ENGINE: an enhancer gene interaction detection algorithm using robust feature extraction

Lou Shaoke

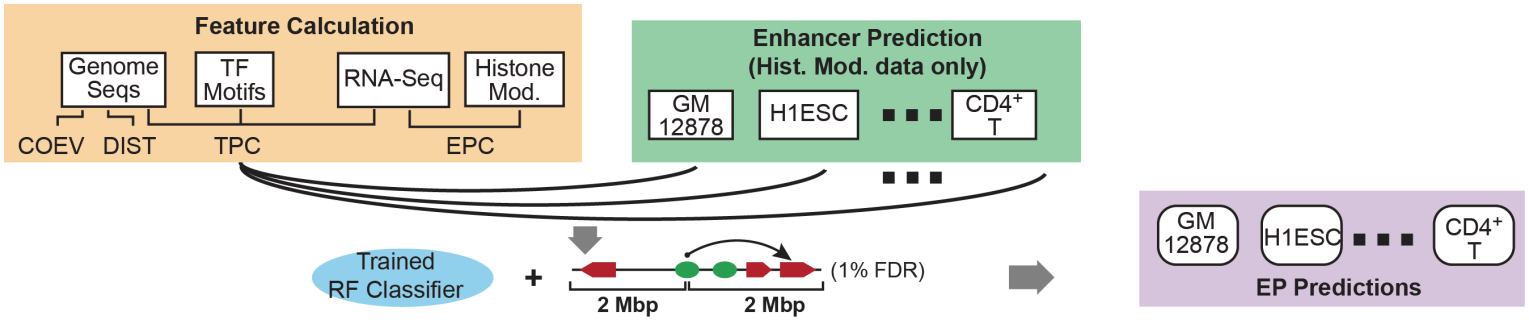
Department of Molecular Biophysics and Biochemistry

loushaoke@gmail.com

April 26, 2016

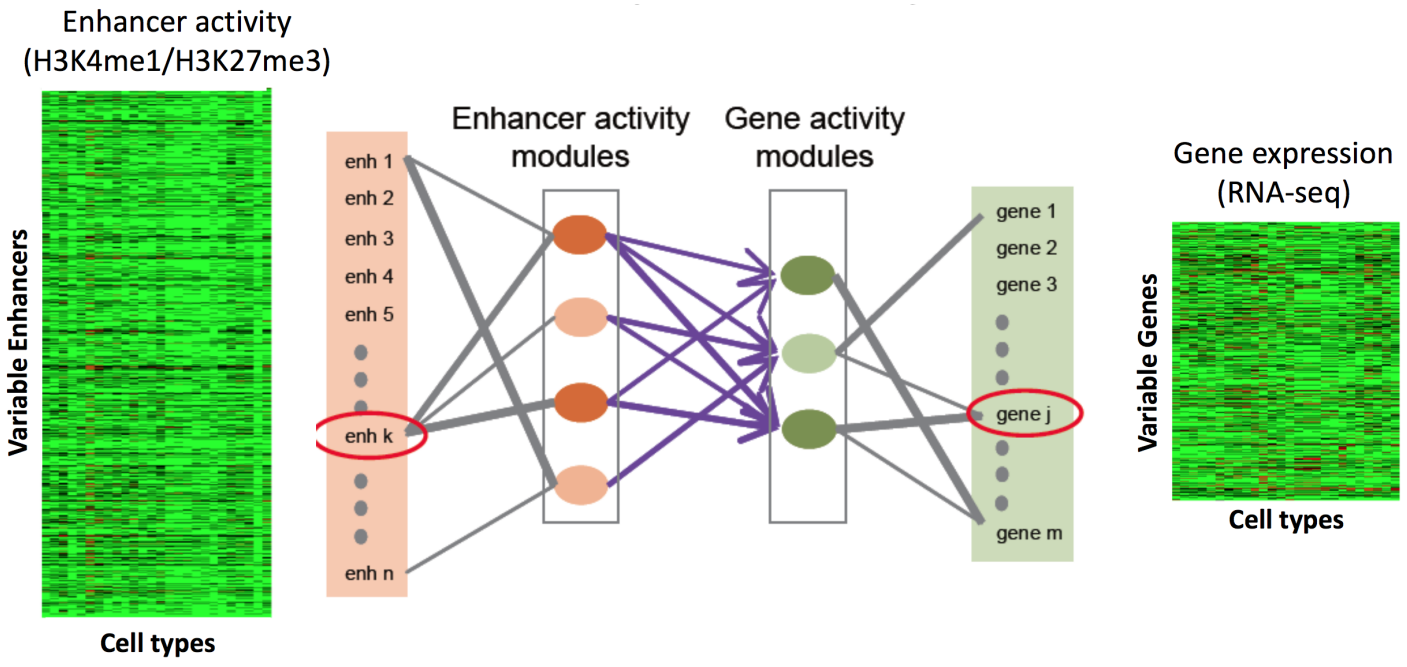
Yale

State of the art



IM-PET: Consider information from 3D gnome interactions, DIST(distance) constrain is a triky feature, boosting AUC from 0.7+ to 0.9+.

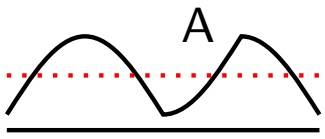
State of the art



LDA:a mixed membership method, didn't use information from 3d genome interaction, and reply on predifined enhancer region, sometimes it has worse aggreements.

Motivation

Average over regions ignore the signal pattern;



No explicitely pairwise pattern recognition;

Local signal normlization;

Structure of Engine

Part1. Why use SURF? Does it work? (done)

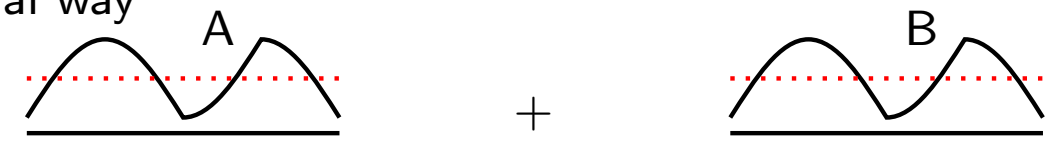
Part2. How Surf works (almost-done)

Part3. Local sequence feature and Kmer cooccurrence (almost-done)

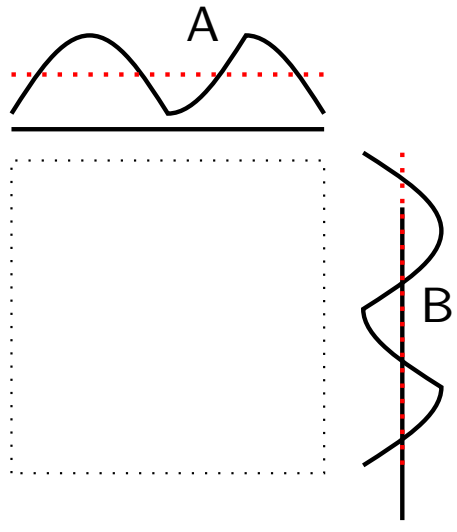
Part4. Dynamics feature with local sequence feature (almost-done)

Part5. Genome-wide search (on-going)

Traditional way

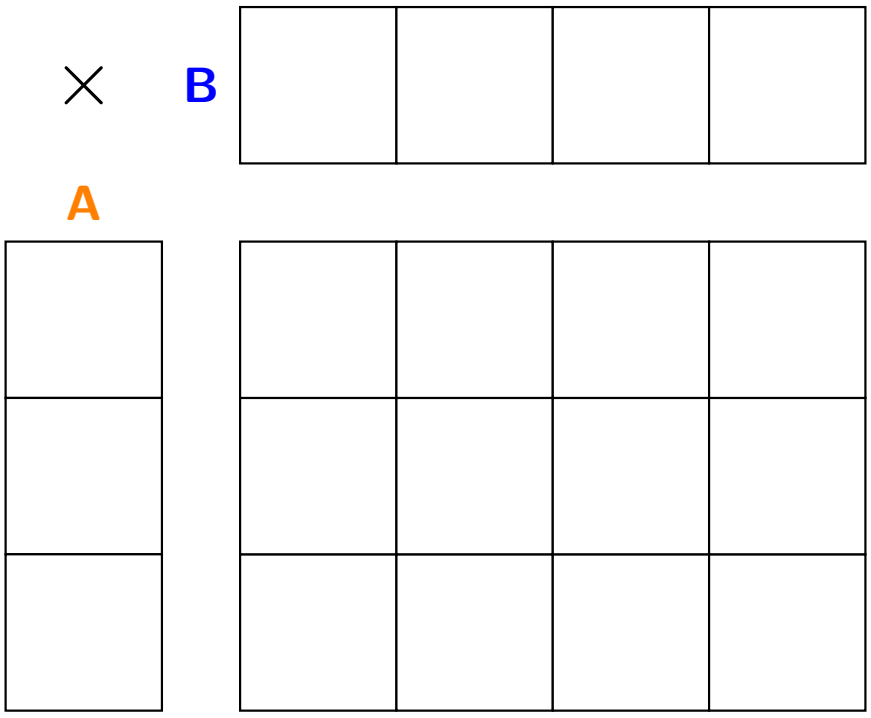


ENGINE



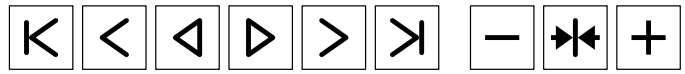
A 5 15 11 0 2 7 14 11

B 9 14 14 15 12 15 7 15

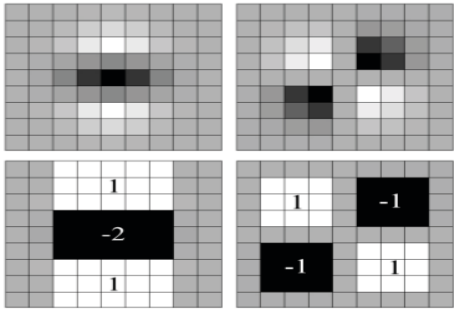


max= undef

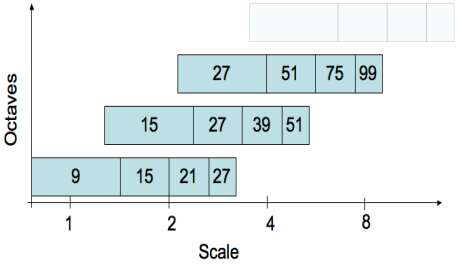
min= undef



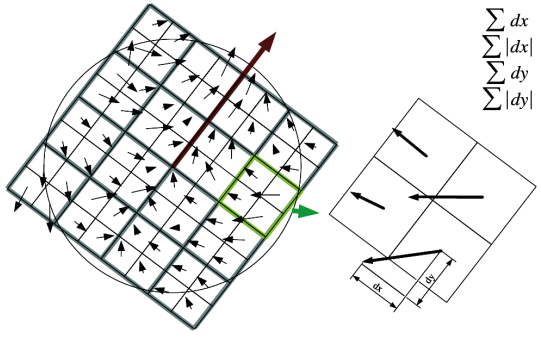
SURF (Speeded Up Robust Feature) is a robust image blob detector and descriptor, first presented by Herbert Bay et al. in 2006.



$$H(x,\sigma) = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix}$$

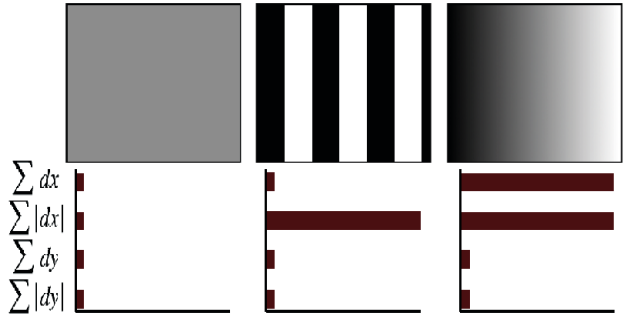


up-scaling filters and scale space



Descriptor based on Sum of Haar Wavelet Responses

$$v = \{\sum d_x, \sum |d_x|, \sum d_y, \sum |d_y|\}$$

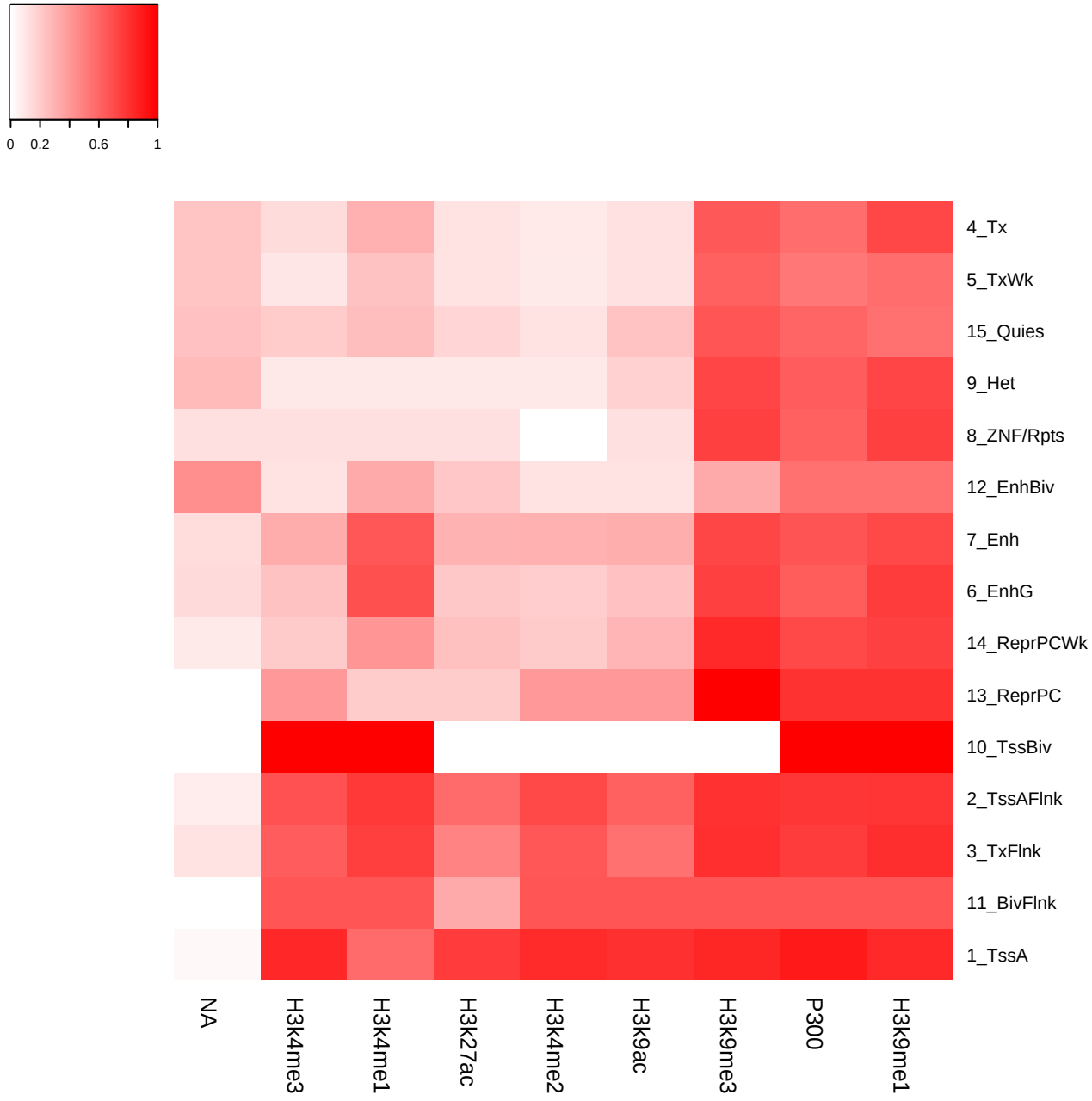


Distance= $\sum (v_2 - v_1)^2$

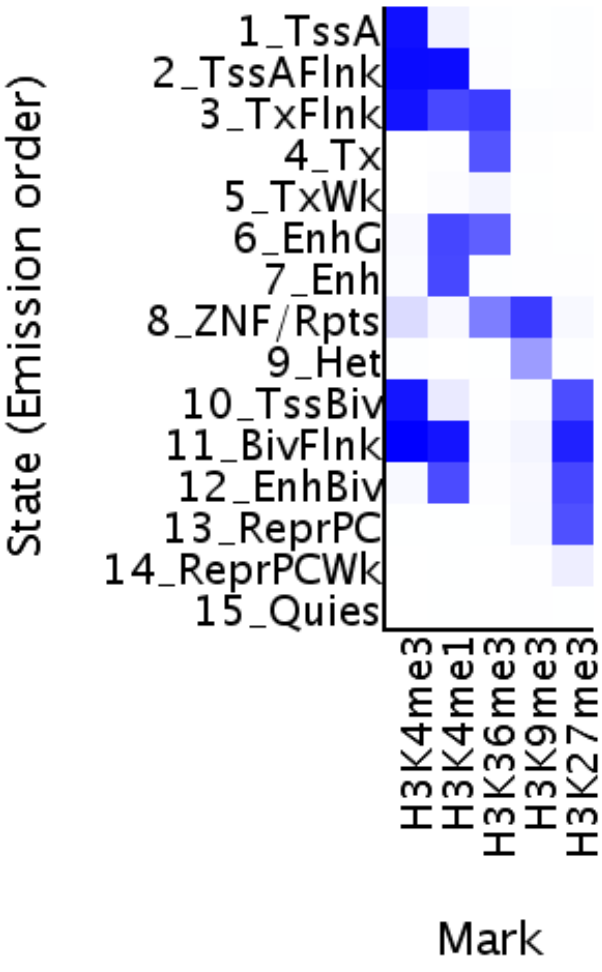
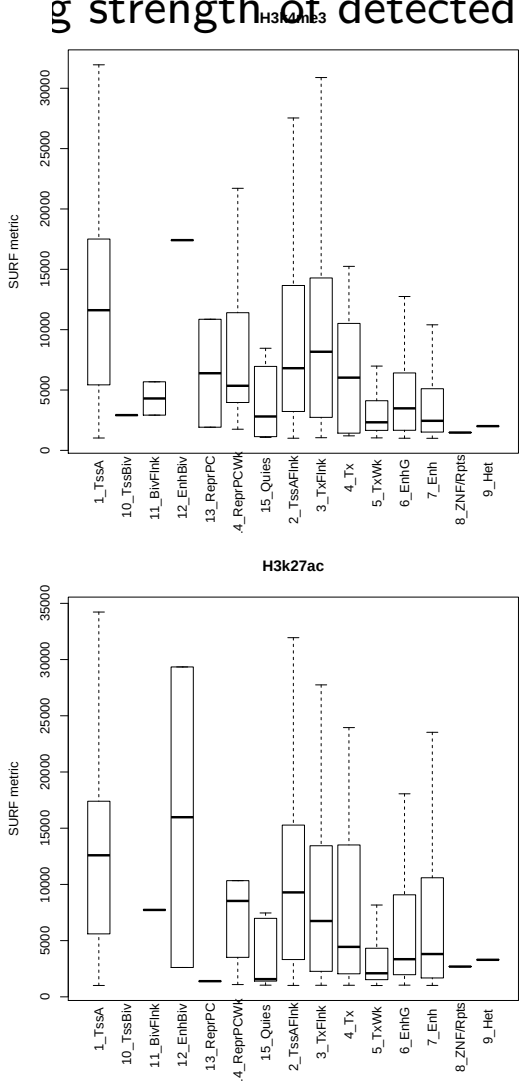
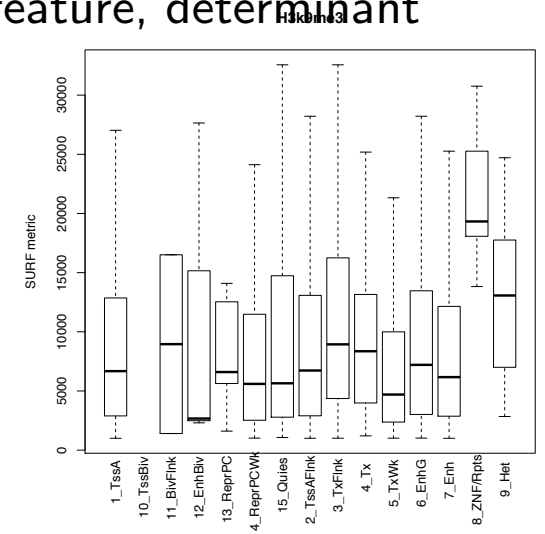
Detection **Description** **Comparison**

Surf feature overlap with ChromHMM

Except 12_EnhBiv, over 85% ChrommHMM functional regions overlap with at least one Surf feature



Surf feature score is consistent with emission prob



g strength of detected feature, determinant