ENGINE: an Enhancer-Gene Interaction dEtection method using robust feature extraction.

Part2: Tuning and feature selection

Lou Shaoke

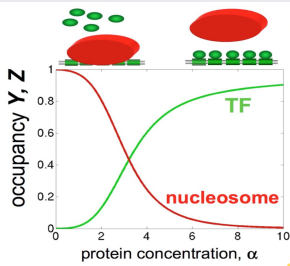Department of Molecular Biophysics and Biochemistry

loushaoke@gmail.com

December 15, 2015

Biological evidence

We have allelic information including: ASE, ASB, ASD(NaseI), ASM, ASH; eQTL, meQTL, dsQTL.

TF Assoc	Nucleosome Assoc	Effect
ASB, ASD1, ASE	ASD2, ASH	QTL

SNV in TF binding region contribute to TF binding (Binding)
The surrounding (buffering) region contribute directly to nucleosome (Buffering)
However, there are conversions between Binding and Buffering

Motivation: Link buffering SNV with TF binding region SNV

A harmony model:

Binding region is thought to visible, and directly affect TF binding (Y);

SNV in Binding region that potential affect Buffering to introduce SNV in Buffering or affect nucelosome binding ($X|Y$)

Buffering region is though to invisible, and indirectly affect TF (X);

SNV in Buffering region that result in TF bound/unbound ($Y|X$)

There are four state: Y, X, $X|Y$, and $Y|X$ with distribution $P(Y|\alpha)$, $P(X|\beta)$, $P(Y|X, \alpha, \beta)$ and $P(X|Y, \alpha, \beta)$, α, β are parameters.

Maximize harmony function:

$H(Y||X) = \int P(Y)P(X|Y)ln(P(X)P(Y|X))dXdY$