

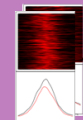
(1) Data Preparation

Group PSSM scores
and histone modification data into
100bp bins for entire genome



(2) Training data

Promoter
(TSS with strong H3K4me3 & PSSM score)



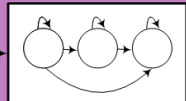
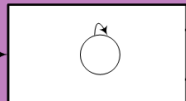
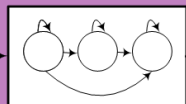
Background
(entire chromosome 1)



Enhancer
(P300 binding sites with high H3K4me1/2 & PSSM score)



(3) Model training



(4) Whole genome scanning (2 log-odds are calculated)

$$\log \frac{P(\mathbf{x} | \Theta^{\text{promoters}})}{P(\mathbf{x} | \Theta^{\text{background}})}$$

$$\log \frac{P(\mathbf{x} | \Theta^{\text{enhancers}})}{P(\mathbf{x} | \Theta^{\text{background}})}$$

(5) TFBS predictions (peak detection)

