Leveraging the Flexibility of Multi-color Imaging of Extremely Long Single-Molecules in NanoChannels for Epigenetic Profile Mapping, Centromere Probing by Hybridization and Sample Multiplexing

Authors: A. Hastie¹, DH. Zhang¹, P. Sheth¹, K. Pham¹, J. Reifenberger¹, X. Zhou¹, G. Pfevalcic¹, C. Escudé², S. Chan¹, E. Lam¹, H. Cao¹
¹BioNano Genomics, San Diego, CA, 92121, USA; ²Muséum National d’Histoire Naturelle, Paris, France

Abstract

Genome mapping by single molecule imaging of fluorescently labeled motif sites on uniformly elongated long DNA molecules in Nano Channel arrays has been used extensively in the last couple of years for improving de novo sequence assemblies through scaffolding sequence contigs and for detections of germ line and somatic structural variation (SV). Several improvements have been recently developed to enhance the throughput and richness of genome mapping data. Here we provide data on multiplexing samples for genome mapping, probing high molecular weight DNA with triplex forming oligonucleotides for mapping and assembly of centromeres as well as for mapping of CpG methylation status in conjunction with genome mapping. As throughput has improved for single molecule data collection in nanochannels, a single flow cell is now able to collect far more data than needed for smaller genomes resulting in a need for multiplexing. By coding samples with different fluorophore combinations, it is possible to mix several samples together in a single flow cell for data collection and demultiplex the samples bioinformatically for downstream analysis. Genome mapping is primarily done by creating single strand nicks using modified restriction enzymes. In order to target other regions of interest that may not have enough information by nick labeling, a hybridization approach was undertaken. This has been applied to the centromere of chromosome 17. From this, the pattern of centromere higher order repeats in single molecules are directly observed and measured. Finally, a new method for CpG methylation status will be presented. Single molecules are labeled at nicking motif sites and, using a second color, at a subset of CpG sites, the unmethylated sites are labeled. This method allows visualization of methylation status in phase with other genomic positions including SVs. These new powerful methods leverage the flexibility of the Irys® System for single molecule linearization and imaging for improved throughput and deeper biological investigation.

Methods

1. IrysPrep kit extraction of long DNA molecules
2. IrysPrep reagents label DNA at specific sequence motifs
3. IrysChip linearizes DNA in NanoChannels
4. Irys automates imaging of single molecules in NanoChannels
5. Molecules and labels detected in images by instrument software
6. Irysview software assembles genome maps

Spectral Multiplexing of Four Samples per Irys Flowcell

The strategy for spectral multiplexing: (A) Label four DNA samples with different mixtures of green and red fluorophores, mix the molecules together (B) Collect data on Irys (C) Different samples will have different ratios of red to green label density (D), which allows for demultiplexing and de novo assembly of each genome separately (E). X-Y scatter plot of four genomes labeled with multiplex label mixtures and run in one Irys flowcell (F). Demultiplexed DNA signals are color-coded for each pool. Mapping of demultiplexed molecules to original genomes shows purity of at least 94%.

Conclusions:
- It is now possible to efficiently multiplex and demultiplex several genomes in a single flow cell of an IrysChip using fluorophore mixture.
- TFO-based labeling provides a flexible method to label genomic DNA for targeted interrogation by Irys.
- Epigenetic status can now be measured on the Irys System and observed in conjunction with structural variation analysis.

Reference:
1) Cao, H., et al., Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. Gigascience (2014); 3(1):34
4) Bénédicte, G.L., et al. Sequence-specific fluorescent labeling of double-stranded DNA observed at the single molecule level. Nucleic Acids Research (2003); 31(20): e125