Bionano Hybrid Scaffold Assemblies Provide High Contiguity and Accuracy

Joyce Lee, Andy Wing Chun Pang, Saki Chan, Jian Wang, Thomas Anantharaman, Alex Hastie
Bionano Genomics, San Diego, California, United States of America

Abstract

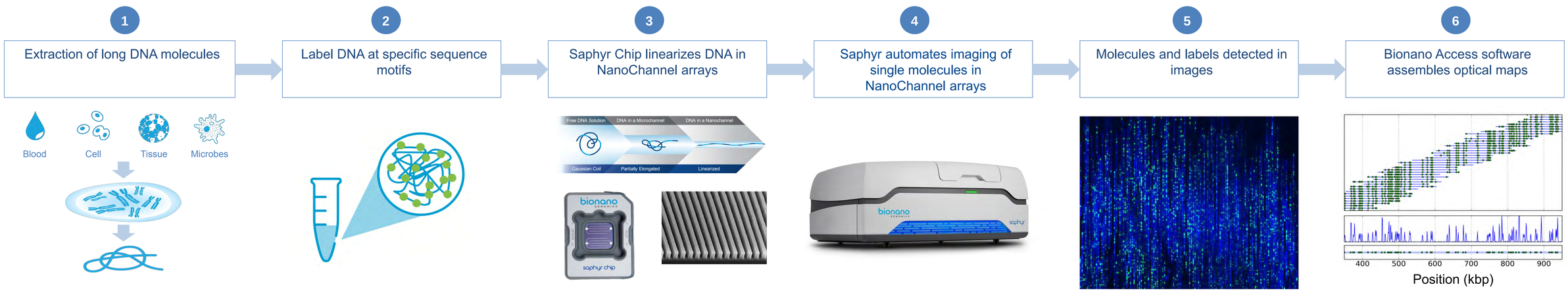
High quality genome assembly of novel genomes has gained momentum in recent years as a result of the advancement of new technologies. To disambiguate homologous regions of these novel genomes, long reads and linked reads are used to assemble contigs. Scaffolding these sequences into chromosome arm or full chromosome length can only be accomplished using Bionano Genomics' long read mapping or one of the Hi-C based methods. In comparing the assemblies of these scaffolding technologies, we demonstrate that Bionano can correct sequence errors and orientation errors generated by other technologies.

Chromosome assemblies can often be achieved using Bionano's new DLS chemistry and/or by combining with the NLRS chemistry. Bionano genome mapping of physically intact molecules of hundreds of kilobases, is unique and generates tens of megabases long contiguous assemblies with the well understood overlap layout consensus algorithms. These assemblies are then used to scaffold sequences into chromosome or chromosome arm length assemblies by the Bionano Hybrid Scaffold pipeline. Alternatively, Hi-C based methods leverage crosslinking of DNA that is in close proximity in vivo through chromatin folding, which is then sequenced using short read sequencing. Since the long range interaction in Hi-C is based on cells at different stages of dynamic biological connections and is encoded by short reads, significant inference is required to reconstruct the interaction information.

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.

Hybird Scaffold Statistic	Bionano	Sequence	Sequence used in Hybrid Scaffold	Hybrid Scaffold	Hybrid + not scaffolded seq
Number of maps	69	10642	2957	24	7866
N50 (Mbp)	116.82	0.50	0.70	119.67	115.61
Total length (Mbp)	1625.61	1591.45	1228.43 (77%)	1575.75	1936.93

Figure 1. Bionano Hybrid Scaffolding provides High Contiguity.

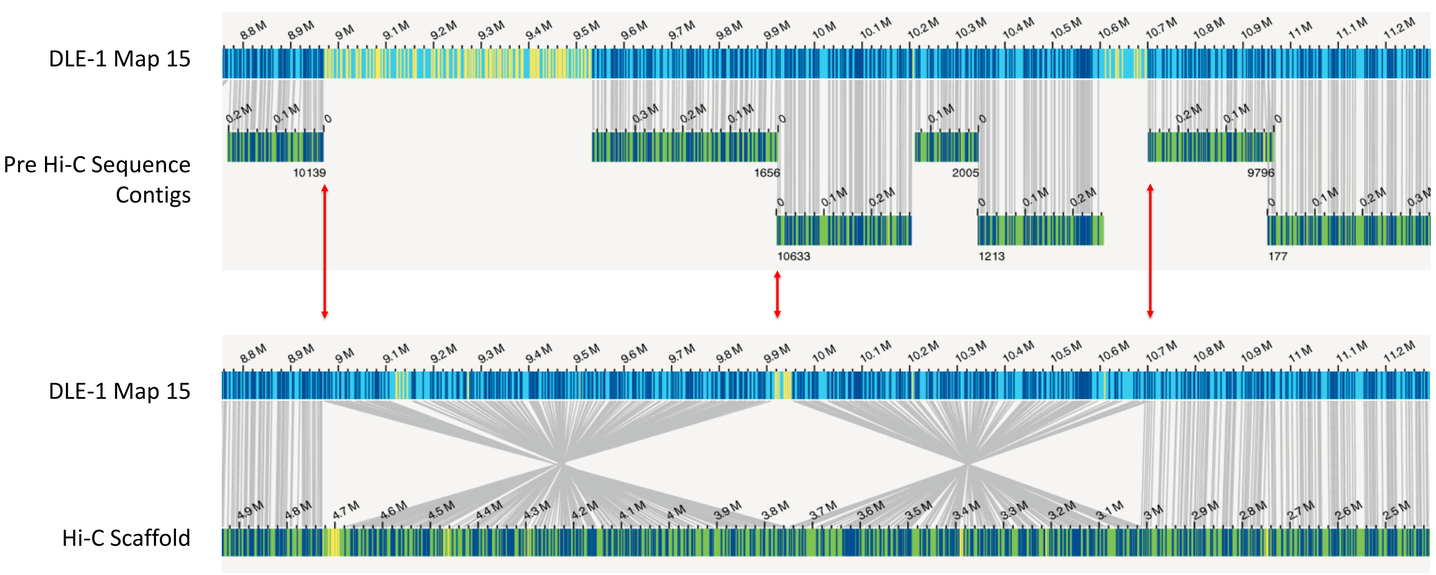


Figure 2. Discrepancies between Bionano maps and sequence assemblies were introduced after Hi-C scaffolding. Bionano automatically corrects errors in sequence scaffolds and also provides superior contiguity.

Results

In comparing Hi-C scaffolds with Bionano DLE Maps, many discrepancies were found. Inverted and rearranged segments of various sizes were identified when aligning the Hi-C scaffolds to the Bionano DLE Maps. Aligning pre Hi-C scaffolded sequences with Bionano DLE maps also shows that most of the divergent breakpoints were at ends of the pre Hi-C scaffolded contigs, suggesting that orientation and arrangement discrepancies were introduced during Hi-C scaffolding (Figure 2). Since the Bionano method leveraged native intact molecules, spanning breakpoints of these segments by Bionano molecules suggests mis-orientations in the Hi-C scaffolds. At regions of discrepancies, a third long read sequencing technology also agrees with Bionano structure, further supporting the accuracy of the Bionano assembly (Figure 3).

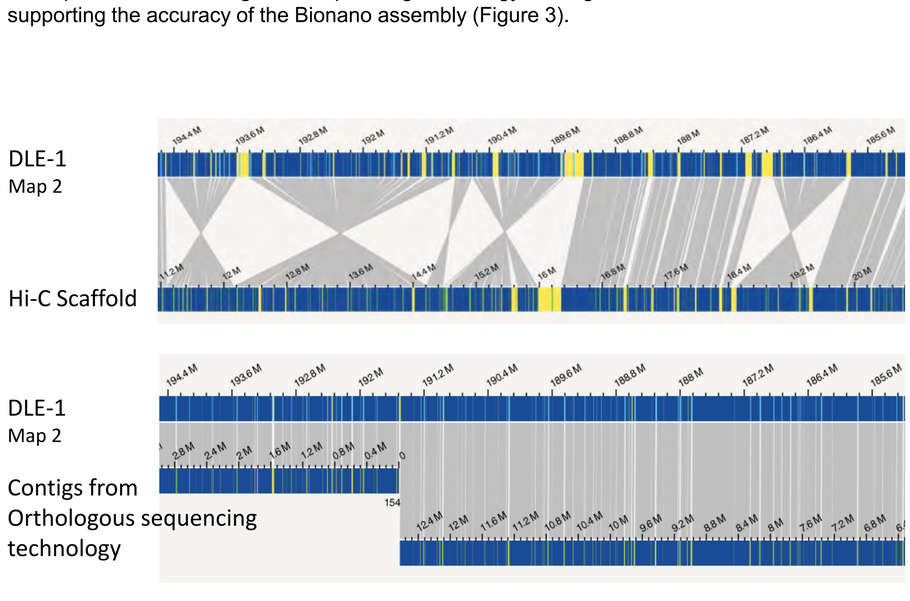


Figure 3. At discrepancy regions between Bionano and Hi-C scaffold, Bionano structure is supported by a third technology of lower contiguity.

Conclusions

Bionano is a rapid, cost effective and highly accurate method for chromosome level *de novo* assembly that can complement other genomic methods. Bionano's high-quality *de novo* assemblies have been a major part of most published genome finishing projects in recent years.

Reference

1) Pendleton, M., Sebra, R., et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nature Methods (2015); e3454
2) Cao, H., et al., Rapid detection of structural variation in a human genome using NanoChannel-based genome mapping technology. Gigascience (2014); 3(1):34
3) 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.