Using Machine Learning to Predict RNA Pol II Interactions of Metastatic Prostate Cancer

Ahmed Abbas[1], Chengcheng Liu[2], Michael Q. Zhang[2], Ram S. Mani[1,3,4]

[1]Department of Pathology, UT Southwestern Medical Center, Dallas, TX 75390, USA

[2]Department of Biological Sciences, Center for Systems Biology, The University of Texas at Dallas, Richardson, TX 75080, USA

[3]Department of Urology, UT Southwestern Medical Center, Dallas, TX 75390, USA

[4]Harold C. Simmons Comprehensive Cancer Center, UT Southwestern Medical Center, Dallas, TX 75390, USA

Background

The three-dimensional (3D) genome organization directly impacts diverse nuclear processes such as transcription, DNA repair, and replication. Therefore, it is crucial to understand how and which distal regulatory element (in the linear genomic distance) interacts with its target gene in the 3D space. There are several experimental methods used to measure the 3D genome interactions, like Hi-C (1), ChIA-PET (2), and others. However, they are costly, time-consuming, and, more importantly, they require millions of cells to perform the experiment, which makes them unsuitable for use with patient biopsies. Thus, it is essential to introduce the machine learning techniques to help solve this problem.

Methods

Performing RNA-seq and ATAC-seq experiments requires a smaller number of cells than those needed for ChIP-seq experiments (3). Thus, they are more suitable for use with human biopsies. In our model, we use normalized ATAC-seq read count and gene expression FPKM values, in addition to genomic distance, to predict the interaction strength between two distal ATAC-seq peaks with at least one of them overlapping with a gene. We trained a gradient boosting regression model on the GM12878 ENCODE cell line and did many experiments to validate the quality of the model.

Results

To measure the performance of our model, we predicted the RNA Pol II ChIA-PET interactions of the HepG2 and K562 cell lines. We obtained high Pearson correlation values between original and predicted interactions. Then, we predicted the interactions between AR+ and NE+ upregulated genes (4, 5) and the ATAC-seq peaks within the 1 Mbp neighborhood of the genes' TSS. We found that the interactions between AR+ upregulated genes and their enhancers are significantly stronger than those between the NE+ upregulated genes and their enhancers in the AR+NE- mCRPC samples. Similarly, we found the interactions overlapping with the NE+ upregulated genes stronger in the AR-NE+ samples. We also calculated super-enhancers (SEs) from mCRPC PDX H3K27ac ChIP-seq samples (6) and identified genes with which they interact.

Conclusion

We proposed a method for predicting RNA Pol II-associated interactions between genes and their distal regulatory regions. Our method showed good performance in predicting the interaction strength of different cell lines. In addition, it performed well in predicting interactions of mCRPC patient samples.

Acknowledgements

This research was supported in part by the computational resources provided by the BioHPC supercomputing facility located in the Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center. The results described here are in part based on data generated by the SU2C/PCF West Coast Dream

Team Project (EGAD00001008991, EGAD00001008487, and EGAD00001009065). AA acknowledges funding support from Prostate Cancer Foundation Young Investigator Award. RSM acknowledges the funding support from National Cancer Institute (NCI)/NIH grants (R01CA245294, R01CA295997), CPRIT Individual Investigator Research Award (RP230382), CPRIT High-Impact/High-Risk Research Award (RP240516), and US Department of Defense Breakthrough Award (W81XWH-21-1-0114).

Conflicts of Interest
No

References

1. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. science. 2009;326(5950):289-93.
2. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, et al. An oestrogen-receptor-α-bound human chromatin interactome. Nature. 2009;462(7269):58-64.
3. Xu D, Forbes AN, Cohen S, Palladino A, Karadimitriou T, Khurana E. Recapitulation of patient-specific 3D chromatin conformation using machine learning. Cell reports methods. 2023;3(9).
4. Lundberg A, Zhang M, Aggarwal R, Li H, Zhang L, Foye A, et al. The genomic and epigenomic landscape of double-negative metastatic prostate cancer. Cancer research. 2023;83(16):2763-74.
5. Shrestha R, Chesner LN, Zhang M, Zhou S, Foye A, Lundberg A, et al. An atlas of accessible chromatin in advanced prostate cancer reveals the epigenetic evolution during tumor progression. Cancer Research. 2024;84(18):3086-100.
6. Pomerantz MM, Qiu X, Zhu Y, Takeda DY, Pan W, Baca SC, et al. Prostate cancer reactivates developmental epigenomic programs during metastatic progression. Nature genetics. 2020;52(8):790-9.