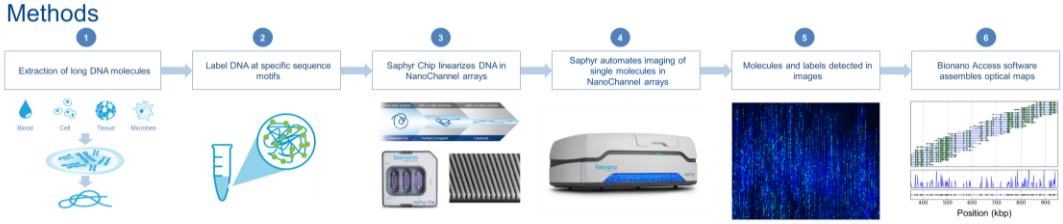
Optical Genome Mapping for Detection of Structural Variants in Constitutional Disease.

A Hastie, A. W. C. Pang, J. Lee, E. T. Lam, T. Anantharaman, H. Sadowski, M. Borodkin
Bionano Genomics, San Diego, California, United States of America

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.

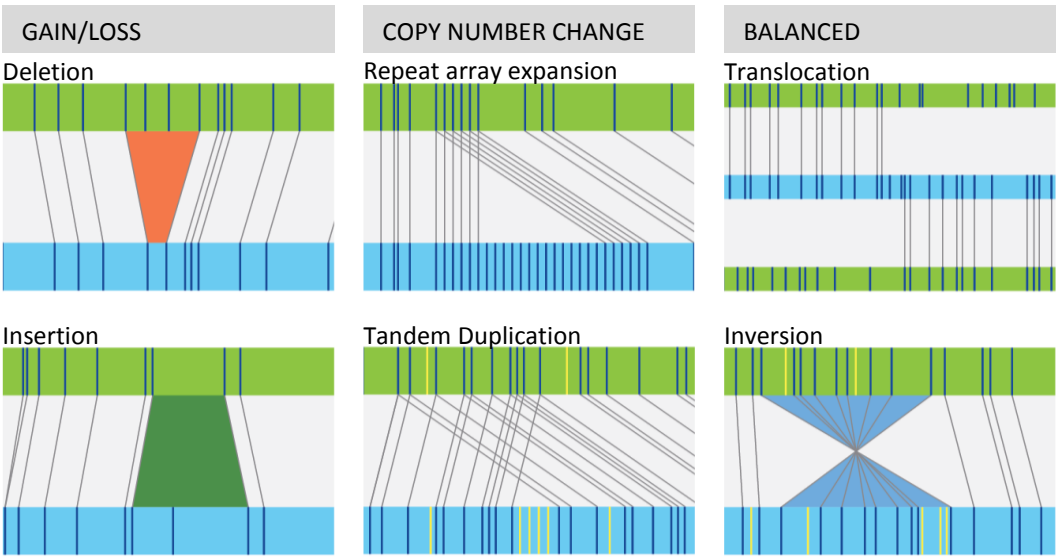
Abstract

Bionano Genome mapping is quickly becoming established as a key method for detection of intractable types of structural variation. Currently, genomes from many different disease states, some of which do not have a molecular diagnosis and others which are very difficult to diagnose are being studied. For example FSHD1, is caused by a repeat collapse of a tandem repeat array with unit sizes of 3.4 kilobases each. This repeat array is generally only be measured by southern blot, a labor intensive and low resolution approach. Another disease type that is difficult to diagnose is triplet expansion diseases such as Fragile X syndrome and Myotonic Dystrophy, these repeat arrays can expand to many kilobases. Microdeletions and microduplications, which cause diseases such as DiGeorge syndrome and other disease syndromes, are detectible by microarrays as well as WGS but they are caused by rearrangements between large segmental duplications that flank the region in question and the segmental duplications are extremely hard to study with conventional tools. Genome mapping can accurately assemble and assay relevant regions for each of these disease classes, even those involving very large segmental duplications. Bionano has built bioinformatics tools to effectively prioritize the ~6000 structural variants based on the estimated frequency in a control population, whether it's inherited or de novo, whether it's somatic and also its proximity to a gene. We provide several examples of pathogenic variants found through Bionano genome mapping.

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.

Detection of structural variation by *de novo* genome map construction and comparison.



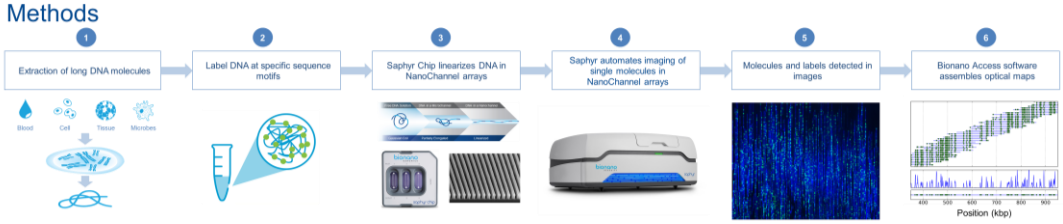
Reference maps (green map bars) are created by processing the human reference sequence (hg19 or hg38) to find and mark all locations of the sequence motif (vertical lines on the green bars) which is recognized by the Bionano optical mapping enzyme (e.g. DLE-1, Nt.BspQI, Nb.BssSI), which occurs every 5 kbp on average. Genome maps (blue bars) are produced from images of labeled DNA molecules >150 kbp.

Insertions and deletions >500 bps are detected by differences in the length of a segment compared to the reference. Tandem repeat expansions can be marked by a labeling enzyme in each unit, when not marked in each unit, a repeat expansion will be called as an insertion and the repeat number could be inferred. Tandem duplication can be direct or inverted, shown is a direct tandem duplication which can be seen by the observation of the same pattern mapping the same reference segment two times in the same orientation. Translocations are seen in maps that map to two different chromosomes and inversion are a balanced event where a map pattern is in a different orientation with respect to the reference.

Optical Genome Mapping for Detection of Structural Variants in Constitutional Disease.

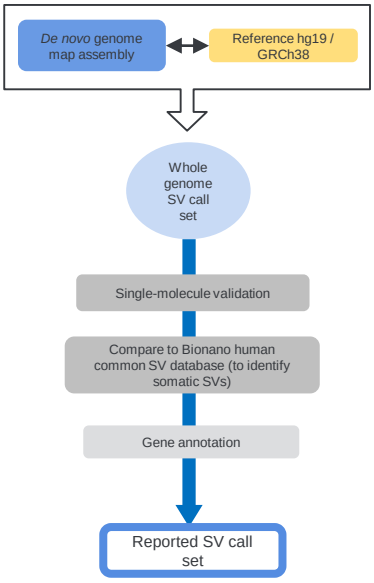
A Hastie, A. W. C. Pang, J. Lee, E. T. Lam, T. Anantharaman, H. Sadowski, M. Borodkin
Bionano Genomics, San Diego, California, United States of America

Methods

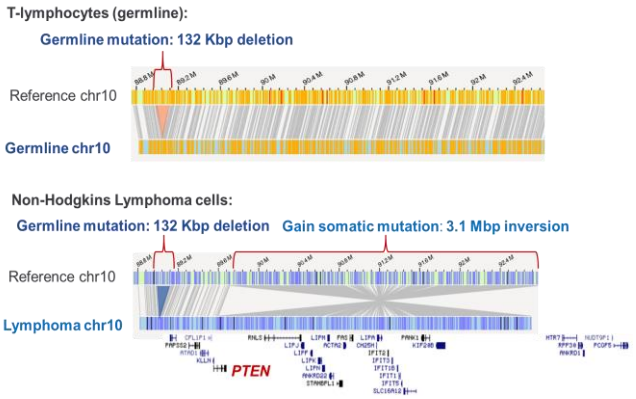


(1) Long molecules of DNA are labeled with Bionano reagents by (2) incorporation of fluorophores at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the Saphyr Chip using NanoChannel arrays (4) Single molecules are imaged by Saphyr and then digitized.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

Variant Annotation Pipeline for annotation and prioritization

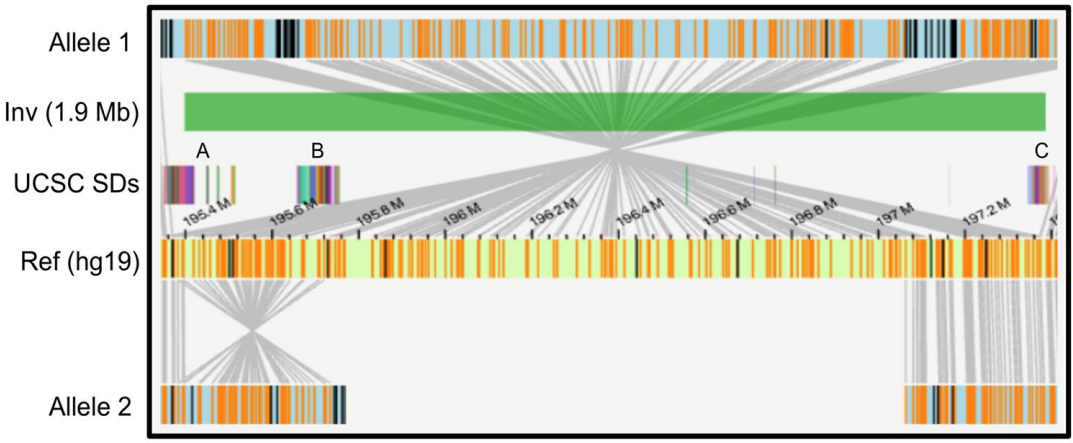


Somatic SV in Non-Hodgkins Lymphoma



Whole genome structural variants are annotated according to the flowchart on the left. By using a control database to estimate the allele frequency in a normal/healthy population, retesting for accuracy by genotyping, and annotating SVs that overlap with genes. This information can be used to quickly prioritize for clinically relevant variants and other potentially relevant variants. On the right, the variant annotation pipeline was used to prioritize variants by their absence from the control database, overlapping genes. In inversion was detected that breaks the PTEN oncogene.

Structure of the reoccurring microdeletion syndrome region 3q29 revealed multiple inversions related to segmental blocks.

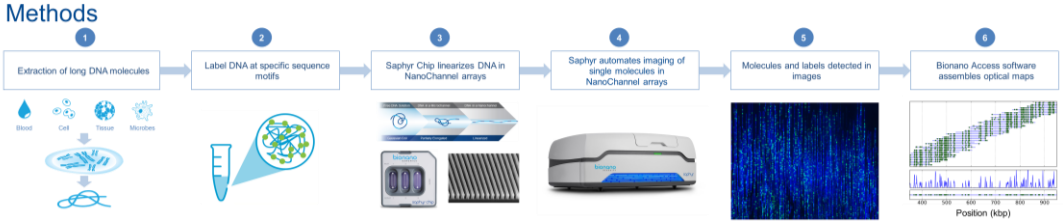


Despite the importance of microdeletions/duplications to disease, very little is known about the flanking segmental duplications that are thought to influence the genomic changes leading to deletion and duplications. Using optical genome mapping we can resolve the structures in patients and other family members. We found multiple inversions in patients and controls apparently mediated by segmental duplications (SDs).

Optical Genome Mapping for Detection of Structural Variants in Constitutional Disease.

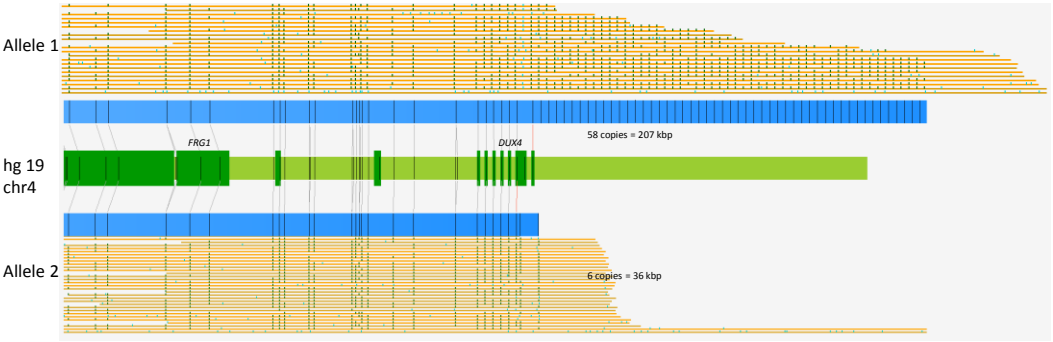
A Hastie, A. W. C. Pang, J. Lee, E. T. Lam, T. Anantharaman, H. Sadowski, M. Borodkin
Bionano Genomics, San Diego, California, United States of America

Methods



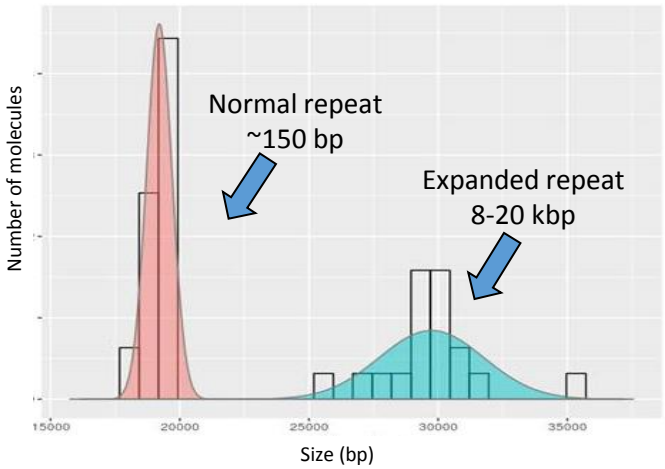
(1) Long molecules of DNA are labeled with Bionano reagents by (2) incorporation of fluorophores at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the Saphyr Chip using NanoChannel arrays (4) Single molecules are imaged by Saphyr and then digitized.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

Facioscapulohumeral (FSHD) muscular dystrophy – extremely long tandem repeat array measurement.



Bionano genome mapping identified copy number difference of the D4Z4 macrosatellite repeat at the subtelomeric region of 4q in an individual with Facioscapulohumeral muscular dystrophy (FSHD). FSHD is associated with expression abnormality of the *DUX4* gene in muscle cells; this gene is located in the D4Z4 repeat locus. The majority of people with FSHD have less than 10 copies of D4Z4. The shortening of the repeats is believed to be linked to the inappropriate expression of the *DUX4* gene. Here, this affected sample carries one allele with a normal copy number (58 copies) and one with a deleterious copy number (6 copies).

Repeat expansion disorders: single molecule measurement of DMPK gene expansion in a Myotonic Dystrophy case

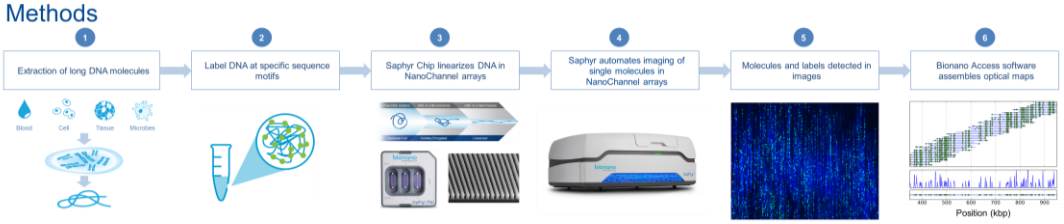


Measurement of DMPK tandem repeat array varies from normal cases of 15-150 bp consisting of a triplet CTG repeat. The array can become unstable and can expand to 1000's of basepairs. In order to provide a molecular diagnosis for Myotonic Distrophy, an expanded repeat array must be measured. NGS sequencing is unable to measure these repeat arrays when extended beyond a critical length. Bionano was used to measure the lenth of a normal repeat of 150 bp (which is within an map interval of 19kbp) and also measure an extended repeat array with varying lenthts between 8 and 20 kbps.

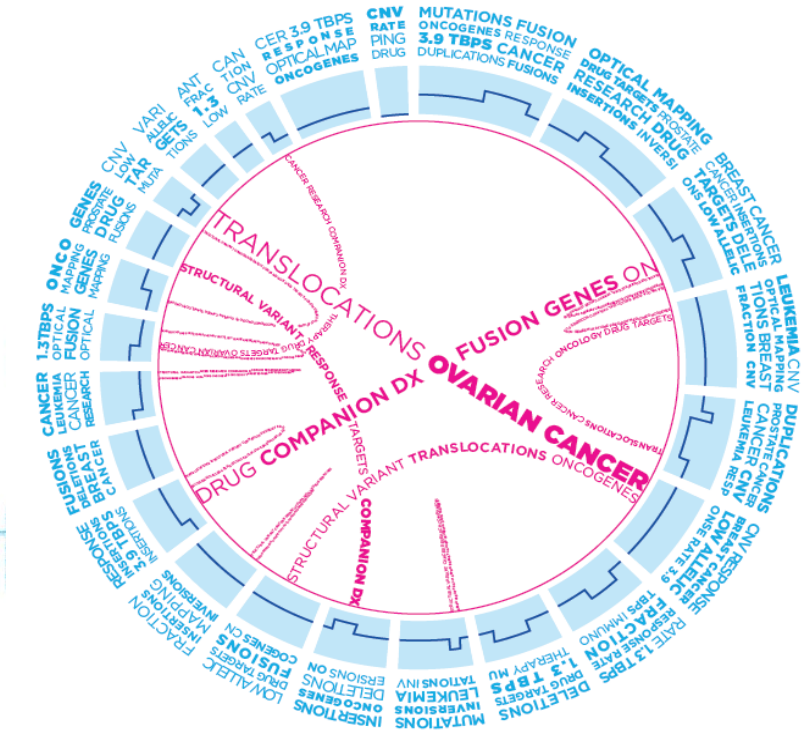
Optical Genome Mapping for Detection of Structural Variants in Constitutional Disease.

A Hastie, A. W. C. Pang, J. Lee, E. T. Lam, T. Anantharaman, H. Sadowski, M. Borodkin
Bionano Genomics, San Diego, California, United States of America

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.



Contact: ahastie@bionanogenomics.com
www.bionanogenomics.com