# bionano GENOMICS

# Whole Genome Imaging for Detection of Structural Variants in Constitutional Disease

J C Lai, A Hastie, J Lee, E T Lam, B Clifford, S Bocklandt, S Oeser, T Anantharaman, H Sadowski, A W Pang
Bionano Genomics, San Diego, California, United States of America

## Abstract

Majority of the human diseases have a genetic underlying component and structural changes to genome often contribute to the complexity of studying genetic disorders. Understanding these structural changes is critical to develop appropriate diagnostic methods as well as for developing therapeutic options. For example, facioscapulohumeral muscular dystrophy (FSHD), is caused by a collapse of a tandem repeat array with unit sizes of 3.4 kilobases each. This repeat array generally can only be measured by southern blot, a labor intensive and low-resolution approach. Another disease family that is difficult to detect is triplet expansion diseases such as fragile X syndrome and myotonic dystrophy, in which repeat arrays can expand to many kilobases. Microdeletions and microduplications, which cause diseases such as DiGeorge syndrome and other syndromes, are detectable by microarrays as well as whole genome sequencing but large rearrangement cannot be detected with conventional technologies or whole genome sequencing. Here, we describe Bionano Whole Genome Imaging which is quickly becoming established as a key technology for detecting complex genome-wide structural changes. The Bionano Saphyr workflow can accurately assemble and assay relevant regions for each of above mentioned disease classes, even those involving large segmental duplications. Bionano has developed bioinformatics tools to effectively prioritize the ~6000 genome wide structural variants based on the estimated frequency in a control population, whether it's inherited or de novo, whether it's somatic and also in proximity to a gene.

## 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. We provide several examples of pathogenic variants found through Bionano whole genome imaging that enhance the understanding of genetic disorders.

## Methods
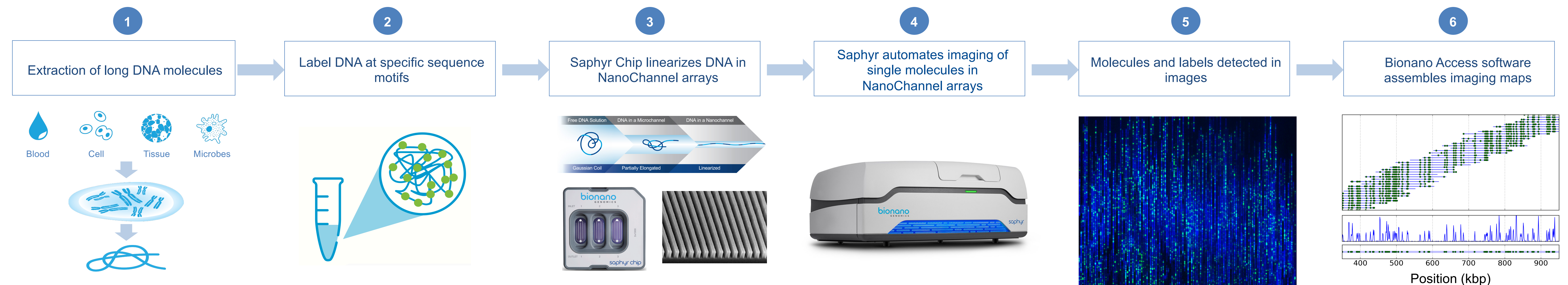


**(1)** Long molecules of DNA are labeled with Bionano reagents by **(2)** incorporation of fluorophores at a specific sequence motif throughout the genome. **(3)** The labeled genomic DNA is then linearized in the Saphyr Chip using NanoChannel arrays **(4)** Single molecules are imaged by Saphyr and 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
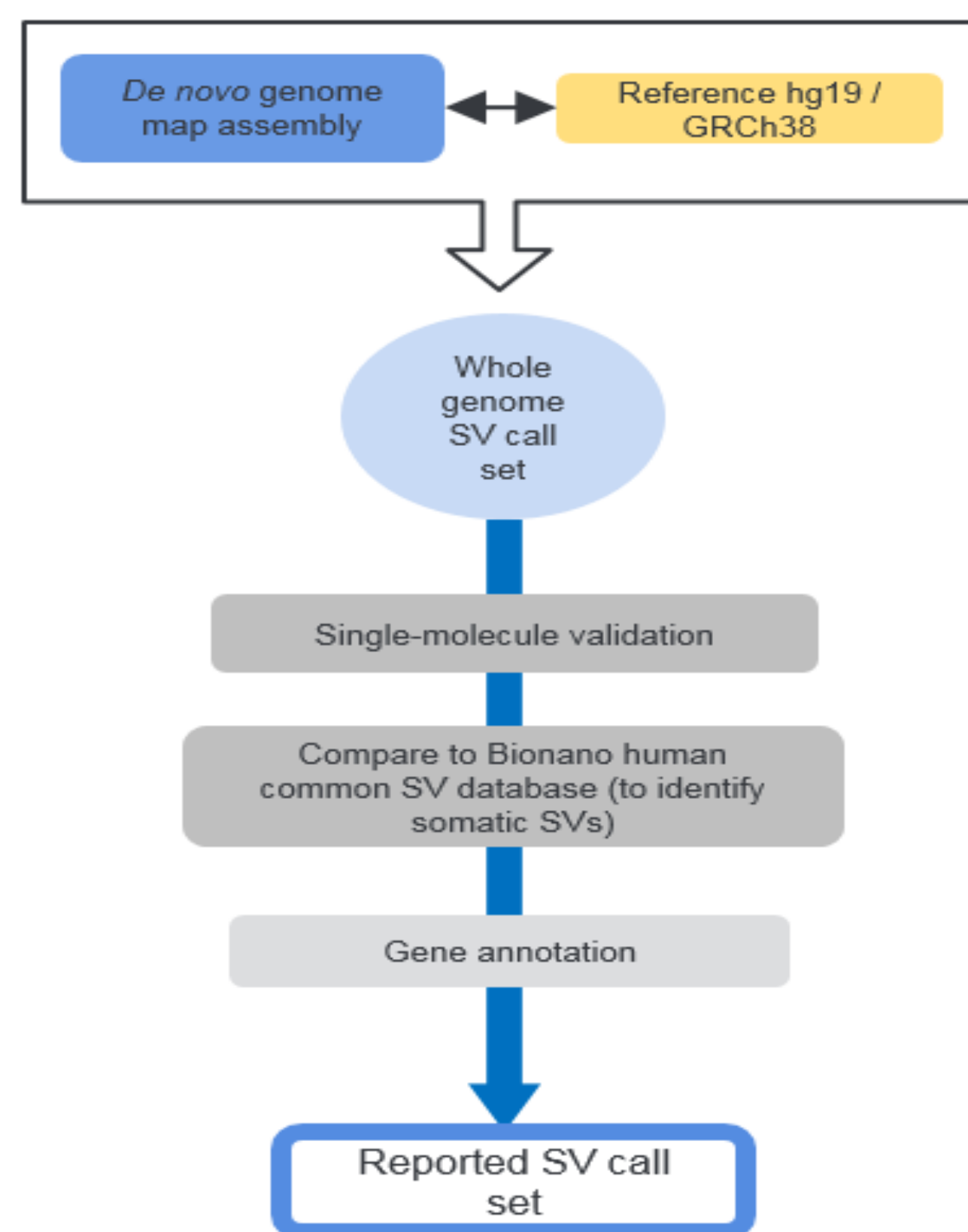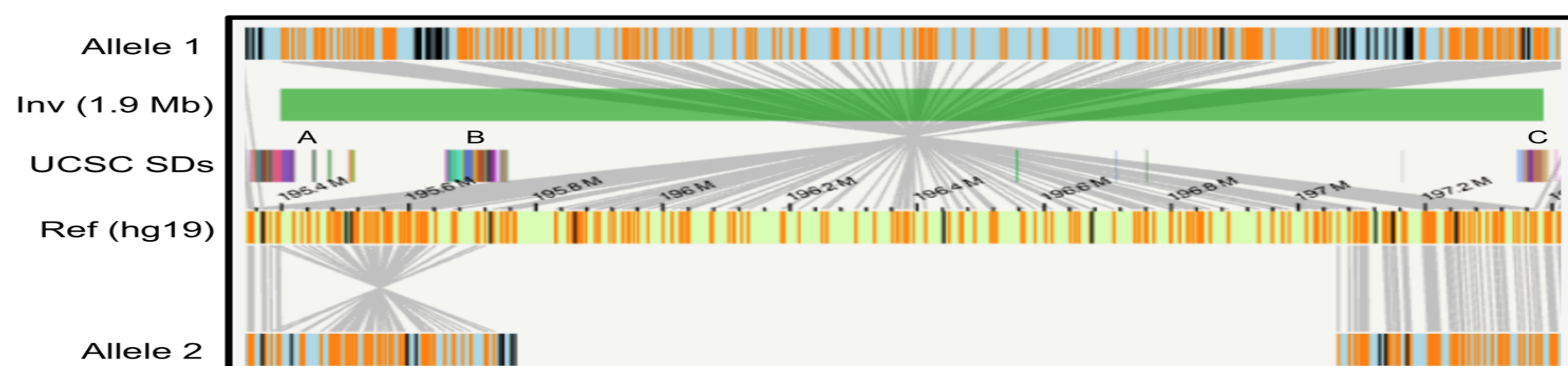


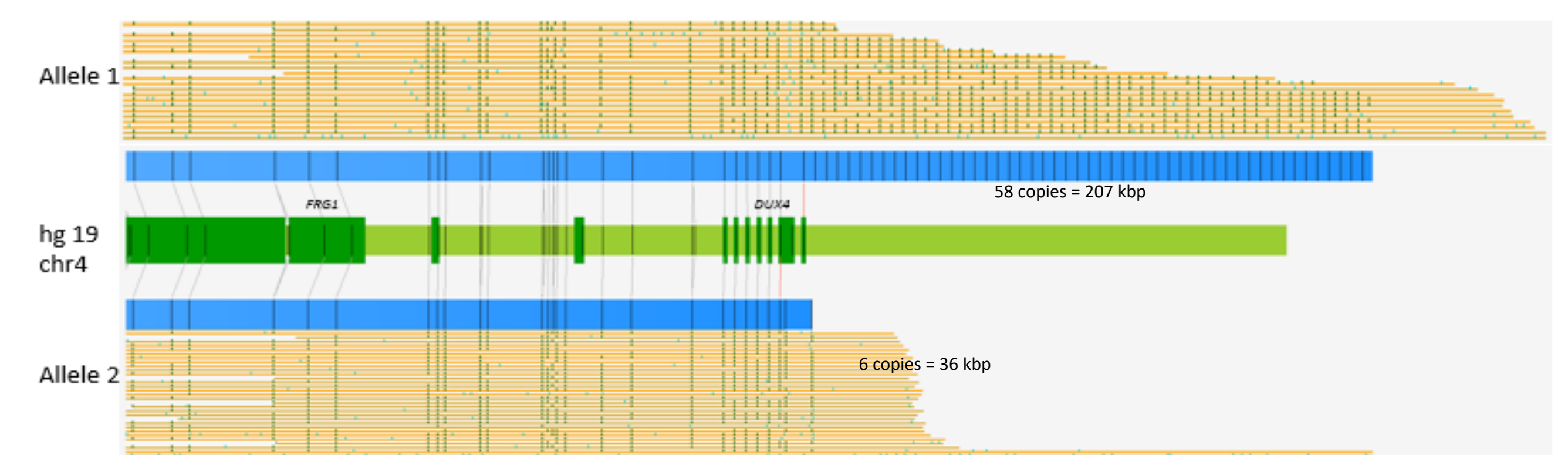## Somatic SV in Non-Hodgkins Lymphoma
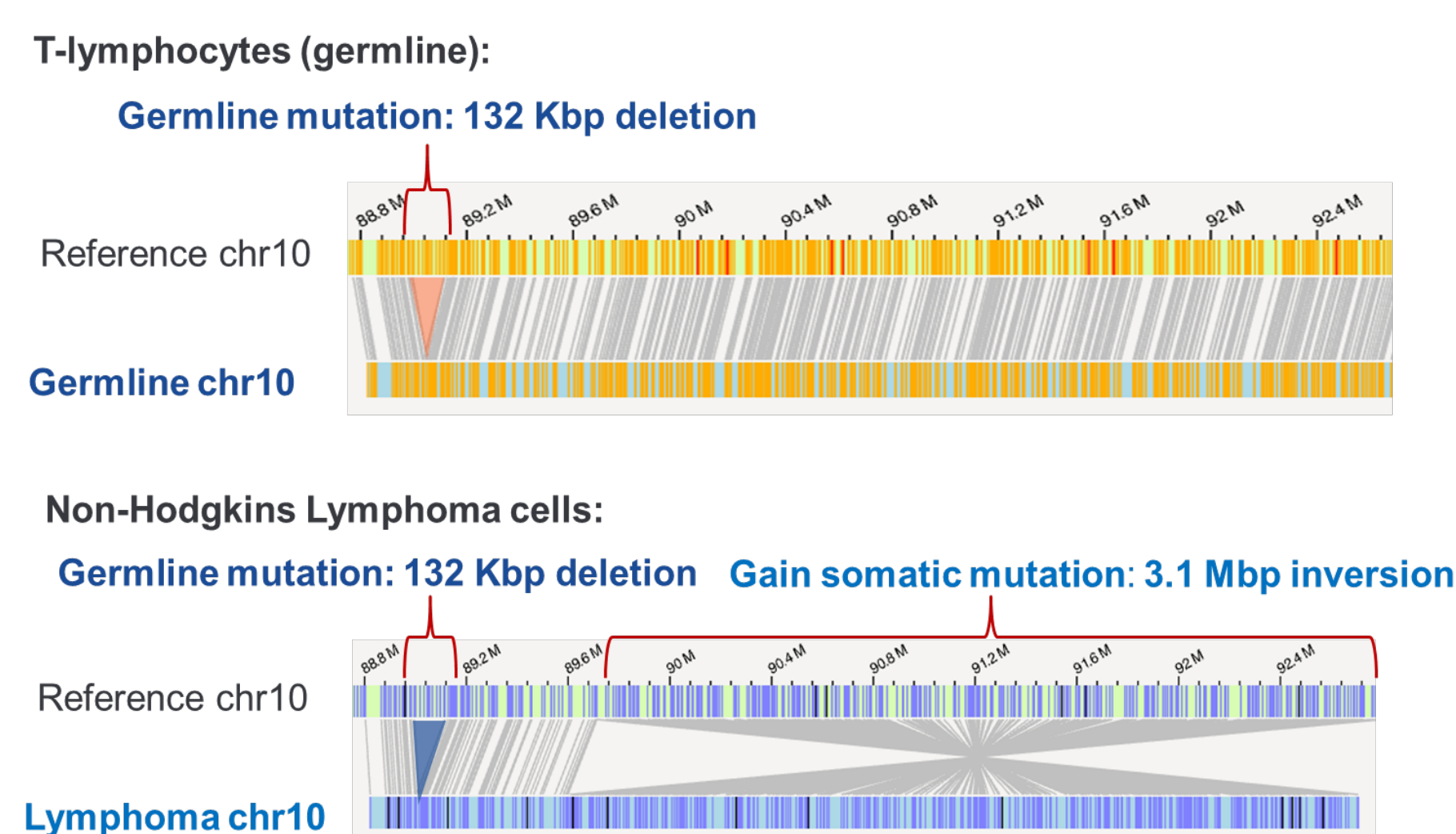


Whole genome structural variants are annotated according to the flowchart on the left. By using a control database to estimate the allele frequency in a normal/healthy population, retesting for accuracy by genotyping, and annotating SVs that overlap with genes. This information can be used to quickly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

## 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
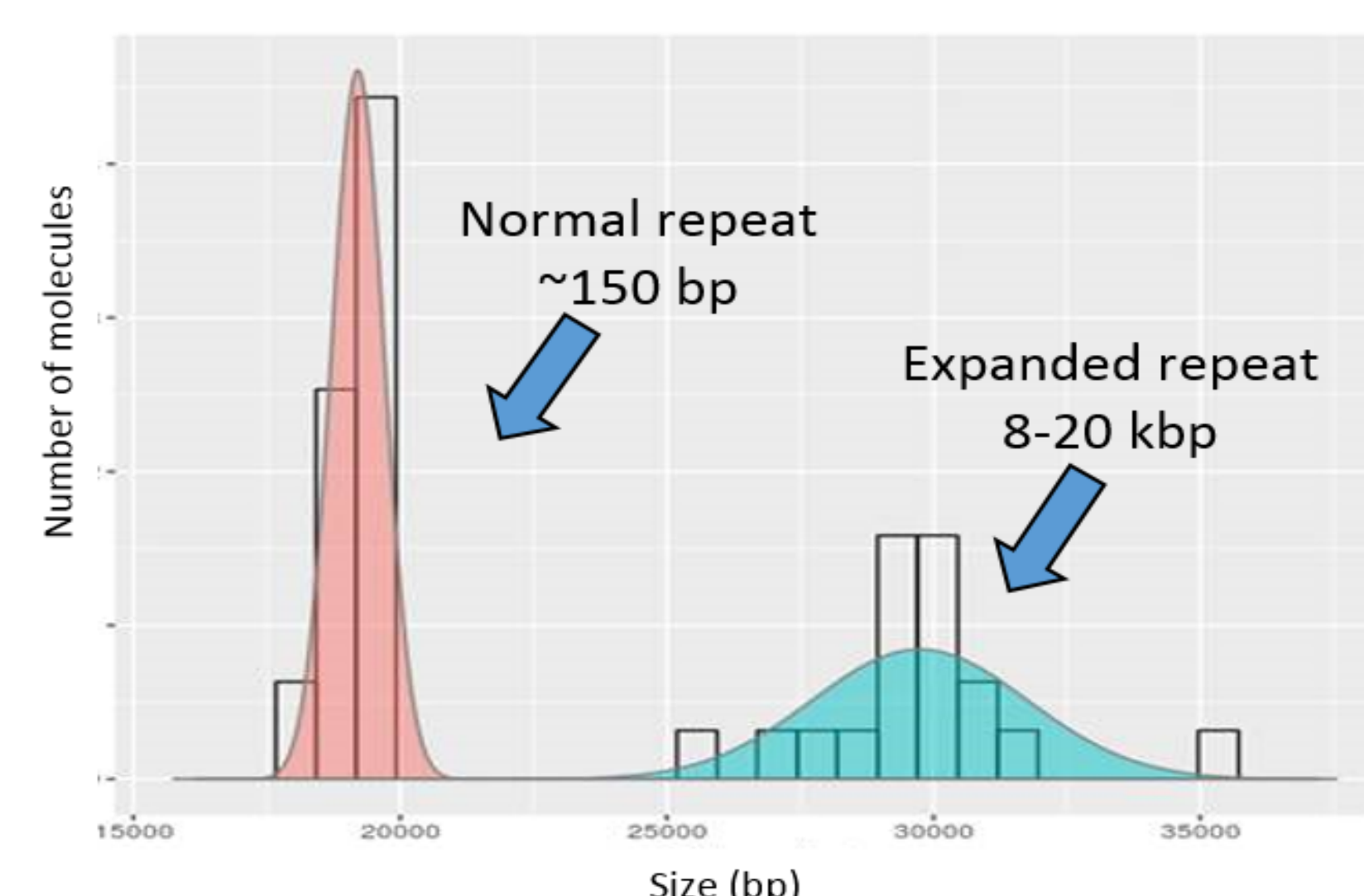


Measurement of DMPK tandem repeat array varies from normal cases of 15-150 bp consisting of a triplet CTG repeat. The array can become unstable and can expand to 1000's of basepairs. In order to provide a molecular diagnosis for Myotonic Distrophy, an expanded repeat array must be measured. NGS sequencing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

## Structure of the reoccurring microdeletion syndrome region 3q29 reveailed 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 genome imaging 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).