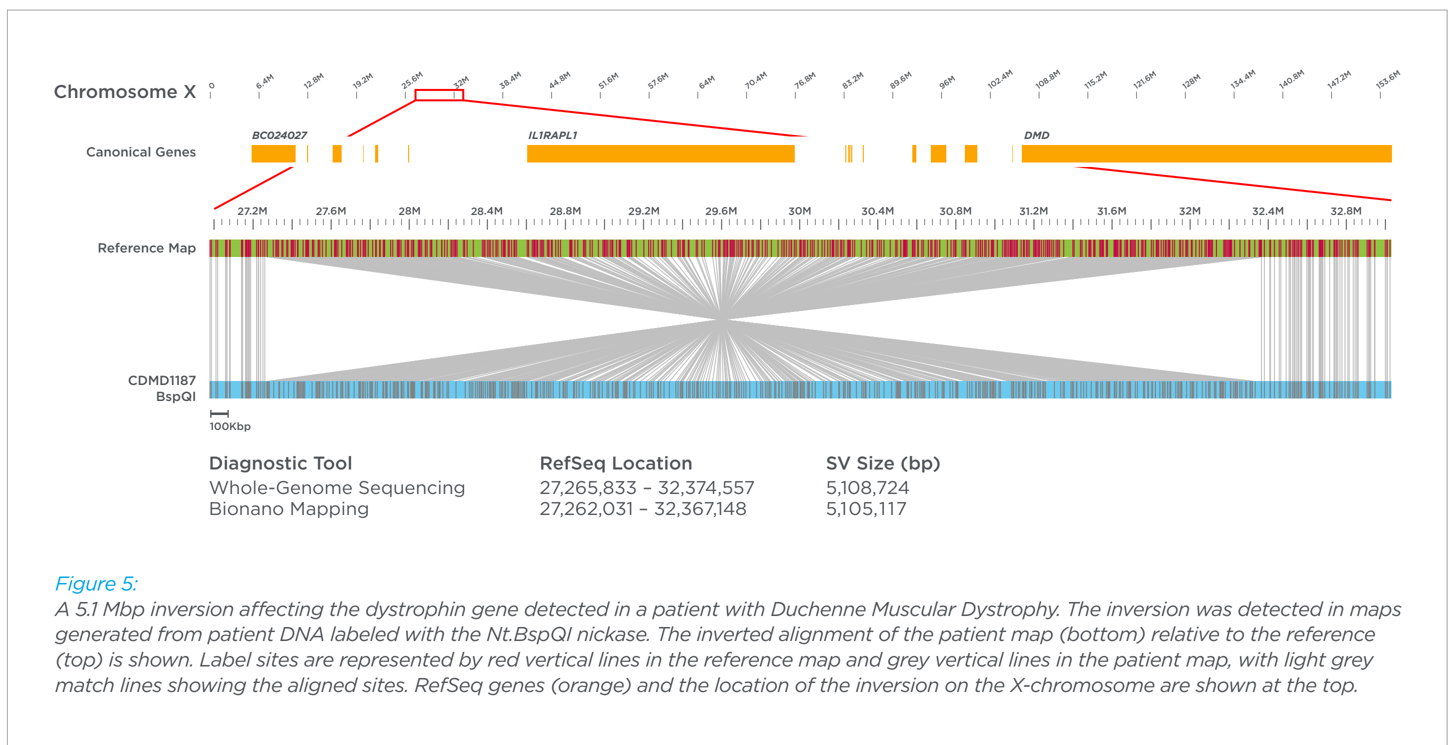

Figure 5:
A 5.1 Mbp inversion affecting the dystrophin gene detected in a patient with Duchenne Muscular Dystrophy. The inversion was detected in maps generated from patient DNA labeled with the Nt.BspQI nickase. The inverted alignment of the patient map (bottom) relative to the reference (top) is shown. Label sites are represented by red vertical lines in the reference map and grey vertical lines in the patient map, with light grey match lines showing the aligned sites. RefSeq genes (orange) and the location of the inversion on the X-chromosome are shown at the top.

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2020, Bionano Genomics Inc.

Diseases such as infertility and recurrent miscarriages, often caused by balanced SVs, are particularly difficult to diagnose and currently require a combination of technologies to clearly identify and dissect the events involved. In a communication at the 2019 edition of European Cytogeneticists Association Meeting, Dr Laila El Khattabi presented an evaluation of the performance of optical mapping with Bionano Genome Imaging and its workflow against the combination of karyotyping and chromosomal microarray.[17] The study included research patient samples bearing a variety of balanced and unbalanced SVs, previously identified by routine cytogenetics, in a context of reproductive or developmental disorders. An initial analysis detected the clinically relevant variants in 25 out of 27 samples, including several balanced translocations detected by Bionano that had been missed by whole genome sequencing. An example of a 5 Mbp inverted duplication flanked by a large deletion identified by Bionano technology is shown in Figure 6 along with the corresponding FISH result, and illustrates the challenge to detect large rearrangements from whole chromosome staining, and the improvement brought by Bionano Genome Imaging. The two SVs that were missed included a pericentromeric translocation that other molecular methods had not been able to detect, and one complex aberration that was partially resolved by Bionano.

Optical mapping with Bionano Genome Imaging is 100% concordant to conventional cytogenetics methods. A recent study, published on bioRxiv[18], found that Saphyr was 100% concordant with the standard of care for the detection of somatic chromosomal abnormalities. A total of 48 patient samples with a combination of myeloid and lymphoid leukemias, were analyzed using standard cytogenetic methods. When all 48 samples were subsequently analyzed with Bionano, the team was able to identify all previously reported aberrations. In addition, Bionano allowed for a better resolution and a more complete picture of complex aberrations. A complex chromothripsis structure was also resolved unambiguously and in other cases, additional fusions were identified, or marker chromosomes of unknown origin were resolved. Generally, optical mapping results were more complete than all three individual previous tests and most likely delivered the true underlying genomic architecture. In addition, the Bionano technology also found 23 potential gene fusions of which only 4 were previously observed. Several other large studies comparing the performance of optical mapping with Bionano Genome Imaging with cytogenetic standard of care tests are currently underway.

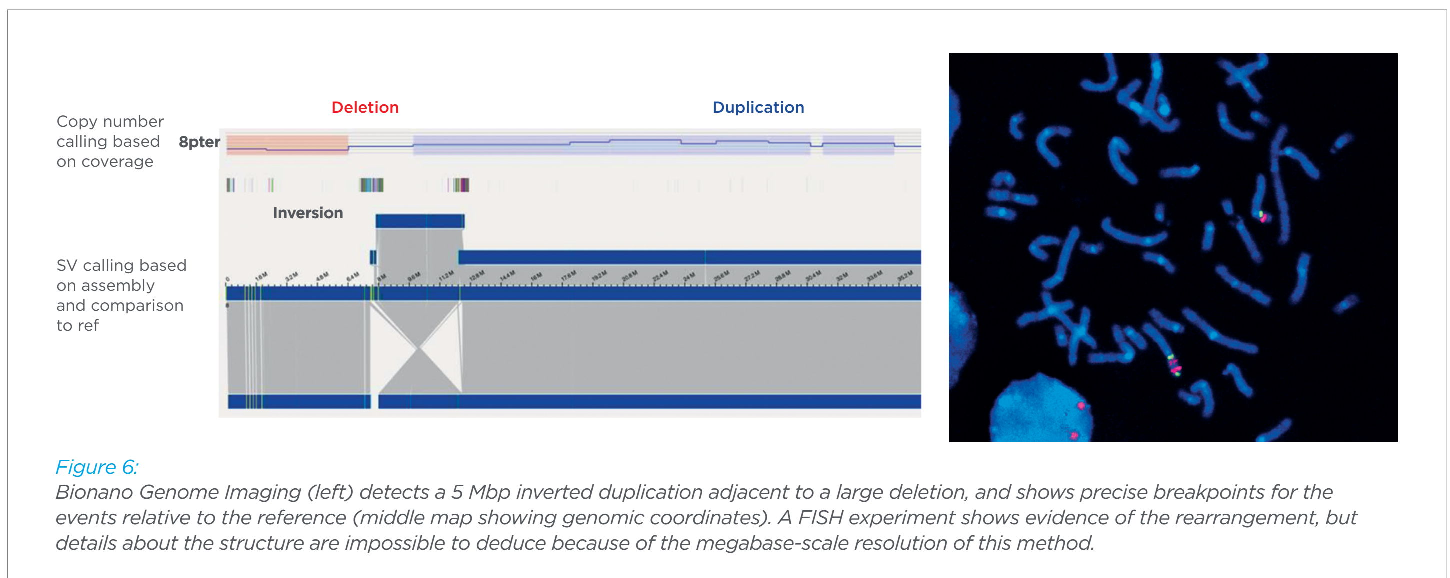

Figure 6:
Bionano Genome Imaging (left) detects a 5 Mbp inverted duplication adjacent to a large deletion, and shows precise breakpoints for the events relative to the reference (middle map showing genomic coordinates). A FISH experiment shows evidence of the rearrangement, but details about the structure are impossible to deduce because of the megabase-scale resolution of this method.

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2020, Bionano Genomics Inc.

Conclusion

Bionano Genome Imaging is the only technology that allows for the highly sensitive detection of all structural variant types, in an unbiased genome-wide manner. By providing a complete and unambiguous picture of the genome structure, it can possibly replace multiple cytogenetic tests that make up the gold standard.

Learn More

You can download detailed technical information about the Saphyr® Genome Imager and SV calling at the Products page on the Bionano Genomics® website: http://www.bionanogenomics.com

References: 1. Sudmant PH (2015). An integrated map of structural variation in 2,504 human genomes. Nature 526, pages 75–81. 2. Emanuel BS, S. T. (2001). Segmental duplications: an 'expanding' role in genomic instability and disease. Nat Rev Genet., 2 (10), 791-800. 3. Brand H, P. V. (2014). Cryptic and complex chromosomal aberrations in early-onset neuropsychiatric disorders. Am J Hum Genet., 95 (4), 454-61. 4. Fernandez TV, S. S.-S. (2012). Rare copy number variants in tourette syndrome disrupt genes in histaminergic pathways and overlap with autism. Biol Psychiatry, 71 (5), 392-402. 5. Butcher NJ, K. T. (2013). Association between early-onset Parkinson disease and 22q11.2 deletion syndrome: identification of a novel genetic form of Parkinson disease and its clinical implications. JAMA Neurol., 70 (11), 1359-66. 6. Crawford DC, P. Z. (2008). LPA and PLG sequence variation and kringle IV-2 copy number in two populations. Hum Hered., 66 (4), 199-209. 7. Bittel DC, Z. X. (2014). Ultra high-resolution gene centric genomic structural analysis of a non-syndromic congenital heart defect, Tetralogy of Fallot. PLoS One, 9 (1), e87472. 8. Bochukova EG, H. N.-S.-S. (2010). Large, rare chromosomal deletions associated with severe early-onset obesity. Nature, 463 (7281), 666-70. 9. Rasmussen HB, D. C. (2012). Genome-wide identification of structural variants in genes encoding drug targets: possible implications for individualized drug therapy. Pharmacogenet Genomics, 22 (7), 471- 83. 10. Huddleston J, Chaisson MJ (2016). Genome Res. Nov 28. pii: gr.214007.116. 11. Hane Lee, J. L.-R.-A. (2014). Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. JAMA, 312 (18), 1880- 1887. 12. Chaisson MJ (2019). Multi-platform Discovery of Haplotype-resolved Structural Variation in Human Genomes. Nature Communications 10, Article 1784. 13. Hastings (2019). European guidelines for constitutional cytogenomic analysis, European Journal of Human Genetics. 14. Levy-Sakin M (2019). Genome maps across 26 human populations reveal population-specific patterns of structural variation. Nature Communications 10, Article 1025. 15. Dai Y (2019). Single-molecule optical mapping enables quantitative measurement of D4Z4 repeats in facioscapulohumeral muscular dystrophy (FSHD). J Med Genet. 2019 Sep 10. 16. Barseghyan H (2017). Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis. Genome Med. 2017; 9: 90. 17. https://bionanogenomics.com/videos/eshg-2019-workshop-series-2/ 18. Neveling et al.bioRxiv. Next generation cytogenetics: comprehensive assessment of 48 leukemia genomes by genome imaging. https://doi.org/10.1101/2020.02.06.935742

19. Reference Theory of Operation:
Bionano Solve Theory of Operation: Structural Variant Calling (Document 30110).
https://bionanogenomics.com/support-page/data-analysis-documentation/

For more information about Bionano Genome Imaging, please contact
info@bionanogenomics.com | 858.888.7600 | bionanogenomics.com

For Research Use Only. Not for use in diagnostic procedures.
Bionano Genomics®, Saphyr®, Saphyr Chip®, Bionano Access®, Bionano EnFocus™ and Irys® are trademarks of Bionano Genomics Inc. All other trademarks are the sole property of their respective owners. © 2020 Bionano Genomics, Inc.

Part #30126 Rev I