De novo motif discovery using MEME-ChIP on [-50bp, 50bp] centered on the peak summit for top 500 peaks ranked by ChIP signal, up to 5 motifs discovered for each dataset.

Training Set

Motif scan using FIMO on [-150bp, 150bp] centered on the peak summit for 501-1000 ranked peaks.

Testing Set 1
Number of peaks with motif, T1

Motif scan using FIMO on 500 GC-matched random regions in the genome excluding the peak regions.

Control Set 1
Number of regions with motif; randomly sample 100 times. mean μ and standard deviation σ

Motif scan using FIMO on [-150bp, 150bp] centered on the peak summit for all peaks ranked 501 and beyond.

Testing Set 2
% of peaks with motif, T2

Motif scan using FIMO on [-450bp,-150bp] and [150bp,450bp] flanking the peak summit for all peaks ranked 501 and beyond.

Control Set 2
% of flanking region with motif, C2

Yes
further analysis

T2 >= 10% AND T2 / C2 >=1.25

No
discard the motif

T2 >= 10% AND T2 / C2 >=1.25

No
discard the motif

No
discard the motif

Wang et al., Figure S1

Figure S1: Flow chart of the de novo motif discovery pipeline.