Sensitive detection of low-allele fraction structural variants in clinical cancer samples

ET Lam, AWC Pang, T Anantharaman, J Wang, T Wang, X Zhang, J Lee, H Sadowski, A Hastie, M 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. Short reads from next-generation sequencing platform provide fragmented information and limited power to detect complex structural variants (SVs). Based on specific labeling and mapping of ultra-high molecular weight (UHMW) DNA, Bionano's single-molecule platform has the potential to detect disease-relevant SVs and give a high-resolution view of tumor heterogeneity.

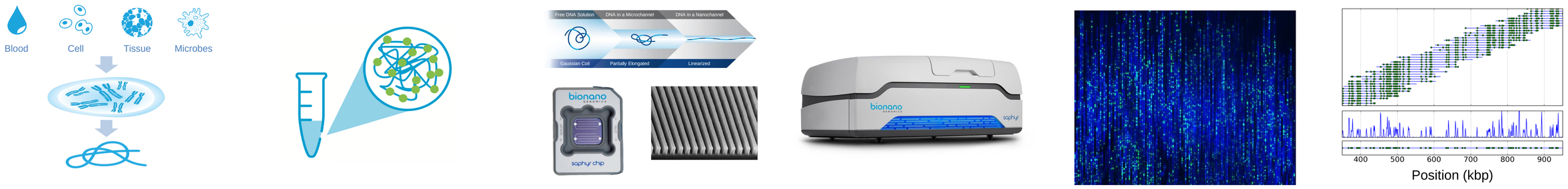
We have developed a pipeline that effectively detects structural variants at low allele fractions. The pipeline uses and merges information from a single-molecule analysis and a *de novo* assembly to detect candidate variants. The single-molecule analysis can be done in mere hours; it includes both single-molecule based SV calling and fractional copy number analysis. The *de novo* assembly, while more computationally intensive, allows for more complete characterization of complex rearrangements. The candidate variants are then annotated and further prioritized based on control data and publically available annotations. Preliminary analyses using simulated and well-characterized cancer samples showed high sensitivity for variants of different types.

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.

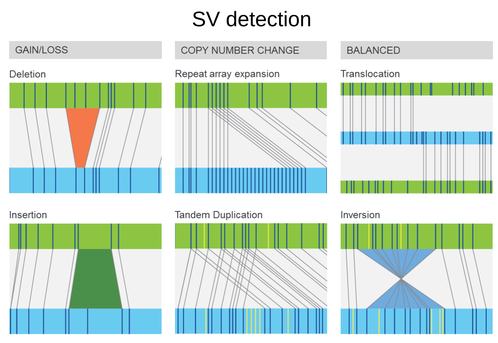
Methods

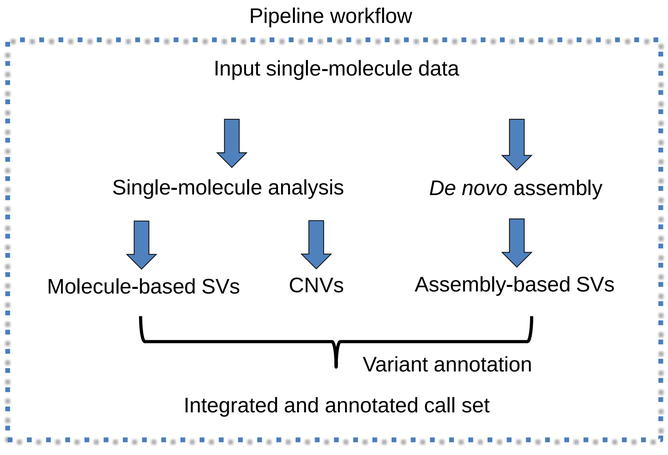
1. Extraction of long DNA molecules
2. Label DNA at specific sequence motifs
3. Saphyr Chip linearizes DNA in NanoChannel arrays
4. Saphyr automates imaging of single molecules in NanoChannel arrays
5. Molecules and labels detected in images
6. Bionano Access software assembles optical maps



(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.

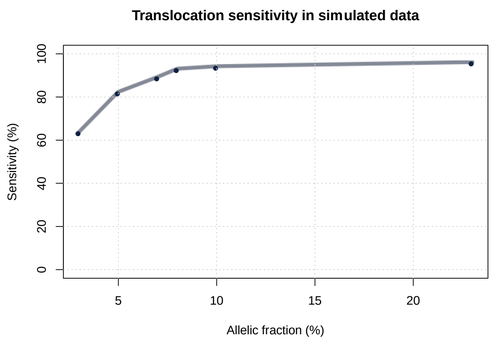
Structural variation analysis



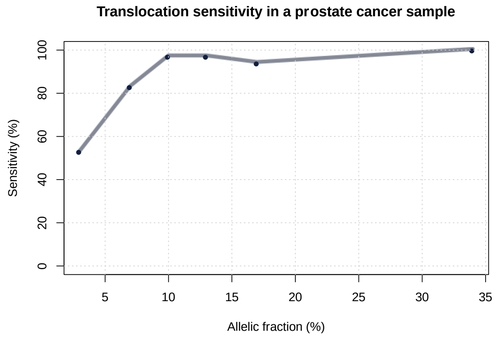


Bionano map data provide critical long-range structural information and are useful for detection of different SV types. Molecule-based and assembly-based calls are integrated and annotated against public databases and control data to facilitate downstream analyses.

SV performance



Molecules were simulated from a modified genome with translocations and mixed with molecules simulated from hg19 at different fractions.

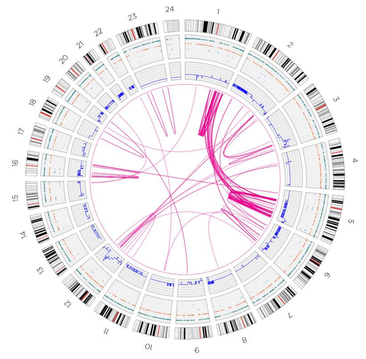


Molecules from a primary prostate cancer sample were mixed with molecules from controls at different fractions.

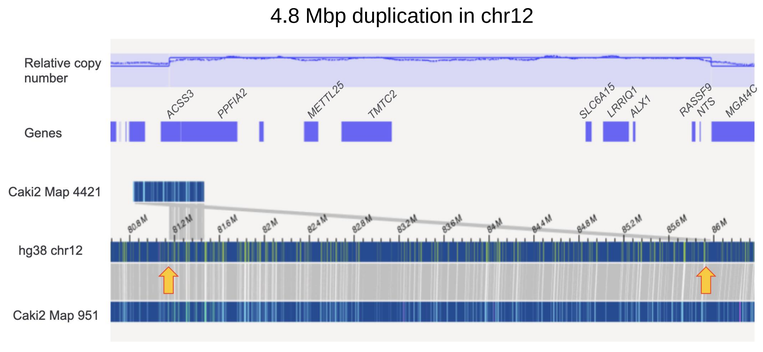
Translocation breakpoint sensitivity based on ***simulated*** data (*top*) and ***real*** data (*bottom*).

The integrated pipeline contains a new SV module and a fraction 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



Track 1 (outer most): ideogram
Track 2: insertion/deletion, inversion breakpoint and duplication calls
Track 3: copy number profile,
Inner circle: translocation breakpoint calls.



Circos visualization of SVs in Caki2 cell line *(top)* and a duplication detected in the sample *(bottom)*.

Advanced visualization options (such as Circos plots) are developed to facilitate exploratory and interactive data analysis. In this sample, we detected a large duplication that overlapped multiple genes and was not found in the Bionano control database.

Conclusions

Understanding of the genome structure is important for disease studies; however, typical short sequence reads are limited in their ability to span across repetitive regions of the genome and to facilitate SV analysis. Bionano provides a streamlined workflow for efficient and comprehensive analysis of the genome structure. Existing bioinformatics tools have been upgraded and new components added to support analysis of heterogeneous samples and detection of low allelic fraction events. Bionano Solve is integrated with Bionano Access, which serves as the graphical user interface for running various downstream analyses. Access' data visualization tools are interactive, and they allow for curation of SVs of interest. Using these tools, researchers can map their genomes of interest 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