Mapping Regulatory Mechanisms and Promoter-Enhancer Interactions using Promoter Capture Hi-C

Discover information about the activity of tens of thousands of promoters to explore how genome regulation impacts human disease.

Introduction

In gene expression, promoter regions work together with enhancers and other gene-regulating elements to drive transcription. Mutations that disrupt these elements can lead to epigenetic changes associated with a number of diseases[1, 2]. Furthermore, the three-dimensional organization of chromatin in these non-coding regions is a critical aspect of disease[3]. Promoters and enhancers are not always close in sequence to their target genes, and their interactions are governed by their positions in space.

Predicting the locations of distal target genes can be a challenge. Promoter capture Hi-C (pcHi-C) is a method that has enabled researchers to systemically identify the interactomes, including in distal regulatory regions, for tens of thousands of promoters[4].

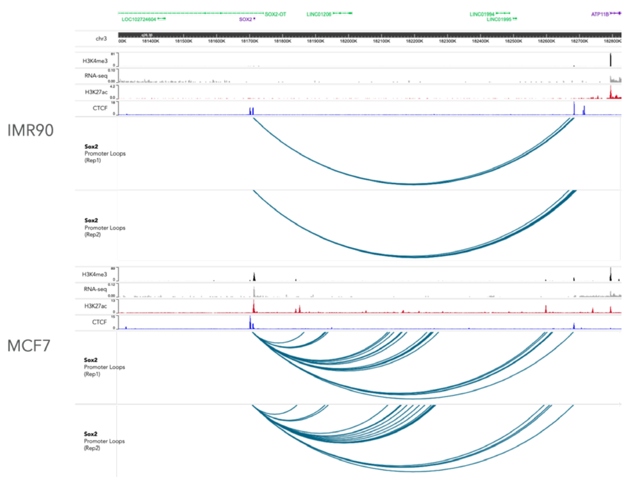


Figure 1. Here pcHi-C captures differential regulation of SOX2 between IMR90 lung fibroblasts and MCF7 breast cancer cells. Blue arches represent promoter interactions.

Promoter capture Hi-C has been used to enrich both mouse and human Hi-C libraries for distal promoter-interacting regions (PIRs) for all promoters in a single experiment, making it a powerful tool for mapping the ensemble of DNA sequences interacting with promoters and identifying promoter-interacting regions with regulatory potential[3,5].

Selected Applications of Promoter Capture Hi-C

Genomics and Epigenomics Researchers: Genomics and Epigenomics researchers are harnessing the power of pcHi-C to delve into the complexities of gene regulation and the influence of non-coding genetic variants on human diseases. This innovative technique is pivotal for understanding the intricate structure and function of the genome. Particularly, it sheds light on the role of the genome's three-dimensional architecture in the control of gene expression. pcHi-C has become an indispensable tool in genome-wide association studies, enabling researchers to pinpoint genes linked to a range of conditions, including cancer and autoimmune diseases, thus offering new avenues for understanding these complex disorders[5].

Researchers in Disease Genetics: In the field of disease genetics, researchers are utilizing pcHi-C to map long-range promoter interactions, a key factor in understanding the genetic basis of various diseases. This method has been instrumental in uncovering genes associated with cardiovascular disease risk and revealing critical aspects of chromatin architecture in colorectal cancer cells. Furthermore, it has been used to characterize promoter-enhancer interactions in human hematopoietic stem cells, both normal and malignant. These studies have led to the identification of previously unknown genes linked to cardiovascular disease, potential therapeutically significant histone modifications in colorectal cancer, and insights into transcription dynamics during the development of normal and cancerous cells, making pcHi-C a valuable tool in disease genetics research[4, 6, 7].

Researchers Analyzing Inflammatory Responses: A group of researchers focusing on inflammatory responses in SARS-CoV-2 infection are integrating pcHi-C data with GWAS meta-analysis data. This integrated approach has been a game-changer in identifying genetic factors associated with COVID-19 susceptibility and severity. By doing so, they were able to compare over 250 genes linked to COVID-19's highly variable disease progression and clinical outcome, many of which were not detected using standard nearest-exon approaches. This novel application of pcHi-C demonstrates its potential in unravelling complex genetic interactions and contributing significantly to our understanding of infectious diseases[8].

Drug Discovery Scientists: Drug discovery scientists are turning to pcHi-C to unearth novel biomarkers for various medical conditions. By analyzing the regulatory interactions between promoters and their target genes, they are able to identify key linkages that could serve as potential therapeutic targets. This aspect of pcHi-C is particularly valuable in the quest for new drugs, as it provides a deeper understanding of the molecular pathways involved in diseases and opens up possibilities for targeted treatments. The application of pcHi-C in this field holds great promise for the development of innovative therapeutic strategies and the discovery of much-needed novel biomarkers[5, 8].

Integrating Promoter Capture Hi-C with Other Data Types

Promoter capture Hi-C can be integrated with various other data types to provide a more comprehensive understanding of the genome's regulatory system.

Genome-Wide Association Study (GWAS) variants are a rich source of data that can illuminate mechanisms of regulation or mis-regulation. However, the large number of variants, particularly non-coding variants, can make their functional assessment challenging. Integrating these data with pcHi-C data can illuminate how GWAS variants are directly impacting gene regulation[9,10].

Assay for Transposable Access Chromatin-using sequencing (ATAC-seq) data can be used to assess the availability of chromatin across the genome. Integrating this data type with pcHi-C data can help identify active regulatory elements that are interacting with promoters and thus potentially regulating gene expression[11, 12, 13, 14].

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) data can be used to identify the binding sites of transcription factors as well as histone modifications across the genome. Integrating ChIP-seq data with pcHi-C data can help reveal the regulatory elements involved in promoter-enhancer interactions and their potential roles in gene regulation[11].

RNA sequencing (RNA-seq) data can be used to measure gene expression levels across the genome. Differential expression between samples can identify candidate genes that may be the focus of other data sets. Integrating this with pcHi-C data can help reveal the functional consequences of promoter-enhancer interactions on gene expression, as well as provide insights into the regulatory mechanisms underlying various biological processes and diseases[15].

By using combinations of RNA-seq, CHiP-seq, and ATAC-seq along with pcHi-C, and integrating these findings with GWAS results, researchers can paint a more complete picture of the genes associated with various diseases, as well as uncover new regulatory loci and their target genes.

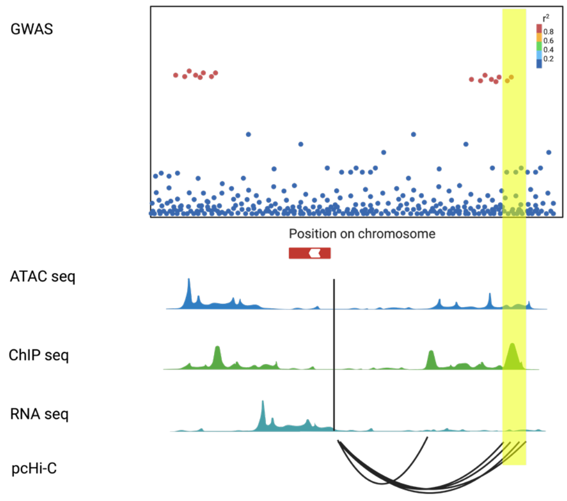


Figure 2. In this visual representation of a multi-omic dataset, GWAS data combined with ATAC-seq, ChIP-seq and RNA seq can help illuminate the functional mechanisms that drive expression of a given gene. With promoter capture Hi-C data, researchers can establish a more direct causal relationship between a promoter and a distal non-coding variant.

Arima Promoter Capture HiC

Arima Genomics provides a range of tools for analyzing pcHi-C data, including the Arima Promoter Capture Panel and the Arima Capture HiC Pipeline. These tools are tailor-made to facilitate the analysis of Arima-generated pcHi-C data and help researchers discover key relationships between promoters and their targets.

Hi-C Prep	Library + Target Capture	Sequencing	Data Analysis
Rapid 6-hour Protocol	Library Prep Module + 1-hour Hybridization	Illumina Next Generation Sequencing	Open and Optimized Bioinformatics Pipelines

Arima Promoter Capture Panel allows researchers to identify key linkages between promoters and their target genes with high cis:trans ratios, high on-target read coverage at low sequencing depths, with minimal variability between replicates. This panel can be used to generate conformational profiles of promoter regions, recognize novel biomarkers, and identify potential therapeutic targets[16].

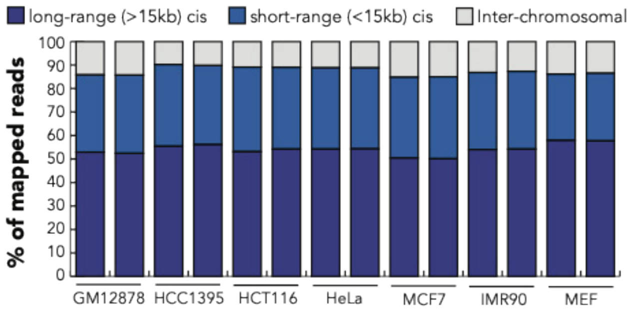


Figure 3. High Cis (signal):Trans (noise) ratio leads to high quality interaction data. The Arima Promoter Capture module was used against 7 sample cell lines (GM12878, HCC1395, HCT116, HeLa, MCF7, IMR90 and MEF). We demonstrate that across all sample types, cis-reads consistently represented 85-90% of all mapped reads. A high cis to trans ratio of mapped reads is a strong indicator of excellent signal:noise which will allow efficient identification of spatially conserved chromatin compartments.

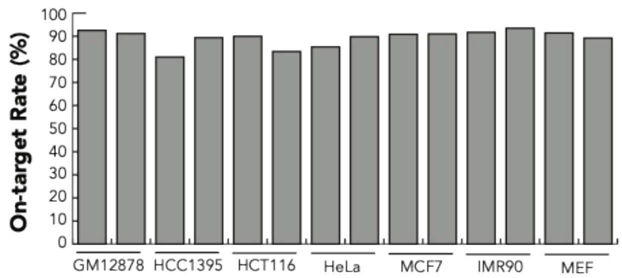
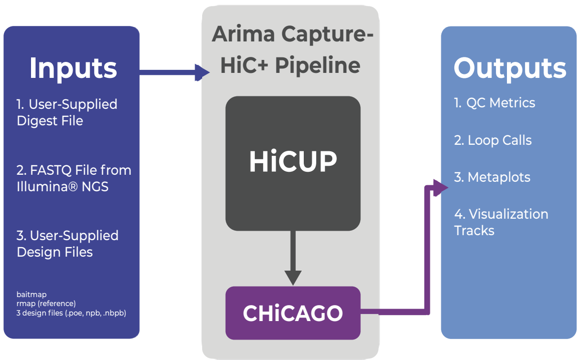


Figure 4. Very high on-target read coverage maximizes sequencing efficiency and reduces sequencing depth requirements. 7 samples were processed using the Arima Capture-HiC+ platform including the Arima Library Prep kit, the Arima HiC+ kit and the Arima Promoter Capture module.

The **Arima Capture HiC pipeline**, hosted on GitHub, is designed for the analysis of capture HiC data utilizing the open-source HiCUP and CHiCAGO pipelines. These pipelines facilitate the generation of pc-HiC QC metrics and the production of data output files, optimizing the process for researchers working with the Arima Genomics pc-HiC kits. Minor changes were made to the default parameters, which have been optimized to improve sensitivity and specificity[17].



> "The ability of Arima-HiC kits to deliver greater insight with less sequencing cost will be critical in enabling more researchers to leverage Hi-C technology for understanding gene regulation within the three-dimensional context of the genome."
>
> – Changhoon Kim, PhD, Chief Technology Officer Macrogen, Inc.

References

1. Matharu, N. and Ahituv, N. (2020). Modulating gene regulation to treat genetic disorders. Nature Reviews Drug Discovery, 19(1), 757-775.
2. Claringbould, A. and Zaugg, J. B. (2021). Enhancers in disease: molecular basis and emerging treatment strategies. Trends in Molecular Medicine, 27(11), 1060-1073.
3. Aljogol, D. et al. (2022). Comparison of capture Hi-C analytical pipelines. Frontiers in Genetics, 13(1).
4. Tomás-Daza, L. et al. (2023). Low input capture Hi-C (liCHi-C) identifies promoter-enhancer interactions at high-resolution. Nature Communications, 14(268).
5. Schoenfelder, S. et al. (2018). Promoter capture Hi-C: high-resolution, genome-wide profiling of promoter interactions. Journal of Visualized Experiments, 136(57320).
6. Montefiori, L. E. et al. (2018). A promoter interaction map for cardiovascular disease genetics. eLife, 7(35788).
7. Saw, A. K. et al. (2022). Combined promoter-capture Hi-C and Hi-C analysis reveals a fine-tuned regulation of 3D chromatin architecture in colorectal cancer. bioRxiv, preprint.
8. Thiecke, M. J. et al. (2021). Prioritisation of candidate genes underpinning COVID-19 host genetic traits based on high-resolution 3D chromosomal topology. Frontiers in Genetics, 12(745672).
9. Zhou, L., Zhao, F. (2018). Prioritization and functional assessment of noncoding variants associated with complex diseases. Genome Med 10, 53.
10. Noboru J. Sakabe et al., (2020). Transcriptome and regulatory maps of decidua-derived stromal cells inform gene discovery in preterm birth. Sci. Adv.6,eabc8696.
11. Lan, X. et al. (2012). Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. Nucleic Acids Research, 40(16), 7690-7704.
12. Advertisement feature: Previewing the functional genome with ATAC-Seq and NGS. Nature Portfolio.
13. Cao, Z. and Gao, G. (2022). Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. Nature Biotechnology, 40(1), 1458-1466.
14. Liu, C. et al. (2023). Single-cell multi-omics integration for unpaired data by a Siamese network with graph-based contrastive loss. 24(5).
15. Yun, H. et al. (2022). Multiomics data integration to reveal chromatin remodeling and reorganization induced by gene mutational synergy. STAR Protocols, 3(4), 101770.
16. Product flyer: Arima Promoter Capture Panel. Arima Genomics.
17. Arima Capture HiC Pipeline using Arima-HiC+ Kit. GitHub.