(a)

Input data

Regression modeling

Cross-validate predictions against RNA-seq expression data

ChIP/DNase-seq data

Genes

Distance from TSS

Predicted expression

Log-linear regression model

SVR model

(b)

Subdivide genes by biological process

Process 1

Process 2

Process n

Construct regression models for each set of genes subdivided by biological process

Repeat (a) for each group of genes

Identify models with significant prediction accuracy by permutation testing

Transcript factors more predictive

Histone modifications more predictive