Bionano Genome Mapping Identifies Large Structural Variants in Cancer and Genetic Disorders

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

Existing technologies including chromosomal microarrays and Whole Genome Sequencing (WGS) diagnose less than 50% of patients with genetic disorders. This leaves a majority of patients without ever receiving a molecular diagnosis.[1,2] Undiagnosed disorders are individually rare but their combined incidence and the associated diagnostic odyssey, with resultant delays in treatment, are a drain on families and the healthcare system. Many of these diseases remain medical mysteries with no root cause or clear basis for treatment.

Two thirds of the human genome consists of repetitive sequences. Exome sequencing accesses just 1.5% of the genome,[3] and Whole Genome Sequencing (WGS) does not align correctly with the repetitive parts of the genome. The most common repetitive sequences in the genome are LINEs, SINEs, retrotransposons and segmental duplications (Figure 1). The short-read sequences Next-Generation Sequencing (NGS) provides, map with poor accuracy to these repeats. Alignment algorithms typically fail to identify the exact genomic location to align these short-reads to. When they do align, the limited 100-150 bp read length and spacing of paired-end reads does not allow for a correct sizing of larger repeats.

Structural variants make up the majority of human genomic variation, but Next-Generation Sequencing technology cannot correctly identify them. Clinical exome sequencing solves approximately 30% of rare diseases.[2] NGS, consisting of Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS) reliably identify single nucleotide variants and small insertions and deletions. However, NGS relies on short-read sequences that are mapped to a reference human genome and fails to identify most large insertions, deletions, and copy-number variations in repetitive regions of the genome. 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. Together, structurally variable regions cover 13% of the genome with individuals showing structural variation of as much as 30 Mbp.[4]

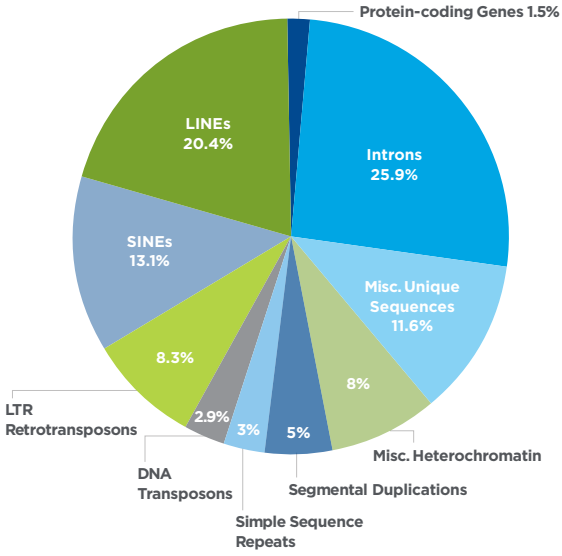


Figure 1:
Repetitive elements in the human genome.[1]

Structural variations (SVs) cause genetic disorders and play an important role in cancer.

The importance of the role of highly repetitive sequences and large structural variations in genetic disorders cannot be overstated. For many known syndromes, clinically relevant large SVs are well characterized. These SVs can be deletions in Prader-Willi syndrome, DiGeorge syndrome and Williams-Beuren syndrome; they can be duplications as in Charcot-Marie-Tooth disease; or inversions such as in Haemophilia A.[5] More recently, large SVs have been found to play a role in neurological diseases, like early-onset neuropsychiatric disorders,[6] Tourette syndrome,[7] and Parkinson's disease;[8] and in coronary heart disease[9] and congenital heart disease.[10] SVs influence obesity[11] and pharmacogenetics.[12]

Cancer cells typically show extreme rearrangements of the genome. Tumors can grow seemingly unrestrictedly by altering a large number of growth factors, growth inhibitors, and cell cycle control points. This happens by making changes in the coding sequence of genes, by altering DNA methylation patterns, and by rearrangements of the genome. SVs are found in most cancer types. Examples are FGFR3-IGH fusion genes in Multiple Myeloma caused by a translocation and deletion, or the Philadelphia Chromosome found in chronic myeloid leukemia.[13]

Bionano genome mapping is the only technology that detects all SV types, homozygous and heterozygous, starting at 500 bp up to millions of bp. Megabase size molecules of genomic DNA are labeled, linearized and uniformly stretched in high density Nanochannel arrays, and imaged on the Saphyr™ or Irys® System. Two different labeling technologies are available for Bionano mapping. In the first—Nick, Label, Repair and Stain (NLRS)—a nicking endonuclease recognizes a specific 6 or 7 basepair sequence and creates a single-strand nick, which is filled with fluorescent nucleotides, repaired, and stained. Bionano's newest 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 16 labels per 100 kbp in the human genome. The label patterns allow each long molecule 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.

Bionano maps are built completely *de novo*, without any reference guidance or bias. This differentiates Bionano from NGS, where short-read sequences 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, like Bionano maps, allow for a completely unbiased, accurate assembly.

Bionano simplifies genome structure analysis. With DLS, genome maps typically reach chromosome-arm lengths. This means that complex genomic rearrangements, like the one shown in Figure 2, 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.

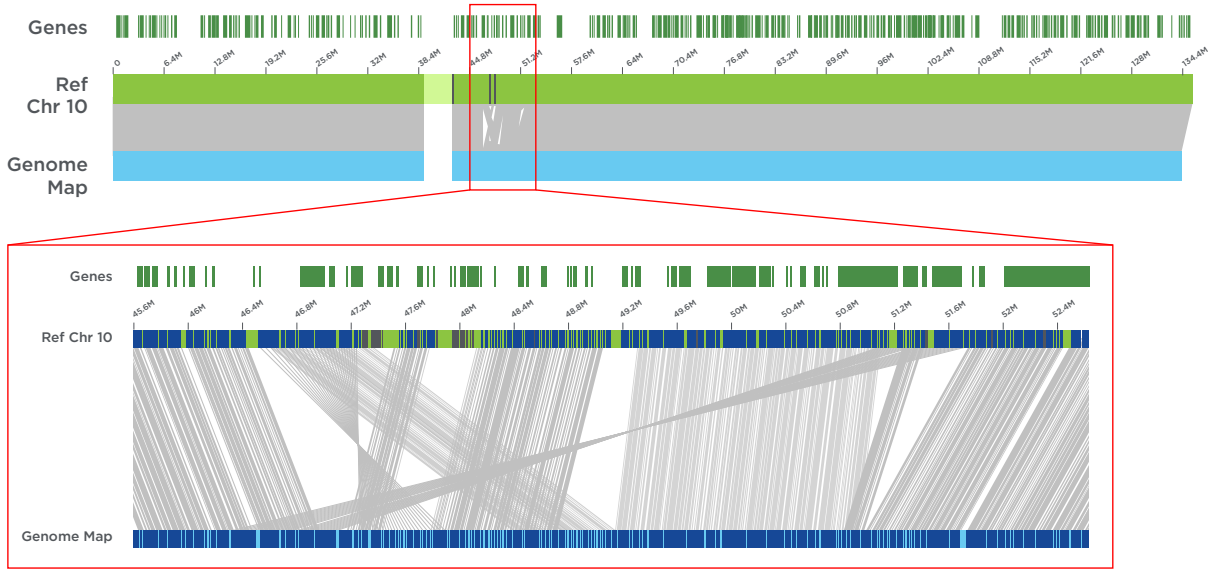


Figure 2:
DLS assembles the entire human chromosome 10 in two single maps, separated by the centromere (top). Gene positions are shown in green. Insert shows multiple large rearrangements assembled into a single map in the 10q11.21 region involved in intellectual disability/developmental delay.

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, with mixed success. With Bionano mapping, megabase-size native DNA molecules are imaged, and most large SVs or their breakpoints can be **observed** directly in the label pattern on the molecules. If a native-state DNA molecule with a specific SV exists, then that SV call cannot be wrong.

Bionano algorithms call SVs by comparing genome structures. To identify a structural variation, a *de novo* genome map assembly can be aligned to a reference genome, or two samples can be aligned to each other directly. 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. 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.

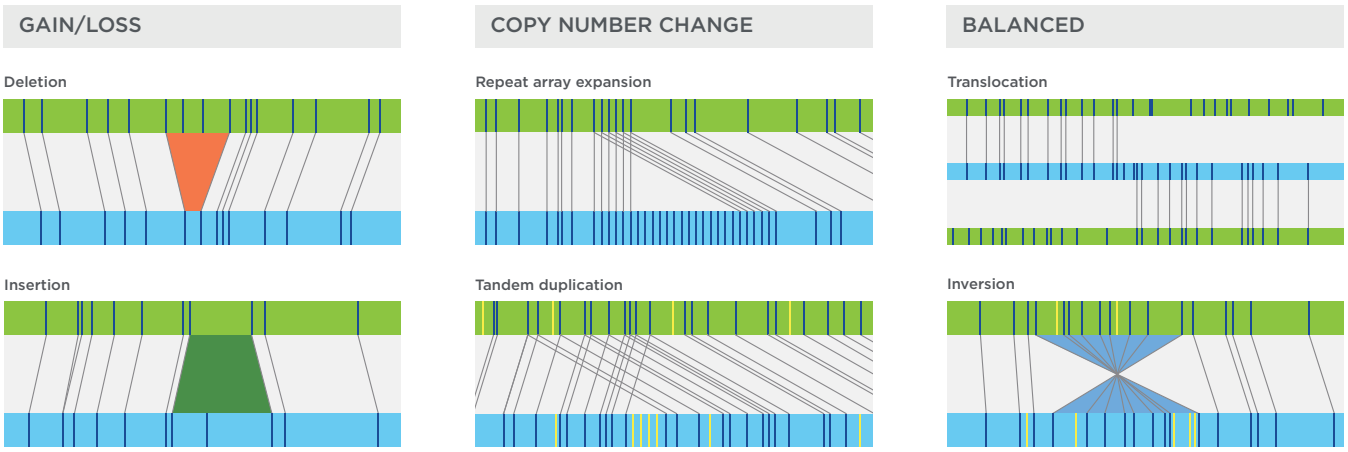


Figure 3:
Structural variant types detected by Bionano mapping. SVs are identified by comparing label patterns in the sample of interest (blue) with those in the reference genome, or in a reference sample (green). Major types detected are:

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, are called as insertions.*

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.*

Large copy number variants (CNV) are detected by an independent molecule-based tool. A per-label 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 500 kbp (Figure 4). The copy number tool has a sensitivity of 97% and PPV of 94% for those large duplications and deletions.

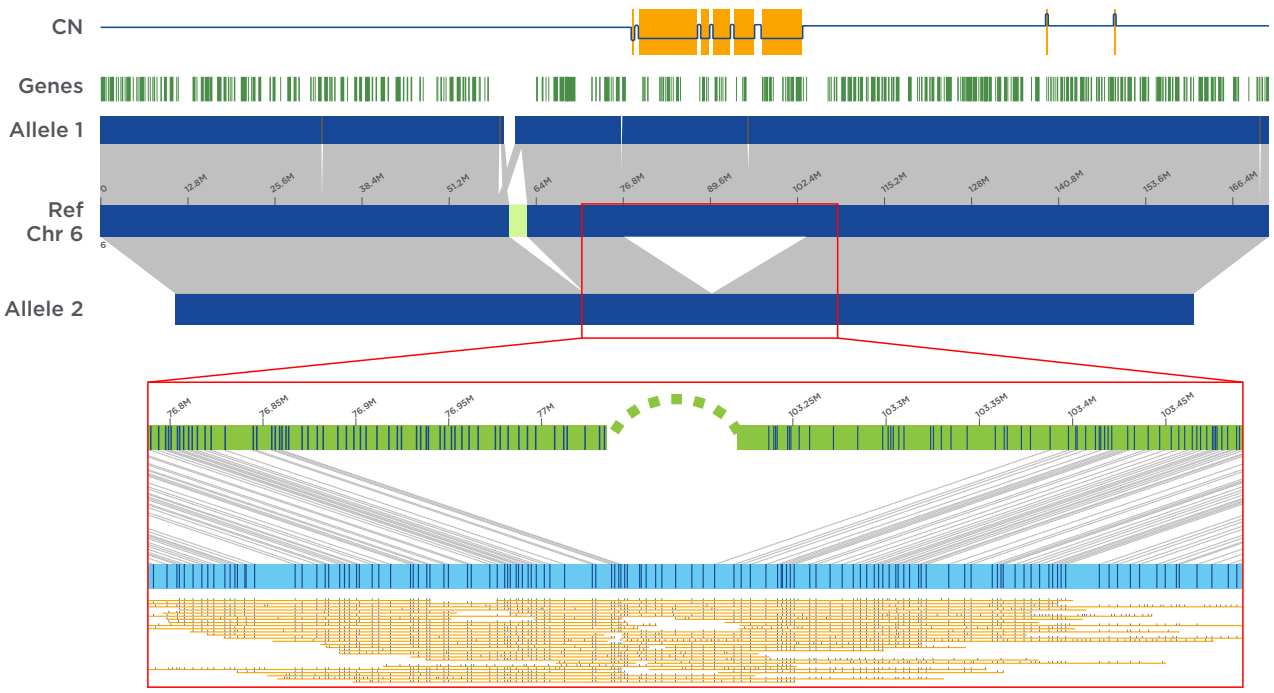


Figure 4:
Chromosome arm-length maps and new copy number (CN) profiling provide a robust tool for large scale SV detection and visualization, here showing a 27 Mbp deletion found in the genome of a leukemia patient. The same deletion is detected independently from the consensus genome map as well (blue map aligned to reference in green). Molecules aligning to the genome map are shown below.

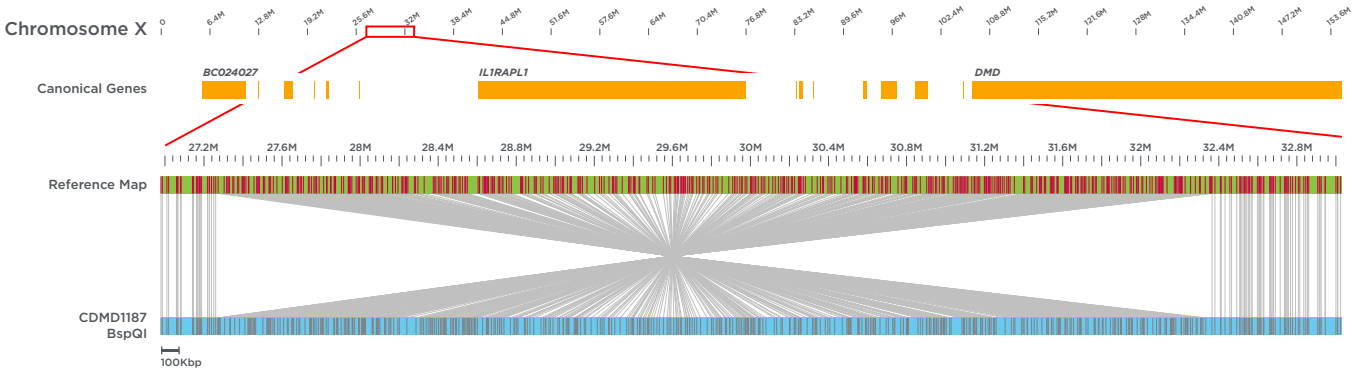
Bionano Genome Mapping has successfully identified large structural variants of clinical significance in genetic disorders and cancer.

Bionano mapping correctly diagnoses genetic disorders: In a publication in Genome Medicine,[14] professor Eric Vilain of Children's National Medical Center, Washington, DC, presents molecular diagnoses using Bionano mapping 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 maps refined the location of deletion break points within introns compared to current PCR-based clinical techniques. They detected heterozygous SVs in carrier mothers of DMD patients as well, demonstrating the ability of Bionano mapping 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).

Bionano mapping identifies genomic rearrangements in prostate cancer: Professor Vanessa Hayes at the Garvan Institute of Medical Research published a complete tumor-normal comparison from a primary prostate cancer.[15] Her team identified 85 large somatic deletions and insertions, of which half directly impact potentially oncogenic genes or regions.

Only one-tenth of these large SVs were detected using high-coverage short-read NGS and bioinformatics analyses using a combination of the best SV calling algorithms for NGS data. A manual inspection of NGS reads corresponding with the Bionano derived target regions verified 94% of the total SVs called with Bionano mapping. Many SVs detected with Bionano



Diagnostic Tool	RefSeq Location	SV Size (bp)
Whole-Genome Sequencing	27,265,833 – 32,374,557	5,108,724
Bionano Mapping	27,262,031 – 32,367,148	5,105,117

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.

were flanked by repetitive sequences, making them all but invisible to short-read sequencing.

Bionano Genome Mapping detects structural variants with a sensitivity and specificity far greater than NGS

Bionano detects seven times more SVs larger than 5 kbp compared to NGS. Professor Pui-Yan Kwok at the University of California, San Francisco, demonstrated the robustness of Bionano mapping for genome-wide discovery of SVs in a trio from the 1000 Genomes Project.[16] Since high quality NGS data on these samples is publicly available, structural variation analysis using short-read data has been performed with over a dozen different algorithms. Using Bionano maps, hundreds of insertions, deletions, and inversions greater than 5 kbp were uncovered, 7 times more than the large SV events previously detected by NGS (Figure 6). Several are located in regions likely leading to disruption of gene function or regulation.

Bionano has exceptional sensitivity and specificity to detect insertions and deletions over a wide size range as demonstrated using simulated data. Insertions and deletions were randomly introduced into an *in silico* map of the human reference genome hg19. The simulated events were at least 500 kbp from each other or N-base gaps. They ranged from 200 bp to 1 Mbp, with smaller SVs more frequent than larger ones.

Based on the edited and the unedited hg19, molecules were simulated to resemble actual molecules collected on a Bionano system and mixed such that all events would be heterozygous. A dataset with 70x effective coverage was generated. The simulated molecules were used as input to the Bionano Solve™ pipeline, SV calls were made, and compared to the ground truth. Figure 7 shows sensitivity and positive predicted value (PPV) for heterozygous insertions and deletions within a large size range.

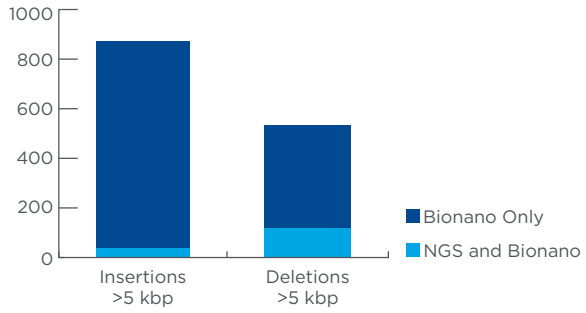


Figure 6:
Bionano mapping was used to detect insertions and deletions >5 kbp in the CEPH trio. The total number of large insertions and deletions detected in the three individuals is shown. The number of SVs found using Bionano was 7x larger than those already detected from NGS data by over a dozen algorithms.

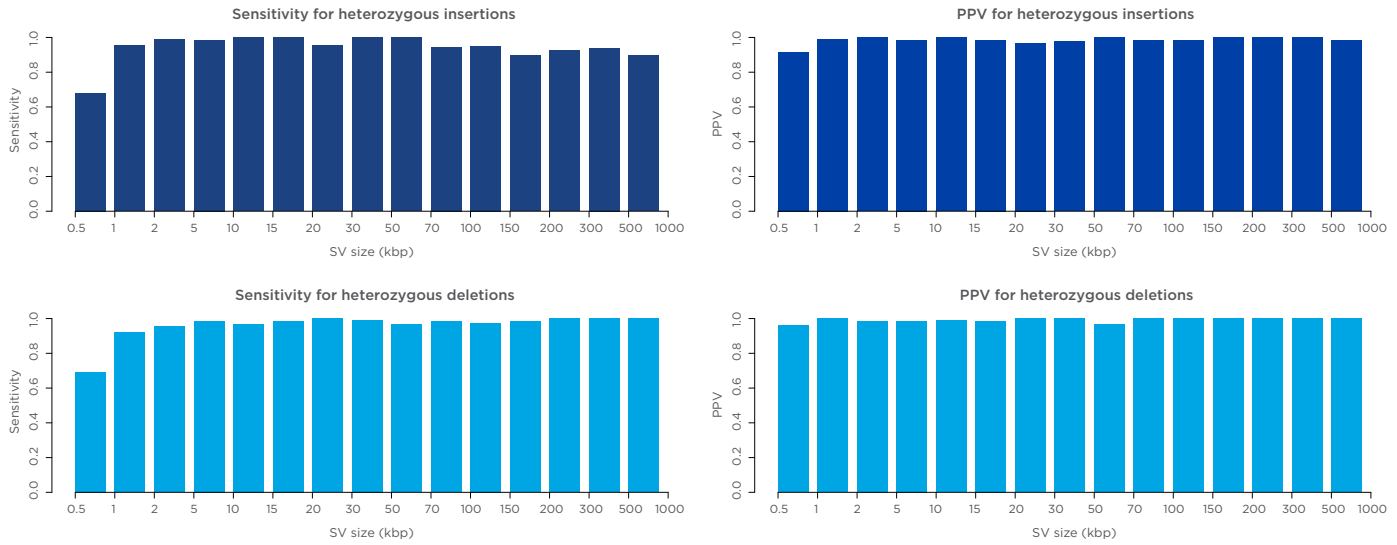


Figure 7:
Heterozygous SV calling performance from a simulated dataset. Molecules were simulated from unedited and edited versions of hg19 (with insertions and deletions of different sizes) and used for assembly and SV calling.

Bionano mapping has exceptional sensitivity and specificity to detect heterozygous insertions and deletions over a wide size range as demonstrated using experimental data. Since there is no perfectly characterized human genome that can be considered the ground truth, a diploid human genome was simulated by combining data from two hydatidiform mole derived cell lines. These moles occur when an oocyte without nuclear DNA gets fertilized by a sperm. The haploid genome in the sperm gets duplicated, and the cell lines resulting from this tissue (CHM1 and CHM13) are therefore entirely homozygous.

Structural variants detected in the homozygous cell lines were considered the (conditional) ground truth. An equal mixture of single molecule data from two such cell lines was assembled to simulate a diploid genome, and SV calls made from this mixture were used to calculate the sensitivity to detect heterozygous SVs.

Table 1 shows the number of insertions and deletions larger than 500 bp detected in the CHM1 and CHM13 homozygous cell lines relative to the reference, and the *in silico* CHM1/13 mixture. SVs detected in CHM1 only or CHM13 only are heterozygous and those detected in both are homozygous. Bionano has a sensitivity of 88% for heterozygous deletions and 81% for heterozygous insertions larger than 500 bp. The largest detected deletion was 4.28 Mbp in size and the largest insertion 412 kbp.

A similar experiment on PacBio long-read sequencing was described recently.[17] Structural variants were called with the SMRT-SV algorithm in CHM1 and CHM13, and compared to those called in an equal mixture of both. The sensitivity to detect homozygous SVs using PacBio was 87%, compared to 99.1% using Bionano. The sensitivity to detect heterozygous SVs using PacBio was only 41%, which is less than half the 84% sensitivity for heterozygous SV detection using Bionano. Even when the PacBio SV calls were limited to insertions and deletions larger than 500 bp, the sensitivity for homozygous SVs was only 81%, and for heterozygous SVs 49% (Table 1).

Duplications over 30 kbp, in direct or inverse orientation, are called with a sensitivity of 97%. Smaller duplication events are detected as well, but if the identity of the inserted sequence cannot be determined they will be labeled insertions.

Bionano genome mapping detects 99% of large inversions. Inversions are the invisible variants, and have traditionally been the hardest to detect structural events. They are balanced, without gain or loss of sequence, and unlike translocations they don't create easily visible changes in genomic context. Inversions often escape detection by traditional cytogenetic techniques. Chromosomal Microarray cannot identify balanced events, and metaphase chromosome spreads can only

visualize some megabase size inversions. Next Generation Sequencing approaches tend to miss inversions because reads from inside the inversion map back to the reference without any indication that the orientation has changed. Detection of the breakpoints often fails, especially if the inversion is flanked by segmental duplications, repeat arrays or other non-unique sequences.

Bionano's imaging of extremely long molecules overcomes these obstacles to identifying inversions. Simulations of thousands of heterozygous inversions of various sizes demonstrated that our SV detection algorithms have 99% sensitivity to detect inversions larger than 30 kbp throughout the genome, and determine the inversion breakpoints within a median distance of 3 kbp from the actual location.

	PacBio				Bionano			
	CHM1 and CHM13 assemblies	Mixture assembly	Sensitivity	PPV	CHM1 and CHM13 assemblies	Mixture assembly	Sensitivity	PPV
Homozygous Insertions	1523	1225	80.4%	96.3%	2278	2261	99.3%	96.3%
Heterozygous Insertions	1487	590	39.7%		2130	1730	81.2%	
Homozygous Deletions	431	355	82.4%	95.0%	740	731	98.8%	95.5%
Heterozygous Deletions	1019	632	62.0%		1391	1223	87.9%	

Table 1:
Two homozygous cell lines, CHM1 and CHM13 were independently de novo assembled and insertions and deletions >500 bp called. Raw data was mixed together, assembled and SVs called (Mixture assemblies column). The sensitivity and positive predictive value (PPV) to detect heterozygous relative to homozygous SVs is shown.

Bionano far outperforms other technologies in the detection of translocations. Thousands of translocations were simulated similarly to insertions and deletions in an *in silico* map of the human reference genome hg19. The sensitivity for heterozygous translocations was shown to be 95% for breakpoint detection in both balanced and unbalanced translocations over 50 kbp. Genome mapping can define the true positions of breakpoints within a median distance of 3.5 kbp, which is approximately 1,000 times more precise than karyotyping and FISH. This accuracy is often sufficient for PCR and sequencing if single nucleotide resolution of the fusion point is desired for subsequent gene function studies.

In addition, translocation detection sensitivity was verified in a large number of samples including two reference samples, NA16736 and NA21891, which are lymphoblast cell lines produced from blood cells from patients.

One patient had a developmental disorders resulting in deafness with DNA repair deficiency caused by a t(9;22) translocation, and a second patient had Prader-Willi syndrome associated with a t(4;15) translocation. Both cell lines had been characterized by traditional cytogenetic methods. Bionano was able to detect both expected translocations as well as the reciprocal translocation breakpoints. Additionally, NA16736 contained a t(12:12) rearrangement which flanked an inverted segmental duplication. In NA21891, one translocation breakpoint could be localized within a gene, resulting in a predicted truncation (Figure 8).

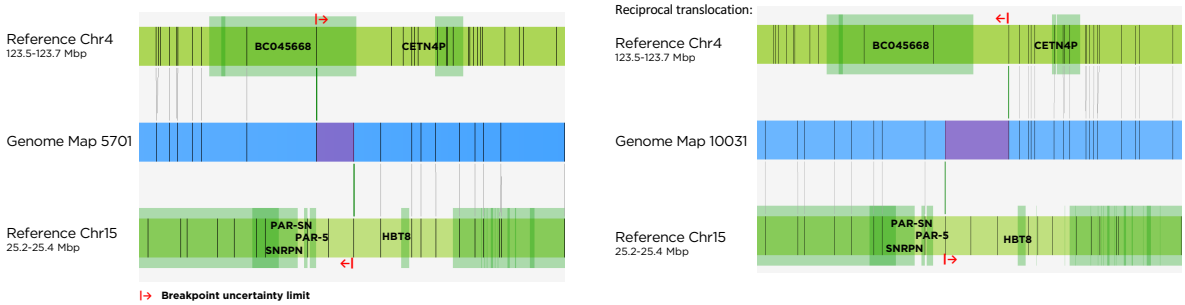


Figure 8:
Example of a translocation detected by Bionano mapping, associated with Prader-Willi syndrome. Blue bars are Bionano maps, and vertical lines represent Nt.BspQI label sites. For each of the reciprocal translocation breakpoints, maps are shown with alignments of the maps to chromosome 4 (top) and chromosome 15 (bottom) of the human reference hg19. Breakpoint resolution can be determined by the distance between matched and unmatched labels.

Bionano's Variant Annotation Pipeline reduces thousands of structural variants to a handful of candidate genes. The Variant Annotation Pipeline (VAP), part of the Bionano Access™ software, streamlines family-based and case-control studies. Using VAP, structural variation calls from multiple samples can be analyzed as a group to detect inherited and *de novo* SVs when comparing variant calls from the parent(s) with those from the child, or to detect somatic mutations when comparing calls from the tumor with the blood from the same patient.

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. 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 Saphyr™, 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.

Conclusion

Bionano genome mapping identifies large structural variants in human genomes. All types of large structural events are detected as heterozygous or homozygous variants with unrivaled sensitivity and specificity.

Learn More

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

References: 1.Miller DT, A. M. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet., 86 (5), 749-64. 2.Hane Lee, J. L.-R.-A. (2014). Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. JAMA, 312 (18), 1880-1887. 3.de Koning AP, G. W. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet, 7 (12), e1002384. 4.Peter H. Sudmant, T. R.-Y.-W.-B. (2016). An integrated map of structural variation in 2504 human genomes. Nature, 526, 75-81. 5.Emanuel BS, S. T. (2001). Segmental duplications: an 'expanding' role in genomic instability and disease. Nat Rev Genet., 2 (10), 791-800. 6.Brand H, P. V. (2014). Cryptic and complex chromosomal aberrations in early-onset neuropsychiatric disorders. Am J Hum Genet., 95 (4), 454-61. 7.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. 8.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. 9.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. 10.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. 11.Bochukova EG, H. N.-S.-S. (2010). Large, rare chromosomal deletions associated with severe early-onset obesity. Nature, 463 (7281), 666-70. 12.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. 13.Krem MM, P. O. (2015). Mechanisms and clinical applications of chromosomal instability in lymphoid malignancy. Br J Haematol, 171 (1), 13-28. 14.Barseghyan H, W.T, (2017). Next-Generation Mapping (NGM): A Novel Approach for Clinical Genetic Diagnosis of Structural Variants. Genome Medicine, in press. 15.Jaratlerdsiri W, C. E (2017). Next generation mapping reveals novel large genomic rearrangements in prostate cancer. Oncotarget. 2017; 8:23588-602. 16.Mak AC, L. Y. (2016). Genome- Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays. Genetics, 202 (1), 351-62. 17.https://www.ncbi.nlm.nih.gov/pubmed/27895111 Huddleston J, Chaisson MJ, Meltz Steinberg K, Warren W, Hoekzema K, Gordon DS, Graves-Lindsay TA, Munson KM, Kronenberg ZN, Vives L, Peluso P, Boitano M, Chin CS, Korlach J, Wilson RK, Eichler EE. Genome Res. 2016 Nov 28. pii: gr.214007.116