Dissecting the contribution of transposable elements to interphase chromosome structure

Liyang Shi[1†*], Zhen Xiao[1*], Xuemeng Zhou[1*], Isaac A. Babarinde[1,2], Xiuling Fu[1,3], Gang Ma[1,4], Zhuoqi Huang[1], Li Sun[1], Jiangping He[5], Alexander Strunnikov[6], Andrew P. Hutchins[1#]

[1]Department of Systems Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, Guangdong, China
[2]Laboratory of Inflammation and Vaccines, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China
[3]Columbia Center for Human Development, Columbia University Irving Medical Center, New York, NY, USA
[4]Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, China
[5]Guangzhou National Laboratory, Guangzhou, Guangdong, China
[6]Guangzhou Institutes of Biomedicine and Health, Guangzhou, Guangdong, China
*Equal contribution
†Present Address: Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Kyoto, Japan, and Department of Human Genetics, McGill University, Montreal, QC, Canada
#Correspondence: andrewh@sustech.edu.cn

References

Babarinde IA, Ma G, Li Y, Deng B, Luo Z, Liu H, Abdul MM, Ward C, Chen M, Fu X *et al* (2021) Transposable element sequence fragments incorporated into coding and noncoding transcripts modulate the transcriptome of human pluripotent stem cells. *Nucleic Acids Res* 49: 9132-9153

Carico ZM, Stefan HC, Justice M, Yimit A, Dowen JM (2021) A cohesin cancer mutation reveals a role for the hinge domain in genome organization and gene expression. *PLoS Genet* 17: e1009435

Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W *et al* (2015) Chromatin architecture reorganization during stem cell differentiation. *Nature* 518: 331-336

Justice M, Carico ZM, Stefan HC, Dowen JM (2020) A WIZ/Cohesin/CTCF Complex Anchors DNA Loops to Define Gene Expression and Cell Identity. *Cell Rep* 31: 107503

Ma G, Fu X, Zhou L, Babarinde IA, Shi L, Yang W, Chen J, Xiao Z, Qiao Y, Ma L *et al* (2025) The nuclear matrix stabilizes primed-specific genes in human pluripotent stem cells. *Nat Cell Biol* 27: 232-245

Theunissen TW, Friedli M, He Y, Planet E, O'Neil RC, Markoulaki S, Pontis J, Wang H, Iouranova A, Imbeault M *et al* (2016) Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* 19: 502-515

Zhang Y, Dali R, Blanchette M (2025) RobusTAD: reference panel based annotation of nested topologically associating domains. *Genome Biol* 26: 129

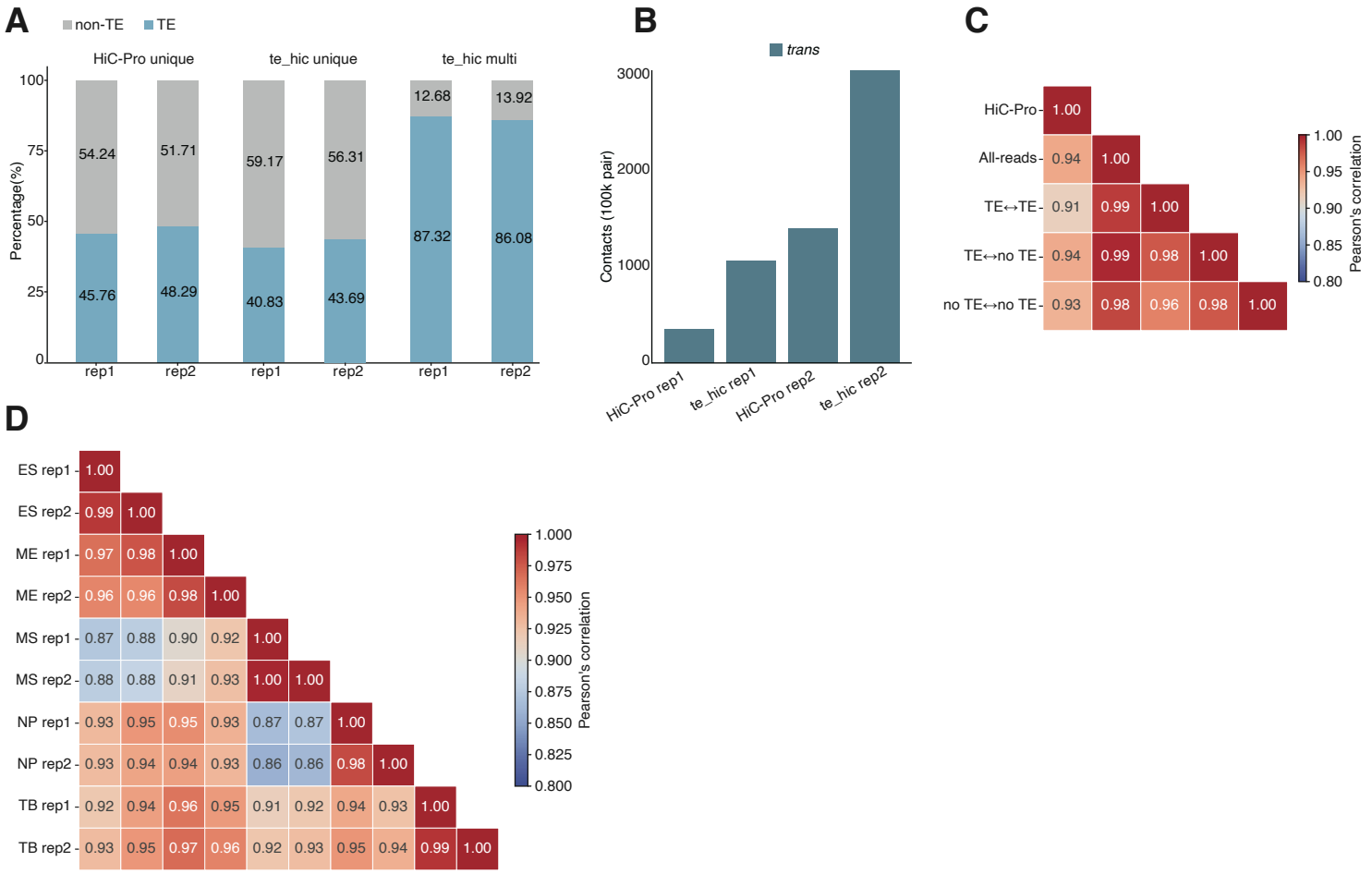
A
non-TE
TE
HiC-Pro unique
te_hic unique
te_hic multi
Percentage(%)
100
75
50
25
0
54.24
45.76
51.71
48.29
59.17
40.83
56.31
43.69
12.68
87.32
13.92
86.08
rep1
rep2
rep1
rep2
rep1
rep2
B
trans
Contacts (100k pair)
3000
2000
1000
0
HiC-Pro rep1
te_hic rep1
HiC-Pro rep2
te_hic rep2
C
HiC-Pro
All-reads
TE↔TE
TE↔no TE
no TE↔no TE
1.00
0.94 1.00
0.91 0.99 1.00
0.94 0.99 0.98 1.00
0.93 0.98 0.96 0.98 1.00
Pearson's correlation
1.00
0.95
0.90
0.85
0.80
D
ES rep1
ES rep2
ME rep1
ME rep2
MS rep1
MS rep2
NP rep1
NP rep2
TB rep1
TB rep2
1.00
0.99 1.00
0.97 0.98 1.00
0.96 0.96 0.98 1.00
0.87 0.88 0.90 0.92 1.00
0.88 0.88 0.91 0.93 1.00 1.00
0.93 0.95 0.95 0.93 0.87 0.87 1.00
0.93 0.94 0.94 0.93 0.86 0.86 0.98 1.00
0.92 0.94 0.96 0.95 0.91 0.92 0.94 0.93 1.00
0.93 0.95 0.97 0.96 0.92 0.93 0.95 0.94 0.99 1.00
Pearson's correlation
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0.800

Fig. S1

Fig. S1. Multimapped reads and replication.

(A) Bar chart of the proportion of multi-mapped and uniquely mapped reads mapping to TEs.

(B) Barplot of the number of *trans* chromatin contacts, where both paired reads are on different chromosomes.

(C) Heatmap of pair-wise Pearson correlations comparing the te_hic generated matrices to the HiC-Pro matrix.

(D) Heatmap of Pearson's correlation of all-read matrices for the indicated biological replicates for hPSCs, ME (mesoderm differentiated cells), MS (Mesenchymal stem cells), NP (Neural progenitors) and TB (Trophoblast cells). Data is from (Dixon *et al*, 2015)

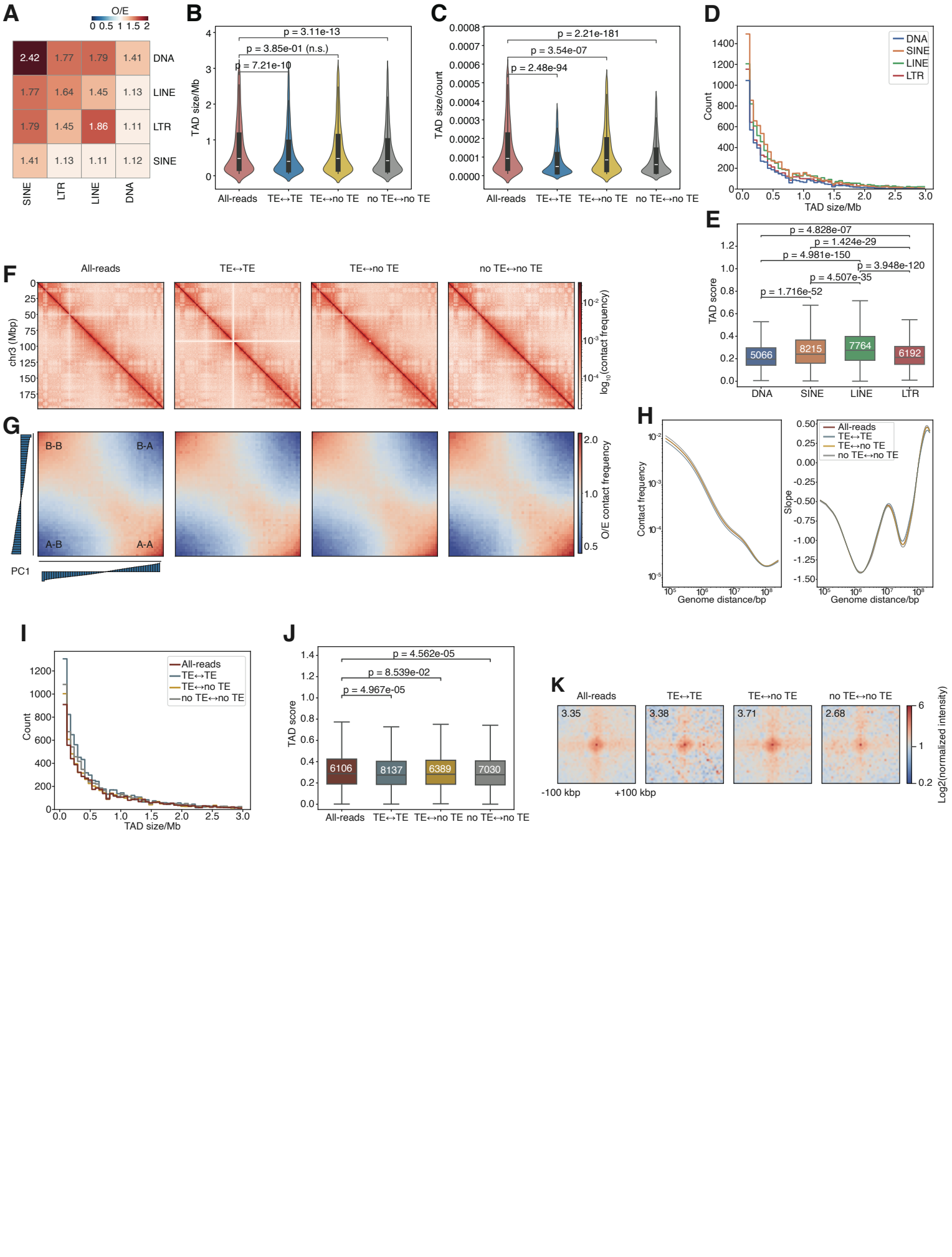
A
O/E
0 0.5 1 1.5 2
2.42 1.77 1.79 1.41 DNA
1.77 1.64 1.45 1.13 LINE
1.79 1.45 1.86 1.11 LTR
1.41 1.13 1.11 1.12 SINE
SINE LTR LINE DNA
B
p = 3.11e-13
p = 3.85e-01 (n.s.)
p = 7.21e-10
TAD size/Mb
All-reads TE↔TE TE↔no TE no TE↔no TE
C
p = 2.21e-181
p = 3.54e-07
p = 2.48e-94
TAD size/count
All-reads TE↔TE TE↔no TE no TE↔no TE
D
DNA
SINE
LINE
LTR
Count
TAD size/Mb
E
p = 4.828e-07
p = 1.424e-29
p = 4.981e-150
p = 3.948e-120
p = 4.507e-35
p = 1.716e-52
TAD score
5066
8215
7764
6192
DNA SINE LINE LTR
F
All-reads TE↔TE TE↔no TE no TE↔no TE
chr3 (Mbp)
log10(contact frequency)
G
B-B
B-A
A-B
A-A
PC1
O/E contact frequency
H
Contact frequency
Genome distance/bp
All-reads
TE↔TE
TE↔no TE
no TE↔no TE
Slope
Genome distance/bp
I
All-reads
TE↔TE
TE↔no TE
no TE↔no TE
Count
TAD size/Mb
J
p = 4.562e-05
p = 8.539e-02
p = 4.967e-05
TAD score
6106
8137
6389
7030
All-reads TE↔TE TE↔no TE no TE↔no TE
K
All-reads TE↔TE TE↔no TE no TE↔no TE
3.35
3.38
3.71
2.68
-100 kbp
+100 kbp
Log2(normalized intensity)

Fig. S2

Fig. S2. 3D genome properties in downsampled matrices, separated by TE superfamily.

(A) Heatmap of the pair-wise Observed/Expected (O/E) for homotypic and heterotypic TE-to-TE contacts.
(B) Violin plots of TAD sizes in the four TE-matrices.
(C) Violin plots of TAD sizes in the four TE-matrices, normalized by the number of TADs in each matrix.
(D) Histogram of TAD sizes generated using matrices containing only reads with the indicated TE superfamily at one or both ends. Contacts assigned with more than one TE superfamilies at one or both ends were discarded.
(E) A boxplot of TAD score for the indicated superfamilies of TEs, as measured by RobusTAD (Zhang *et al*, 2025). Significance is from a two-tailed Mann-Whitney U test.
(F) The Hi-C matrices were downsampled to the same size as the smallest matrix (noTE-to-noTE). Normalized contact density heatmaps for chromosome 3 using the downsampled subsets of downsampled reads: (1) all-reads, (2) TE-to-TE, (3) TE-to-noTE, and (4) noTE-to-noTE matrices.
(G) Saddle plots showing genome-wide average O/E interaction frequency and A/B compartmentalization of the four downsampled matrices.
(H) Plots showing contact frequency as a function of genome distance for the four downsampled matrices as in panel B. The left plot shows the overall contact frequency (distance in base pairs between the mapped read pairs), and the right plot shows the slopes.
(I) Histogram of TAD domain sizes in megabase pairs for the four downsampled matrices.
(J) Boxplot of the TAD score estimated for each of the four downsampled matrices by RobusTAD. The number in white indicates the number of TAD observed in each matrix. Significance is from a two-tailed Mann-Whitney U test.
(K) Aggregate Pileup Analysis (APA) plots centered on the loop coordinate in the downsampled matrices. The flanking 100 kbp is shown. The value in the top left corner is the score of the central pixel.

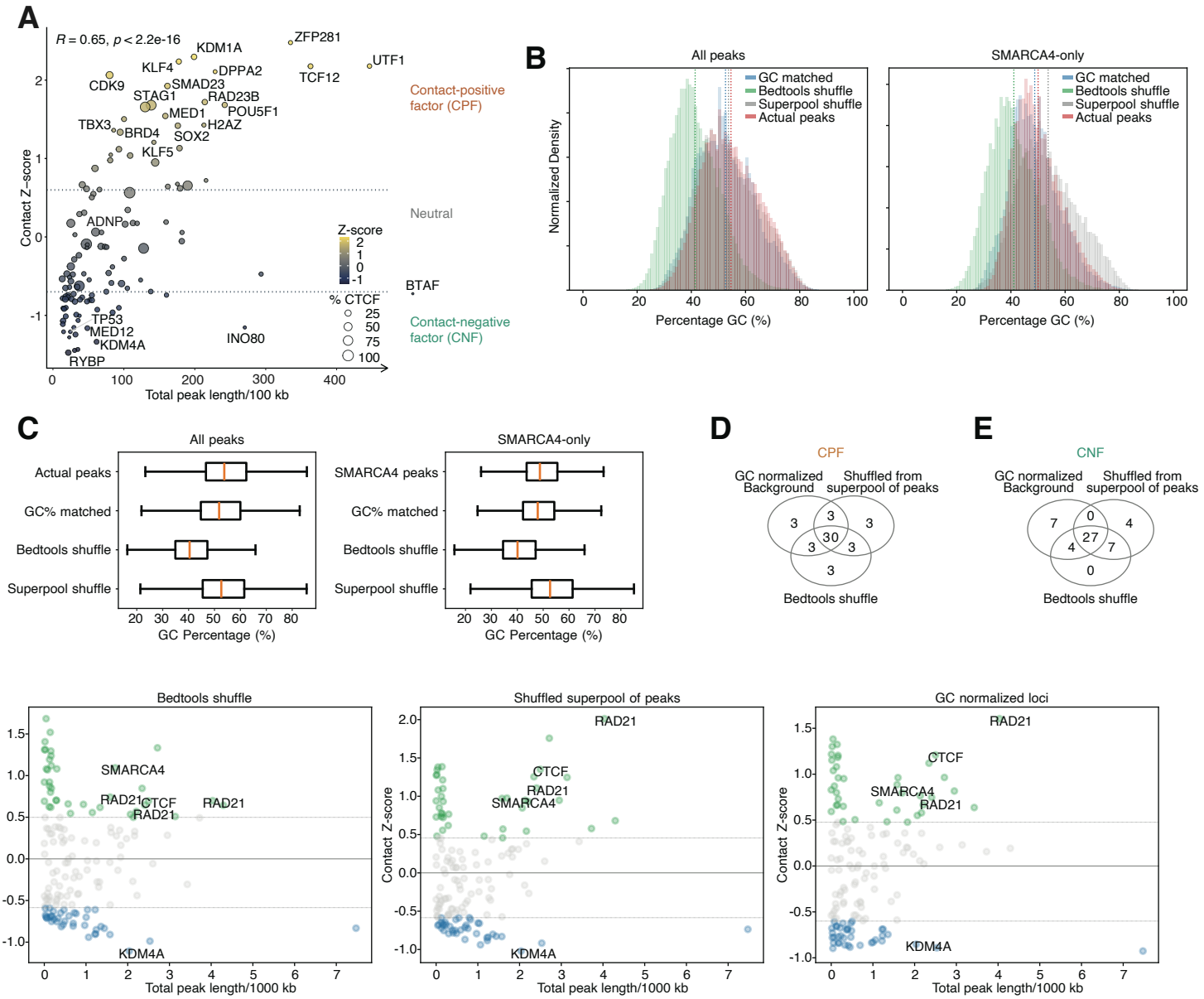
A
R = 0.65, p < 2.2e-16
KDM1A
ZFP281
UTF1
KLF4
DPPA2
TCF12
CDK9
SMAD23
STAG1
RAD23B
MED1
POU5F1
TBX3
BRD4
H2AZ
SOX2
KLF5
ADNP
TP53
MED12
KDM4A
RYBP
INO80
BTAF
Contact-positive factor (CPF)
Neutral
Contact-negative factor (CNF)
Z-score
% CTCF
Contact Z-score
Total peak length/100 kb
B
All peaks
SMARCA4-only
GC matched
Bedtools shuffle
Superpool shuffle
Actual peaks
Normalized Density
Percentage GC (%)
C
All peaks
SMARCA4-only
Actual peaks
GC% matched
Bedtools shuffle
Superpool shuffle
SMARCA4 peaks
GC Percentage (%)
D
CPF
GC normalized Background
Shuffled from superpool of peaks
Bedtools shuffle
E
CNF
GC normalized Background
Shuffled from superpool of peaks
Bedtools shuffle
F
Bedtools shuffle
Shuffled superpool of peaks
GC normalized loci
SMARCA4
RAD21
CTCF
KDM4A
Contact Z-score
Total peak length/1000 kb

Fig. S3

Fig. S3. Contact Z-score in mouse pluripotent stem cells and impact of different normalization strategies on the designation of CPFs and CNFs.

(A) Scatter plot of contact Z-score versus normalized peak size (in base pairs) for mouse PSCs. Only homotypic ETF-ETF contacts are considered. The size of the dots indicates the percentage overlap with CTCF binding sites. Hi-C data was from GSM4074304 (Justice *et al*, 2020) and GSM4647542 (Carico *et al*, 2021).

(B) Histograms of GC content for all peaks (left) and SMARCA4-only peaks (right), using three background strategies. (1) Bedtools shuffle (the original strategy), (2) GC normalized background and (3) peaks selected randomly from the superset of all peaks. The dotted lines indicate the mean of the distribution for each histogram.

(C) Boxplots of the GC percentage for the backgrounds described in **panel B**.

(D) Venn diagrams of the number of CPFs detected using three background strategies.

(E) As in **panel D**, but for CNFs.

(F) Scatter plots of contact Z-score versus normalized peak size (in base pairs) for hPSCs. Each plot was generated using one of the three backgrounds: Bedtools shuffle, shuffled selection from a superpool of all peaks, or matched GC normalized peaks. Note that ETFs can appear more than once if multiple ChIP-seq datasets were analyzed.

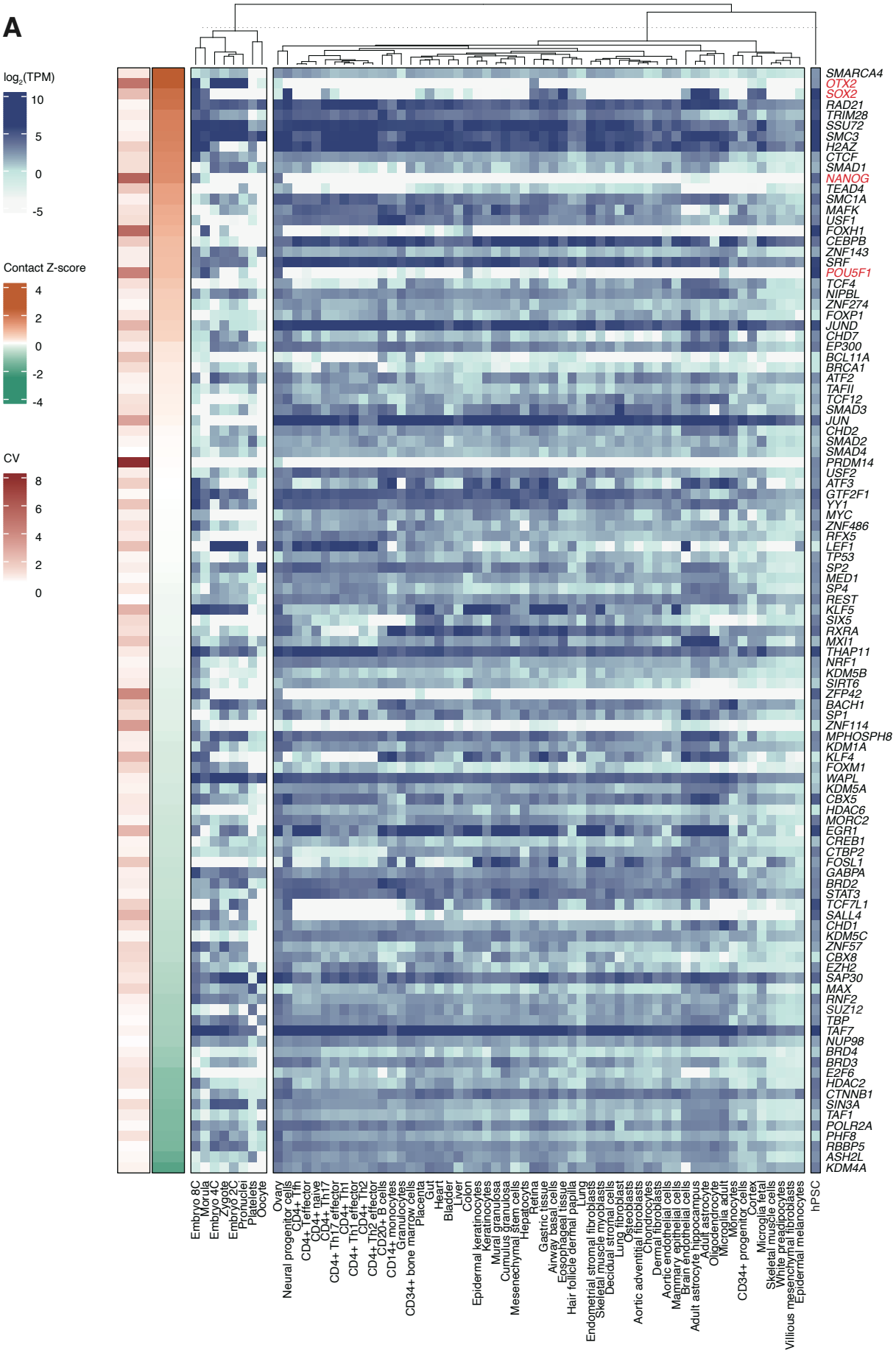
A
log2(TPM)
10
5
0
-5
Contact Z-score
4
2
0
-2
-4
CV
8
6
4
2
0
SMARCA4
OTX2
SOX2
RAD21
TRIM28
SSU72
SMC3
H2AZ
CTCF
SMAD1
NANOG
TEAD4
SMC1A
MAFK
USF1
FOXH1
CEBPB
ZNF143
SRF
POU5F1
TCF4
NIPBL
ZNF274
FOXP1
JUND
CHD7
EP300
BCL11A
BRCA1
ATF2
TAFII
TCF12
SMAD3
JUN
CHD2
SMAD2
SMAD4
PRDM14
USF2
ATF3
GTF2F1
YY1
MYC
ZNF486
RFX5
LEF1
TP53
SP2
MED1
SP4
REST
KLF5
SIX5
RXRA
MXI1
THAP11
NRF1
KDM5B
SIRT6
ZFP42
BACH1
SP1
ZNF114
MPHOSPH8
KDM1A
KLF4
FOXM1
WAPL
KDM5A
CBX5
HDAC6
MORC2
EGR1
CREB1
CTBP2
FOSL1
GABPA
BRD2
STAT3
TCF7L1
SALL4
CHD1
KDM5C
ZNF57
CBX8
EZH2
SAP30
MAX
RNF2
SUZ12
TBP
TAF7
NUP98
BRD4
BRD3
E2F6
HDAC2
CTNNB1
SIN3A
TAF1
POLR2A
PHF8
RBBP5
ASH2L
KDM4A
Embryo 8C
Morula
Embryo 4C
Zygote
Embryo 2C
Pronuclei
Platelets
Oocyte
Ovary
Neural progenitor cells
CD4+ Tfh
CD4+ Teffector
CD4+ naive
CD4+ Th17
CD4+ Th17 effector
CD4+ Th1
CD4+ Th1 effector
CD4+ Th2
CD4+ Th2 effector
CD20+ B cells
CD14+ monocytes
Granulocytes
CD34+ bone marrow cells
Placenta
Gut
Heart
Bladder
Liver
Colon
Epidermal keratinocytes
Keratinocytes
Mural granulosa
Cumulus granulosa
Mesenechymal stem cells
Hepatocytes
Retina
Gastric tissue
Airway basal cells
Esophageal tissue
Hair follicle dermal papilla
Lung
Endometrial stromal fibroblasts
Skeletal muscle myoblasts
Decidual stromal cells
Lung fibroblast
Osteoblasts
Aortic adventitial fibroblasts
Chondrocytes
Dermal fibroblasts
Aortic endothelial cells
Mammary epithelial cells
Breast epithelial cells
Adult astrocyte
Adult astrocyte cortex
Oligodendrocyte
Microglia adult
Monocytes
CD34+ progenitor cells
Cortex
Microglia fetal
Skeletal muscle cells
White preadipocytes
Vilious mesenchymal fibroblasts
Epidermal melanocytes
hPSC

Fig. S4

Fig. S4. Cell type-specific expression of ETFs.

(A) Heatmap of expression of the genes for the ETFs explored in this study. Expression was measured in hPSCs versus a panel of somatic cell types. Rows of the heatmap is ordered based on their contact Z-score from **Fig. 3A**. RNA-seq data is from the reanalysis performed in (Babarinde *et al*, 2021).

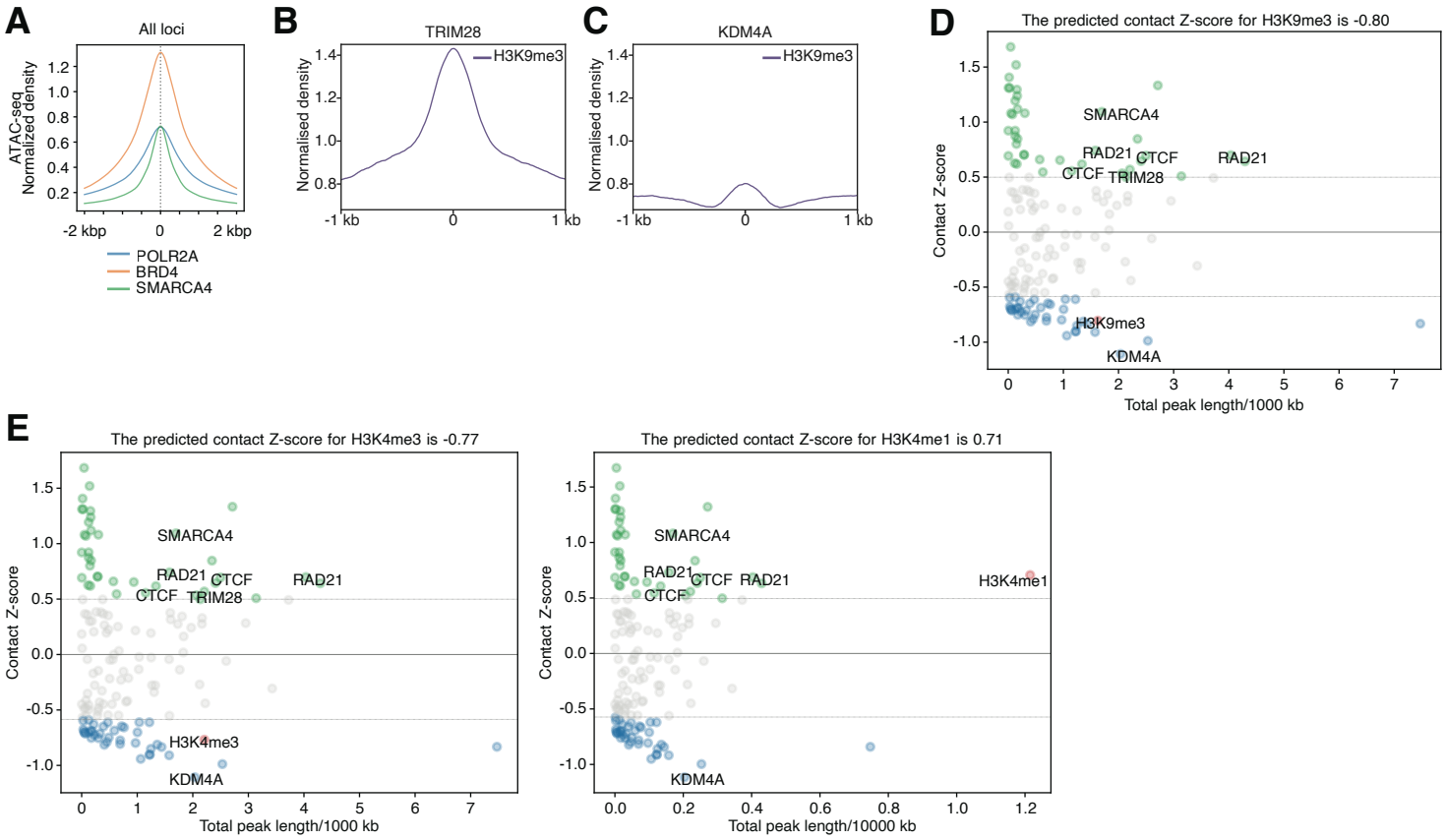
A
All loci
ATAC-seq
Normalized density
-2 kbp
0
2 kbp
POLR2A
BRD4
SMARCA4
B
TRIM28
H3K9me3
Normalised density
-1 kb
1 kb
C
KDM4A
H3K9me3
D
The predicted contact Z-score for H3K9me3 is -0.80
Contact Z-score
SMARCA4
RAD21
CTCF
RAD21
CTCF
TRIM28
H3K9me3
KDM4A
Total peak length/1000 kb
E
The predicted contact Z-score for H3K4me3 is -0.77
SMARCA4
RAD21
CTCF
RAD21
CTCF
TRIM28
H3K4me3
KDM4A
Total peak length/1000 kb
The predicted contact Z-score for H3K4me1 is 0.71
SMARCA4
RAD21
CTCF
RAD21
CTCF
H3K4me1
KDM4A
Total peak length/10000 kb

Fig. S5

Fig. S5. CPFs are associated with enhancers and inversely associated with H3K9me3.

(A) Chromatin accessibility at POLR2A, BRD4 and SMARCA4 bound loci in hPSC, derived from ATAC-seq data. ATAC-seq data is from GSE242351 (Ma *et al.*, 2025).

(B) Pileup of H3K9me3 ChIP-seq data at TRIM28 binding sites. Data was taken from GSE84382 (Theunissen *et al*, 2016).

(C) Pileup of H3K9me3 ChIP-seq data at KDM4A binding sites. Data was taken from GSE84382 (Theunissen *et al.*, 2016).

(D) Predicted contact Z-score for H3K9me3. H3K9me3 data from hPSCs was embedded into the contact Z-score prediction map using the contactZ tool as described in the methods. Data was taken from GSE84382 (Theunissen *et al.*, 2016). Note that ETFs can appear more than once if multiple ChIP-seq datasets were analyzed.

(E) Predicted contact Z-score for H3K4me3 and H3K4me1. Epigenetic data was taken from (Ma *et al.*, 2025) (Theunissen *et al.*, 2016). Note that ETFs can appear more than once if multiple ChIP-seq datasets were analyzed.

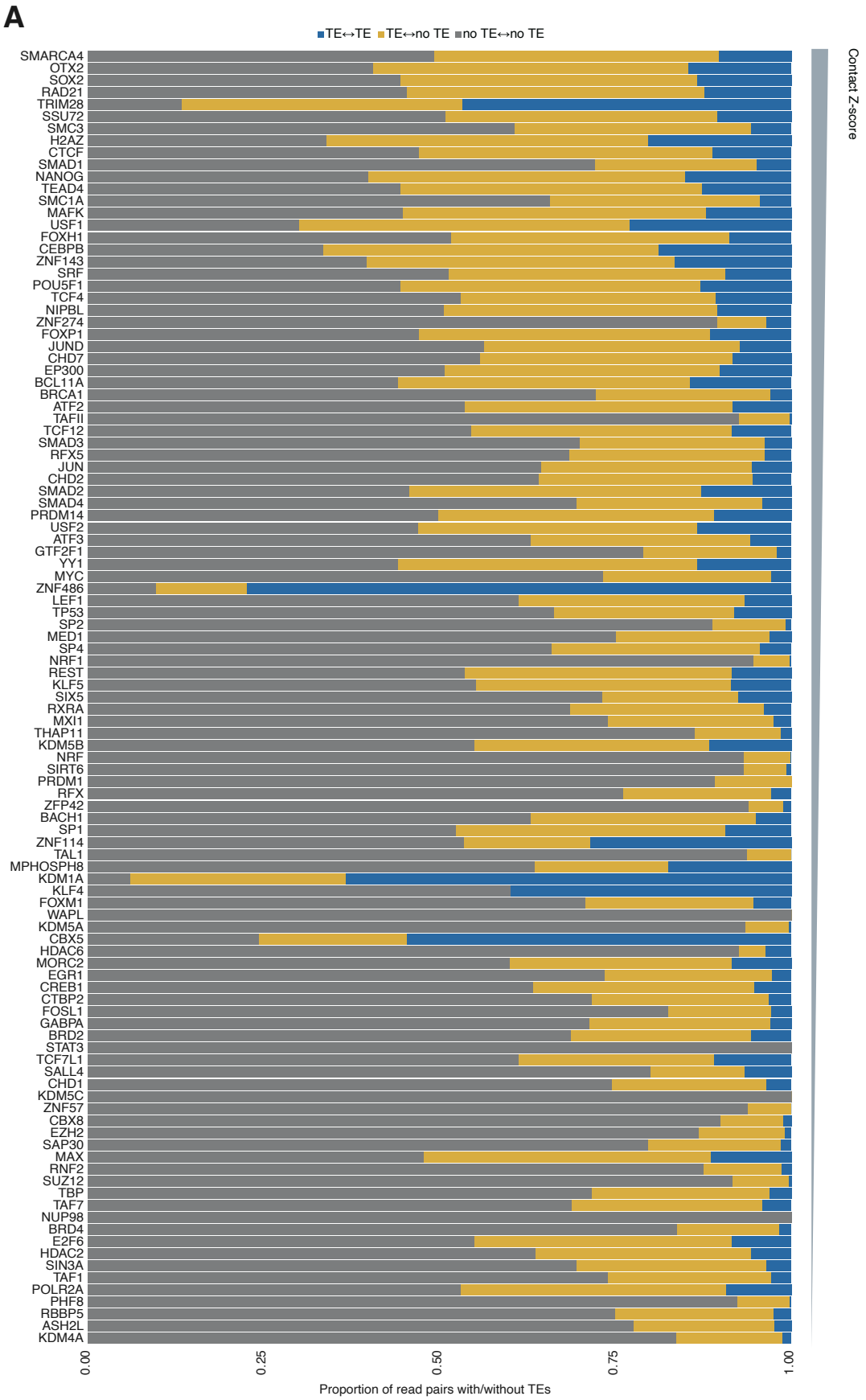
A
TE↔TE
TE↔no TE
no TE↔no TE
Contact Z-score
SMARCA4
OTX2
SOX2
RAD21
TRIM28
SSU72
SMC3
H2AZ
CTCF
SMAD1
NANOG
TEAD4
SMC1A
MAFK
USF1
FOXH1
CEBPB
ZNF143
SRF
POU5F1
TCF4
NIPBL
ZNF274
FOXP1
JUND
CHD7
EP300
BCL11A
BRCA1
ATF2
TAFII
TCF12
SMAD3
RFX5
JUN
CHD2
SMAD2
SMAD4
PRDM14
USF2
ATF3
GTF2F1
YY1
MYC
ZNF486
LEF1
TP53
SP2
MED1
SP4
NRF1
REST
KLF5
SIX5
RXRA
MXI1
THAP11
KDM5B
NRF
SIRT6
PRDM1
RFX
ZFP42
BACH1
SP1
ZNF114
TAL1
MPHOSPH8
KDM1A
KLF4
FOXM1
WAPL
KDM5A
CBX5
HDAC6
MORC2
EGR1
CREB1
CTBP2
FOSL1
GABPA
BRD2
STAT3
TCF7L1
SALL4
CHD1
KDM5C
ZNF57
CBX8
EZH2
SAP30
MAX
RNF2
SUZ12
TBP
TAF7
NUP98
BRD4
E2F6
HDAC2
SIN3A
TAF1
POLR2A
PHF8
RBBP5
ASH2L
KDM4A
0.00
0.25
0.50
0.75
1.00
Proportion of read pairs with/without TEs
B
ATAC-seq
Cluster 3
Cluster 2
Cluster 1
ATAC-seq
Normalized density
3.5
3.0
2.5
2.0
1.5
1.0
-1 kbp
ETF&TE
1 kbp
Contact-positive factor (CPF)
Contact-negative factor (CNF)
Neutral

Fig. S6

Fig. S6. Proportion of ETF-bound Hi-C read pairs overlap with TEs.

(A) Cumulative bar plot of the proportion of TEs at one or both ends of each ETF-specific contact, bars are ranked by the contact Z-score from **Fig. 3A**.

(B) Chromatin accessibility at ETFs bound to TEs. ETFs were divided into CPF, CNF and neutral factors, as defined in **Fig. 4A**, against the three clusters of TEs, as defined in **Fig. 4B**. The three clusters correspond to: cluster 3; TEs associated with CPF ETFs, cluster 2, TEs weakly associated with CPF ETFs, and cluster 1; TEs not associated with CPF ETFs. The central location of each pileup shows an ETF bound to a specific TE in the indicated class. Data is from (Ma *et al.*, 2025).

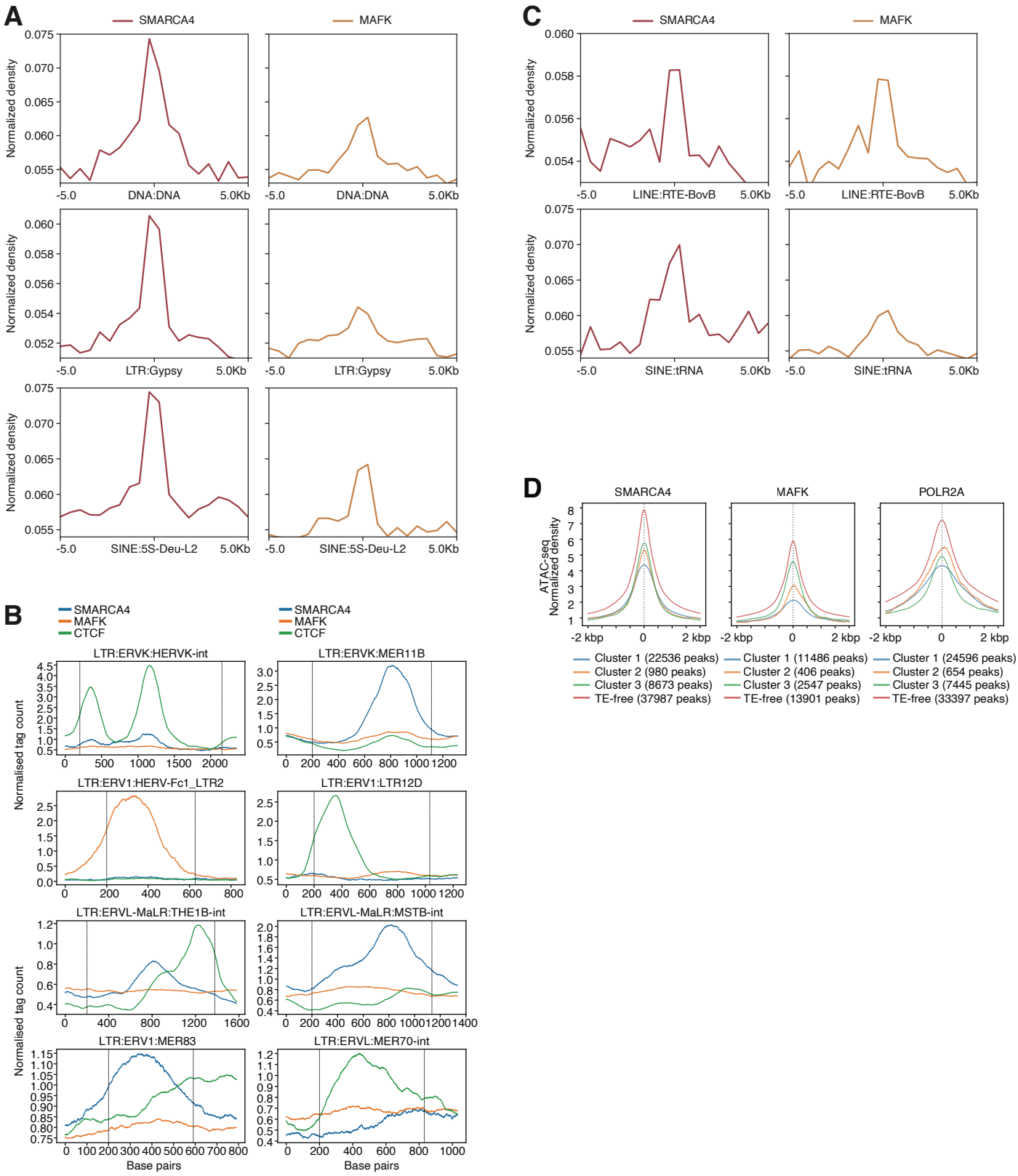
A
SMARCA4
MAFK
Normalized density
DNA:DNA
LTR:Gypsy
SINE:5S-Deu-L2
C
LINE:RTE-BovB
SINE:tRNA
B
SMARCA4
MAFK
CTCF
Normalised tag count
LTR:ERVK:HERVK-int
LTR:ERVK:MER11B
LTR:ERV1:HERV-Fc1_LTR2
LTR:ERV1:LTR12D
LTR:ERVL-MaLR:THE1B-int
LTR:ERVL-MaLR:MSTB-int
LTR:ERV1:MER83
LTR:ERVL:MER70-int
Base pairs
D
SMARCA4
MAFK
POLR2A
ATAC-seq
Normalized density
-2 kbp
0
2 kbp
Cluster 1 (22536 peaks)
Cluster 2 (980 peaks)
Cluster 3 (8673 peaks)
TE-free (37987 peaks)
Cluster 1 (11486 peaks)
Cluster 2 (406 peaks)
Cluster 3 (2547 peaks)
TE-free (13901 peaks)
Cluster 1 (24596 peaks)
Cluster 2 (654 peaks)
Cluster 3 (7445 peaks)
TE-free (33397 peaks)

Fig. S7

Fig. S7. Pileups of SMARCA4 and MAFK ChIP-seq on selected TEs.

(A) Pileups of selected cluster 3 TE subfamilies. The flanking 5 kbp is shown centered on the TE sequence.

(B) Binding profiles of CTCF, SMARCA4 and MAFK on selected TE subfamilies. The TEs were aligned by strand, and the 5' ends of the match, and their locations re indicated with the dotted lines. The flanking 200 bp are shown for comparison.

(C) Pileups of selected cluster 2 TE subfamilies. The flanking 5 kbp is shown centered on the TE sequence.

(D) Pileups of ATAC-seq data for loci bound by SMARCA4, MAFK or POL2RA at the three clusters of TEs, as defined in **Fig. 4B**. The three classes correspond to: cluster 3; TEs associated with CFP-associated ETFs, cluster 2, TEs weakly associated with CFP-associated ETFs, and cluster 1; TEs not associated with CFP-associated ETFs. ATAC-seq data is from (Ma *et al.*, 2025).

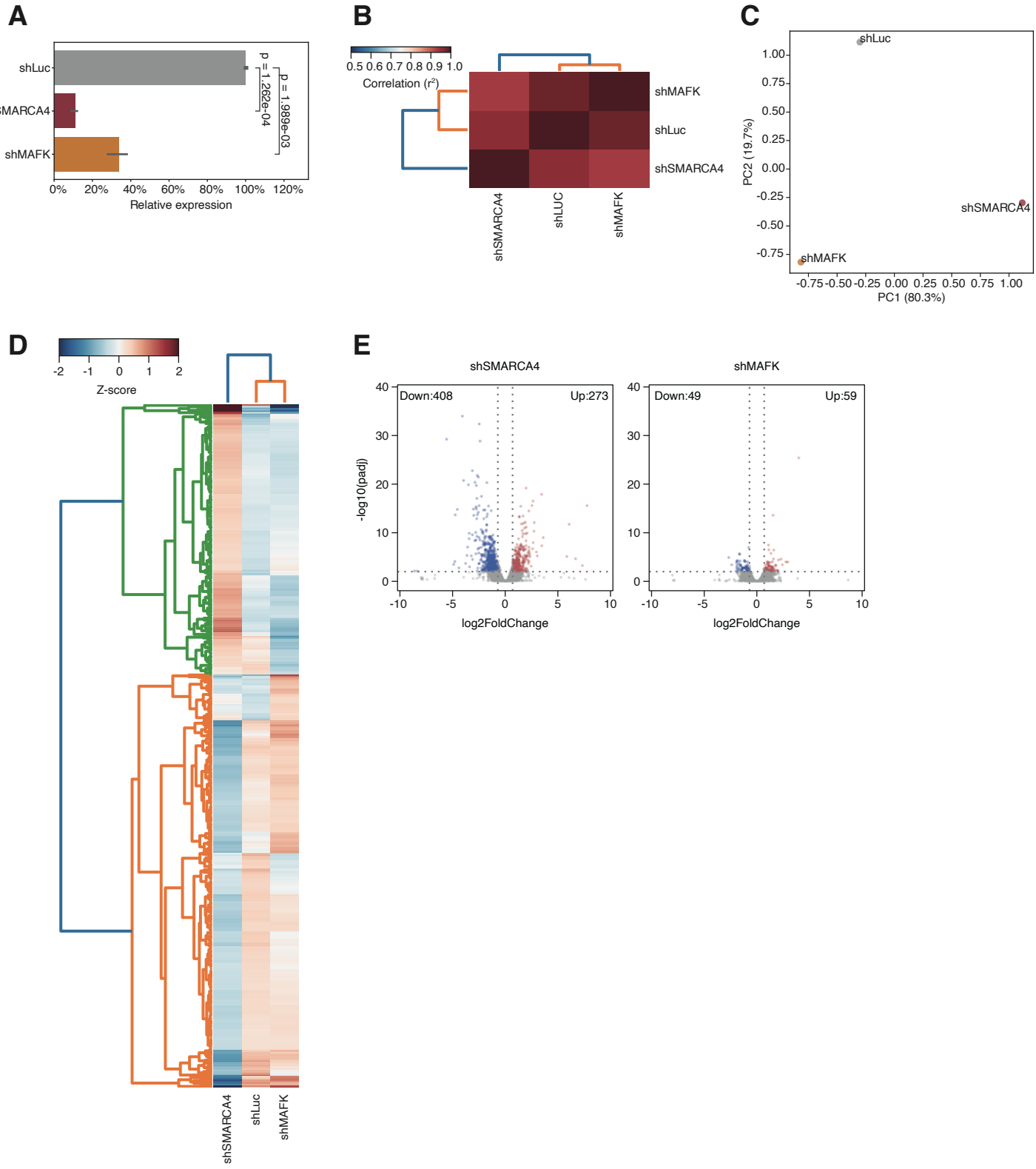
A
shLuc
shSMARCA4
shMAFK
p = 1.262e-04
p = 1.989e-03
0%
20%
40%
60%
80%
100%
120%
Relative expression
B
0.5 0.6 0.7 0.8 0.9 1.0
Correlation (r²)
shMAFK
shLuc
shSMARCA4
shSMARCA4
shLUC
shMAFK
C
shLuc
shSMARCA4
shMAFK
PC2 (19.7%)
PC1 (80.3%)
D
-2 -1 0 1 2
Z-score
shSMARCA4
shLuc
shMAFK
E
shSMARCA4
shMAFK
Down:408
Up:273
Down:49
Up:59
-log10(padj)
log2FoldChange
log2FoldChange

Fig. S8

Fig. S8. Knockdown of candidate CPFs *SMARCA4* and *MAFK*.

(A) RT-qPCR for the indicated genes in the indicated knockdowns. Statistical significance is from a two-tailed unpaired Student's t-test.

(B) Coefficient of determination (R^2) of merged RNA-seq samples.

(C) Principal component analysis (PCA) of merged RNA-seq samples.

(D) Heatmap of all differentially expressed genes in the indicated knockdown.

(E) Volcano plots for all differentially expressed genes in sh*SMARCA4* and sh*MAFK* versus sh*Luc* control. A gene was considered differentially expressed if it had a fold-change of at least 2-fold and a Bonferroni-Hochberg corrected p-value (padj) less than 0.01.

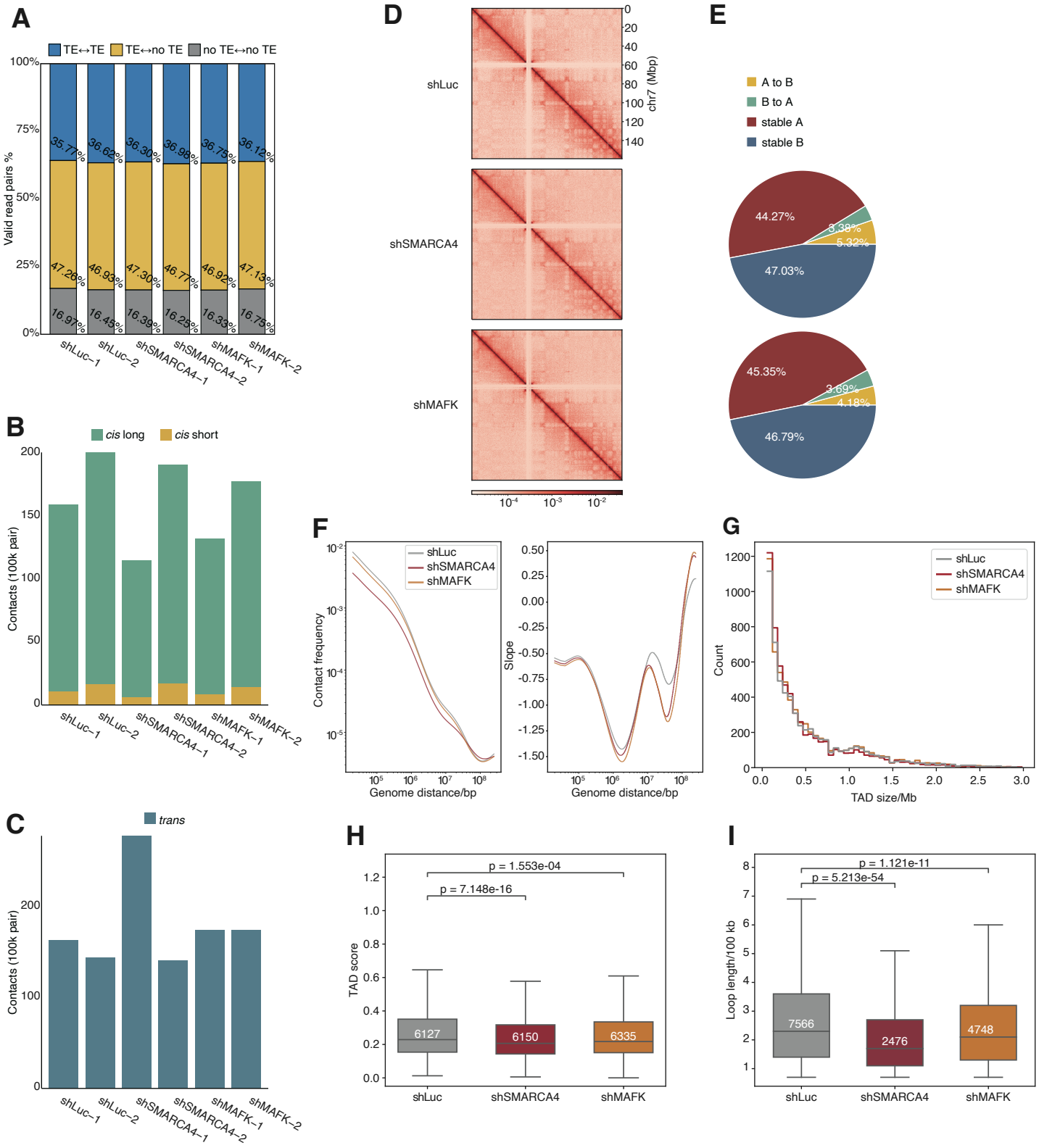
A
TE↔TE
TE↔no TE
no TE↔no TE
Valid read pairs %
100%
75%
50%
25%
0%
35.77%
36.62%
36.30%
36.98%
36.75%
36.12%
47.26%
46.93%
47.30%
46.77%
46.92%
47.13%
16.97%
16.45%
16.39%
16.25%
16.33%
16.75%
shLuc_1
shLuc_2
shSMARCA4_1
shSMARCA4_2
shMAFK_1
shMAFK_2
B
cis long
cis short
Contacts (100k pair)
200
150
100
50
0
C
trans
Contacts (100k pair)
200
100
0
D
shLuc
shSMARCA4
shMAFK
chr7 (Mbp)
0
20
40
60
80
100
120
140
E
A to B
B to A
stable A
stable B
44.27%
3.38%
5.32%
47.03%
45.35%
3.69%
4.18%
46.79%
F
shLuc
shSMARCA4
shMAFK
Contact frequency
Genome distance/bp
Slope
Genome distance/bp
G
shLuc
shSMARCA4
shMAFK
Count
TAD size/Mb
H
p = 1.553e-04
p = 7.148e-16
TAD score
6127
6150
6335
shLuc
shSMARCA4
shMAFK
I
p = 1.121e-11
p = 5.213e-54
Loop length/100 kb
7566
2476
4748
shLuc
shSMARCA4
shMAFK

Fig. S9

Fig. S9. Hi-C metrics for the knockdown of the predicted CPFs *SMARCA4* and *MAFK*

(A) Cumulative bar chart showing the proportion of reads assigned to each matrix by te_hic.

(B) Number of detected chromatin contacts in the indicated Hi-C knockdowns. *cis*-short was defined as within 20 kbp, and *cis*-long was defined as over 20 kbp.

(C) Bar chart showing the number of *trans* contacts in the indicated knockdown.

(D) Heatmaps of the normalized contact intensity of chromosome 7 for the indicated knockdowns using the all-reads matrices generated by te_hic.

(E) Pie charts showing the proportion of compartment switches in the indicated knockdowns, relative to the sh*Luc* control knockdown. Compartments were defined at 150 kb resolution. Only bins with same sign in PC1 values in two replicates were kept.

(F) Contact frequency against genomic distance. Right panel: slopes of the curves.

(G) Histogram of the TAD domain size distribution in the indicated knockdowns.

(H) Boxplot showing the TAD boundary strength, and the number of boundaries detected (white). Statistical significance is from a two-tailed Mann-Whitney U test.

(I) Boxplots of loop length of the indicated knockdowns. Number of loops are indicated (white). Statistical significance is from a Mann-Whitney U test.

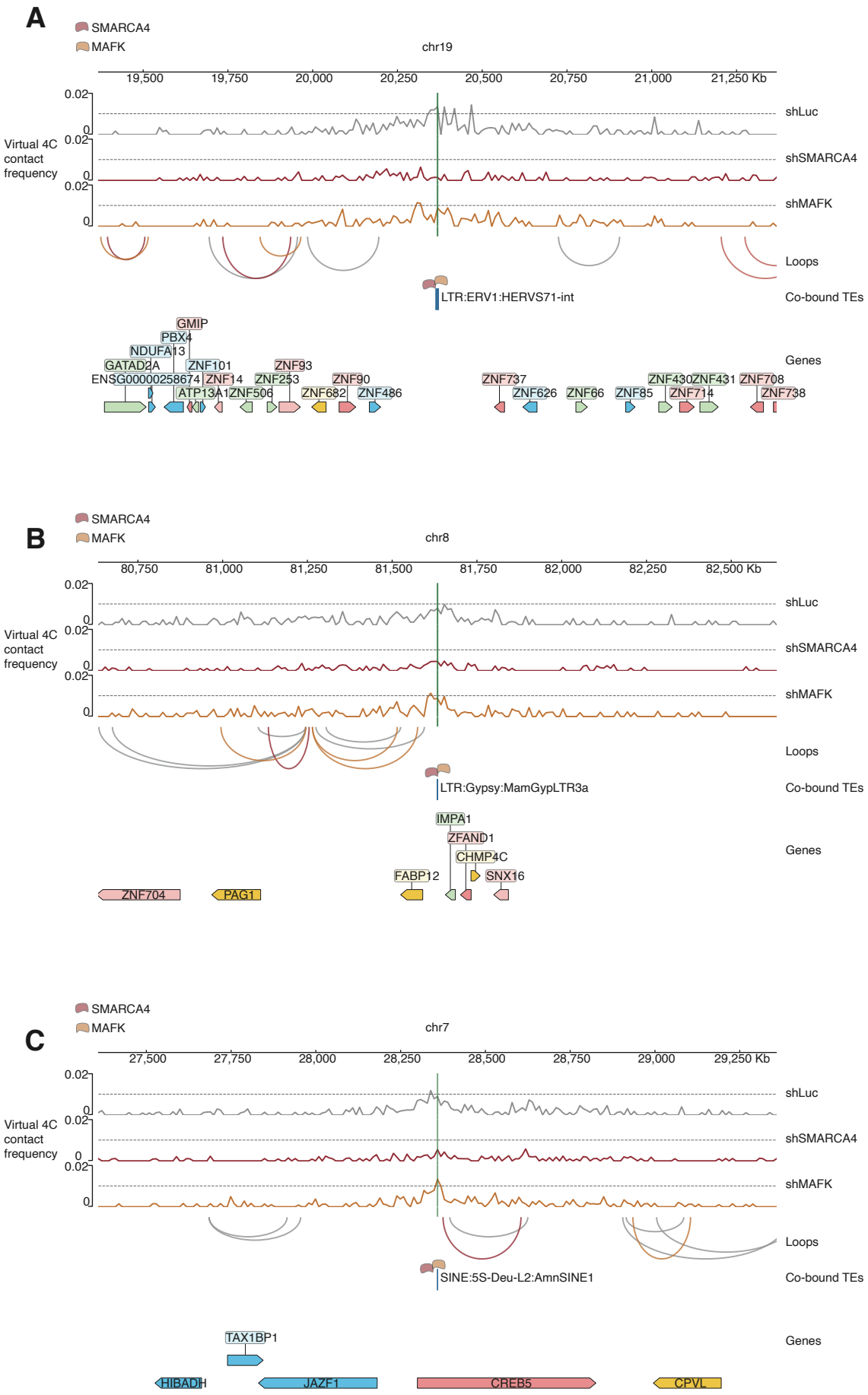
A
SMARCA4
MAFK
chr19
19,500
19,750
20,000
20,250
20,500
20,750
21,000
21,250 Kb
Virtual 4C contact frequency
0.02
0
shLuc
shSMARCA4
shMAFK
Loops
LTR:ERV1:HERVS71-int
Co-bound TEs
GMIP
PBX4
NDUFA13
GATAD2A
ZNF101
ZNF93
ENSG00000258674
ZNF14
ZNF253
ZNF90
ATP13A1
ZNF506
ZNF682
ZNF486
ZNF737
ZNF626
ZNF66
ZNF85
ZNF430
ZNF431
ZNF708
ZNF714
ZNF738
Genes
B
SMARCA4
MAFK
chr8
80,750
81,000
81,250
81,500
81,750
82,000
82,250
82,500 Kb
Virtual 4C contact frequency
0.02
0
shLuc
shSMARCA4
shMAFK
Loops
LTR:Gypsy:MamGypLTR3a
Co-bound TEs
IMPA1
ZFAND1
CHMP4C
FABP12
SNX16
ZNF704
PAG1
Genes
C
SMARCA4
MAFK
chr7
27,500
27,750
28,000
28,250
28,500
28,750
29,000
29,250 Kb
Virtual 4C contact frequency
0.02
0
shLuc
shSMARCA4
shMAFK
Loops
SINE:5S-Deu-L2:AmnSINE1
Co-bound TEs
TAX1BP1
HIBADH
JAZF1
CREB5
CPVL
Genes

Fig. S10

Fig. S10. Virtual 4C views at selected loci for the CPFs.

(A) Virtual 4C views of sh*Luc*, sh*SMARCA4*, and sh*MAFK* Hi-C data centered on a cluster 3 TE, LTR:ERV1:HERVS71-int on chromosome 19, as the viewpoint showing the flanking 1 Mbp from the center. Both SMARCA4 and MAFK is bound to the TE. Co-binding of SMARCA4 and MAFK is indicated.

(B) As in **panel A** but centered on an LTR:Gypsy:MamGypLTR3a on chromosome 8.

(C) As in **panel A** but centered on a SINE:5S-Deu-L2:AmnSINE on chromosome 7.