

# Comprehensive Structural Analysis of Cancer Genomes by Genome Mapping

Ernest T Lam, Andy WC Pang, Thomas Anantharaman, Jian Wang, Xinyue Zhang, Tom Wang, Alex R Hastie, Scott Way, Henry B Sadowski, Mark Borodkin  
Bionano Genomics, San Diego, California, United States of America

## Abstract

Tumors are often comprised of heterogeneous populations of cells, with certain cancer-driving mutations at low allele fractions in early stages of cancer development. Effective detection of such variants is critical for diagnosis and targeted treatment. However, typical short sequence reads are limited in their ability to span across repetitive regions of the genome and to facilitate structural variant (SV) analysis. Based on specific labeling and mapping of ultra-high molecular weight (UHMW) DNA, we developed a single-molecule platform that has the potential to detect disease-relevant SVs and give a high-resolution view of tumor heterogeneity.

We developed a simplified DNA isolation and sample preparation workflow that preserves the DNA integrity and maximizes the labeling efficiency. The DNA labeling workflow takes approximately 5 hours, with little hands-on time. Single molecules are labeled at specific motifs and analyzed in massively parallel nanochannels. The single-molecule maps are used in a bioinformatics pipeline that effectively detects structural variants at low

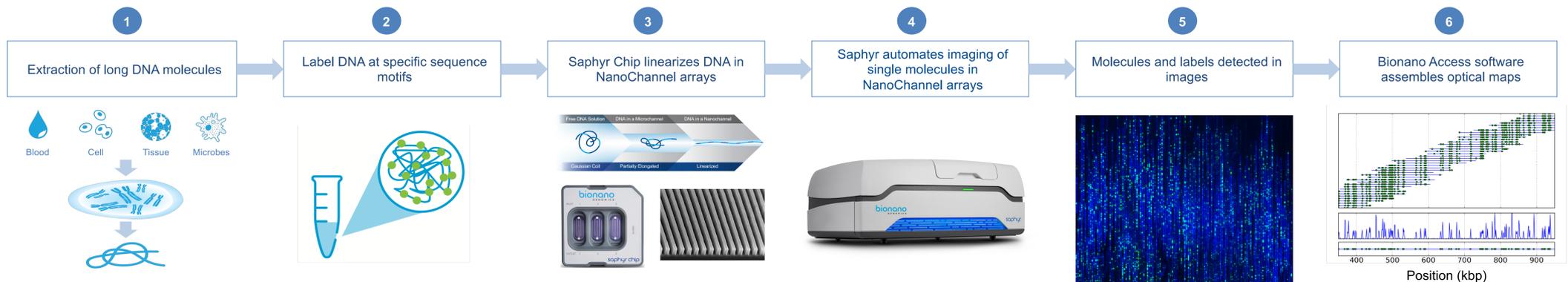
allele fractions. It includes single-molecule based SV calling and fractional copy number analysis. Preliminary analyses using simulated and well-characterized cancer samples showed high sensitivity for variants of different types at as low as 5% allele fractions. The candidate variants are then annotated and further prioritized based on control data and publically available annotations. The data are imported into a graphical user interface that includes new visualization tools (such as Circos diagrams) for real-time interactive visualization and curation.

Bionano offers sample preparation, DNA imaging and genomic data analysis technologies combined into one streamlined workflow that enables high-throughput genome mapping on the Bionano Saphyr system. Together, these components allows for efficient analysis of any genome of interest.

## Background

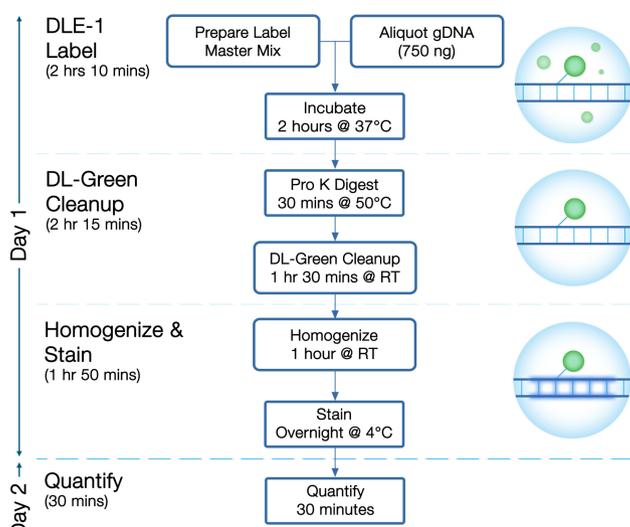
Generating high-quality finished genomes replete with accurate identification of structural variation and high completion (minimal gaps) remains challenging using short read sequencing technologies alone. The Saphyr™ system provides direct visualization of long DNA molecules in their native state, bypassing 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.

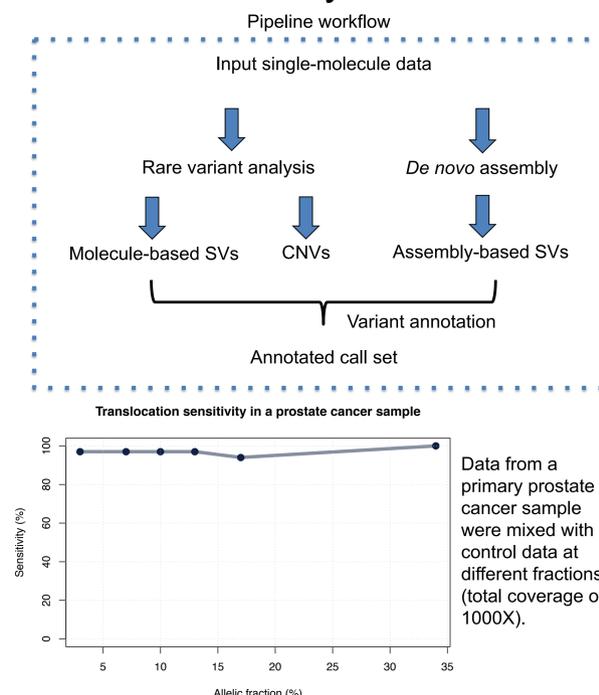
## Sample preparation



Sequence-specific, direct labeling of megabase-size genome DNA for Bionano mapping.

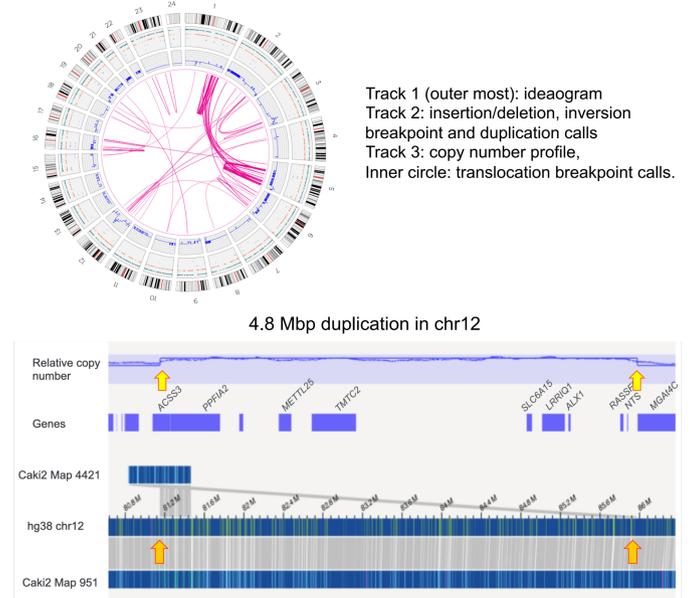
The Direct Label Enzyme (DLE-1) is the first of a new class of DLS enzymes. It shows high sequence specificity and labeling efficiency, and its labeling density is suitable for mapping a wide range of organisms. Molecules are fluorescently labeled with DL-Green, and their length N50s are routinely larger than 300 kbp. The entire labeling workflow requires roughly 5 hours, with little hands-on time.

## Analysis



The integrated pipeline contains a new SV module and a fractional copy number analysis module designed for detection of low allelic fraction SVs. Analyses based on simulated and real data showed high sensitivity and PPV at different allelic fractions.

## Visualization



Circos visualization (top) and a duplication detected in Caki2 (bottom).

Advanced visualization options (such as Circos plots) facilitate exploratory and interactive data analysis. In this sample, we detected a large duplication (elevation in the copy number profile and inference based on map alignment pattern) that overlapped multiple genes and was not found in the Bionano control database.

## Conclusions

Understanding of the genome structure is important for disease studies. It is particularly relevant for analysis of cancer genomes, which often contain complex rearrangement events. Traditional approaches like FISH and karyotyping have limited throughput and resolution. Bionano provides a streamlined workflow for efficient and comprehensive analysis of the genome structure. We developed a sample preparation and labeling protocol that maximizes labeling efficiency and preserves the DNA integrity. We also developed bioinformatics tools that enable detection of low allelic fraction structural variants and fractional copy number variants. The variants are then annotated against known genes and control data. These tools are integrated with Bionano Access, which serves as the graphical user interface for running various downstream analyses. Access's data visualization tools are interactive, and they allow for curation of SVs of interest. Using these tools, researchers could map a cancer genome in a single assay for under \$500 per sample.

## Reference

Mak AC et al. Genome-wide structural variation detection by genome mapping on nanochannel arrays. *Genetics* (2016); 202: 351-62.  
Cao H et al. Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. *GigaScience* (2014); 3(1):34