

## General Motif Characteristics:

### Top Similar Looking Motifs:

CTCF.p2	0.000
JASPAR.CTCF.wm	0.018
HCMC.CTCF_f2.wm	0.022
HOMER.CTCF(Zf).CD4+CTCF-ChIP-Seq	0.028
HOMER.BORIS(Zf).K562-CTCF-ChIP-Seq	0.091
ENCODE.CTCF	0.092
HTSELEX.CTCF.C2H2.full.monomeric.wm1	0.124
UNIPROBE.UW.Motif.0003.wm	0.186
HTSELEX.Hic1.C2H2.DBD.dimeric.wm1	0.278
TFAP4.p2	0.279

### Statistics:

- Motif name: CTCF.p2
- Enrichment: 28.6647235043
- log-Likelihood Ratio: 1677.83361028
- Area under precision recall curve: 0.9099

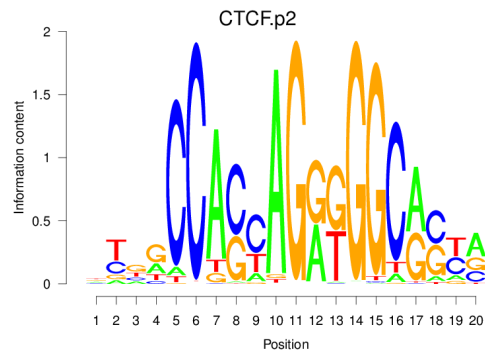


Figure 1: Logo of Weight Matrix

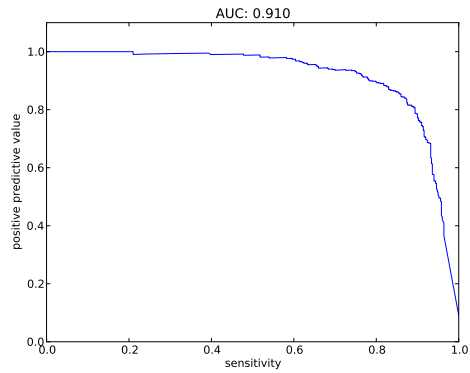


Figure 2: Precision and Recall curve

## 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: 32452/63141
- 51.40 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: 5.229

Correlation between peak Z-score and number of binding sites at peak: 0.5032

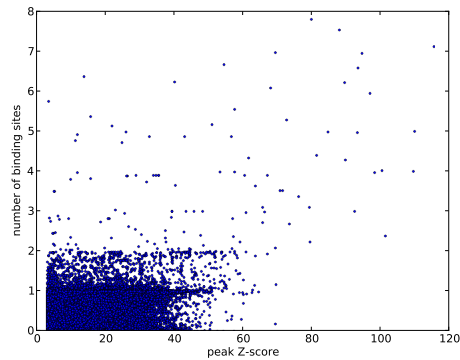


Figure 3: Scatter plot of peak Z-score versus number of binding sites at peak

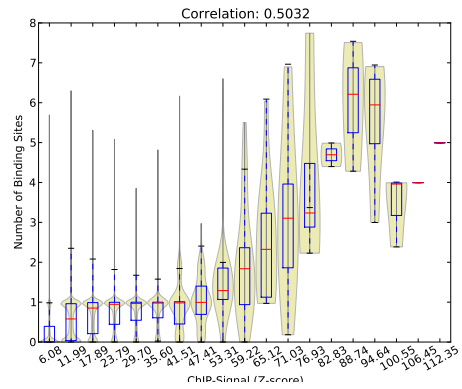


Figure 4: Violin plot of peak Z-score versus number of binding sites at peak

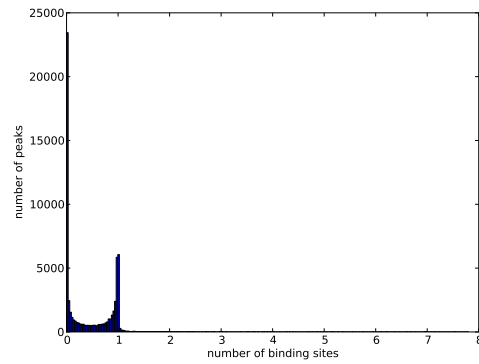


Figure 5: Histogram of number of binding sites at peaks

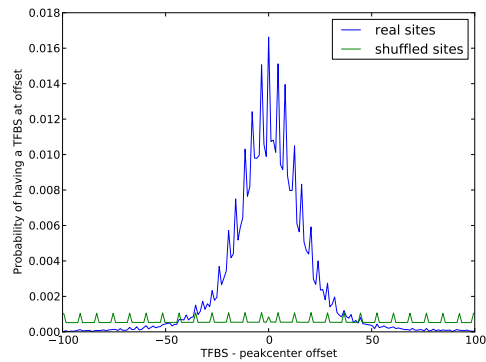


Figure 6: Histogram of distances of TFBS to its nearest peak center

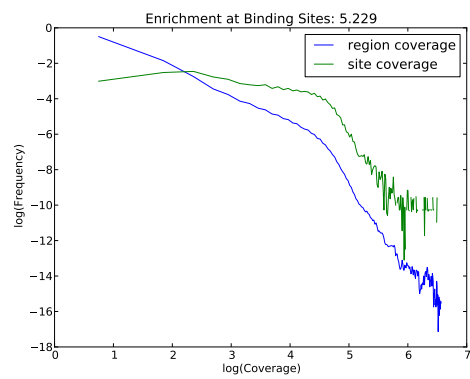


Figure 7: Histograms of coverage at sites and in total regions