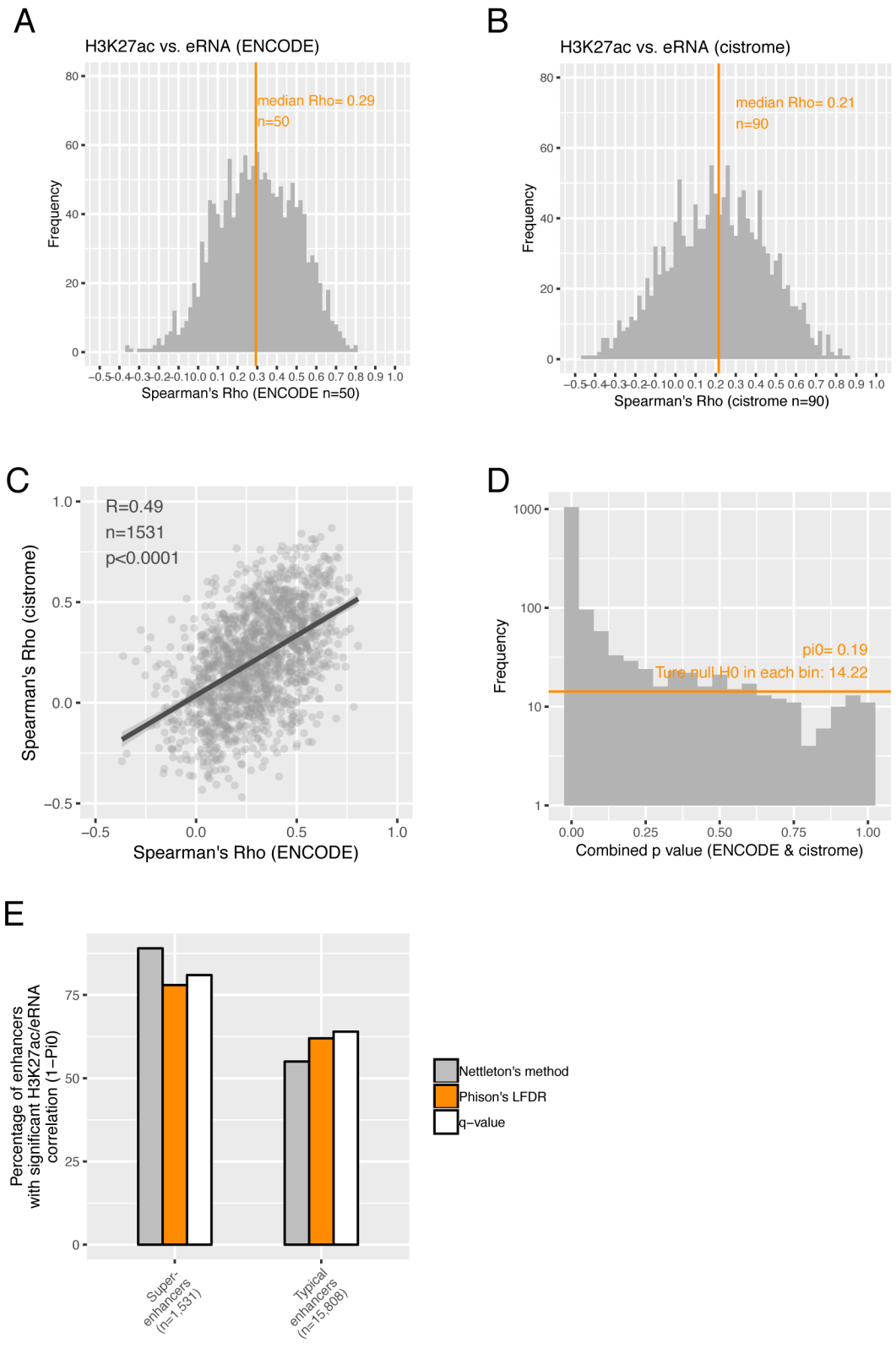


Figure S1. The cross-sample correlation between eRNA expression and H3K27ac level, Related to Figure 1

(A, B) Correlations between H3K27ac modification level and eRNA expression (RPKM) were calculated for each of the 1,531 core super-enhancers in the ENCODE database (**A**) and the cistrome database (**B**). **(C)** Scatter plot of the 1,531 pairs of Spearman's *Rho*s in **A** (x-axis) and **B** (y-axis). **(D)** For each super-enhancer, we combined the p-values from the two datasets using Fisher's method before subjecting them to a q-value calculation. Distribution of q-values indicates the estimated pi0, the percentage of true null H0 (no positive correlation between H3K27ac and eRNA for the super-enhancer studied). Pi0 was calculated using Storey's q-value. We also calculated pi0 using Phison's LFDR and Nettleton's method to estimate the error of pi0 computation. Pi0 was estimated to be 19% ± 3.5%.

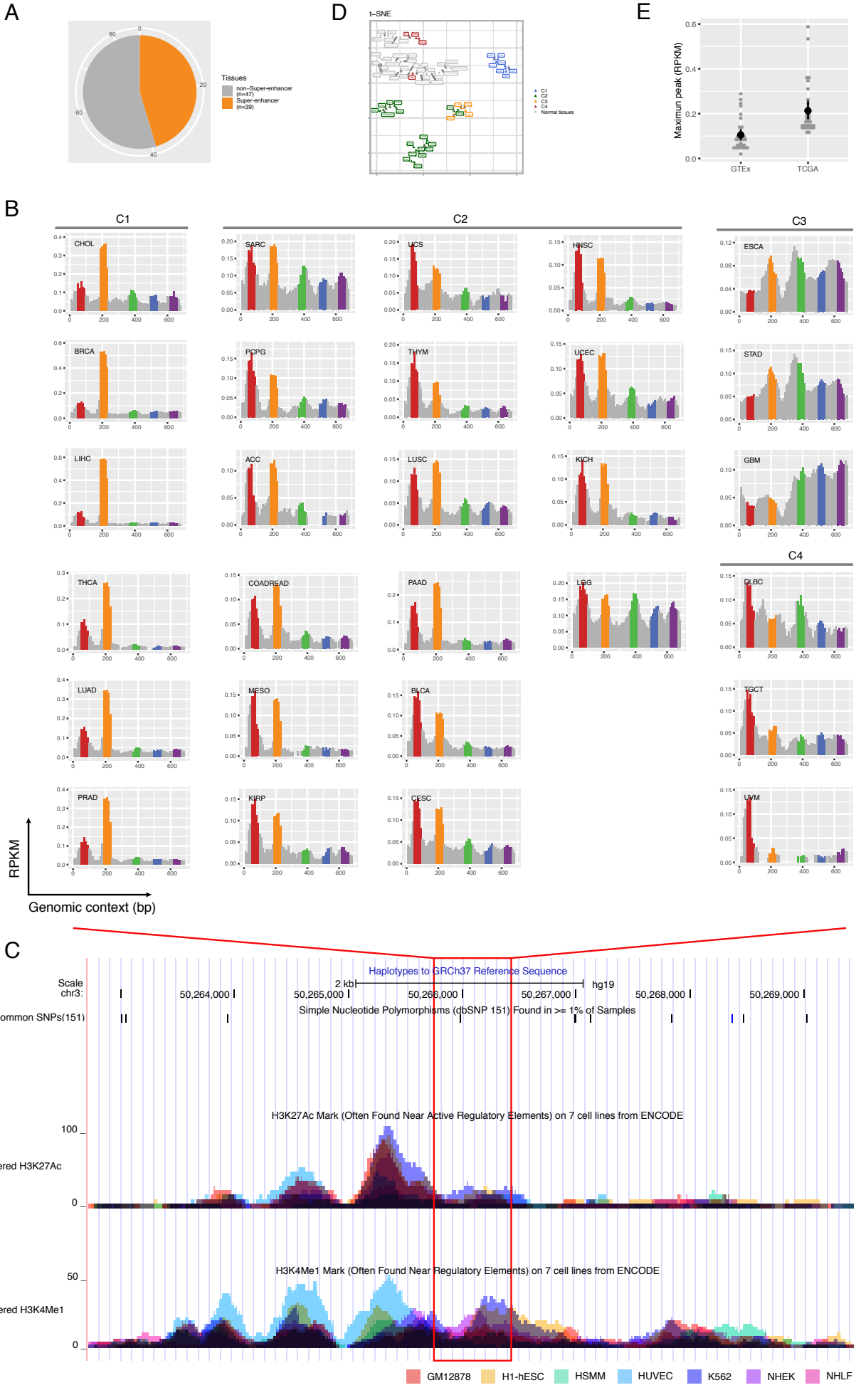

A
Tissues
non−Super−enhancer (n=47)
Super−enhancer (n=39)
D
t−SNE
C1
C2
C3
C4
Normal tissues
E
Maximun peak (RPKM)
GTEx
TCGA
B
C1
C2
C3
C4
CHOL
SARC
UCS
HNSC
ESCA
BRCA
PCPG
THYM
UCEC
STAD
LIHC
ACC
LUSC
KICH
GBM
THCA
COADREAD
PAAD
LGG
DLBC
LUAD
MESO
BLCA
TGCT
PRAD
KIRP
CESC
UVM
RPKM
Genomic context (bp)
C
Haplotypes to GRCh37 Reference Sequence
Scale
chr3:
2 kb
hg19
50,264,000
50,265,000
50,266,000
50,267,000
50,268,000
50,269,000
Common SNPs(151)
Simple Nucleotide Polymorphisms (dbSNP 151) Found in >= 1% of Samples
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K4Me1
GM12878
H1-hESC
HSMM
HUVEC
K562
NHEK
NHLF

Figure S2. The eRNA expression profile of a super-enhancer region chr3:50,265,725-50,266,396, Related to Figure 2

(A) The super-enhancer status of chr3:50265725-50266396 in the original 86 tissue and cell types. **(B)** The eRNA expression in 28 cancer types not included in **Figure 1A**. **(C)** The H3K27ac and H3K4me1 status of chr3:50265725-50266396 and its flanking 3 kb regions in 7 ENCODE cell lines. **(D)** The t-SNE plot of the eRNA expression in the 32 TCGA cancer types and 31 GTEx normal tissues. **(E)** Each grey point represents the average RPKM on chr3:50,265,725-50,266,396 in one of the 32 TCGA cancer types or 31 GTEx body tissues. The median and SE of the indicated group of samples are highlighted in black.

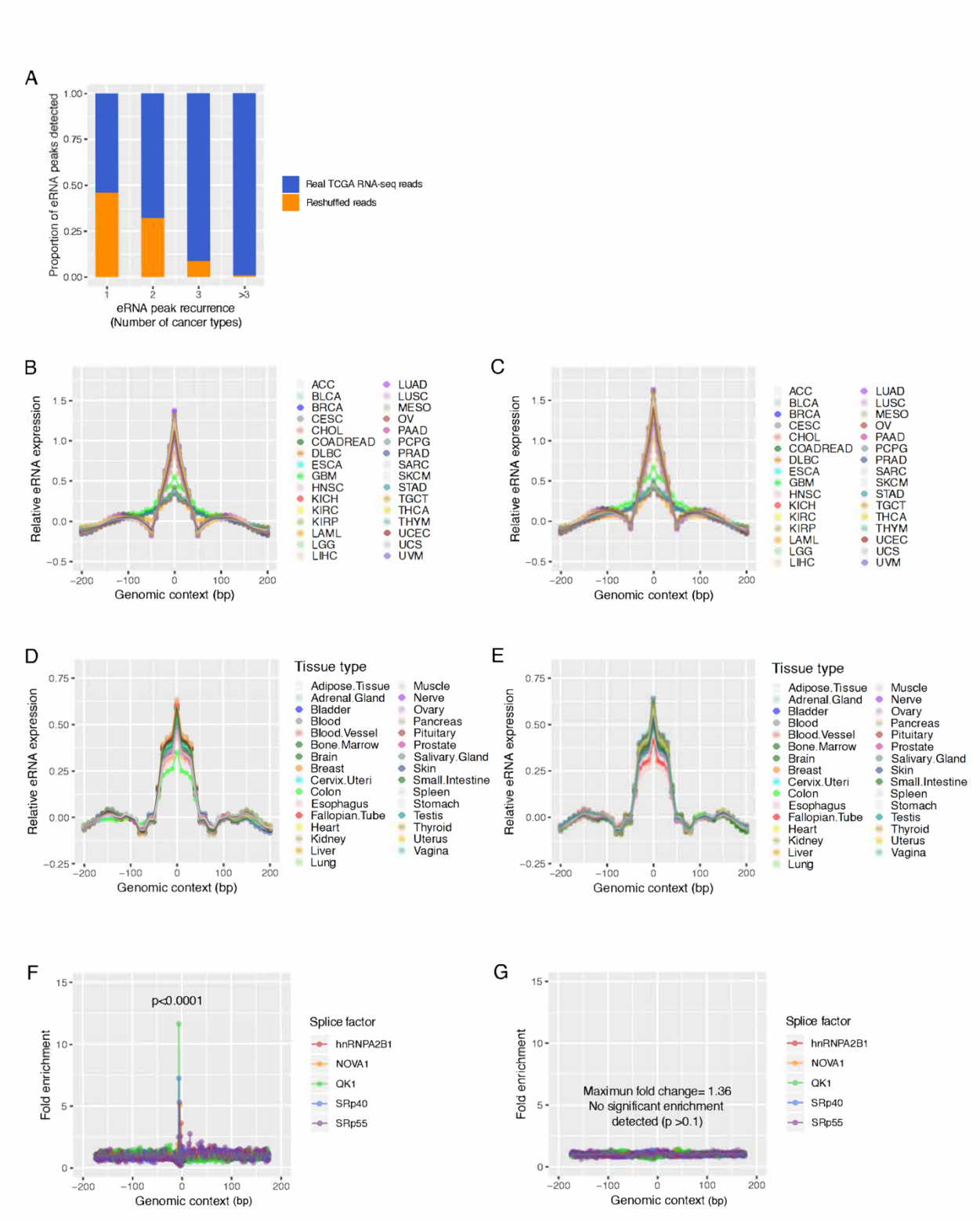


Figure S3. Recurrent eRNA expression peaks in super-enhancers, Related to Figure 3

(A) Estimated FDR of peak detection when an eRNA peak is observed in the indicated number of cancer types. The RNA-seq reads mapped to the core super-enhancer regions in each cancer type

were randomly reshuffled before the same peak identification procedure was performed to estimate the number of peaks generated by transcription noise. FDR was then estimated by comparing the RPKM of peaks identified using permutated reads or real data (see Online Methods). **(B, C)** are the same as in Figure 3A, except that the search range for local maximum RPKMs was extended to 400 bp **(B)** or 600 bp **(C)**. **(D, E)** are the same as in Figure 3B, except that the search range for local maximum RPKMs was extended to 400 bp **(D)** or 600 bp **(E)**. **(F)** The enrichment of splicing factor binding motifs near the exon-intron (or intron-exon) junctions for protein-coding genes. Five splicing factors with $p<1\times10^{-4}$ and fold enrichment >3 are plotted. **(G)** The same analysis as in **F**, except that the 200 bp flanking sequences of 29,828 super-enhancer peaks were used. The same five splicing factors in **F** are displayed. Notably, the first 20 and last 20 bps were not included in enrichment analysis since SFmap requires at least 20 bp sequences for motif searching.

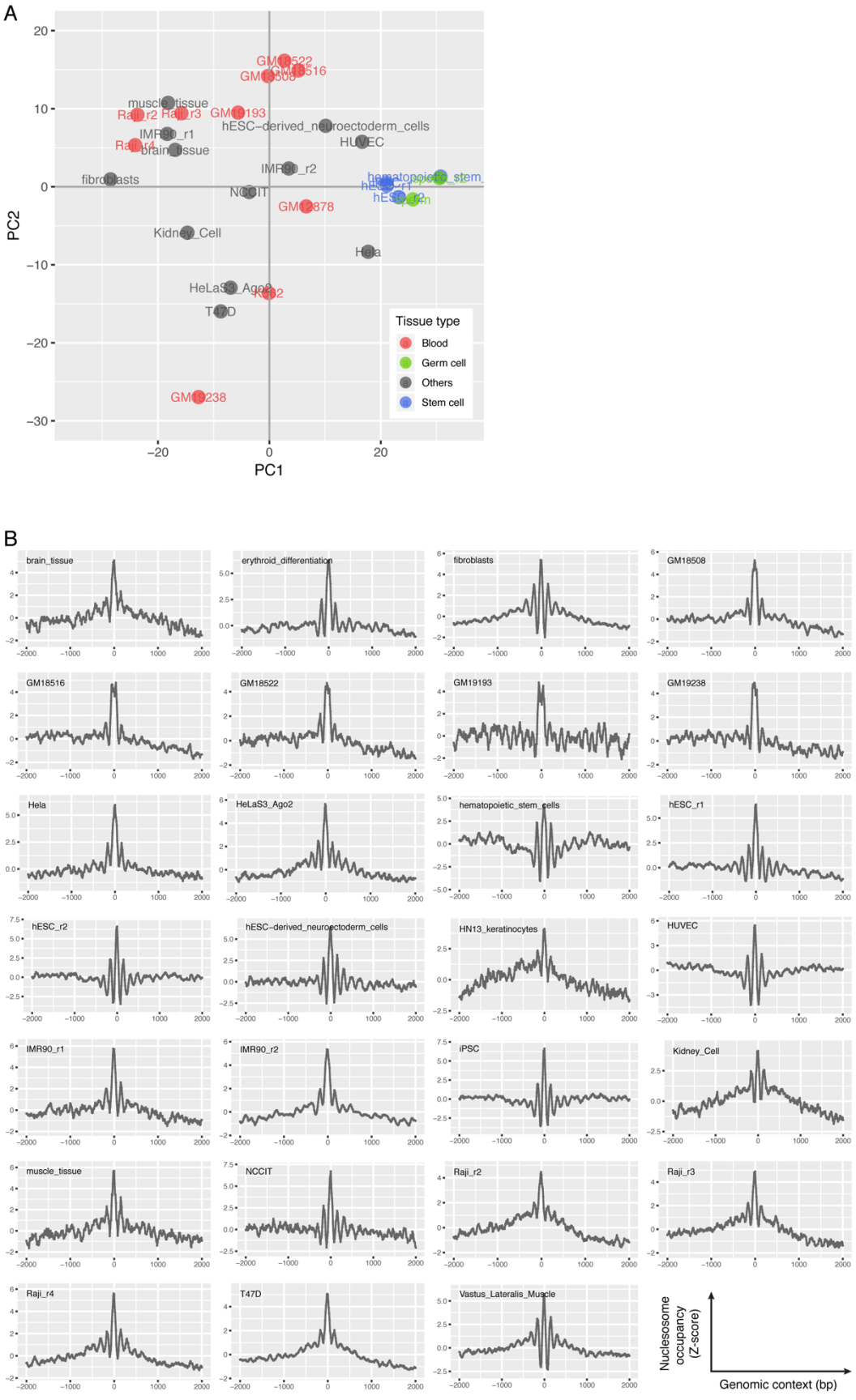

A
PC2
PC1
Tissue type
Blood
Germ cell
Others
Stem cell
B
Nuclesosome occupancy (Z-score)
Genomic context (bp)

Figure S4. Nucleosome positioning around the super-enhancer eRNA peaks, Related to Figure 4

(A) Principal component analysis (PCA) on the genome-wide nucleosome profiles of 29 human tissue/cell types. The PC1 and PC2 explain 33.7% and 12.5% variations, respectively. **(B)** The nucleosome positioning in the flanking 2 kb region of the 29,828 super-enhancer peaks in the 27 human tissue/cell types. The calculation is the same as in Figure **4A** (top panel), except that the surveyed region was extended to 2 kb on each side, and the signal for each tissue is plotted separately.

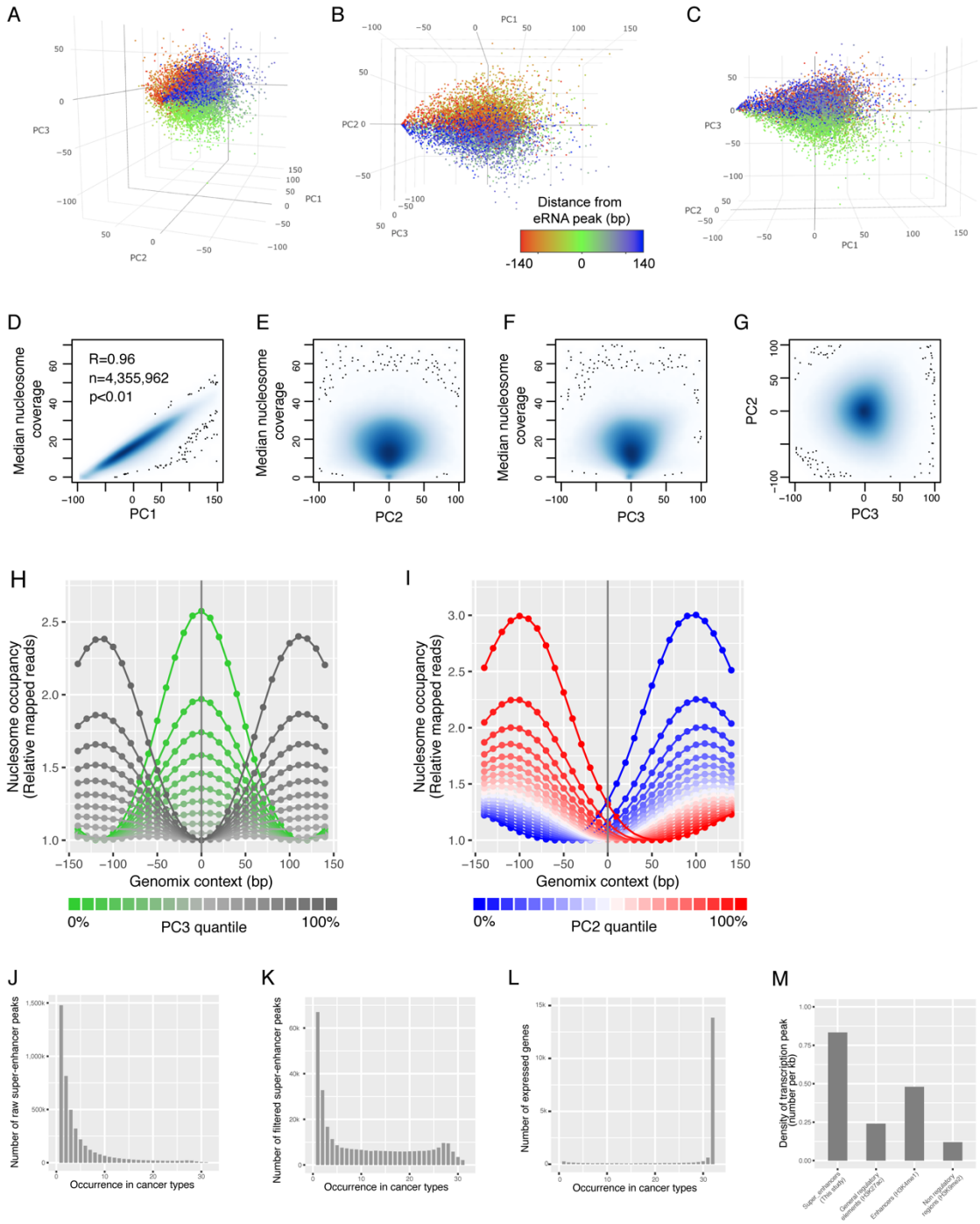
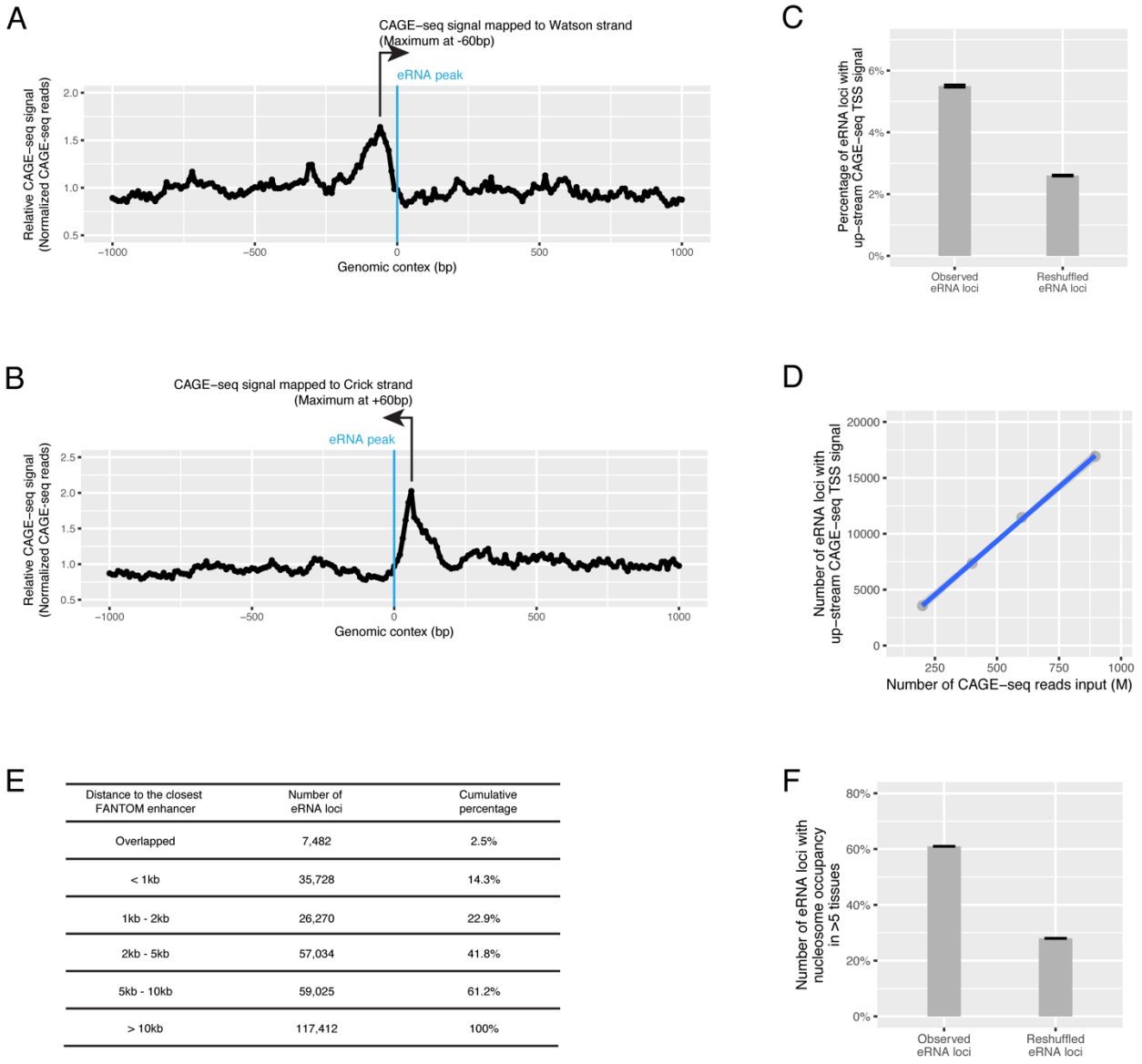


Figure S5. Nucleosomes positioned near the eRNA peaks in super-enhancers, Related to Figure 5

(A, B, and C) 3D scatterplots of the PCA result shown in Figure **5A**, except that the data points are visualized from different angles. Only the first 10,000 points of the ~4 million peaks are plotted. **(D, E, and F)** Scatterplot of PC1, PC2, and PC3 versus the median number of MNase-seq reads mapped to the flanking 140 bp DNA of the ~4 million loci of the local maximum eRNA RPKMs. (**G**) Scatterplot of PC2 versus PC3 for the ~4 million loci. **(H)** PC3 represents the phase

difference between eRNA expression and the local nucleosome positioning. The ~4 million loci were evenly divided into 20 quantiles according to the PC3 (from negative/green to positive/grey). For each quantile, the 280 bp sequences were aligned with the loci of the local maximum eRNA RPKMs at the center (0 bp). Each point represents a 10 bp window. The mean number of mapped MNase-seq reads was calculated for all 10 bp windows with the same relative positions. The resulting mean signals were normalized into relative nucleosome occupancies (y-axis) by dividing them with the minimum values of the 280 bp sequence. **(I)** same as **H**, except that the quantile of PC2 was considered (form negative/blue to positive/red). (**J**) The occurrence of the ~4 million loci being identified as the local maximum PPKM in the 32 TCGA cancer types. (**K**) The occurrence of the 302,951 peaks with PC3<0 and relative height >0.05 in the 32 TCGA cancer types. (**L**) The occurrence of all protein-coding genes expressed in the 32 TCGA cancer types. A gene is defined as expressed in a cancer type if it has an RSEM>10 in >10% of the tumors of that cancer type. (**M**) The density of transcriptional peaks with PC3<0 and relative height >0.05 in different genomic regions. We divided the human genome into non-overlapping 1 kb fragments and called the average intensities of the indicated markers in each fragment from all the ENCODE bigwig files available in the ENCODE portal. For each marker, we selected fragments with signals larger than 2-fold of the corresponding input and with no significant signal (1-fold) observed for any other markers. These regions were subjected to transcription peak identification using the same criterion defining the 302,951 super-enhancer eRNA loci. Peak density was counted as the number of peaks per 1 kb genomic DNA.

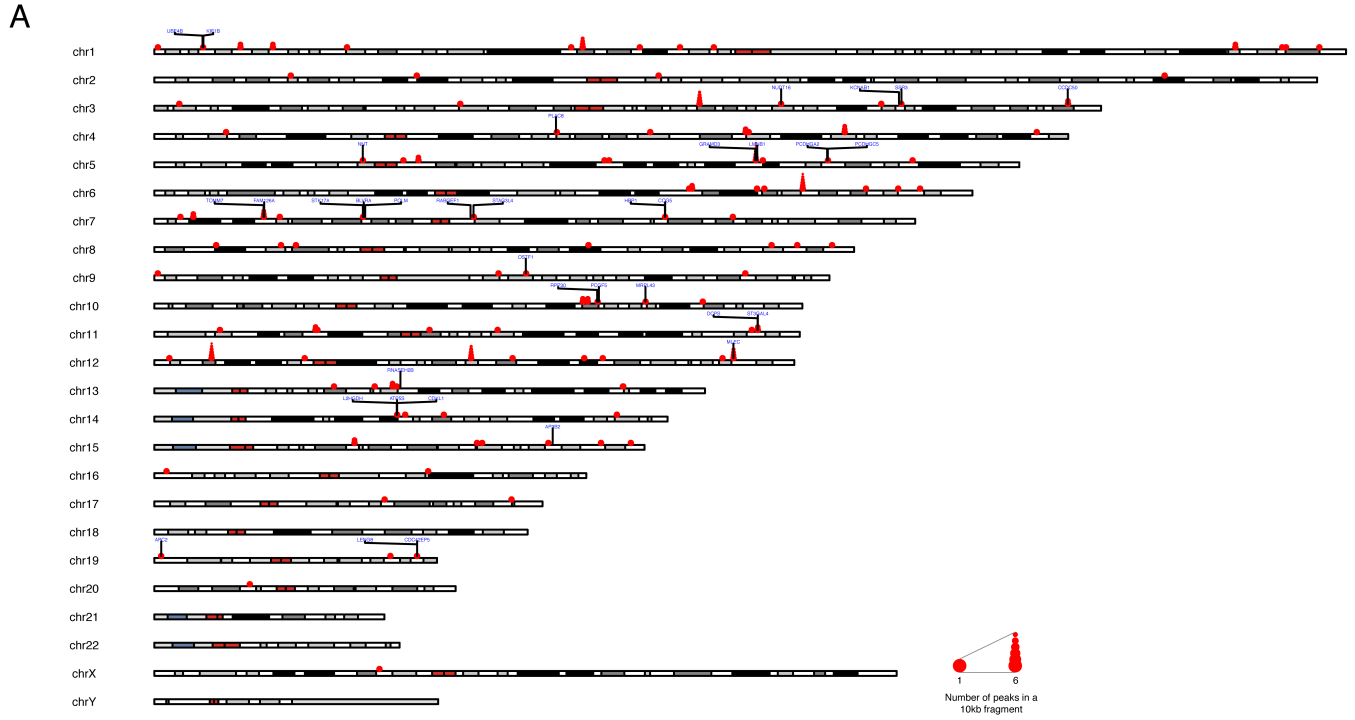


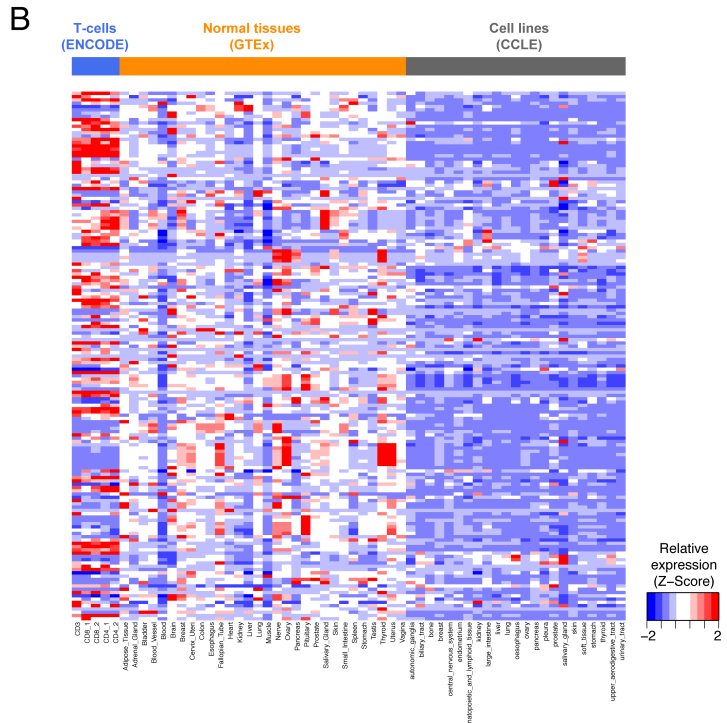
Distance to the closest FANTOM enhancer	Number of eRNA loci	Cumulative percentage
Overlapped	7,482	2.5%
< 1kb	35,728	14.3%
1kb - 2kb	26,270	22.9%
2kb - 5kb	57,034	41.8%
5kb - 10kb	59,025	61.2%
> 10kb	117,412	100%

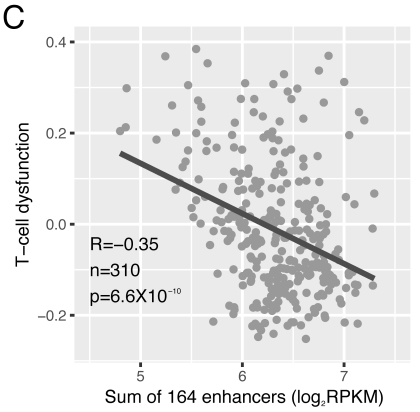
Figure S6. Comparison of eRNA loci identified using RNA-seq with FANTOM enhancers, Related to Figure 5

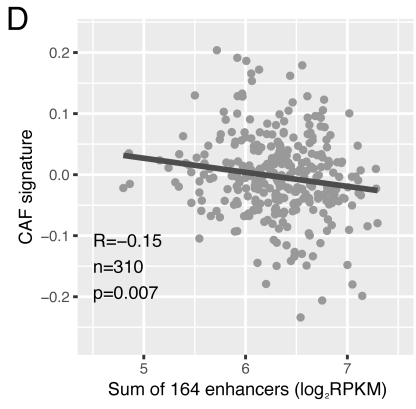
(A, B) The FANTOM CAGE-seq signals on the Watson strand **(A)** and Crick strand **(B)** of the genomic DNA flanking the eRNA loci identified in this study. The FANTOM enhancer CAGE-seq reads were mapped to the human genome, and the nucleotide at the 5′ end of each mapped read was considered the start of an eRNA transcript. The 302,951 super-enhancer eRNA loci and their flanking 1 kb DNA were aligned with the eRNA peaks at the center. Each point represents a 10 bp window with indicated distance to the eRNA loci. The total number of eRNA start sites detected by CAGE-seq in all 10 bp windows with the same relative positions was counted and the average of super-enhancer regions. **(C)** The percentage of eRNA loci with at least one enhancer FANTOM CAGE-seq read mapped to its up-stream -90 to -10 bp on either strand. The reshuffled result is based on when all eRNA loci were randomly assigned to the human genome. **(D)** The same as in **C**, except the input FANTOM CAGE-seq reads were down-sampled to the indicated number. **(E)** The percentage of eRNA loci with well-position nucleosomes detected in their ± 20

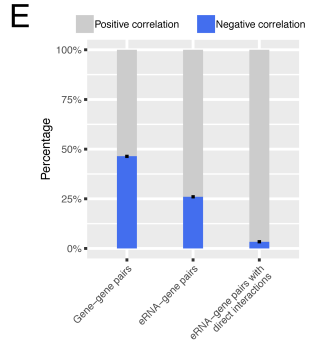
bp region in ≥5 of the 15 paired-end MNase-seq profiles using DANPOS2.0. The background was calculated by randomly assigning the eRNAs on the human genome. **(F)** The number of eRNA loci with nucleosome occupancy in >5 tissues. The background was calculated by randomly shuffled eRNA loci.

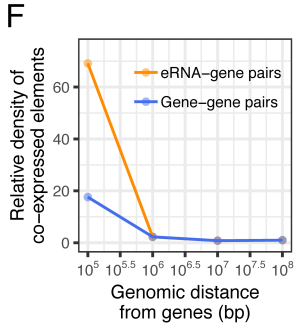

A
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chrY
1
6
Number of peaks in a 10kb fragment

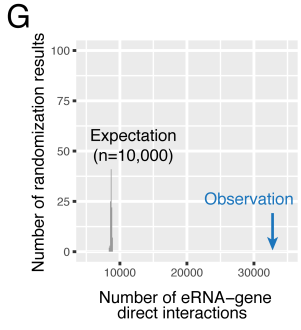

B
T-cells (ENCODE)
Normal tissues (GTEx)
Cell lines (CCLE)
Relative expression (Z-Score)
−2
2


C
T−cell dysfunction
0.4
0.2
0.0
−0.2
R=−0.35
n=310
p=6.6X10−10
5
6
7
Sum of 164 enhancers (log2RPKM)


D
CAF signature
0.2
0.1
0.0
−0.1
−0.2
R=−0.15
n=310
p=0.007
5
6
7
Sum of 164 enhancers (log2RPKM)


E
Positive correlation
Negative correlation
Percentage
100%
75%
50%
25%
0%
Gene−gene pairs
eRNA−gene pairs
eRNA−gene pairs with direct interactions


F
Relative density of co−expressed elements
eRNA−gene pairs
Gene−gene pairs
60
40
20
0
10^5
10^5.5
10^6
10^6.5
10^7
10^7.5
10^8
Genomic distance from genes (bp)


G
Number of randomization results
100
75
50
25
0
Expectation (n=10,000)
Observation
10000
20000
30000
Number of eRNA−gene direct interactions

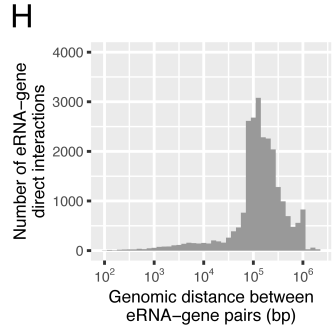

H
Number of eRNA−gene direct interactions
4000
3000
2000
1000
0
10^2
10^3
10^4
10^5
10^6
Genomic distance between eRNA−gene pairs (bp)

Figure S7. The genomic distribution and molecular function of the 164 eRNA loci, Related to Figure 6

(A) The genomic distribution of the 164 eRNA loci (red dots) differentially expressed in a cohort of 28 melanoma patients receiving anti-PD1 immunotherapy and of the 36 genes targeted by eQTLs in the 164 eRNA loci. **(B)** The RPKM of the 164 eRNAs in primary T-cells, GTEx normal tissues, and CCLE cancer cell lines. For GTEx and CCLE data, each column represents the mean of all samples/cell-lines of the indicated tissue of origin. (**C, D)** Correlation between the T-cell dysfunction signature and CAF function signature from the TIDE database and the sum of $\log_2$RPKM of the 164 eRNA loci in a cohort of 310 TCGA melanoma samples. **(E)** The direction of eRNA-gene co-expression. An eRNA-gene pair or gene-gene pair was considered as co-expressed if they consistently co-expressed/reversed in at least 3 cancer types. In each cancer type, we required a p-value with Bonferroni correction to be <0.01 and an absolute Spearman's *Rho* to be >0.3. **(F)** The probability of any possible eRNA-gene pair or gene-gene pair to be co-expressed when they are at an indicated distance. The probability for pairs located on different chromosomes was used as the background for normalization. **(G)** The number of co-expressed eRNA-gene pairs connected by Hi-C loops in 3D Genome Browser (n = 32,298). The distribution of the expected number was calculated by randomly assigning the eRNAs to the human genome 10,000 times. **(H)** The distribution of the genomic distance between any eRNA and its partner gene in the 32,298 pairs.