Broad Range Chromosomal Abnormality Detection through Bionano Genome Mapping

Alex Hastie, Andy Wing Chun Pang, Joyce Lee, Ernest T Lam, Xinyue Zhang, Thomas Anantharaman, Warren Andrews, Henry Sadowski, Mark Borodkin, Han Cao

Bionano Genomics, San Diego, California, USA

Abstract

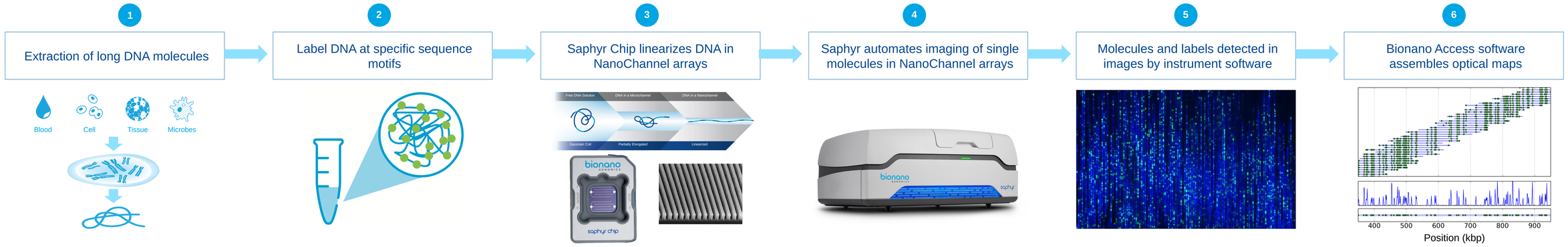
Current methods for detection of balanced structural variation can be broken down into two categories: traditional cytogenetics and molecular methods. Cytogenetics may include chromosomal karyotyping, fluorescence in situ hybridization (FISH), chromosomal microarray and adaption of them. Molecular methods primarily include NGS sequencing based methods. Bionano genome mapping, an optical mapping approach in a NanoChannel array system, is a method that combines the advantages of different categories while solving many of the limitations. Compared to cytogenetics, Bionano mapping is high throughput and removes manual interpretation, it also has much higher resolution, detecting balanced events as small as about 30 kbp compared to multi-megabases needed for cytogenetic approaches, and unbalanced events starting at 500 bp. NGS based methods often are limited by read lengths that cannot provide unambiguous information across repeat elements longer than individual reads. This limitation results in significantly reduced sensitivity for balanced variation and abnormalities as well as for insertions and even some categories of larger deletions. This is particularly true in highly medically relevant locations where segmental duplications mediate chromosomal abnormalities.

Genome mapping using Bionano Genomics' Saphyr System offers high-throughput, genome-wide visualization of extremely long DNA molecules in their native form. It allows researchers to interrogate genomic structural variations (SVs) in the range of 500 base pairs and above. We present several in silico and biological validation experiments that demonstrate the sensitivity and specificity of Bionano mapping for the detection of insertions, deletions, inversions and translocations compared to benchmark studies using short read and long read sequencing. We also show a new direct labeling method that dramatically improves assembly contiguity such that chromosome arms can be assembled into single maps. Bionano mapping is a fast and cost effective method for the detection of a broad range of traditionally refractory SVs across the human genome.

Background

Generating high-quality finished genomes 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 platform provides the ability to correctly position and orient sequence contigs into chromosome-scale scaffolds and detect a large range of structural variation with very high efficiency.

Methods



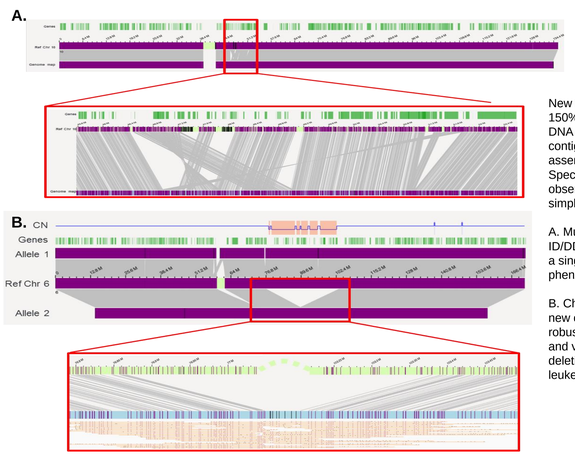
(1) Long molecules of DNA are labeled with Bionano reagents by (2) incorporation of fluorophores at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the Saphyr Chip using NanoChannel arrays (4) Single molecules are imaged by Saphyr and then digitized.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

New Direct Labeling System

Sample	Molecule N50 > 150 kbp (Kbp)	Bionano Map N50 (Mbp)
NA12878	293	55.9
Human Blood	307	56.9
Maize B73	260	100.0
Durum Wheat	364	13.0
Farro	300	32.7
Strawberry	241	13.3
Kakapo	247	69.3
Hummingbird	310	38.7
Blackbird	243	21.6
Fish	245	22.3
Ferret	262	66.1
Pig	335	65.2
Soybean	246	23.0
Brassica	270	12.4
Mouse	280	101.0

Organisms de novo assembled using Bionano direct labeling chemistry (DLS). De novo assemblies often cover whole chromosome arms, only broken at centromeres and other low complexity regions which are longer than molecule length.

DLS performs with equivalent sensitivity and specificity for structural variation when compared to previous NLR genome mapping at 50% the cost and time needed.



New Bionano direct labeling chemistry has 150% higher information density and no DNA damage allows unprecedented contiguity of up to 100+ Mbp NG50, often assembling whole chromosome arms. Spectacular contiguity from directly observed ultra long molecules gives a simple and robust picture of the genome.

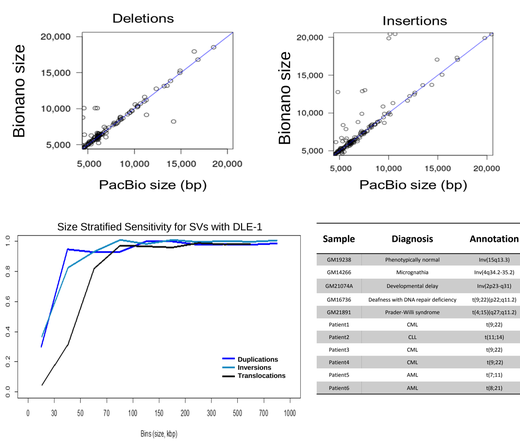
A. Multiple large SVs are seen in the ID/DD region of 10q11.21 assembled onto a single map (here in an individual with no phenotype).

B. Chromosomal arm length maps and new copy number profiling provide a robust tool for large scale SV detection and visualization, here showing a 27 Mbp deletion found in the genome of a leukemia patient.

Validation of Bionano Structural Variation Detection

	PACBIO				BIONANO			
	CHM1 and CHM13 assemblies	Mixture assembly	Sensitivity	PPV	CHM1 and CHM13 assemblies	Mixture assembly	Sensitivity	PPV
Homozygous Insertions	467	353	75.6%	96.1%	707	700	99.0%	97.9%
Heterozygous Insertions	586	252	43.0%		663	554	83.6%	
Homozygous Deletions	221	183	82.8%	94.9%	269	268	99.6%	97.1%
Heterozygous Deletions	501	337	67.3%		517	477	92.3%	

Two Homozygous cell lines, CHM1 and CHM13 were independently *de novo* assembled and SVs called. The molecules were mixed together, assembled and SVs called (mixture column). In the simulated diploid assembly, CHM1 only and CHM13 only SVs are heterozygous and those detected in both are homozygous SVs. Results are compared to a similar experiment using PacBio data (filtered to the same size, >1.5 kbp) for SV calling (Huddleston et al., 2016)



SV size comparison between overlapping Bionano and PacBio SV calls shows good size concordance, verifying that Bionano SVs are accurate.

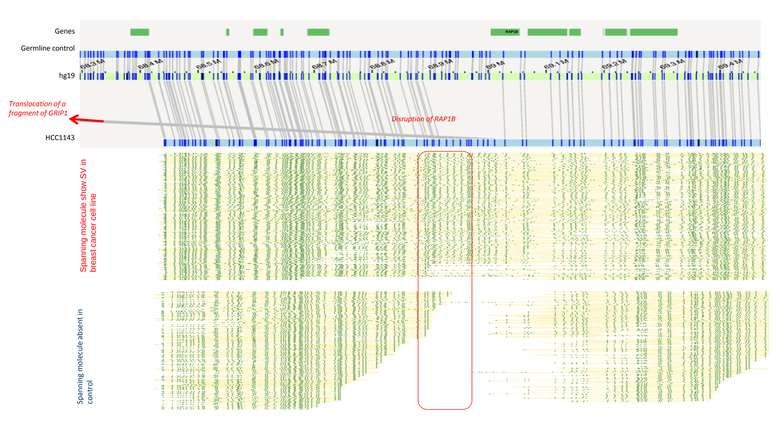
Sample	Diagnosis	Annotation	Found by Bionano?
GM19238	Phenotypically normal	Inv(15q13.3)	Yes
GM14266	Micrognathia	Inv(4q34.2-35.2)	Yes
GM21074A	Developmental delay	Inv(2p23-q31)	Yes
GM16736	Deafness with DNA repair deficiency	t(9;22)(q22;q11.2)	Yes
GM21891	Prader-Willi syndrome	t(4;15)(q27;q11.2)	Yes
Patient1	CML	t(9;22)	Yes
Patient2	CLL	t(11;14)	Yes
Patient3	CML	t(9;22)	Yes
Patient4	CML	t(9;22)	Yes
Patient5	AML	t(7;11)	Yes
Patient6	AML	t(8;21)	Yes

Inversion, duplication and translocations were simulated from hg19 by selecting fragments of the genome with characteristics of biological data. Simulated molecules were de novo assembled and aligned to hg19 to find structural variation. Sensitivity as a function of size for inversions, duplications and translocations is plotted. To verify performance, we tested 11 samples with known translocations and inversions and were able to find all expected events.

Automated Somatic Structural Variation Discovery

Somatic SVs	HCC1143 Tumor vs LCL	Control LCL vs LCL
Insertion	20	2
Deletion	71	1
Inversion breakpoint	37	1
Translocation	45	0
Total	173	4

(Top) The table shows the number of somatic SVs detected in a breast tumor cell line sample (HCC1143) and absent in the matched blood cell line sample. SVs detected in HCC1143 but not detected in HCC1143BL were retested through molecule alignments to the SV with tumor and normal molecules to confirm somatic nature. As a negative control, the NA12878 sample was run twice, and there were no SV uniquely detected. (Bottom) A translocation is shown that disrupts RAP1B, a member of the RAS Oncogene Family in HCC1143. The spanning molecule pileups show strong molecule support for the translocation in the tumor but no evidence in the control cell line.



Conclusion

Bionano genome mapping with DLS reaches unprecedented genome assembly contiguity in a single assay for under $500. The high accuracy and ultra-long DNAs being analyzed provide the most broad range of structural variation sensitivity, from 500 bp up to cytogenetic scale variation, capturing insertions, deletions, inversions, translocations, repeat array expansion and collapse. Direct observation of ultra-long native DNA molecules results in higher confidence in structural variation detection by removing the reliance of inferring variation needed for analysis of short read data. This broad array of SV sensitivity, high performance, high throughput, low cost and automated bioinformatic workflow have stimulated utilization of the technology for multiple disease studies including intellectual disability, developmental disorder, FSHD, leukemia, myeloma, and other cancers.

References

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