

## SUPPLEMENTARY FIGURE LEGENDS

1. Examples of BET-inhibited gene prediction by ROSE super-enhancers.
  - A. TMC3, a gene that is associated with a super-enhancer (shown in red) but is not down-regulated by JQ1.
  - B. VREB3, an example of a JQ1 down-regulated gene that is not super-enhancer associated even though the gene is surrounded by highly enriched BRD4 and H3K27ac chromatin.
  - C. RIPK4, a gene that is super-enhancer associated but not JQ1 down-regulated.
  - D. F13A1, a JQ1 down-regulated gene that has a super-enhancer nearby, but is not detected as a super-enhancer target gene since the super-enhancer is located further than 50kb from the transcription start site. Narrow H3K27ac enriched regions that are not super-enhancers are located closer to the transcription start site of this gene.
2. Regulatory potential distance parameter effect on prediction of BET-inhibition in BET-inhibited cells, including 5 DLBCL cells lines (DHL6, HBL1, LY1, LY4, and TOLEDO), one liver cancer cell line (HepG2) and one malignant peripheral nerve sheath tumor cell line (90-8TL). The area under the curve of BET-inhibited genes changes with the decay rate in the regulatory potential function. The optimum range for the  $\frac{1}{2}$  potential lies in the range between 6kb and 20kb.
3. Performance comparison of the regulatory potential, relative regulatory potential and ROSE super-enhancer associated genes in the prediction of BET-inhibited genes. Receiver operator curves in 5 DLBCL cell lines (LY1, LY4, DHL6, HBL1, and TOLEDO) and in the malignant peripheral nerve sheath tumor cell line 90-8TL consistently show that the best and worst methods are respectively the relative regulatory potential and ROSE. Inclusion of topologically associating domains does not impact the prediction performance to a significant degree.
4. BET-inhibited gene prediction performance for the regulatory potential calculated using all reads and a regulatory potential that includes only reads confined to H3K27ac ChIP-seq peak regions. Receiver operator curves in 5 DLBCL cell lines (LY1, LY4, DHL6, HBL1, and TOLEDO) and the liver cancer cell line HepG2 show no significant difference between the two predictions.
5. Gene Ontology analysis of the 500 genes with the highest median regulatory potentials across H3K27ac ChIP-seq compendium of 365 samples.
6. Precision-recall performance comparison of the regulatory potential, relative regulatory potential and ROSE super-enhancer associated genes in the prediction of BET-inhibited genes in DLBCL (LY1, LY4, TOLEDO, DHL6, and HBL1) and in the liver cancer cell line HepG2.
7. Regulatory potentials and relative regulatory potentials correlate with statistical significance (log q-value) of BET-inhibited genes in DLBCL (LY1, LY4, DHL6, HBL1, and TOLEDO) and in the liver cancer cell line HepG2. Boxplots represent the bin median, the interquartile range and whiskers that extend to the most extreme data point that is not more than the interquartile range from the box.
8. Regulatory potentials and relative regulatory potentials correlate with fold-change in gene expression of BET-inhibited genes in DLBCL (LY1, LY4, DHL6, HBL1, and TOLEDO) and in the liver cancer cell line HepG2. Boxplots represent the bin median, the interquartile

range and whiskers that extend to the most extreme data point that is not more than the interquartile range from the box.

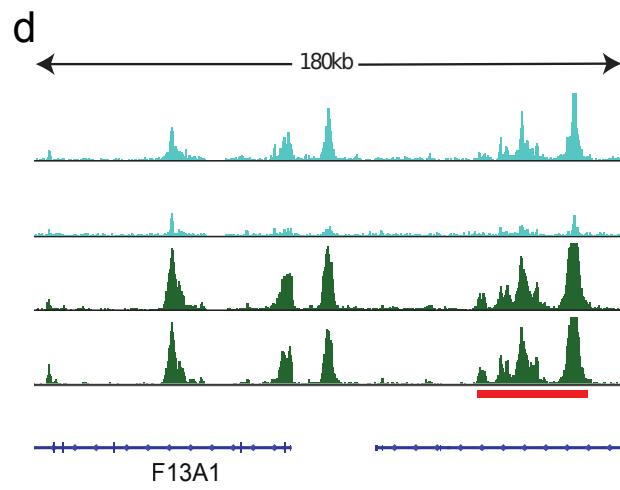
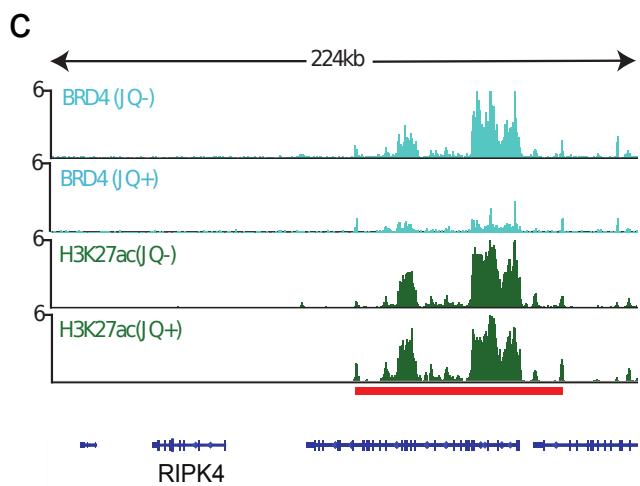
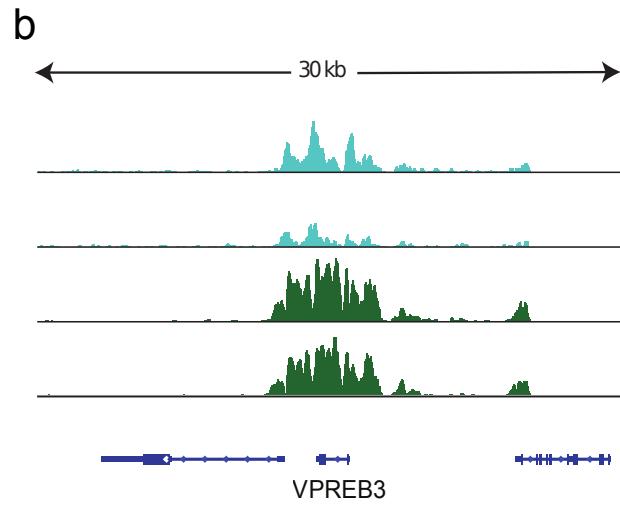
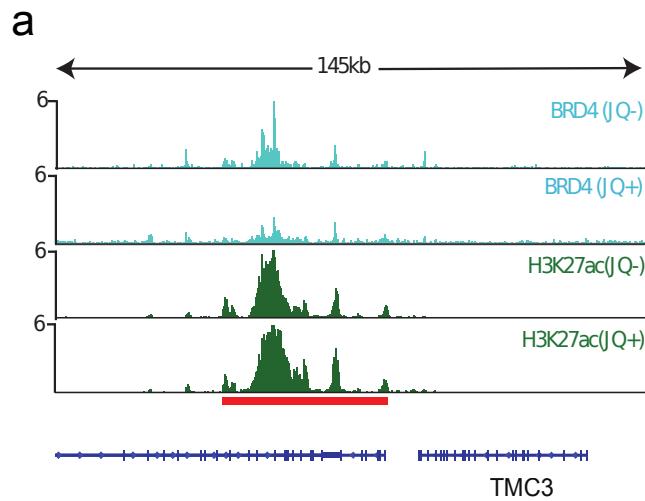
9. Gene Ontology analysis of the 500 genes with the highest relative regulatory potential or ROSE super-enhancer index in 14 diverse H3K27ac samples. Genes with high relative regulatory potentials are more highly enriched in cell type related GO categories than super-enhancer associated genes.
10. Number of samples selected by MARGE-express. Cross-validation AUCs increase then level off as more samples are added to the MARGE-express logistic regression model.
11. MARGE reweighted H3K27ac profiles in the identification of AR binding sites from testosterone induced genes in the LNCaP prostate cancer cell line.

## SUPPLEMENTARY TABLES

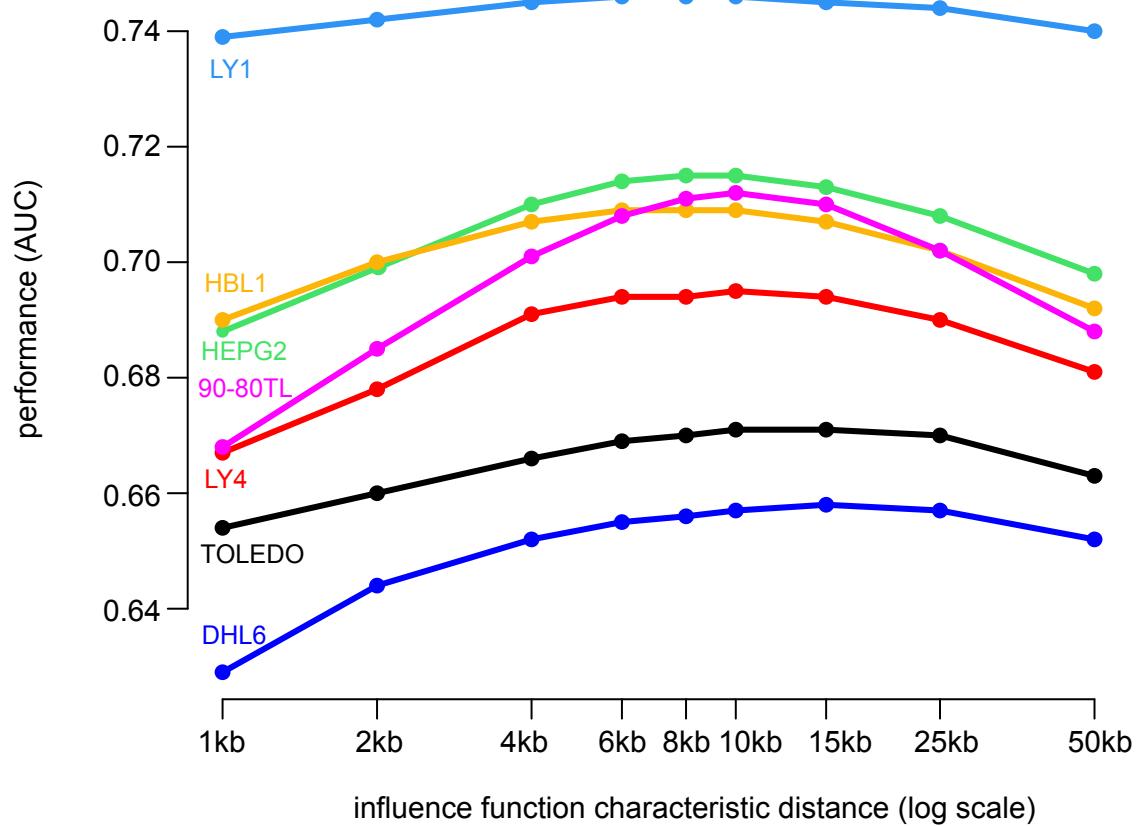
1. Description of human H3K27ac ChIP-seq data sets. Table includes the following information for each H3K27ac ChIP-seq sample: internal dataset identifier; cell lineage; whether the sample is derived from a cell line or from tissue; whether the sample is cancer or normal; the cell line name; tissue/organ; the cell line classification; detailed tissue type; Gene Expression Omnibus (GEO) accession numbers for ChIP and control samples.
2. Description of mouse H3K27ac ChIP-seq data sets. Table includes the following information for each H3K27ac ChIP-seq sample: internal dataset identifier; cell lineage; whether the sample is derived from a cell line or from tissue; whether the sample is cancer or normal; the cell line name; tissue/organ; the cell line classification; detailed tissue type; GEO accession numbers for ChIP and control samples.
3. Relative regulatory potentials derived from human H3K27ac ChIP-seq data sets. Table includes the following information: chromosome; position of transcription start site on human reference genome hg38; RefSeq ID; GeneSymbol; strand; median regulatory potential across H3K27ac samples in the human compendium; relative regulatory potentials of samples indexed by sample ID (to be cross-referenced with Supplementary Table S1).
4. Relative regulatory potentials derived from mouse H3K27ac ChIP-seq data sets. Table includes the following information: chromosome, position of transcription start site on mouse reference genome mm10; RefSeq ID; GeneSymbol; strand; median regulatory potential across H3K27ac samples in mouse compendium; relative regulatory potentials of samples indexed by sample ID (to be cross-referenced with Supplementary Table S2).
5. Median relative regulatory potentials of transcription factors, chromatin regulators and kinases across human neuronal, lymphoblastoid, and embryonic stem cell types.
6. Samples selected by MARGE-express for *AR*, *ESR1*, *GR*, *PPARG*, *POU5F1* and *NOTCH* examples. Table includes sample IDs (to be cross-referenced with Supplementary Table S1), MARGE-express model coefficients, cell line, cell type, tissue, condition, cancer type and GEO accession numbers.
7. MARGE-express predictions for up-regulated gene sets in *AR*, *ESR1*, *GR*, *PPARG*, *POU5F1* and *NOTCH* examples. Table includes Gene Symbol, RefSeq ID, MARGE-express prediction values and whether (1) or not (0) the gene is in the input gene set.

8. MARGE-cistrome predictions for *AR*, *ESR1*, *GR*, *PPARG*, *POU5F1* (*OCT4*) and *NOTCH* examples. Table includes chromosome, start position of DHS region (hg38), end position of DHS region, DHS identifier, MARGE-cistrome predictions for *AR*, *ESR1*, *GR*, *PPARG*, *POU5F1* (*OCT4*) and *NOTCH*, and whether (1) or not (0) DHS regions overlap with ChIP-seq peaks for the respective factors.
9. Samples selected by MARGE-express for LNCaP-abl siRNA knockdown experiments. Table includes sample IDs (to be cross-referenced with Supplementary Table S1), MARGE-express model coefficients, cell line, cell type, tissue, condition, cancer type and GEO accession numbers.
10. MARGE-express predictions for LNCaP-abl siRNA knockdown experiments of *AR*, *E2F1*, *EZH2*, *FOXA1*, *FOXM1*, *KDM1A*, *MALAT1*, *RAD21*, and *UTX*. Table includes RefSeq ID, GeneSymbol, MARGE-express predictions and whether (1) or not (0) the gene is in the input gene sets. MARGE-express predictions based on public H3K27ac data alone and public data combined with in-house H3K27ac ChIP-seq data are provided.
11. MARGE-cistrome predictions for *AR*, *E2F1* and *FOXA1* siRNA knockdown experiments in LNCaP-abl cells. Table includes DHS ID, chromosome, center position of DHS region (hg38), MARGE-cistrome predictions for *AR*, *E2F1* and *FOXA1* binding, and whether (1) or not (0) DHS regions overlap with ChIP-seq peaks for the respective factors. MARGE-cistrome predictions based on public H3K27ac data alone and public data combined with in-house H3K27ac ChIP-seq data are provided.

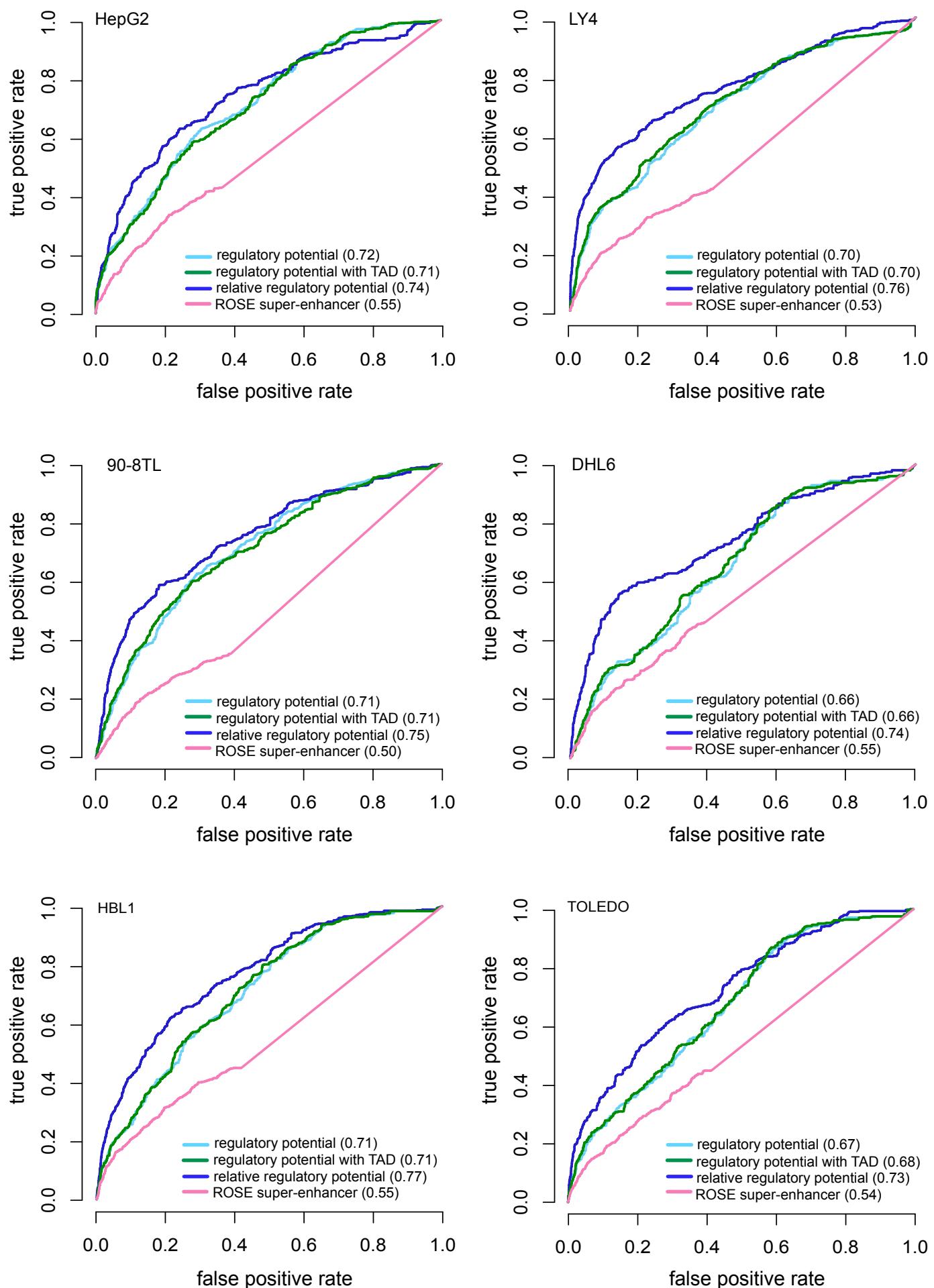
# Supplementary Figure 1



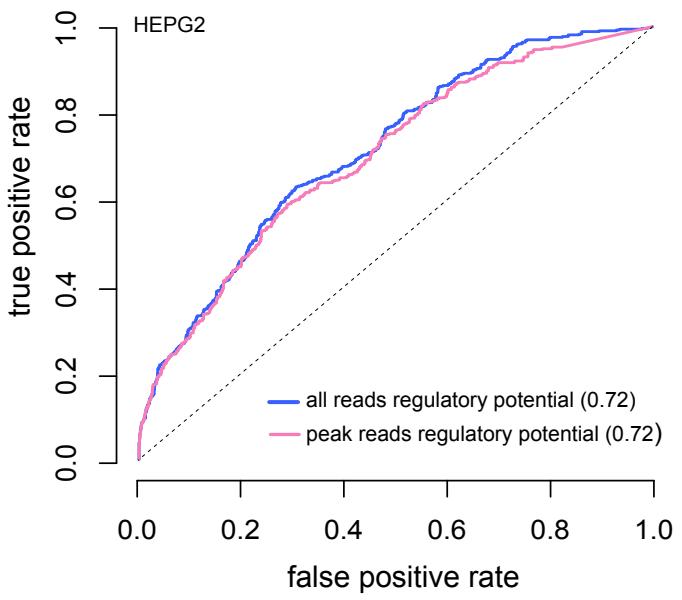
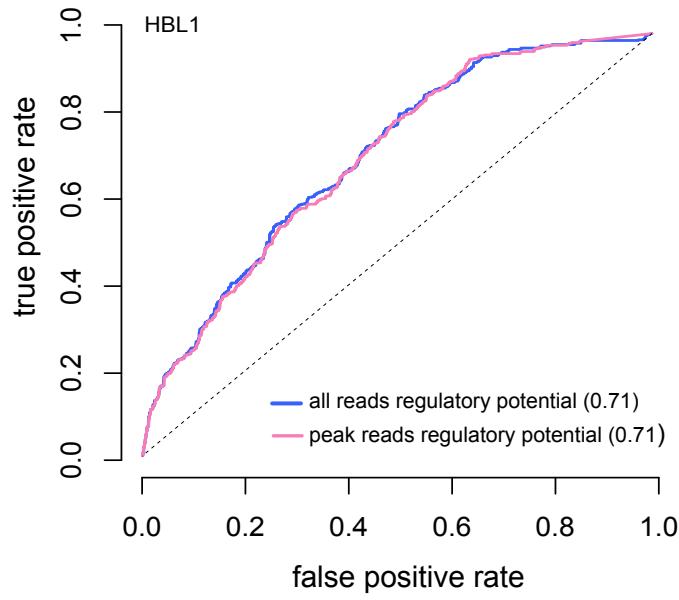
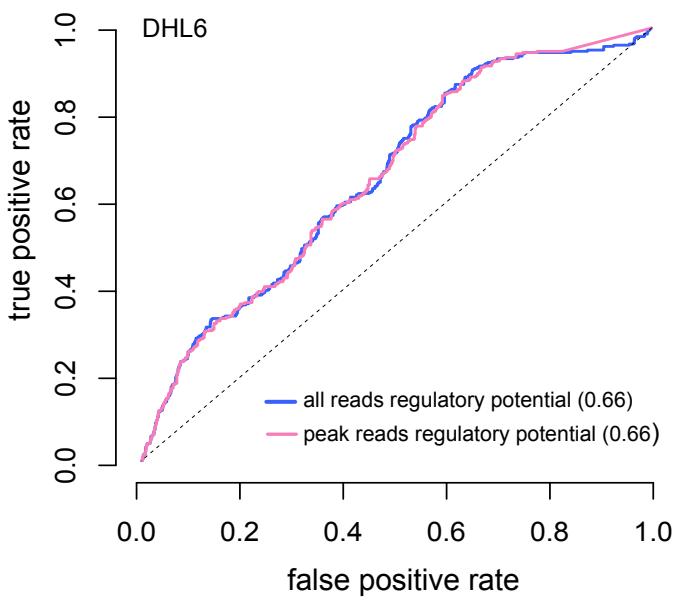
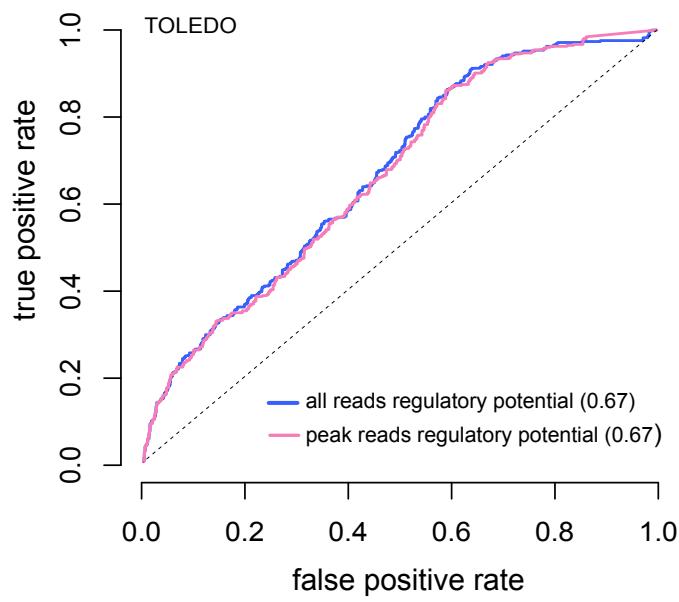
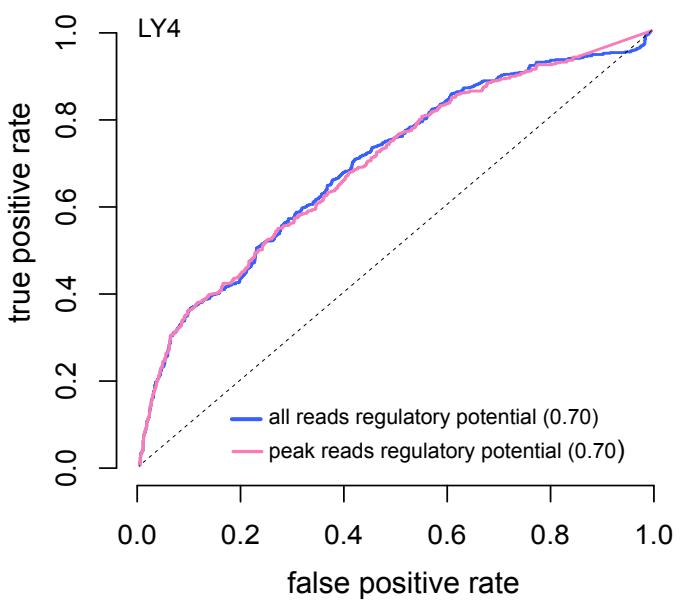
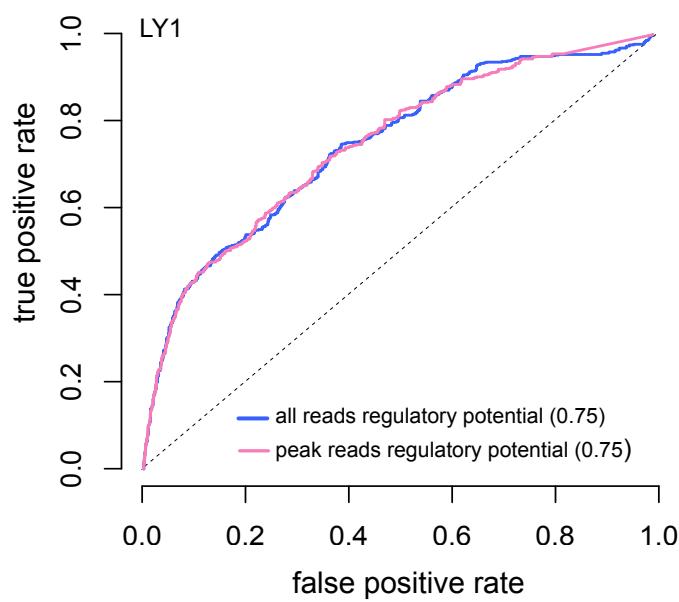
## Supplementary Figure 2



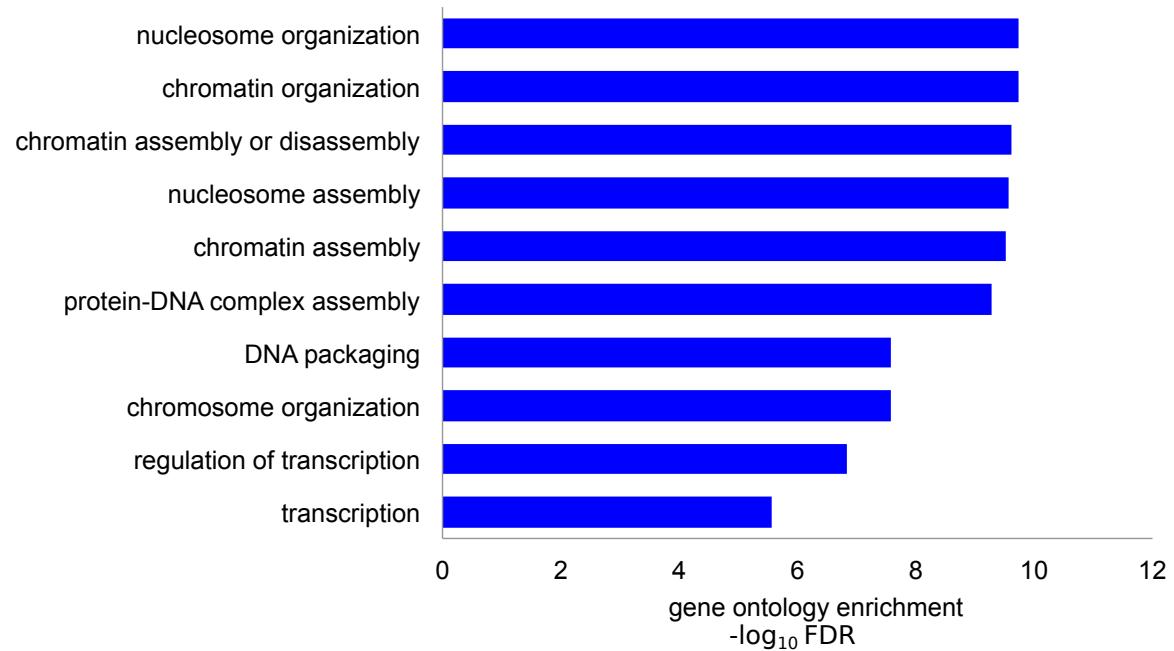
# Supplementary Figure 3



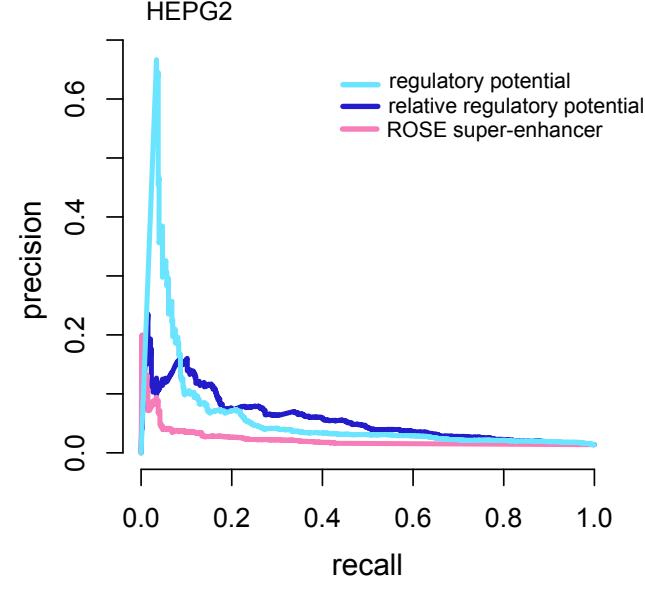
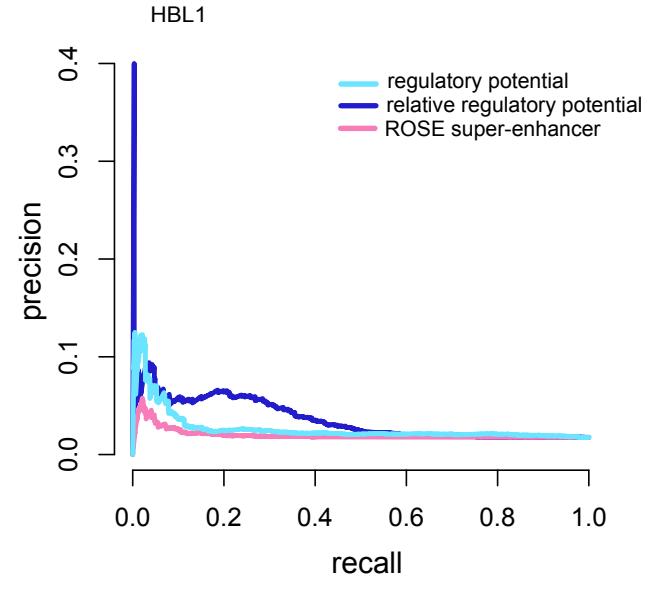
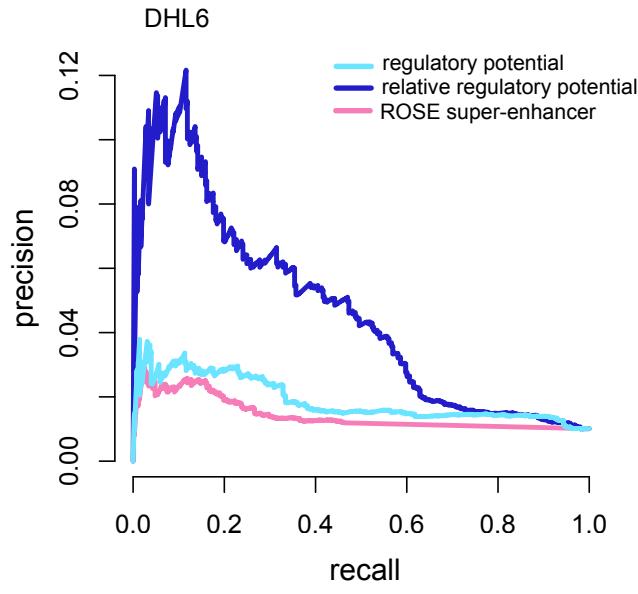
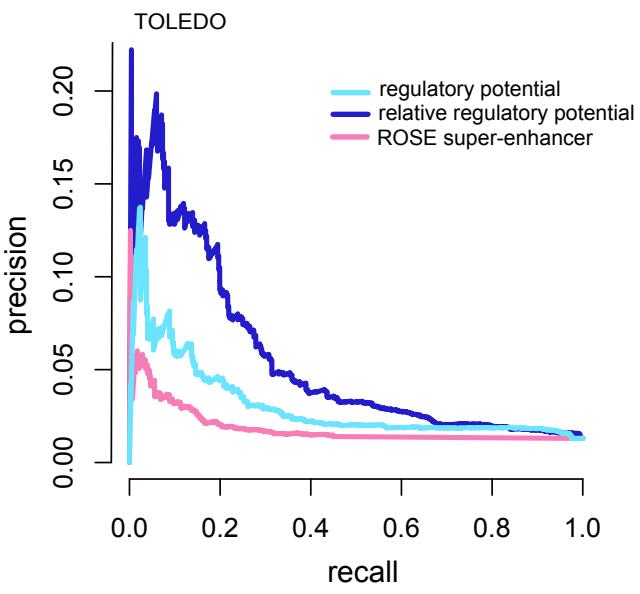
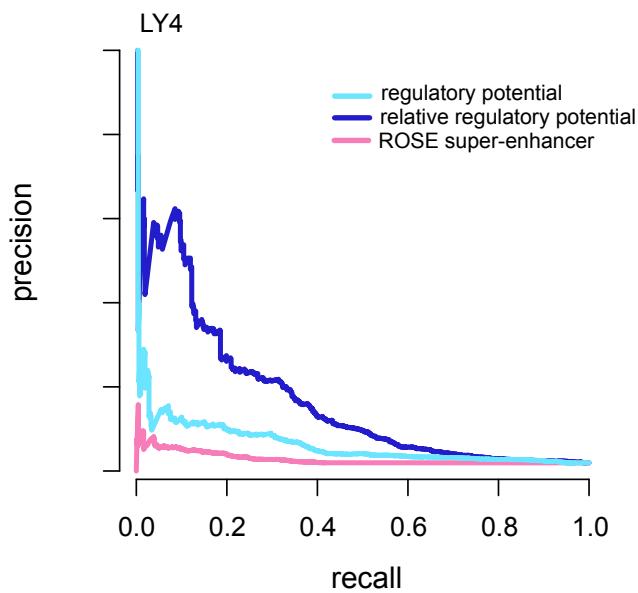
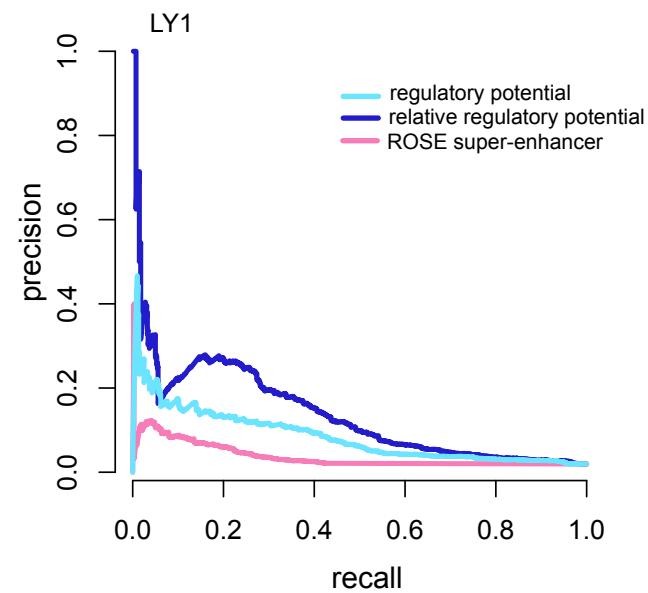
## Supplementary Figure 4



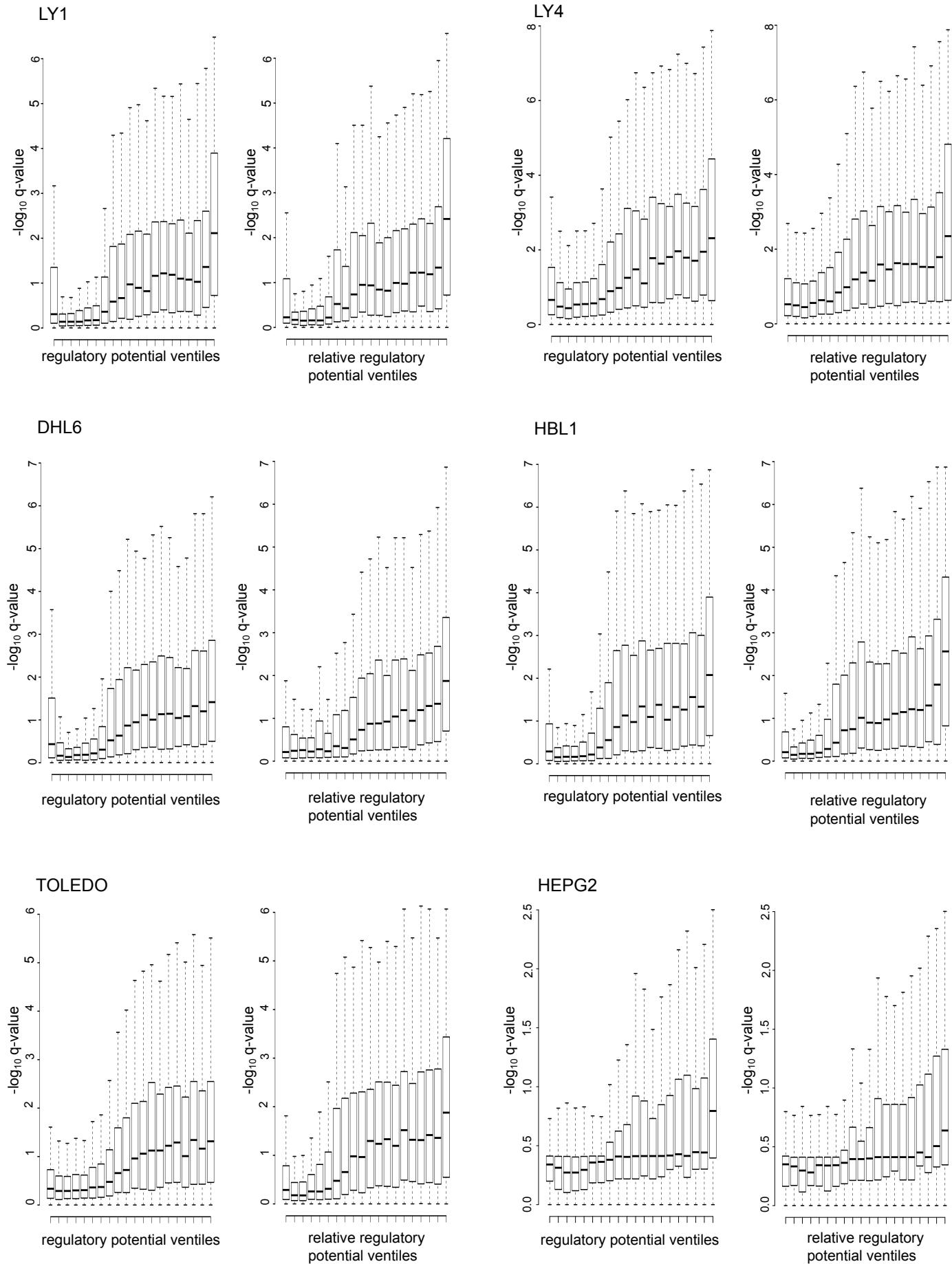
## Supplementary Figure 5



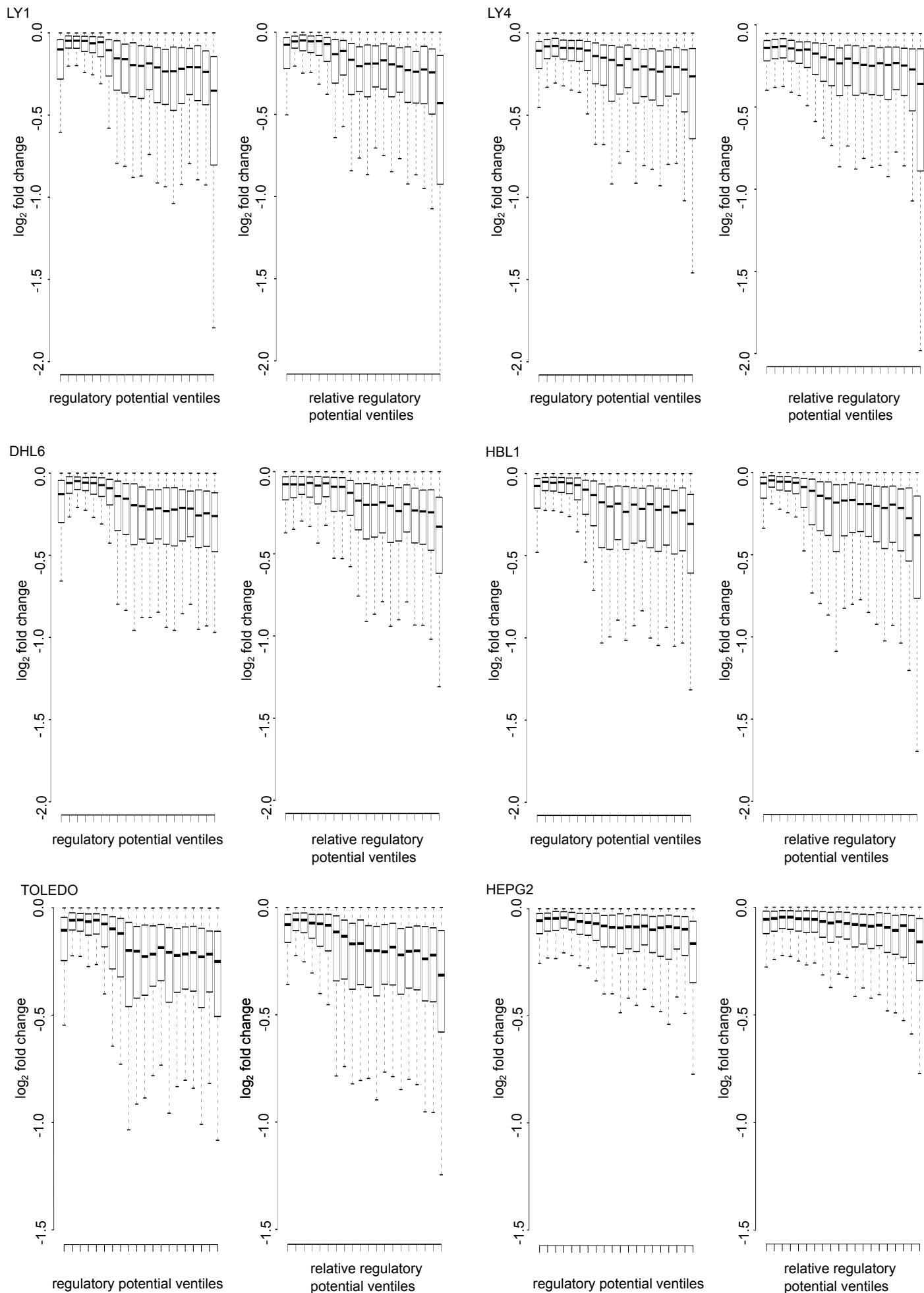
## Supplementary Figure 6



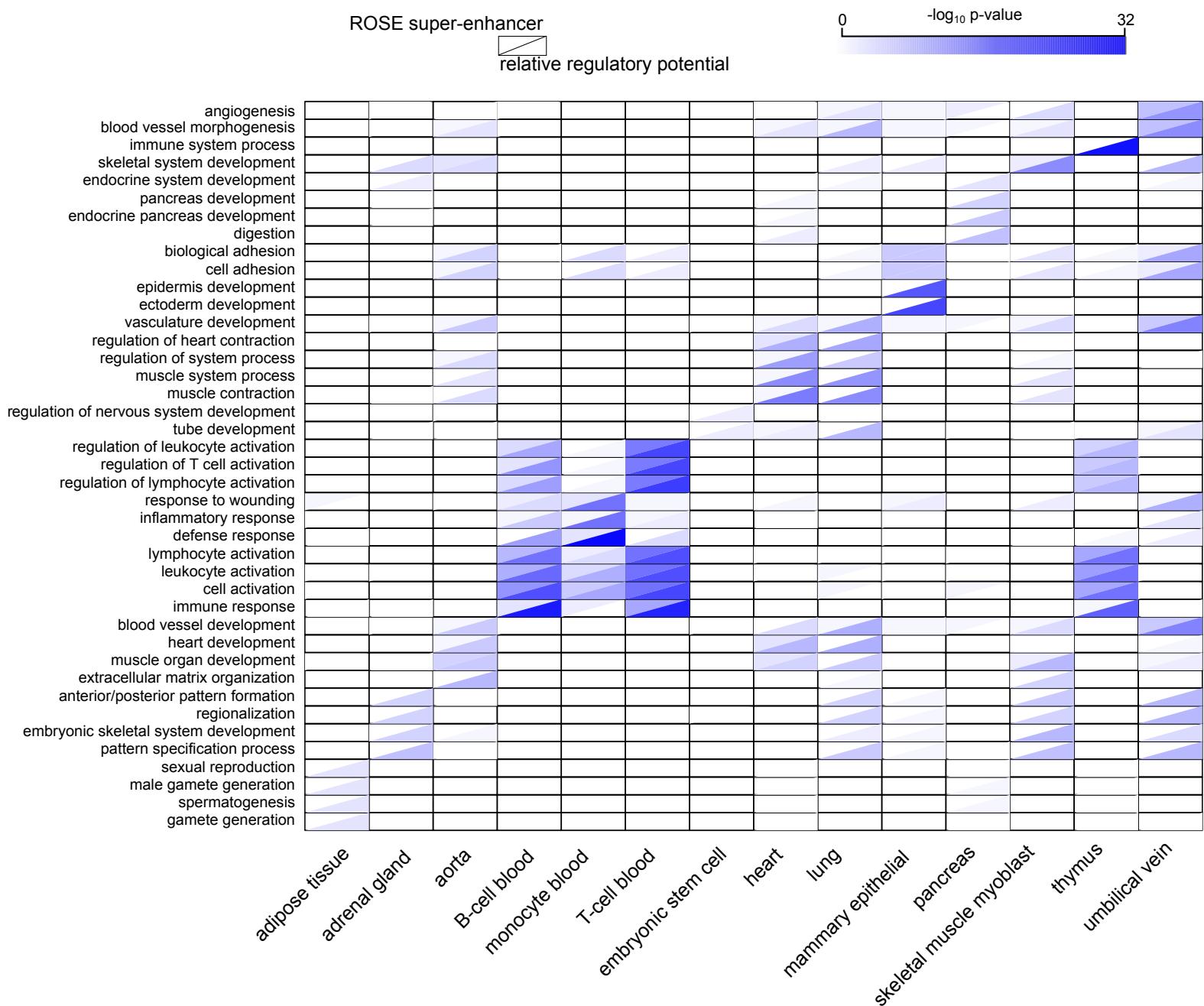
## Supplementary Figure 7



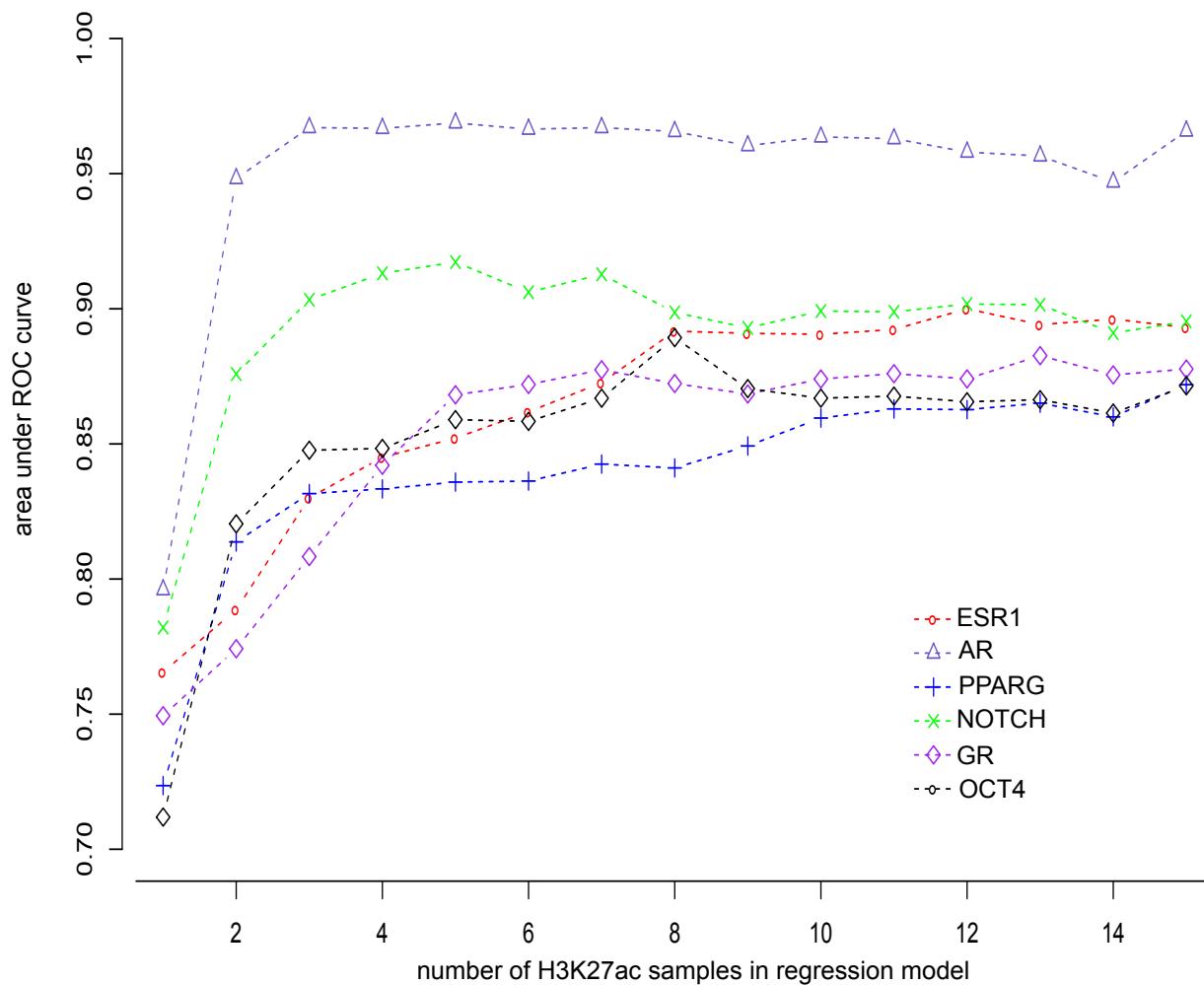
## Supplementary Figure 8



# Supplementary Figure 9



## Supplementary Figure 10



## Supplementary Figure 11

