

Attend and Predict: Understanding Gene Regulation by Selective Attention on Chromatin

Tymor Hamamsy, Tristan McRae, Mamie Wang, Timothy Wu

Attend and predict: understanding gene regulation by selective attention on chromatin

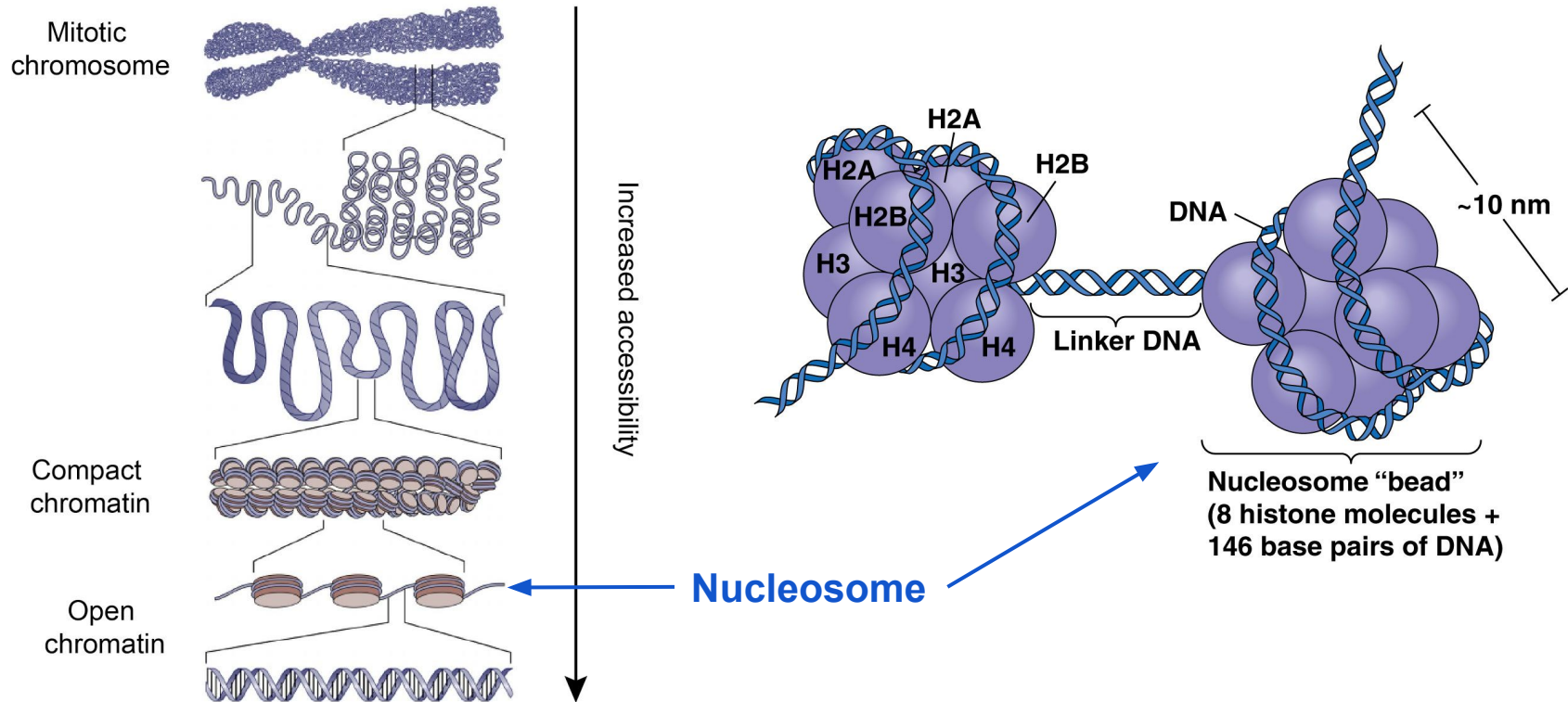
Biological significance

Current state of the problem

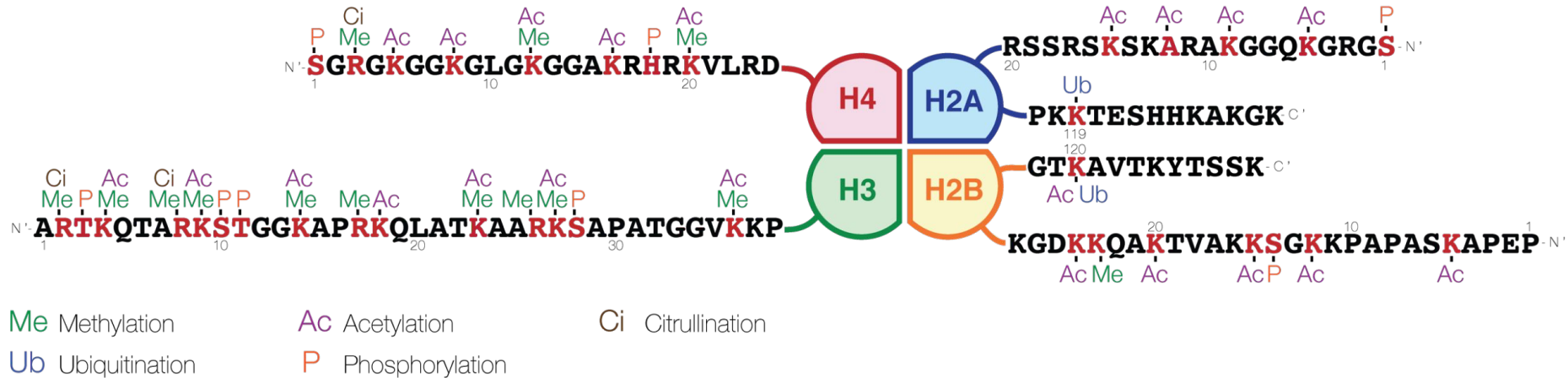
Solution and model architecture

Results and evaluation

The nucleosome is a fundamental component of chromatin



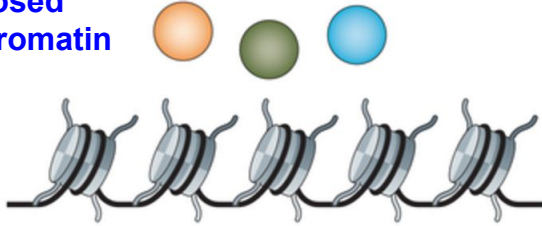
Histones tails are regulated by diverse modifications



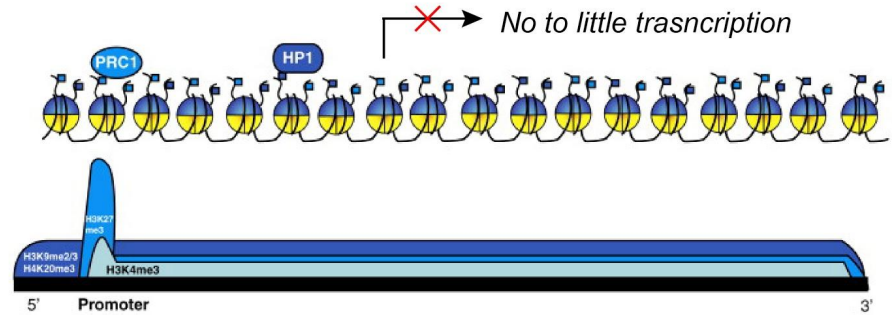
The 'histone code' hypothesis: the transcription of genetic information encoded in DNA is in part regulated by chemical modifications to histone proteins.

Histone modifications are dynamic and change in gene activation

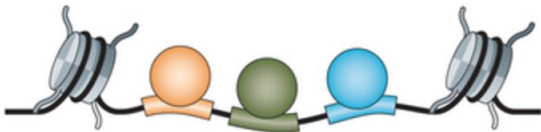
Closed
chromatin



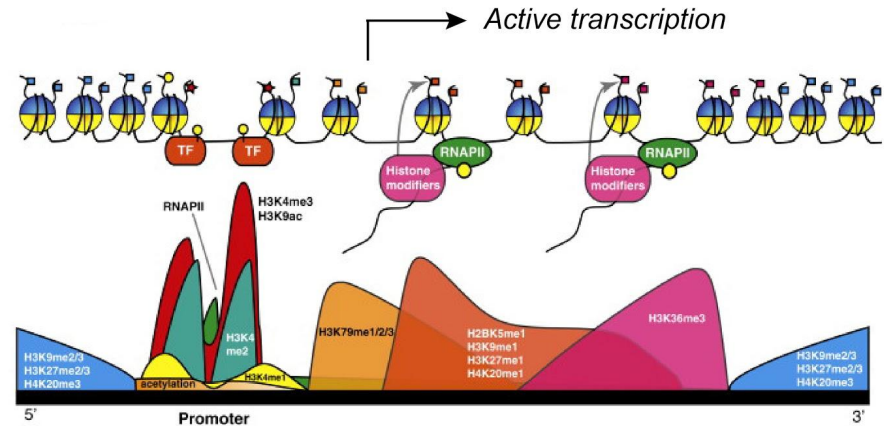
Inactive gene



Open
chromatin



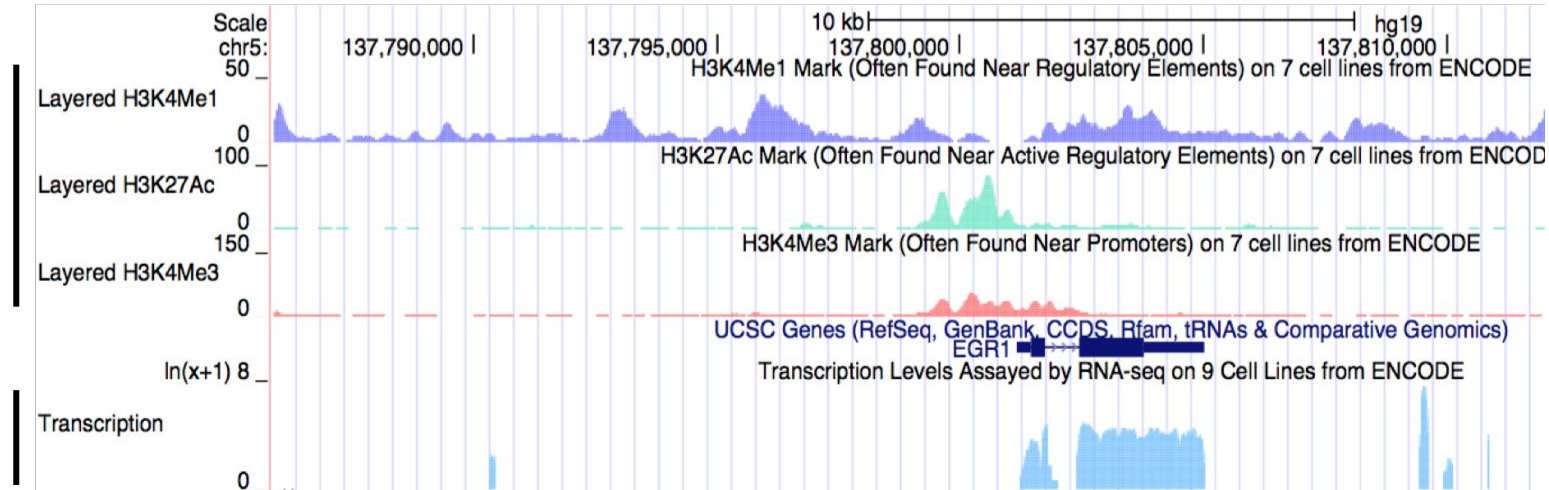
Active gene



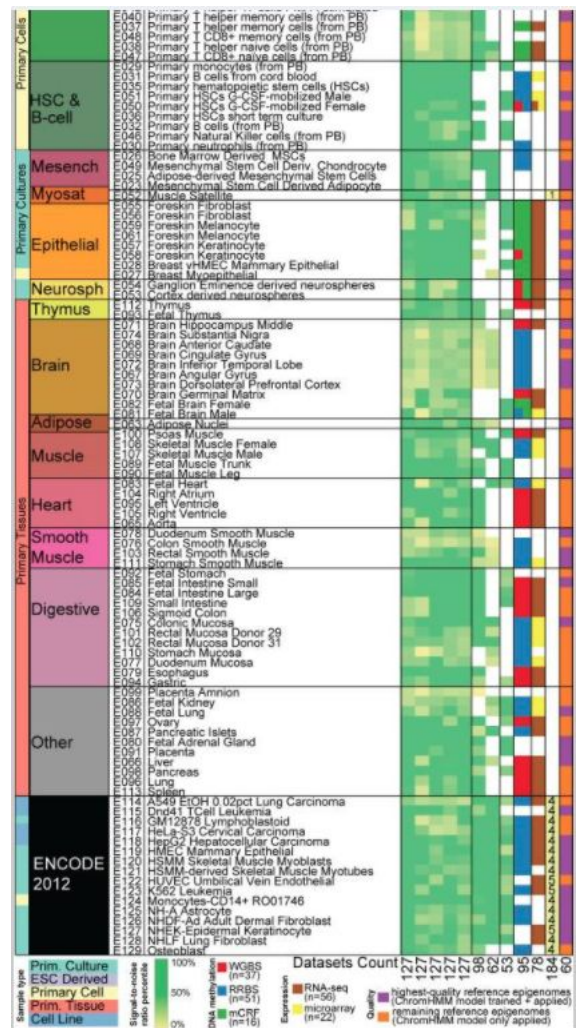
Statement of the classification problem

Given: **Histone Modification (HM) marks**

Predict: **gene expression**
(mRNA level)



Roadmap Epigenomics Project



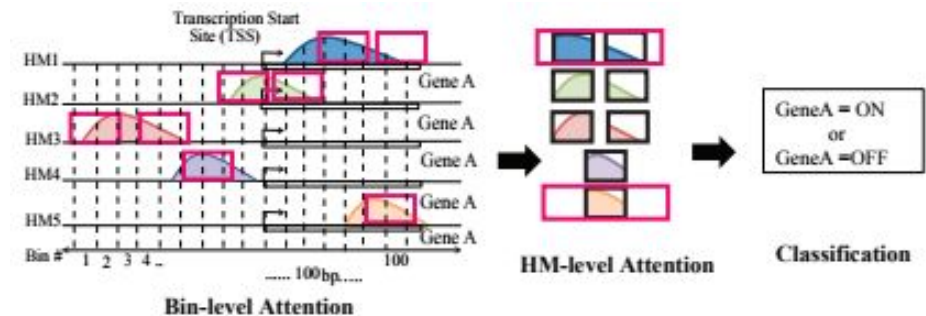
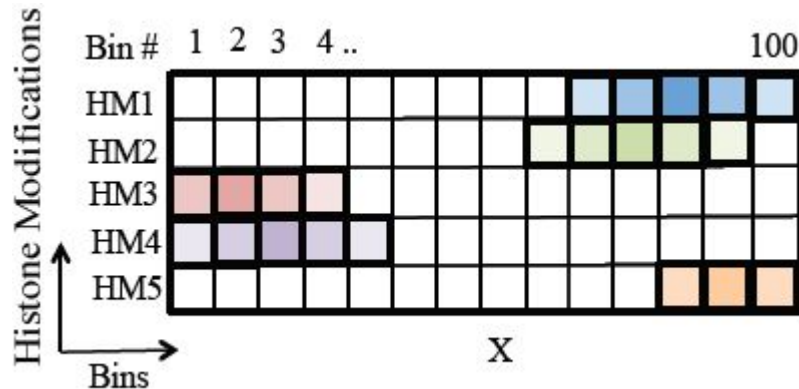
epigenomes. *Nature*, 518(7539), 317–330

Data preprocessing

Features: 10,000 bp regions around TSS with gene expression data for 56 different cell types; broken into 100 bins; 5 histone modification (HM) marks profiled for each bin position

Train/validation/test split: Three separate, equal-sized folds for training (6601 genes), validation (6601 genes), and testing (6600 genes)

Output: binary variable of gene expression being high or low; defined by the median gene expression



Key challenges for learning gene expression from chromatin

Genome-wide chromatin signals are spatially structured, high-dimensional and highly modular

Difficult to show how chromatin marks work together in controlling gene expression

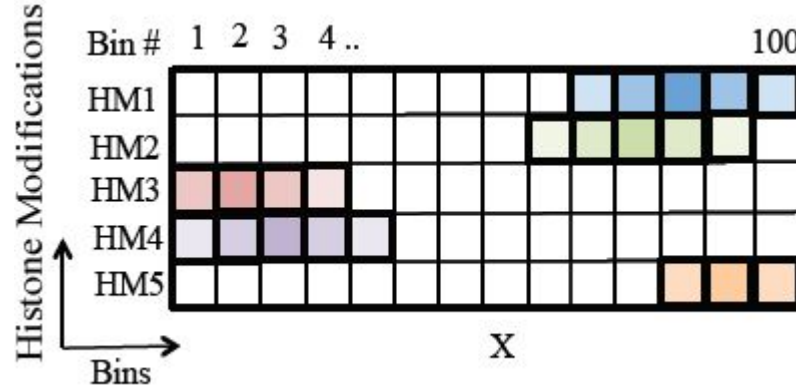
AttentionChrome: an attention-based deep learning approach that can model and interpret dependencies among chromatin factors for controlling gene regulation

Pitfalls of previous methods

Computational Study	Unified	Non-linear	Bin-Info	Representation Learning			Prediction	Feature Interact	Interpretable
				Neighbor Bins	Whole	Re-gion			
Linear Regression ([14])	×	×	×	×	✓		✓	×	✓
Support Vector Machine ([7])	×	✓	Bin-specific	×	✓		✓	✓	×
Random Forest ([10])	×	✓	Best-bin	×	✓		✓	×	×
Rule Learning ([12])	×	✓	×	×	✓		×	✓	✓
DeepChrome-CNN [30]	✓	✓	Automatic	✓	✓		✓	✓	×
AttentiveChrome	✓	✓	Automatic	✓	✓		✓	✓	✓

Deep neural net architecture

Data enters in matrix



Data Passes through several Bidirectional LSTMs and attention layers

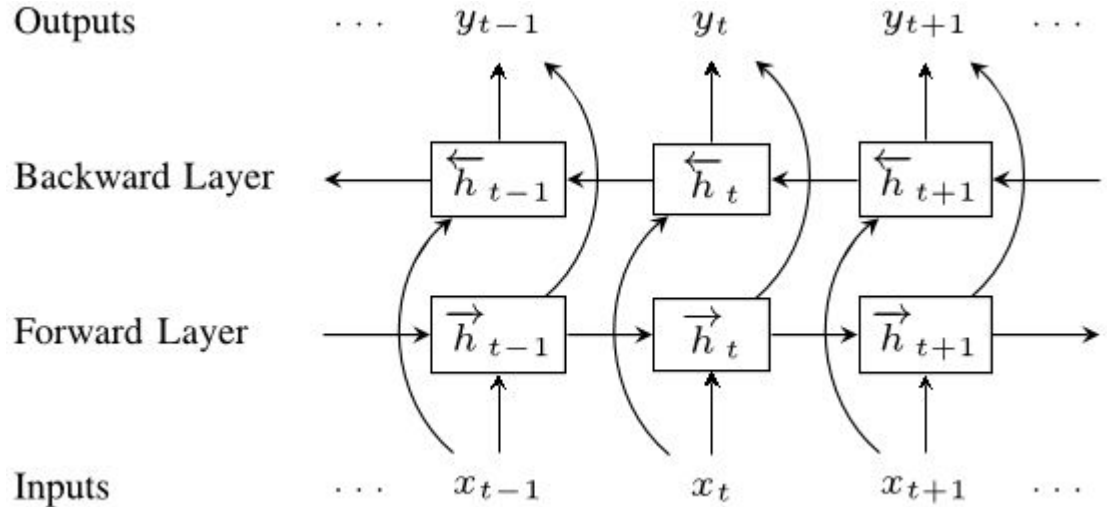
Softmax classifier predicts probability of gene being expressed

Bidirectional LSTM

Patterns exist in both forward and backward direction

BiLSTM captures both

Outputs are concatenated



Attention mechanism - concept

Find areas of interest in data

Success in natural language processing and computer vision

Visualization and Interpretability



Image Credit: *Show, Attend and Tell: Neural Image Caption Generation with Visual Attention*. Xu et al.

A dog is standing on a hardwood floor.

Attention mechanism - implementation

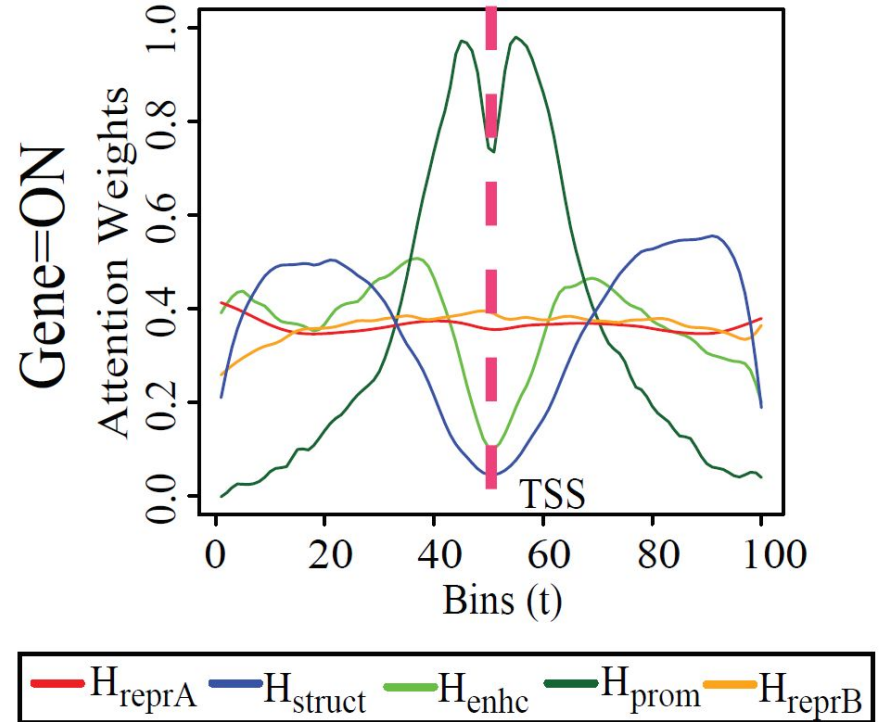
Relevance score for each input position

Relevance learned during training

Normalized and turned into weights

Inputs are multiplied by weights

Emphasizes important areas



Softmax Classifier

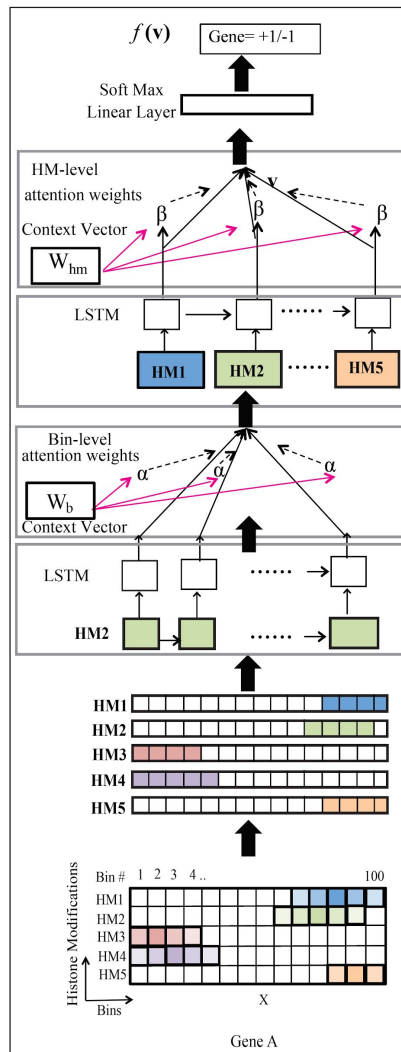
HM Level Attention

HM Level BiLSTM (!)

Bin Level Attention

Bin Level BiLSTM

Input



(g) Classification

(f) HM-level Attention

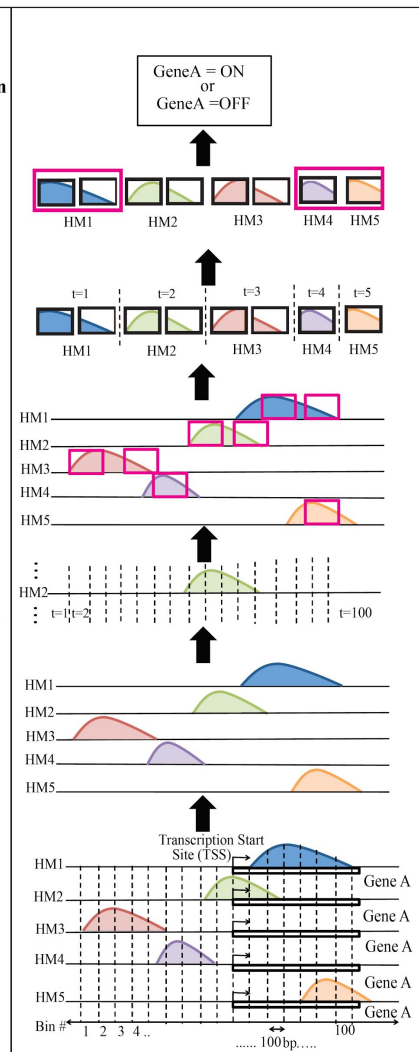
(e) HM-level Encoder
LSTM for all HMs

(d) Bin-level Attention

(c) Bin-level Encoder
LSTM for each HM

(b) Input

(a) Feature Generation



Results: AUC performance for variations of model and baselines

	Baselines		AttentiveChrome Variations				
Model	CNN[30]	LSTM	CNN-Attn	CNN- α, β	LSTM-Attn	LSTM- α	LSTM- α, β
Mean	0.8008	0.8052	0.7622	0.7936	0.8100	0.8133	0.8115
Median	0.8009	0.8036	0.7617	0.7914	0.8118	0.8143	0.8123
Max	0.9225	0.9185	0.8707	0.9059	0.9155	0.9218	0.9177
Min	0.6854	0.7073	0.6469	0.7001	0.7237	0.7250	0.7215
Improvement over CNN (out of 56 cell types)		36	0	16	49	50	49

Results: AUC performance for variations of model and baselines

	Baselines	
Model	CNN[30]	LSTM
Mean	0.8008	0.8052
Median	0.8009	0.8036
Max	0.9225	0.9185
Min	0.6854	0.7073
Improvement over CNN (out of 56 cell types)	36	

Standard Baselines:

LSTM: LSTM without adding any attention

CNN: Temporal (1-D) CNN

No consideration of the modular property of each HM mark

Results: AUC performance for variations of model and baselines

Model	AttentiveChrome Variations				
	CNN-Attn	CNN- α, β	LSTM-Attn	LSTM- α	LSTM- α, β
CNN-Attn: Baseline CNN with one attention layer; max-pooling layer removed	0.7622	0.7936	0.8100	0.8133	0.8115
Median	0.7617	0.7914	0.8118	0.8143	0.8123
Max	0.8707	0.9059	0.9155	0.9218	0.9177
LSTM-Attn: Baseline LSTM with one attention layer	0.6469	0.7001	0.7237	0.7250	0.7215
Improvement over CNN (56 cell types)	0	16	49	50	49

Results: AUC performance for variations of model and baselines

CNN- α , β : 1 CNN/mark; 1 α -attention/mark; β -attention to combine HM				AttentiveChrome Variations			
				CNN- α, β	LSTM-Attn	LSTM- α	LSTM- α, β
Mean	0.8008	0.8052	0.7622	0.7936	0.8100	0.8133	0.8115
LSTM- α : 1 LSTM/mark; 1 α -attention/mark	0.8036	0.8036	0.7617	0.7914	0.8118	0.8143	0.8123
	0.9185	0.8707		0.9059	0.9155	0.9218	0.9177
Min	0.6854	0.7073	0.6469	0.7001	0.7237	0.7250	0.7215
LSTM- α , β : 1 LSTM/mark; 1 α -attention/mark; β -attention to combine HM				16	49	50	49

Results: AUC performance for variations of model and baselines

	Baselines		AttentiveChrome Variations				
Model	CNN[30]	LSTM	CNN-Attn	CNN- α, β	LSTM-Attn	LSTM- α	LSTM- α, β
Mean	0.8008	0.8052	0.7622	0.7936	0.8100	0.8133	0.8115
Median	0.8009	0.8036	0.7617	0.7914	0.8118	0.8143	0.8123
Max	0.9225	0.9185	0.8707	0.9059	0.9155	0.9218	0.9177
Min	0.6854	0.7073	0.6469	0.7001	0.7237	0.7250	0.7215
Improvement over CNN (out of 56 cell types)		36	0	16	49	50	49

Results: Using attention scores for interpretation

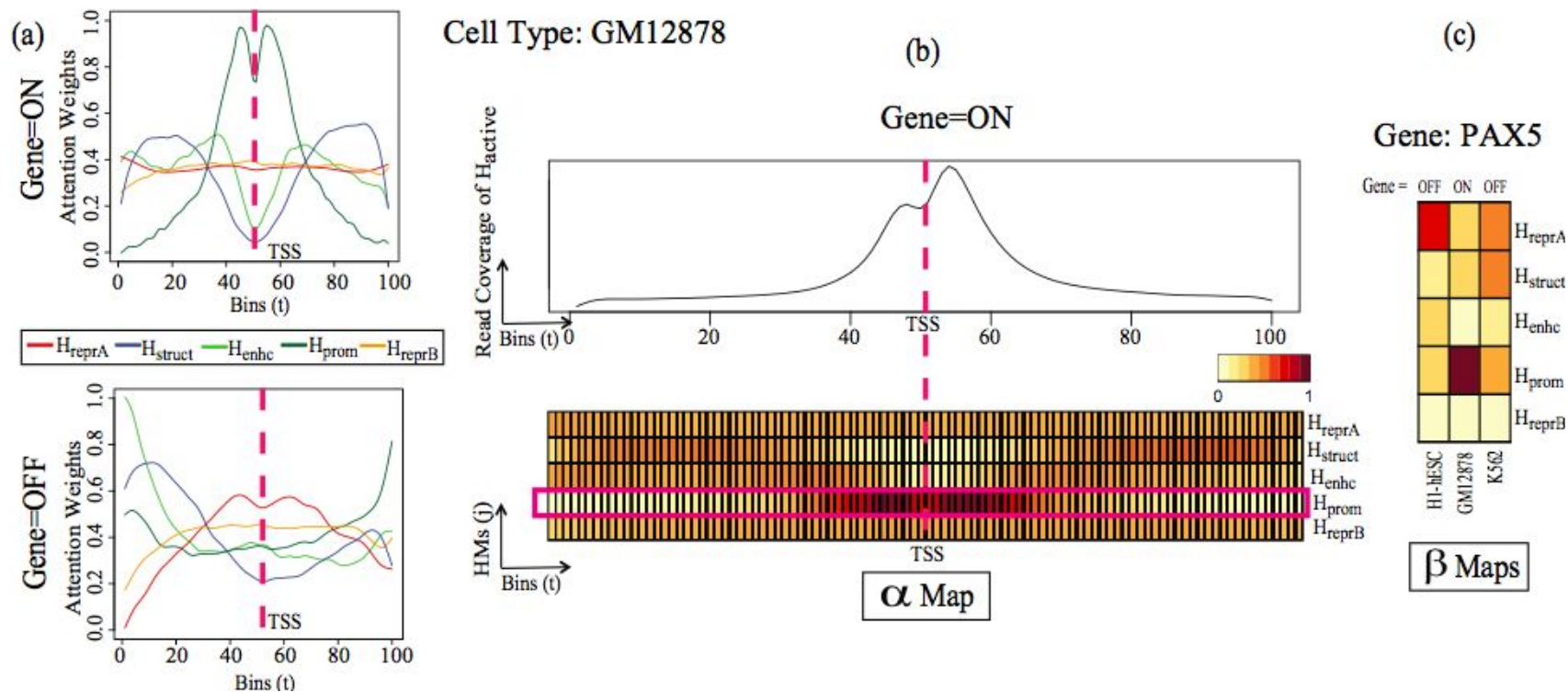
H3K27ac (H_{active}) is a good indicator of active gene region

Genome-wide reads of H3K27ac available for three cell types and used to validate interpretation

Viz. Methods	H1-hESC	GM12878	K562
α Map (LSTM- α)	0.8523	0.8827	0.9147
α Map (LSTM- α, β)	0.8995	0.8456	0.9027
Class-based Optimization (CNN)	0.0562	0.1741	0.1116
Saliency Map (CNN)	0.1822	-0.1421	0.2238

Pearson correlation of importance weight assigned for H_{prom} by different visualization techniques and H_{active} for predicted “ON” genes across three cell types

Results: Using attention scores for interpretation



(a) Average attention weights when predicting gene=ON and gene=OFF (b) top: cumulative H_{active} signal across active genes, bottom: bin-level α -weights heatmap (c) HM-level β -weights heatmap

Commentary

Summary:

- Attention-based deep learning approach provides more accurate predictions and attention weights provide a good interpretation for prediction

Critiques:

- Use of LSTM might not be necessary to model HM level relationships
- Conclusions not generalizable across cell-type (averaged AUC obscures type-wise performance)
- No evaluation metric other than AUC (might overlook class imbalance)
- The validation of interpretation utility of the attention weights by comparison of the averaged profiles is dubious