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

Irys Next-Generation Mapping to Identify Medically-Relevant Genomic Structural Variation

AWC Pang, J Lee[1], S Chan, W Andrews, T Anantharaman, A Hastie, Han Cao

BioNano Genomics, San Diego, California, United States of America

Abstract

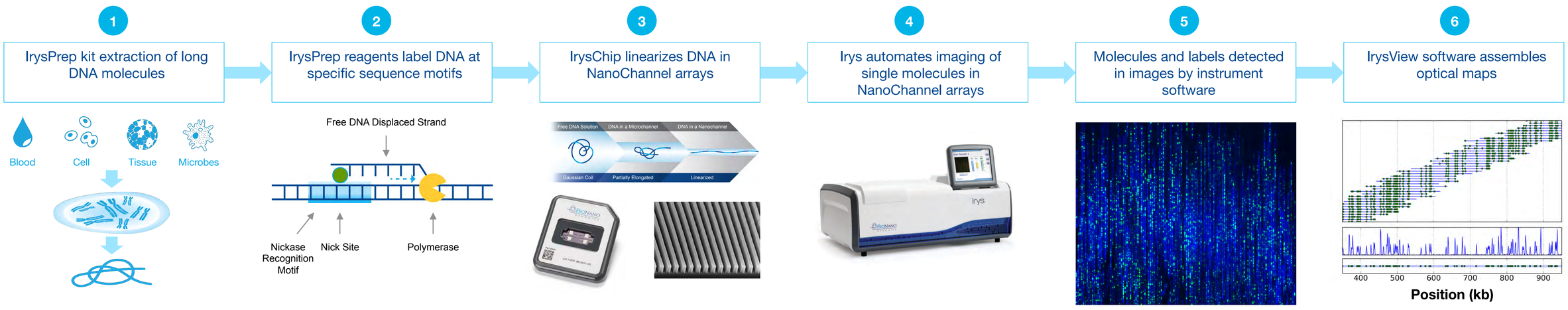
Structural variation (SV) detection is fundamental to understanding the genetic basis of diseases. While karyotype, cytogenetic, and conventional molecular detection approaches are robust, they can be manually intensive, biased towards targeted loci, and cannot elucidate copy number of long repeats. The BioNano Genomics Irys® System offers an unbiased and sensitive method to detect large variants, named next-generation mapping (NGM). It uses NanoChannel arrays to linearize DNA hundreds of kilobases in size, high resolution imaging, and haplotype-aware *de novo* assembly to construct optical maps megabases in length. These maps generally capture over 4,000 insertions, deletions (>1 kbp), translocations and inversions in a human sample.

Here we present a novel software that examines SV calls made by IrysSolve to identify disease-associated SVs. To ascertain high-confidence calls, it searches for genomic elements such as centromeres and segmental duplications that may confound SV-calling, and evaluates quality scores of the assembly, alignments, and SVs detected. To further ascertain medically-relevant candidates, it filters the remaining calls with over 536,000 SVs collected from 81 mapping experiments from phenotypically "normal" individuals, and compares with calls from disease-specific databases. Using this pipeline, whose total runtime is only a few hours, we can efficiently focus on several dozens of significant candidates for further analysis. We ran 10 samples with hematological malignancy on the Irys system, applied the candidate-finding software, and identified known rearrangements such as the t(9;13) in leukemia/ lymphoma and t(9;22) in CML. Moreover, we uncovered novel mutations ranging in size from a small 4.2 kbp insertion disrupting an ALL gene *CNOT3* to a 35.4 Mbp chr13 deletion. In conclusion, Irys technology may replace conventional detection approaches with one standardized platform, aid the discovery of functional SVs, and improve our understanding of the mechanisms of diseases.

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 optical 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 analysis.

Methods



(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.

	Component of the variant annotation pipeline
Genomic features	Overlap N-gaps
	Span centromere
	Overlap segmental duplication
Integrity of assembly	Flag variant breakpoints with low assembly scores
	Flag variant breakpoints with low molecule coverage
Integrity of the variant	Identify sub-optimal alignments
	Flag low variant with confidence score
Candidate search	Compare against calls found in BNG control sample db*
	Compare against public control db (DGV)
	Compare against cancer databases

Table 1. The components of variant annotation pipeline. The pipeline is run to remove potential false positive and common variants found in control samples, and to identify relevant candidates.

Structural Variation	Count	Size (Mbp)
Insertion	314,979	719.1
Deletion	219,274	1,109.6
Inversion breakpoint	1,371	-
Total	535,624	1,828.7

Table 2. Catalog of structural variation collected from 81 phenotypically "normal" human samples. SV calls from disease samples were compared against this database to remove common variants.

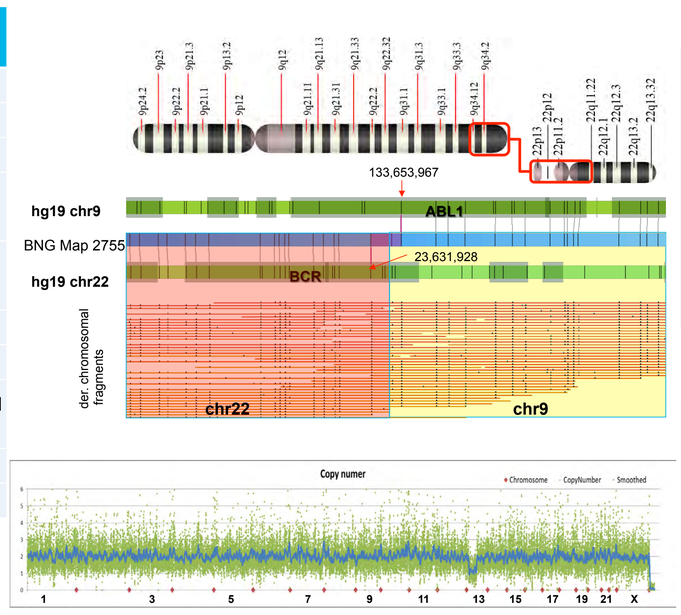


Figure 1. Examples of large chromosomal aberrations detected by mapping.
(Top) A Philadelphia translocation t(9;22) was detected in leukemia cancer cells. The map was aligned to the public reference assembly hg19, and the resulting alignments show a conjoined junction between chr9 and chr22, creating a fusion gene called *BCR-ABL1*.
(Bot) Whole genome depth of coverage shows chr13 monosomy. In addition, to comparing assembled optical maps, we also aligned ~100x of molecules from a cancer sample to hg19, to detect copy number alterations. The green dots represent raw copy number count, while the blue dots represents smoothed copy number counts after removal of local noise.

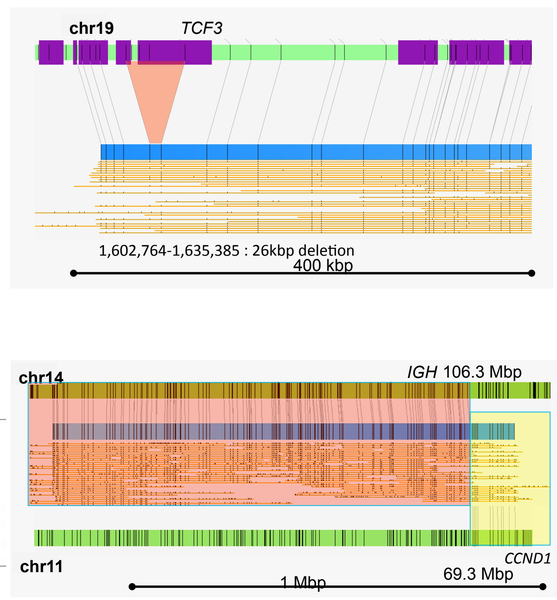


Figure 2. Examples of other deleterious events captured.
(Top) A 26 kbp deletion impacting the *TCF3* gene. Mutation at the gene may be involved in lymphoid malignancies. The gene is known to be involved in other translocation events such as B-cell ALL (t(1;19), with *PBX1*), childhood leukemia (t(19;19), with *TFPT*) and acute leukemia (t(12;19), with *ZNF384*).
(Bot) A t(11;14) translocation detected. The *IGH-CCND1* conjoined gene is formed from this translocation that is known to be associated with diseases involving B-lineage lymphocytes.

Sample	Insertion	Deletion	Inversion	Inter-chrom Translocation†	Intra-chrom Translocation†
S1	3380	1504	20	16	6
S2	3813	1605	34	13	11
S3	3034	1519	19	11	2
S4	3723	1920	31	17	10
S5	4427	2016	31	12	8
S6	3243	1535	23	11	6
S7	3101	1327	27	14	5
S8	2479	1093	15	6	11
S9	3314	1560	44	11	16
S10	2605	1192	26	15	13

Table 3. Total number of SV detected in 10 genomes with hematologic malignancy. BioNano optical mapping technology can reliably detect SV greater than 1 kbp in size.

Sample	Cancer associated events detected	Insertion disrupting cancer genes	Deletion disrupting cancer genes	Additional CNV/ translocation
S1	t(9;22): *ABL-BCR* translocation*			
S2	chr13 deletion*			t(9;13)
S3	t(11;14): *IGH-CCND1* translocation*	*CNOT3*	*TCF3*	
S4	chr20 monosomy	*CNOT3*	*TRIM33*	
S5	chr12 trisomy*			
S6		*NOTCH1*	*NUTM2B*	
S7		*NOTCH1*	*NUTM2B*	
S8	chr13 deletion			
S9				chr18 trisomy (low clonality)
S10	t(9;22): *ABL-BCR* translocation*			

Table 4. Cancer-associated mutations discovered in each sample. By running the variant-annotation pipeline, we are able to identify cancer-associated events among the plethora of events list in Table 3. In particular, the calls indicated by asterisks are also detected by targeted FISH experiments.

Conclusions

The BioNano Irys® System uses NanoChannel arrays to image DNA for genome mapping and SV detection. Irys accurately detected genetic mutation hallmarks in samples with hematologic malignancies. In addition, it detected aberrations that overlap with cytogenetic experiments. We are in the process of analyzing a broad catalog of SVs called by Irys that are missed by FISH, short-read technologies and SNP genotyping microarrays. Our results shown here indicate that the Irys platform is able to capture relevant mutations in highly complex genomes, and has the potential to provide an easy solution for detecting genetic diseases.

Acknowledgement: We like to thank Jonathan Diver from Genoptix Medical Laboratory for providing the samples and performing the FISH 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
2) 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.
3) Lam, E.T., et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nature Biotechnology (2012); 10: 2303
4) 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
5) Xiao, M et. al. Rapid DNA mapping by fluorescent single molecule detection. Nucleic Acids Research (2007); 35:e16.