Graph-Based Neural Networks for Modeling Long-Range Epigenetic Regulation

Xavi Loinaz (advised by Prof. Ritambhara Singh)

Agenda

1 - Problem Motivation

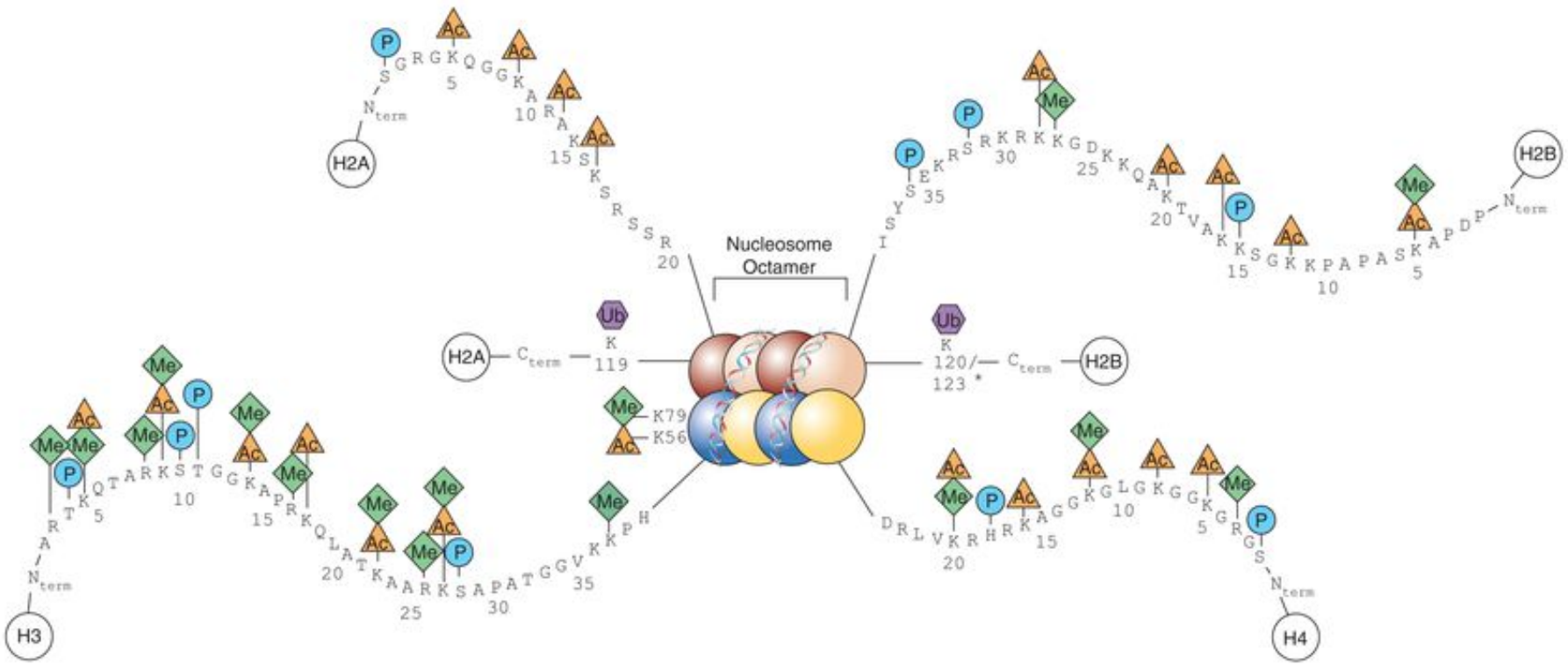
2 - Method: GC-MERGE

3 - Method: XL-MERGE

4 - Discussion

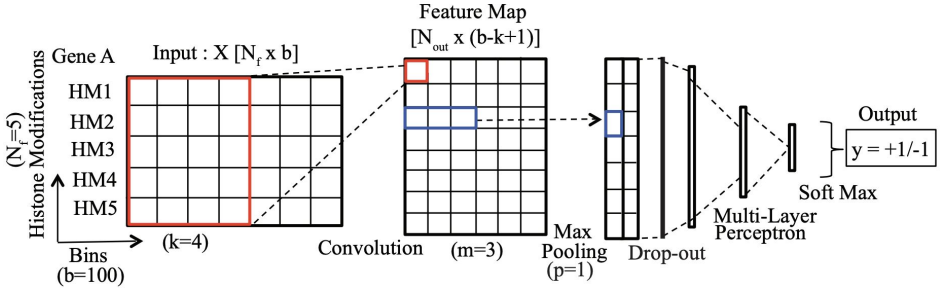
Problem Motivation

Histone Marks

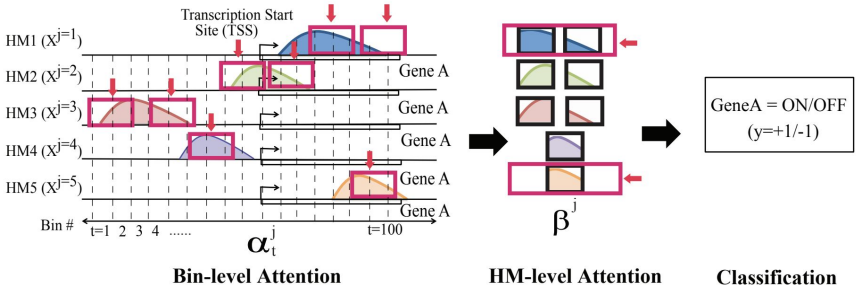


Prior Methods in Deep Learning

Histone marks → Gene expression prediction



DeepChrome (Singh et al.)



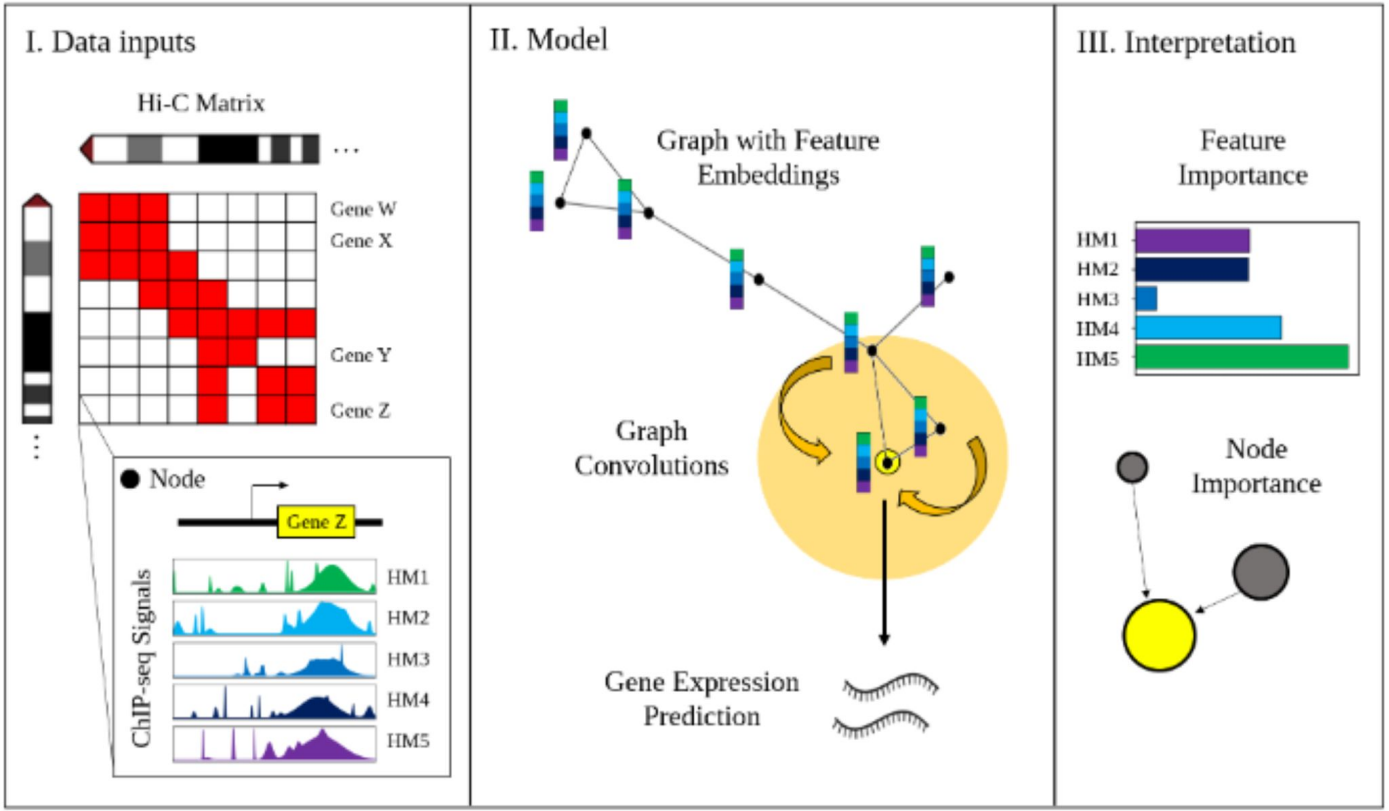
AttentiveChrome (Singh et al.)

But what about long-range regulation effects?

Method: GC-MERGE

Graph **C**onvolutional **M**odel of **E**pigentic **R**egulation of **G**ene **E**xpression (**GC-MERGE**)

Overview


I. Data inputs
Hi-C Matrix
Gene W
Gene X
Gene Y
Gene Z
Node
Gene Z
ChIP-seq Signals
HM1
HM2
HM3
HM4
HM5
II. Model
Graph with Feature Embeddings
Graph Convolutions
Gene Expression Prediction
III. Interpretation
Feature Importance
HM1
HM2
HM3
HM4
HM5
Node Importance

Results and Limitations

3 Results, 3 Limitations

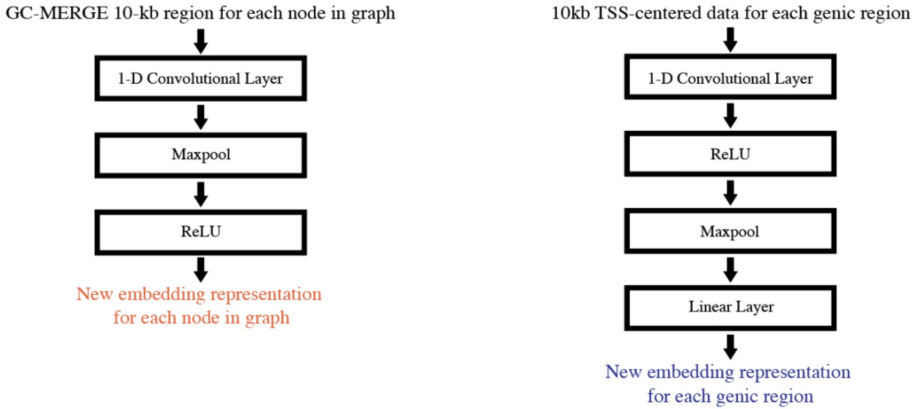
- Binary prediction: AUROC from 0.88 to 0.91
  - Still outperformed by best models
- Some evidence that long-range interactions help predictions
  - Shuffling long-range neighbors intrachromosomally does not worsen performance
- Some proof-of-concept evidence to validate known long-range regulations
  - Interpretation results have high variance

Method: XL-MERGE

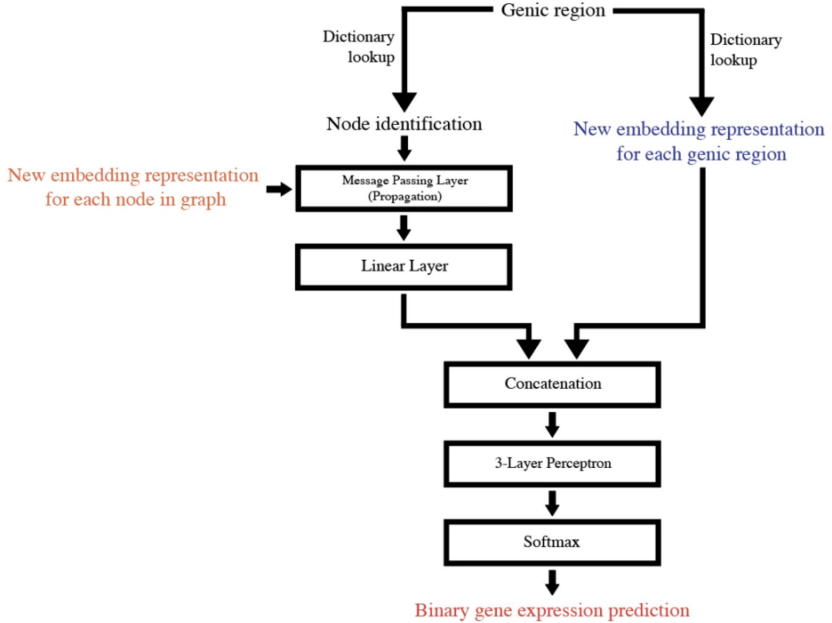
Xavier **L**oinaz's **M**odel of **E**pigentic **R**egulation of **G**ene **E**xpression (**XL-MERGE**)

Architecture

Pre-embedding Formulation in XL-MERGE



Gene Expression Prediction Formulation in XL-MERGE



Results and Improvements

3 Results, 3 Improvements

- Binary prediction: AUROC from 0.91 to 0.932
  - Ties or outperforms best models
- Stronger evidence that long-range interactions contribute to predictions
  - Ablation analysis for shuffled neighbors shows great difference in prediction power
- No interpretation results yet
  - Model gives simpler, more realistic possibilities for interpretation

Discussion

Conclusions/Further Directions

Conclusions

- Evidence for effectively capturing long-range interactions in model
- Potential for simpler tractable interpretation to be performed on model

Further Directions

- Figure out interpretation angle
- Test more mechanisms of extracting long-range regulation information
- Train different weights of model at different rates