Toward Rapid Identification of Coding Fusions and Structural Rearrangements in Cancer Genomes: Multiple Myeloma First

M Reardon[1], H Dai[1], M Saghibini[1], X Yang[1], A Pang[1], H VanSteenhouse[1], T Dickinson[1], A Nooka[2], JL Kaufman[2], SM Matulis[2], L Zhang[2], DF Saxe[2], LH Boise[2], KP Mann[2], DL Jaye[2], S Lonial[2], H Cao[1], MR Rossi[2]

[1]BioNano Genomics, San Diego, California, USA, [2]Emory University, Atlanta, GA, USA

Abstract

Karyotype and FISH have been standard diagnostic tools in monitoring response to treatment and disease progression in hematological disorders, including multiple myeloma, for over 40 years. However, with the availability of array and NGS technologies in most clinical diagnostic laboratory settings, it is time to consider evaluating the use of more modern methods in diagnosing plasma cell dyscrasias. Whereas RNA-Seq, SNP-CN arrays and deep sequencing cancer panels are routinely run as a cost effective workflow for clinical trials, the validation of translocations and structural rearrangement by PCR and/or FISH is generally cumbersome. BioNano Genomics Irys System utilizes nanochannel technology and high resolution imaging of high molecular weight DNA for whole genome mapping of translocations and copy number abnormalities. As proof of concept, the Irys system was used to analyze the multiple myeloma cell line, KMS11, and will demonstrate the utility, speed and accuracy in detecting structural rearrangements in this line compared to whole genome sequencing, RNA-seq and SNP-CN datasets. In addition to validating hallmark genetic aberrations of multiple myeloma in this proof of concept study, the use of the system with unknown CD138+ enriched bone marrow samples from five consented patients currently on clinical trials is presently being analyzed and the data generated will be compared to FISH, G-band karyotype and SNP-CN arrays. These proof of concept data demonstrate that the Irys system fills a much needed role between cytogenetics and NGS/arrays and has the potential to be a disruptive innovation with broad applicability to genome research and refinement of normal variation and disease.

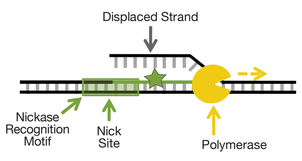
Background

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 Irys platform 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 whole genome. The resulting order and orientation of sequence elements in the map can be used for anchoring NGS contigs and structural variation detection.

Methods

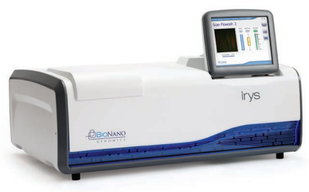
(1) DNA is labeled with IrysPrep™ reagents by incorporation of fluorophore-labeled nucleotides at a specific sequence motif throughout the genome. (2) The labeled genomic DNA is then linearized in the IrysChip™ nanochannels and single molecules are imaged by Irys. (3) Irys performs automated data collection and image processing. (4) Molecules are labeled with a unique signature pattern that is uniquely identifiable. (5) Molecules are assembled into genome maps and downstream analysis of maps is performed with the IrysView™ software suite.

1) IrysPrep reagents label DNA at specific sequence motifs



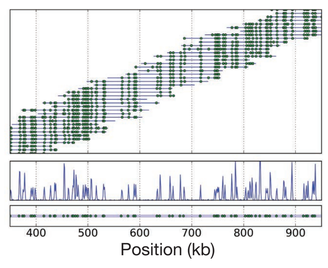
2) IrysChip linearizes DNA in NanoChannels

3) Irys automates imaging of single molecules in NanoChannels

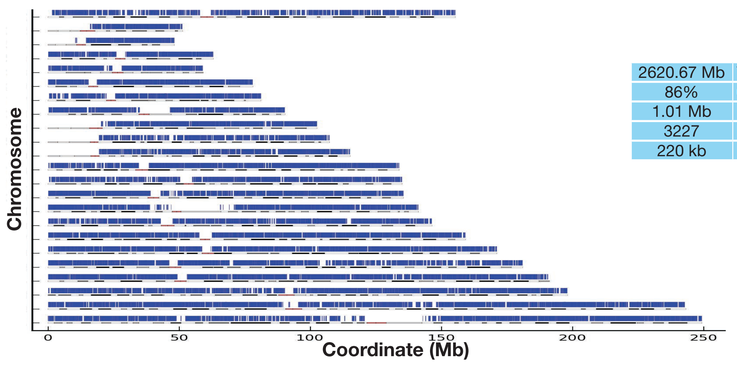


4) Molecules and labels detected in images by instrument software

5) IrysView software assembles genome maps



De Novo Genome Map Assembly

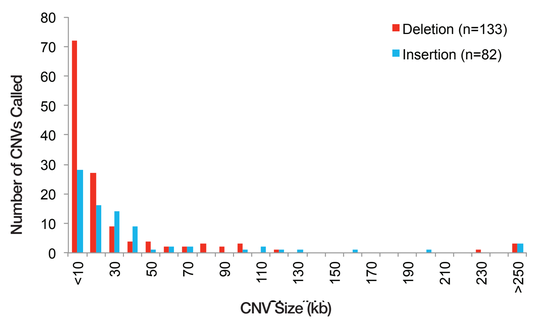


2620.67 Mb	Total Assembled Map Length
86%	Overlap with hg19
1.01 Mb	Map N50
3227	N Contigs
220 kb	Molecule N50

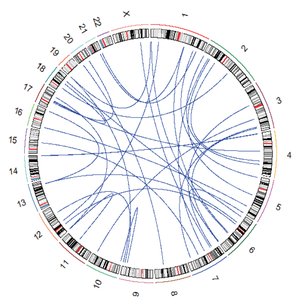
This is the genome coverage of a full *de novo* assembly for the multiple myeloma cell line, KMS11, using all molecules > 150 kb (at ~50x depth). White regions indicate N-base gaps in the reference (hg19) that cannot be aligned to.

Structural Variation and Translocation Discovery

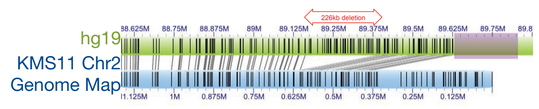
Genome-Wide CNVs

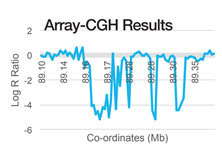


Widespread Translocations



2p11.2 Deletion with Irys and a-CGH



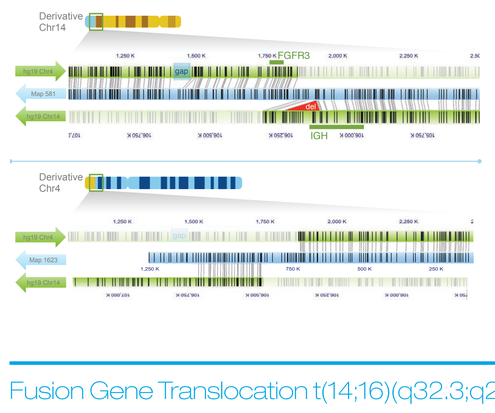


Following *de novo* assembly of genome maps, structural variation across a broad range of sizes refractory to many high throughput and short-read technologies was detected. Insertions are called by the presence of novel label sites and expansion of adjacent labels. Deletions are evident by the absence of label sites or narrowing of inter-label segments. CNVs are evident with positional and contextual information visualized in a genome browser, for interpretation that may be more clear than noisy array data.

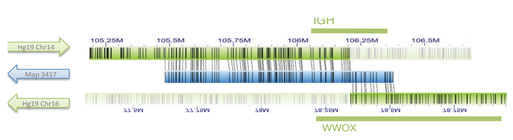
Translocations can manifest as aberrant expression of existing genes, nascent expression of resulting new fusion genes, and potential oncogenic activity, yet are difficult to detect with current methods. Irys long molecules retain positional information essential to discover translocations across the genome, as seen in the circos plot of the KMS11 multiple myeloma cell line.

Translocations: Reciprocal and Unbalanced

Balanced Translocation t(4;14)



Fusion Gene Translocation t(14;16)(q32.3;q23)



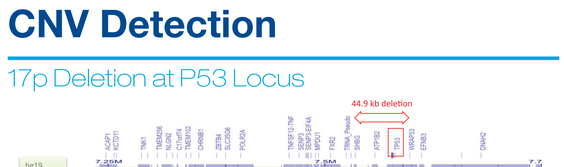
Archetypal variants are precisely detected by Irys genome map analysis.

The t(4;14)(p16.3;q32.3) reciprocal translocation results in FGFR3/IGH fusion and MMSET dysregulation. FGFR3 can be dysregulated in myeloma by t(4;14) into the vicinity of IgH enhancers, now separated to influence gene expression on either side of the 4p16 genomic breakpoints. The KMS11 cell line is positive for t(4;14) and expresses variants of FGFR3 that are constitutively activated to some extent. We also see a deletion in the IGH locus. This could be incident to t(4;14) (commonly seen) or simply an SV in the IGH, which is a variable region in humans. In t(4;14) MMSET is located on the reciprocal chromosome placing it in close proximity to and presumably upregulated by the 5' IgH enhancer. Both FGFR3 and MMSET gene targets have potential oncogenic activity.

t(14;16)(q32.3;q23) is another translocation involving the IgH locus. The breakpoints identified on 16p23 occur over a region centromeric to c-maf, a proto-oncogene dysregulated in t(14;16), within an intron of an oxidoreductase gene, WWOX/FOR. This region is a common fragile site and t(14;16) inactivates one allele of the WWOX/FOR gene which has been implicated as a tumor suppressor gene in several solid tumors.

CNV Detection

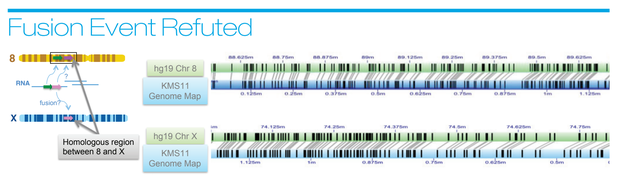
17p Deletion at P53 Locus



In addition to the translocations above, Irys genome map analysis easily detected the 17p deletion which includes the TP53 locus. This gene codes for the tumor suppressor p53, known as "the guardian of the genome" because of its role in conserving stability by preventing genome mutation. Here a deletion removes one copy of P53. More than 50 percent of human tumors contain a mutation or deletion of the TP53 gene.

Long Molecules Are More Unique Than RNA-Seq Reads

Fusion Event Refuted



Long molecules collected on Irys contain more uniqueness than short RNA-Seq reads, and are less succeptable to false calls caused by homologous sequence. In this case, a similar region on chr X caused a fusion gene to be called with sequencing, whereas there was no evidence of translocation in Irys genome maps.

Conclusions

The Irys System, which utilizes nanochannel technology and high resolution imaging of high molecular weight DNA for whole genome mapping of translocations and copy number variations, can quickly and accurately detect three hallmark genetic aberrations in the multiple myeloma cell line KMS11 as well as refute a fusion gene called with sequencing. In addition to these archetypal variants and the false call correction, we are currently analyzing the broad range of structural variants called by Irys which include large scale insertions, deletions and complex rearrangements that are missed in many high throughput and short-read technologies. We are presently running unknown CD138+ enriched bone marrow samples from five consented patients currently on clinical trials in order to demonstrate the use of the system when compared to FISH, G-band karyotype and SNP-CN arrays. These proof of concept data demonstrate that the Irys system can reveal order and orientation of functionally relevant components of complex genomes and therefore fills a gap between cytogenetics and NGS/arrays. This technology displays the potential to be a disruptive innovation with broad applicability to genome research and refinement of normal variation and disease.

References

1) 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.
2) Lam, E.T., et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nature Biotechnology (2012); 10: 2303
3) Das, S. K., et al. Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. Nucleic Acids Research (2010); 38: 8
4) Xiao, M., et. al. Rapid DNA mapping by fluorescent single molecule detection. Nucleic Acids Research (2007); 35:e16.
5) Zhu, L., et. al. Fibroblast growth factor receptor 3 inhibition by short hairpin RNAs leads to apoptosis in multiple myeloma. Molecular Cancer Therapeutics (2005); 4: 787.