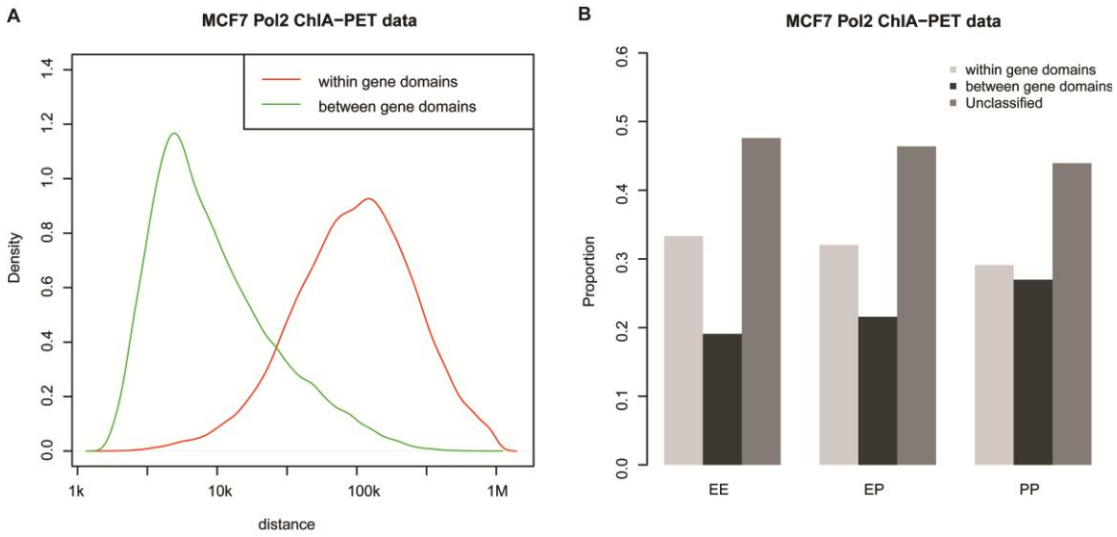


Figure S1. (A) Log10-distance distribution between the two ends of interaction cluster within gene domains (Group A) and between gene domains (Group B). (B) Proportion of EE, EP and PP interactions that belong to Group A or Group B interactions. We defined regulatory domain of a gene as the two flanking regions (-10kb upstream of TSS and 10kb downstream of 3'UTR) plus gene body. Interactions within gene regulatory domains (denoted as Group A interactions) were those that both ends were located within the regulatory domain of the same gene and not overlapping with regulatory domain of any other genes. Interactions between gene domains (denoted as Group B interactions) were those that crossed at least two promoters of different genes.

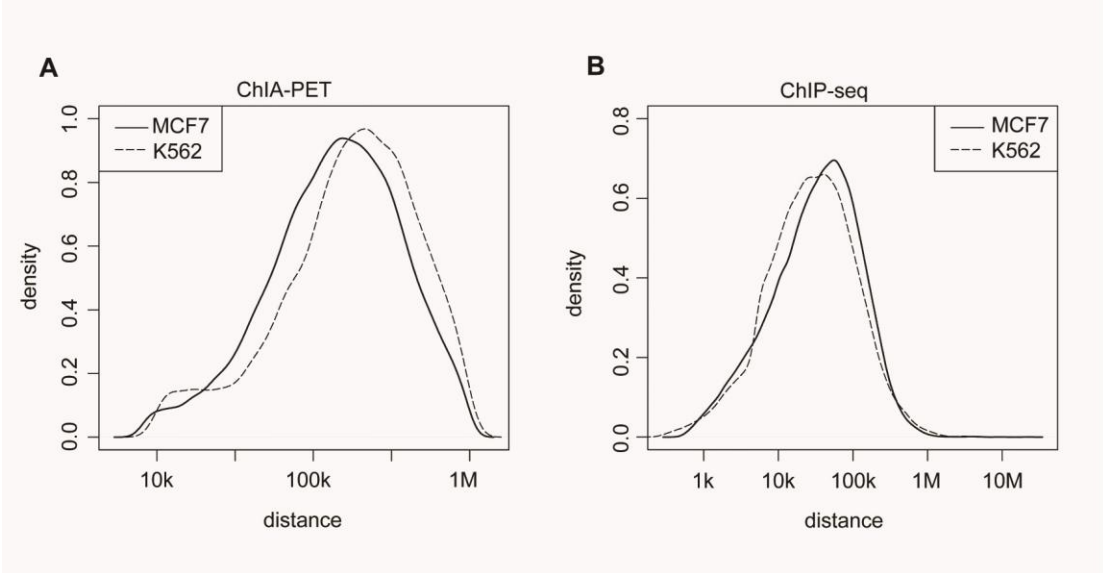


Figure S2. (A) Log10-distance distribution of the two ends of PETs 2+ clusters of MCF7 and K562 CTCF ChIA-PET data. Interactions overlapping with H3K4me1 or H3K4me3 peaks were removed. The peak position was 5.2 for MCF7 and 5.3 for K562. (B) Log10-distance distribution of neighboring CTCF binding sites that did not overlap with either H3K4me1 or H3K4me3 peaks in MCF7 and K562 cell lines. The peak position was 4.7 for MCF7 and 4.6 for K562.

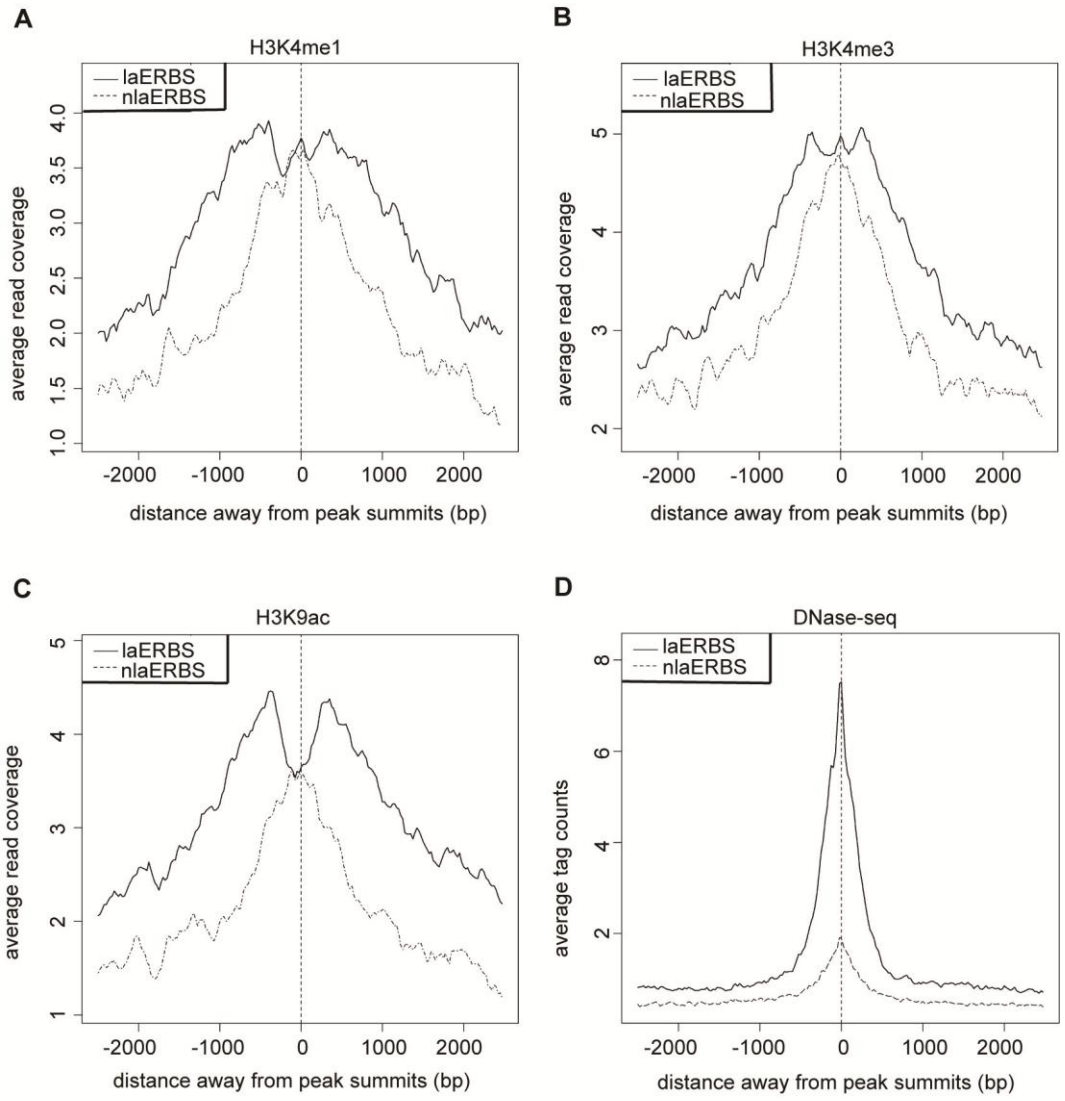
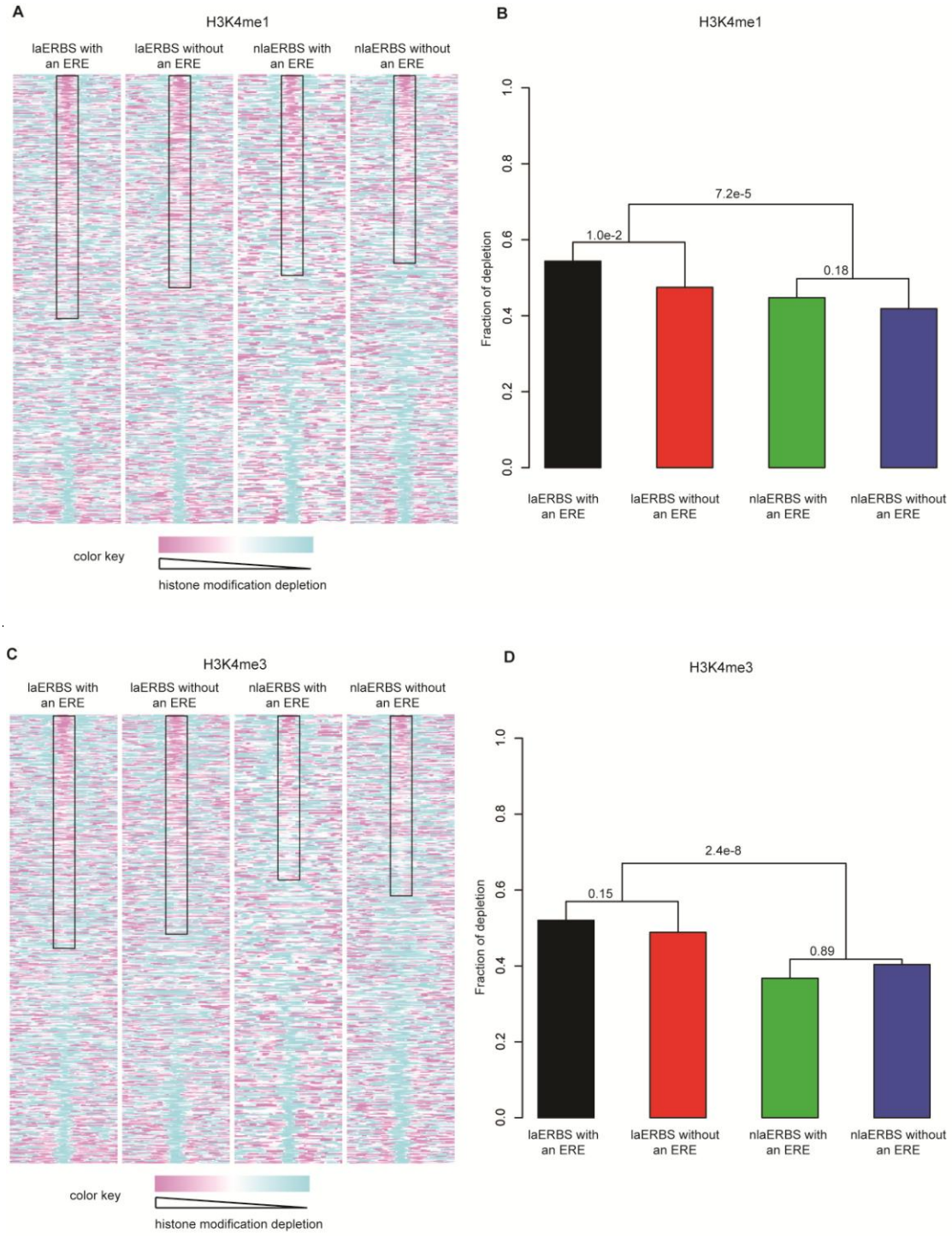


Figure S3. Average H3K4me1 (A), H3K4me3 (B) and H3K9ac (C) read coverage against input surrounding laERBSs and nlaERBSs. Average DNase-seq tag counts surrounding laERBSs and nlaERBSs (D).


A
H3K4me1
laERBS with an ERE
laERBS without an ERE
nlaERBS with an ERE
nlaERBS without an ERE
color key
histone modification depletion
B
H3K4me1
Fraction of depletion
0.0
0.2
0.4
0.6
0.8
1.0
7.2e-5
1.0e-2
0.18
laERBS with an ERE
laERBS without an ERE
nlaERBS with an ERE
nlaERBS without an ERE
C
H3K4me3
laERBS with an ERE
laERBS without an ERE
nlaERBS with an ERE
nlaERBS without an ERE
color key
histone modification depletion
D
H3K4me3
Fraction of depletion
0.0
0.2
0.4
0.6
0.8
1.0
2.4e-8
0.15
0.89
laERBS with an ERE
laERBS without an ERE
nlaERBS with an ERE
nlaERBS without an ERE

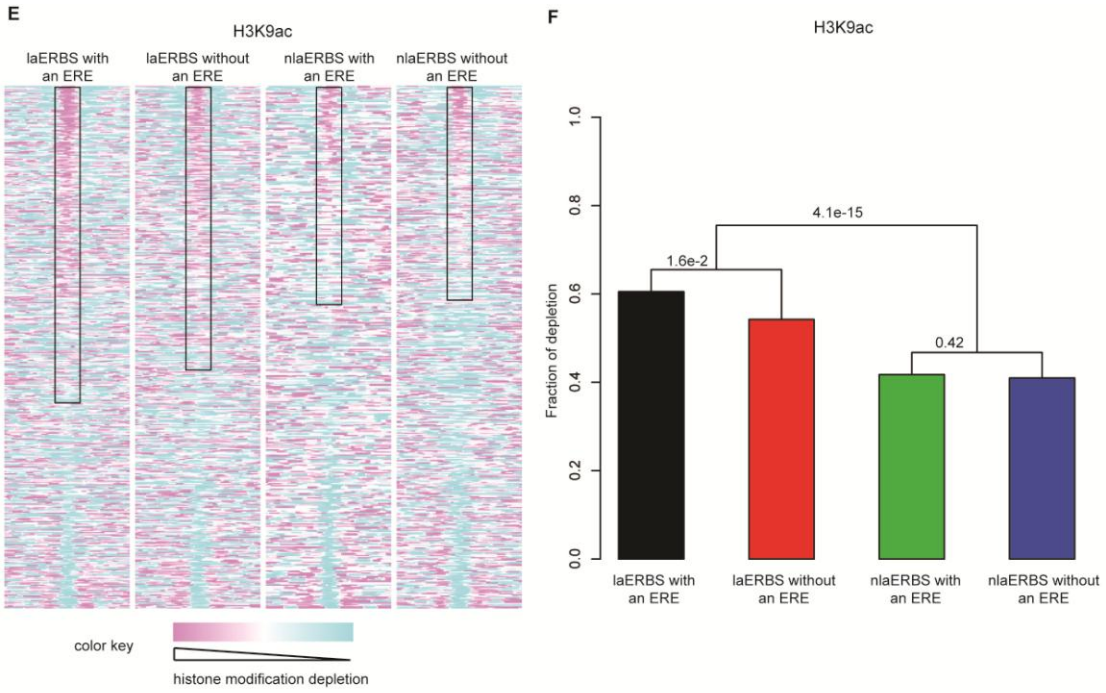


Figure S4. Heatmap show of H3K4me1 (A), H3K4me3 (C) and H3K9ac (E) log2 transformed ratio of read coverage against input around summits of four types of ERBSs: laERBS with an ERE, laERBS without an ERE, nlaERBS with an ERE and nlaERBS without an ERE. Each row represented signal strengths of $\pm$2.5kb region around the summit of ERBS. To make the signal values comparable between different ERBSs, signal values in each row were scaled between 0 and 1. To quantify central histone modification depletion level, we defined $H_{i,j}$ as the differences between log2 transformed ratio of read-coverage against input in central region ($\pm$100bp region relative to the peak summit) to average of read-coverage against input in the two flanking regions (-400bp to -200bp and +200bp to +400bp relative to peak summit) of ERBS i for histone modification j. Then each row (A), (C) and (E) was ranked by $H_{i,j}$. Black rectangle marked those ERBSs with $H_{i,j} < 0$. Barplot show of faction of H3K4me1 (B), H3K4me3 (D) and H3K9ac (F) depletion (defined as $H_{i,j} < 0$) in the center of four types of ERBSs: laERBS with an ERE, laERBS without an ERE, nlaERBS with an ERE and nlaERBS without an ERE. P-values were given by one-tailed Fisher test.

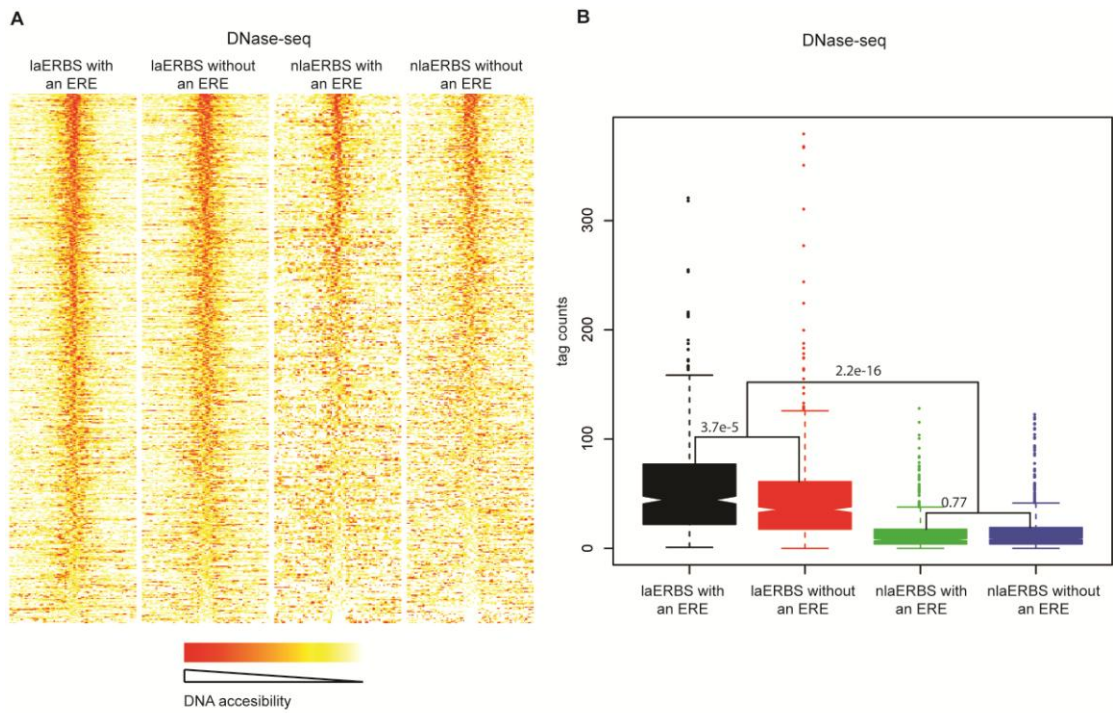


Figure S5. (A) Heatmap show of DNase-seq tag counts around the summit of four types of ERBSs: laERBSs with and without an ERE and nlaERBSs with and without an ERE. Each row was simply ranked by scaled tag counts in the central 200bp region of ERBS. (B) Boxplot of tag counts in the central 200bp region of four types of ERBSs. P-values were given by one-tailed Wilcox test.

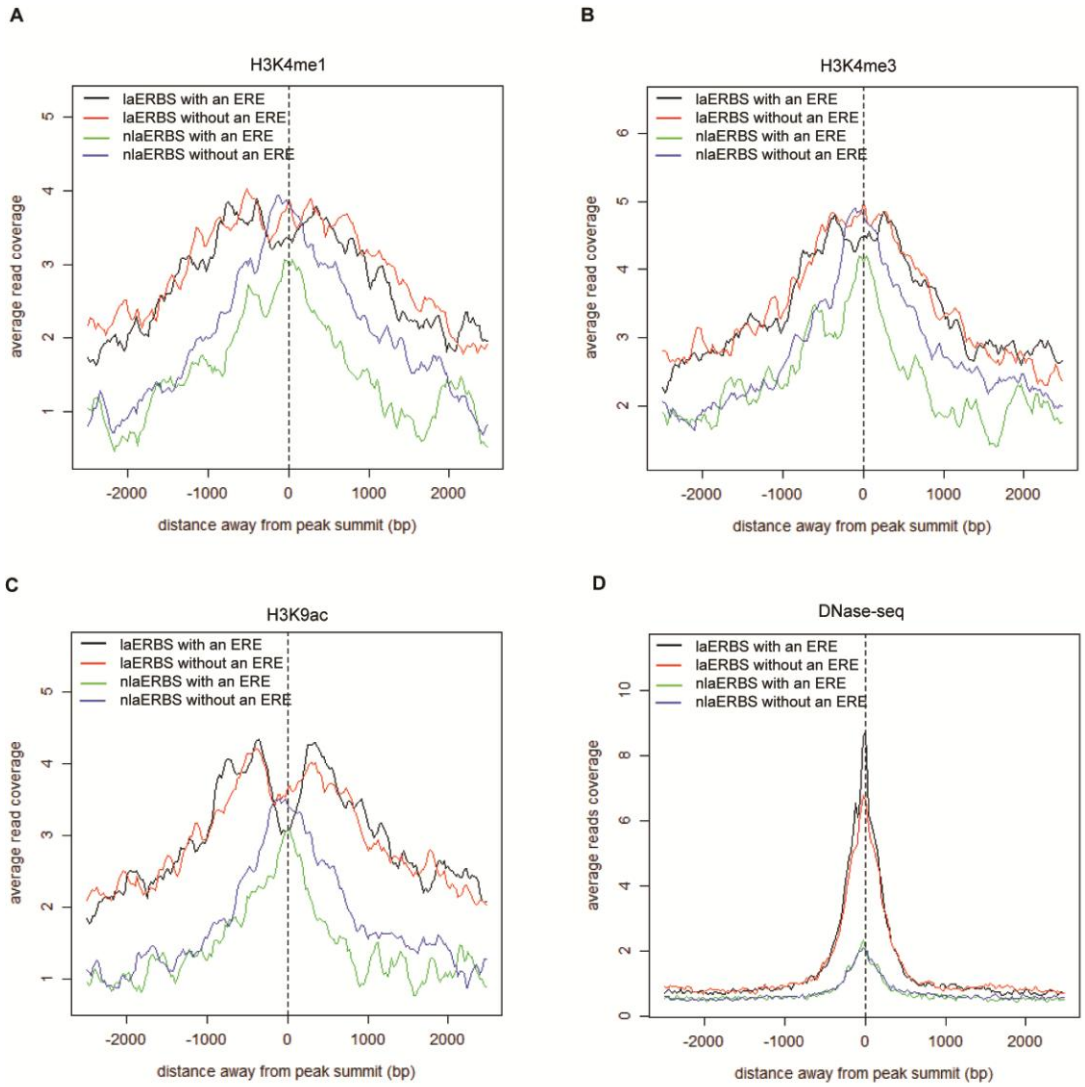


Figure S6. This figure is a complement for Figure 2 in the main text. Compared with Figure 2, ERBSs located within 2kb away from annotated gene TSS were not included. Average H3K4me1 (A), H3K4me3 (B) and H3K9ac (C) read-coverage against input surrounding laERBSs with and without an ERE and nlaERBSs with and without an ERE. Average DNase-seq tag counts surrounding laERBSs with and without an ERE and nlaERBSs with and without an ERE (D).

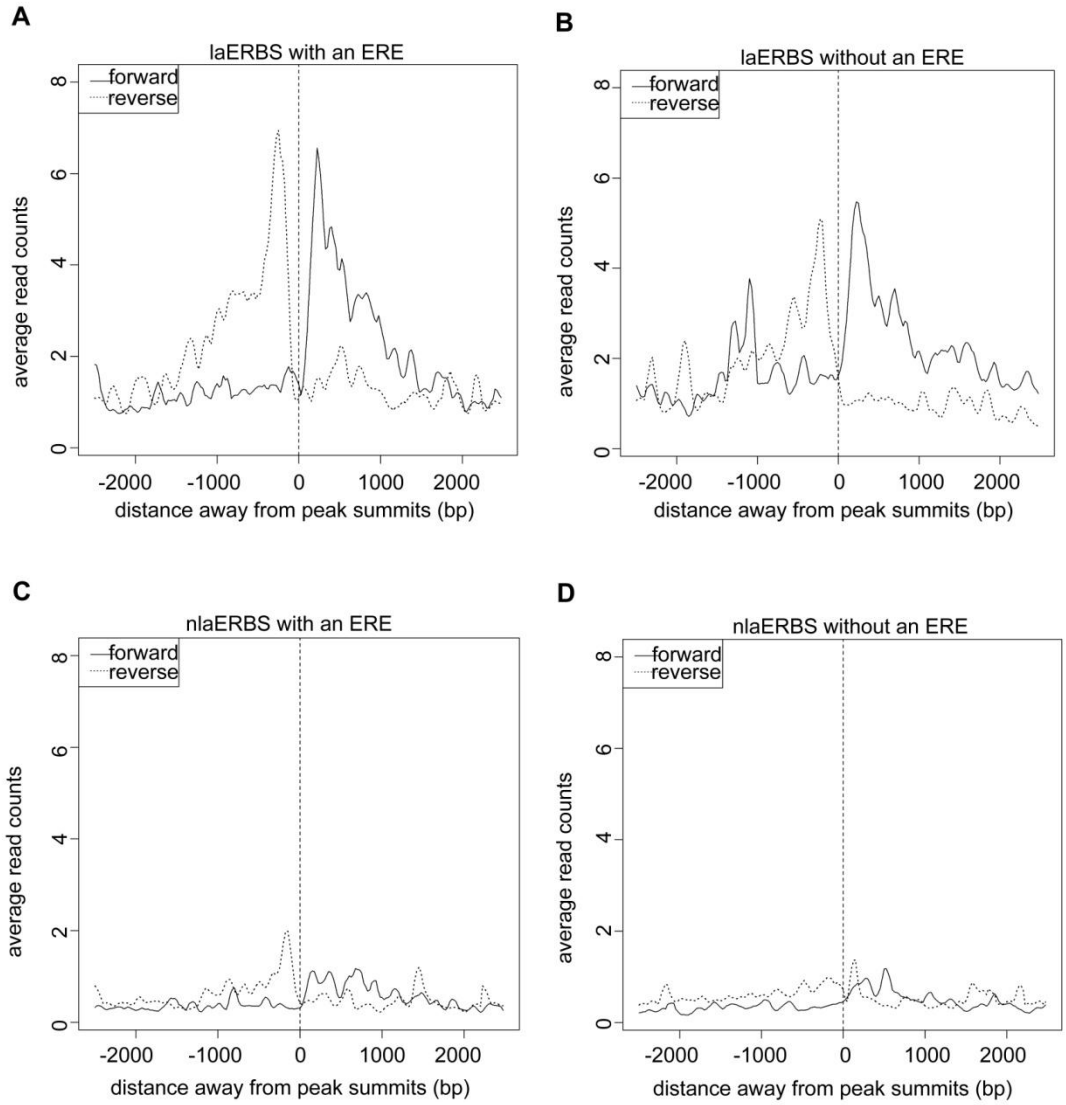


Figure S7. Average GRO-seq signal surrounding laERBSs with and without an ERE and nlaERBSs with and without an ERE.

	laERBS with an ERE (552)	laERBS without an ERE (651)	nlaERBS with an ERE (400)	nlaERBS without an ERE (803)
H3K4me1	300	309	179	336
H3K4me3	287	318	147	324
H3K9ac	334	353	167	329

Table S1. Numbers of ERBSs with $H_{i,j} < 0$ in four types of ERBSs. $H_{i,j}$ was defined the same as in Figure S4

	laERBS with an ERE (552)	laERBS without an ERE (651)	nlaERBS with an ERE (400)	nlaERBS without an ERE (803)
FoxA1	512	595	193	400
GATA3	512	599	175	341
AP2γ	464	560	147	451
p300	276	265	27	60

Table S2. Numbers of ChIP-seq binding peaks that overlapped with four types of ERBSs.

Variable	coefficient	p-value
ERα	1.52	<2.2e-16
FoxA1	0.39	0.000377
GATA3	0.36	0.039454
AP2γ	-0.17	0.176848
DNase-seq	0.68	1.69e-05
H3K4me1	-0.34	0.009967
H3K4me2	-0.20	0.037300
H3K4me3	0.11	0.547894
H3K27me3	-0.44	9.32e-05
P300	-0.88	7.38e-06
Pol2	0.24	0.124450
Log10-distance	-3.72	<2.2e-16
Inverse distance	-1.02969	0.002403

Table S3. Trained parameters for all the training set with all the features selected to predict interactions. P-value is given by R with logistic regression.

Variable	coefficient	p-value
ERα	1.51	< 2e-16
FoxA1	0.39	6.00e-05
AP2γ	-0.32	0.00306
DNase-seq	0.65	1.75e-06
H3K4me2	-0.18	0.01308
H3K27me3	-0.68	3.76e-13
Log10 distance	-3.63	< 2e-16
Inverse distance	-0.98	0.00317

Table S4. Trained parameters for all the training set with eight features selected to predict interactions. P-value is given by R with logistic regression.

	Reported (3113)	Predicted & Unreported (8805)	Non-predicted & unreported (85284)
PETs 2+ cluster	1239	2759	7687
PETs 3+ cluster	904	1932	4749

Table S5. Numbers of reported, predicted and unreported, and rest of the candidate (non-predicted and unreported) ERBS-ERBS pairs overlapped with Pol2 PETs 2+ and PETs 3+ clusters, respectively.

	Reported (113)	Predicted & Unreported (475)	Non-predicted & unreported (87500)
PETs 2+ cluster	57	214	14790
PETs 3+ cluster	47	174	9847

Table S6. Numbers of reported, predicted and unreported, and rest of the candidate (non-predicted and unreported) ERBS-promoter pairs overlapped with Pol2 PETs 2+ and PETs 3+ clusters, respectively.

	Predicted (374)	Rest (9098)	p-value
12h up-regulated	33	245	1.265e-08
24h up-regulated	21	189	8.205e-05
48h up-regulated	29	418	0.005724
12h down-regulated	0	61	1
24h down-regulated	0	36	1
48h down-regulated	3	126	0.8896

Table S7. Enrichment comparison of E2 induced differential expressed gene in predicted ERα target and rest of the candidate genes. Tests were performed in R by one-tailed Fisher test.