Dec 2023

Building the Gene Regulatory Network using Deep Learning

Ashwini Suriyaprakash

Advised by Lei Xiong, Computational Biology Group (Prof. Manolis Kellis)

Eric and Wendy Schmidt Center Funded Research and Innovation Scholar

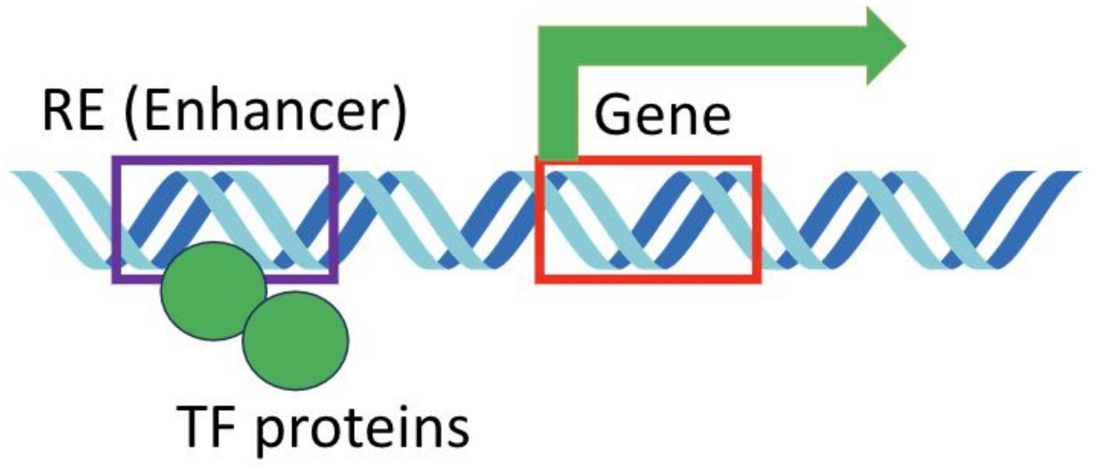
MIT EECS

SuperUROP

Advanced Undergraduate Research Opportunities Program

Introduction

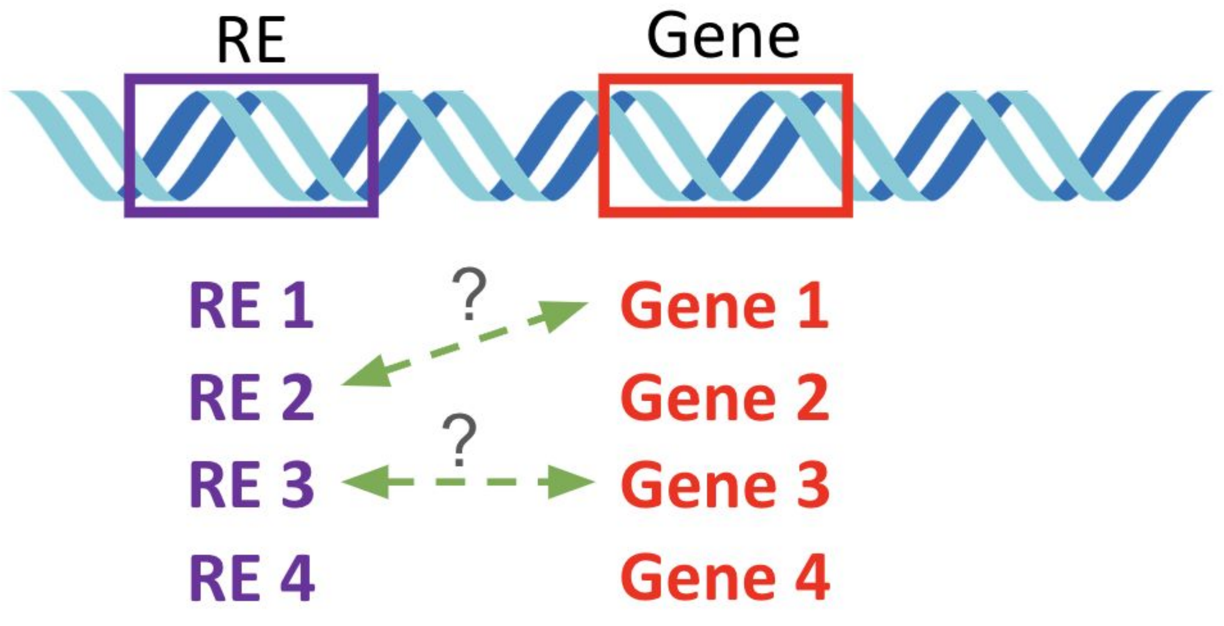
Gene regulatory networks (GRNs) consist of genes, regulatory elements (REs), and transcription factor (TF) proteins



Understanding GRNs is important for investigating disease origins and treatment

- Most disease-related mutations occur in non-coding regions (eg. REs such as enhancers) and cause target genes to be expressed in an incorrect amount
- Can use drugs/gene-editing techniques to correct enhancer behavior

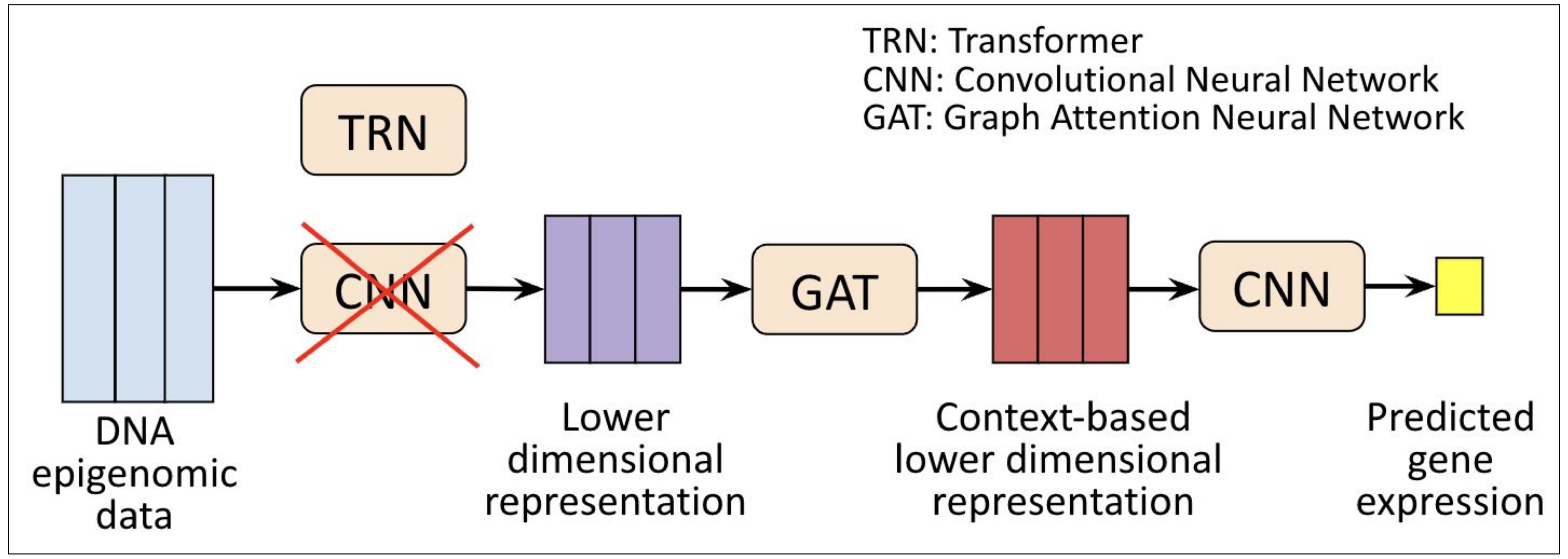
Problem: Many interactions between REs and genes still remain unknown



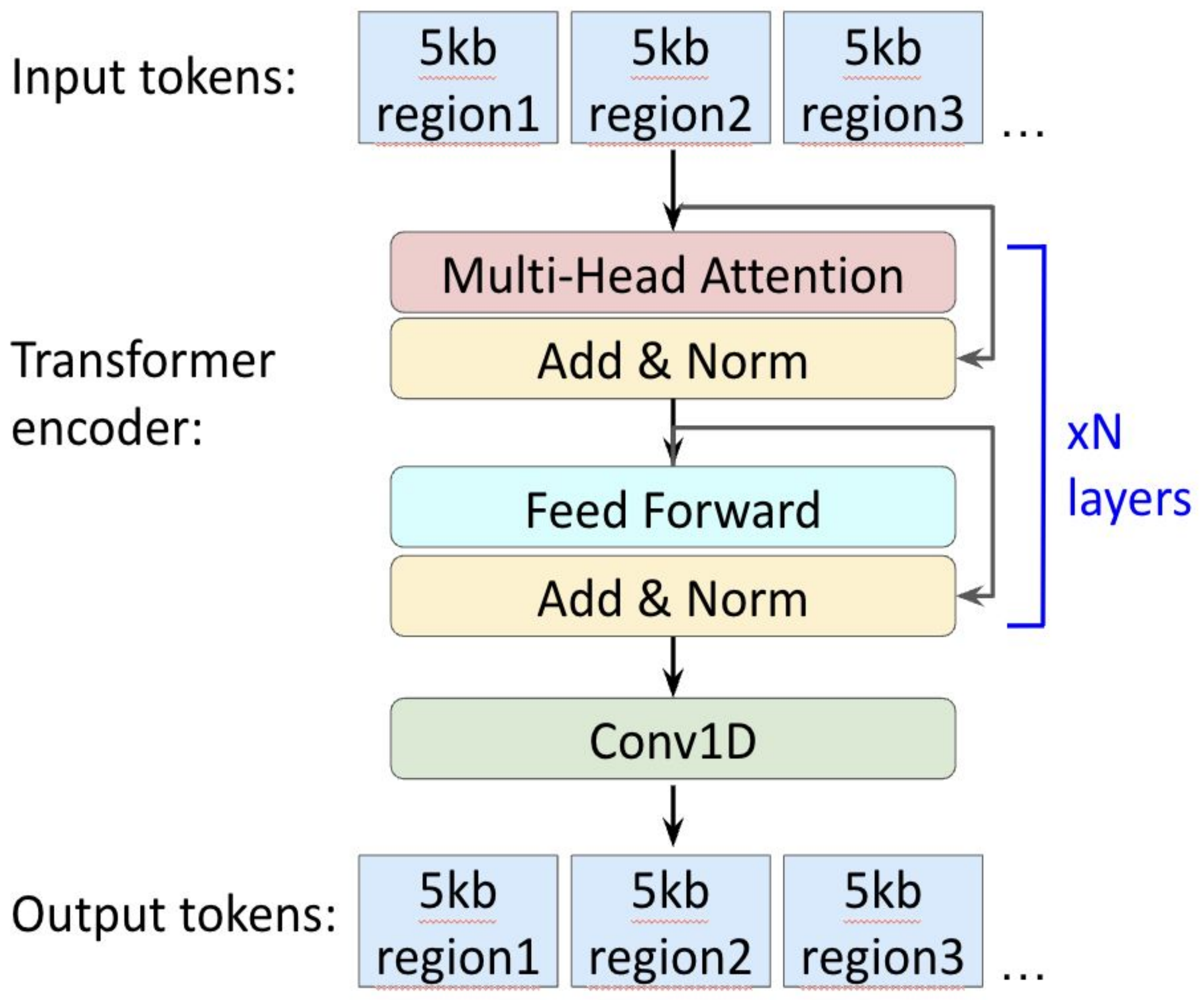
Goal: Discover interactions using deep-learning

Methods

Approach: Improve prediction accuracy of RE-Gene interactions by employing Transformer in existing GraphReg method (Alireza, et. al)



Transformer Architecture

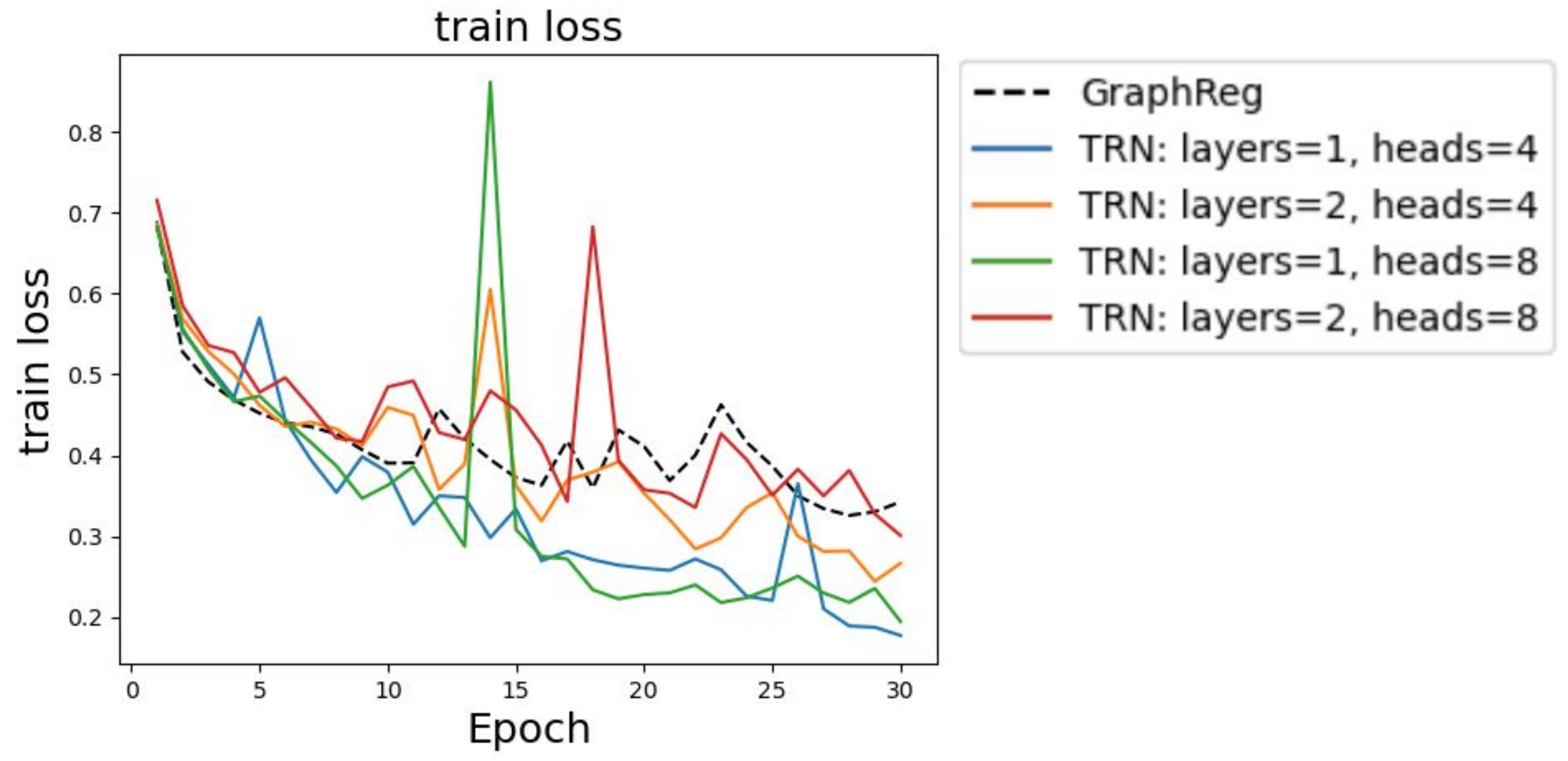


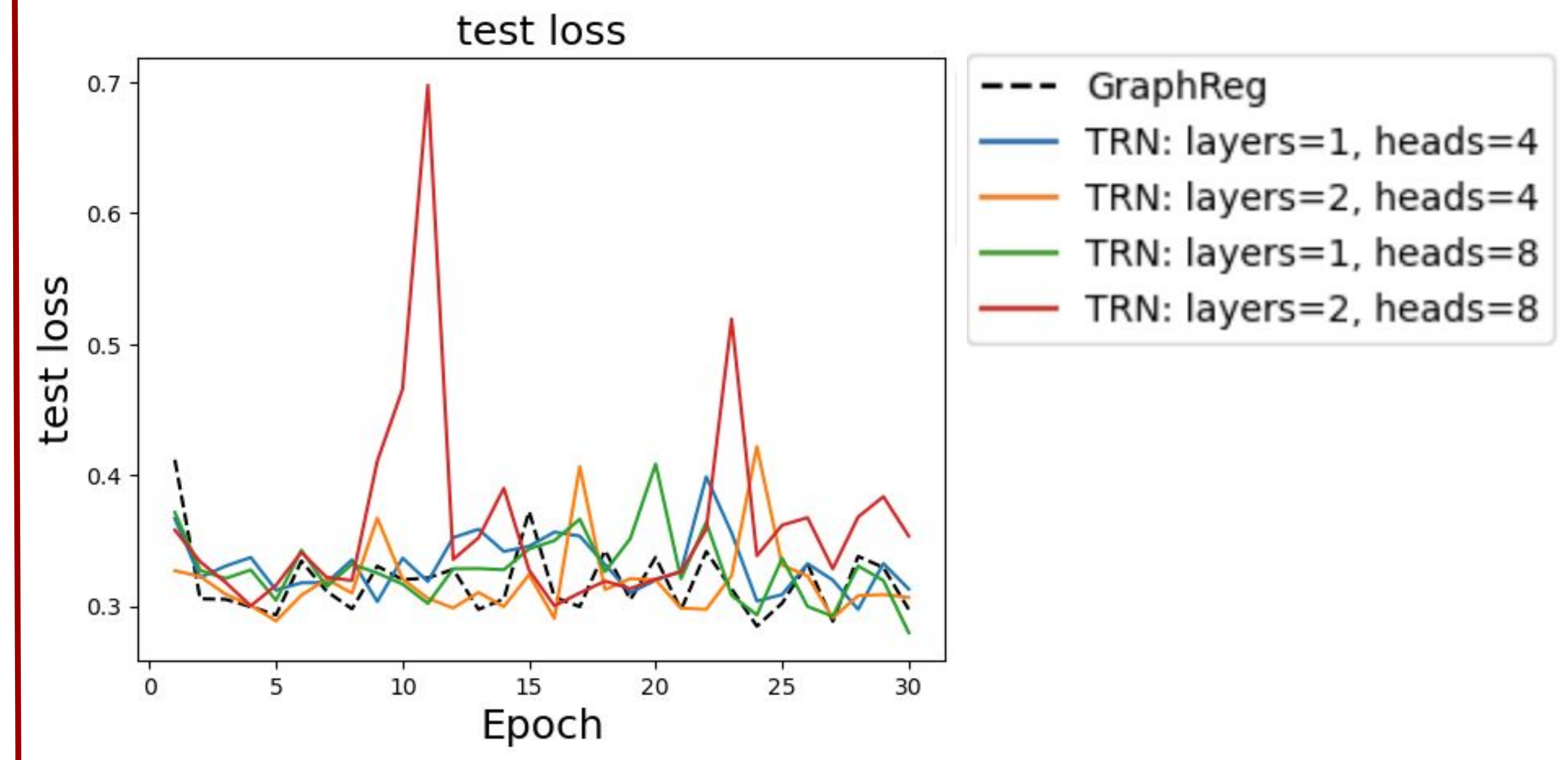
Self-Attention Matrix (example)

Unlike CNNs which learn relationships between a genomic region and its direct neighbors, Transformers learn relationships between a region and many other regions upstream and downstream by computing scores in the self-attention matrix below.

	5kb region1	5kb region2	5kb region3	…
5kb region1	0.71	0.04	0.07	
5kb region2	0.1	0.65	0.01	
5kb region3	0.2	0.01	0.6	
⋮				

Results





- For training, Transformer performs better compared to GraphReg, while testing losses are comparable
- Interestingly, Transformer's training performance decreases as # of attention layers/heads increases

Next Steps

1. Add positional encoding to the Transformer
2. Use scores in self-attention matrix to link REs with genes and compare results to GraphReg
3. Augment input data characterizing a genomic region (eg. ATAC-Seq data)