Structural Variation Analysis Using Nanochannel Genome Mapping to Evaluate Genome Integrity after Induction of Pluripotency of Human Fibroblasts

EH Cho[1], H Dai[1], A Pang[1], R Williams[2], K Nazor[2], K Bhutani[3], N Schork[3], J Loring[2]
[1]BioNano Genomics, San Diego, CA
[2]Center for Regenerative Medicine, The Scripps Research Institute, La Jolla, CA
[3]Human Biology, J. Craig Venter Institute, La Jolla, CA

Abstract

Four key genes, POU5F1, SOX2, KLF4, and MYC, are commonly used for reprogramming human fibroblasts into stem cells. Induced pluripotency of human fibroblasts into stem cells currently relies on three different transfection methods of these key genes: retroviral, Sendai virus, and mRNA. The latter two methods are non-integrating methods, while the former method integrates into the genome. Retroviral transfection to induce pluripotency is currently the most efficient method, but due to the potential damaging effects of viral integration into the chromosomes, it is thought to be of high risk for future clinical applications.

To investigate whether structural changes result during the process of induction of pluripotency in stem cell genomes we used nanochannel genome mapping technology from BioNano Genomics (BNG) to compare structural differences among cell lines derived by the three reprogramming methods and the parental genome. By producing *de novo* genome map assemblies of each genome, comparing to a reference genome, and cross comparing all four assemblies, we determined sample-specific SV calls for each of the three methods. After automated assembly and comparison, followed by manual verification, 1 sample-specific SV was found in the mRNA induced sample, with respect to the Hg19 human reference genome. Examination of this structural variation found a potentially malignant X-linked deletion.

Subsequent studies confirming these results using orthogonal methods, such as next-generation sequencing, could provide strong evidence for continued pursuit of clinical applications using any of these methods. Furthermore, confirmation of no significant SV differences using the more efficient retroviral transfection method may give more confidence in the use of this method in clinical therapeutic iPSC applications.

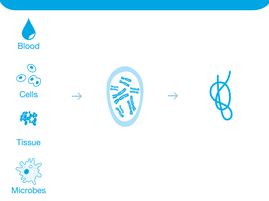
Background

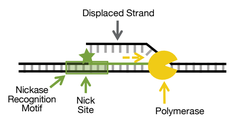
Generating high quality finished genomes replete with accurate identification of structural variation and high completion (minimal gaps) remains challenging using short read sequencing technologies alone. Instead, Irys technology provides direct visualization of long DNA molecules in their native state,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

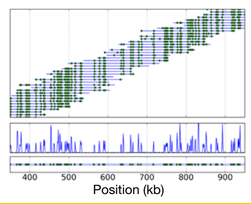
Methods

(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.

1) IrysPrep Kit extraction of long DNA molecules → 2) IrysPrep reagents label DNA at specific sequence motifs → 3) IrysChip linearizes DNA in NanoChannels → 4) Irys automates imaging of single molecules in NanoChannels → 5) Molecules and labels detected in images by instrument software → 6) IrysView software assembles genome maps



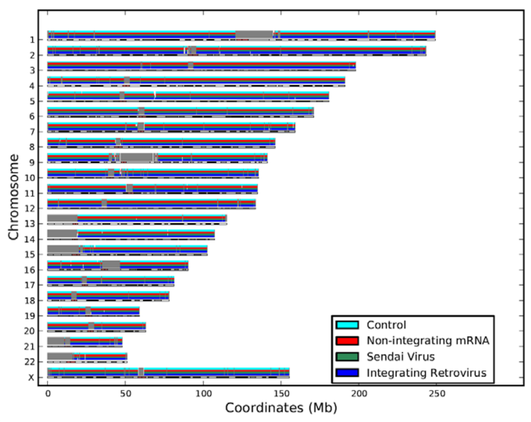




De Novo Assemblies and Genome Map Alignments

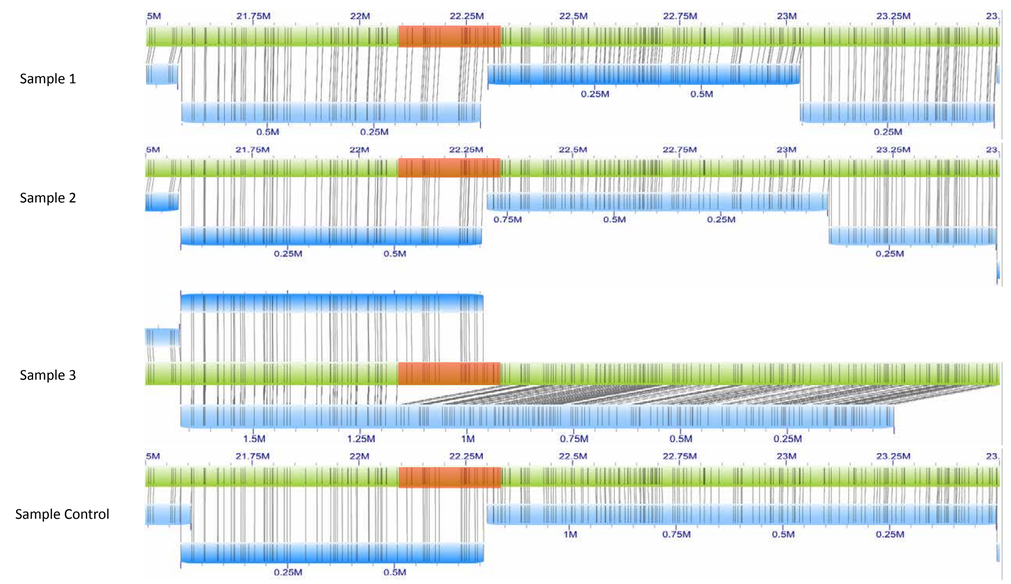
	Control	Sample 1	Sample 2	Sample 3
Molecules > 150 kb (Gb)	156.98	160.69	160.69	160.66
Molecule N50 (kb)	284.15	261.35	259.91	255.55
Molecule Coverage (X)	50.71	51.91	51.91	51.90
Assembly Contigs (#)	3,886	3,832	3,930	3,799
Total Contig Len (Mb)	2,855.30	2,820.79	2,829.19	2,847.77
Contig N50 (Mb)	0.93	0.95	0.90	0.97
Total Contig Len / Ref Len	0.92	0.91	0.91	0.92
Total Unique Len / Ref Len	0.89	0.89	0.88	0.89

*Control: Parental cell line
*Sample 1: Retrovirus-integrating
*Sample 2: Sendai Virus-non-integrating
*Sample 3: mRNA-non-integrating



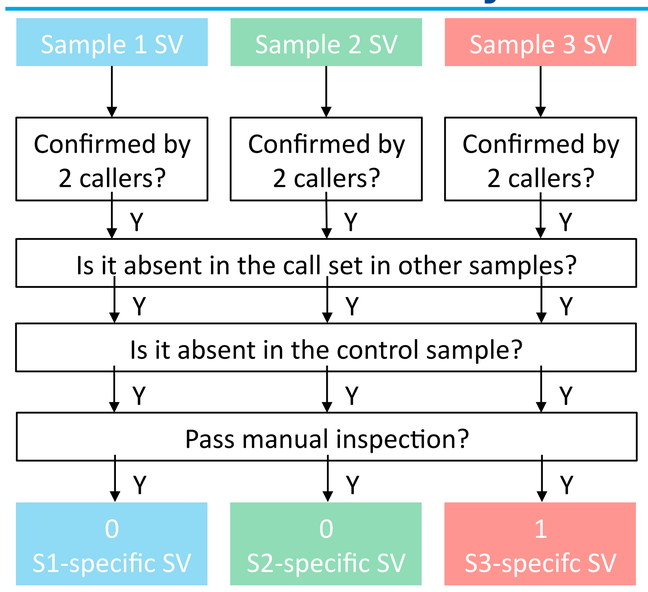
Data collected from Irys platform for *de novo* assemblies for the four cell lines (left); and alignment of genome maps of the four cell line genomes to GRch37 reference map (right). Ideogram and Giemsa banding is plotted at the bottom of each chromosome in grey scale, with centromeres highlighted in light red. "N" base gaps in the reference are shaded with grey.

Structural Variation Confirmed by Visualization



Of the four cell lines tested, one sample-specific structural variation call was found in the mRNA transfected sample: sample 3-specific 228,786 bp deletion at chrX: 22,150,079-22,386,899, potentially truncating the phosphate-regulating endopeptidase homolog, X-linked (PHEX) gene. This sample-specific call was confirmed by visualization of the assembly.

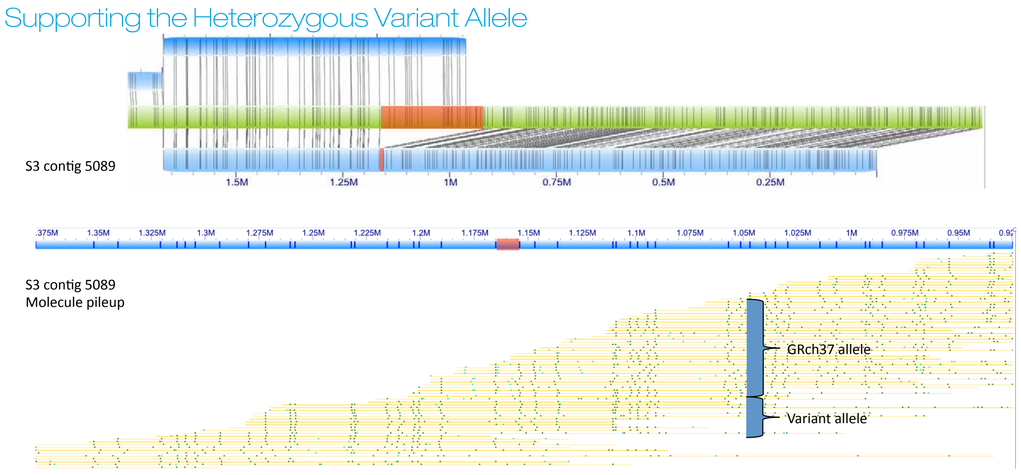
Structural Variation Analysis Workflow and Results



The workflow for identifying cell line-specific structural variation. From the list of insertions and deletions detected by BNG Irys platform, one high-confidence deletion was determined to be specific to the mRNA transfected cell line.

Single Molecule Pileup

Supporting the Heterozygous Variant Allele



The sample-specific call was further confirmed by looking at the single-molecule pileup that is output from the Irys platform. There are long molecules supporting the variant allele on genome map. It is important to note that this is a heterozygous variant, as there are molecules supporting the GRch37 allele.

Conclusions

The BioNano Genomics Irys platform was used to visualize extremely long (>150 kb) DNA molecules for direct characterization of complex structural events in the genomes of three pluripotency-induced stem cell lines and their parental control. Approximately 50× coverage of each cell line was collected on the Irys instrument and analyzed with the standard Irys de novo assembly and structural variation analysis workflow. And expanded workflow was developed to identify cell line-specific structural variations. From this analysis, one high-confidence deletion was determined to be specific to sample 3 from the list of insertions and deletions detected by BNG Irys platform, a ~228kb deletion in chromosome X. This deletion was confirmed through visualization and curation of assembled contigs and single DNA molecules and was found to be a heterozygous deletion as evidenced by the presence of the GRch37 allele. By comparing the de novo assemblies of three different methods of inducing pluripotency, we were able to identify a deletion impacting the PHEX gene in chromosome X in the mRNA transfected cell line. Deletion of PHEX gene could potentially manifest in X-linked hypophosphatemic rickets, an X-linked dominant disease. Next, we plan to compare these data with next-generation sequencing (NGS) data currently being analyzed.

References

1) Nguyen, J. V. Genomic Mapping: A statistical and algorithmic analysis of the optical mapping system. (2010).
2) Anantharaman, T. A Probabilistic Analysis of False Positives in Optical Map Alignment and Validation 1. in WABI2001 (2001).
3) Valouev, A., Schwartz, D. C., Zhou, S. & Waterman, M. S. An algorithm for assembly of ordered restriction maps from single DNA molecules. Proc. Natl. Acad. Sci. U. S. A. 103, 15770–15775 (2006).
4) 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.
5) Lam, E.T., et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nature Biotechnology (2012); 10: 2303
6) 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
7) Xiao, M et. al. Rapid DNA mapping by fluorescent single molecule detection. Nucleic Acids Research (2007); 35:e16.