Bionano Genomics' Next-Generation 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 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.

To close this diagnostic sensitivity gap and get a better understanding of the genetic causes of disease, we need better tools to access the entire genome, and large translational research studies to apply these tools to the discovery of novel biomarkers. Genetic disorders for which no molecular basis is currently known are either caused by genomic events that are poorly detected with current technology, events occurring in inaccessible parts of the genome, or a combination of events that is too complex to analyze using existing tools. Better molecular tools are needed to analyze the entire range of genomic variations. Armed with such tools, large translational research studies are needed to identify disease correlated biomarkers spanning all genomic variants in patients with genetic disorders.

Two thirds of the human genome consists of repetitive sequences. Exome sequencing accesses just 1.5% of the genome,[3] **and Wh ole 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.

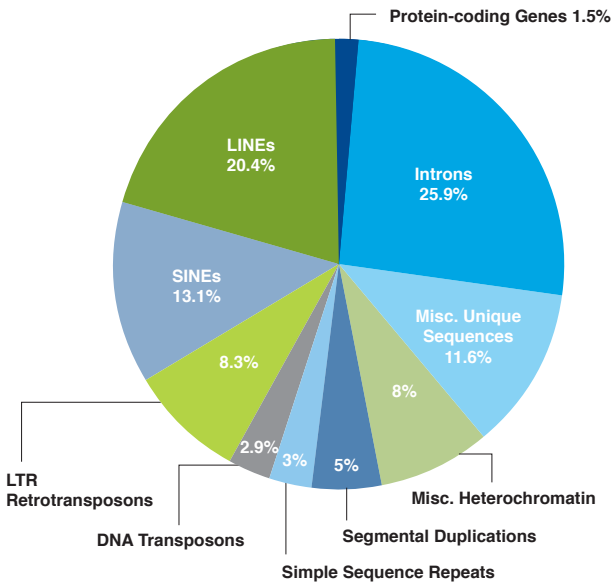

Figure 1:
Repetitive elements in the human genome.[1]

Structural variants make up the majority of human genomic variation, but Next-Generation Sequencing technology can't correctly identify them. Clinical exome sequencing solves about 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, or copy-number variations in repetitive regions of the genome. It is incapable of easily detecting other structural variations (SVs) such as inversions and translocations. Together, structural variable regions cover 13% of the genome and individuals show structural variation covering as much as 20-30 Mbp between each other.[4]

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

The importance of the role of large structural variants 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 Genomics' Next-Generation Mapping (NGM) is the only technology that can show you all SV types, homozygous and heterozygous, starting at 1000 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. Using a nicking endonuclease, a specific 6 or 7 basepair sequence is labeled approximately 10 times per 100 kbp. 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 can be created using different endonucleases to generate broader coverage and higher label density.

Bionano maps are built completely de novo, without any reference guidance or bias. This differentiates NGM 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's SVs are observed, 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 NGM, megabase-size native DNA molecules are imaged, and most large SVs or their breakpoints (in the case of inter-chromosomal translocations) 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 Next-Generation Mapping has successfully identified large structural variants of clinical significance in genetic disorders and cancer.

NGM correctly diagnoses genetic disorders: Professor Eric Vilain at the University of California, Los Angeles, presents the first molecular diagnoses using Bionano NGM of patients with Duchenne Muscular Dystrophy. His team successfully mapped the intronic breakpoints within the Dystrophin gene, identifying deletions 30-300 kbp and insertions ranging from 10-150 kbp. Multiplex ligation-dependent probe amplification (MLPA) was used to validate the size and location of the SVs. The deletions were found in heterozygous (carrier) mothers as well, as illustrated in Figure 2.

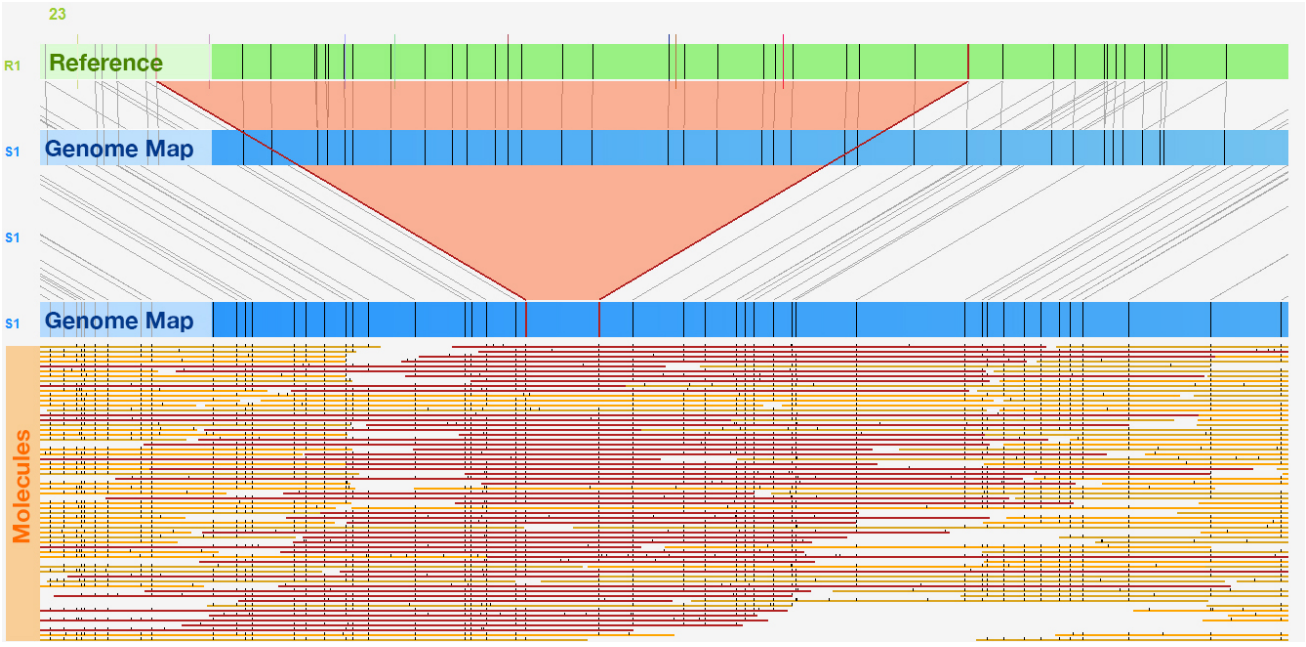

Figure 2:
250 kbp deletion detected in the mother of a patient with Duchenne Muscular Dystrophy. Vertical lines represent the label patterns detected on long genomic DNA molecules from the mother (blue) aligned to the reference (hg 19, green). The top blue genome map aligns perfectly with the reference above. The bottom blue bar shows a 250 kbp deletion (orange triangle) of a part of the Dystrophin gene. Individual molecules imaged in the mother showing the deletion are displayed as horizontal orange and red lines below the assembled genome map.

NGM maps complete genomic rearrangement in prostate cancer: Professor Vanessa Hayes at the Garvan Institute of Medical Research reports the first complete tumor-normal map from a primary prostate cancer. Her team identified 85 large somatic deletions and insertions, of which half directly impact genes or gene regions. One such insertion, disrupting a gene known to be involved in cancer, is shown in Figure 3.

Only one-tenth of these large SVs were detectable using high-coverage short-read NGS and automated five-tooled bioinformatic analyses. A manual inspection of NGS reads corresponding with the Bionano derived target regions and *de novo* assembled scaffolds verified 94% of the total SVs called with NGM.

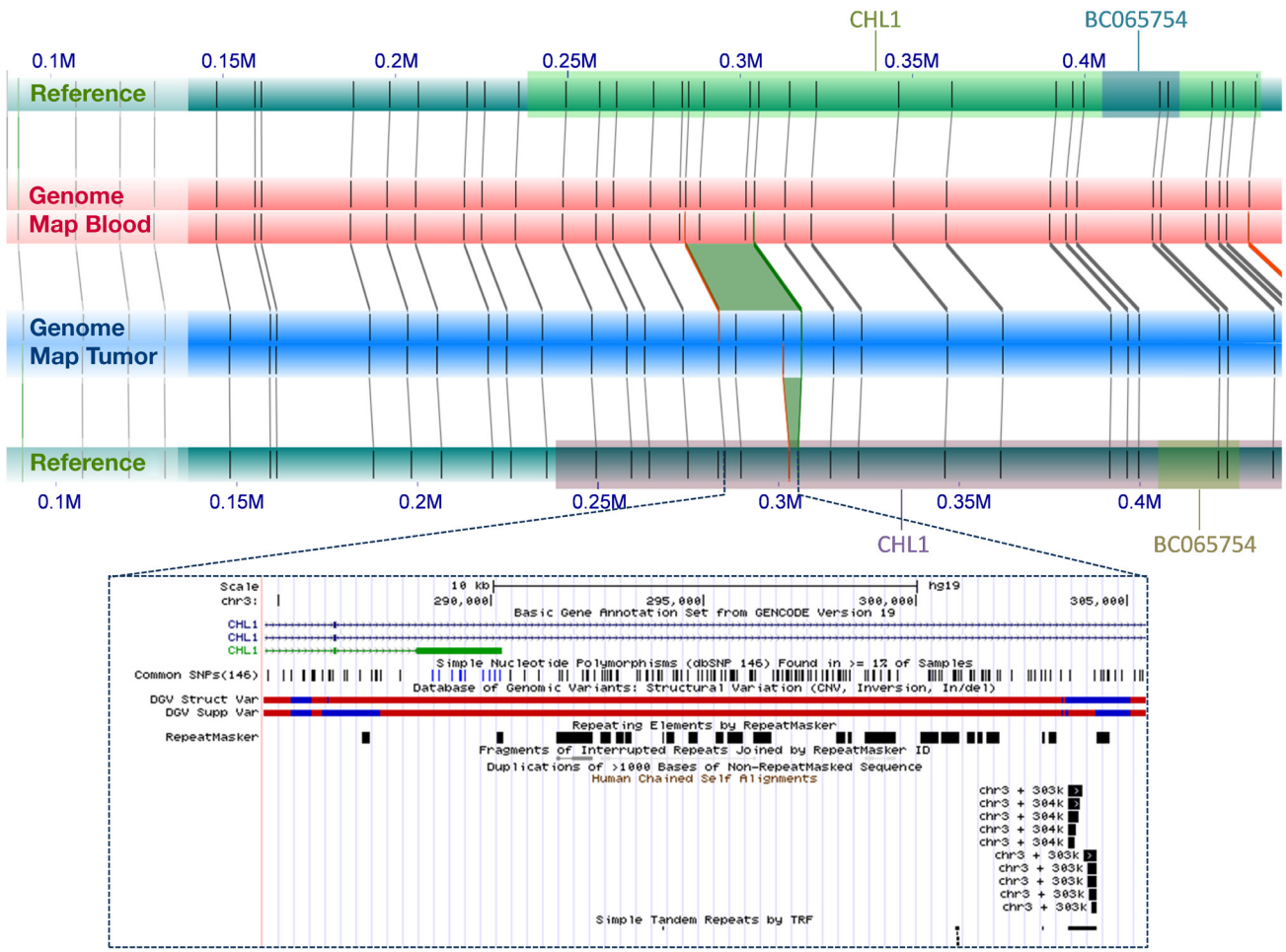

Figure 3:
A 4-kbp somatic insertion within the CHL1 gene on chromosome 3 identified in the prostate tumor of UP2153 using NGM. The tumor map (blue track) shows a 2.5-kbp insertion (Chr3: 302.9 - 305.4 kbp) relative to hg19 (blue track), defined by a tandem repeat interval (inset). However, direct comparison of the tumor to genome maps derived from blood of the same patient (red track) found a larger 4-kbp insertion.

Bionano Next-Generation Mapping detects structural variants with a sensitivity and specificity far greater than NGS

NGM 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 NGM for genome-wide discovery of SVs in a trio from the 1000 Genomes Project.[14] 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 3). Several are located in regions likely leading to disruption of gene function or regulation.

NGM 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. Two sets of molecules were simulated, each labeled with a different nicking endonuclease. Datasets with 70x effective coverage were generated. The simulated molecules were used as input to the Bionano Solve™ pipeline and SV calls were made by combining the single-enzyme SV calls from both nicking endonucleases using the SV Merge algorithm. SV calls were compared to the ground truth.

Figure 4 shows sensitivity and positive predicted value (PPV) for heterozygous insertions and deletions within a large size range. SV size estimates were typically within 500 bp of the actual SV sizes, while reported breakpoints were typically within 10 kbp of the actual breakpoint coordinates. Additional large insertions (>200 kbp) were found but classified as end-calls.

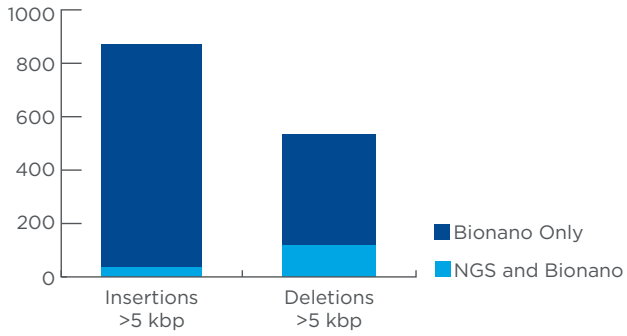

Figure 3:
NGM was used to detect insertions and deletions >5 kbp in the CEPH trio. The total number of large insertions and deletions detected in the three individuals is shown. The number of SVs found using Bionano was 7x larger than those already detected from NGS data by over a dozen algorithms.

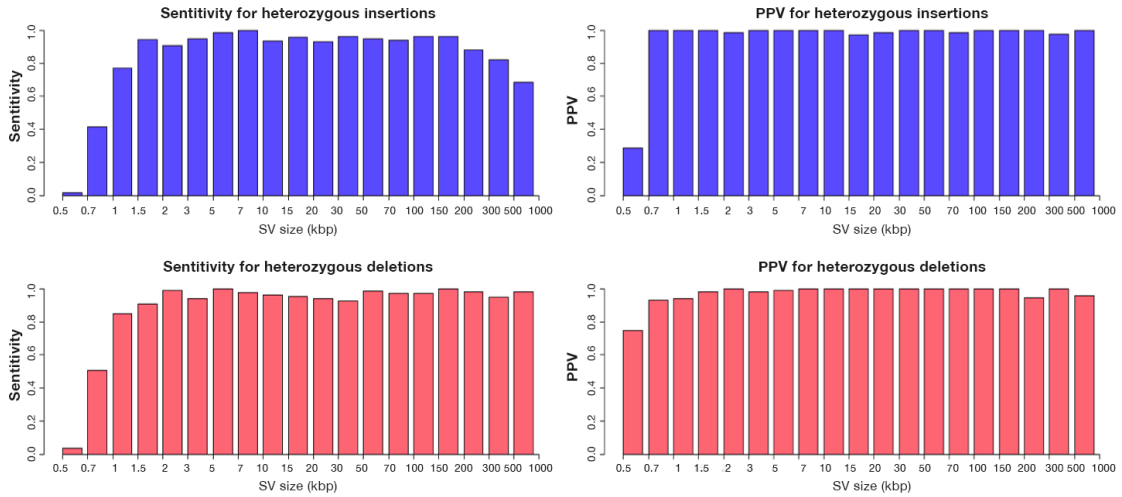

Figure 4:
Heterozygous SV calling performance from a simulated dataset. Molecules were simulated from unedited and edited versions of hg19 (with insertions and deletions of different sizes) and used for assembly and SV calling.

NGM 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 1.5 kbp 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. NGM has a sensitivity of 92% for heterozygous deletions and 84% for heterozygous insertions larger than 1.5 kbp. 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.[15] 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 99.2% using Bionano. The sensitivity to detect heterozygous SVs using PacBio was only 41%, which is less than half the 86% sensitivity for heterozygous SV detection using Bionano. Even when the PacBio SV calls were limited to insertions and deletions larger than 1.5 kbp, the sensitivity for homozygous SVs was only 78%, and for heterozygous SVs 54% (Table 1).

	PacBio				Bionano			
	CHM1 and CHM13 assemblies	Mixture assembly	Sensitivity	PPV	CHM1 and CHM13 assemblies	Mixture assembly	Sensitivity	PPV
Homozygous Insertions	467	353	75.6%	96.1%	707	700	99.0%	97.9%
Heterozygous Insertions	586	252	43.0%		663	554	83.6%	
Homozygous Deletions	221	183	82.8%	94.9%	269	268	99.6%	97.1%
Heterozygous Deletions	501	337	67.3%		517	477	92.3%	

Table 1:
Two homozygous cell lines, CHM1 and CHM13 were independently de novo assembled and insertions and deletions >1.5 kbp called. Raw data was mixed together, assembled and

NGM 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 98% for breakpoint detection in both balanced and unbalanced translocations. Genome mapping can define the true positions of breakpoints within a median distance of 2.9 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 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. NGM 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 5).

Conclusion

Bionano Next-Generation 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

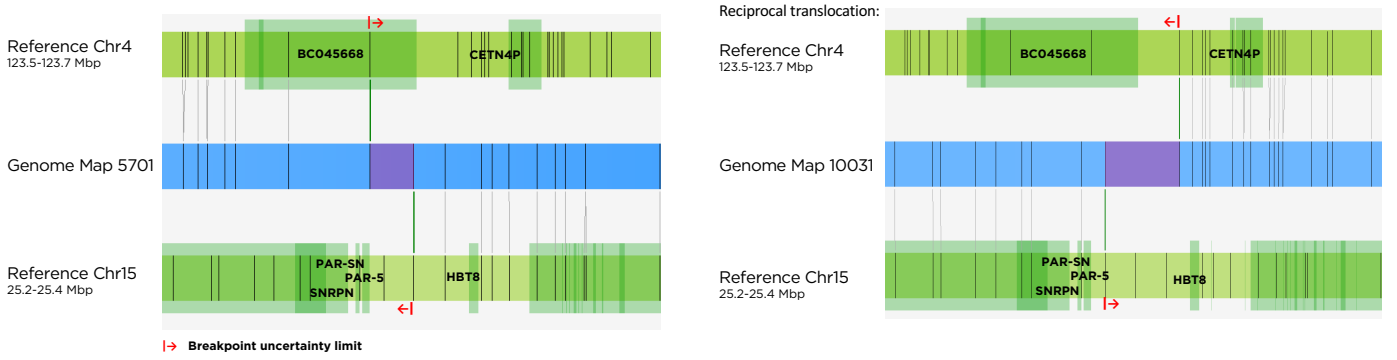

Figure 5:
Example of a translocation detected by NGM, associated with Prader-Willi syndrome. Blue bars are Bionano maps, and vertical lines represent Nt.BspQI label sites. For each of the reciprocal translocation breakpoints, maps are shown with alignments of the maps to chromosome 4 (top) and chromosome 15 (bottom) of the human reference hg19. Breakpoint resolution can be determined by the distance between matched and unmatched labels.

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.Mak AC, L. Y. (2016). Genome- Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays. Genetics, 202 (1), 351-62. 15.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.

For general information about the Saphyr™ and Irys® System, please contact:
info@bionanogenomics.com | 858.888.7600 | bionanogenomics.com

Part #30126 Rev A