

## General Motif Characteristics:

### Top Similar Looking Motifs:

ENCODE.ZNF143-ext	0.128	
HOMER.GFY-Staf.Promoters	0.259	
HCMC.RUNX3.si.wm	0.271	
ENCODE.CTCF-ext	0.288	
RBPJ.p2	0.309	
ARID5B.p2	0.318	
HTSELEX.SOX8.HMG.full.dimeric.wm2	0.322	
HCMC.DDIT3_f1.wm	0.327	
HTSELEX.RUNX3.RUNX.DBD.dimeric.wm1	0.329	
HTSELEX.SOX8.HMG.DBD.dimeric.wm2	0.329	

### Statistics:

- Motif name: denovo\_WM\_6
- Enrichment: 32.8072228599
- log-Likelihood Ratio: 1745.32435015
- Area under precision recall curve: 0.9061

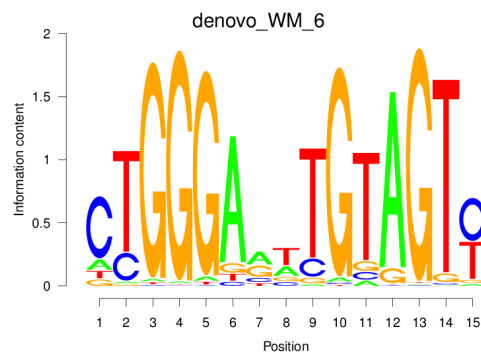


Figure 1: Logo of Weight Matrix

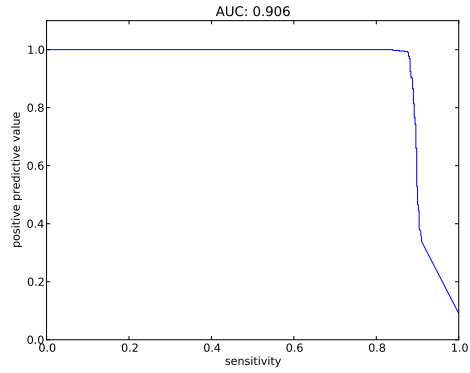


Figure 2: Precision and Recall curve

## Statistics for TFBS - Peak Centering and Peak Posterior - Peak Height Correlation:

### 0.1

Overall statistics:

- Number of true peaks out of total number of peaks: 3129/36955
- 8.47 percent are true.
- Cut-off: minimum summed posterior of 0.20
- Peak plots contain TFBS of posterior  $\geq 0.20$

Statistic for centering of TFBSs at peak centers:

- Enrichment at binding sites: 20.9514

Correlation between peak Z-score and number of binding sites at peak: 0.6123

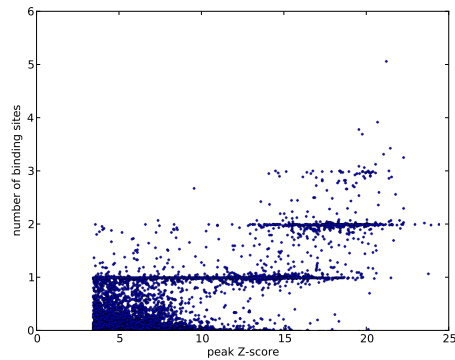


Figure 3: Scatter plot of peak Z-score versus number of binding sites at peak

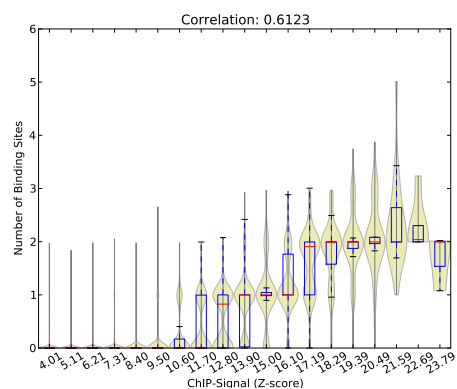


Figure 4: Violin plot of peak Z-score versus number of binding sites at peak

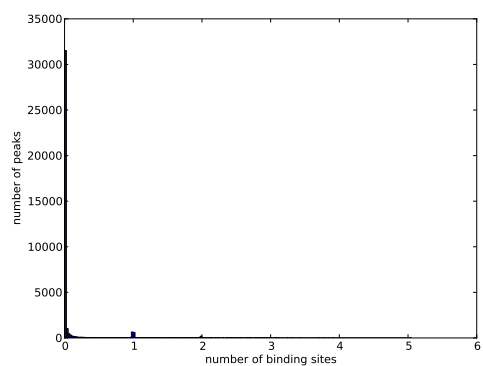


Figure 5: Histogram of number of binding sites at peaks

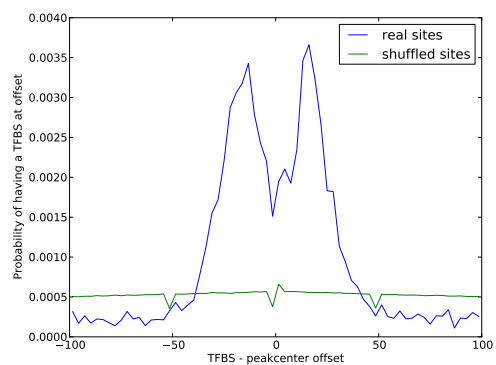


Figure 6: Histogram of distances of TFBS to its nearest peak center

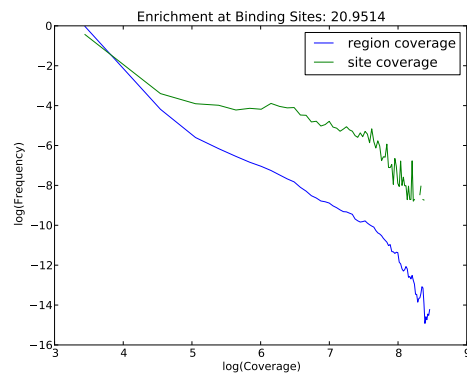


Figure 7: Histograms of coverage at sites and in total regions