Identifying Functional Genomic Regions in iPSC Reprogramming with ChromHMM

UCLA

Gideon Shaked[1], Jingyuan Fu[2], Jason Ernst[2,3,4]

[1]BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA, [2]Department of Computer Science, UCLA, [3]Department of Biological Chemistry, UCLA, [4]Department of Computational Medicine, UCLA

Abstract

We used **ChromHMM** to map and analyze **chromatin states**—patterns of histone modifications that indicate **functional elements** and **regulatory activities** in the genome. By applying this to **ATAC-seq** data with **cell type clusters**, we identified key genomic regions that putatively play roles in the **iPSC reprogramming** process.

Introduction

Reprogramming human fibroblasts into **iPSCs** is a key process in **regenerative medicine**, involving major **epigenomic changes**. These alterations are essential for activating **pluripotency-associated genes** and repressing **lineage-specific genes**, enabling the transition from a differentiated state to a **pluripotent** one.

Methods

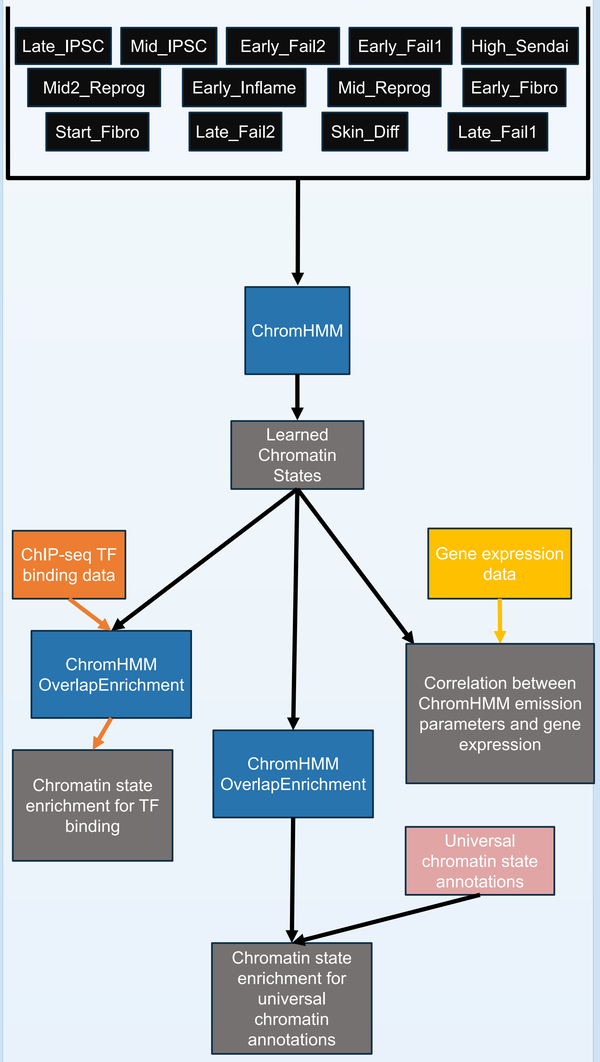

Figure 1. Overview of the methodology used for chromatin state analysis in this study. The diagram illustrates the workflow starting from the ATAC-seq data (black), followed by the usage of ChromHMM (blue) with public data (orange, yellow, pink) and the resulting analysis (brown).

Results

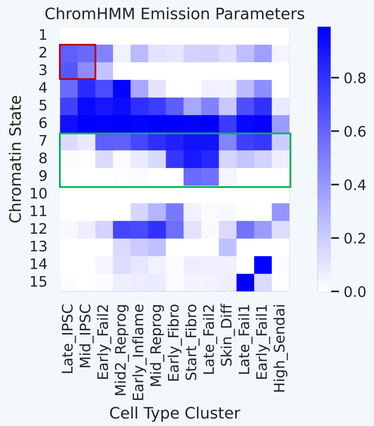

Figure 2. ChromHMM emission parameters. ChromHMM states derived from segmentation of ATAC signal, based on cluster cell groups. Each square represents the probability of the relevant ATAC-seq cluster occurring in the relevant state.

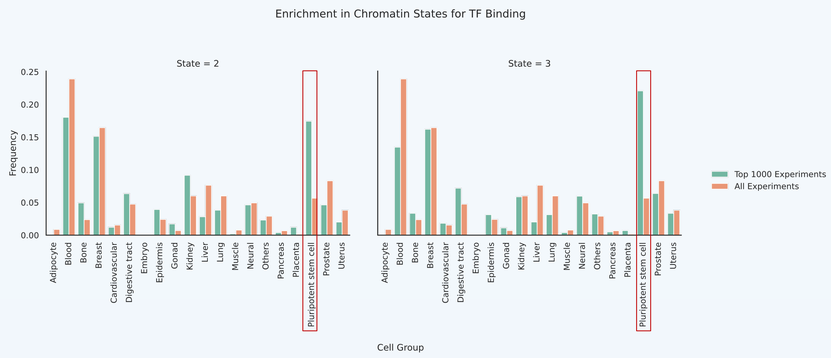

Figure 3. Enrichment of cell type-specific transcription factors in states 2 and 3. Transcription factor (TF) tracks were downloaded from ChIP-Atlas and experiments with <1,000 peaks were filtered out. For states 2 and 3, the top 1,000 experiments were disproportionally performed with pluripotent stem cells.

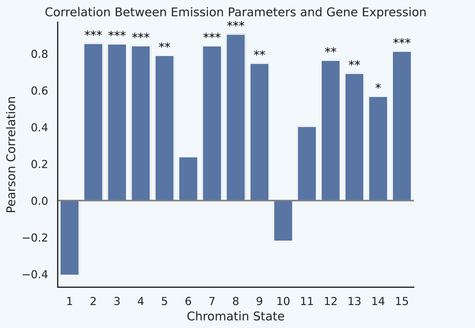

Figure 4. Pearson correlation between ChromHMM emission parameters and average total gene expression levels in each state. One asterisk (*) indicates p < 0.05, two asterisks (**) indicate p < 0.01, and three asterisks (***) indicate p < 0.001.

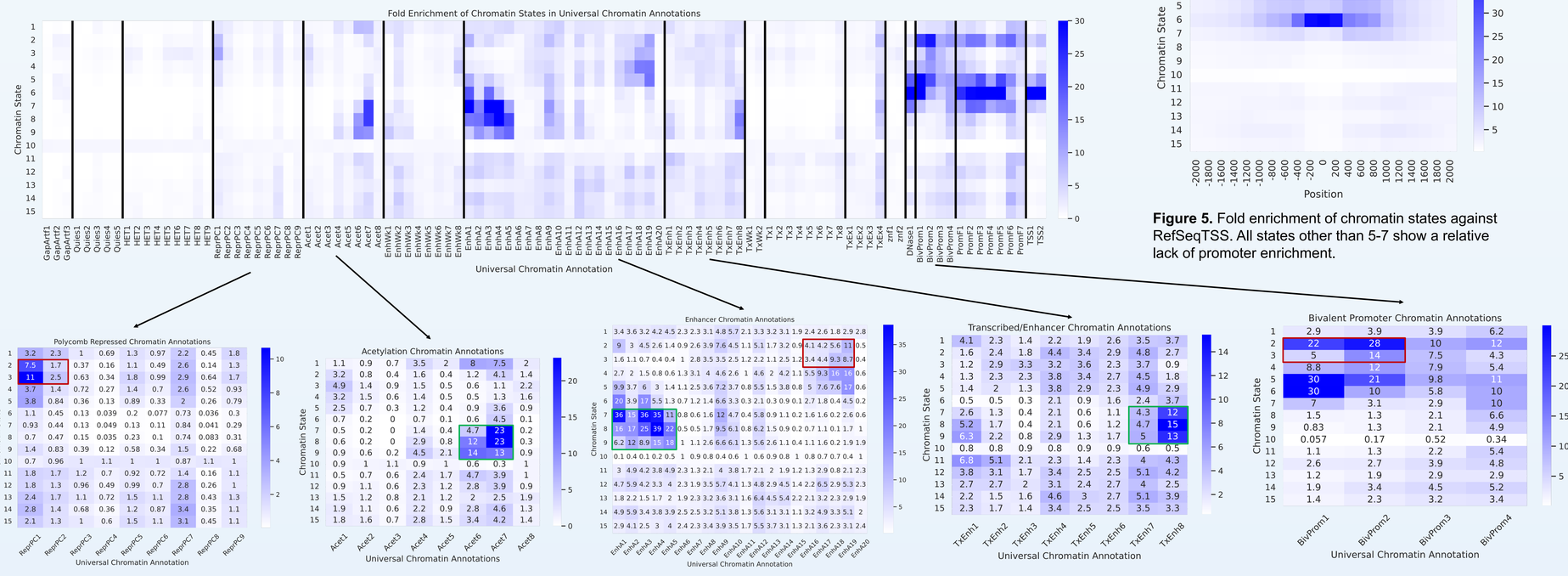

Figure 5. Fold enrichment of chromatin states against RefSeqTSS. All states other than 5-7 show a relative lack of promoter enrichment.

Figure 6. Fold enrichment of chromatin states against universal chromatin state annotations (Vu and Ernst 2022). Universal chromatin state annotations are epigenomic patterns with similar functional significance in most cell types. Red boxes indicate fold enrichments that suggest that the boxed states may be relevant to the maintenance of pluripotency in iPSCs. Green boxes indicate fold enrichments that suggest that the boxed states may contain enhancers that cause a progressive blockade of iPSC reprogramming at multiple stages. For the top row heatmap, fold enrichments larger than 30 are shown as 30.

Conclusion

- **Bivalent Chromatin and Polycomb Repressed Regions:** Our analysis revealed that there are distinct regions enriched for both bivalent promoters and polycomb repressed regions, indicating that these regions are likely involved in maintaining pluripotency in iPSCs.
- **Potential Enhancer Activity in Maintaining Pluripotency:** The data also suggest that those same regions are enriched for cell type-specific transcription factors, which putatively indicates that there are enhancers in these regions that are also involved in maintaining pluripotency in iPSCs.
- **Enhancer States and Reprogramming Blockade:** The data indicate that several putative enhancer states may be responsible for progressively blockading iPSC reprogramming at different stages.

References

Ernst, Jason, and Manolis Kellis. 2010. “Discovery and Characterization of Chromatin States for Systematic Annotation of the Human Genome.” *Nature Biotechnology* 28 (8): 817–25. https://doi.org/10.1038/nbt.1662.

Ernst, Jason, and Manolis Kellis. 2017. “Chromatin-State Discovery and Genome Annotation with ChromHMM.” *Nature Protocols* 12 (12): 2478–92. https://doi.org/10.1038/nprot.2017.124.

Vu, Ha, and Jason Ernst. 2022. “Universal Annotation of the Human Genome through Integration of over a Thousand Epigenomic Datasets.” *Genome Biology* 23 (1): 9. https://doi.org/10.1186/s13059-021-02572-z.

Zou, Zhaonan, Tazro Ohta, and Shinya Oki. 2024. “ChIP-Atlas 3.0: A Data-Mining Suite to Explore Chromosome Architecture Together with Large-Scale Regulome Data.” *Nucleic Acids Research* 52 (W1): W45–53. https://doi.org/10.1093/nar/gkae358.

Acknowledgements

The authors would like to thank B.I.G. Summer for funding this research. I would also like to thank Justin Langerman for his extensive advice on the nature of the iPSC reprogramming process. This work used computational and storage services associated with the Hoffman2 Cluster which is operated by the UCLA Office of Advanced Research Computing’s Research Technology Group.