Detection of Complex Rearrangements in Cancer Genomes: A Study on Multiple Myeloma

AWC Pang, H Dai, M Saghbini, EH Cho, A Hastie, X Yang, T Dickinson, A Nooka[1], JL Kaufman[1], SM Matulis[1], L Zhang[1], DF Saxe[1], LH Boise[1], KP Mann[1], DL Jaye[1], S Lonial[1], H Cao, MR Rossi[1]
BioNano Genomics, San Diego, CA, USA, [1]Emory University, Atlanta, GA, USA

Abstract

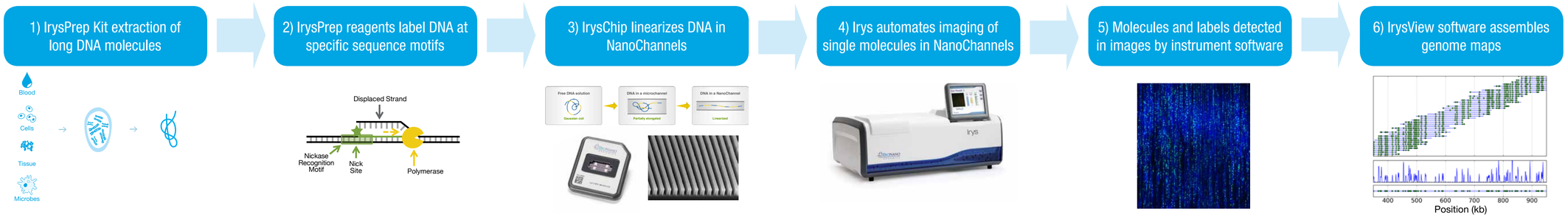
Large DNA rearrangements are known to be associated with the initiation and progression of multiple myeloma. While karyotype and FISH are routinely used for diagnosis and prognosis of this disease, microarrays, RNA- and DNA sequencing are also used for variant discovery. In our experience, short-read sequencing has limited ability to detect translocation events reliably, and the validations of those calls by FISH or PCR is manually intensive. Here we present the BioNano Genomics Irys System that utilizes nanochannel technology to linearize long DNA molecules of hundreds of kilobases. It uses high resolution imaging for whole genome mapping and de novo assembly. Leveraging on ultra-long assembled genome maps, the platform is able to detect large DNA rearrangements such as translocations, amplifications and deletions. As a proof of concept, we ran the cell line KMS11 on the Irys System, and we detected known translocation events, such as t(4;14), as well as a gene-fusion event in this sample. In addition, we developed a pipeline that would use our long-length molecules to validate and potentially phase complex rearrangements detected by sequencing. By examining molecule depth profile, we identified multiple gross genomic abnormalities in copy number. We further ran additional multiple myeloma clinical samples on the Irys platform, and identified numerous rearrangements. Based on comparisons with variants found by SNP arrays, whole genome and whole transcriptome sequencing, we are confident that the BioNano Genomics Irys System will enable genome assembly finishing and rearrangement discoveries, expand our view of genome architecture, and improve our understanding of the molecular mechanisms which drive hematological malignancies and solid tumors.

Background

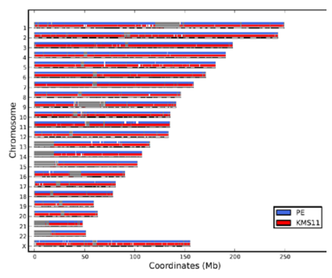
Generating high quality finished genomes replete with accurate identification of structural variation and high completion (minimal gaps) remains challenging using short read sequencing technologies alone. Instead, Irys technology provides direct visualization of long DNA molecules in their native state, 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) Long molecules of DNA is 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™ nanochannels 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.



De Novo Assemblies of Cancer Genomes

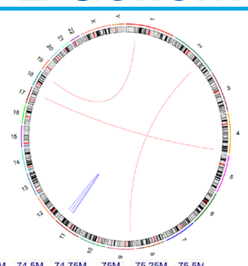


The left panel shows the genome coverage by the *de novo* assemblies of the multiple myeloma cell line KMS11 and a multiple myeloma patient sample, called PE. Both assemblies were constructed using molecules > 150 kb. Grey loci represent N-base gaps in the hg19 reference.

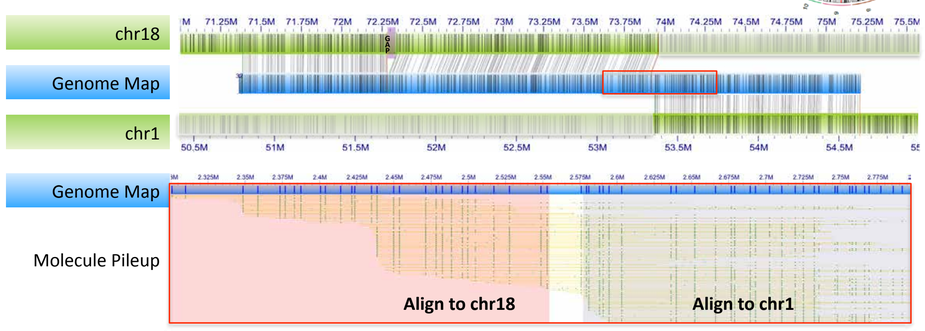
	KMS11	PE
Total Assembled Contig Length	2.62 Gb	2.80 Gb
Contig N50	1.01 Mb	1.043 Mb
% hg19 Overlapping BNG Assemblies	86%	88.3%

Translocation Profiles of the PE Genome

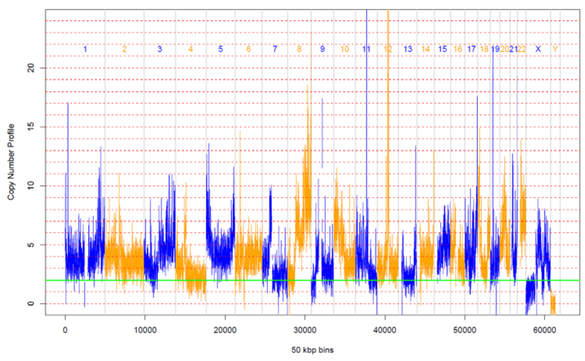
Translocations were detected in the PE genome. By aligning our long-range genome maps to the reference hg19, we were able to detect high-confidence intra-chromosomal (blue) and inter-chromosomal (red) translocations. One example is translocation t(1;18), which was detected by a 3.85Mb genome map. Note that there are multiple long molecules that span across the translocation junction.



Translocation t(1;18) Detected



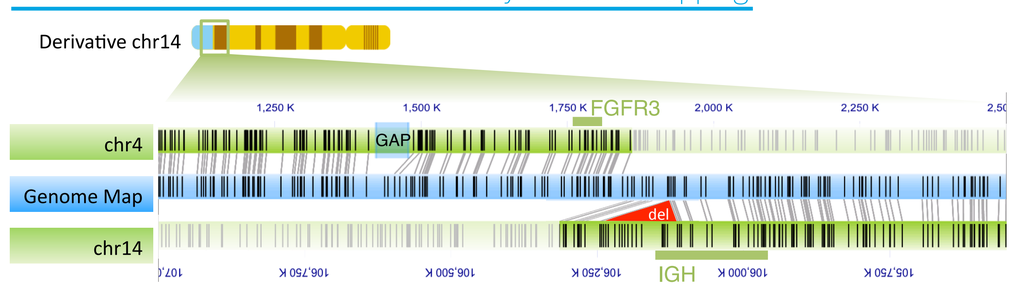
Detecting Copy Number Aberrations



Chromosomal aberrations can also be detected by significant deviations in the number of molecule alignments compared to diploid regions. Here, we show a coverage profile of a breast cancer sample, and we see that there are significant amplifications and deletions throughout the genome. The green line represents the expected diploid level. This result shows another utility of this technology for cancer research.

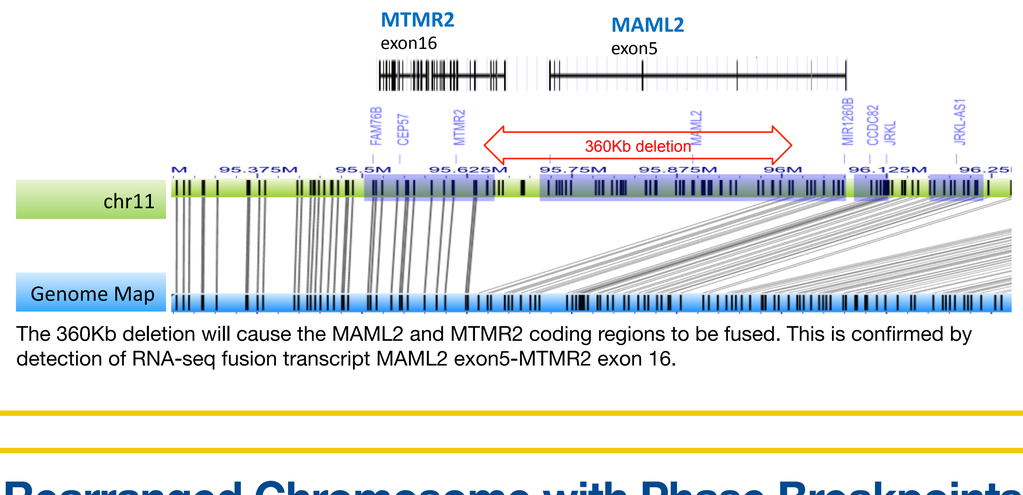
Structural Variation in KMS11

Known KMS11 Translocation Detected by Genome Mapping



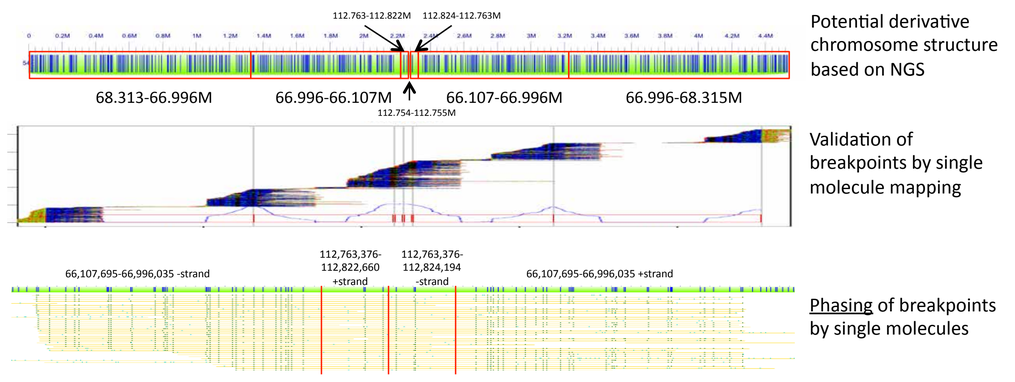
Our data show that KMS11 is positive for t(4;14), a known variant in this cell line. The FGFR3 can be dysregulated in myeloma by the translocation, as it brings the gene into the vicinity of IGH enhancers. The deletion in the IGH locus could be incidental to KMS11 or simply a common variant in the human population.

Deletion at 11q21 (95.65-96M) Caused MAML2-MTMR2 Fusion Transcript



The 360Kb deletion will cause the MAML2 and MTMR2 coding regions to be fused. This is confirmed by detection of RNA-seq fusion transcript MAML2 exon5-MTMR2 exon 16.

Rearranged Chromosome with Phase Breakpoints



Potential derivative chromosome structure based on NGS

Validation of breakpoints by single molecule mapping

Phasing of breakpoints by single molecules

Single molecules alignment to each of four fragments of chr12, proving the structure of the derivative chromosome in this breast cancer sample.

Conclusions

The Irys System utilizes nanochannel technology to image high molecular weight DNA for genome mapping of translocations, amplifications and deletions, and it accurately detected hallmark genetic mutations in multiple myeloma samples. We also demonstrate the usage of long molecules to validate and phase breakpoints identified by sequencing. In addition, we are currently analyzing the broad range of structural variation called by Irys that are missed by short-read technologies and SNP arrays. These proof of concept data demonstrate that that the Irys platform can reveal relevant mutations in complex genomes, thus filling in the gap between cytogenetics and NGS/microarrays, and has broad applicability in genome research.

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.