Efficient Structural Variation Detection and Annotation Using Bionano Genome Mapping

A. W. C. Pang[1], J. Lee[1], A. Hastie[1], E. Lam[1], E. Chan[2], V. Hayes[2], H. Cao[1], M. Borodkin[1]

1) Bionano Genomics, San Diego, California, USA; 2) Garvan Institute of Medical Research, Darlinghurst, Australia

Abstract

Structural variation (SV) detection is fundamental to understanding cancer genomes. 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.

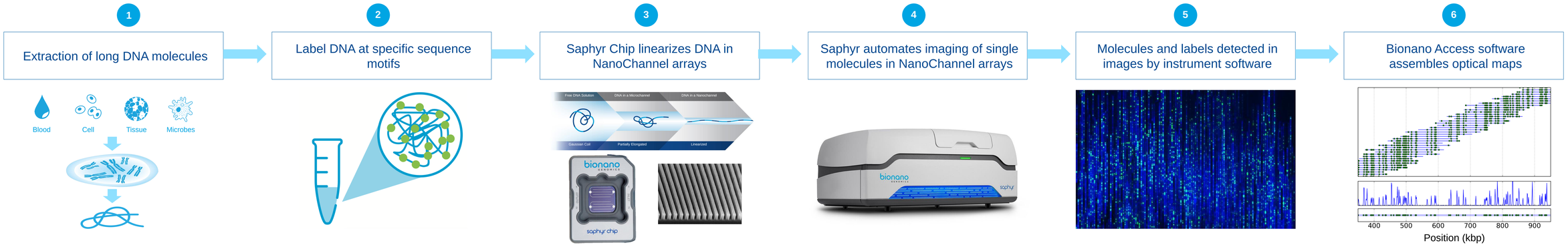
Bionano Genomics' Saphyr System offers a sensitive method for detecting large SVs. DNA molecules larger than 100 kbp are extracted, labelled at specific motifs, and linearized through NanoChannels arrays for subsequent visualization. Molecule images are digitized and the genome *de novo* assembled, creating megabase long Bionano maps. While our calling algorithm can sensitively detect all types of SVs in these long maps, we also use a variant annotation workflow to specifically uncover rare and sample-specific mutations. To determine variant frequency in a tumor normal experimental design, it compares the cancer sample's calls against over 600,000 SVs from >160 humans with no observable diseases. To identify somatic mutations, the workflow can compare against a control sample, and examines whether the cancer mutations are present in low fraction among the control's molecules. Using this pipeline, whose runtime is only a few hours, we can efficiently focus on several dozen significant candidates for further analysis.

We ran multiple solid and hematologic cancer samples. First, we generated a highly contiguous genome map assembly on a solid tumor sample. The ultra long maps – with lengths encompassing entire chromosomal arms – were able to capture large inversion events, and balanced and unbalanced translocations. In a separate study, we discovered known rearrangements such as the t(9;22)(q34.12;q11.23) in chronic myeloid leukemia (CML). Moreover, we uncovered a large megabases long deletion on chr13 by using our copy number detection tool. In conclusion, Bionano maps can discover 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. Bionano mapping 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.

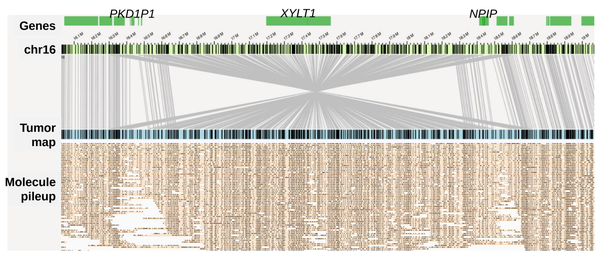
Methods



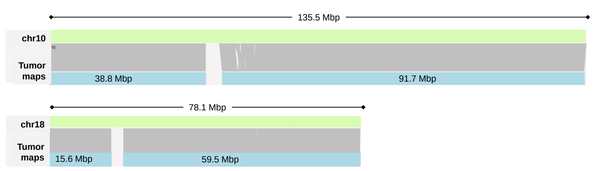
(1) Long molecules of DNA are labeled with Bionano reagents by (2) incorporation of fluorophores at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the Saphyr Chip using NanoChannel arrays (4) Single molecules are imaged by Saphyr and then digitized.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

Study 1: Direct labeling experiment of a solid tumor sample

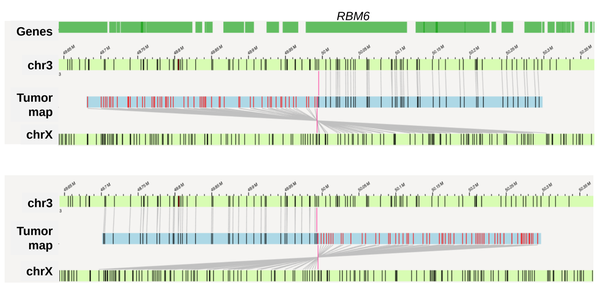
DLE1 labeled sample	Solid tumor
Data collected (molecules >150 kb)	367 Gbp
Assembly size (haplotype-aware)	5.87 Gbp
Genome map N50	**71.0 Mbp**
SV called against hg19	
Insertions	3,894
Deletions	1,799
Inversion breakpoints	95
Interchromosomal translocations	33
Intrachromosomal translocations	21



A 2.3 Mbp inversion on chr16 captures by a 32.2 Mbp *de novo* assembled tumor genome map. Note that the breakpoints contain some additional sequences that are novel to the reference genome hg19. The presence of overlapping alignments indicates homology (in opposite direction) between the two inversion breakpoints.

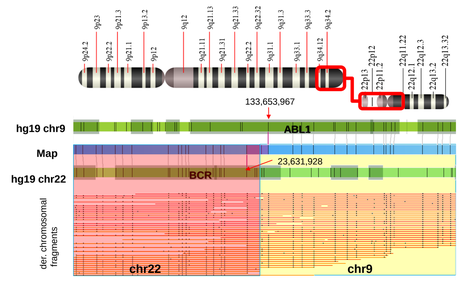


Two examples of ultra long genome maps spanning entire arms of chromosomes 10 and 18. The maps are only separated by centromeres.

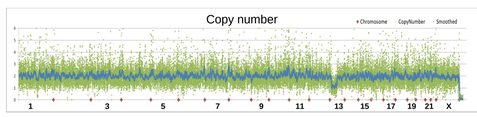


(Top) A 596.9 kbp genome map captures a t(3;X) translocation. The translocation breakpoint interrupts the gene RBM6 on chr3. RBM6 codes for a RNA-binding protein 6, and mutations and copy number mutations have been found in many types of cancer.
(Bottom) An alternate map is able to capture the reciprocal translocation event.

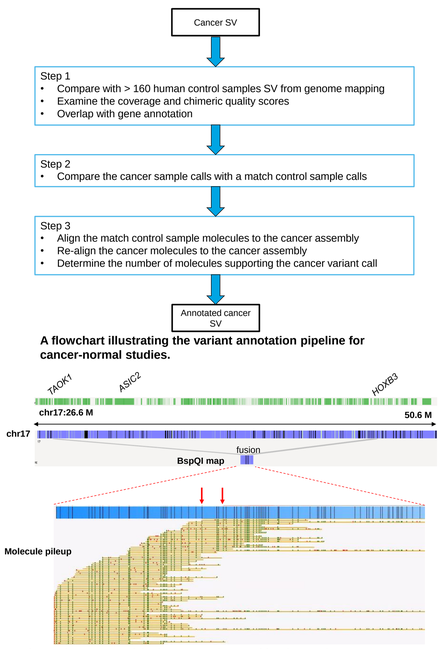
Study 2: SV results on samples with hematological malignancy



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*.



A whole genome depth of coverage shoring a large chr13 deletion. The green dots represent raw copy number count, while the blue dots represent smoothed copy number count after the removal of local noise



A flowchart illustrating the variant annotation pipeline for cancer-normal studies.

An intra-chromosomal fusion event detected in chr17. The two genes located at the junctions are *TAOK1* and *HOXB3*. The red arrows indicate the variant junction on the genome map.

Conclusions

We demonstrate that the Saphyr system can be used to accurately detect genetic mutation hallmarks in samples with solid tumor and hematologic malignancies. These includes large rearrangements ranging from interchromosomal translocations, within chromosome fusions, to copy number alterations. Researchers can perform mapping experiments to uncover somatic variants by comparing with our control sample database and a matching non-tumor sample. 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 cancer studies.

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.