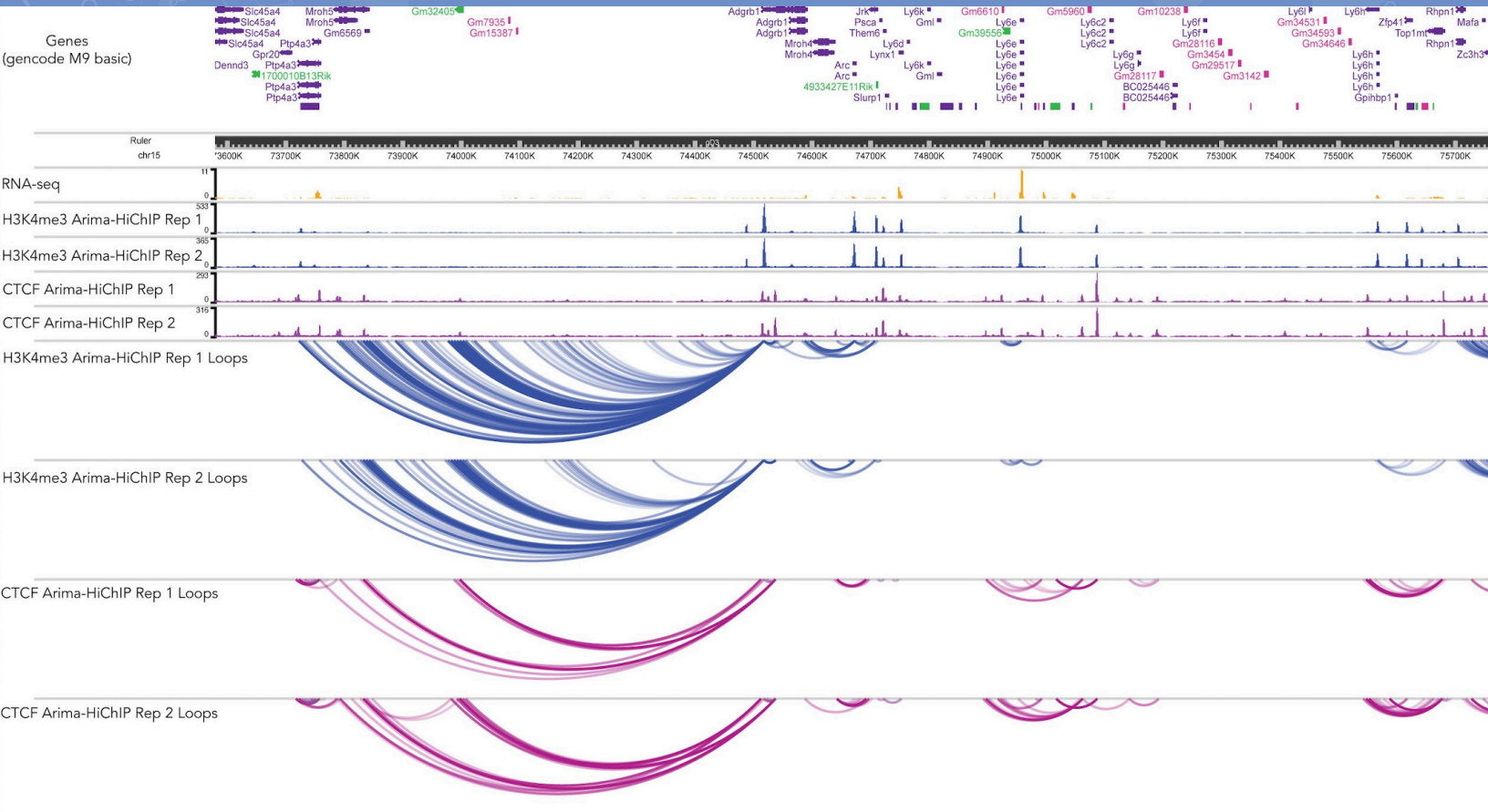
ARIMA GENOMICS

Arima-HiChIP

Multi-Omic Analyses of Transcriptional Protein Binding and Chromatin Interactions



1. Introduction

In early 2020, Arima Genomics introduced the Arima-HiChIP protocol which leverages the Arima HiC+ kit. At the time of launch, the assay was validated against cell lines and two transcription factors. In this updated technical note, we demonstrate Arima-HiChIP utility for transcription factors (validated against CTCF), tissues (validated against mouse brain tissue) and now includes an updated bioinformatic pipeline and user guide that combines ChIP peak calls with HiC loops for a streamlined, multi-omics assay.

Three-dimensional (3D) chromatin conformation and protein:DNA interactions both have a profound impact on numerous biological processes. HiC sample preparation and chromatin immunoprecipitation (ChIP) are techniques that, when combined with next generation sequencing, deliver 3D spatial genomic information and 1D protein-binding information respectively. Combined, these data yield long-range chromatin interactions associated with the enriched protein factor of interest.

Running ChIP and HiC assays separately poses various challenges with regard to sample consumption and sequencing cost. While Arima-HiChIP, reduces sequencing cost through antibody-enrichment and lowers the sample burden via a single assay, there remains added challenges with regard to sample compatibility, antibody-target support, and complex bioinformatics. Here we introduce Arima-MAPS 2.0~~+~~, which combines FEATHER, MACS2 and MAPS for streamlined production of chromatin loops associated with immunoprecipitated protein in a single pipeline.

2. Materials & Methods

2.1 DNA samples

Testing was performed internally and externally using GM12878 lymphoblastoid cells and mouse brain tissue to demonstrate compatibility with cell lines and tissue. Antibodies tested include H3K27ac (Active Motif, P/N: 91193 or 91194), H3K4me3 (Millipore, P/N:04-745) and CTCF (Active Motif Cat # 91285) - demonstrating utility in both histone and transcription factors.

2.2 Arima-HiC+ for HiChIP Library Preparation

The Arima-HiChIP protocol is a 3-day protocol that begins with primary cells, cell cultures or fresh or frozen tissue (Figure 1). The process begins with Arima-HiC, in which chromatin undergoes crosslinking and proximity ligation, shearing and antibody binding (Day 1). The following day features chromatin immunoprecipitation in which antibody-bound chromatin is isolated and purified, resulting in enriched proximally ligated DNA (Day 2). Arima-HiChIP libraries are prepared using a pre-validated, commercially available kit (Day 3).

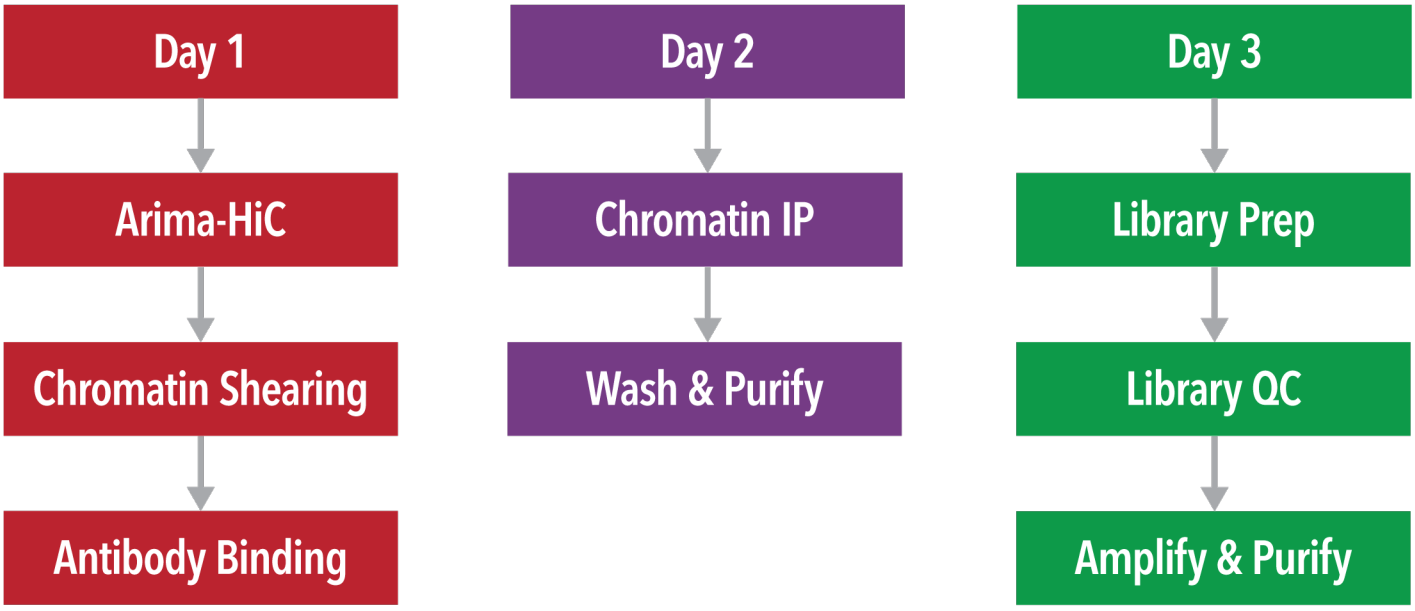


Figure 1: The Arima HiChIP pre-sequencing workflow.

2.3 Sequencing and Analysis

Arima-HiChIP libraries first underwent shallow-sequencing to assess the quality of 3D long-range interactions and 1D ChIP enrichment, and determine the estimated number of reads needed for deep sequencing. Libraries are first sequenced at shallow depth (~1M reads pers ample) using paired-end mode on Illumina next generation sequencers.. The data is then mapped to a reference genome using Arima-MAPS 2.0, a modified version of the open-source MAPS pipeline2 adapted for Arima-HiChIP data. Two signals were enumerated: (1) long-range cis interactions that are captured from proximity ligation, and (2) the fraction of reads enriched at ChIP peaks. The product of these two features (3D and 1D) in the HiChIP data were then used to determine the sequencing depth needed for each sample to obtain sufficient long-range cis interactions anchored at each ChIP peak for reproducible chromatin looping analyses. All of the aforementioned data features and statistics (plus several more) are tallied in a convenient quality control table output by the Arima-MAPS 2.0 pipeline, and can be copy/pasted into the Arima-HiChIP QC Worksheet for tabulation and analysis.

A set of ChIP peaks are needed for HiChIP analysis. The Arima-MAPS 2.0 pipeline automatically calls ChIP peaks from the HiChIP data using MACS2. The MACS2 parameters have been optimized for high precision and recall based on benchmarking to known ChIP-Seq peaks.

After assessing the quality of shallow-sequencing output and determining the required sequencing depth needed for more in-depth analysis, samples undergo deep sequencing using the same sequencing parameters described previously. Arima-HiChIP data were again mapped to a reference genome using Arima-MAPS 2.0. In addition to the quality metrics described above, the deep sequencing data allows the discovery of thousands of protein-associated chromatin loops at a genome-wide scale using the Arima-MAPS 2.0 pipeline.

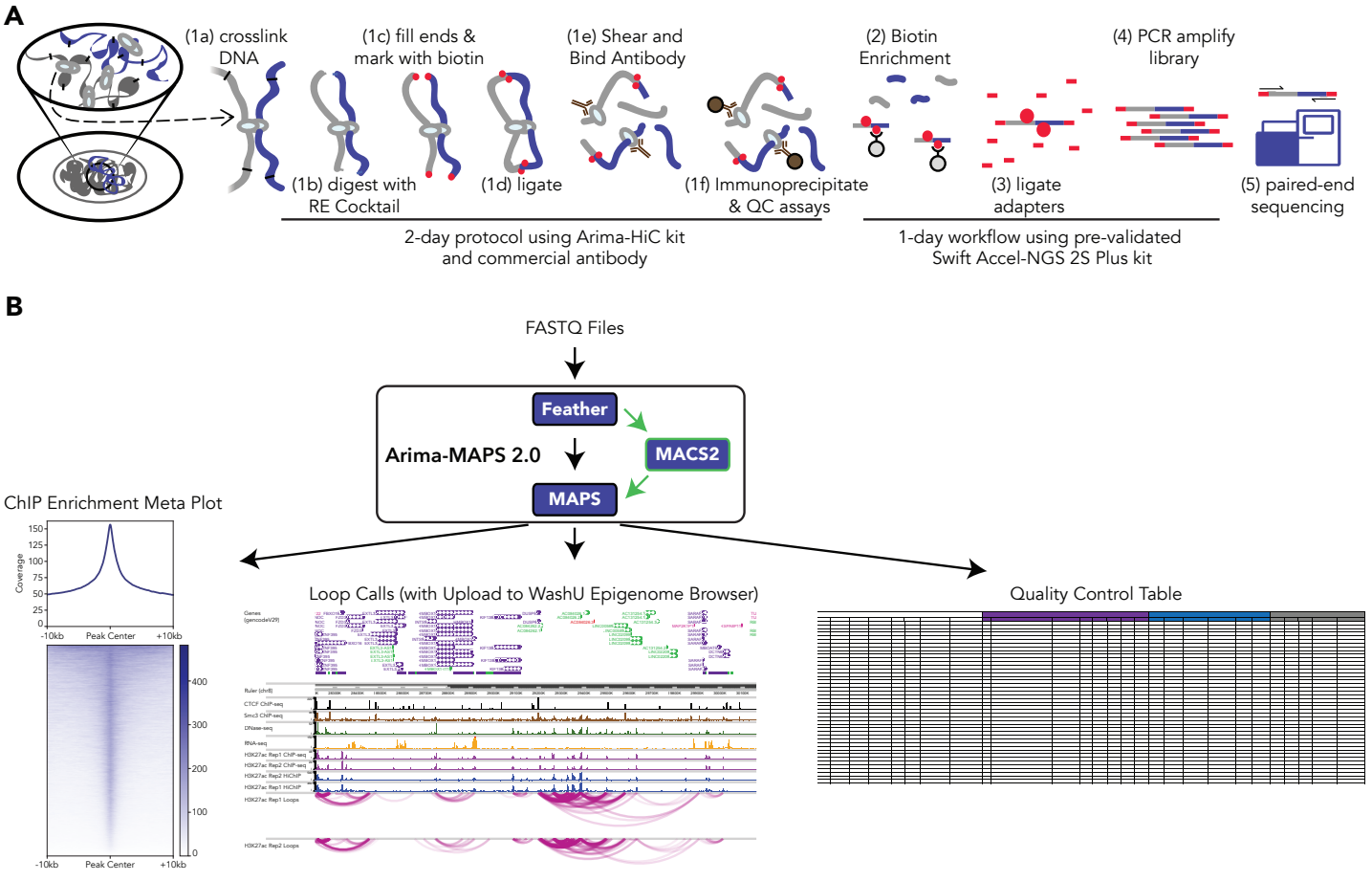


Figure 2: Arima-HiChIP workflow now includes Arima-MAPS 2.0. A) The Arima-HiChIP workflow is a streamlined protocol that results in immunoprecipitated biotin-labeled proximally ligated DNA that was associated with the immunoprecipitated protein target. The immunoprecipitated DNA is enriched for biotin and prepared as a library and sequenced in paired-end mode on Illumina sequencing instruments. B) Arima-HiChIP data can be ana- lyzed using the Arima-MAPS pipeline, producing ChIP enrichment metaplots and heatmaps around known ChIP peaks (if available), loop calls that can be uploaded to the WashU Epigenome Browser, and a QC table for high level and detailed analysis of data quality.

3 Results

3.1 HiChIP yields ChIP-enriched long-range interactions

One histone factor (H3K4me3) and one transcription factor (CTCF) were assessed using the Arima HiC+ kit and an updated Arima-HiChIP protocol. To illustrate the type of 1D ChIP enrichment and 3D chromatin looping data obtained from Arima-HiChIP, a snapshot of the MYC locus is provided (Figure 4). Plotted below the chromatin, DNA accessibility, and transcriptome tracks obtained from ENCODE are the 1D coverage tracks of the Arima-HiChIP data. The H3K27ac HiChIP 1D signal is nearly identical compared to the H3K27ac ChIP-seq signal directly above, suggesting not only the strong signal to noise of the ChIP enrichment in the HiChIP data, but the similar capture of 1D protein localization peaks compared to conventional ChIP-seq.

In developing an updated HiChIP protocol, histone factor HiChIP (HF HiChIP) was validated against H3K4me3 and mouse brain. Fold enrichment ranged between 6-11X for internally generated data within cell lines and tissues. An average 6X fold enrichment was achieved for cell lines and 11X for tissues. Cis-reads (signal) represented 83%. The signal:noise ratio is measured at 5X.

Transcription factor HiChIP was validated with CTCF against GM12878 cell line and mouse brain tissue. Fold enrichment of this target ranged between 11–23X (avg. 21X for cell lines and avg. 12X for tissues) for internally generated data.

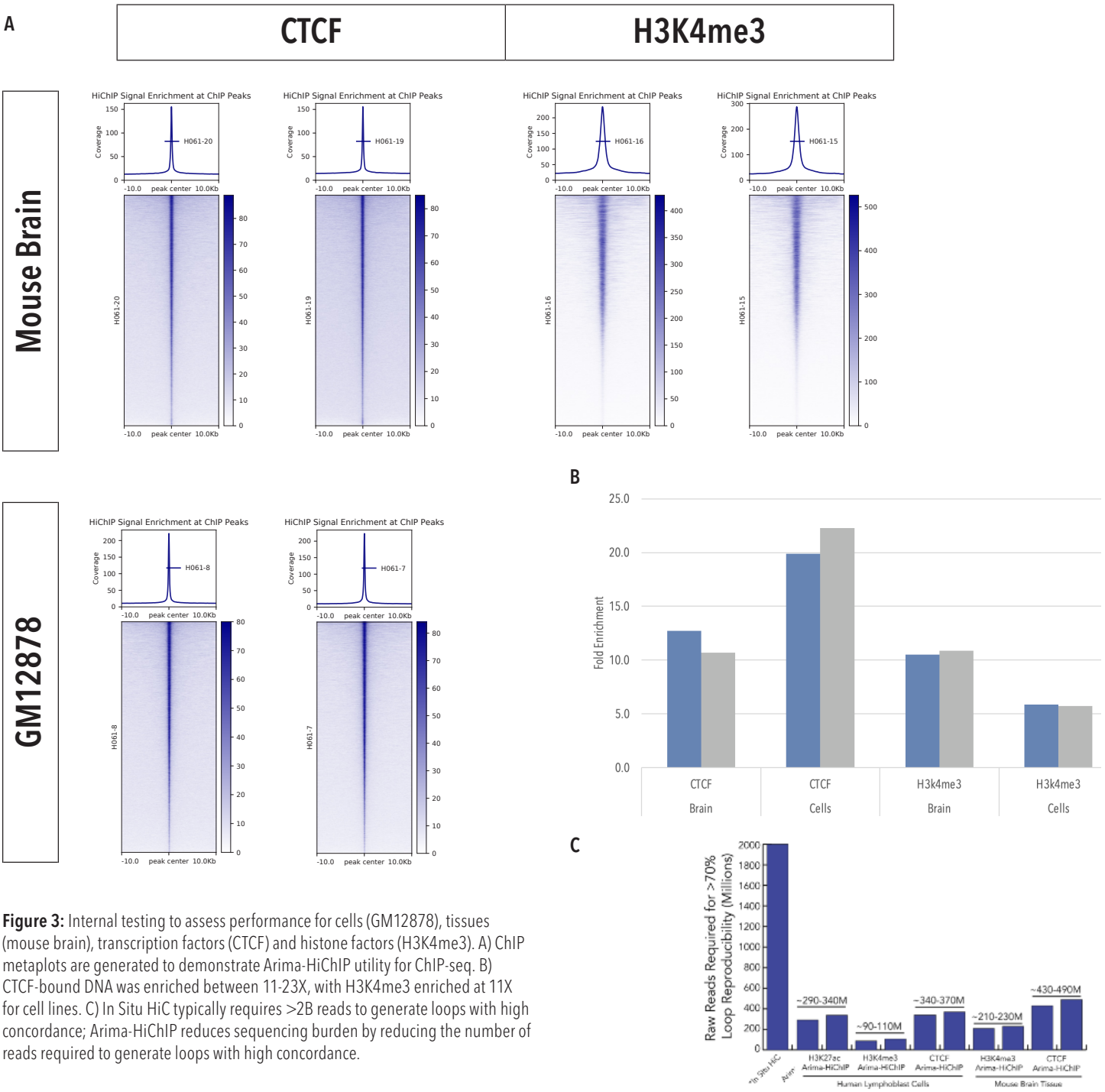


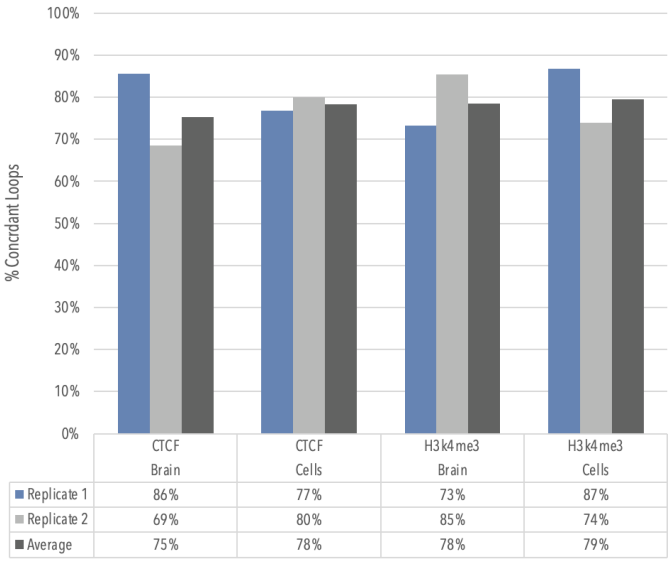
Figure 3: Internal testing to assess performance for cells (GM12878), tissues (mouse brain), transcription factors (CTCF) and histone factors (H3K4me3). A) ChIP metaplots are generated to demonstrate Arima-HiChIP utility for ChIP-seq. B) CTCF-bound DNA was enriched between 11-23X, with H3K4me3 enriched at 11X for cell lines. C) In Situ HiC typically requires >2B reads to generate loops with high concordance; Arima-HiChIP reduces sequencing burden by reducing the number of reads required to generate loops with high concordance.

3.2 Assay demonstrates strong reproducibility & sensitivity

Reproducibility is measured using "replicate loop concordance" between two identical samples. Loops are generated using Arima-MAPS 2.0 bioinformatics pipeline. This metric compares the presence of loop calls between two identical samples. For TF HiChIP in cell culture produced 77% average concordance vs, 75% in mouse brain tissue. For HF HiChIP, 78% and 79% average concordance was produced for mouse brain tissue and cell culture respectively.

Sensitivity is measured by comparing generated loops against those produced by Rao et al for GM12878 cell lines. >70% of CTCF Arima HiChIP.

A



	CTCF Brain	CTCF Cells	H3k4me3 Brain	H3k4me3 Cells
Replicate 1	86%	77%	73%	87%
Replicate 2	69%	80%	85%	74%
Average	75%	78%	78%	79%

B

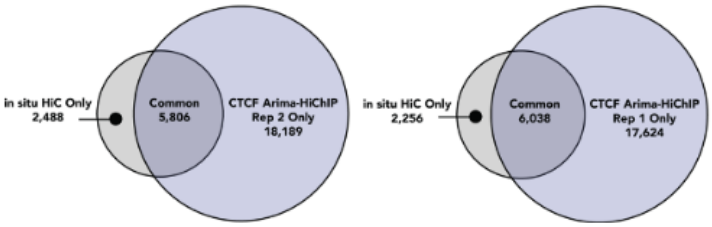


Figure 4: CTCF Arima HiChIP reproducibility and sensitivity. A) % concordance is calculated by comparing the number of loops generated by Replicate 1 that are also found in Replicate 2, and the number of loops generated in Replicate 2 that are also found in Replicate 1. For all samples and targets, average concordance is 75% or greater. B) CTCF Arima-HiChIP captured >70% of loops generated by genome-wide in-situ HiC data published by Rao et al.

4 Conclusions

4.1 Arima-HiChIP is a true multi-omics wet and dry lab workflow

The release of the Arima-HiChIP protocol in early 2020 introduced a major advancement in multi-omics technology with the introduction of a robust and reliable solution for a combined HiC (3D) + ChIP-seq (1D) assay workflow. However, analysis of the data output required researchers to either (1) recruit bioinformaticians to help with ChIP peak-calling, or (2) learn how to perform the analysis themselves.

The Arima-MAPS 2.0 bioinformatics pipeline expands upon the wetlab multi-omics solution by combining ChIP-peak calling, leveraging MACS2, with HiC loop calls. This enables researchers to perform a pre-validated, standardized bioinformatics workflow – resulting in reduction in time from sample-to-answer. With this solution in place, customers can publish sooner and more reliably than before.

4.2 Arima-HiChIP uncovers cell-type specific chromatin loops associated with active promoters.

To illustrate the utility of Arima-HiChIP data in exploring differential gene regulation across experimental samples, we identified a significantly different chromatin looping landscape associated with the active promoter of ITGB1BP1 (Fig.4). While this gene is expressed in both cell types as evidenced by the H3K4me3 peak at the promoter region and transcriptional signal across the gene body, the chromatin loops are largely skewed downstream in lymphoblasts towards a series of other active promoters (evidenced by H3K4me3) and putative enhancers (evidenced by H3K27ac ChIP signal but not H3K4me3). Strikingly, the chromatin loops are largely skewed upstream in IPSC-derived neurons, towards a broad range of H3K27ac signal and the active promoter of a neuron-specific gene, ASAP2. These interactions occur upstream, despite several other active promoters and putative enhancers downstream that are also observed in lymphoblast cells. This observation may be described as "enhancer-promoter switching". One possible mechanism facilitating this differential looping landscape could be the CTCF peak at the ITGB1BP1 gene found only in IPSC-derived neurons.

Further analysis of chromatin loops also exemplifies two additional modalities of long-range gene regulation. For example in the H3K4me3-associated loops anchored at ITGB1BP1, it is observed that some promoter-enhancer loops skip over genes, demonstrating the well-known observation that enhancers do not always regulate their nearest gene (REF). It is also observed that promoters significantly interact with other promoters. Taken together, the joint analysis of architectural protein occupancy, chromatin activity, transcription, and chromatin looping provide a more comprehensive view of dynamic gene regulatory mechanisms across cell types.

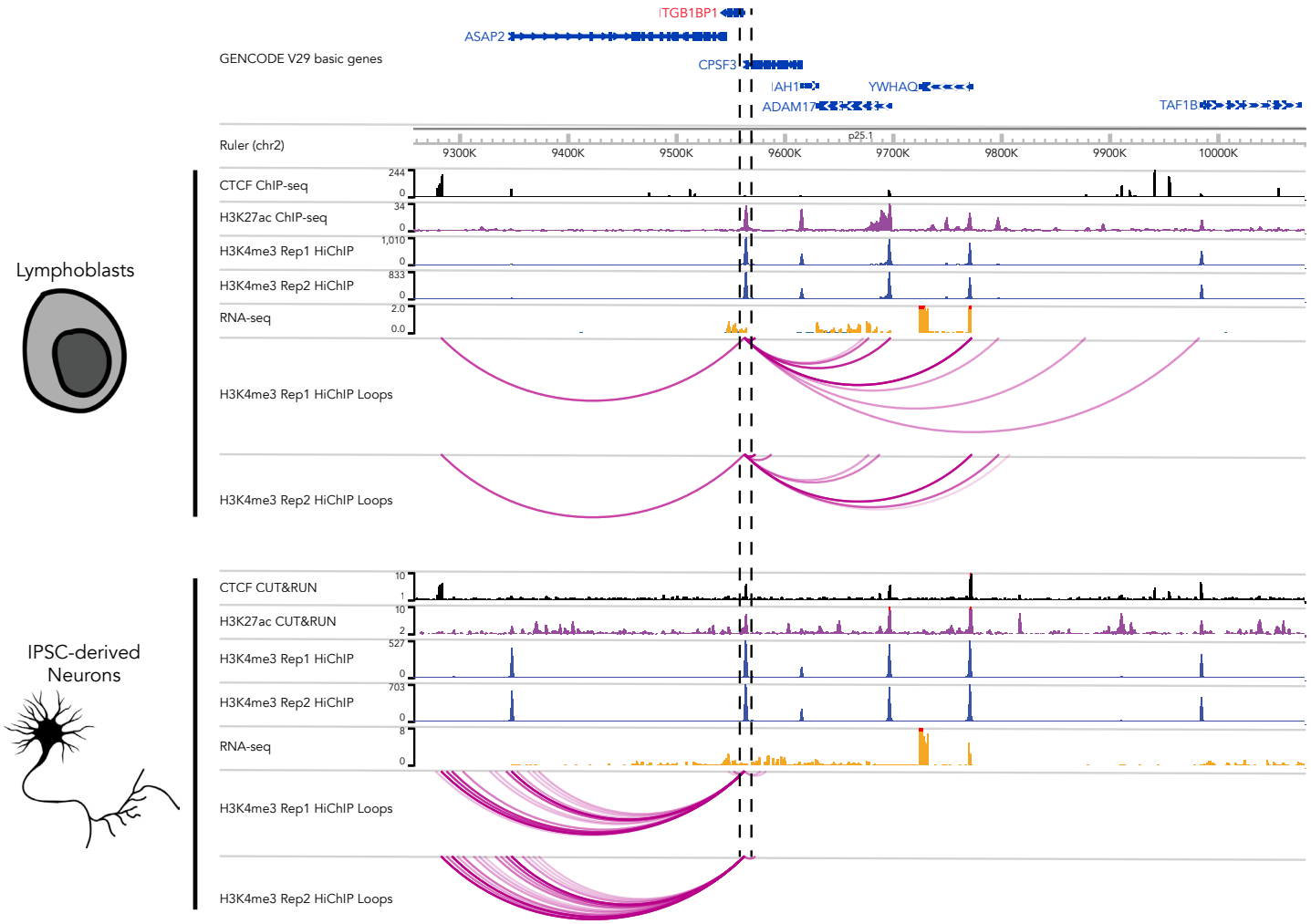


Figure 5. Arima-HiChIP uncovers cell-type specific loops associated with active promoters. WashU Epigenome Browser snapshot of the ITGB1BP1 locus in human lympho- blast cells (top half) and IPSC-derived neurons (bottom half). Shown for each cell type is CTCF occupancy, H3K27ac ChIP-seq, RNA-seq, H3K4me3 HiChIP 1D coverage, and H3K4me3 HiChIP loops. ChIP and RNA-seq data in lymphoblasts were obtained from ENCODE, and CUT&RUN and RNA-seq data in IPSC-derived neurons were obtained from the Yin Shen Lab (UCSF). Of note, only loops anchored at the ITGB1BP1 promoter are shown and other loops not associat- ed with ITGB1BP1 promoter region are masked for illustrative and comparative purposes. The darker the purple arcs, the more statistically significant the loop.

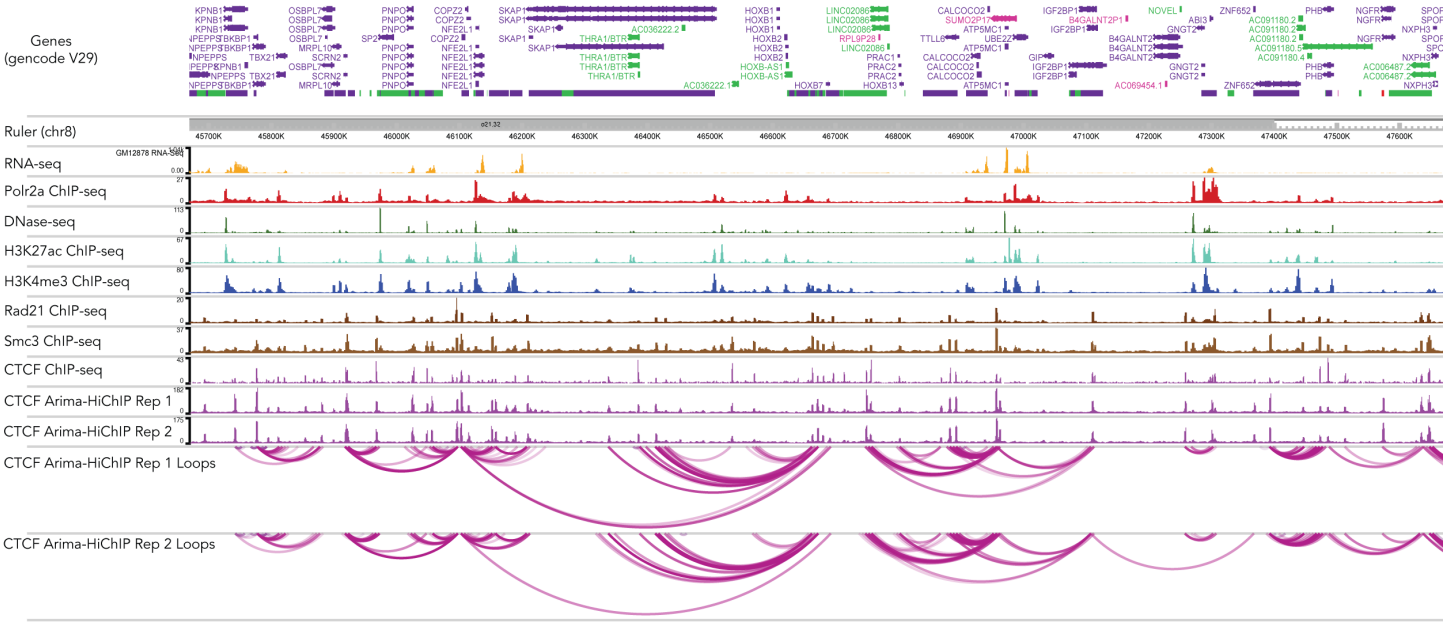


Figure 6. Arima-HiChIP uncovers long-range chromatin interactions associated with actively transcribed regions.r WashU Epigenome Browser snapshot of the HOXB locus in human lymphoblasts. Also shown is CTCF occupancy via standard ChIP-Seq, RNA-seq, CTCF HiChIP 1D coverage, and CTCF HiChIP loops. ChIP and RNA-seq data in lymphoblasts were obtained from ENCODE.

4.3 Ability to understand both transcription and histone factors

When Arima-HiChIP launched in early 2020, it revolutionized the study of gene regulation by enabling cost-effective, robust analysis of histone-factor bound chromatin and corresponding long-range implications. While histone factor Arima HiChIP is more suitable as input to chromatin segmentation models which seek to classify chromatin regions into functional categories, transcription factor Arima-HiChIP can interrogate proteins which are thought to associate with specific DNA sequences to influence the rate of transcription (cite Encode).

With the introduction of support for transcription factor HiChIP as validated through CTCF HiChIP, researchers can further expand upon epigenetic / gene regulation studies by enriching protein elements that are tied to transcription (and gene expression). In the latest protocol update, we demonstrate that Arima-HiChIP reliably reproduces 1D ChIP-seq peaks, but also illuminates corresponding chromatin structure (Figure 4).

5. Acknowledgements

We thank the NHGRI for generously funding the Arima-HiC kit development project 1R44HG009584-01.

6. References

1. https://github.com/ijuric/MAPS/tree/master/Arima_Genomics
2. Juric et al, 2019
3. Fang et al, 2016

ARIMA GENOMICS • 6404 Nancy Ridge Drive, San Diego, CA 92121
Office: 858.275.6463 • info@arimagenomics.com • arimagenomics.com