NanoChannel Array Based Genome Mapping Reveals "Dark Matter' in the Genome: Enabling Comprehensive Genome Wide Large Structural Variation Analysis in Population Studies

H Cao, A Hastie, A W Pang, E T Lam, W Andres, T Anantharaman, T Liang, J Lee, K Pham, S Chan, M Saghbini, X Zhou, J Wang, ZY Zhu, M Austin, M Borodkin

BioNano Genomics, San Diego, CA

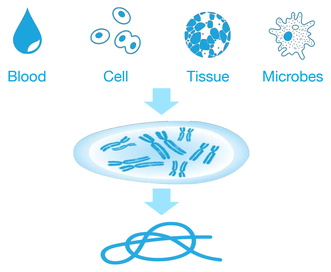
Abstract

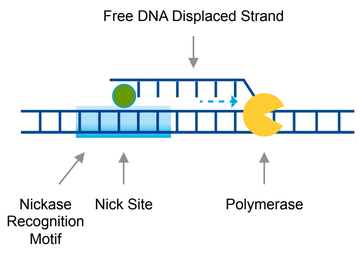
Structural variations (SV) have been well established to be associated with complex traits and diseases; Large SV (> 1 kbp) analysis of human genomes has been limited to date by technical shortcomings, i.e. karyotyping, FISH, microarrays and NGS. Karyotyping and FISH present low resolution, microarrays are limited to imbalanced CNVs and show a narrow dynamic range, low resolution and short read spans. Next – Generation Mapping (NGM) using BioNano Genomics Irys® System allows to comprehensively analyze whole genomes for SVs > 1 kbp, including balanced events, in a cost-effective and high-throughput manner. NGM allows for the *de novo* discovery and detection of large SVs in patient cohorts and population studies, which is needed to potentially uncover genomic structural causes of Mendelian and complex diseases such as cancer and neurobehavioral disorders.

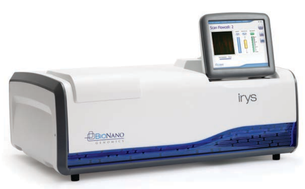
Here we demonstrate the robustness of NGM for genome-wide discovery of SVs in the CEPH trio from the 1000 Genomes Project with a 96% Mendelian concordance rate. We uncovered 100s of insertions, deletions, and inversions greater than 5 kbp, 7 times more than the large SV events previously detected by NGS. A large portion of those is novel, and some are located in the regions likely leading to disruption of gene function or regulation. We have also analyzed a trio of Ashkenazi Jewish descent from the NIST GIAB project, where we have found hundreds of inversions, insertions, and deletions, including large deletions in a clinically relevant UGT2B17 gene locus in the mother and son (See also Poster #3156F: "Next-Generation Mapping, a Highly Sensitive and Accurate Method for Interrogation of Clinically Relevant Structural Variation" by Hastie et al) as well as SV difference in different ethnic population. We have also demonstrated the direct detection of complex large segmental duplication with different location and orientation, as well as translocation events in clinically validated leukemia and multiple myeloma samples.

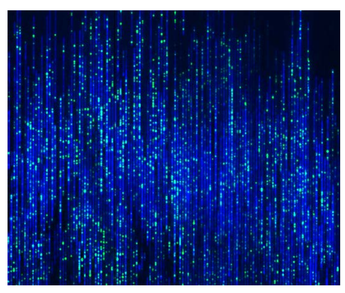
Methods

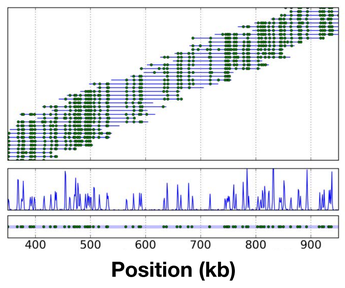
1. IrysPrep kit extraction of long DNA molecules
2. IrysPrep reagents label DNA at specific sequence motifs
3. IrysChip linearizes DNA in NanoChannel arrays
4. Irys automates imaging of single molecules in NanoChannel arrays
5. Molecules and labels detected in images by instrument software
6. IrysView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





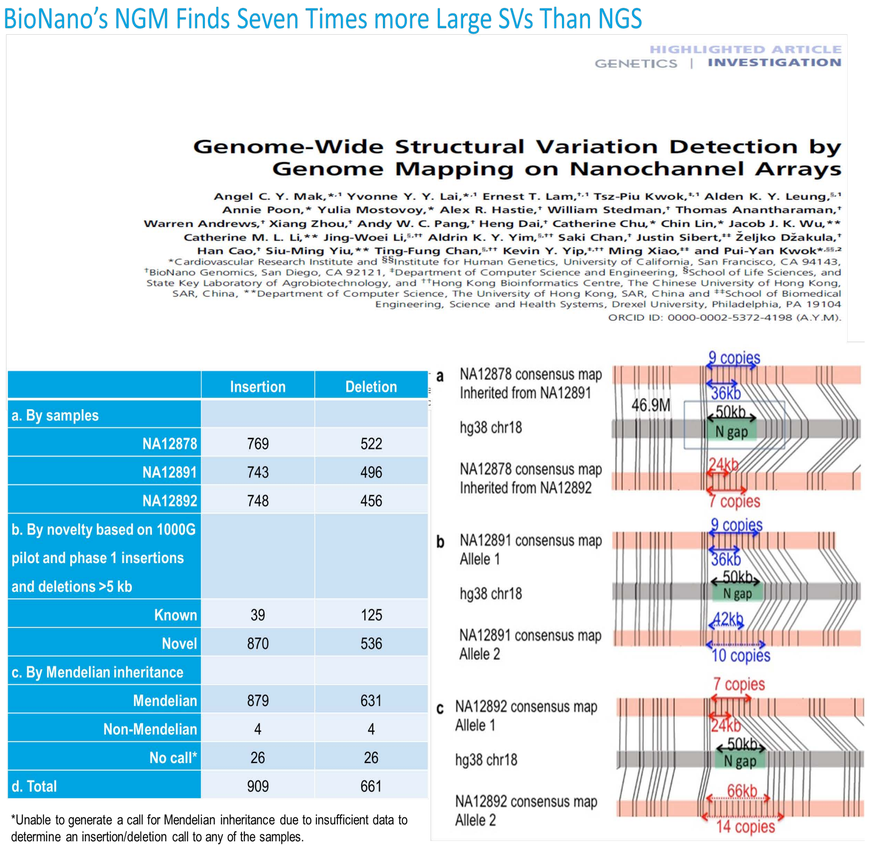






(1) Long molecules of DNA are labeled with IrysPrep® reagents by (2) incorporation of fluorophore labeled nucleotides at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the IrysChip® using NanoChannel arrays and single molecules are imaged by Irys. (4) 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.

BioNano's NGM Finds Seven Times more Large SVs Than NGS



HIGHLIGHTED ARTICLE
GENETICS | INVESTIGATION

Genome-Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays

Angel C. Y. Mak,*,† Yvonne Y. Y. Lai,*,† Ernest T. Lam,*,† Tsz-Piu Kwok,*,† Alden K. Y. Leung,‡,§,† Annie Poon,* Yulia Mostovoy,* Alex R. Hastie,† William Stedman,† Thomas Anantharaman,† Warren Andrews,† Xiang Zhou,† Andy W. C. Pang,† Heng Dai,† Catherine Chu,* Chin Lin,* Jacob J. K. Wu,** Catherine M. L. Li,** Jing-Woei Li,**,†† Aldrin K. Y. Yim,**,†† Saki Chan,† Justin Sibert,†† Željko Džakula,† Han Cao,† Siu-Ming Yiu,** Ting-Fung Chan,**,†† Kevin Y. Yip,**,†† Ming Xiao,** and Pui-Yan Kwok*,‡‡,2

*Cardiovascular Research Institute and §§Institute for Human Genetics, University of California, San Francisco, CA 94143, †BioNano Genomics, San Diego, CA 92121, ‡Department of Computer Science and Engineering, §School of Life Sciences, and State Key Laboratory of Agrobiotechnology, and ††Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, SAR, China, **Department of Computer Science, The University of Hong Kong, SAR, China and ‡‡School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA 19104

ORCID ID: 0000-0002-5372-4198 (A.Y.M).

	Insertion	Deletion
a. By samples		
NA12878	769	522
NA12891	743	496
NA12892	748	456
b. By novelty based on 1000G pilot and phase 1 insertions and deletions >5 kb		
Known	39	125
Novel	870	536
c. By Mendelian inheritance		
Mendelian	879	631
Non-Mendelian	4	4
No call*	26	26
d. Total	909	661

*Unable to generate a call for Mendelian inheritance due to insufficient data to determine an insertion/deletion call to any of the samples.

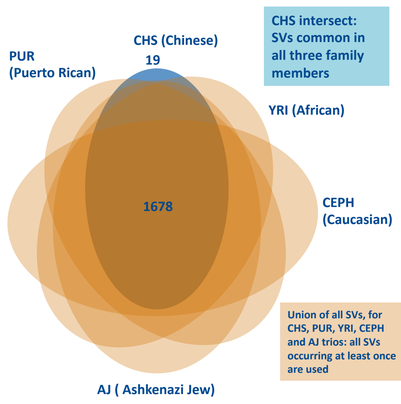
Genomic SV Profiling in Sample Cohorts and Population

SV (>2 kb) positions in 22 Euploid individuals. SVs are plotted across the genome. Of a total of 24,360 SVs, 2655 SVs are only identified in a single individual, while 5537 SVs were common to at least 20 of 22 of the genomes (257 loci).

Unique SVs in Trios of Different Family and Ethnicities – "The Clan Signature"

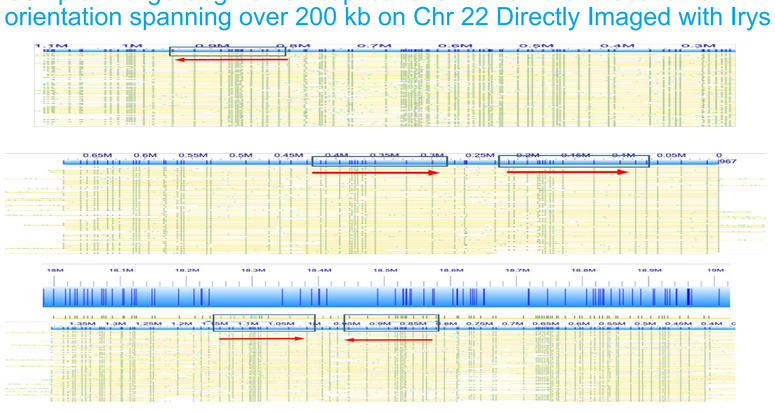
	Total SVs# of All Three	SVs Shared by all Three	SVs Unique to Family*
Caucasian^	7685	1159	5
Ashkenazi	9558	1673	16
Puerto Rican	9213	1582	16
Han Chinese	9360	1678	19
Nigerian	10285	1607	71

Variation called against GRC38 Reference
^Caucasian was not assembled with haplotype assembler
*Total SVs shared by all three, not present in any other individual

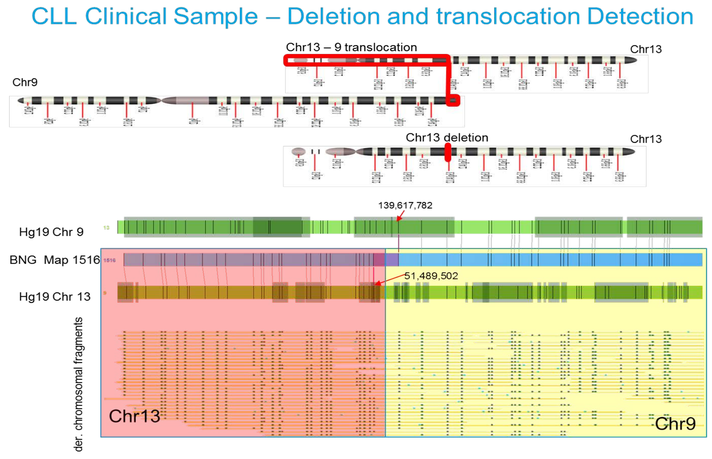


In comparison, 17+ different algorithm and methods were used for 1000 Genomes Project

Complex Large Segmental Duplications with different location and orientation spanning over 200 kb on Chr 22 Directly Imaged with Irys



CLL Clinical Sample – Deletion and translocation Detection



Conclusions

Bionano's Irys System Next-Generation Mapping is a very powerful and versatile genome analysis tool providing native long range information of complex genomic architecture at single molecule level . It can be used for stand alone long range SV discovery and detection in human genomes (1Kb to >1Mbp), as well as for *de novo* whole genome assembly and hybrid scaffolding. Irys helps orient and align fragmented sequencing contigs, close assembly gaps and extend into regions hard to sequence.

We have shown that NGM is the robust and effective method for the detection of biologically relevant and complex SVs detection in human genomes. In the era of precision medicine, information on whole genome SVs within disease population cohorts is critical. These SVs are important in addition to conventional SNP analysis, to study the effects of a full spectrum of genomics variations in complex traits and human disease.

Reference

1) Mostovoy J et al. A hybrid approach for de novo human genome sequence assembly and phasing Nature Methods (2016)
2) Pendleton, M., Sebra, R., et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nature Methods (2015); e3454
3) Mak AC et al., Genome-Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays Genetics (2016)
4) Zook, J., et al. Extensive sequencing of seven human genomes to characterize benchmark reference materials. Scientific Report (2016)
5) Cao, H., et al. Rapid Detection of Structural Variation in a Human Genome using Nanochannel-based Genome Mapping Technology. Giga Science (2014); 3(1):34
6) Lam, E.T., et al. Genome mapping on Nanochannel arrays for structural variation analysis and sequence assembly. Nature Biotechnology (2012); 30(8):7713) Lam, E.T., et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nature Biotechnology (2012); 10: 2303

©2015 BioNano Genomics. All rights reserved.