

Contiguous Assembly of Bionano DLS Chemistry Reveals Structural Variants at Challenging Regions

Joyce Lee¹, Andy Wing Chun Pang¹, Thomas Anantharaman¹, Ernest T Lam¹, Alex Hastie¹, Mark Borodkin¹
¹Bionano Genomics, San Diego, California, United States of America

Abstract

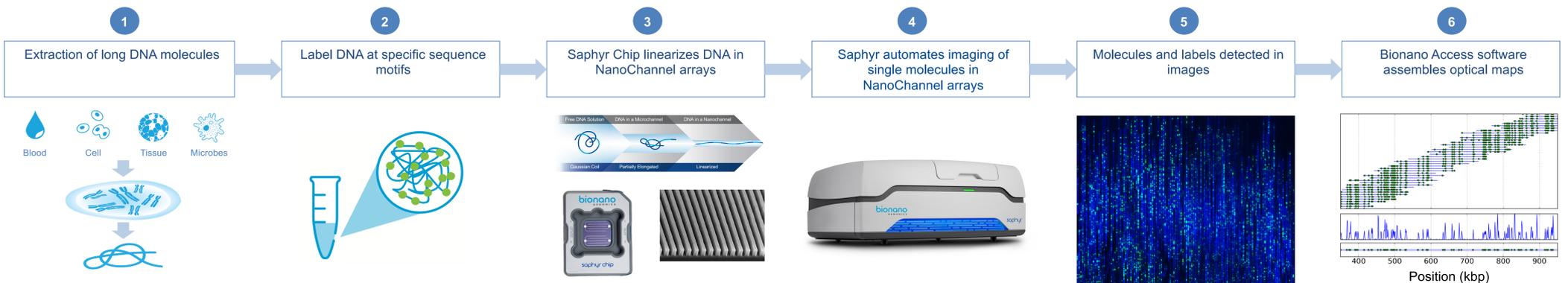
The human reference, hg38, released in 2013 is by far the best annotated human assembly and has been used in many clinical studies. This current reference is about 3.2 Gbp; it is a haploid representation of the human genome with about 200 Mbp of gaps. Since a complete and accurate anchor is very important for clinical studies, closing of these gaps and accurate representation of challenging regions, such as tandem repeats and segmental duplication regions, are important. With the advancement of technologies, recent studies have also made an effort in identifying structural variants that are due to the incompleteness or misrepresentation of the hg38 reference.

One of the advanced technologies is Bionano Genomics' Saphyr™ system and its direct labelling and staining (DLS) chemistry; physically intact molecules of this technology have N50 length of about 250 kbp and maximum length of about 2.5 Mbp. Challenging regions of the human genome, including regions that are incomplete or repeats that are collapsed in the reference, become accessible and easy to assemble with high accuracy using these ultra-long molecules. Moreover, with these contiguous molecules, assemblies with N50 of about 60 Mbp are assembled with the well understood overlap layout consensus algorithms. Such contiguous assembly provides the basis for high sensitivity in structural variation detection, which is very important in clinical studies.

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.

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.

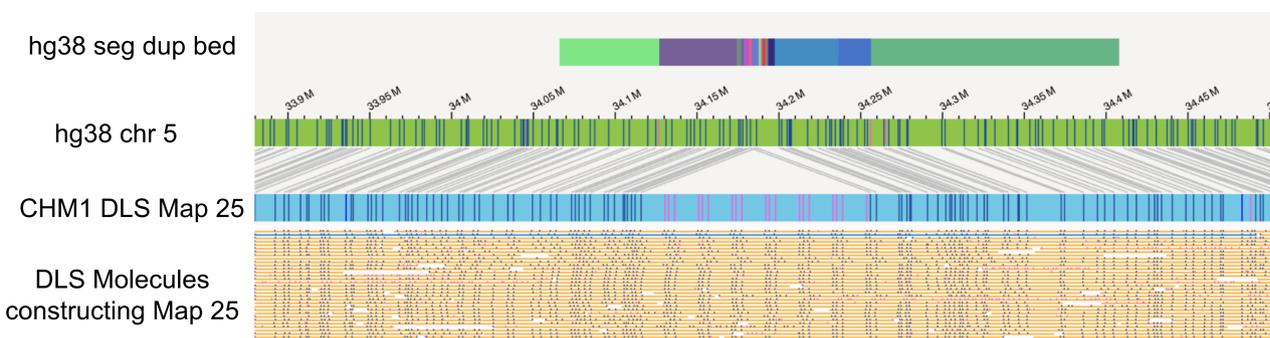


Figure 2. A 144 kbp repeat expansion at a segmental duplication region is uniquely detected by Bionano. Long DLS molecules spanned the repeat expansion and the whole region of segmental duplications.

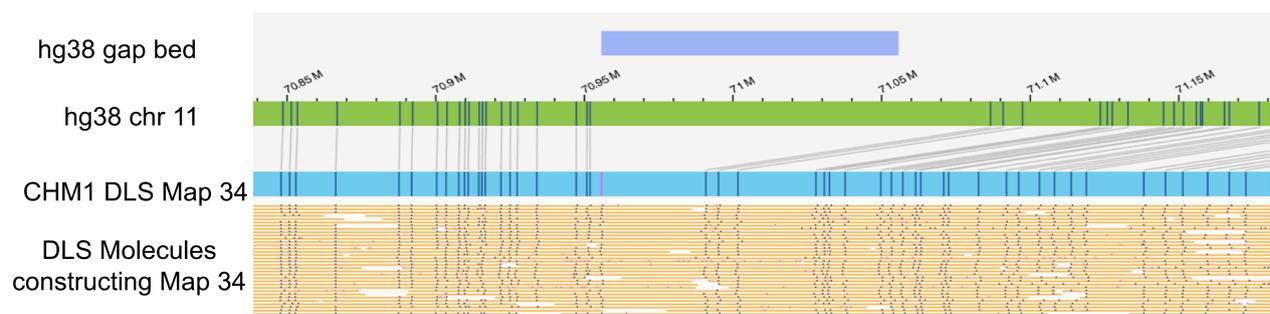


Figure 3. An over estimation of gap length in hg38 by 96 kbp in CHM1. Sequence technology is only able to assemble through unique or short repeats while Bionano can unambiguously span large complex regions.

Results

One of the latest efforts to close gaps and resolve challenging regions in the hg38 reference used a combination of multiple platforms and analyzed structural variants called from fifteen genomes (*Audano et al.*). Using a 2.99 Gbp CHM1 assembly with N50 of 60 Mbp constructed from Bionano DLS molecules, structural variants with respect to hg38 were extracted. About 78% of the 811 deletions and 1989 insertions overlapped the structural variants found in Audano's study (Figure 1); with about 37% and 46% of the Bionano unique insertions and deletions involved in segmental duplication regions respectively (Figure 2), and with about 2.5% and 28% of the Bionano unique insertions and deletions overlapped sequence gap regions in hg38, respectively (Figure 3). Most of the Bionano calls >15 kbp do not have concordant overlaps despite strong intact DLS molecule support; this might be caused by the uniqueness in detecting large structural variants using ultra-long molecules.

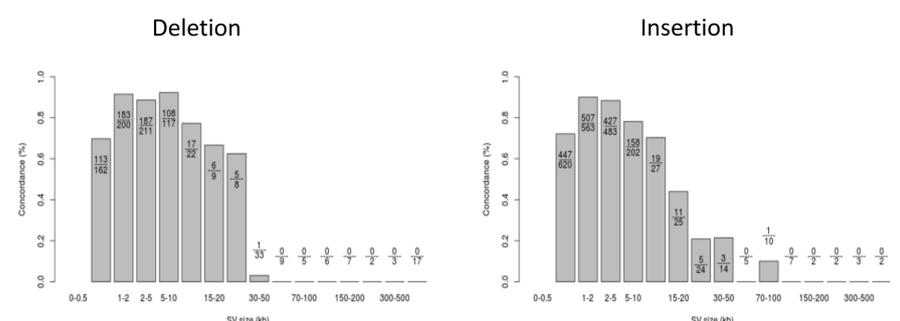


Figure 1. About 78% of Bionano detected insertions and deletions are concordant with those in Audano et al. Most insertions and deletions >15kbp are detected uniquely by Bionano.

Conclusions

In conclusion, Bionano DLS chemistry provides contiguous assemblies with high accuracy; given such assembly, structural variants in challenging regions are revealed and well supported by intact ultra-long molecules. This helps in resolving challenging regions and covering gaps in the current hg38 reference.

Reference

- Audano, Peter A., et al. "Characterizing the Major Structural Variant Alleles of the Human Genome." *Cell* (2019).
- Pendleton, M., Sebra, R., et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nature Methods* (2015); e3454
- Cao, H., et al., Rapid detection of structural variation in a human genome using NanoChannel-based genome mapping technology. *Gigascience* (2014); 3(1):34
- 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.