Use of Bionano Optical Maps to Identify Medically-Relevant Genomic Variation

AWC Pang, J Lee, S Chan, W Andrews, T Anantharaman, A Hastie, H Cao

Bionano Genomics, San Diego, California, United States of America

Abstract

Structural variation (SV) detection is fundamental to understanding genetic diseases. While karyotyping and conventional molecular detection approaches are robust, they can be manually intensive, biased towards targeted loci, and cannot determine the copy number of long repeats. Bionano Genomics' Saphyr™ System offers an unbiased and sensitive method for detecting large SVs. Bionano's next-generation mapping (NGM) technique, relies on NanoChannel arrays to linearize and image high-molecular weight DNA, and *de novo* assemble megabases-long genome maps that preserve haplotypes with correct phasing. This method can capture over 4,000 insertions, deletions (>1kbp), translocations, and inversions in a single human sample.

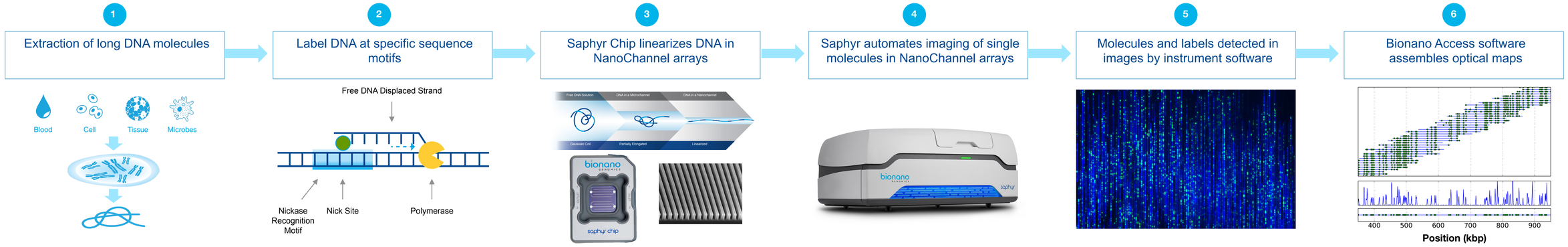
We present a novel software that examines genome map calls to identify candidate disease-associated SVs. To search for high-confidence calls, it searches for genomic elements such as centromeres and segmental duplications that may confound typical SV-calling algorithms, and evaluates quality scores of the assembly, alignments, and SVs detected. To further ascertain medically-relevant candidates, it filters the remaining calls with over 443,000 SVs collected from 144 individuals with no observable disease phenotypes, and compares with calls from disease-specific databases. Using this pipeline, whose total runtime is only a few hours, we can efficiently focus on several dozens of significant candidates for further analysis.

We ran specimens from individuals with hematological malignancies, applied the candidate SV-finding software, and identified known rearrangements such as the t(9;11)(q21.3;q23.3) in acute myeloid leukemia (AML) and t(9;22)(q34.12;q11.23) in chronic myeloid leukemia (CML). Moreover, we uncovered novel mutations ranging in size from a small 4.2 kbp insertion disrupting an acute lymphoblastic leukemia (ALL) gene *CNOT3* to a large 9.6 Mbp deletion at 17q25. Also, we applied NGM on rare and currently undiagnosed diseases. Using a family-based study design, we looked for rare or *de novo* SVs in affected individuals. In one quartet family with neurological disorder, we identified a 174 kbp heterozygous deletion shared by the two affected children and the carrier father. This deletion overlaps *ABCD3*, a member of the ATP-binding cassette gene family. In addition, in one affected sample, we detected a rare 364 kbp heterozygous inversion, which was also inherited from the father. Further experimental validation of these candidate SVs and comparison with the phenotypes are needed to establish disease association. In conclusion, Bionano optical maps can aid the discovery of functional SVs, and improve our understanding of the mechanisms of diseases.

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 whole genome. The resulting order and orientation of sequence elements in the map can be used for anchoring NGS contigs and structural variation detection.

Methods



(1) Long molecules of DNA are labeled with Bionano reagents by (2) incorporation of fluorophore labeled nucleotides 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 molecule are imaged and then digitized by the Saphyr instrument. (5) Molecules are labeled with a unique signature pattern that is uniquely identifiable and useful in assembly into genome maps. (6) Bionano maps may be used in a variety of downstream analysis using Bionano Access software.

Study 1: 10 samples with hematological malignancy

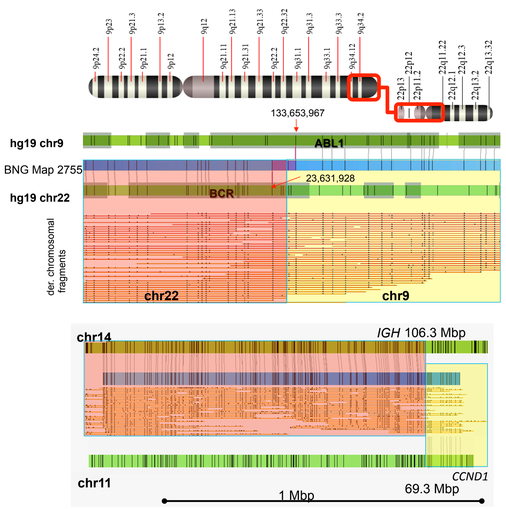


Figure 1. Examples of large chromosomal aberrations detected by mapping. (Top) A Philadelphia translocation t(9;22) was detected in leukemia cancer cells. The map was aligned to the public reference assembly hg19, and the resulting alignments show a conjoined junction between chr9 and chr22, creating a fusion gene called *BCR-ABL1*.
(Bottom) A t(11;14) translocation detected. The *IGH-CCND1* conjoined gene is formed from this translocation that is known to be associated with diseases involving B-lineage lymphocytes.

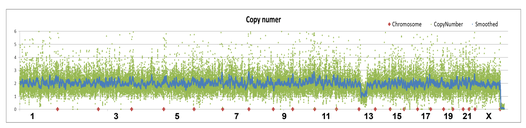
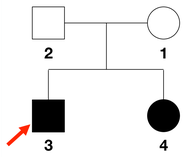


Figure 2. Whole genome depth of coverage shows chr13 monosomy. In addition to comparing assembled optical maps, we also aligned ~100x of molecules from a cancer sample to hg19, to detect copy number alterations. The green dots represent raw copy number count, while the blue dots represents smoothed copy number counts after removal of local noise.

Sample	Cancer associated events detected	Insertion disrupting cancer genes	Deletion disrupting cancer genes	Additional CNV/ translocation
S1	t(9;22): *ABL-BCR* translocation*			
S2	chr13 deletion*			t(9;13)
S3	t(11;14): *IGH-CCND1* translocation*	*CNOT3*	*TCF3*	
S4	chr20 monosomy	*CNOT3*	*TRIM33*	
S5	chr12 trisomy*			
S6		*NOTCH1*	*NUTM2B*	
S7		*NOTCH1*	*NUTM2B*	
S8	chr13 deletion			
S9				chr18 trisomy (low clonality)
S10	t(9;22): *ABL-BCR* translocation*			

Table 1. Cancer-associated mutations discovered in each sample. By running the variant-annotation pipeline, we are able to identify cancer-associated events. In particular, the calls indicated by asterisks are also detected by targeted FISH experiments.

Study 2: A quartet family with two children with neurological disorder



Bionano experiments were performed on the four members of this family. We are able to detect over 2500 SVs in each sample (**Table 2**). Our automated variant annotation pipeline determined the heredity of the calls, and selected 40-50 candidate rare variants (**Table 3**). **Figure 3** shows a deletion affecting the ABCD3 gene, which involves in the transport of fatty and bile acids into the peroxisome. Mutation in this gene is associated with an autosomal recessive disorder called congenital bile acid synthesis defect 5.

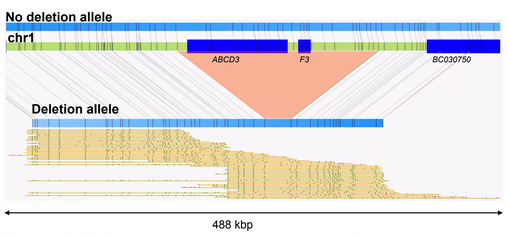


Figure 3. A 174.4 kbp rare heterozygous paternally-inherited deletion shared between the proband and the affected sister.

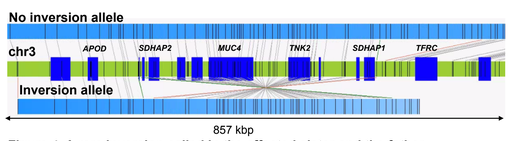


Figure 4. A rare inversion called in the affected sister and the father.

Sample	3 (Proband)	4 (Affected Sister)	1 (Mother)	2 (Father)
Data collected (molecules >150 kb)	402.40 Gbp	554.00 Gbp	465.11 Gbp	300.42 Gbp
Assembly size	2.86 Gbp	2.93 Gbp	2.90 Gbp	2.89 Gbp
Genome map N50	0.88 Mbp	1.60 Mbp	1.43 Mbp	1.32 Mbp
SV calls against hg19				
Insertions	1775	1862	1796	1762
Deletions	776	848	851	786
Inversions	15	20	14	20

Table 2. Assembly statistics and SV calls summary.

	3 (Proband)	4 (Affected Sister)
Insertions	10	20
Deletions	36	37
Inversions	0	1

Table 3. Rare variants identified in the proband and the affected sister. The calls are a subset of the Table 2 after removal of common calls found in Bionano's catalog of SV collected from >100 phenotypically "normal" samples.

Conclusions

We demonstrate that the Saphyr system can be used to accurately detect genetic mutation hallmarks in samples with hematologic malignancies. We were able to find known calls from cytogenetic experiments, and also detected novel aberrations. Especially useful for rare disease studies, researchers using this system can uncover rare and *de novo* variants by comparing with our control sample database and the unaffected parents, respectively. In the neurological disorder project, the Bionano SV workflow was able to automatically and reliably identify about 50 rare variants in each of the affected children. Our results shown here indicate that the Saphyr system can capture a broad spectrum of variation with functional importance, and can provide easy solutions for disease studies.

Acknowledgements: We would like to thank Jonathan Diver from Genoptix Medical Laboratory for providing the cancer samples and performing the FISH analysis, and Eric Vilain from the University of California, Los Angeles for providing the quartet samples.

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) Lam, E.T., et al. Genome mapping on NanoChannel arrays for structural variation analysis and sequence assembly. Nature Biotechnology (2012); 10: 2303
4) 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
5) Xiao, M et. al. Rapid DNA mapping by fluorescent single molecule detection. Nucleic Acids Research (2007); 35:e16.