

# Improve HiC Resolution Based on Relationship Between Epigenetic Features and Chromatin Interaction Structures



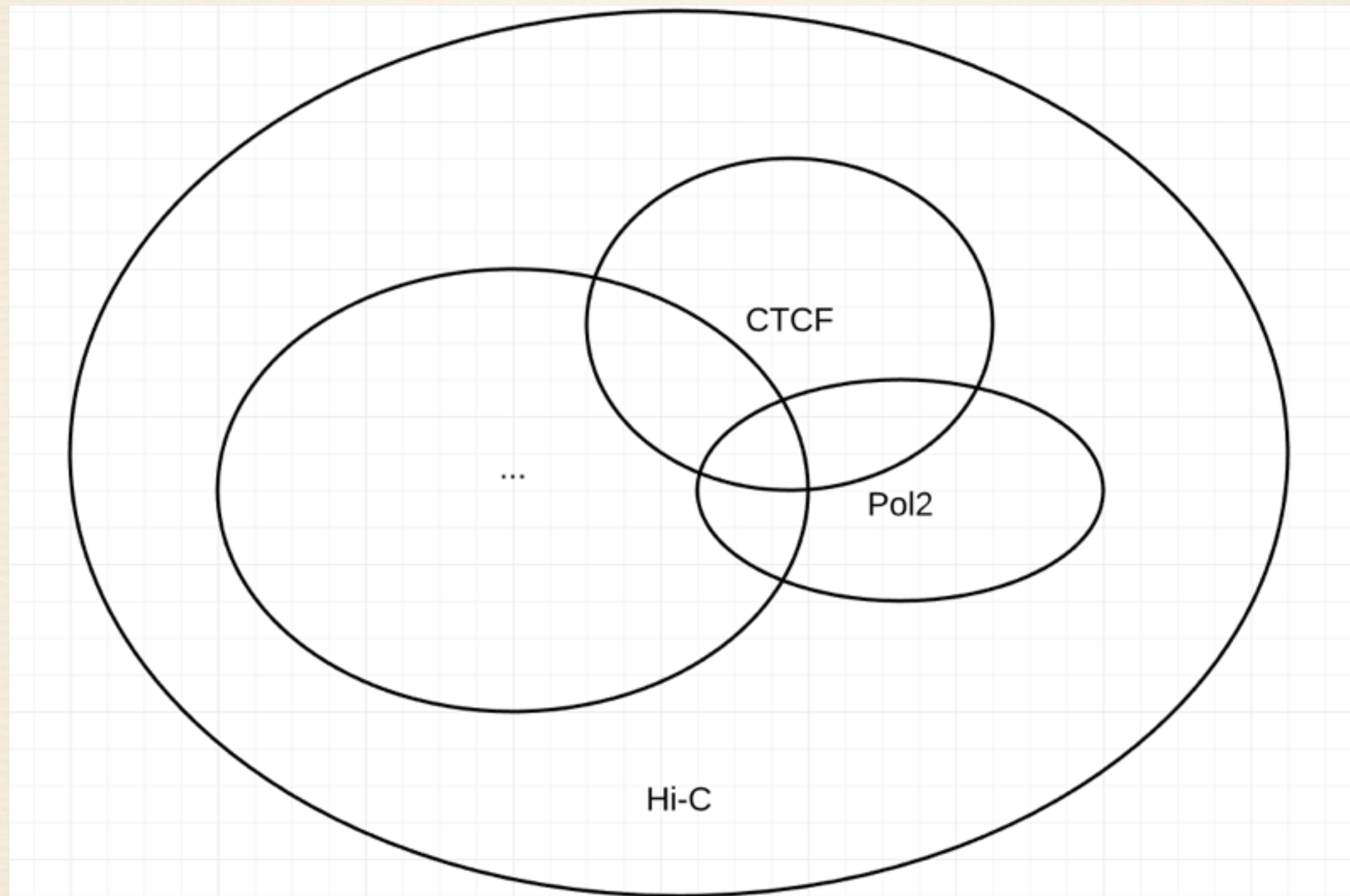
*Minzhen Yi*  
*12/10/2015*



# Motivations

- ❖ Resolution for Hi-C experiment data is relatively low<sup>1</sup>
- ❖ ChIA-PET can only detect partial interaction in each experiment<sup>2</sup>
- ❖ Significance of genome-wide chromatin interaction map
- ❖ Difficulties in improving Hi-C resolution by experiment

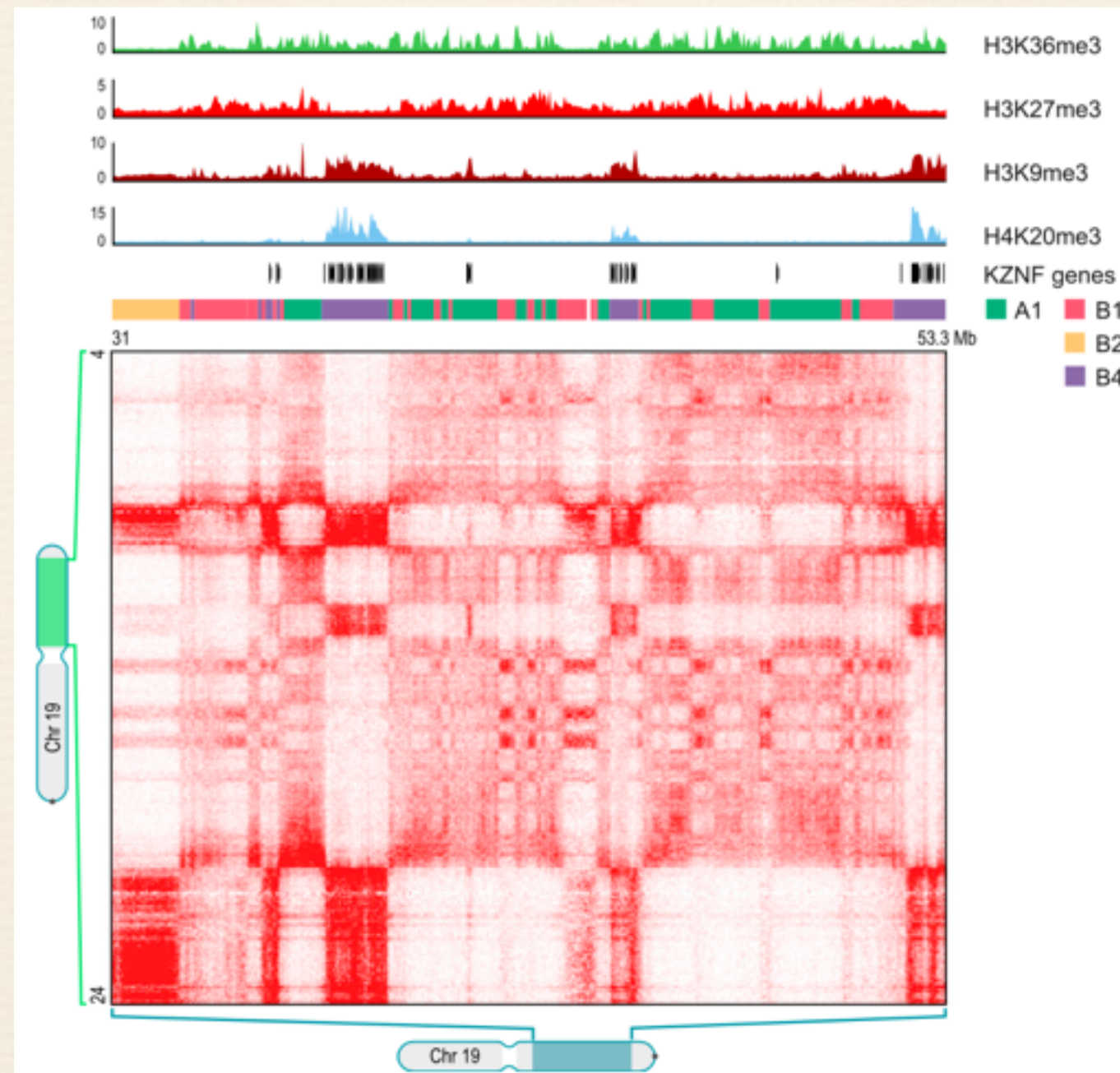
# Idea illustrated by Venn Diagram



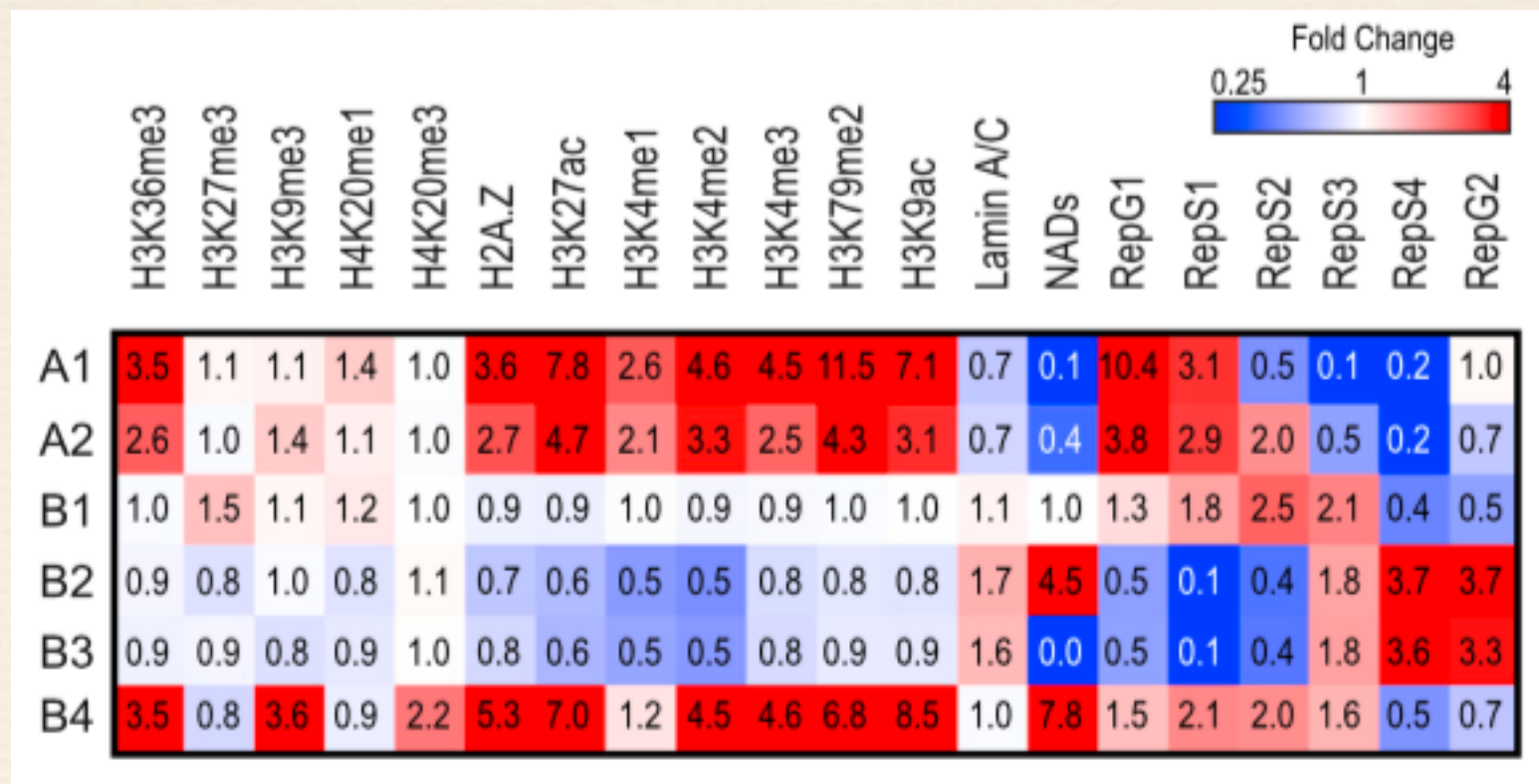


# Assumptions

- ❖ Chromatin interaction structures are related to epigenetic features<sup>3</sup>







# Data Resource

- ❖ ChIP-seq: Encode ChIP-seq Experiment Matrix
- ❖ Other data reflecting epigenetic features (DNase, Replication timing): Encode
- ❖ ChIA-PET: Encode, GEO, published articles
- ❖ Hi-C: GEO, publish articles



# Data Format

- ❖ To be selected: narrowpeak, broadpeak, bam, bed
- ❖ Selected: bam at first; narrowpeak, bed at present

# Failed Models

- ❖ Nonlinear Regression: Y: n by n interaction matrix, X: m by n feature matrix, A: m by m coefficient matrix

$$Y = XAX^T$$

- ❖ Multi Linear Regression: Y: 1 by n interaction vector, A: 1 by m\*m coefficient matrix, F: m\*m by n interaction matrix

$$Y = AF \quad y_i = \sum_{j=1}^m \sum_{k=1}^m a_{jk} h_{ij} t_{ik}$$

Note that m is the number of features, n is the number of fragment pairs



# F-test for Linear Relationship

$$\text{cov}(X,Y) = E[X - E(X)][Y - E(Y)]$$

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{D(X)D(Y)}}$$

$$L_{yy} = \sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N y_i^2 - N\bar{y}^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$$

TSS

ESS

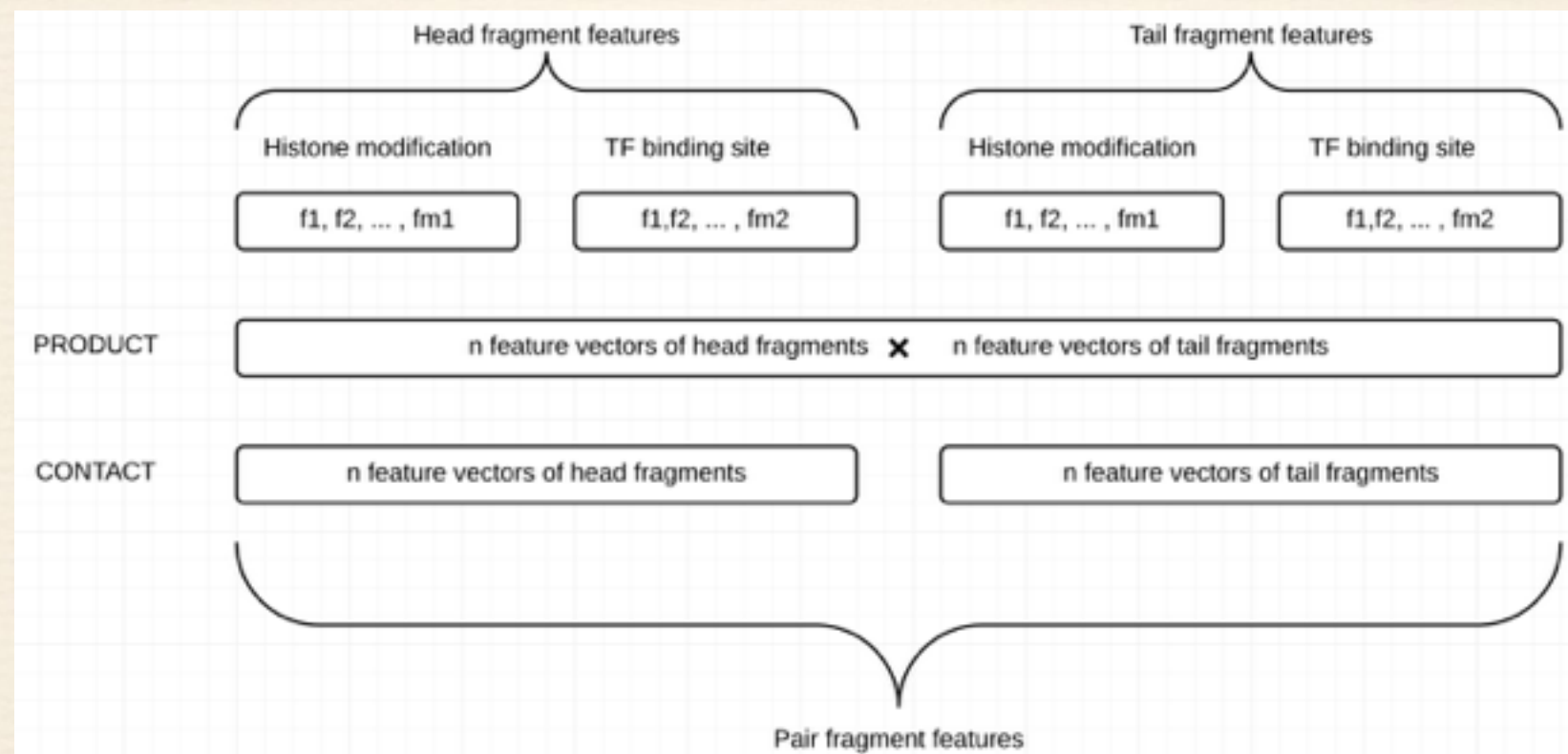
RSS

$$F = \frac{ESS / f_E}{RSS / f_R} = \frac{(n - m^2 - 1) ESS}{m^2 RSS}$$



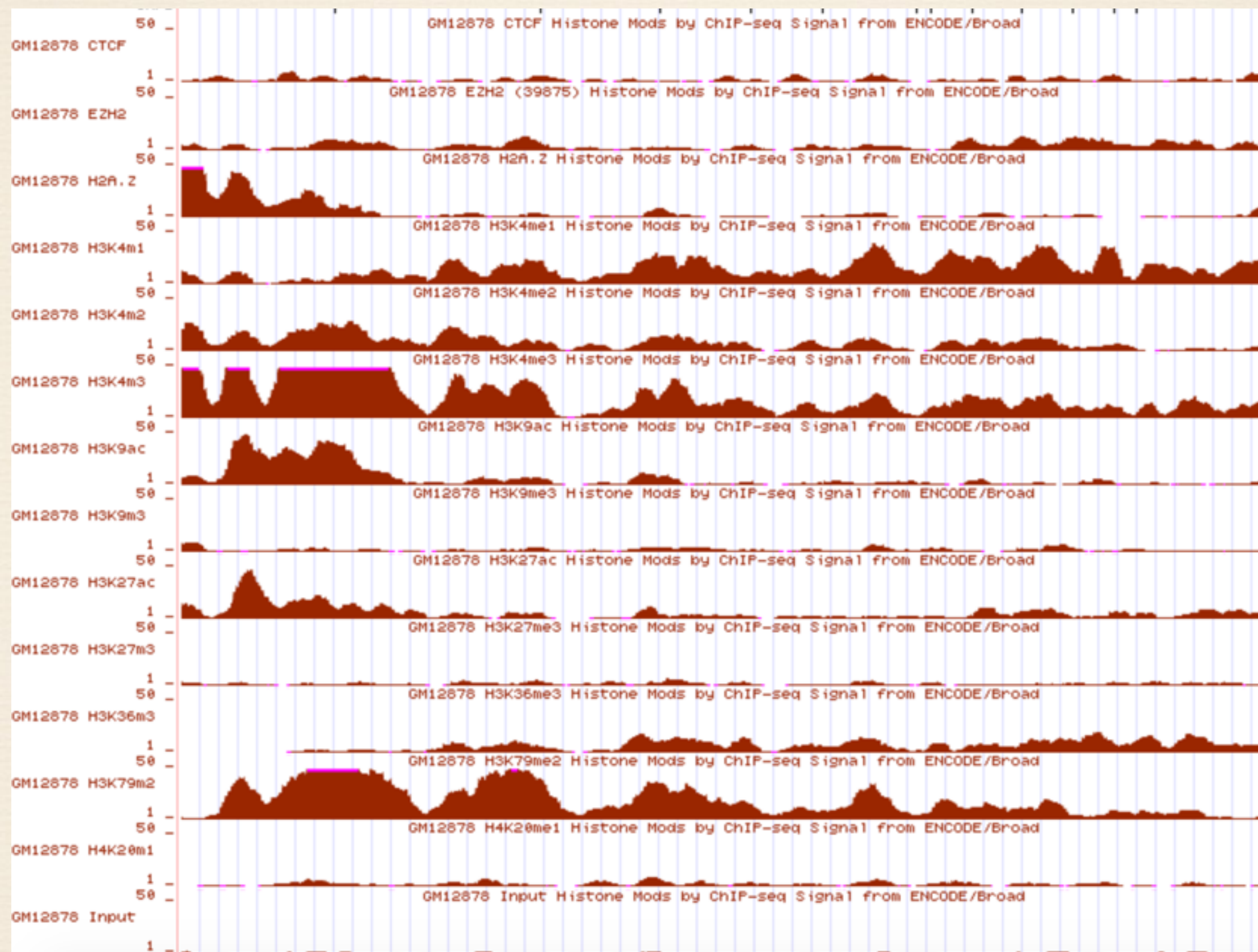
# Methods<sup>4</sup>

- ❖ Model: Binary Classification
- ❖ Two strategies to represent an interaction pair: PRODUCT and CONTACT



- ❖ Feature quantification: continuous or binary

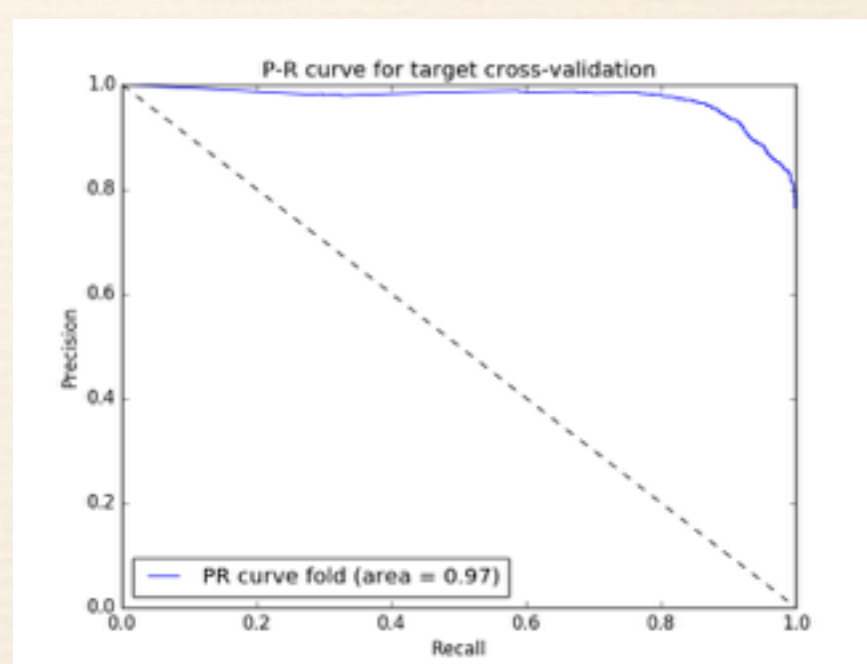
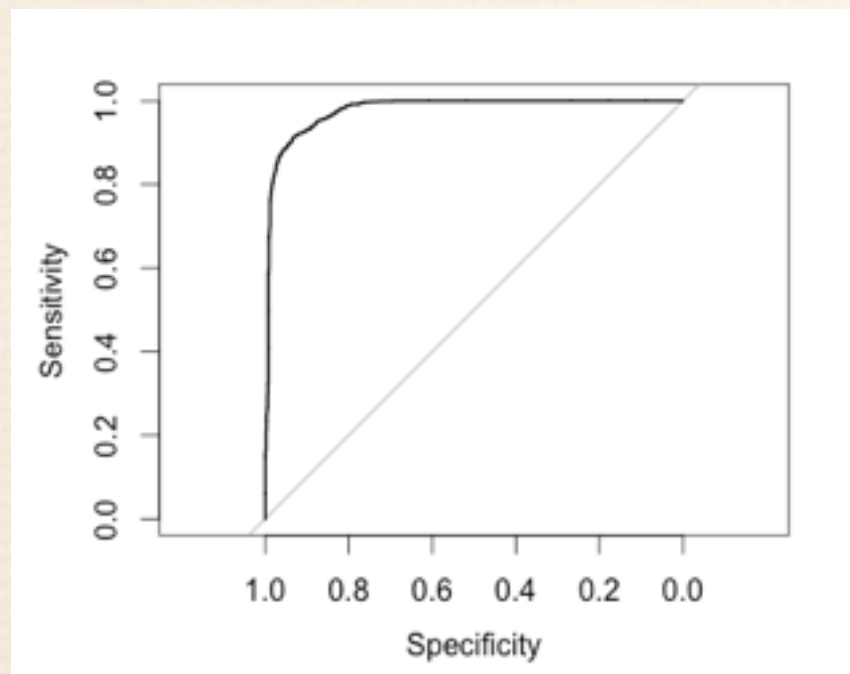
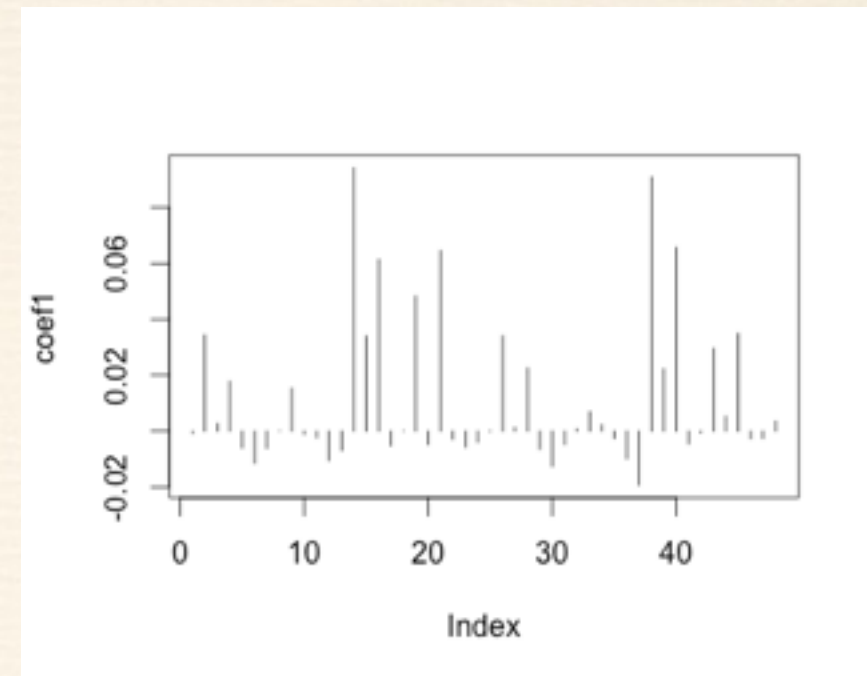
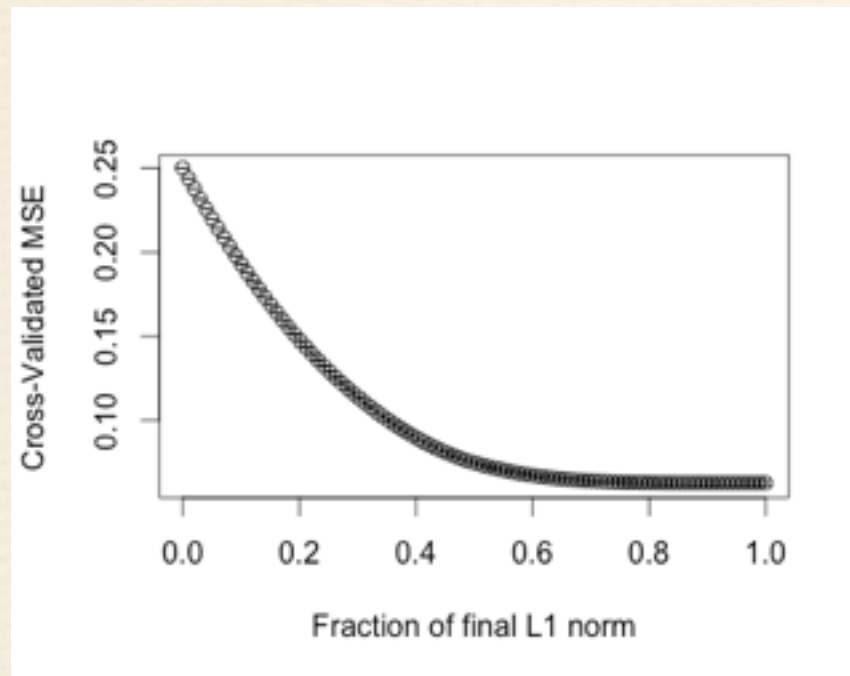






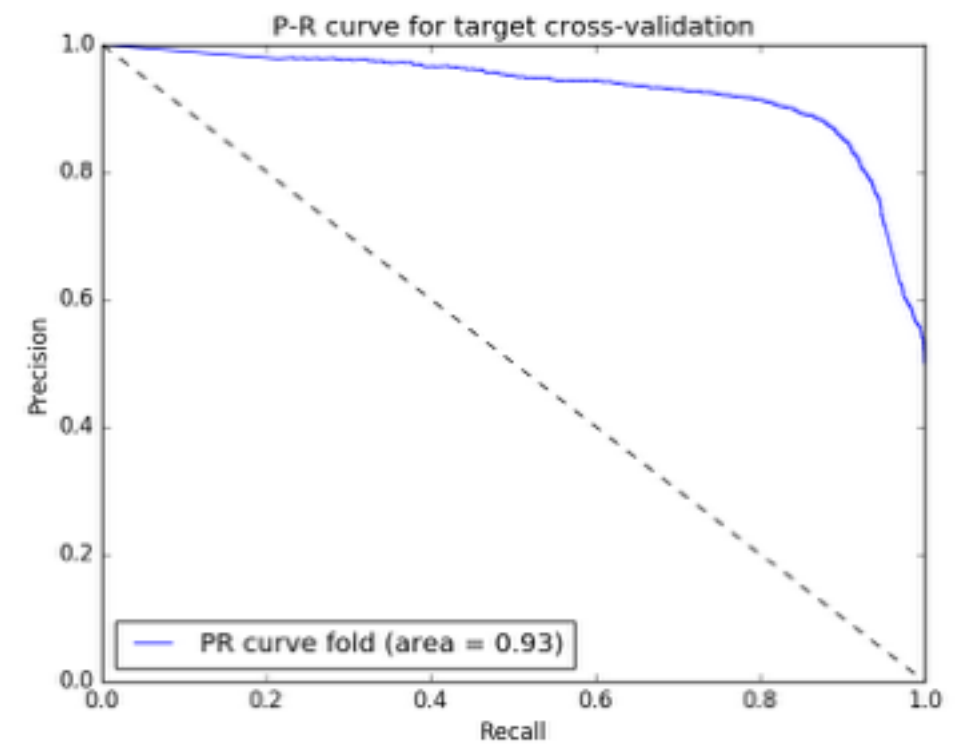
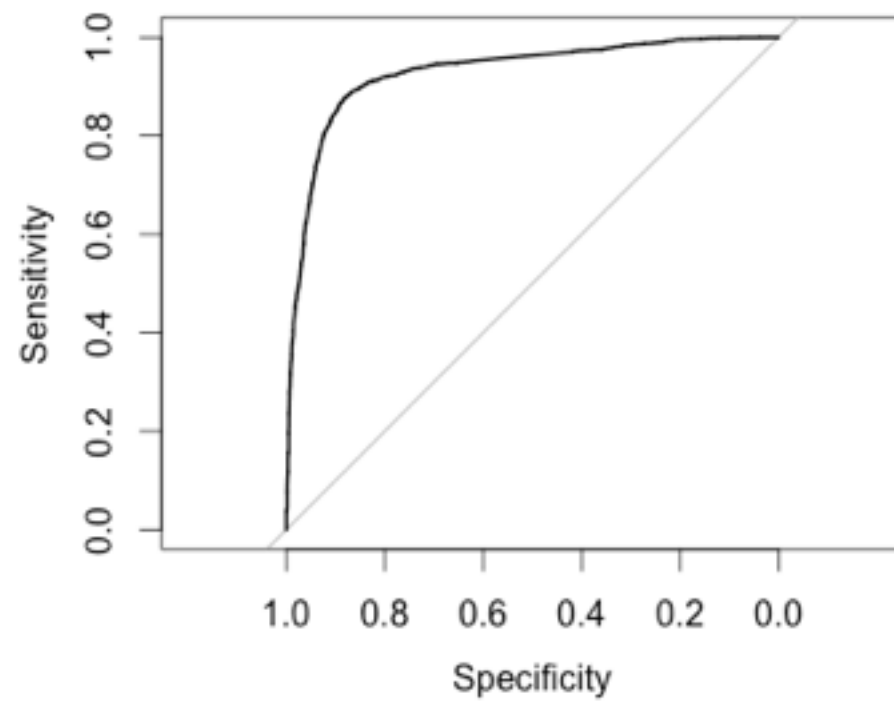
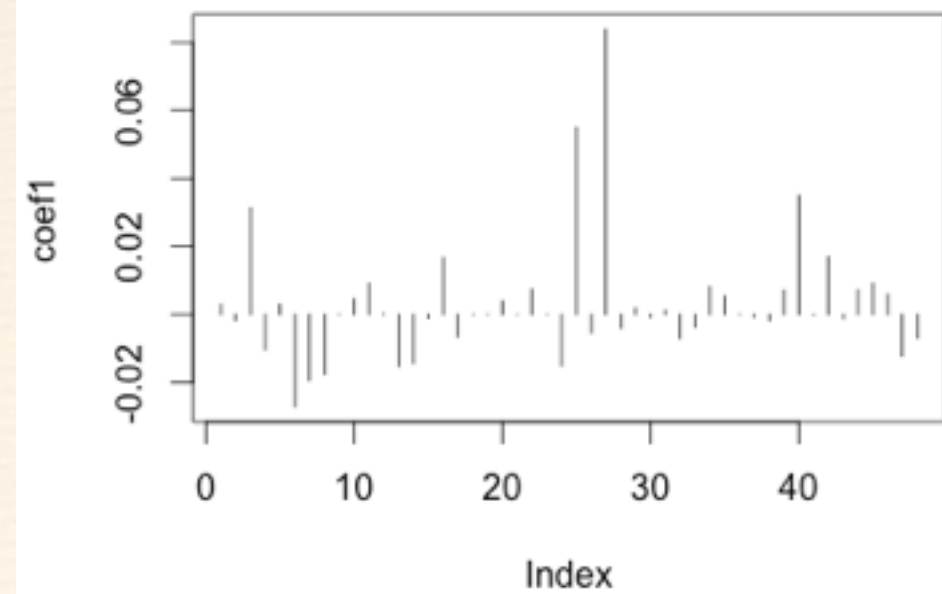
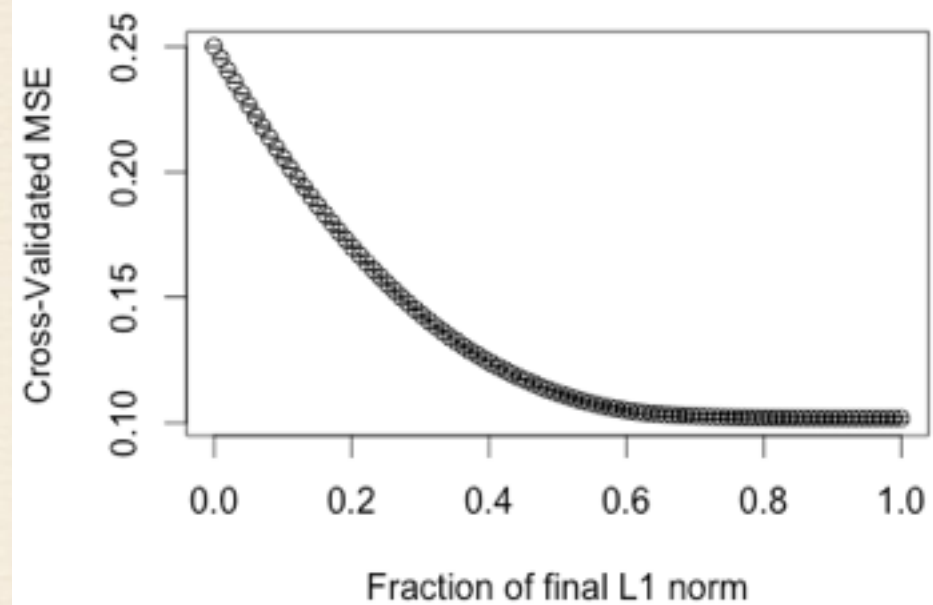
# Results

GM12878, ChIA-PET CTCF





# K562, ChIA-PET CTCF





# Reference

1. Melissa J. Fullwood et al. An oestrogen-receptor--bound human chromatin interactome. *Nature* **462**, 58-64 (5 November 2009)
2. Erez Lieberman-Aiden. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* Vol. 326 no. 5950 pp. 289-293 (9 October 2009)
3. Suhas S.P. Rao et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* Vol. 159 Issue 7 pp. 1665-1680 (18 December 2014)
4. A predictive modeling approach for cell line-specific long-range regulatory interactions