# Understanding the Genomic Architecture of Cancer Genomes

Ernest T Lam[1], Alex R Hastie[1], Marcin B Imielinski[2], Cheng-Zhong Zhang[2], Jeremiah Wala[2], Željko Džakula[1], Han Cao[1]
[1]BioNano Genomics, Research and Development, San Diego, CA
[2]Broad Institute of Harvard and MIT, Center for Biomedical Informatics, Cambridge, MA

## Abstract

Understanding the genetic architecture of cancer requires whole-genome and integrative approaches. Cancers often feature genomic alterations that range from single-base changes to large-scale structural rearrangements. Having a complete catalogue of mutations in cancer is crucial for identifying key drivers and providing accurate diagnosis, prognosis, and targeted therapy.

Next-generation sequencing (NGS) platforms have limited power to decipher large, complex structural variants frequently observed in cancer. Genome mapping represents a complementary technology that provides critical structural information. It involves high-throughput analysis of single molecules spanning hundreds of kilobases in NanoChannels. Long-range information is preserved and direct interrogation of complex structural variants made possible. Leveraging the strengths of these complementary platforms provides a more comprehensive view of a cancer genome.

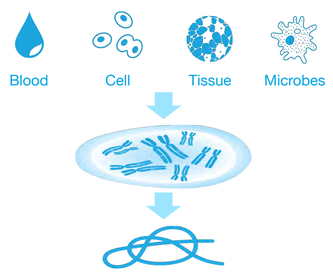
Here, we present our analysis of a well-studied and highly rearranged cancer genome. We constructed completely *de novo* genome map assemblies with N50 lengths of more than 1 Mb, allowing for discovery of rearrangement events. We derived normalized copy number profiles of matched tumor-control pairs based on genome mapping data. We observed that tumor samples had highly variable copy number profiles, corresponding to focal and chromosome-scale changes. In addition, we found that integrating NGS and next-generation mapping data provided a comprehensive view of a cancer genome.

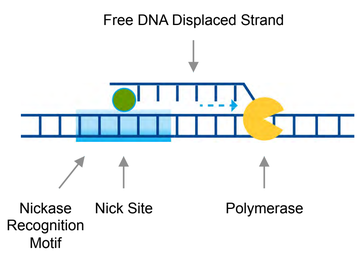
## Background

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® System 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.

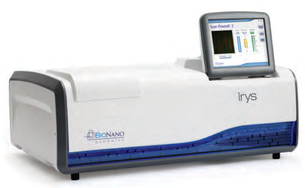
## Methods

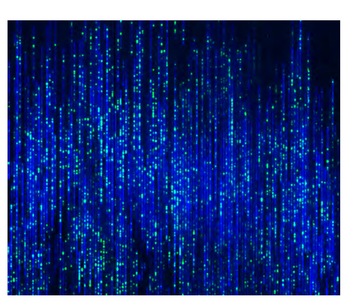
1. IrysPrep® kit extraction of long DNA molecules
2. IrysPrep® reagents label DNA at specific sequence motifs
3. IrysChip® linearizes DNA in NanoChannels
4. Irys® automates imaging of single molecules in NanoChannels
5. Molecules and labels detected in images by instrument software
6. IrysView® software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

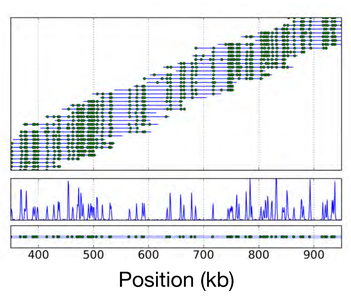






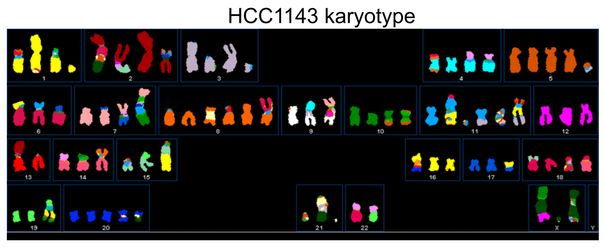



(1) Long molecules of DNA is labeled with IrysPrep® reagents by (2) incorporation of fluorophore labeled nucleotides at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the IrysChip® NanoChannels and single molecules are imaged by Irys®. (4) Single molecule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

## Analysis of HCC1143 (and paired control)

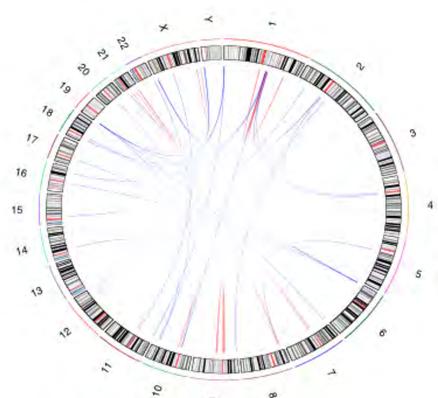
A hypo-tetraploid breast cancer cell line, extremely rearranged, with 48 chromosomal abnormalities.

HCC1143 karyotype



## Putative inter-chromosomal translocation calls

- 17 calls in control (9 shared by case)
- 39 calls in case, hence 30 putative translocations unique to cancer sample
- One example involves a putative t(1:16) translocation, overlapping with NBPF20 and HYDIN, both of which have been associated with breast cancer.
- Other genes involved:
  - HECW1: suppresses ErbB4
  - EPHB1: aberrant expression in human cancers



## HCC1143 Raw data and assembly summary statistics

**Single-molecule stats**

| | | |
|---|---|---|
| Input molecules | Gb | 273.11 (~85X) |
| Avg mol. length | kb | 310.71 |

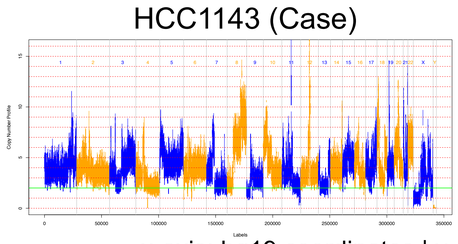
***De novo* assembly stats**

| | | |
|---|---|---|
| #Maps | N | 2797 |
| Assembly size | Gb | 2.9 |
| Map N50 | Mb | 1.42 |
| %Aligned maps | % | 99% |
| Overlap with hg19 | % | 88% |

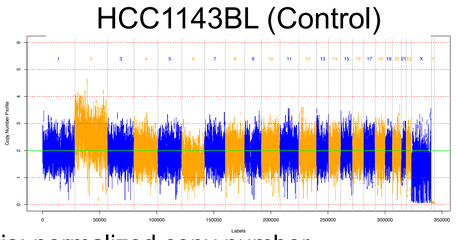
## CNV profiling

CNV profiling provides a first glance of the genome. Single-molecule are aligned to reference *in silico* map; coverage counts are then normalized. Dramatic fluctuation is observed in cancer sample compared to control.

HCC1143 (Case)



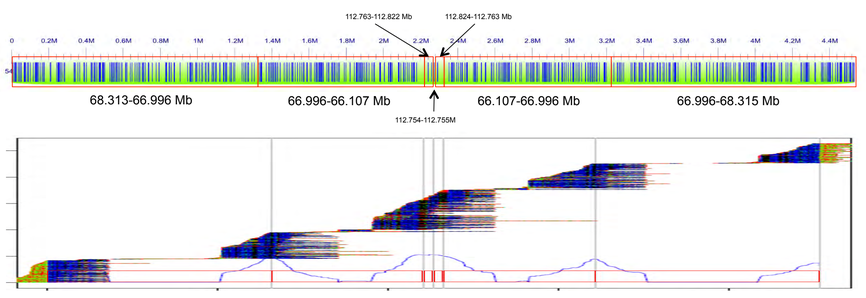
HCC1143BL (Control)



x-axis: hg19 coordinates | y-axis: normalized copy number

## Phasing of translocation breakpoints

Putative derivative chromosome structure based on NGS data can be validated and phased by single-molecule mapping




## Conclusions

Analysis of cancer genomes remains one of the major challenges in genomic analysis due to their genome complexities. Current NGS approaches can detect individual translocation breakpoints but are limited by their short-read lengths and insert sizes. Next-generation mapping provides additional long-range information about the genome structure, making access to high-coverage population analysis of cancer genomes a more feasible pursuit. For more information about next-generation mapping, also see Posters #1832T, #3118T, #1632F and #2496F.

## Reference

1. Cao, H., et al. Rapid detection of structural variation in a human genome using NanoChannel based genome mapping technology. GigaScience (2014); 3 (December 2014): 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.

©2015 BioNano Genomics. All rights reserved.