Structural Variant Detection with Hi-C Technology

Discover and detect structural variants throughout the genome to gain improved insights into disease mechanisms.

Structural Variants Impact Human Health and Disease

Genomic studies have shown that structural variants (SVs) contribute significantly to disease and disease susceptibility. Typically, SVs are genomic alterations >50 bp in length—including translocations, inversions, insertions, deletions, and duplications—and account for most of the genetic variation between human haplotypes[1]. The average human genome contains more than 20,000 SVs, which occur in both coding and non-coding regions[2]. Notably, large SVs are more than 30 times as likely to affect the expression of a gene when compared to single nucleotide variants[3]. Still, the extent of the effects of SVs on gene regulation remains largely uncharacterized.

SVs are pathogenic in numerous disease processes and are considered a hallmark of cancer genomes[1]. Notably, an extensive study of 2,658 cancers across 38 tumor types showed that 95% of characterized tumor samples contain one or more structural variants[4]. Whether at the individual genome level or involving entire chromosomes, the rearrangement of genetic material is an important form of somatic mutations and can lead to oncogenesis through various mechanisms, many of which remain unclear. It is critical to discover and detect SVs throughout the genome to gain a comprehensive understanding of disease mechanisms.

Arima Hi-C Enables SV Detection with Short-Read Sequencing

One of the mechanisms by which SVs can impact the regulation of cancer genes is by altering the 3-dimensional (3D) genome organization. While traditional short-read sequencing is limited in detecting structural variants, combining high-throughput chromatin conformation capture (Hi-C) technology with short-read sequence data enables the detection of SVs[5]. This approach can detect and discover large-scale rearrangements through breakpoint identification. These translocations can be identified on Hi-C contact frequency heat maps and have a characteristic butterfly appearance[6].

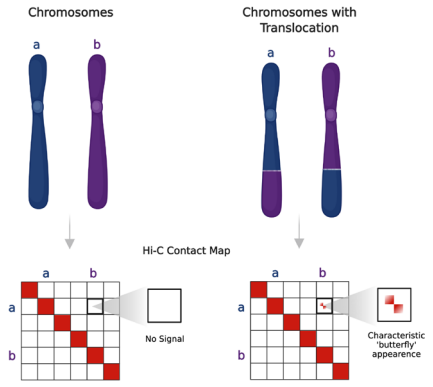


Figure 2. Chromosomal translocations can be identified in Hi-C contact frequency heat maps, where interchromosomal translocations appear as a characteristic butterfly pattern*.

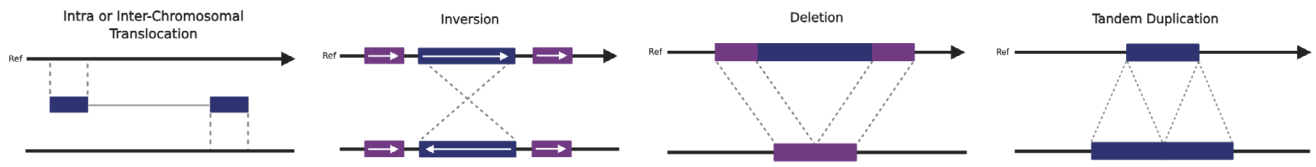


Figure 1. Structural variants result from the addition, subtraction, or rearrangement of sequence in the genome*.

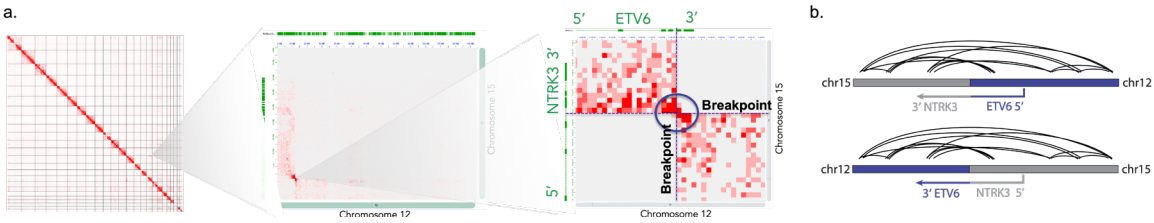


Figure 3. A gene fusion was identified from a confirmed fibrosarcoma FFPE sample using the Arima-HiC+ FFPE kit and Arima-SV bioinformatics pipeline. Hi-C heat maps showing relative spatial proximity and breakpoints between chr12 and chr15 resulting in a reciprocal gene fusion between Neurotrophic Receptor Tyrosine Kinase 3 (*NTRK3*) and ETS Variant Transcription Factor 6 (*ETV6*) (a). Pictorial representation of the reciprocal *NTRK3-ETV6* gene fusion (b)[8].

Even in cancer genomes, Hi-C technology provides high accuracy for identifying inter- and intrachromosomal translocations and rearrangements with relatively shallow sequencing requirements[6,7]. For example, Hi-C can detect deletions, tandem duplications, and inversions ≥1 Mb, with breakpoint resolution of 1kb and the potential to pinpoint the breakpoint to 1bp resolution. Hi-C also enables the detection of complex rearrangements in the genome by linking multiple SVs, overcoming unalignable junctions, and mapping SVs inside of repeats[6].

With Arima Hi-C technology and the Arima-SV bioinformatics pipeline, scientists can now visually and algorithmically identify structural variants in genes known to have clinical significance with high concordance to orthogonal datasets[8].

Customer Success: Detecting Structural Variants in Peripheral Blood

Identifying the chromosome translocation that drives B-cell acute lymphoblastic leukemia (B-ALL) is typically carried out by bone marrow biopsy – a painful and invasive procedure. However, scientists at the University of Calgary demonstrated that structural variants are detectable in peripheral blood samples using Arima Hi-C technology, reducing the need for invasive bone marrow biopsies[6].

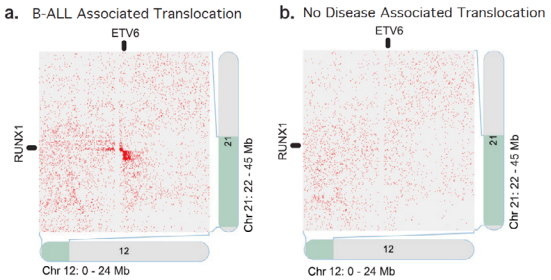


Figure 4. Hi-C contact maps from a peripheral blood sample with a known structural variant for B-ALL, specifically a balanced translocation at the *ETV6* locus on chr12 and the *RUNX1* locus on chr21 (a). A blood sample with no known variant at that locus demonstrates the absence of abnormal signal in the Hi-C data (b)[6].

Customer Success: Altered Genome Conformation Yields New Drug Targets

Using Arima genome-wide Hi-C, an international team of scientists was able to identify chromosomal rearrangements – including the formation of new topologically associating domains, known as a neo-TAD – in LAMC1 and other genes. The team was subsequently able to identify the regulatory mechanisms underlying the aberrant expression of these genes as essential for ependymoma tumorigenesis. These findings suggest these genes may provide novel therapeutic targets[8].

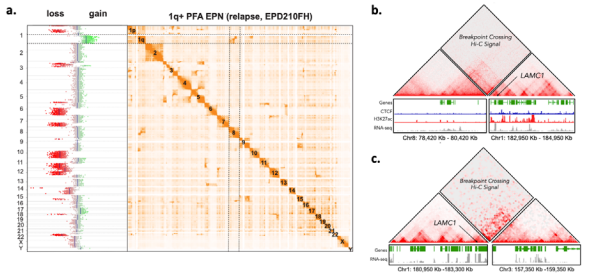


Figure 5. Genome-wide Hi-C heat maps were used to identify complex inter-chromosomal structural variants, including an inversion that involves chr1q and chr8 (a), which results in the formation of a neo-TAD in cultured cells that alters the regulatory environment of the *LAMC1* gene (b). Similarly, in a PFA relapse FFPE tumor sample, a structural variant involving chr1q and chr3 results in altered *LAMC1* regulation (c), which activates gene transcription relative to expression in primary PFA tumors[9].

End-to-End Solution for Structural Variant Detection

Arima Genomics Hi-C technology allows scientists to detect and discover SVs in a broad range of sample types, including formalin fixed paraffin embedded (FFPE) tissues. Although FFPE samples are a critical archival and clinically relevant sample type, they typically perform poorly in most molecular assays due to DNA damage between and within DNA molecules. However, using the Arima-HiC+ sample preparation kits and the Arima-SV bioinformatics pipeline, scientists can:

- Use a variety of sample types, including FFPE samples, fresh/frozen tissue, blood, and cell culture
- Detect coding and non-coding structural variants using Hi-C data and Arima Genomics advanced bioinformatics pipeline
- Link structural variants to impacts on gene function by mapping 3D interactions between the variant and the affected gene

Workflow Summary

Whether using the standard Arima-HiC+ kit, or the Arima-HiC+ FFPE kit, our workflow lets you go from sample to discovery quickly and easily[10,11].

Sample Prep
using standard Arima-HiC+ or Arima-HiC+ FFPE

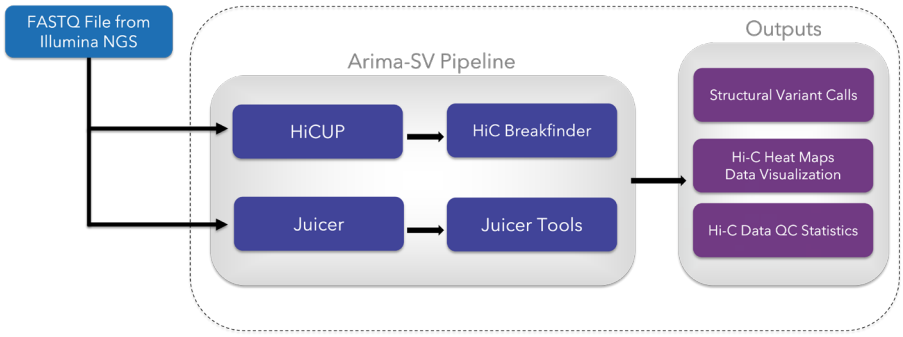
Library Prep
with pre-validated protocols

Sequencing
with standard NGS platforms

Data Analysis
with Arima-SV Pipeline

Bioinformatics Pipeline

The Arima-SV pipeline incorporates the critical Hi-C analysis tools into a portable and scalable bioinformatics workflow that is easy to install and run[12].



"Using Arima Hi-C technology and the new FFPE sample preparation method and bioinformatics tools, we've been widely successful in detecting structural variants in a variety of tumor samples. We are hopeful that additional insights gained with this approach will lead to improved understanding of disease mechanisms and, ultimately, the development of new therapeutic options for people with cancer."

– Matija Snuderl, MD, Director of Molecular Pathology and Diagnostics, NYU Langone Medical Center

References

1. Eichler, E. E. (2019). Genetic variation, comparative genomics, and the diagnosis of disease. New England Journal of Medicine, 381(1), 64–74.

2. Chaisson, M. J. P., et al. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nature Communications, 10(1), 1784.

3. Chiang, C., et al. (2017). The impact of structural variation on human gene expression. Nature Genetics, 49(5), 692–699.

4. Li, Y., et al. (2020). Patterns of somatic structural variation in human cancer genomes. Nature, 578(7793), 112–121.

5. Dixon, J. R., et al. (2018). Integrative detection and analysis of structural variation in cancer genomes. Nature Genetics, 50(10), 1388–1398.

6. Mallard, C., et al. (2022). Hi-C detects genomic structural variants in peripheral blood of pediatric leukemia patients. Molecular Case Studies, 8(1), a006157.

7. Harewood, L., et al. (2017). Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. Genome Biology, 18(1), 125.

8. Schmitt, A., (2021). Webinar: Leveraging 3D genomics for SV detection and gene regulation analysis in cancer genomes: technology update. Arima Genomics.

9. Okonechnikov, K., et al. (2020). Oncogenic 3D genome conformations identify novel therapeutic targets in ependymoma. Research Square Preprint. In Review.

10. User Guide: Arima-HiC for Animal Tissues. Arima Genomics.

11. User Guide: Arima-HiC for Formalin Fixed Paraffin Embedded (FFPE) Tissues. Arima Genomics.

12. Bioinformatics User Guide: Arima Structural Variant Pipeline. Arima Genomics.

Connect with an Arima Genomics scientist to plan your next project.

AN-001-022122 | *Figure created with BioRender.com

ARIMA GENOMICS