Next-Generation Cytogenetics: Bionano Whole Genome Mapping Resolves Structural Variants in Heterogeneous Cancer Samples

Detect all types of structural events at low allele frequency with unrivaled sensitivity and specificity.

Genome instability is a hallmark of all cancer cells and includes somatic mutations ranging from a single base pair to entire chromosomes. A clear identification of such alterations at low allele fraction is the key to molecular diagnostics and provides precious information on tumor origin, patient prognosis, and therapeutic options available.

Tumors can grow seemingly unrestrictedly through a multistep carcinogenesis that alters large numbers of proto-oncogenes and tumor suppressor genes. These alterations include changes in the coding sequence, in DNA methylation patterns, and structural rearrangements of the gene itself or neighboring regulatory sequences. Large structural variants make up the majority of genetic variation in cancer with significant consequences.

An interesting example is the Philadelphia chromosome, observed in more than 90% of chronic myeloid leukemia patients: a large translocation leads to a fusion of the BCR and ABL genes, driving the expression of a fusion protein promoting uncontrolled cellular growth. The identification of such fusion lead to the design of a specific tyrosine kinase inhibitor (Gleevec®) which targets this fusion protein and significantly improves patient outcome. Interestingly, resistance to Gleevec by cancer cells is acquired by a duplication of the BCR-ABL fusion gene which leads to an overexpression of the fusion protein, titrating down the chemotherapy.[1] Such structural variants are frequently encountered in leukemias as well as solid tumors.

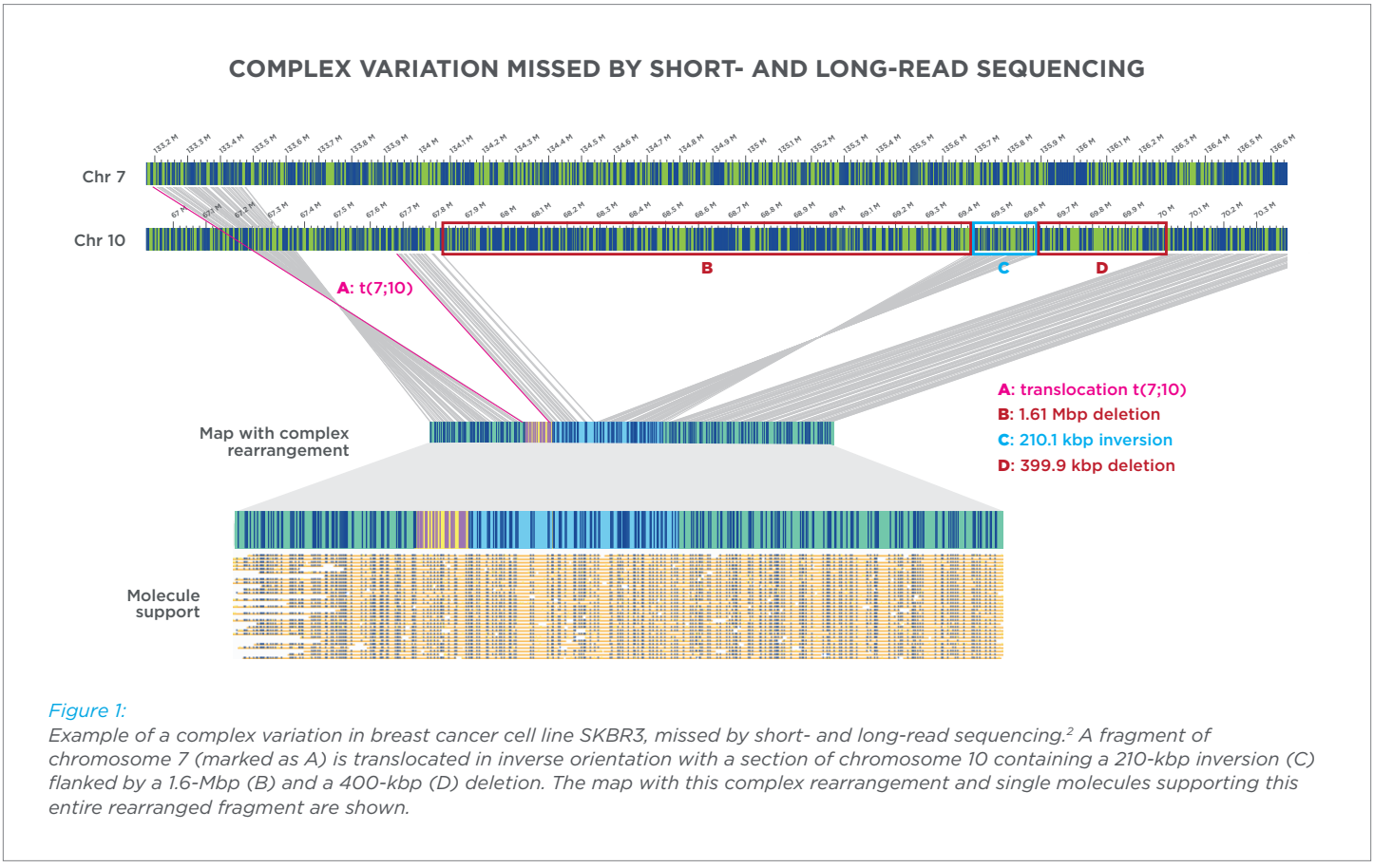

Figure 1:
Example of a complex variation in breast cancer cell line SKBR3, missed by short- and long-read sequencing.[2] A fragment of chromosome 7 (marked as A) is translocated in inverse orientation with a section of chromosome 10 containing a 210-kbp inversion (C) flanked by a 1.6-Mbp (B) and a 400-kbp (D) deletion. The map with this complex rearrangement and single molecules supporting this entire rearranged fragment are shown.

Advances in sequencing technologies have barely changed the way structural variants are detected. Next Generation Sequencing (NGS) reliably identifies 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.

Long read sequencing lacks the throughput, read length and sensitivity to identify complex rearrangements in heterogeneous samples. While long-read sequencing has improved significantly over the years, and occasional reads reach hundreds of kilobasebairs, median read lengths are typically still in the 10 to 30 kbp range, which is not sufficient to span longer repetitive areas of the genome or elucidate large, complex events. Sequencing-based approaches lack the sensitivity to reliably pick up heterozygous structural variants,[3] making the detection of rare variants in cancer subclones impossible. In their current state, long-read sequencing technologies require too large of a sacrifice in cost, time, and coverage depth to be suitable for clinical diagnosis in cancer or genetic disease.

These limitations make direct visualization of the DNA the most reliable approach for the identification of structural variants to date. Therefore, the way structural variants are detected in cancer samples barely evolved over in the past decades and mainly relies on cytogenetic methods such as karyotyping, Fluorescent in situ Hybridization (FISH) and array Comparative Genomic Hybridization (aCGH). Unfortunately, none of these methods alone can address complex cases due to technical limitations (Table 1) and need to be combined to provide a complete therapeutic and prognostic assessment of the tumor.

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.[4]

Bionano whole genome mapping is the only technology that detects all SV types, homozygous and heterozygous, starting at 500 bp, and up to millions of bp. Megabase size molecules of genomic DNA are extracted, labeled, linearized and uniformly stretched in high density Nanochannel arrays, and imaged on the Saphyr® system. The Saphyr instrument generates up to 1300 Gbp, or 400x coverage of molecules larger than 150 kbp, for each of up to three flowcells per 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 16 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.

Bionano optical 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, like Bionano maps, 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, with mixed success. Like karyotyping and FISH, Bionano mapping relies on direct visualization of megabase-size native DNA, 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 simplifies genome structure analysis. With DLS, genome maps typically reach chromosome-arm lengths. 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 (Figure 1).

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 (Figure 2). 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.

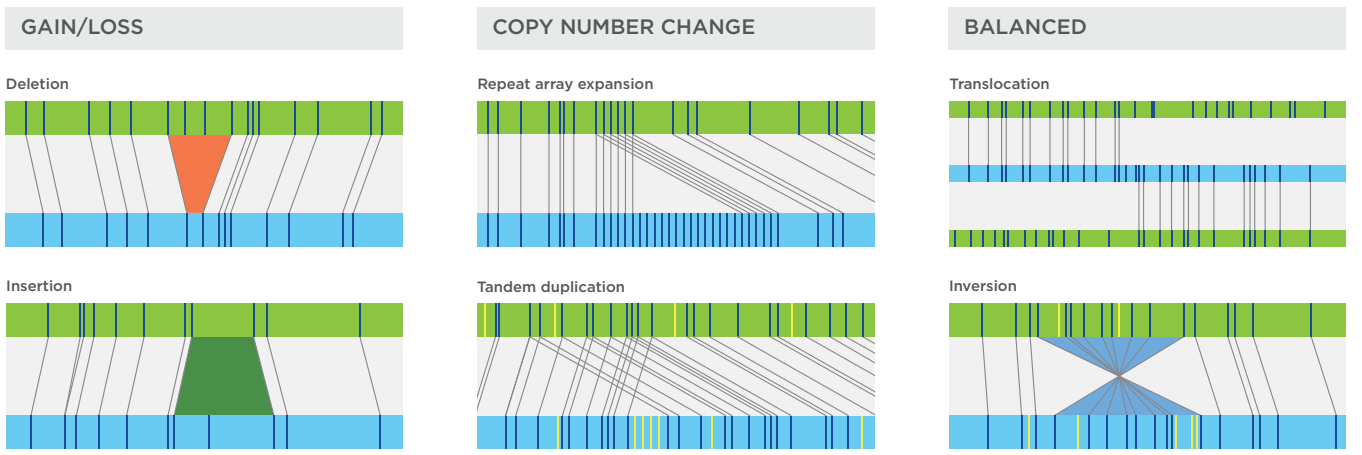

Figure 2:
Structural variant types detected by Bionano mapping. SVs are identified by comparing label patterns in the sample of interest (blue) with those in the reference genome, or in a reference sample (green). Major types detected are:

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, 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.*

Large copy number variants (CNV) are detected at low allele fraction 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 250 kbp.

Bionano Genome Mapping successfully identified large structural variants of clinical significance in solid tumors and leukemias.

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

Bionano genome mapping identifies numerous previously unrecognizable structural variants in leukemia. Professor James Broach, Director of the Penn State Hershey Institute for Personalized Medicine, showed in a prepublication[6] that Bionano mapping detected large numbers of gains, losses, insertions, inversions and translocations that were not previously identified by karyotyping. Fig 3 shows a 3.1-Mbp somatic inversion detected in Non-Hodgkins lymphoma cells that disrupts the PTEN tumor suppressor gene. Several compounds that rescue PTEN mutations are currently in trials. Additionally, Bionano resolved large events that karyotype couldn't identify, such as a large insertion event that Bionano identified as an inverted duplication, and the identity of translocated chromosome fragments. Combined with NGS-based algorithms, Bionano identified structural variants that affected a number of leukemia associated genes as well as cancer driver genes not previously associated with leukemia, and genes not previously associated with cancer. A number of variants only affected intergenic regions, highly enriched for neighboring cancer associated genes. Analysis of TCGA data indicates that the status of several of the recurrently mutated genes identified in this study significantly affect survival of AML patients.

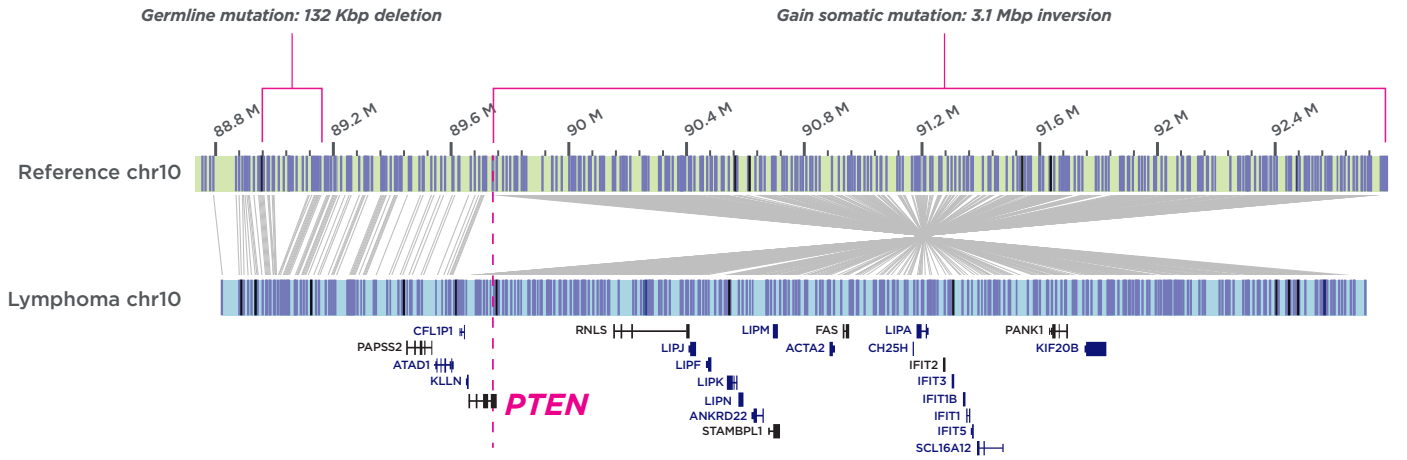

Figure 3:
A 3.1-Mbp somatic inversion in Non-Hodgkins lymphoma cells that disrupts the PTEN tumor suppressor gene, identified only with Bionano and missed by karyotype, NGS.

Bionano optical mapping refines clinically relevant SVs and CNVs in leukemia. Professor Alex Hoischen at Radboud UMC in Nijmegen presented a pilot study on leukemia genomes analyzed with Bionano, as part of a larger clinical validation study of 100 leukemia samples. All clinically relevant SVs and CNVs previously identified with a combination of karyotype, aCGH and FISH were detected by Bionano mapping. Additionally, translocation breakpoints were refined thousandfold to allow for the detection of gene fusions, and the precise overlap with CNVs shown in a single assay. Fig 4 shows a 3-way Philadelphia Chromosome involving chromosomes 9, 14 and 22.

Bionano genome mapping detects structural variants at low allele fraction with unprecedented sensitivity and specificity.

Tumors are often comprised of heterogeneous populations of cells, with certain cancer-driving mutations at low allele fractions in early stages of cancer development. Bionano developed a new single-molecule SV pipeline that is able to detect low-allelic fraction SVs at high sensitivity and PPV, by directly aligning single molecules to a reference genome and calling structural variants when detected in just a few molecules. The pipeline is particularly suited for analysis of cancer samples and samples with germline mosaicism.

Bionano detects extremely rare variants in heterogeneous samples. Since there is no perfectly characterized complex cancer sample that can be considered the ground truth, the Bionano Solve® Rare Variant Pipeline was tested using simulated data, and by creating a dilution series of actual cancer samples, to determine the sensitivity of SV detection at decreasing allele fraction.

Edited haploid genomes with random structural variant events were simulated based an in-silico DLE-1 map of the human reference build hg19. Then, simulated molecules from the edited genome and from hg19 with realistic error characteristics were generated according to empirically derived error and size characteristics such that the simulated molecules would resemble actual molecules collected on a Bionano system.

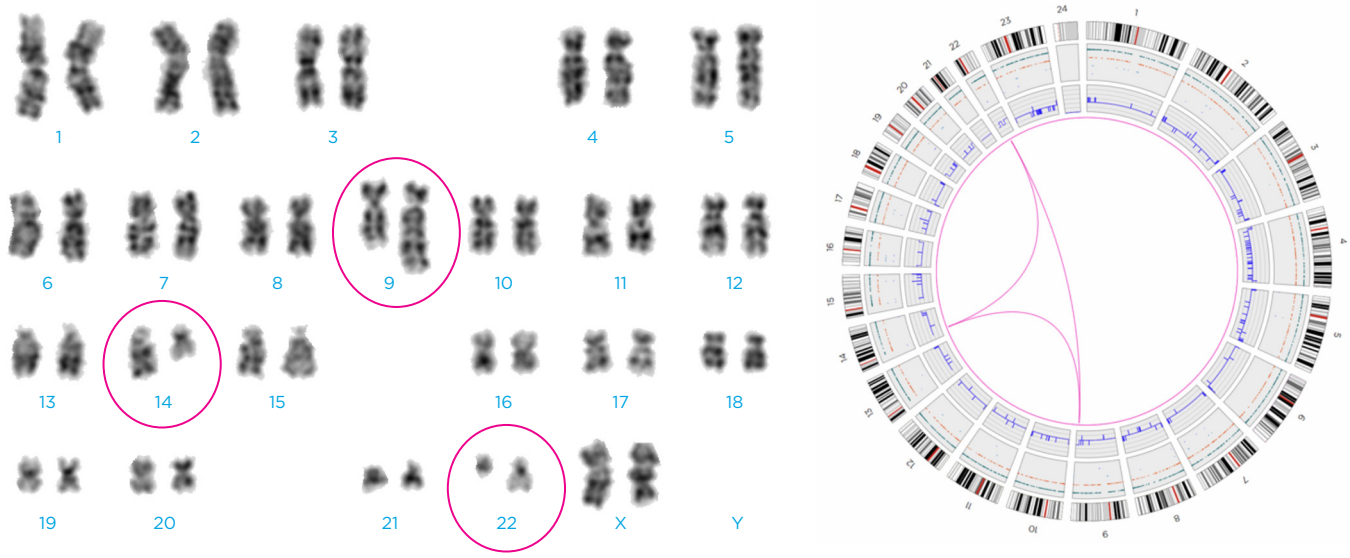

Figure 4:
Bionano mapping identifies a 3-way Philadelphia Chromosome involving chromosomes 9, 14 and 22. The karyotype, and Bionano Access' generated Circos plot are shown side by side. The translocations are represented by the red lines connecting chromosomes 9, 14 and 22.

For insertions and deletions, about 1600 of each type were randomly introduced into an in-silico map of hg19, ranging from 200 bp to 1 Mbp. For translocations, about 900 segments ranging from 50 kbp to 1 Mbp were randomly selected and inserted elsewhere in the genome. For inversions, about 900 segments of 5 kbp to 1 Mbp in size were randomly selected and inverted to simulate inversions. For duplications, about 900 intervals of 5 kbp to 1 Mbp in size were randomly selected. For each, an extra copy of the sequence was inserted in tandem next to the original interval in the same or opposite orientation. All events were at least 500 kbp from each other or N-base gaps.

To simulate a diploid genome, two sets of molecules were mixed in equal proportions. To simulate a genome with a specific average allele fraction, the molecules were mixed at different proportions. The mixture molecule sets were then used as input to the Bionano Solve Rare Variant Pipeline. The resulting SV calls were compared to the ground truth to assess detection sensitivity and positive predictive value (PPV).

At 300x effective coverage, which corresponds to the usable data generated from one flowcell of a Saphyr Chip,® Bionano mapping has:

- **90%** sensitivity to detect **translocations** > 70 kbp at 5% allele fraction
- **90%** sensitivity to detect **inversions** > 70 kbp at 5% allele fraction
- **90%** sensitivity to detect **deletions** > 5 kbp at 5% allele fraction
- **90%** sensitivity to detect **insertions** 5-50 kbp at 5% allele fraction
- **90%** sensitivity to detect **duplications** > 100 kbp at 5% allele fraction

The Rare Variant Pipeline 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, 5% allele fraction corresponds to a heterozygous variant present in 10% of diploid cells.

Additionally, the fractional copy number tool provides 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.

Bionano's Variant Annotation Pipeline reduces thousands of structural variants to those of clinically relevance. The Variant Annotation Pipeline (VAP), part of the Bionano Access® software, streamlines cancer studies. Using VAP, structural variation calls from multiple samples can be analyzed as a group to detect somatic mutations when comparing calls from the tumor with the blood from the same patient.

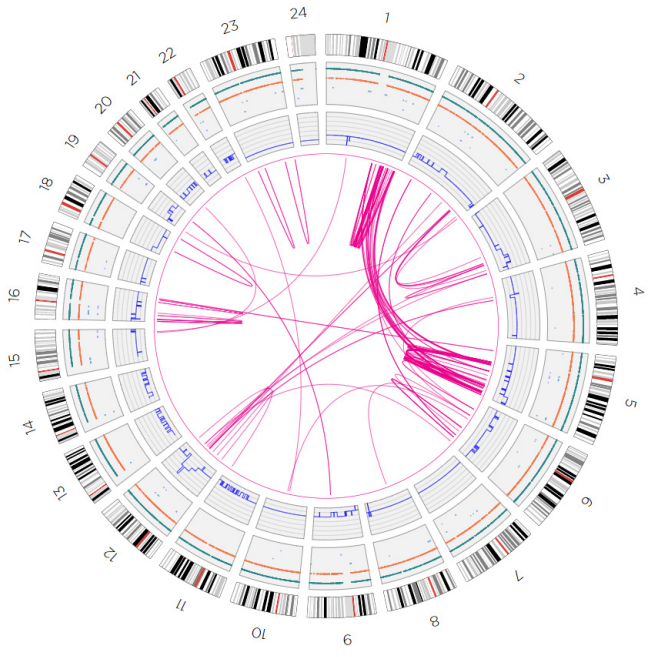

Figure 5:
Circos visualization of SVs in Caki2 cell line
Track 1 (outer most): *cytobands*
Track 2: *insertion/deletion, inversion breakpoint and duplication calls*
Track 3: *copy number profile*
Inner circle: *translocation breakpoint calls*

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 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 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. 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 cancer-associated genes, or with any custom gene list (Figure 5).

The Genoox Integrated SV Platform automatically validates and refines Bionano SV calls using NGS data, and provides clinical annotation for all variants. The genome analysis and annotation company Genoox has developed an integrated pipeline that uses Bionano's ultra-sensitive and specific SV calls to guide alignment of NGS reads. This combined solution validates more than 85% of Bionano's SVs and typically improves the breakpoint precision to less than 20 bp. Their AI-powered classification engine classifies both SVs and SNVs based on clinical evidence, and generates a combined actionable clinical report. This Bionano-NGS integrated solution streamlines the analysis of all variants, affecting single basepairs all the way to entire chromosomes.

Conclusion

Bionano optical genome mapping is the only technology that allows for the highly sensitive detection of all structural variant types present at low allele fraction in heterogenous cancer samples, in an unbiased genome-wide manner. By providing a complete and unambiguous picture of the cancer genome structure, it can identify prognostic markers not currently monitored, and enable a complete characterization of the cancer genome in single test, potentially replacing multiple cytogenetic tests that make up the gold standard. Importantly, Bionano detects SVs affecting genes not previously associated with specific tumor types, enabling the discovery of new biomarkers for patient stratification and targets for the development of innovative new therapies.

Learn More

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

References: 1. Pecorino et al. Molecular biology of cancer mechanisms, 3rd edition. 2. Nattestad et al., Genome Research, 2018 Aug;28(8):1126-1135. 3. Huddleston J, Chaisson MJ, et al. Genome Res. 2016 Nov 28. pii: gr.214007.116. 4. Hastings et al. European guidelines for constitutional cytogenomic analysis, European Journal of Human Genetics, 2019. 5. Jaratlerdsiri W, et al. Oncotarget. 2017; 8:23588-602. 6. Xu J et al. bioRxiv 2019; doi: https://doi.org/10.1101/563270

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