Creating an Epigenomic Map of the Heart

Giulio Duregon, Joby George, Jonah Poczobutt

NYU Center For Data Science

Introduction

Chromosomes contain genetic information in tightly coiled DNA, called chromatin. Most chromatin is so tightly coiled that it is inaccessible to transcription factors that affect gene expression. These transcription factors bind to Cis-Regulatory Elements (CREs). CREs are short DNA sequences that serve as binding sites for transcription factors if they are accessible (chromatin is open). Therefore, open chromatin regions influence gene expression.

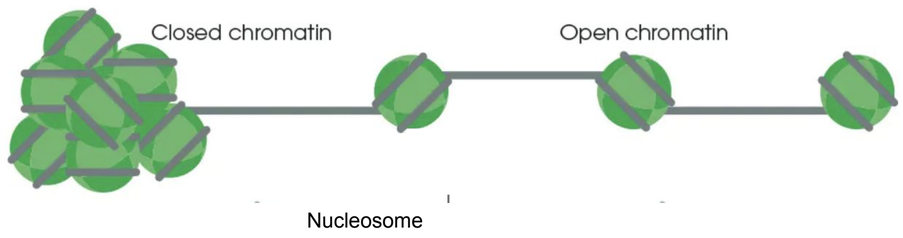

Figure 1: *Chromosomal Structure*

Differential accessibility is a key concept in genetic links to diseases. Differential accessibility is when an open chromatin region is present in one group of samples in a specific location of the genome, and not present in another group of samples.

Our research goal was to create an epigenomic map of the human heart, identifying a universe of CREs and those which are differentially accessible between conditions.

Data & Related Work

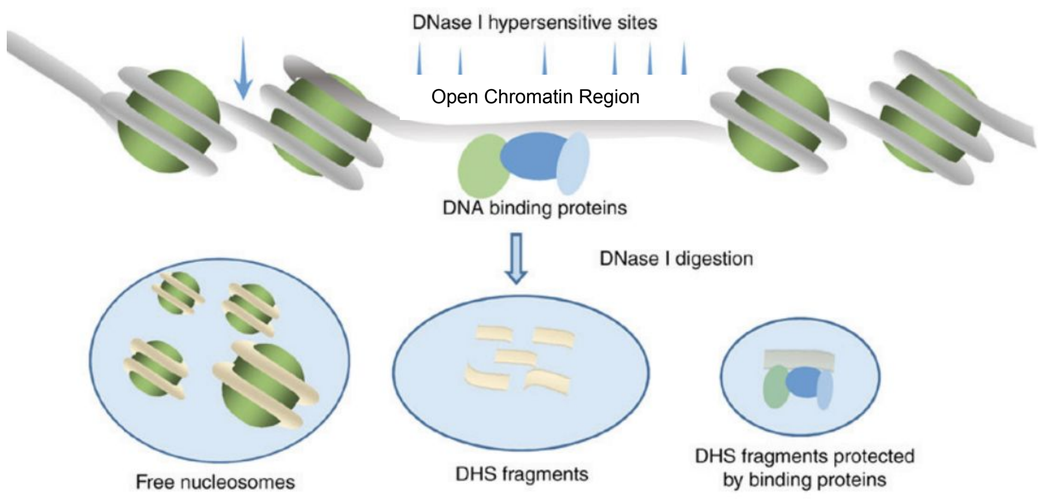

Figure 2: *DNAse-Seq data gathering technique*

The DNAse-Seq technique, developed in 2008, greatly improved researchers ability to identify open chromatin regions and their associated CREs.

All our Data come from the ENCODE project and include high throughput RNA-seq and DNase-seq data. These data include sequences and coordinates of where the given sequence aligns to a reference genome (HG38). Our dataset was comprised of 14 biological samples, each coming from an adult male or female, and from the atrium or ventricle tissue.

Results

Our work led to the identification of over 330k DNase Hypersensitive Sites, that can be thought of as a functional universe for examining differences in chromatin accessibility for the human heart. We identified 21k of these sites as being differentially accessible between males and females in our samples. Genes identified as differentially expressed are much more likely to have at least one differentially accessible DHS in close proximity (50% for males, 33% in females) than the general population of genes. (16% for males, 20% for females)

Condition	DA Regions Indentified
Male V Female	6,426
Atrium V Ventricle	13,672
M. Ventricle V F. Ventricle	21,830

Figure 5: *Analysis between conditions results*

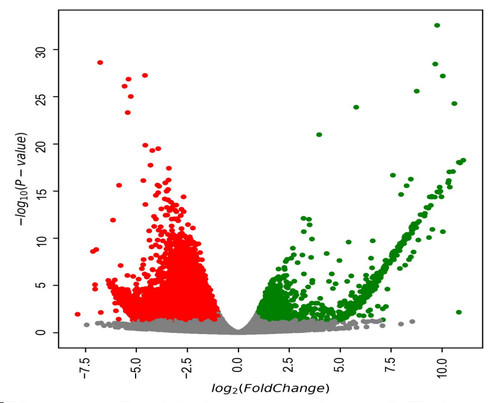

Figure 6: *Volcano plot of DA regions between Male and Female*

Conclusion

Better understanding the role of CREs in the human genome and discrepancies in the accessibility of these elements between groups is vital for better understanding genetic factors contributing for heart disease. Further work building on our results would likely incorporate known relationships between certain genes and their associated phenotypic relationships. One could also use our consensus map to correlate differential accessibility with certain health outcomes between individuals or groups, illuminating relationships between CREs and health outcome data.

Acknowledgements

We would like to thank our project mentors, Drs. Aravinda Chakravarti, and Dongwon Lee for their guidance in our research process. Their experience with functional genetic mapping was immensely helpful when presenting intermediate analyses. Additionally, we would like to thank the Capstone Project supervisors for their support and providing this exciting research opportunity.

References

1. *Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. 2008 High-resolution mapping and characterization of open chromatin across the genome. Cell 132(2):311–322*
2. *Chen Y, Chen A Unveiling the gene regulatory landscape in diseases through the identification of DNase I hypersensitive sites Biomedical Reports, 11(3): 87-97*
3. *I. Dunham, et al. 2012. An integrated encyclopedia of DNA elements in the human genome Nature, 489(7414):57–74*
4. *Yang Liao, Gordon K Smyth, and Wei Shi. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Research, 47(8):e47–e47*
5. *M.I. Love, W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. Genome Biol, 15(550)*
6. *Pugh BF Zhang Z. 2011. High-resolution genome-wide mapping of the primary structure of chromatin Cell, 144(2):175–186*

Methodology

Figure 3: *Methodology workflow. Data is acquired through REST API requests to Encode, and processed in two separate workflows, one for DNase datasets, and one for RNA datasets.*

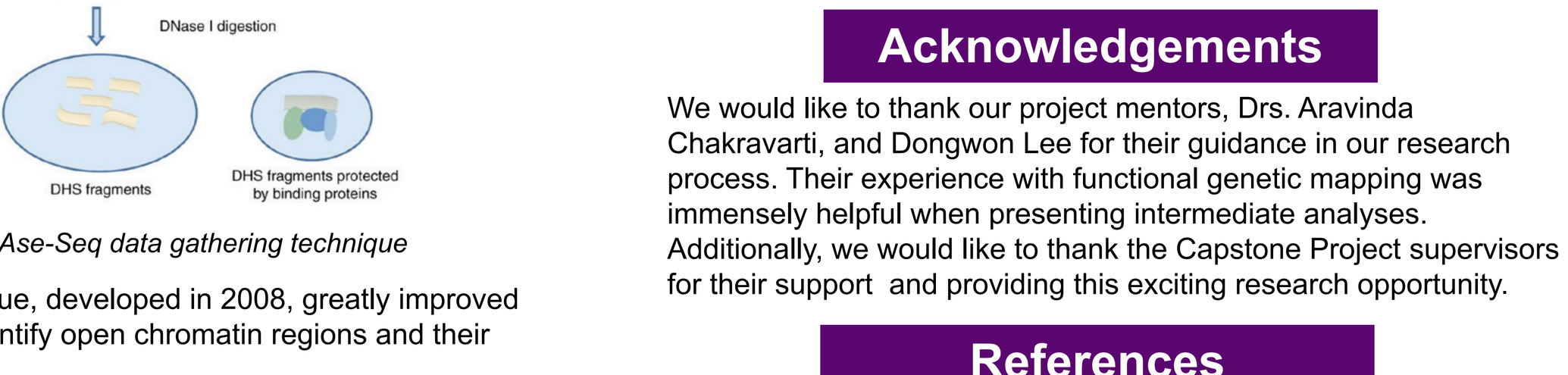

Figure 4: *Consensus Map Demonstration with Orange highlighting to show a differentially accessible peak in Male Samples.*