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

A.W.C. Pang[1], J. Lee[1], K. Hong[1], T. Anantharaman[1], E.T. Lam[1], Y. Delpu[1], **S. Bocklandt[1]**, S. Marin[1], A. Hastie[1], M. Borodkin[1]

[1]Bionano Genomics, San Diego, California, United States of America

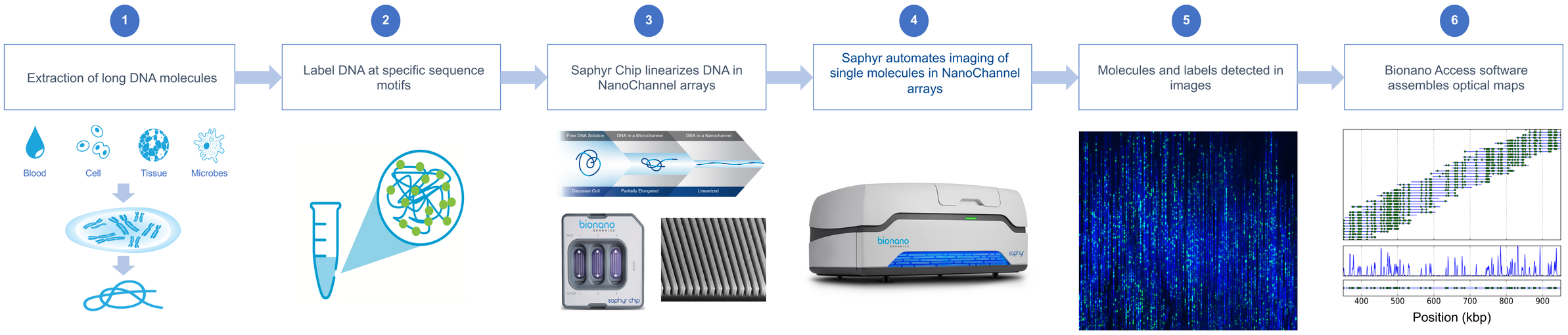
Abstract

Introduction: The ability to identify structural variants (SVs) is crucial in cancer genetics. Karyotype and cytogenetics are manually intensive. Microarrays and sequencing cannot detect calls in segmental duplications and repeats, and miss balanced variants and low-frequency mutations.
Materials and Methods: We describe the Bionano Genomics's Saphyr platform to identify SVs in cancer genomes. DNA >100 kbp is extracted, labelled at specific motifs, and linearized through NanoChannel arrays. Molecule images are digitized and *de novo* assembled, creating chromosomal-arm scale genome maps. Cancer mutations >500 bp are detected by aligning the molecules or the genome maps to the public reference.
Results: Over the past 12 months, the power of Bionano's cancer workflow has been demonstrated on nearly 50 various cancers, including leukemia, breast, ovarian, prostate, pancreatic, among others. While the number of SVs varies among samples, we typically observe >3500 calls per genome. Among leukemia samples, we captured the *BCR-ABL1* translocation as well as deletions impacting tumor suppressor genes such as *PTPN14* and *ESRRG*. We resolved the structure of large duplications (790 kbp) disrupting *BRCA1* in early-onset breast cancers, found the amplification of *MYC* in lung cancers.
Conclusions: In conclusion, with one platform, Saphyr can discover a broad range of traditionally refractory but relevant SVs, and improves our understanding of cancer.

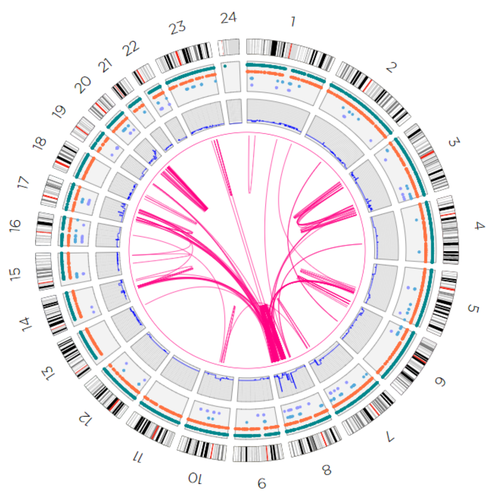
Background

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



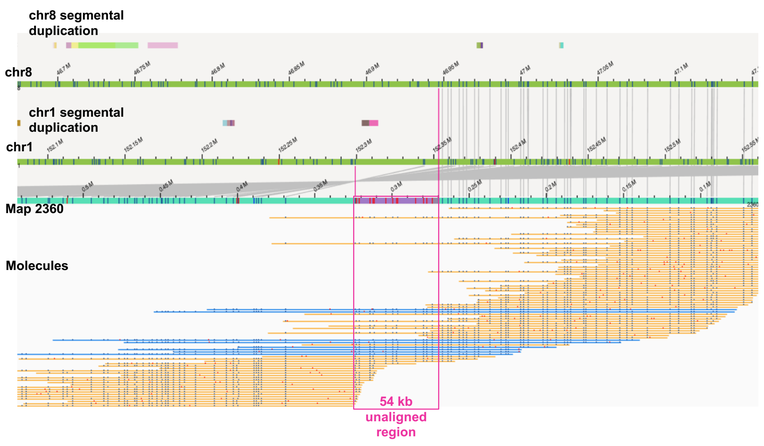
(1) Long molecules of DNA are labeled with Bionano reagents by (2) incorporation of fluorophores at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the Saphyr Chip using NanoChannel arrays (4) Single molecules are imaged by Saphyr and then digitized.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



Sample	SK-BR-3
Data collected (molecules >150 kb)	974 Gbp
Assembly size (haplotype-aware)	5.79 Gbp
Genome map N50	45.9 Mbp
SV called against hg38*	
Insertions	2256
Deletions	1057
Duplications	95
Inversions	148
Interchromosomal translocations	20
Intrachromosomal translocations	59

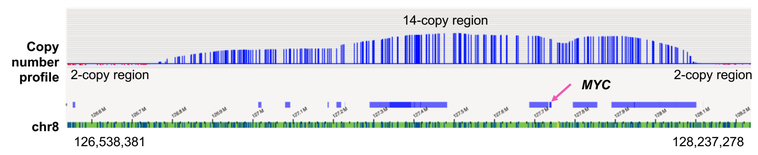
Insertion
Deletion
Inversion
Duplication
Translocation
CNV

Overview of SVs detected in a breast cancer sample

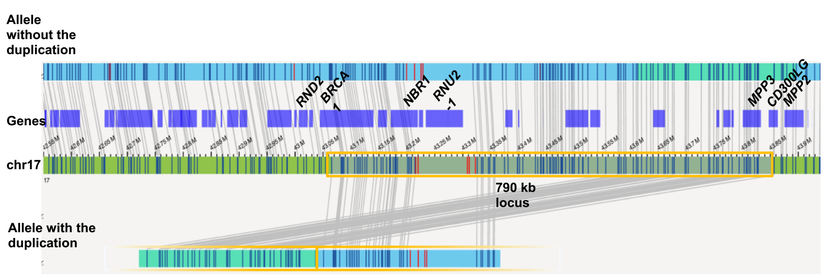


Example of a translocation in SK-BR-3 missed by sequencing in the study Nattestad et al., 2018. With Bionano's long molecules, we were able to construct the translocation allele. We observe a 54 kbp novel sequences between the translocation junctions, and the one of the junctions coincides with segmental duplications.

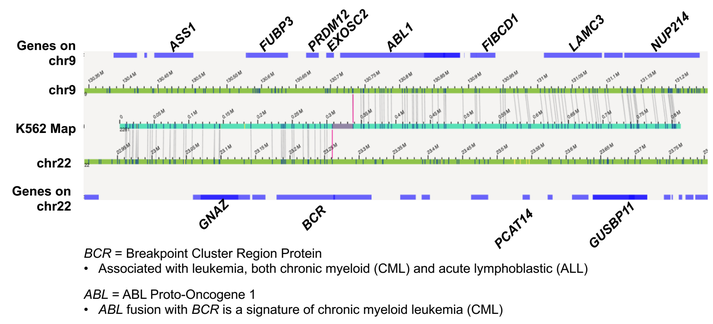
Change in molecule depth of coverage can identify amplifications and deletions



A 1.33 Mbp copy number amplification at the *MYC* region in SK-BR-3



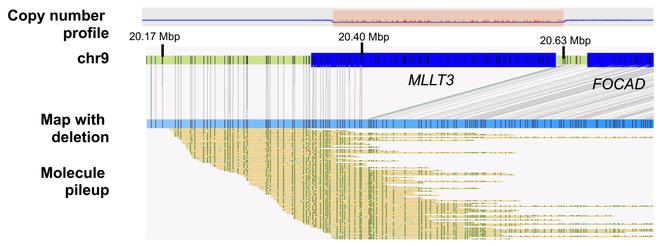
A 790 kbp tandem duplication found in a breast cancer sample. The de novo assembly captured the duplication breakpoint, which indicates that *BRCA1* fused to *CD300LG*.



BCR = Breakpoint Cluster Region Protein
- Associated with leukemia, both chronic myeloid (CML) and acute lymphoblastic (ALL)

ABL = ABL Proto-Oncogene 1
- *ABL* fusion with *BCR* is a signature of chronic myeloid leukemia (CML)

The t(9;22) translocation detected in the chronic myeologenous leukemia (CML) sample K-562. Note that the breakpoints are at *BCR* on chr22 and at *ABL1* on chr9.



MLLT3 (MLLT3 Super Elongation Complex Subunit)
- Involves in transcriptional misregulation in cancer.

A 222 kbp deletion identified in an acute lymphoblastic leukemia (ALL) sample. The deletion is called independently by the copy number profile and by genome mapping

Conclusions

We demonstrate that the Saphyr system can be used to accurately detect genetic mutation hallmarks in samples with cancer. These include large rearrangements ranging from translocations, within chromosome fusions, to copy number alterations. Researchers can perform experiments to uncover somatic variation by comparing with Bionano control sample database, or against a matched pair sample. Furthermore, Bionano SV pipelines can detect SVs with complex breakpoint structures that are recalcitrant to detection by other technologies. 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.
5) Nattestad, M. et al. Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. Genome Research (2018); 28:1126.