De Novo Assembly of Extremely Long Single-Molecule
Genome Maps Imaged in Irys NanoChannel Arrays

Abstract
Despite significant advances in shotgun sequencing technology, de novo genome assemblies using only short read data are generally incomplete due to the complexity found in large genomes. This complexity, consisting mainly of large duplications and repetitive regions, hinders sequence assembly and subsequent comparative analysis. We present a single-molecule genome analysis system based on a NanoChannel Array™ technology that resolves these sequence ambiguities. This technology provides high-throughput sequence motif maps of single molecule fragments hundreds of kilobases in size. Capitalizing on the information encoded in the extremely long single molecule maps, assembly algorithms unique to this format deliver high-resolution whole-genome sequence motif maps. Parallel assembly by shotgun sequencing and sequence motif mapping offers a multi-scale pathway for whole genome de novo assembly solving many of the ambiguities inherent in using short read assembly alone. Here we detail the BioNano genome assembly of Drosophila melanogaster, a eukaryotic model organism having a diploid genome estimated to be ~190 Mb haploid size. The genetic material was isolated from embryos. In this case, a high-quality genome draft exists, and the BioNano assembly is compared to the reference. Some structural variations are detected and previously unordered sequence contigs are anchored highlighting the utility of parallel multi-scale genome assembly for finishing applications. With this demonstration, de novo assembly of the human genome using BioNano technology is within reach.

Background
Generating high quality finished genomes replete with accurate identification of structural variation and high completion (mininal gaps) remains extremely challenging using short read sequencing technologies alone. Instead, Irys™ technology provides direct visualization of long DNA molecules in their native state, avoiding the statistical assumptions that are normally used to force sequence alignments of low uniqueness elements. The resulting order and orientation of sequence elements are demonstrated in anchoring NGS contigs and structural variation detection.

Methods
(1) DNA is labeled with IrysPrep™ reagents by incorporation of fluorophore labeled nucleotides at a specific sequence motif throughout the genome. (2) The labeled genomic DNA is then linearized in the IrysChip™ nanochannels and single molecules are imaged by Irys. (3) Single molecule data are collected and detected automatically. (4) Molecules are labeled with a unique signature pattern that is uniquely identifiable and useful in assembly into genome maps. (5) Maps may be used in a variety of downstream analysis using the IrysView™ software suite.

Conclusions
BioNano Genomics Irys enables direct visualization of single-molecule, extremely long DNA for the direct observation and measurement of genome complexities. This system permits accurate genome-wide assembly and detection of structural variants that typically confound short read genome assembly and comparative genomic analysis. Here we demonstrate the structural variation and genome assembly capabilities of the IrysChip nanochannel array and Irys imaging system to characterize a number of complex elements in a common model organism, Drosophila melanogaster. In addition to assembly of a high-coverage physical map, several examples detail the system’s ability to resolve heterochromatin, repetitive regions, and copy number variation.

References
3) Dai, S. et al. Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. Nucleic Acids Research (2010); 38: 8