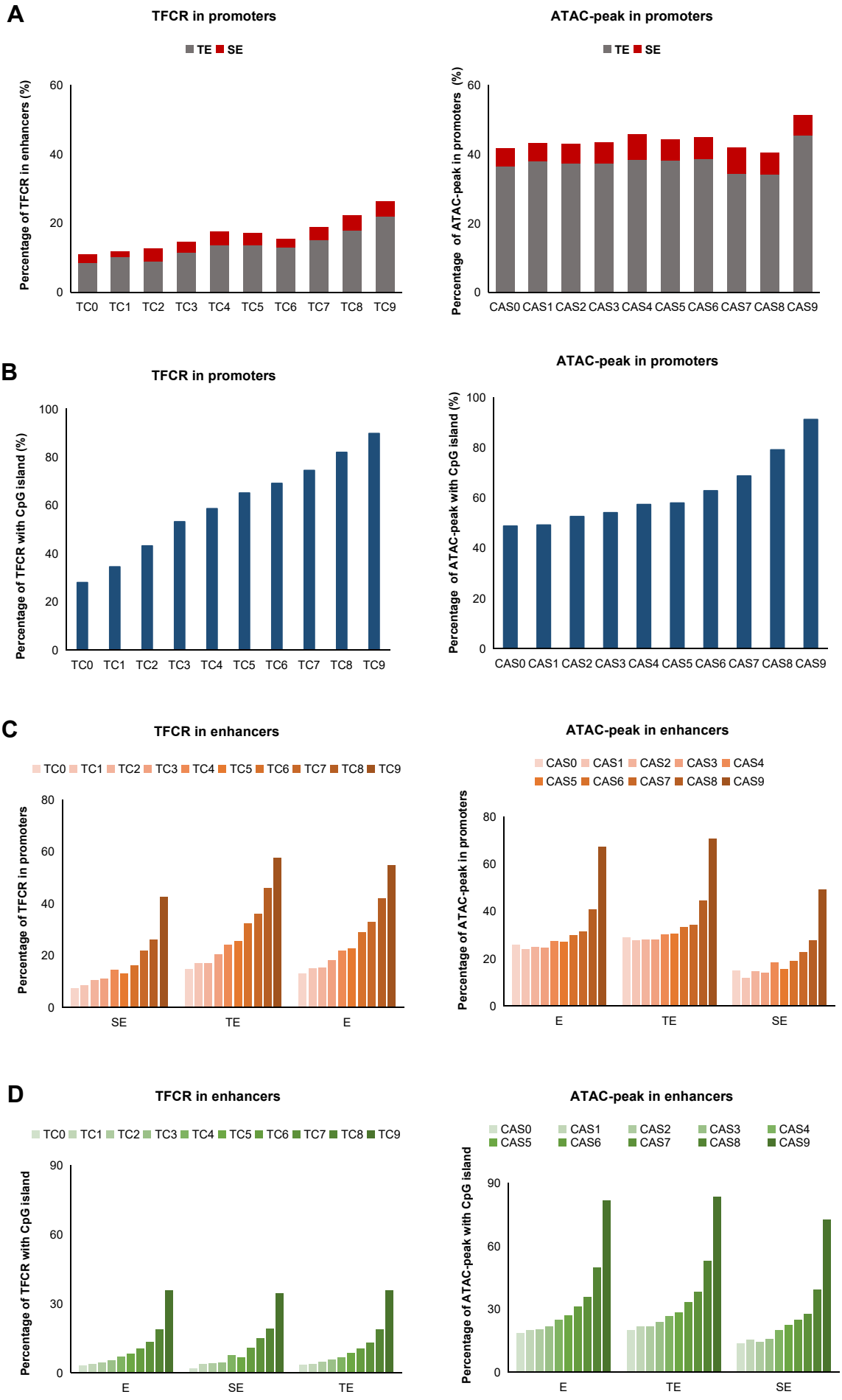

A
TFCR in promoters
TE SE
Percentage of TFCR in enhancers (%)
60
40
20
0
TC0 TC1 TC2 TC3 TC4 TC5 TC6 TC7 TC8 TC9
ATAC-peak in promoters
TE SE
Percentage of ATAC-peak in promoters (%)
60
40
20
0
CAS0 CAS1 CAS2 CAS3 CAS4 CAS5 CAS6 CAS7 CAS8 CAS9
B
TFCR in promoters
Percentage of TFCR with CpG island (%)
100
80
60
40
20
0
TC0 TC1 TC2 TC3 TC4 TC5 TC6 TC7 TC8 TC9
ATAC-peak in promoters
Percentage of ATAC-peak with CpG island (%)
100
80
60
40
20
0
CAS0 CAS1 CAS2 CAS3 CAS4 CAS5 CAS6 CAS7 CAS8 CAS9
C
TFCR in enhancers
TC0 TC1 TC2 TC3 TC4 TC5 TC6 TC7 TC8 TC9
Percentage of TFCR in promoters
80
60
40
20
0
SE TE E
ATAC-peak in enhancers
CAS0 CAS1 CAS2 CAS3 CAS4
CAS5 CAS6 CAS7 CAS8 CAS9
Percentage of ATAC-peak in promoters
80
60
40
20
0
E TE SE
D
TFCR in enhancers
TC0 TC1 TC2 TC3 TC4 TC5 TC6 TC7 TC8 TC9
Percentage of TFCR with CpG island
90
60
30
0
E SE TE
ATAC-peak in enhancers
CAS0 CAS1 CAS2 CAS3 CAS4
CAS5 CAS6 CAS7 CAS8 CAS9
Percentage of ATAC-peak with CpG island
90
60
30
0
E TE SE

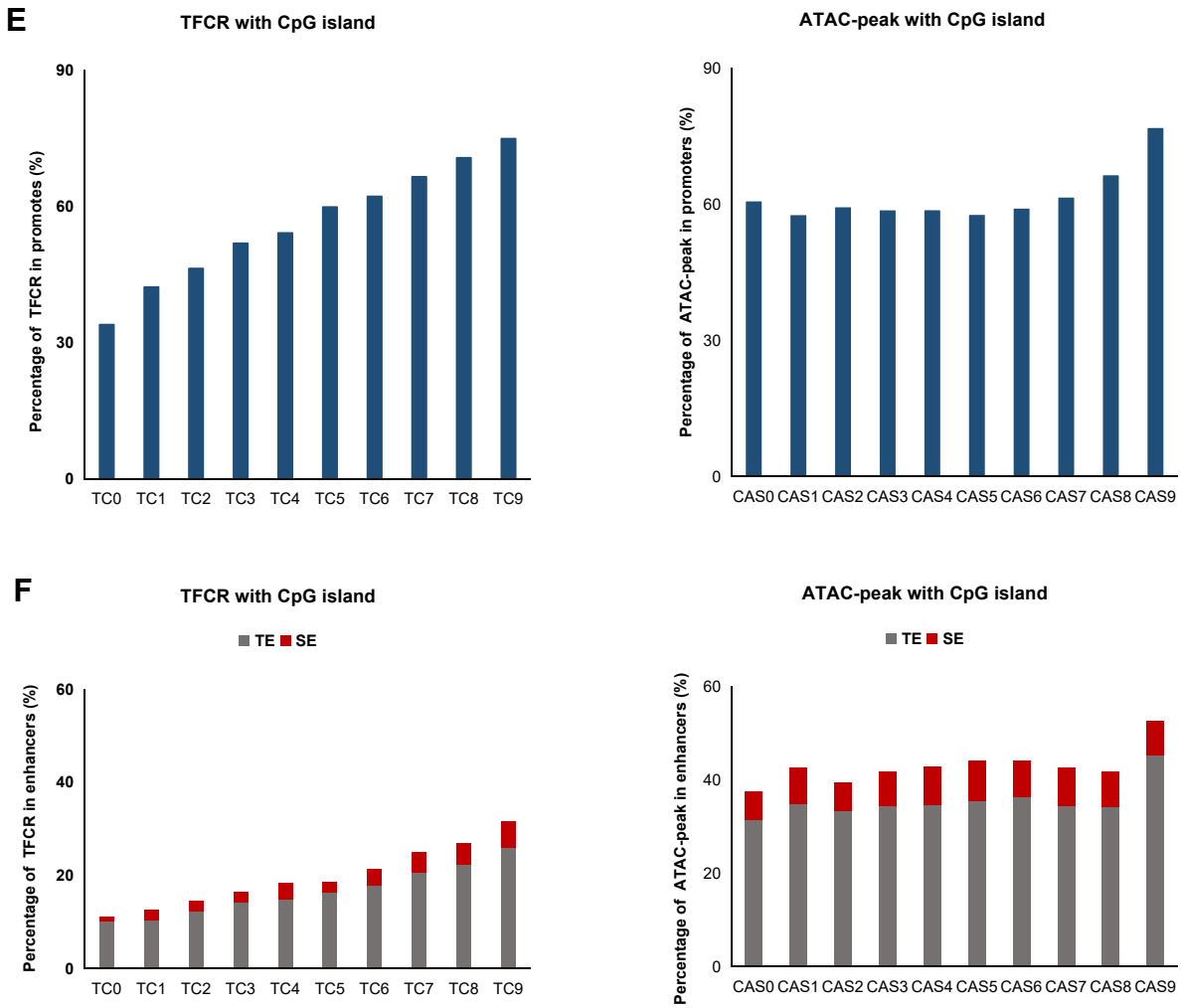
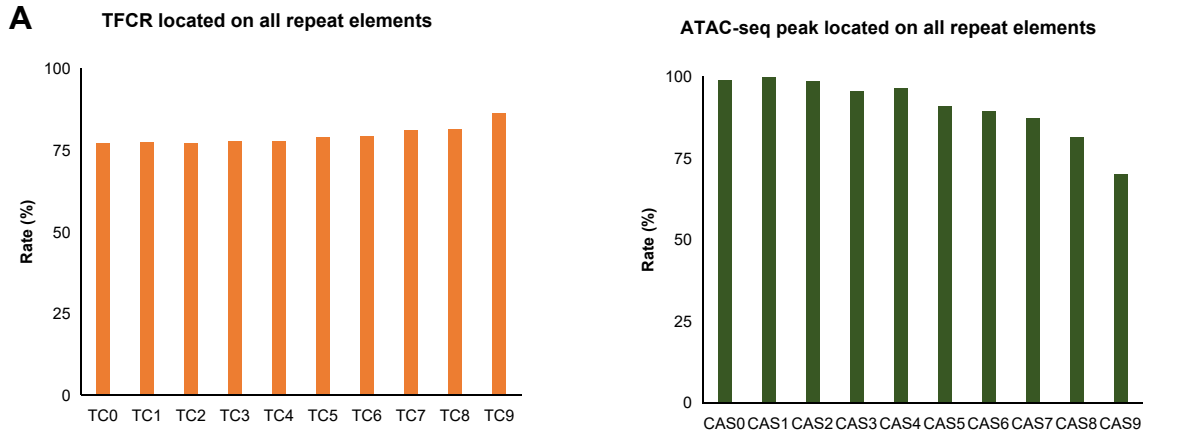
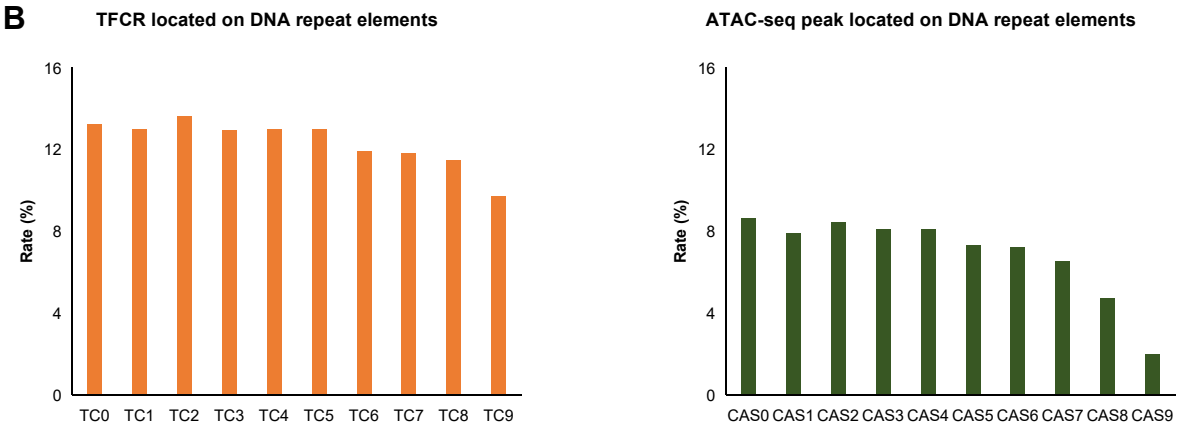
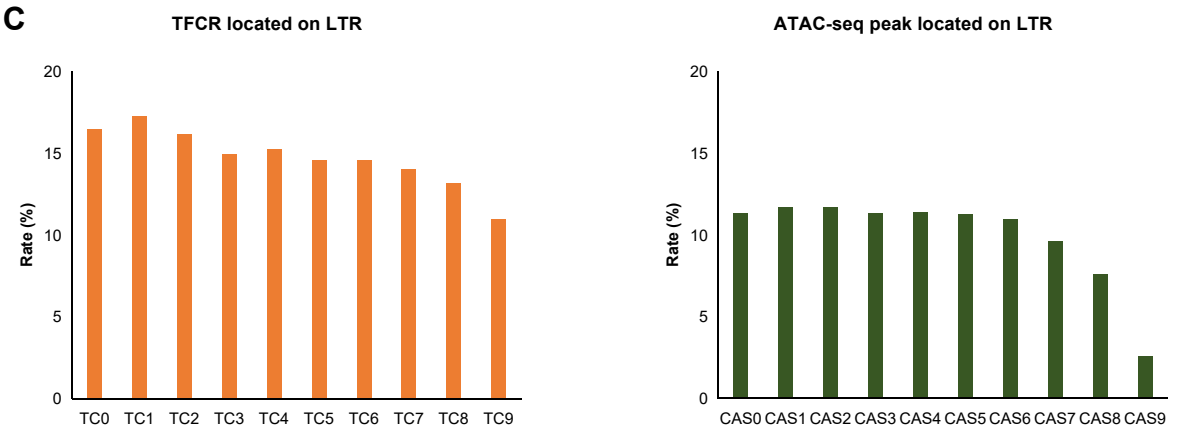
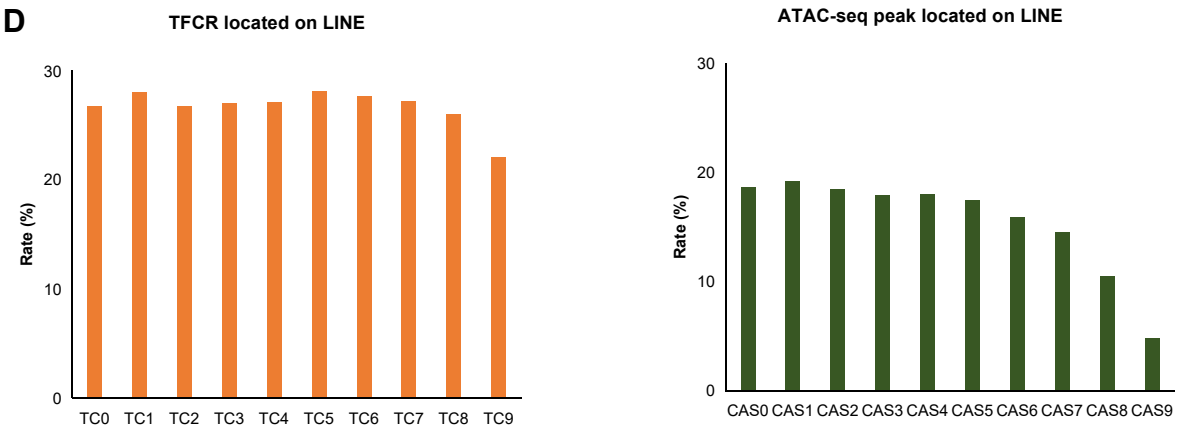


Figure S1 Features of TFCRs and ATAC-seq peaks at genomic locations. (A) Barplots shows the proportion of TFCRs (left) or ATAC-seq peaks (right) overlapped with promoter, distributed in enhancer region. (B) Barplots shows the proportion of TFCRs (left) or ATAC-seq peaks (right) overlapped with promoter, distributed in CpG island. (C) Barplots shows the proportion of TFCRs (left) or ATAC-seq peaks (right) overlapped with enhancer, distributed in promoter region. (D) Barplots shows the proportion of TFCRs (left) or ATAC-seq peaks (right) overlapped with enhancer, distributed in CpG island. (E) Barplots shows the proportion of TFCRs (left) or ATAC-seq peaks (right) overlapped with CpG island, distributed in promoter region. (F) Barplots shows the proportion of TFCRs (left) or ATAC-seq peaks (right) overlapped with CpG island, distributed in enhancer region.


A
TFCR located on all repeat elements
ATAC-seq peak located on all repeat elements
Rate (%)
100
75
50
25
0
TC0 TC1 TC2 TC3 TC4 TC5 TC6 TC7 TC8 TC9
CAS0 CAS1 CAS2 CAS3 CAS4 CAS5 CAS6 CAS7 CAS8 CAS9


B
TFCR located on DNA repeat elements
ATAC-seq peak located on DNA repeat elements
Rate (%)
16
12
8
4
0
TC0 TC1 TC2 TC3 TC4 TC5 TC6 TC7 TC8 TC9
CAS0 CAS1 CAS2 CAS3 CAS4 CAS5 CAS6 CAS7 CAS8 CAS9


C
TFCR located on LTR
ATAC-seq peak located on LTR
Rate (%)
20
15
10
5
0
TC0 TC1 TC2 TC3 TC4 TC5 TC6 TC7 TC8 TC9
CAS0 CAS1 CAS2 CAS3 CAS4 CAS5 CAS6 CAS7 CAS8 CAS9


D
TFCR located on LINE
ATAC-seq peak located on LINE
Rate (%)
30
20
10
0
TC0 TC1 TC2 TC3 TC4 TC5 TC6 TC7 TC8 TC9
CAS0 CAS1 CAS2 CAS3 CAS4 CAS5 CAS6 CAS7 CAS8 CAS9

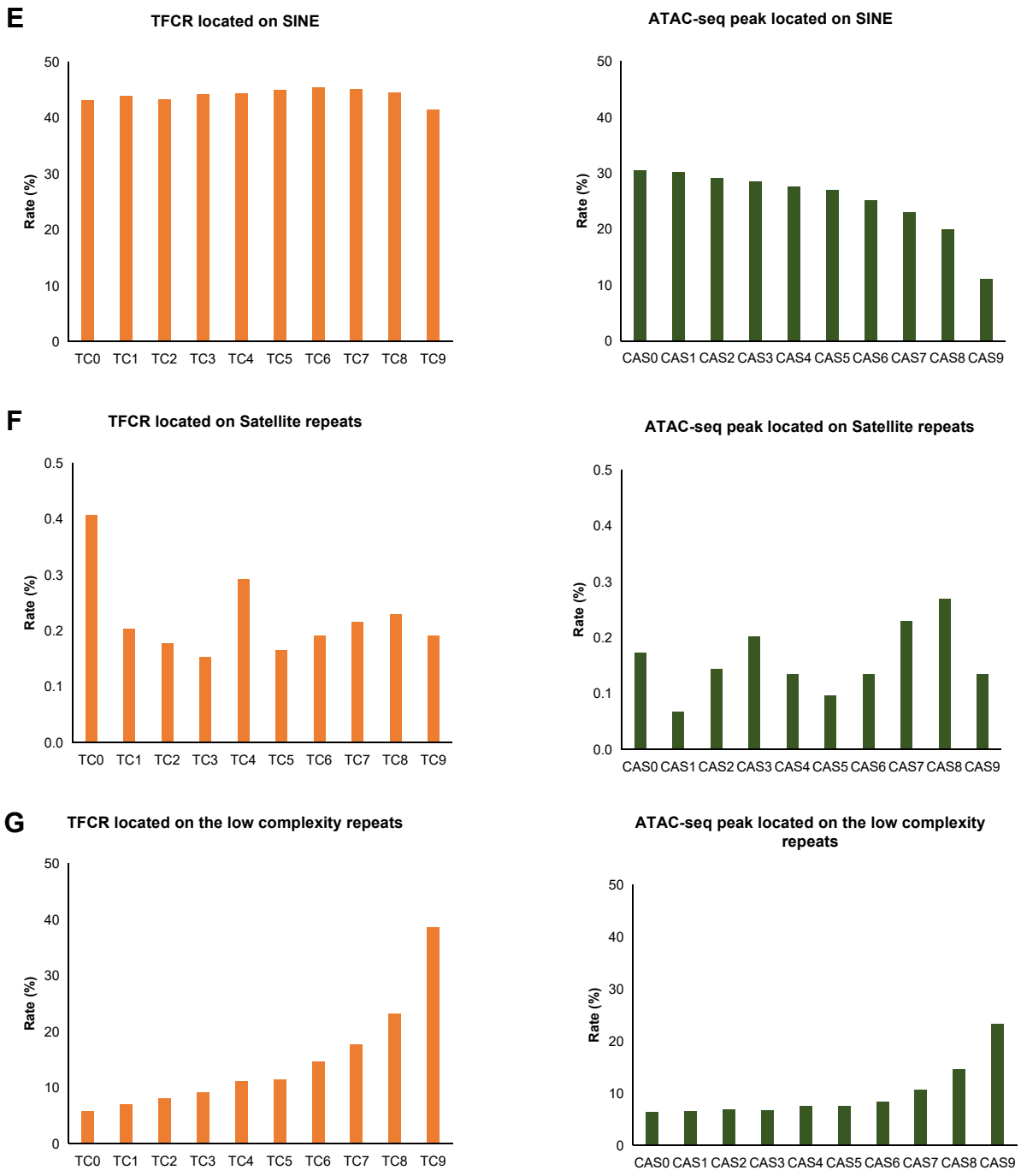


Figure S2 Relationship between TFCR or ATAC-seq peak and repeat elements in the genome.

(A-G) correspond to all repeat elements, DNA repeat elements, LTR, LINE, SINE, Satellite repeats and the low complexity repeats, respectively.

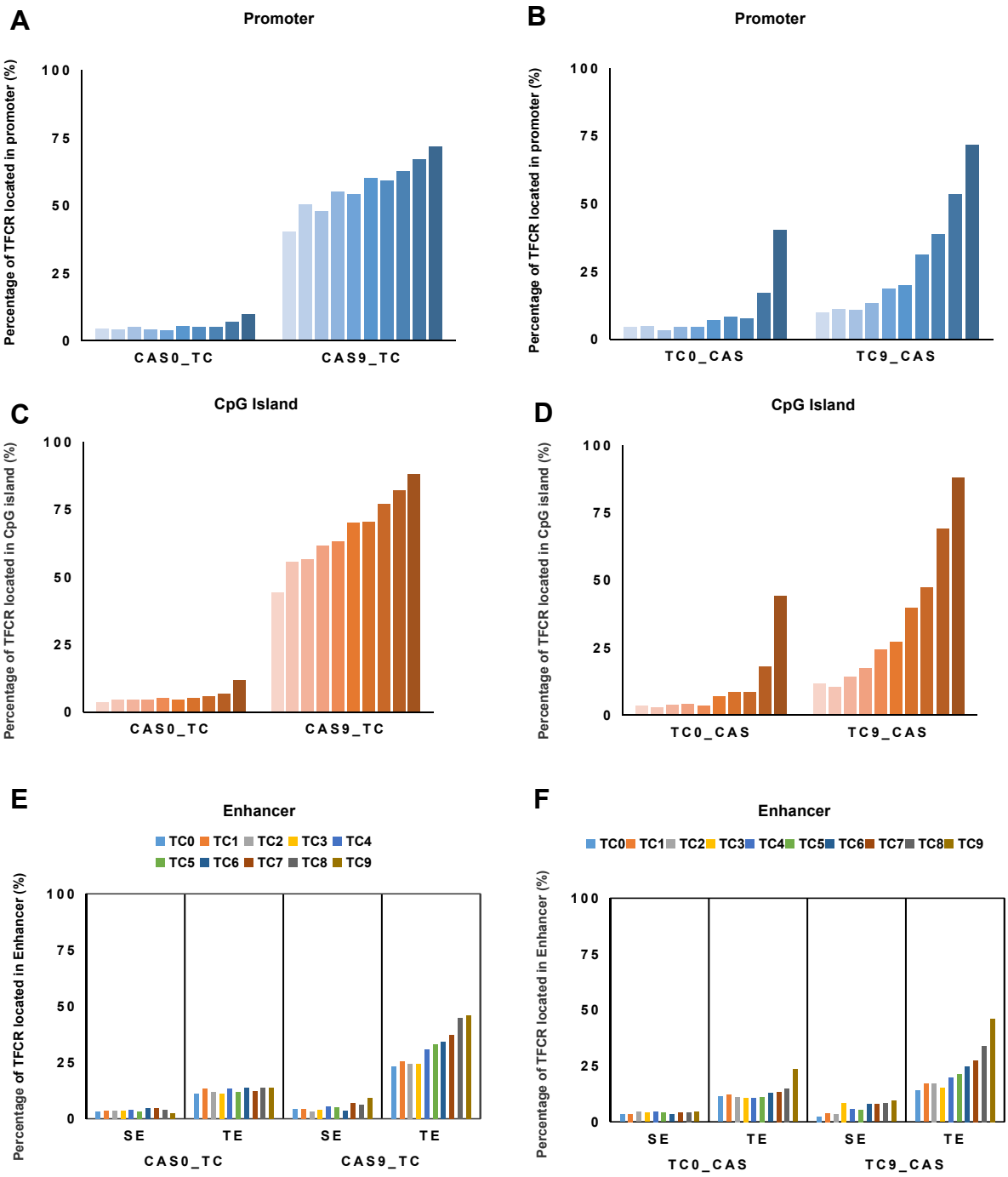


Figure S3 The distribution of TFCR with transcription factor complexity and chromatin accessibility imbalance on the main regulatory elements. (A-F) the proportion of TFCRs with different chromatin accessibility (CAS0 and CAS9) or transcription factor complexity (TC0 and TC9) near gene promoters (A, B), CpG island (C, D), and enhancers (E, F) as a function of altered TC or CAS grade.

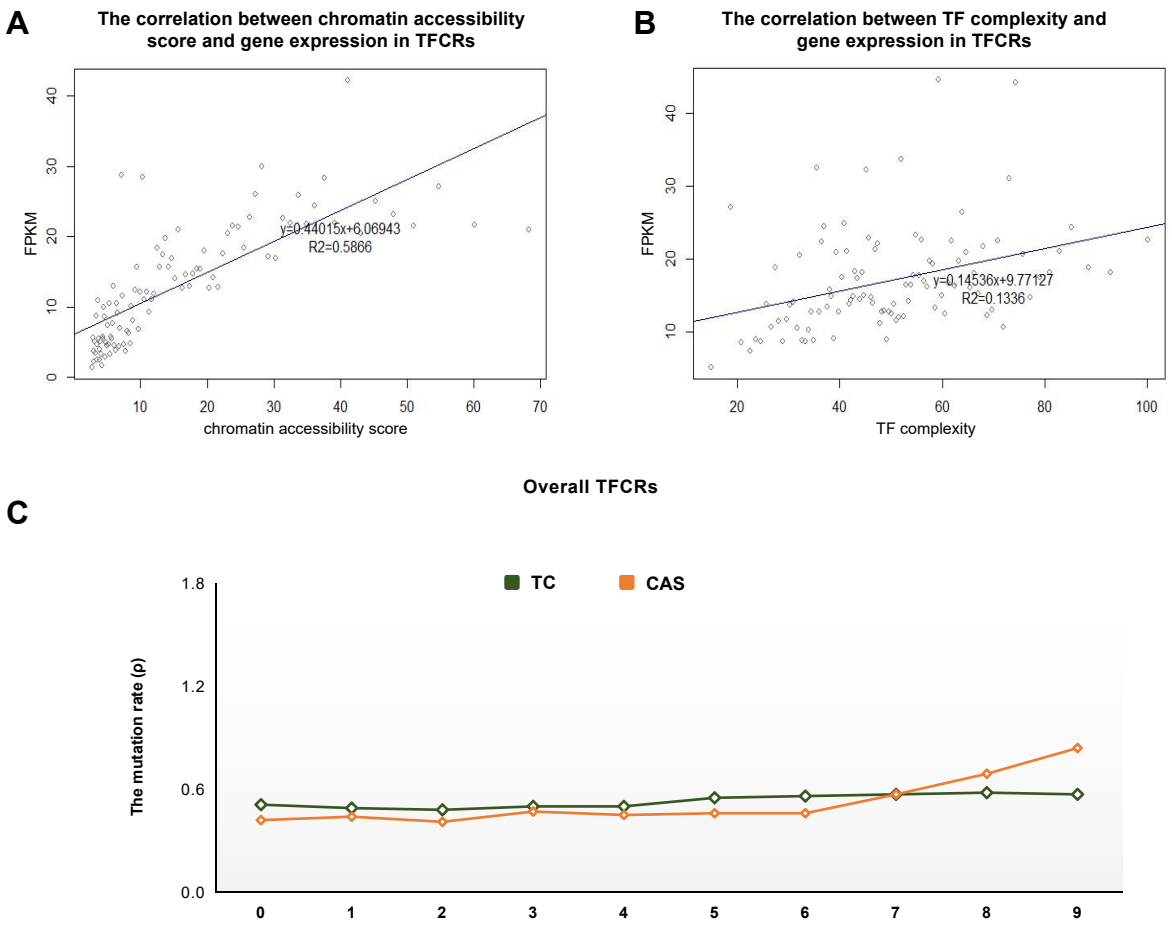
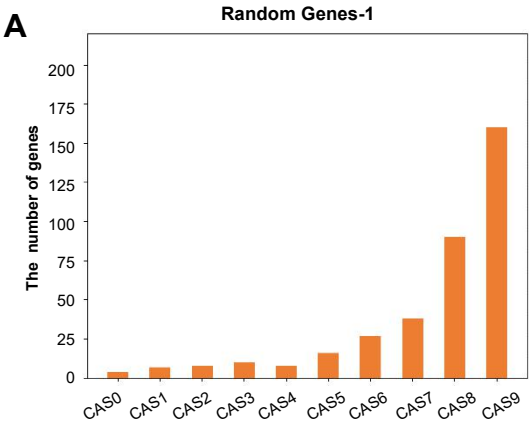
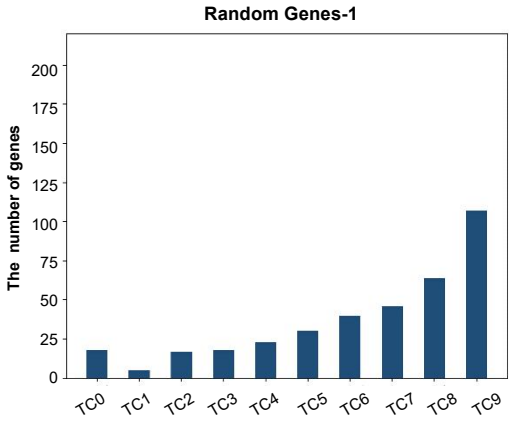
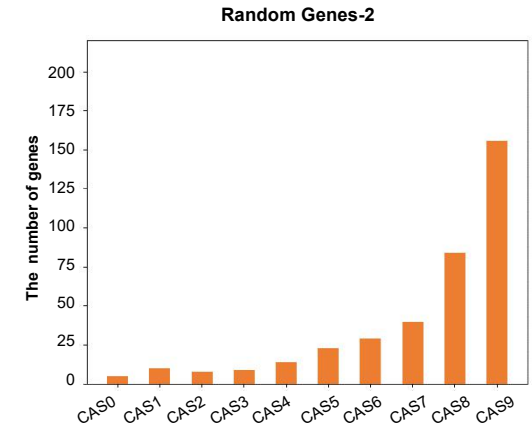


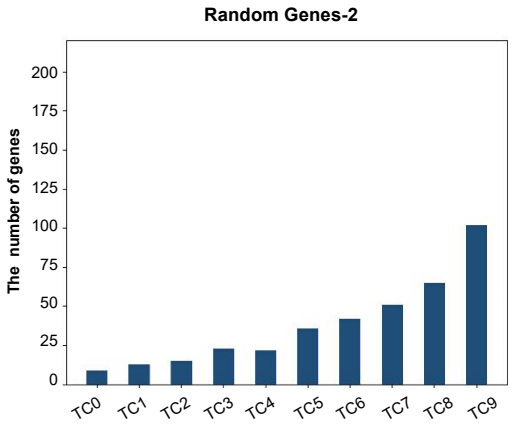
Figure S4 Correlation analysis of chromatin accessibility score and TF complexity with gene expression in TFCRs. Scatter plots show that chromatin accessibility score (A, R^2 = 0.59) and TF complexity (B, R^2 = 0.13) are positively correlated with FPKM in TFCRs. (C) The line graph shows the overall mutation rate of TFCRs (CAS: Orange; TC: Darkgreen).

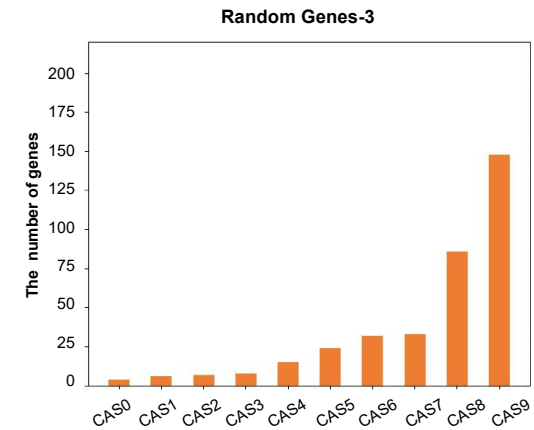
A

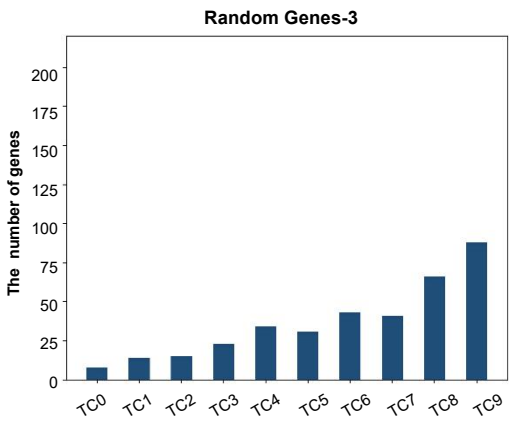

Random Genes-1
The number of genes
0
25
50
75
100
125
150
175
200
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9

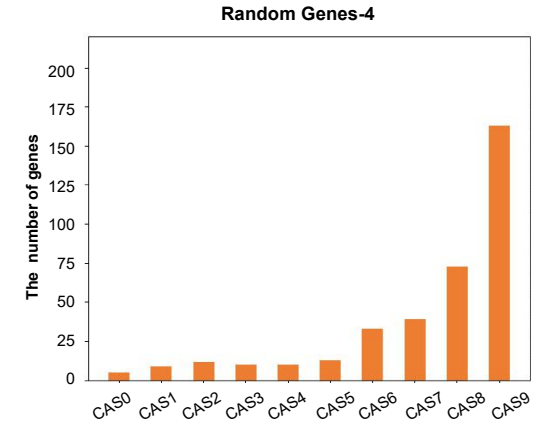

Random Genes-1
The number of genes
0
25
50
75
100
125
150
175
200
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9

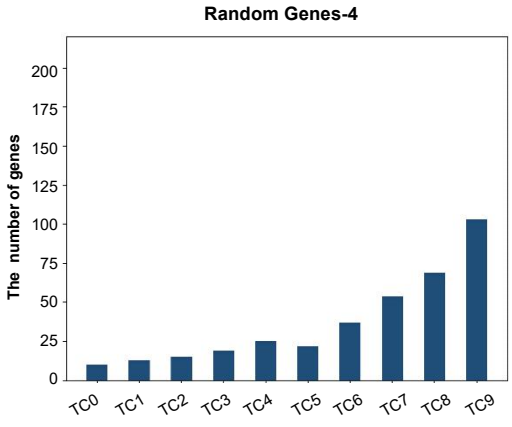

Random Genes-2
The number of genes
0
25
50
75
100
125
150
175
200
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9


Random Genes-2
The number of genes
0
25
50
75
100
125
150
175
200
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9

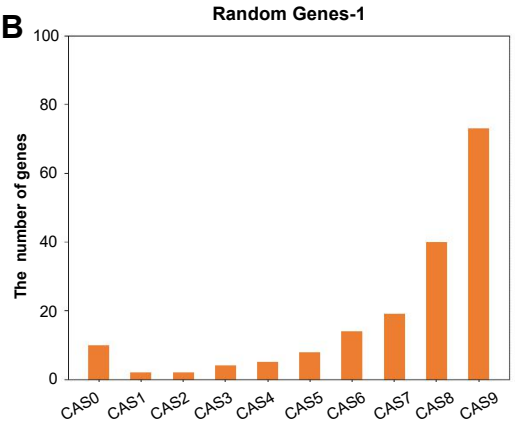

Random Genes-3
The number of genes
0
25
50
75
100
125
150
175
200
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9

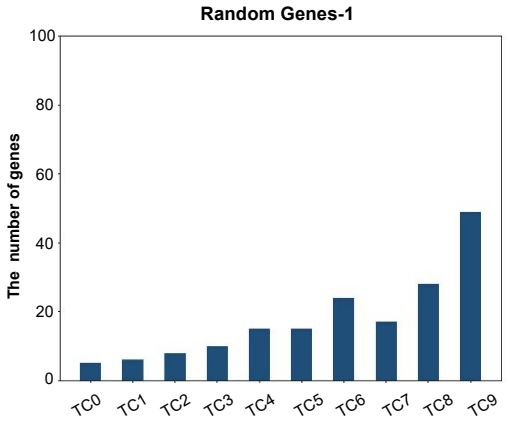

Random Genes-3
The number of genes
0
25
50
75
100
125
150
175
200
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9

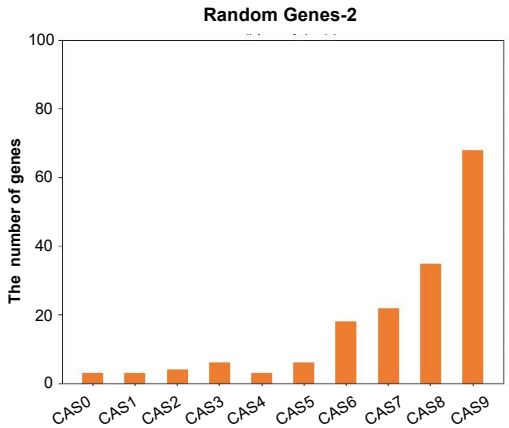

Random Genes-4
The number of genes
0
25
50
75
100
125
150
175
200
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9

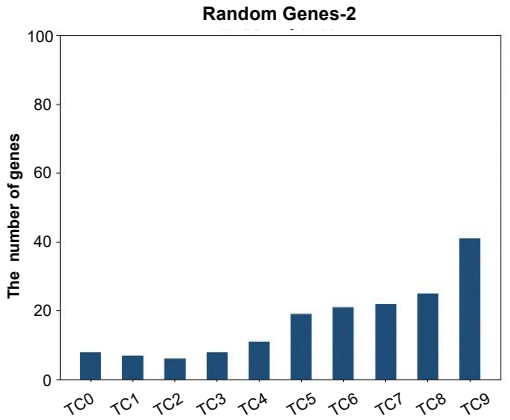

Random Genes-4
The number of genes
0
25
50
75
100
125
150
175
200
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9

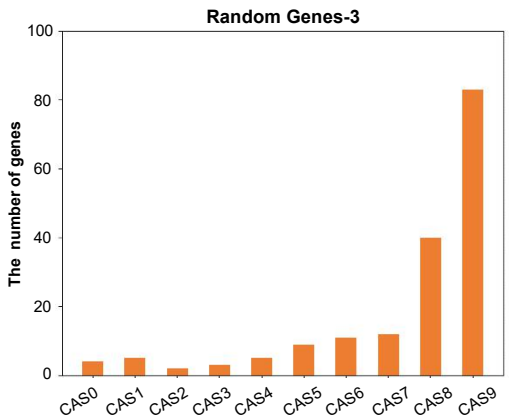
B

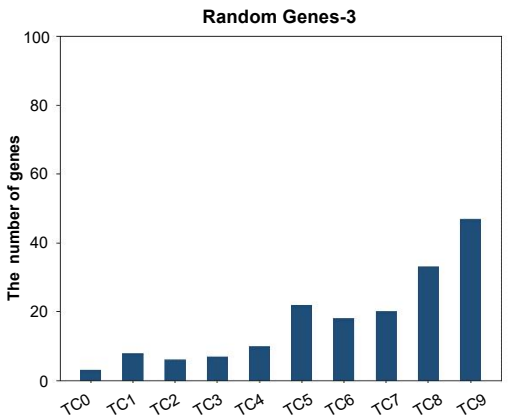

Random Genes-1
The number of genes
0
20
40
60
80
100
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9

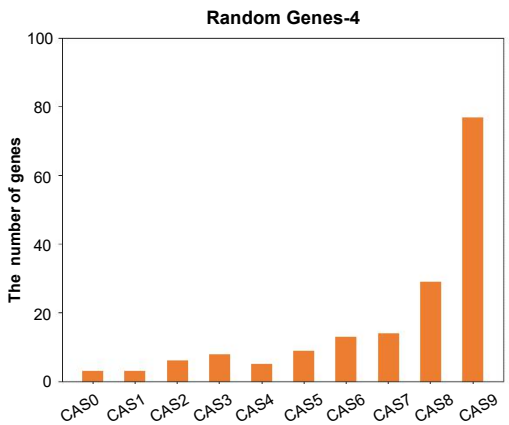

Random Genes-1
The number of genes
0
20
40
60
80
100
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9

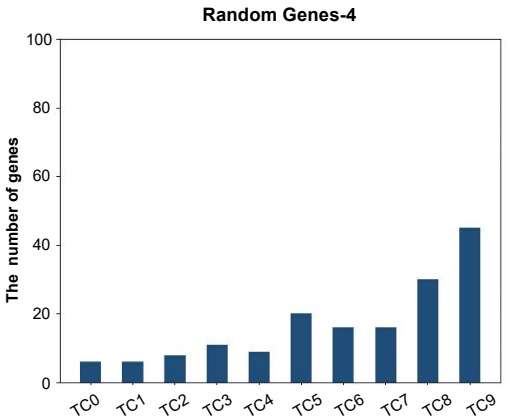

Random Genes-2
The number of genes
0
20
40
60
80
100
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9


Random Genes-2
The number of genes
0
20
40
60
80
100
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9

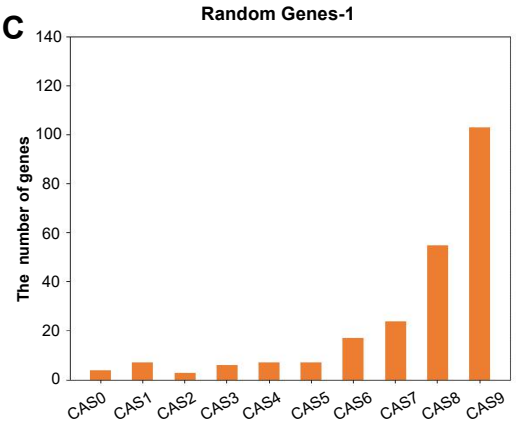

Random Genes-3
The number of genes
0
20
40
60
80
100
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9

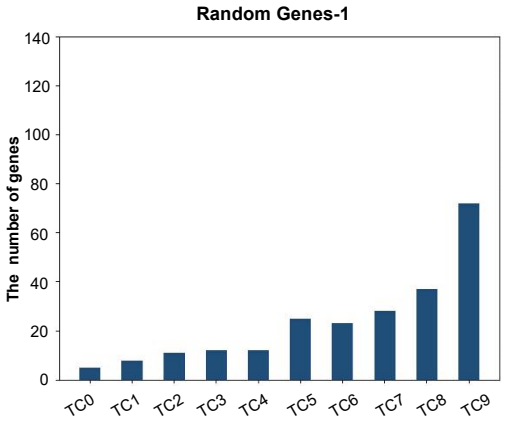

Random Genes-3
The number of genes
0
20
40
60
80
100
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9

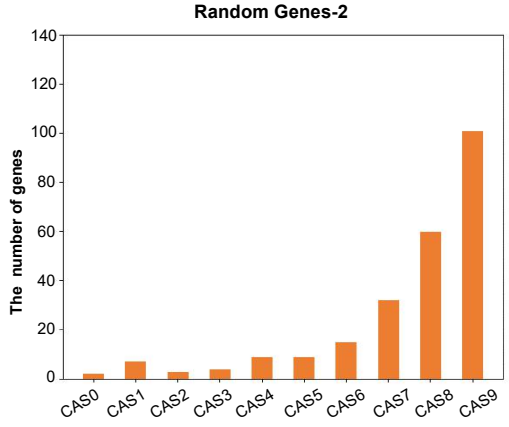

Random Genes-4
The number of genes
0
20
40
60
80
100
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9

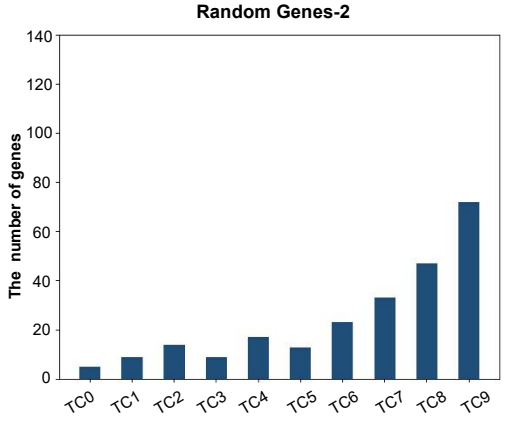

Random Genes-4
The number of genes
0
20
40
60
80
100
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9

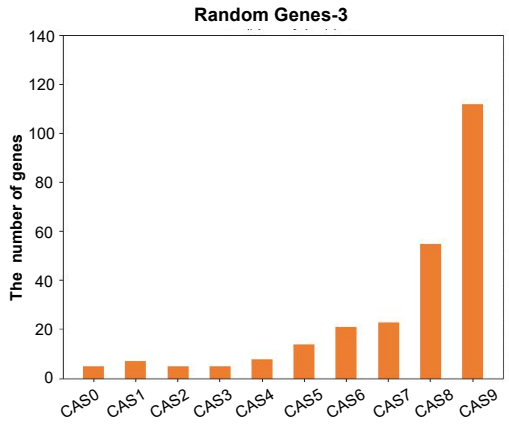
C

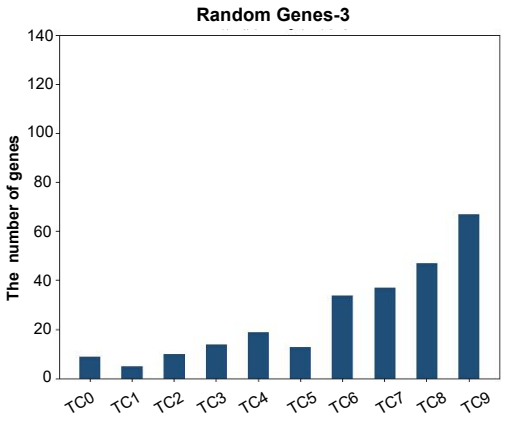

Random Genes-1
The number of genes
0
20
40
60
80
100
120
140
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9

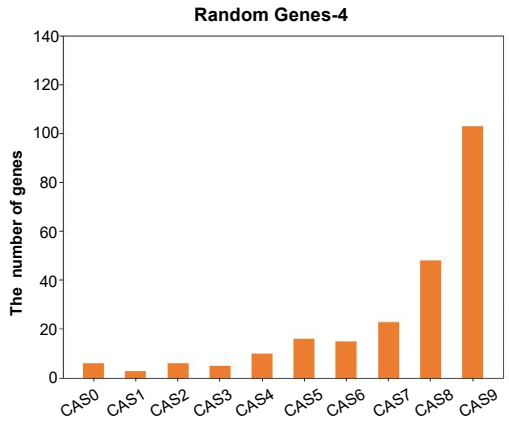

Random Genes-1
The number of genes
0
20
40
60
80
100
120
140
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9

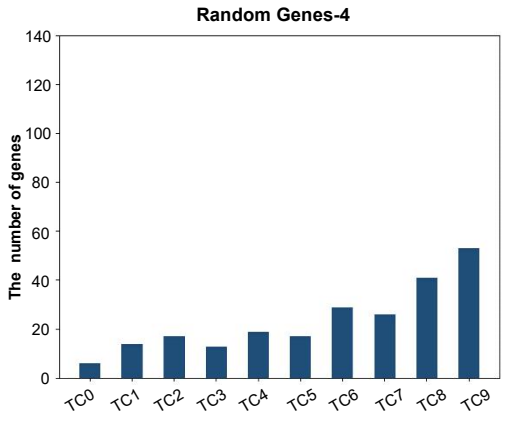

Random Genes-2
The number of genes
0
20
40
60
80
100
120
140
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9


Random Genes-2
The number of genes
0
20
40
60
80
100
120
140
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9


Random Genes-3
The number of genes
0
20
40
60
80
100
120
140
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9


Random Genes-3
The number of genes
0
20
40
60
80
100
120
140
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9


Random Genes-4
The number of genes
0
20
40
60
80
100
120
140
CAS0
CAS1
CAS2
CAS3
CAS4
CAS5
CAS6
CAS7
CAS8
CAS9


Random Genes-4
The number of genes
0
20
40
60
80
100
120
140
TC0
TC1
TC2
TC3
TC4
TC5
TC6
TC7
TC8
TC9

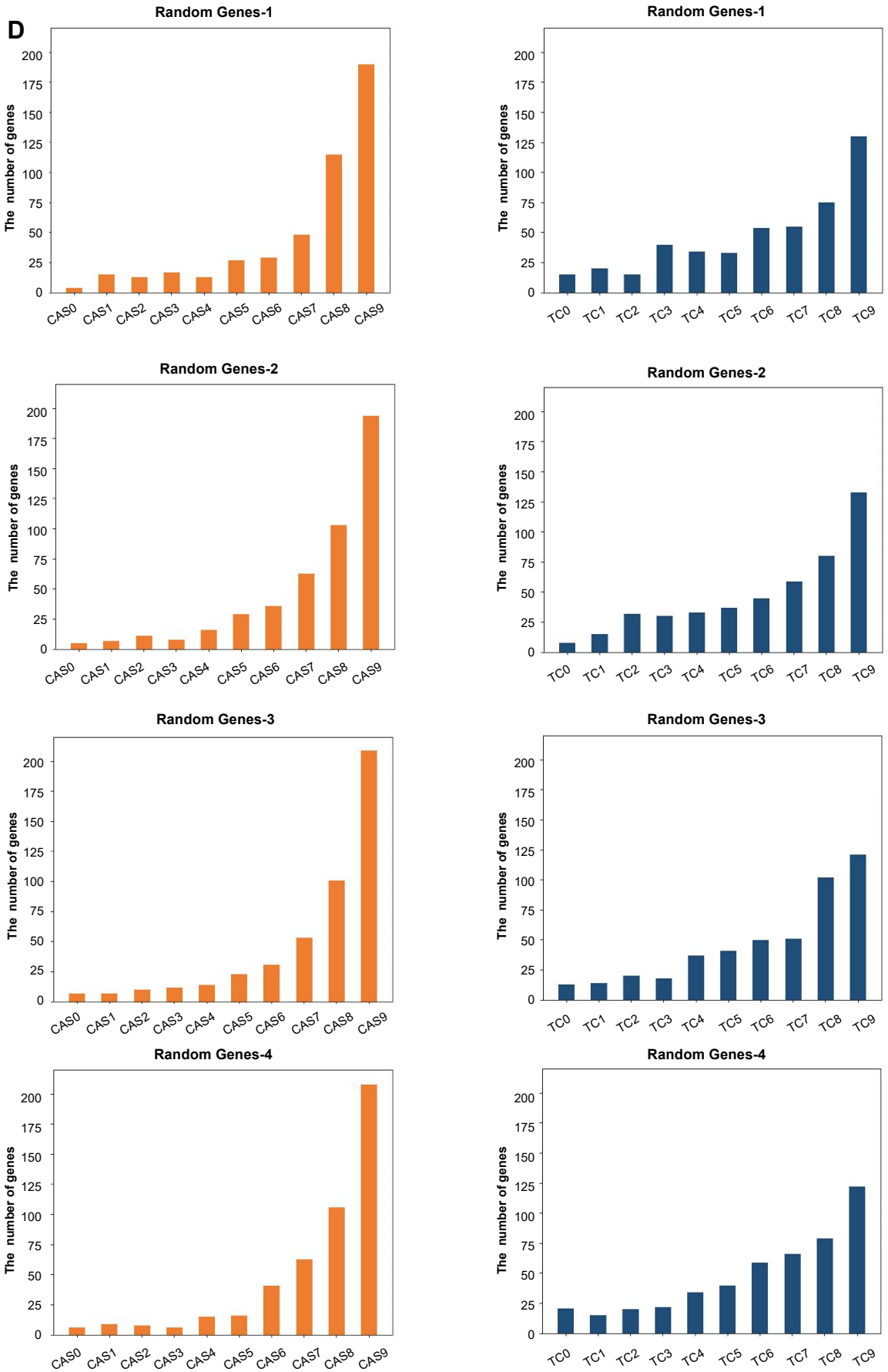


Figure S5 Random controls for tumor key genes. Generating random gene sets with equal numbers to the literature reported cancer driver gene sets and examination of the distribution of these genes across TFCR different TF complexity (TC, Darkblue) groups versus chromatin accessibility score (CAS, Orange) groups. Four cancer driver gene sets are shown as examples in which A to D correspond to the cancer driving genes reported in Nature (2014), Nature Reviews (2018), Cell (2020) and Nature Reviews Cancer (2020), respectively. Each control experiment was repeated for more than four times.

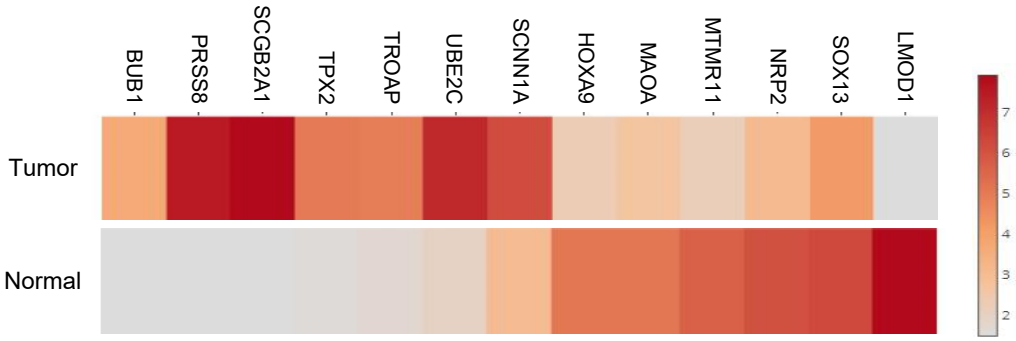
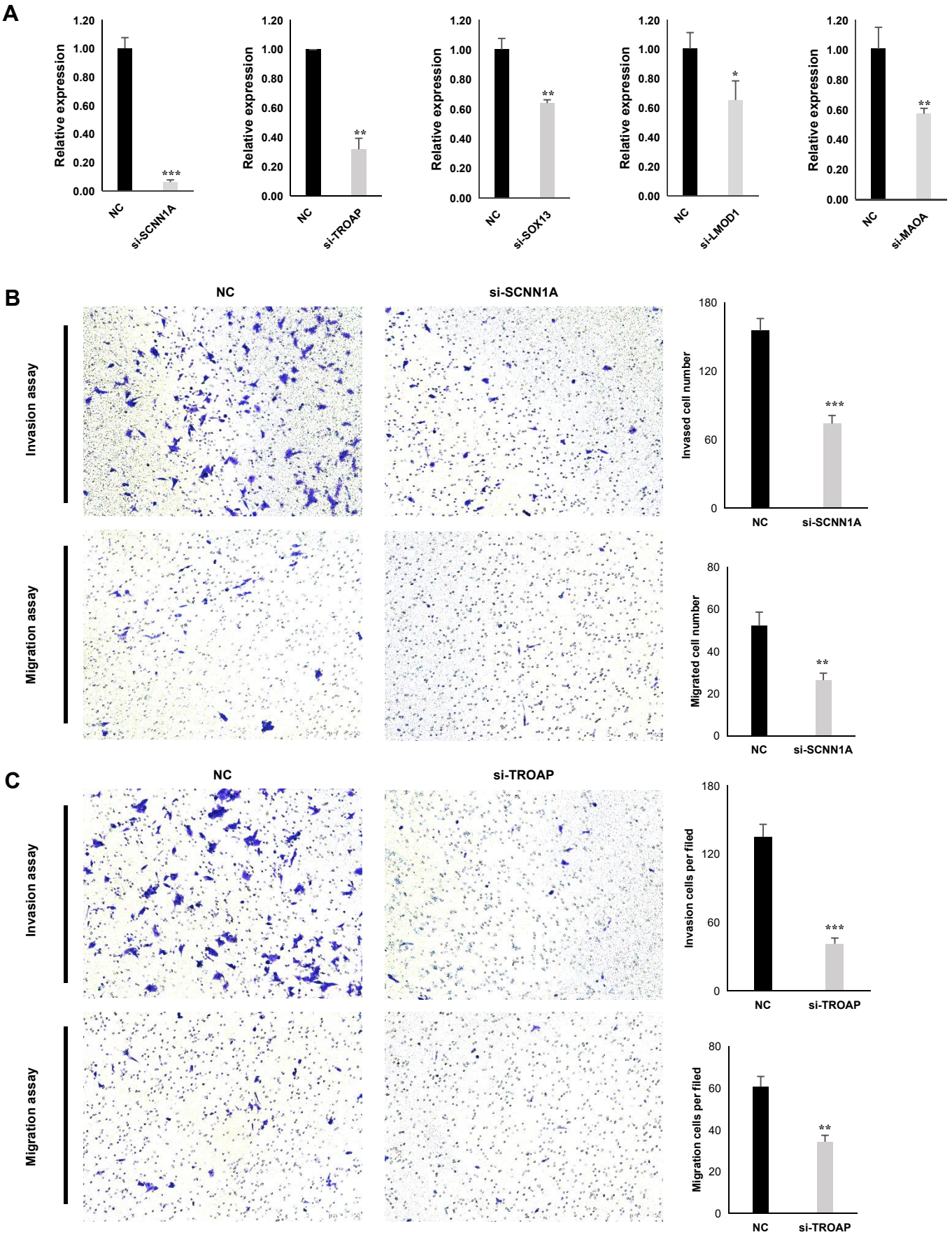


Figure S6 Heatmap of differential expression of candidate cancer key genes. GEPIA2 was used to investigate the expression of candidate cancer key genes in tumor tissues and normal tissues, among which 13 (*LMOD1*, *SOX13*, *NRP2*, *MTMR11*, *MAOA*, *HOXA9*, *SCNN1A*, *UBE2C*, *TROAP*, *TPX2*, *SCGB2A1*, *PRSS8* and *BUB1*) showed statistically significant differences ($log_2FC \geq 3$, $P < 0.001$).


A
Relative expression
1.20
1.00
0.80
0.60
0.40
0.20
0.00
***
NC
si-SCNN1A
**
si-TROAP
si-SOX13
*
si-LMOD1
si-MAOA
B
NC
si-SCNN1A
Invasion assay
Migration assay
Invased cell number
180
120
60
0
Migrated cell number
80
40
20
C
si-TROAP
Invasion cells per filed
Migration cells per filed

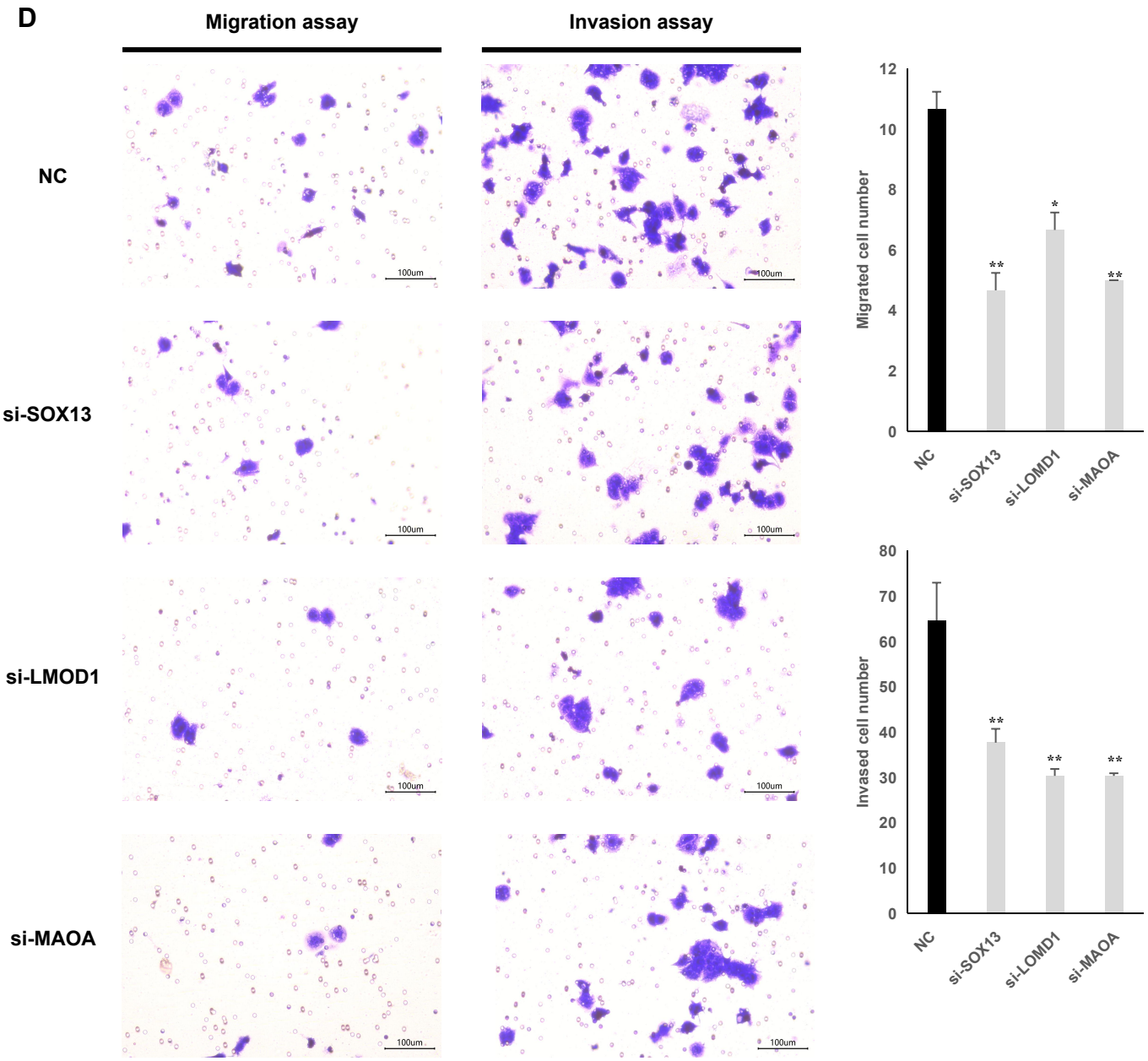


Figure S7 Effect of *SCNN1A*, *TROAP*, *SOX13*, *MAOA* and *LMOD1* knockdown on migration and invasion of HEC-1-A. (A) Effect of *SCNN1A*, *TROAP*, *SOX13*, *MAOA* and *LMOD1* knockdown on (B-D) invasion and migration potential of HEC-1-A, detected by transwell assay. Data are expressed as the mean ± standard deviation of three independent experiments. * P < 0.05; ** P < 0.01; *** P < 0.001. NC, negative control.

Table.S1 Genome overview of TFCRs and ATAC-seq peaks

Classification	Interval number	Total length (bp)	Proportion covered with genome	Mean	Standard deviation
TFCR	78820	76320782	2.51%	968.29	299.86
ATAC-seq peaks	104723	52570946	1.73%	502	0

The table shows the number of TFCRs and ATAC-seq peaks, the total number of bases, whole-genome coverage ratio, and mean and standard deviation of the number of bases between different TFCRs or ATAC-seq peaks.

Table.S2 Mutation rate of TFCR (TC0-CAS9)

Classification	Total length (bp)	Mutation numbers	Mutation rate (ρ)
TFCRs (Belong to both TC0 and CAS9)	296150	342	1.15 ‰
TFCRs	76320782	40757	0.53 ‰
ATAC-seq peaks	52570946	29993	0.57 ‰
Genomes	3036303846	886377	0.29 ‰

The table shows the differences in mutation rates (ρ) among different taxonomic regions of the genome (such as TFCRs belonging to both TC0 and CAS9, all TFCRs, ATAC-seq peaks and the whole-genome).

Table.S3 Small inference RNA sequence

siRNA	Sequence (5'-3')
si-NC1	CGUACGCGGAAUACUUCGA TT UCGAAGUAUUCCGCGUACG TT
si-NC2	UUCUCCGAACGUGUCACGUTT ACGUGACACGUUCGGAGAATT
si-HOXA9	UCCUCCAGUUGAUAGAGAATT UUCUCUAUCAACUGGAGGATT
si-TROAP	GAAACCACCGCUCAAUAUUTT AAUAUUGAGCGGUGGUUUCTT
si-SCNN1A	GGAGUGGCCAAAGUCAACA TT UGUUGACUUUGGCCACUCC TT
si-SOX13	GCCAUAGAGAAGCUGUUGUTT ACAACAGCUUCUCUAUGGCTT
si-MAOA	GAAAGCUGAUCGACUUGCUTT AGCAAGUCGAUCAGCUUUCTT
si-LMOD1	GUCUAGAGUAGCCAAAUAUTT AUAUUUGGCUACUCUAGACTT

Table.S4 Primes used in qPCR expirement

Gene	Primer
GAPDH	F: TCGGAGTCAACGGATTTGGT
	R: TTCCCGTTCTCAGCCTTGAC
HOXA9	F: TACGTGGACTCGTTCCTGCT
	R: CGTCGCCTTGGACTGGAAG
TROAP	F:CCTCCGGGGTGTATCTCCTAC
	R: ACGGCGCACGATGTAACAG
SCNN1A	F:AGGGGAACAAGCGTGAGGA
	R: GGTGGAACTCGATCAGGGC
SOX13	F: CACAGATGACAGGCACTCGG
	R: TCTCTGGGGCACGAACAAGT
MAOA	F: TTCAGGACTATCTGCTGCCAA
	R: GGTCCCACATAAGCTCCACC
LMOD1	F: GTAAAAGGGGAGCGTAGGAAC
	R: CTCGGGTGTTTTGGTCTTGCT