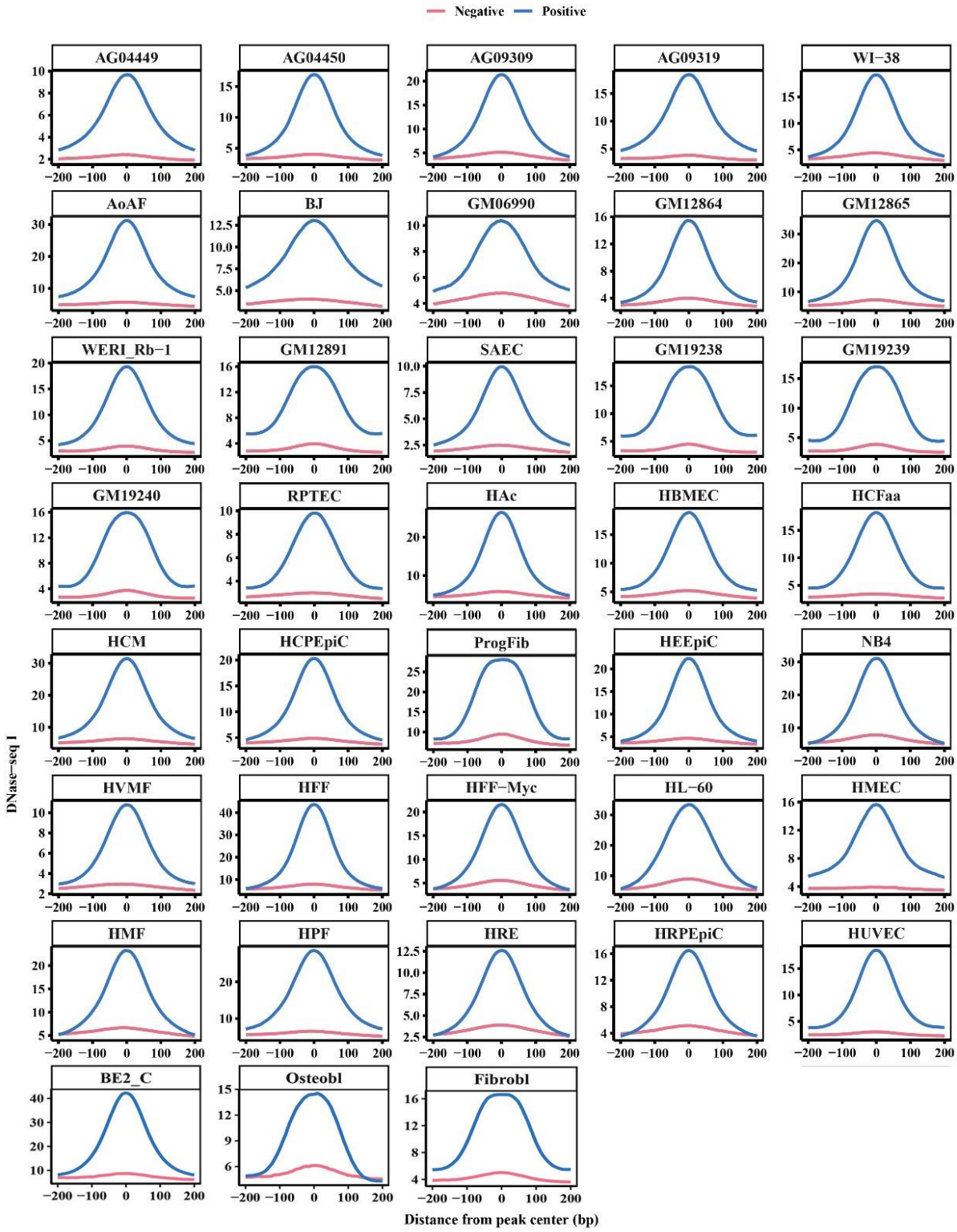
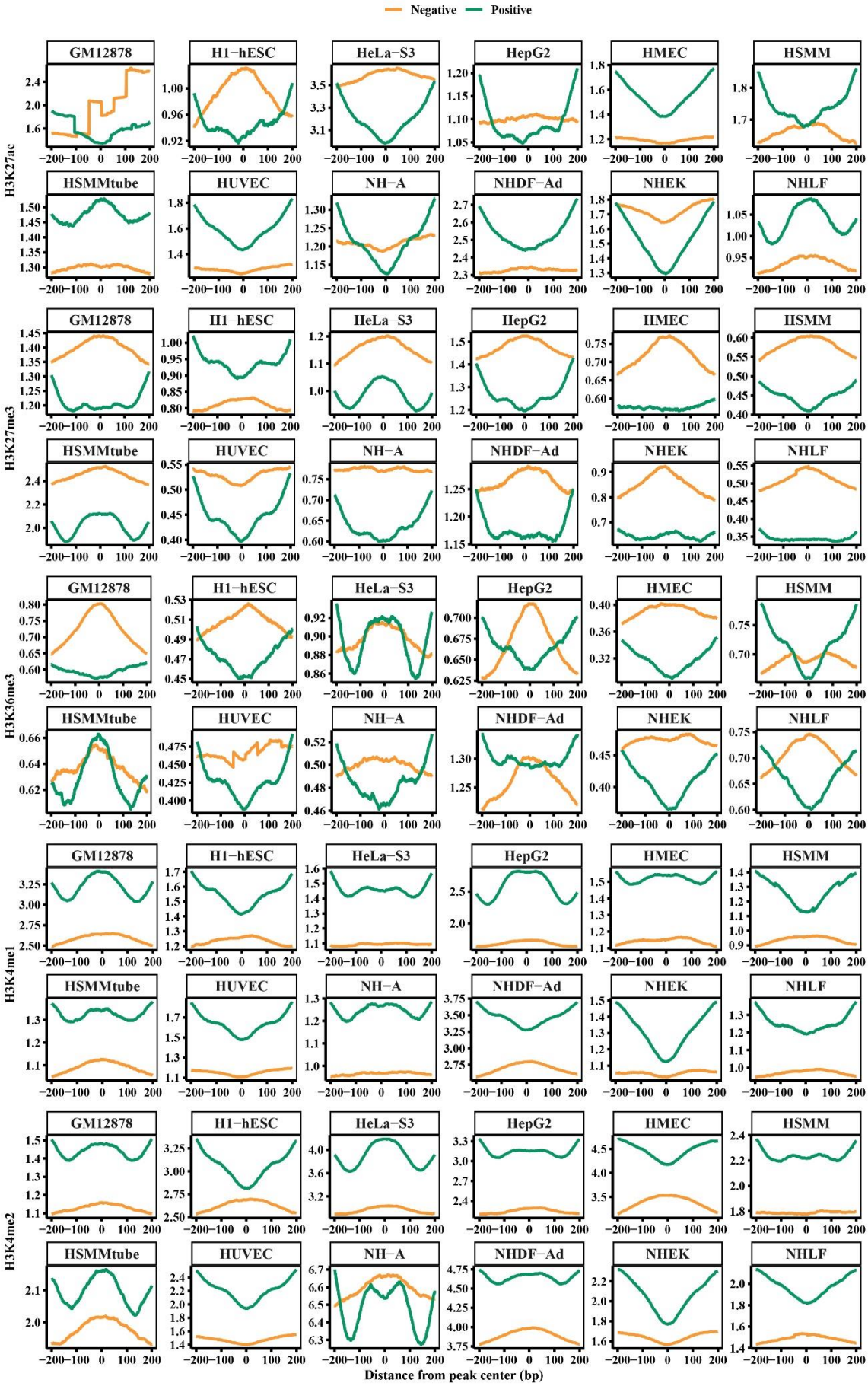


Figure S1. Comparison of signal distribution of Chromatin accessibility in CTCF core binding zone between the positive and negative sets (38 cell lines).


Negative
Positive
H3K27ac
H3K27me3
H3K36me3
H3K4me1
H3K4me2
GM12878
H1−hESC
HeLa−S3
HepG2
HMEC
HSMM
HSMMtube
HUVEC
NH−A
NHDF−Ad
NHEK
NHLF
−200 −100 0 100 200
Distance from peak center (bp)

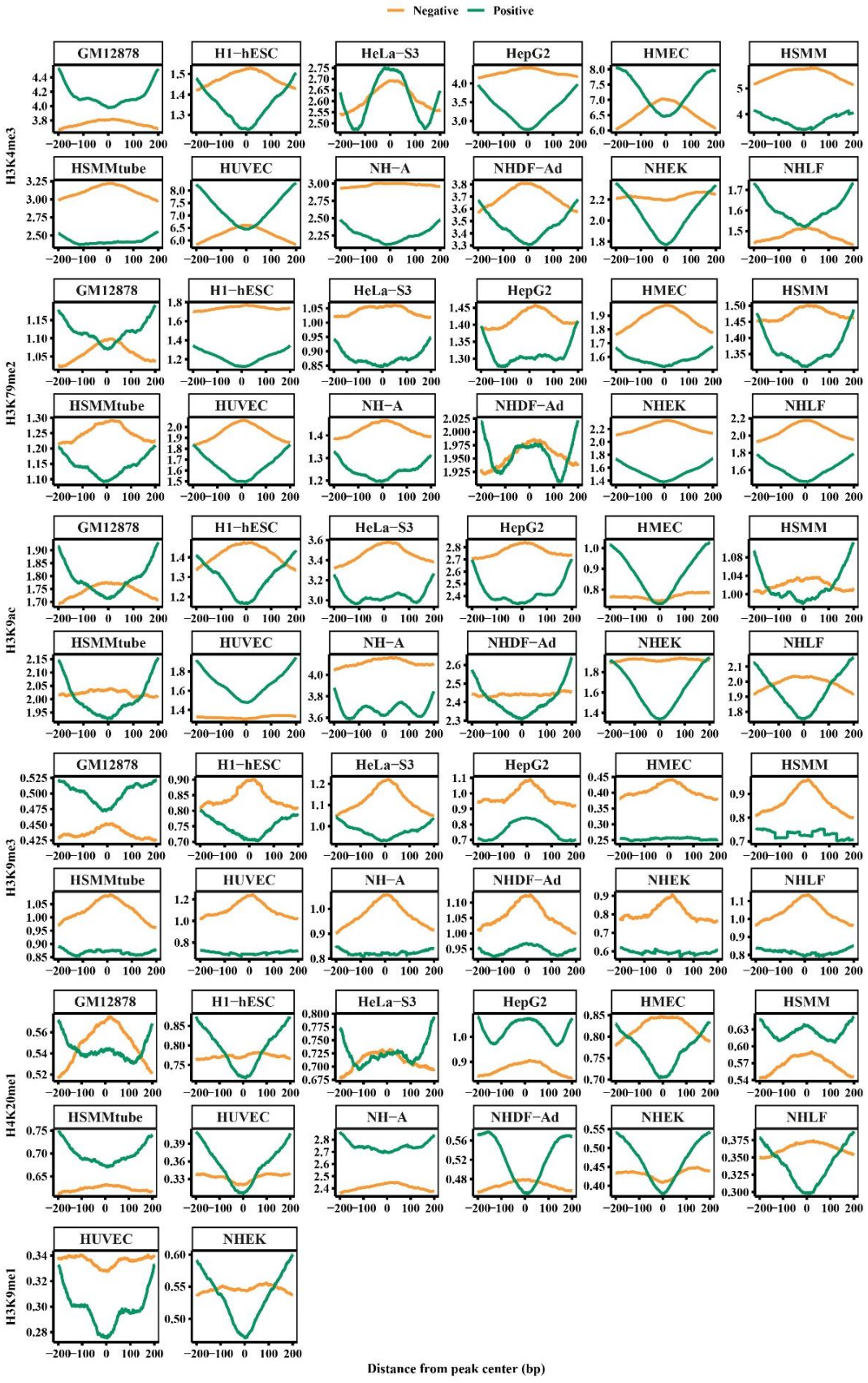


Figure S2. Comparison of signal distribution of 11 HMs in CTCF core binding zone between the positive and negative sets.

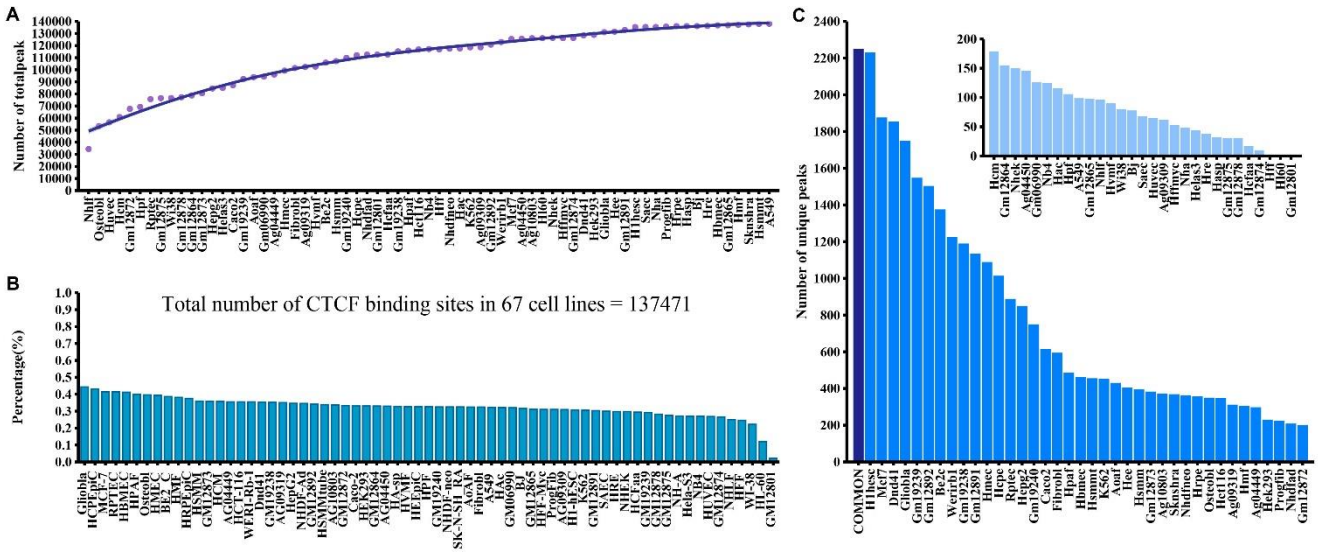


Figure S3. The peak number in 67 cell lines. (A) The peak number increases with the addition of each cell line. The arrangement of cell lines on the horizontal axis is random. (B) The percentage of peaks in one cell line to total peaks in the Complete_peak dataset. The abscissa displays the names of cell lines, and the ordinate is the percentage. (C) The number of peaks shared across 67 cell lines or unique to one cell line. The abscissa displays the names of cell lines, where “COMMON” refers to peaks shared across 67 cell lines. The ordinate gives the number of peaks in the Uni_peak dataset of each cell line.

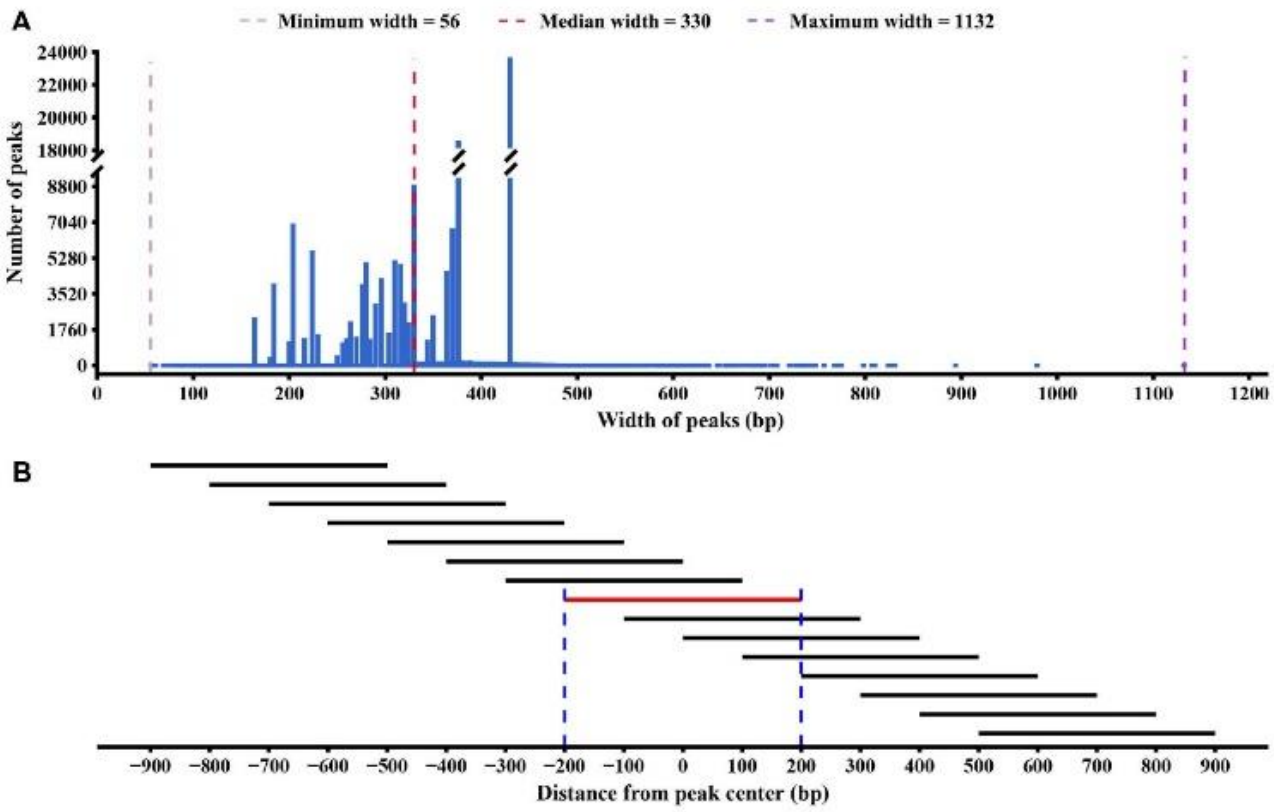


Figure S4. Determination of feature collection region. (A) The width distribution of peaks provided by the ChIP-seq data in the ENCODE project. The maximum, minimum, and median widths are showed in the top. (B) The 15 windows for extracting features, ranging from 900 bp upstream to 900 bp downstream. The red window represents the CTCF core binding zone.