DEEP LEARNING BASED MULTI-VIEW MODEL FOR DECIPHERING GENETIC REGULATORY KEYWORDS

Pramod Bharadwaj Chandrashekar, Li Liu
pbchandr@asu.edu, liliu@asu.edu
Biomedical Informatics, College of Health Solutions, Arizona State University, Scottsdale, AZ;

INTRODUCTION

Gene regulation involves controlling of gene expressions. As they are key for various traits and diseases, they are used in clinical setting as prognostic, diagnostic, and therapeutic markers. Factors affecting gene regulation involve DNA sequence of the gene, transcription factors binding to the genes, histone modifications, etc. Identifying the regulatory codes and its effect on gene expression is a huge challenge in precision medicine and requires computational approaches.

OBJECTIVES

Develop a novel deep learning method to not only predict gene expression levels but also identify the regulatory regions which contributes towards regulation.

METHODS

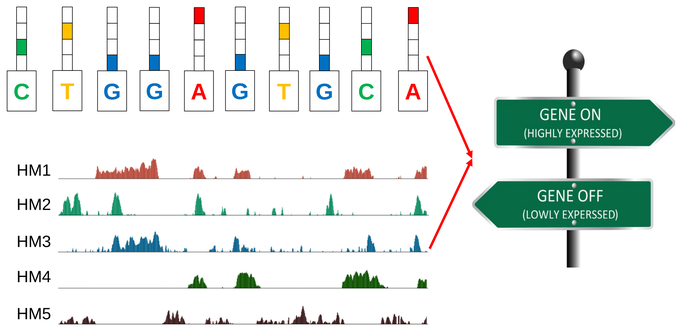

Fig 1: Problem Formulation

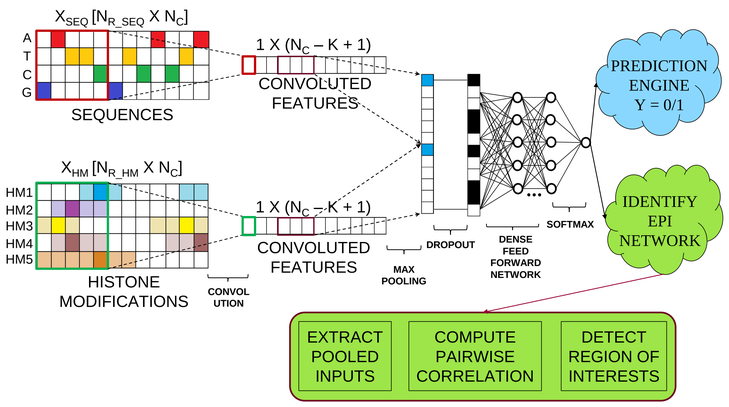

Fig 2: Architecture of deep multi-view CNN model

Multi-view deep learning model can **accurately classify gene expression** using genetic and epigenetic signals.

The deep learning model can also **accurately identify regions with high histone modification activity** which we hypothesize to be enhancer and promoter regions.

Take a picture to download poster and slides

RESULTS

Gene sequences and five histone modification tracks (H3K4me1, H3K4me3, H3K9me3, H3K27me3, and H3K36me3) for 18779 genes belonging to lung tissue were analyzed. Median expression level was used to separate the RPKM values into two classes.

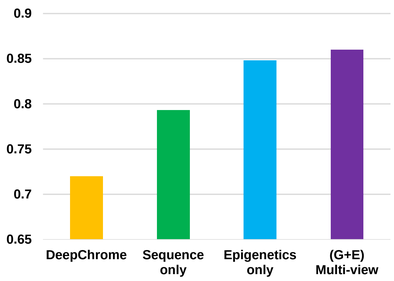

Fig 3: AUC score comparison of various models

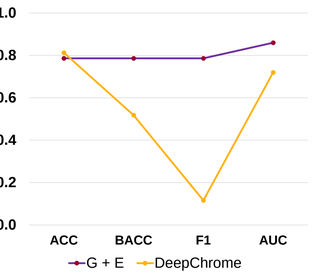

Fig 4: Performance of our model with DeepChrome

Histone modifications	Low Expression Genes		High Expression Genes	
	Enhancers	Promoters	Enhancers	Promoters
H3K9me3	**38**	**39**	0	0
H3K27ac	2	1	**88**	**90**
H3K36me3	**40**	**41**	10	10
H3K4me1	9	8	2	0
H3K27me3	11	9	0	0

Table 1: Comparison of HM activity among high vs low gene expression as identified by our model

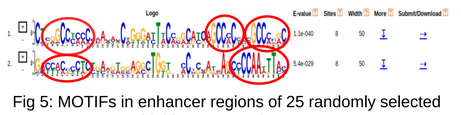

Fig 5: MOTIFs in enhancer regions of 25 randomly selected highly expressed genes

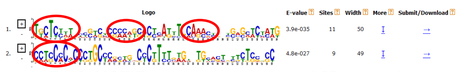

Fig 6: MOTIFs in promoter regions of 25 randomly selected highly expressed genes

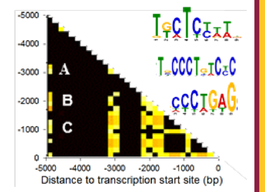

Fig 7: 3 enhancer-promoter pairs found by decomposing CNN model of FTL gene

CONCLUSION AND DISCUSSION

- Our novel method has the best performance in predicting gene expression compared to al existing methods
- We observe that even though combining gene sequences and histone signals produce the best result, histone signals contribute in majority towards the prediction.
- We also show that our method can be used to identify enhancer-promoter interaction pairs which has the highest histone activity.
- We intend to apply our model to other tissues and discover unknown regulatory mechanisms.