

## General Motif Characteristics:

### Top Similar Looking Motifs:

ENCODE.CTCF	0.007	
HCMC.CTCF_f2.wm	0.068	
JASPAR.CTCF.wm	0.070	
HOMER.BORIS(Zf).K562-CTCF-Seq	0.077	
CTCF.p2	0.082	
HOMER.CTCF(Zf).CD4-CTCF-Seq	0.088	
HTSELEX.CTCF.C2H2.full.monomeric.wm1	0.106	
UNIPROBE.UW.Motif.0003.wm	0.116	
TFAP2B.p2	0.250	
UNIPROBE.Zic1_primary.wm	0.255	

### Statistics:

- Motif name: denovo\_WM\_22
- Enrichment: 22.8400308209
- log-Likelihood Ratio: 1564.25736711
- Area under precision recall curve: 0.9325

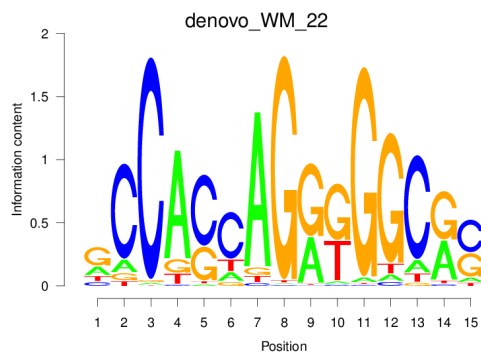


Figure 1: Logo of Weight Matrix

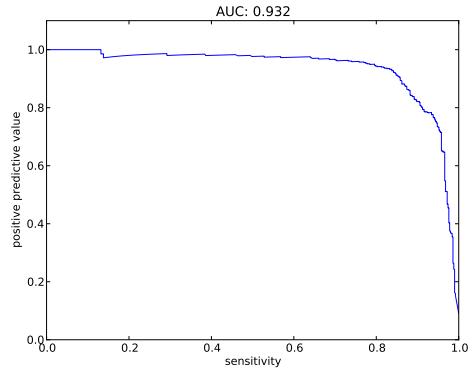


Figure 2: Precision and Recall curve

## 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: 30640/44017
- 69.61 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: 4.3159

Correlation between peak Z-score and number of binding sites at peak: 0.4698

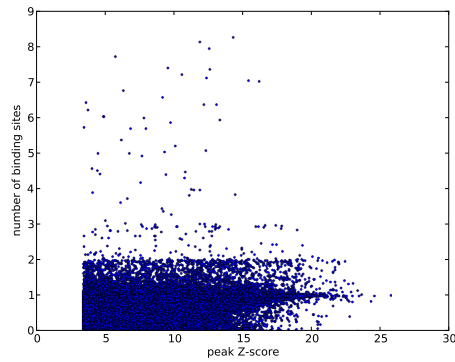


Figure 3: Scatter plot of peak Z-score versus number of binding sites at peak

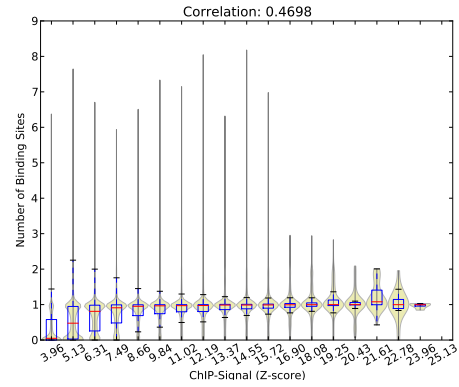


Figure 4: Violin plot of peak Z-score versus number of binding sites at peak

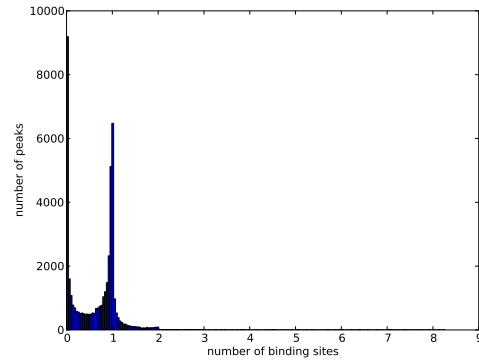


Figure 5: Histogram of number of binding sites at peaks

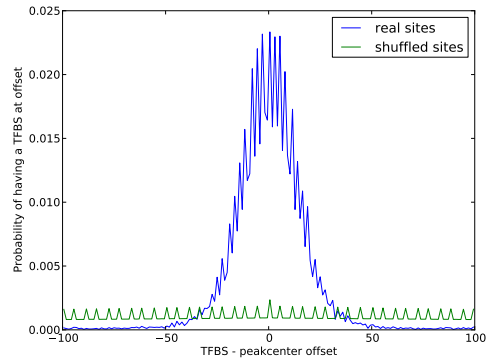


Figure 6: Histogram of distances of TFBS to its nearest peak center

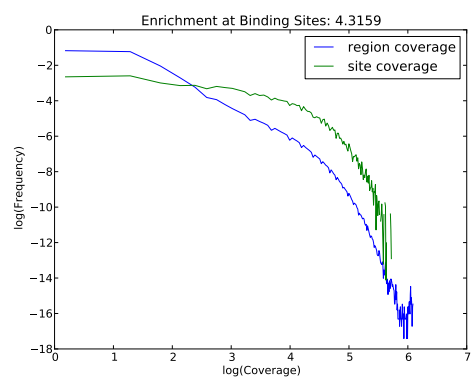


Figure 7: Histograms of coverage at sites and in total regions