Bionano Genomics “Sample to Answer” workflow for single molecule analysis of variation in genome structure.

S. Bocklandt, H. Sadowski, E. Lam, A. Pang, T. Anantharaman, A. Hastie, G. Pljevaljcic, S. Way, P. Lynch, W. Willemse and M. Borodkin
Bionano Genomics, San Diego, California, United States of America

Abstract

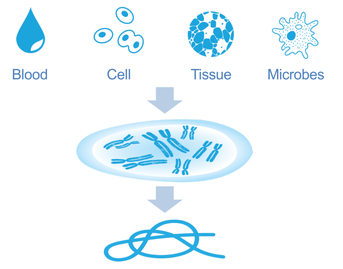
Complex structural changes are often observed in cancer and other diseases. Bionano optical mapping of very long single DNA molecules provides critical structural information for comprehensive genome analysis. The extracted DNA molecules are labeled at specific motifs and analyzed in massively parallel nanochannels. The single-molecule maps can be assembled into chromosome arm-length maps. The maps can be used for structural variation (SV) detection and elucidation of complex rearrangements. Bionano presents an optimized workflow for cancer and constitutional disorders that includes DNA isolation and labeling, DNA imaging, and genomic data analysis. The workflow starts with the isolation of ultra-high molecular weight (UHMW) genomic DNA using Bionano Prep SP. This couples solution-based lysis with a purification step that leverages a novel process to bind, wash and elute UHMW genomic DNA. This entire protocol can be conducted with 3 hours of hands-on time on a batch of 6 samples, allowing 12 samples to be processed in one day. The eluted material is ready to label by day 2 and contains high quality UHMW DNA, that includes Mbp-sized molecules. After DNA isolation, an enzymatic labeling approach, Bionano Prep DLS (Direct Label and Stain), preserves the integrity of the DNA while labeling sequence motifs across the whole genome. The labeled DNA molecules flow through the Saphyr Chip and are imaged by the Saphyr system. Up to 3.9 Tbp of coverage can be collected per Saphyr Chip, allowing for the processing of multiple samples and/or very high depth for cancer analysis. The data analysis tools provided with Bionano Access take advantage of this data for genome assembly and SV analysis applications. At standard coverage, we provide unprecedented sensitivity to heterozygous SVs. At high coverage, the single molecule pipeline also uncovers SVs that occur at a 5% allelic frequency. These analyses can be performed on local resources or on Bionano Compute On Demand, an economical hosted offering. Combining all these elements, allows cancer researchers, for example, to get whole genome low allelic SVs from multiple samples, in less than 5 days, sample-to-SVs.

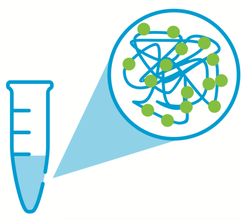
Background

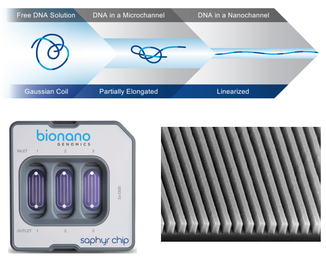
Generating high-quality finished genomes replete with accurate identification of structural variation and high completion (minimal gaps) remains challenging using short read sequencing technologies alone. The Saphyr™ system provides direct visualization of long DNA molecules in their native state, bypassing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

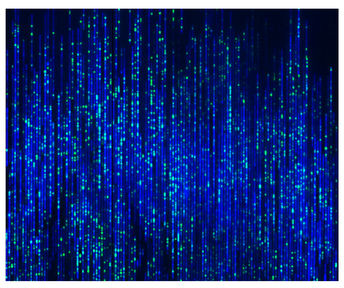
Methods

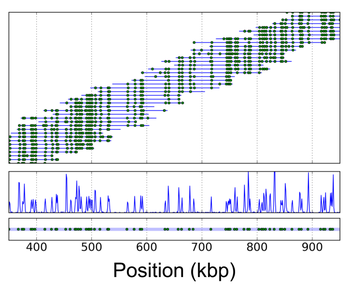
1. Extraction of long DNA molecules
2. Label DNA at specific sequence motifs
3. Saphyr Chip linearizes DNA in NanoChannel arrays
4. Saphyr automates imaging of single molecules in NanoChannel arrays
5. Molecules and labels detected in images
6. Bionano Access software assembles optical maps





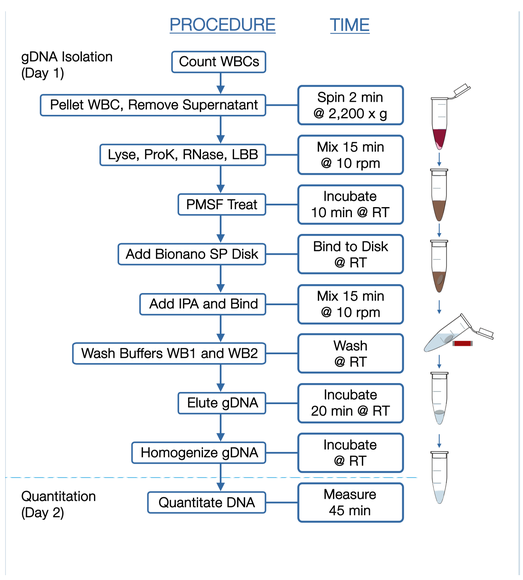


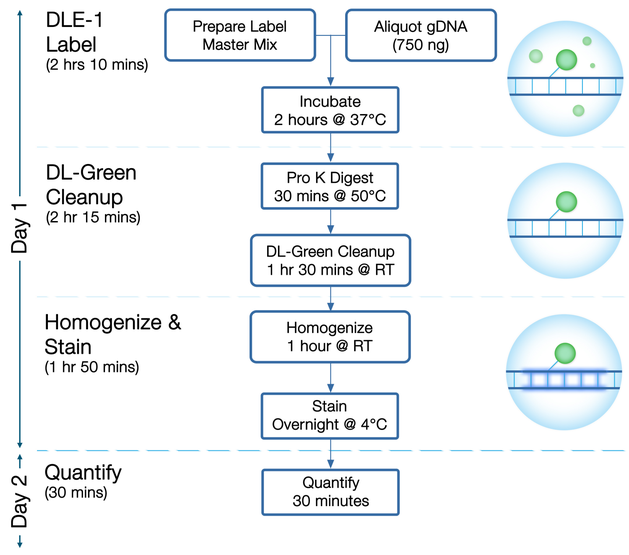




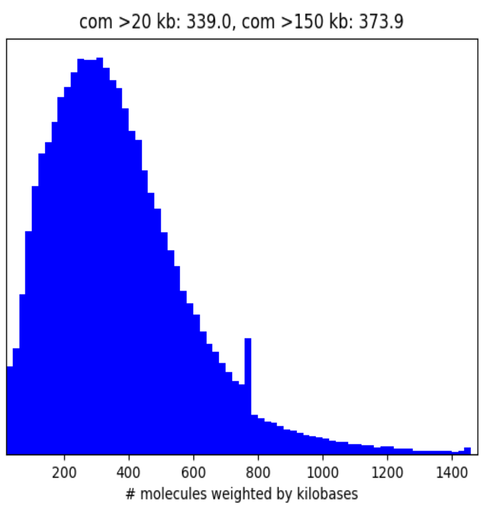
(1) Long molecules of DNA are labeled with Bionano reagents by (2) incorporation of fluorophores at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the Saphyr Chip using NanoChannel arrays (4) Single molecules are imaged by Saphyr and then digitized.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

Bionano SP DNA Isolation of Various Sample Types, DLS Labeling, Saphry Single Molecule Metrics & Size Distribution of Labeled DNA from GM12878 Cells





Sample Type	gDNA [ ] (ng/uL)	Gbp / Scan	N50		Label Density	Map Rate
			> 20kbp	> 150kbp		
Fresh Blood (4 users, 8 bloods, 24 samples)						
Mean	59.0	36.1	233.4	300.5	15.4	89.6%
SD	12.1	9.3	58.4	42.0	0.4	3.6%
Frozen Blood (3 users, 17 bloods, 34 samples)						
Mean	57.4	27.8	203.0	264.6	15.2	90.2%
SD	12.0	8.5	28.5	22.1	0.3	2.5%
Fresh Cells (4 users, 2 cultures, 10 samples)						
Mean	113.4	31.8	283.2	317.7	14.9	93.1%
SD	16.2	14.0	39.0	37.4	0.3	0.8%
Frozen Cells (4 users, 2 cultures, 12 samples)						
Mean	106.9	21.4	264.4	319.4	15.2	89.6%
SD	29.9	11.7	40.2	32.1	0.4	5.2%
Frozen Heparin BMAs (2 users, 8 BMAs, 28 samples)						
Mean	60	30.8	253.5	306.5	15.6	90.1%
SD	13.0	15.7	53.9	50.1	0.6	2.6%



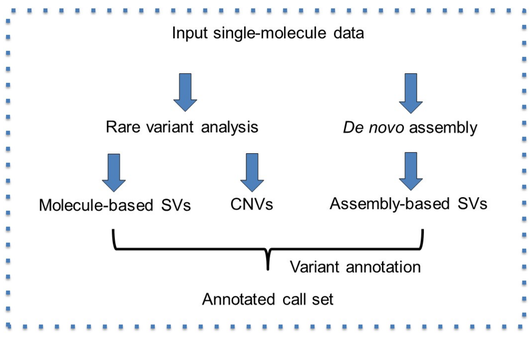
Data Analysis on Bionano Compute (Server or On Demand) and Data Visualization with Bionano Access

Bionano Compute Options

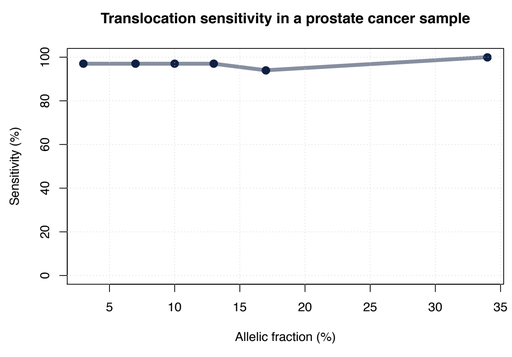
COMPUTE SERVER
- Expect to run servers for >25% of the time
- Have consistent loads
- Internet access not permitted

COMPUTE ON-DEMAND
- Have variable workloads
- Capital budget concerns
- Receive data from service providers
- Work on large genomes

Pipeline workflow

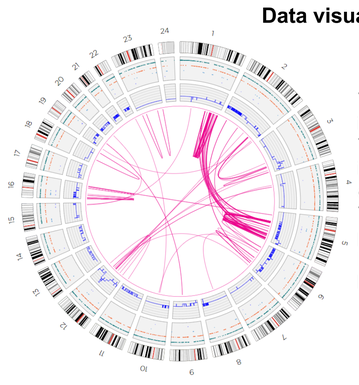


Rare Variant Pipeline



Data from a primary prostate cancer sample were mixed with control data at different fractions (1000X effective coverage).

Data visualization



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

Conclusions

Understanding 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; sample to answer in under 5 days.

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