View this poster online at:
http://bionanogenomics.com/science/posters/

Next-Generation Mapping: A Highly Sensitive and Accurate Method for Interrogation of Clinically Relevant Structural Variation

A Hastie, A W Pang, J Lee, E T Lam, T Anantharaman, W Andrews, M Saghbini, H Cao
BioNano Genomics, San Diego, CA

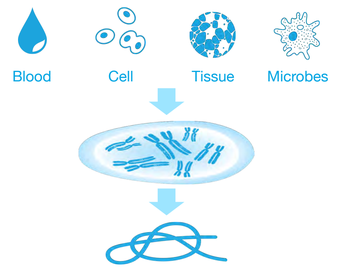
Abstract

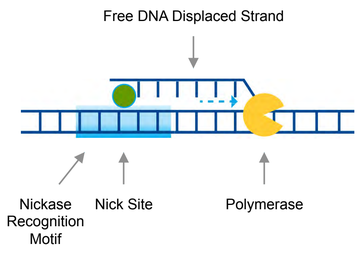
Structurally complex loci underlie many diseases. These loci can be very challenging to resolve by currently available methods such as karyotyping, Arrays, PCR-based tests and next-generation sequencing (NGS). Next-generation mapping (NGM) using the BioNano Genomics Irys® System offers a high-throughput, genome-wide method, based on direct visualization of extremely long genomic fragments, able to interrogate genome structural variations (SVs) in the range of one kilobase pairs to hundreds of kilobase pairs. The Irys System uses extremely long range information to span interspersed and even long tandem repeats making it suitable for elucidating the structure and copy number of complex regions of the human genome, such as complex pseudogene and paralogous gene families. Clinically relevant regions often contain genes with paralogs and other complex repetitive structures complicating the interpretation of data and diagnosis of disease. We have used two hydatidiform mole cell lines (homozygous human cell lines) to simulate a diploid genome and measure sensitivity to homozygous and heterozygous SV detection. We found 87% sensitivity to heterozygous SVs and 99% sensitivity to homozygous SVs (kbp and up). We also present several examples of SVs at complex genetic loci by NGM, such as those made up of tandem repeats, paralogous gene families, and loci flanked by segmental duplications. Examples of variants at tandem repeats are tRNAs, kringle IV, and D4Z4. An important variable length tandem repeat is D4Z4, which is associated with facioscapulohumeral muscular dystrophy (FSHD). FSHD is strongly associated with a low copy number (< 10 units), occurring in 95% of FSHD cases. Copy number of tandem repeats is extremely hard to measure accurately with available methods, but we show that NGM using the Irys System can accurately measure the copy number of D4Z4 as well as haplotype the repeat array and differentiate the gene from a paralog that occurs on another chromosome. A second class of complex SVs involves genes with paralogs such as amylase and UGT2B17, two genes whose copy number have been shown to be associated with human health.

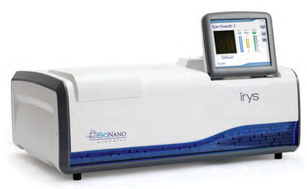
Here we demonstrate that NGM using the Irys System is proving to be a highly accurate and sensitive method for detection of clinically relevant SVs.

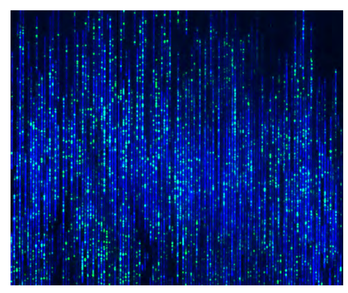
Methods

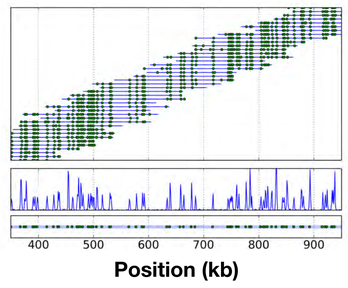
1. IrysPrep kit extraction of long DNA molecules
2. IrysPrep reagents label DNA at specific sequence motifs
3. IrysChip linearizes DNA in NanoChannel arrays
4. Irys automates imaging of single molecules in NanoChannel arrays
5. Molecules and labels detected in images by instrument software
6. IrysView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











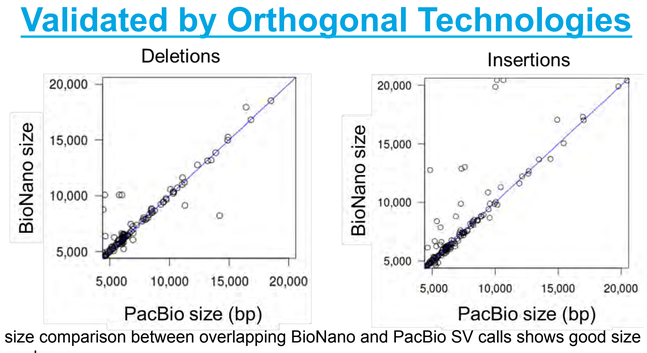
(1) Long molecules of DNA are labeled with IrysPrep® reagents by (2) incorporation of fluorophore labeled nucleotides at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the IrysChip® using NanoChannel arrays and single molecules are imaged by Irys. (4) Single molecule data are collected and detected automatically. (5) Molecules are labeled with a unique signature pattern that is uniquely identifiable and useful in assembly into genome maps. (6) Maps may be used in a variety of downstream analysis using IrysView® software.

Next Generation Mapping Detects Structural Variation With High Sensitivity and Specificity

	CHM1 and CHM13 assemblies	Mixture assembly	Sensitivity	PPV
Homozygous Insertions	880	876	99.5%	94.2%
Heterozygous Insertions	798	680	85.2%	
Homozygous Deletions	229	227	99.1%	95.5%
Heterozygous Deletions	574	467	81.4%	

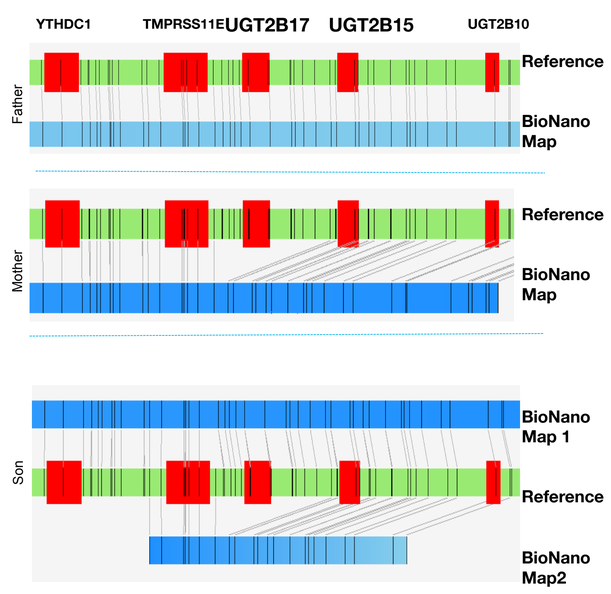
Two Homozygous cell lines, CHM1 and CHM13 were independently *de novo* assembled and SVs called. Raw data was mixed together, assembled and SVs called (mixture column). In the simulated diploid assembly, CHM1 only and CHM13 only SVs are considered heterozygous and those detected in both are homozygous SVs.

Next Generation Mapping Can be Cross Validated by Orthogonal Technologies



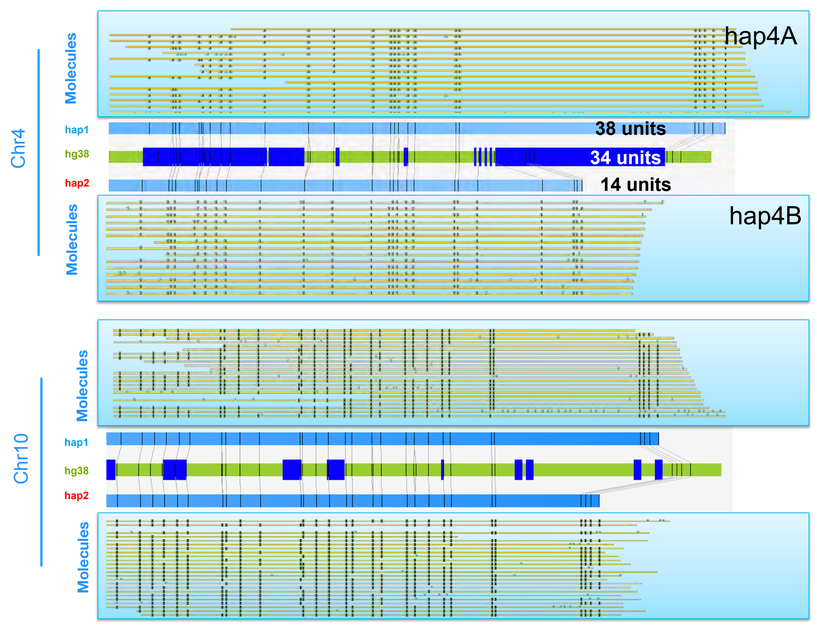
SV size comparison between overlapping BioNano and PacBio SV calls shows good size concordance.

Gene Deletion in Mother and Son Genomes from a PGP Trio of Ashkenazi Decent



A 117 kbp deletion shows the missing UDP glucuronosyltransferase 2 family, polypeptide B17 gene (*UGT2B17*). Deletion of *UGT2B17* has been reported to result in better quality of osteopathic health as well as higher testosterone and estradiol levels. *UGT2B17* is believed to produce an important antigen involved in graft versus host disease (McCarroll).

D4Z4 Array Length Measurement Haplotype and Paralog Resolution



D4Z4 repeat arrays occur at the subtelomeric region of chromosome 4q and 10q. 4q has two haplotypes, 4qA and 4qB, if the array length of 4qA falls below ~5 copies, Facioscapulohumeral (FSHD) muscular dystrophy can occur. In order to diagnose a pathogenic repeat array, the array for each allele or paralog must be measured and differentiated. The figure shows the differentiation and measurement of each allele on chromosome 4 and paralogs on chromosome 10 (this is a non pathologic measurement) in genome maps (blue bars) and single molecules shown below.

Conclusions

De novo detection of insertions, inversions and translocations is inefficient and inconsistent using commercially available technologies. Next generation mapping is efficient, cost effective, sensitive and specific for structural variation detection. Most importantly, next-generation mapping can be used for the detection of clinically relevant genomic features otherwise difficult to detect, including insertions/ CNVs, SVs involving paralogous regions and segmental duplications (e.g. UGT2B17),tandem repeat arrays (e.g. D4Z4), and balanced events (translocation and inversions).

References

1) Lam, E.T., *et* al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nature Biotechnology (2012); 10: 2303
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) Chaisson et.al. (2015) Resolving the complexity of the human genome using single molecule sequencing. Nature. 517, 608-611
4). McCarroll, S.A., *et* al., Donor-recipient mismatch for common gene deletion polymorphisms in graft-versus-host disease. Nat Genet. (2009) 2009 Dec;41(12):1341-4.
5) Menard, V., *et* al., Copy-number variations (CNVs) of the human sex steroid metabolizing genes UGT2B17 and UGT2B28 and their associations with a UGT2B15 functional polymorphism. Hum. Mutat. (2009) 30: 1310-1319.