# Comprehensive Detection of Germline and Somatic Structural Mutation in Cancer Genomes by Bionano Genomics Optical Mapping

A.W.C. Pang, J. Lee, T. Anantharaman, E.T. Lam, A. Hastie, M. Borodkin

Bionano Genomics, San Diego, California, United States of America

## Abstract

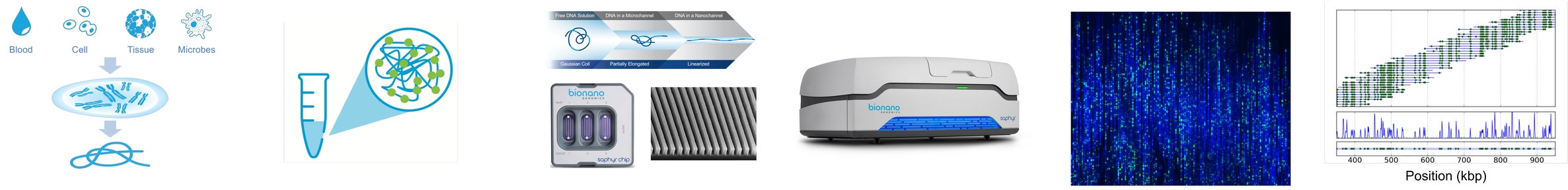
In cancer genetics, the ability to identify constitutive and low-allelic fraction structural variants (SVs) is crucial. Conventional karyotype and cytogenetics approaches are manually intensive. Microarrays and short-read sequencing cannot detect calls in segmental duplications and repeats, often miss balanced variants, and have trouble finding low-frequency mutations. We describe the use of Bionano Genomics's Saphyr platform to comprehensively identify SVs for studying cancer genomes. DNA >100 kbp is extracted, labelled at specific motifs, and linearized through NanoChannel arrays for visualization. Molecule images are digitized and *de novo* assembled, creating chromosomal arm scale genome maps. Somatic mutations can be identified by running the variant annotation pipeline that compares the cancer sample's assembly SVs against >600,000 SVs in Bionano's control sample SV database, and against a matched control sample's SVs, if available. Also, two new Bionano pipelines leverage these long molecules to identify additional somatic SVs: the copy number variation (CNV) and the molecule mapping pipelines. By examining the coverage-depth of molecules alignment to the public reference, the pipeline can identify megabases long CNVs. Similarly, clusters of split-molecule alignments can reliably find translocations and other rearrangements. We applied this suite of discovery tools to identify SVs in a well-studied adenocarcinoma cell line SK-BR-3. We collected data from the tumor, constructed contiguous assemblies (N50 45.9 Mbp), and called >3,600 SVs in the genome. Then, we classified 231 as somatic by comparing against the control sample database. We observed that chromosomes 5, 8 14, 17 and 19 are enriched in somatic rearrangements, harboring multiple balanced, unbalanced translocations and putative gene-fusions. The two new pipelines further increased sensitivity to rearrangements, for example they captured a *MYC* duplication. We apply these thorough approaches to multiple well-studied cancer lines to identify novel SVs missed by previous studies. In conclusion, with one comprehensive platform, Saphyr can discover a broad range of traditionally refractory but relevant SVs, and further improves our understanding of cancer.

## Background

Generating high-quality finished genomes replete with accurate identification of structural variation and high completion (minimal gaps) remains challenging using short read sequencing technologies alone. The Saphyr™ system provides direct visualization of long DNA molecules in their native state, bypassing the statistical inference needed to align paired-end reads with an uncertain insert size distribution. These long labeled molecules are *de novo* assembled into physical maps spanning the entire diploid genome. The resulting provides the ability to correctly position and orient sequence contigs into chromosome-scale scaffolds and detect a large range of homozygous and heterozygous structural variation with very high efficiency.

## Methods

1. Extraction of long DNA molecules
2. Label DNA at specific sequence motifs
3. Saphyr Chip linearizes DNA in NanoChannel arrays
4. Saphyr automates imaging of single molecules in NanoChannel arrays
5. Molecules and labels detected in images
6. Bionano Access software assembles optical maps




**(1) Long molecules of DNA are labeled with Bionano reagents by (2) incorporation of fluorophores at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the Saphyr Chip using NanoChannel arrays (4) Single molecules are imaged by Saphyr and then digitized. (5) Molecules are uniquely identifiable by distinct distribution of sequence motif labels (6) and then assembled by pairwise alignment into *de novo* genome maps.**

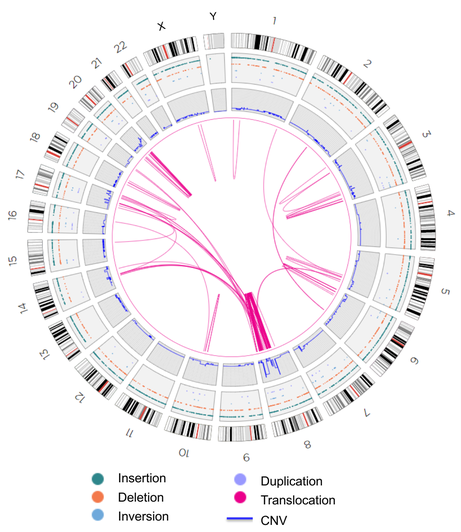
## Case study: Identifying SVs in adenocarcinoma cell line SK-BR-3

### *De novo* assembly and SV detected by comparing the assembly against GRCh38

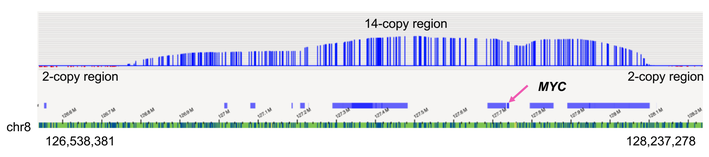
**1**

| Sample | SKBR3 |
|---|---|
| Data collected (molecules >150 kb) | 974 Gbp |
| Assembly size (haplotype-aware) | 5.79 Gbp |
| Genome map N50 | 45.9 Mbp |
| SV called against GRCh38* | |
| Insertions | 2256 |
| Deletions | 1057 |
| Duplications | 95 |
| Inversions | 148 |
| Interchromosomal translocations | 20 |
| Intrachromosomal translocations | 59 |

**2**



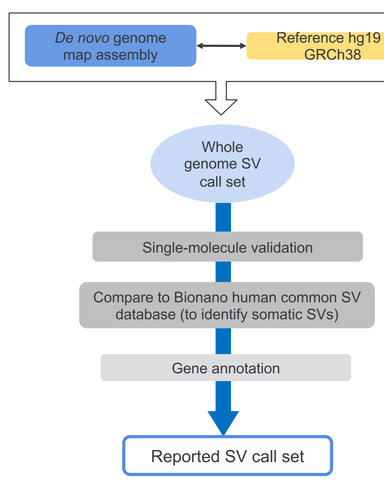

**3**




1. **Results of the *de novo* assembly and SV detected**
2. **Circos plot indicates that chromosomes 5, 8 14, 17 and 19 are enriched in somatic rearrangements**
3. **A 1.33 Mbp copy number amplification at the *MYC* region**

### Variant annotation pipeline identifies somatic mutations

**4**

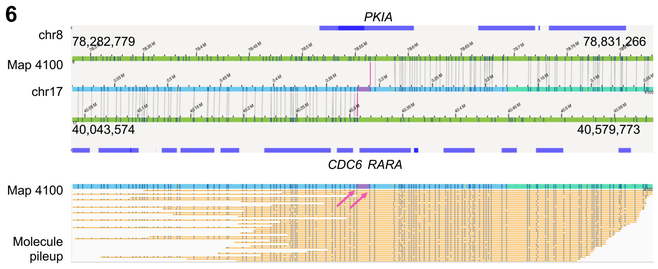



**5**

| Somatic variants identified | SKBR3 |
|---|---|
| Insertions | 8 |
| Deletions | 73 |
| Duplications | 43 |
| Inversions | 28 |
| Interchromosomal translocations | 20 |
| Intrachromosomal translocations | 59 |

4. **Workflow of the annotation pipeline**
5. **The number of candidate somatic mutations**
6. **An optical map captures a t(8;17) translocation**

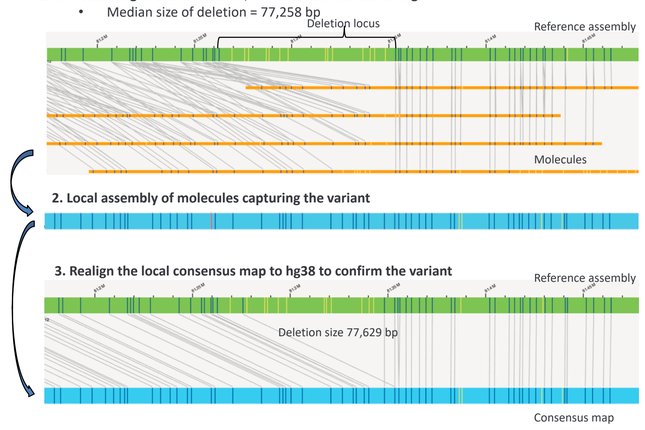
**6**




- Molecule pileup shows strong molecule support across the translocation breakpoints (pink arrows), thus confirming the variant allele
- *PKIA*: protein kinase inhibitor alpha
- *CDC6:* Cell division control protein 6
- *RARA*: retinoic acid receptor alpha; associated with acute promyelocytic leukemia

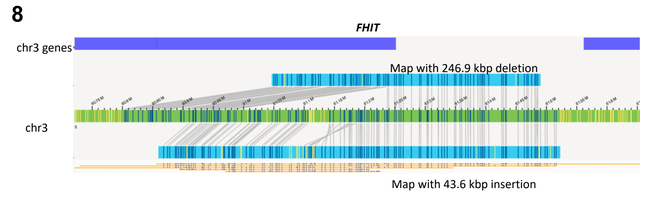
### Rare variant pipeline detects low allelic variants

**7**




7. **Illustration of the rare variant pipeline**
8. **Example a 246.9 kbp deletion and a 43.6 kbp insertion. Both calls are detected uniquely by the rare variant pipeline**

**8**




- *FHIT*: Fragile histidine triad: involved in purine metabolism; found in about 50% of all esophageal, stomach and colon carcinomas

## Conclusions

We demonstrate that the Saphyr system can be used to accurately detect genetic mutation hallmarks in samples with cancer. These includes large rearrangements ranging from translocations, within chromosome fusions, to copy number alterations. Researchers can perform mapping experiments to uncover somatic variants by comparing with our control sample database. Furthermore, our rare variant pipeline enables us to identify lower allelic mutation. Our results indicate that the Saphyr system can capture a broad spectrum of variation with functional importance, and can provide easy solutions for cancer studies.

## Reference

1) Cao, H., et al., Rapid detection of structural variation in a human genome using NanoChannel-based genome mapping technology. Gigascience (2014); 3(1):34
2) Hastie, A.R., et al. Rapid Genome Mapping in NanoChannel Arrays for Highly Complete and Accurate De Novo Sequence Assembly of the Complex Aegilops tauschii Genome. PLoS ONE (2013); 8(2): e55864.
3) Das, S. K., et al. Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. Nucleic Acids Research (2010); 38: 8
4) Xiao, M et. al. Rapid DNA mapping by fluorescent single molecule detection. Nucleic Acids Research (2007); 35:e16.