

1 Supplemental files for

2 **MAE-seq refines regulatory elements across the genome**

3 This files includes:

4 MAE-seq screening vector.

5 Figure S1 Supporting materials related to MAE-seq development.

6 Figure S2 Epigenetic characteristics and chromatin accessibility of MAE-seq loci.

7 Figure S3 Activity validation of MAE-seq enhancers with or without left or right
8 flanking sequences in HEK 293T cell.

9 Figure S4 Participate in the ratio of chromatin loop novel enhancers and the GO
10 analysis of known enhancers.

11 Figure S5 Ce1 to Ce3 knockout results supplement to figure 5.

12 Figure S6 Epigenetic characterization and taxonomic analysis of HEK 293T.

13 Figure S7 The number of TFs bound to long sequences and the expression of target
14 cofactors after interference with cells.

15 Table S1 Location of randomly verified MAE-seq enhancers in different cells.

16 Table S2 Public data sources used in this study.

17 Table S3 The genomic locations of MAE-seq enhancers and long-segment enhancers
18 in the “LER” experiment.

19 Table S4 The enhancer target gene location for verification, sgRNA for CRISPR
20 activation/knockout experiments, and QPCR detection primers.

21 Table S5 The MAE-seq enhancer location information (mm9 reference genome,
22 known and novel classifications have been performed on it) identified in mESCs.

23 Table S6 The MAE-seq enhancer location information (hg19 reference genome,
24 known and novel classifications have been performed on it) identified in HEK 293T.

25 Table S7 The MAE-seq enhancer location information (mm9 reference genome,
26 known and novel classifications have been performed on it) identified in C2C12.

27

28 MAE-seq screening vector sequence:

29 CTGCATAATGAAAGACCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGCCATTTTGAAGGCATGGAAAAATA
30 CATAACTGAGAATAGAAAAGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAAGCGGATATC
31 TGTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTC
32 GTGGTAAGCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTC
33 TAGAGAACCATCAGATGTTTCCAGGGTGCCCCAAGGACCTGAAATGACCCTGTGCCTTATTGAACTAACCAATCAG
34 TTCGTTCTCGCTTCTGTTTCGCGCGCTTCTGCTCCCCGAGCTCAATAAAAGAGCCCAACCCCTCACTCGGGGCGC
35 CAGTCCCTCCGATTGACTGAGTCGCCCCGGGTACCCGTGTATCCAATAAACCCCTCTTGCAGTTGCATCCGACTTGTGGT
36 CTCGCTGTTTCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCGTGAGCGGGGGTCTTTCATTGGGGGCTCGTCC
37 GGGATCGGGAGACCCCTGCCCAGGGACCAACGACCCACCAACCGGGAGGTAAAGCTGGCCAGCAACTTATCTGTGTCT
38 GTCCGATTGTCTAGTGTCTATGACTGATTTTATGCGCCTGCGTCGGTACTAGTTAGCTAACTAGCTCTGTATCTGGCGG
39 ACCCGTGGTGGAAGTACGAGTTCGGAACACCCGGCCGCAACCCCTGGGAGACGTCCCAGGGATCTCGGGGGCCGT
40 TTTTGTGGCCCGACCTGAGTCCAAAAATCCCGATCGTTTGGACTCTTTGGTGCACCCCCCTTAGAGGAGGGATATG
41 TGGTCTGGTAGGAGACGAGAACCTAAACAGTTCCCGCCTCCGTCTGAATTTTGTCTTTCGGTTTGGGACCGAAGC
42 CGCGCCGCGCGTCTTGTCTGCTGCAGCATCGTTCTGTGTTGTCTCTGTCTGACTGTGTTTCTGTATTGTCTGAAAAT
43 ATGGGCCCCGGCCAGACTGTTACCACTCCCTTAAGTTTGACCTTAGGTCAGTGAAGATGTCGAGCGGATCGCTCA
44 CAACCAGTCGGTAGATGTCAAGAAGAGACGTTGGGTTACCTTCTGCTCTGCAGAATGGCCAACCTTTAACGTCGGA
45 TGGCCGCGAGACGGCACCTTTAACCGAGACCTCATCACCCAGGTTAAGATCAAGGTCTTTTCACTGGCCCGCATG
46 GACACCCAGACAGGTCCCTTACATCGTGACCTGGGAAGCCTTGGCTTTTGACCCCCCTCCTGGGTCAAGCCCTT
47 GTACACCTAAGCCTCCGCCTCCTCTTCTCCATCCGCCCCGTCTCTCCCCCTGAACCTCCTCGTTTCGACCCCGCT
48 CGATCTCCCTTTATCCAGCCCTCACTCCTTCTTAGCGCCCCCATATGGCCATATGAGATCTTATATGGGGCACCCC
49 CGCCCCCTGTAAACTTCCCTGACCTGACATGACAAGAGTTACTAACAGCCCCCTCTCTCAAGCTCACTTACAGGCT
50 CTCTACTTAGTCCAGCACGAAGTCTGGAGACCTCTGGCGGCAGCCTACCAAGAACAACTGGACCGACCGGTGGTAC
51 CTCACCTTACCGAGTCGGCGACACAGTGTGGGTCCGCCGACACCAGACTAAGAACCTAGAACCTCGCTGGAAAG
52 GACCTTACACAGTCCTGCTGACCACCCCAACGCCCTCAAAGTAGACGGCATCGCAGCTTGGATACACGCCGCCCA
53 CGTGAAGGTGCCGACCCCGGGGTGGACCATCTCTAGACTGCCGGATCCAGTGTGGTGGTACGGGAATTCTGC
54 AGTCGACGGTACCGCGGGCCCGGGATCCACCGGTGCGCCACCAATAAAATATCTTTATTTTATTACATCTGTGTGTG
55 GTTTTTTGTGTGAATCGATAGTACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAACTAGCAAAATAG
56 GCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTGGCCTAACTGGCCGGTACCTGAGCTCGCTAGCCTCG
57 AGGATATCAAGATCTGGCCTCGGCGGCCAAGCTTAGACACTAGAGGGTATATAATGGAAGCTCGACTTCCAGCTTGG
58 CAATCCGGTACTGTTGGTAAAGCCACCATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCA
59 TGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACCGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCC
60 CCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCTGGGACATCCTGTC
61 CCCTCAGTTTATGTACGGCTCCAAGGCCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCTTCC
62 CCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCC
63 TGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCCTCCGACGGCCCCGTAATGCAGAA
64 GAAGACCATGGGCTGGGAGGCCCTCCTCCGAGCGGATGTACCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCA
65 GAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGC
66 AGCTGCCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACCATCGTGGAACA
67 GTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAATAATTCTAGAGTCGGGGC
68 GGCCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAGAA
69 AATGCTTTATTGTGAAATTGTGATGCTATTGCTTTATTGTAAACATTATAAGCTGCAATAAAACAAGTTAACACAA
70 CAATTGCATTATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAACCTCTACAAAT
71 GTGGTAAGCGGCCGCCAGCACAGTGGTCGACGATAAAATAAAAGATTTTATTAGTCTCCAGAAAAAGGGGGGAAT

72 GAAAGACCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGCCATTTTGAAGGCATGGAAAAATACATAACTG
73 AGAATAGAGAAGTTCAGATCAAGGTCAGGAACAGATGGAACAGCTGAATATGGGCCAAAACAGGATATCTGTGGTAA
74 GCAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGAACAGCTGAATATGGGCCAAAACAGGATATCTGTGGTAA
75 CAGTTCCTGCCCCGGCTCAGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGAGAAC
76 CATCAGATGTTTCCAGGGTGCCCCAAGGACCTGAAATGACCCTGTGCCCTATTTGAACTAACCAATCAGTTCGCTTC
77 TCGCTTCTGTTGCGCGCTTCTGCTCCCCGAGCTCAATAAAAGAGCCCACAACCCCTCACTCGGGGCGCCAGTCTC
78 CGATTGACTGAGTCGCCCGGTACCCGTGTATCCAATAAACCCCTTTGCAGTTGCATCCGACTTGTGGTCTCGCTGTT
79 CCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCGTGAGCGGGGGTCTTTCATTCTGCATTAATGAATCGGCCAAC
80 GCGCGGGGAGAGGCGGTTTGCCTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCCG
81 CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGA
82 ACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCC
83 GCCCCCTGACGAGCATCAGAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAACCCGACAGGACTATAAAGATACC
84 AGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTT
85 TCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAG
86 CTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACATCGTCTTGAGTCCAACCC
87 GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGAGGCGGTGCT
88 ACAGAGTTCCTGAAGTGGTGGCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCC
89 AGTTACCTTCGGAAGAGTGGTAGCTCTTGATCCGGCAAAACAAACCACCGTGGTAGCGGTGGTTTTTTTGT
90 GCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA
91 GTGGAACGAAAACTCACGTAAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTGC
92 CGGCCGCAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTAT
93 CTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTT
94 ACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTATCAGCAATAAACCCAGC
95 CAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAA
96 GCTAGAGTAAGTAGTTTCGCCAGTTAATAGTTTTCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTC
97 GTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAA
98 AAGCGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCA
99 GCACTGCATAATTCTCTTACTGTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC
100 TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAA
101 CTTTAAAGTGCTCATCATTGAAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGT
102 TCGATGTAACCCACTCGTGACCCAACTGATCTTCAGCATCTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAAC
103 AGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGAAATGTTGAATACTCATACTCTTCCTTTTCAA
104 TATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTGAATGTATTTAGAAAAATAAACAAATAG
105 GGGTCCGCGCACATTC

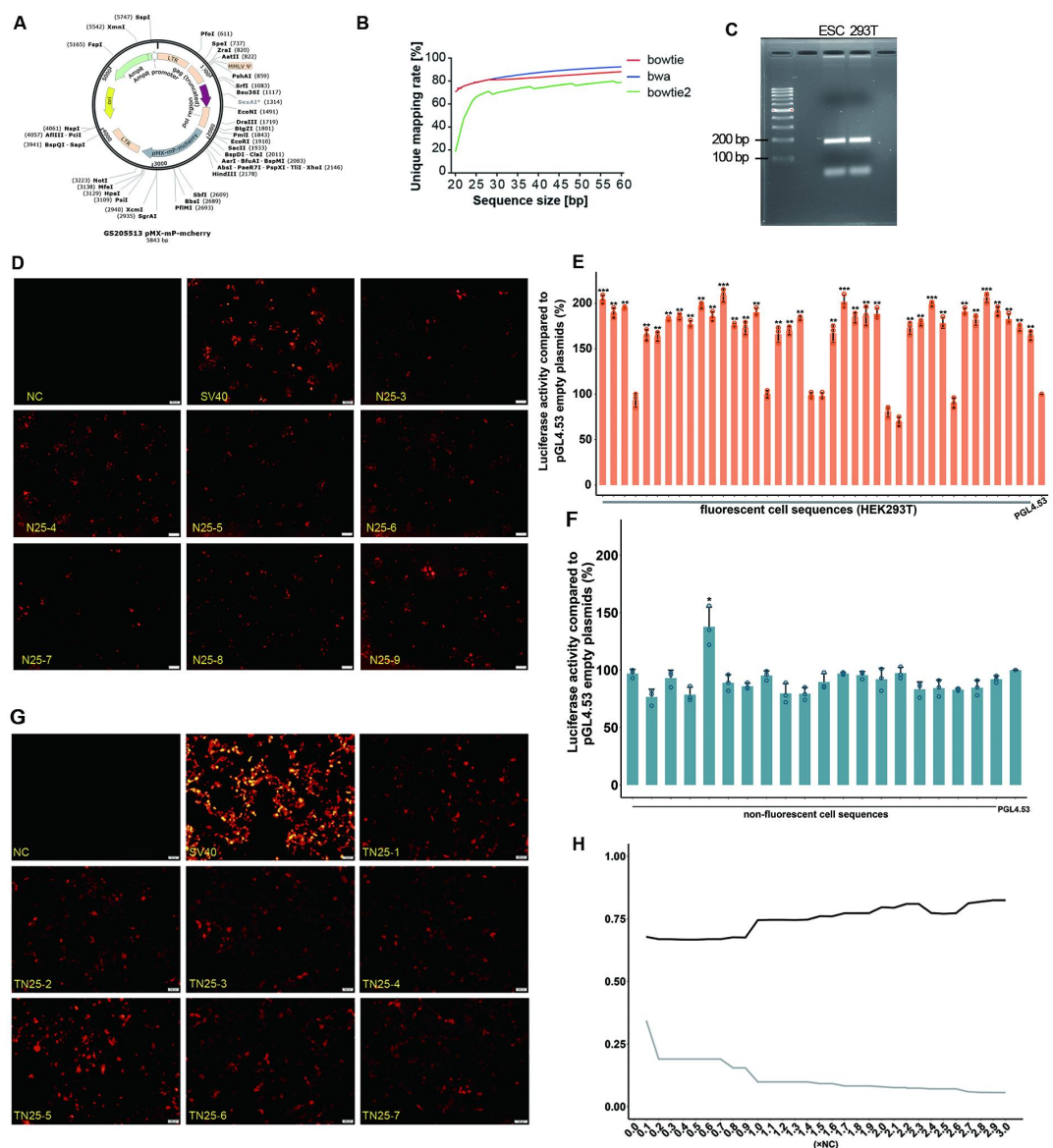


Fig. S1. Supporting materials related to MAE-seq development.

(A) Plasmid map of pMX-mp-mcherry reporter vector.

(B) Test for sequence length and alignment software. DNA sequences of 20-60 base pairs length are randomly extracted from the hg19 genome. Statistics of unique mapping rates are conducted with bowtie, bowtie2 and bwa software, respectively.

(C) Amplification of sequencing library of positive cells. ESC, represents positive mESCs; 293T, represents positive HEK 293T cells.

(D) Fluorescence expression in mESCs Validation of N25-3 to N25-9 supplemented to figure 1C.

(E) 40 MAE-seq loci were randomly selected in HEK293T for activity validation using the luciferase system. empty pGL4.53 plasmid was used as the control for baseline luciferase activity. The y axis represents the percentage of luciferase activity compared to pGL4.53 empty plasmids in the respective cells (n = 3 biological independent samples; bars show mean value $\pm$ s.e.m.; ns: non-significant, **:p<0.01, calculated using t-test).

(F) 20 sequences enriched in non-fluorescent cells were validated for enhancer activity by the luciferase system. empty pGL4.53 plasmid was used as the control for baseline luciferase activity. The y axis represents the percentage of luciferase activity compared to pGL4.53 empty plasmids in the respective cells (n = 3 biological independent samples; bars show mean value $\pm$ s.e.m.; ns: non-significant, *:p<0.05, calculated using t-test).

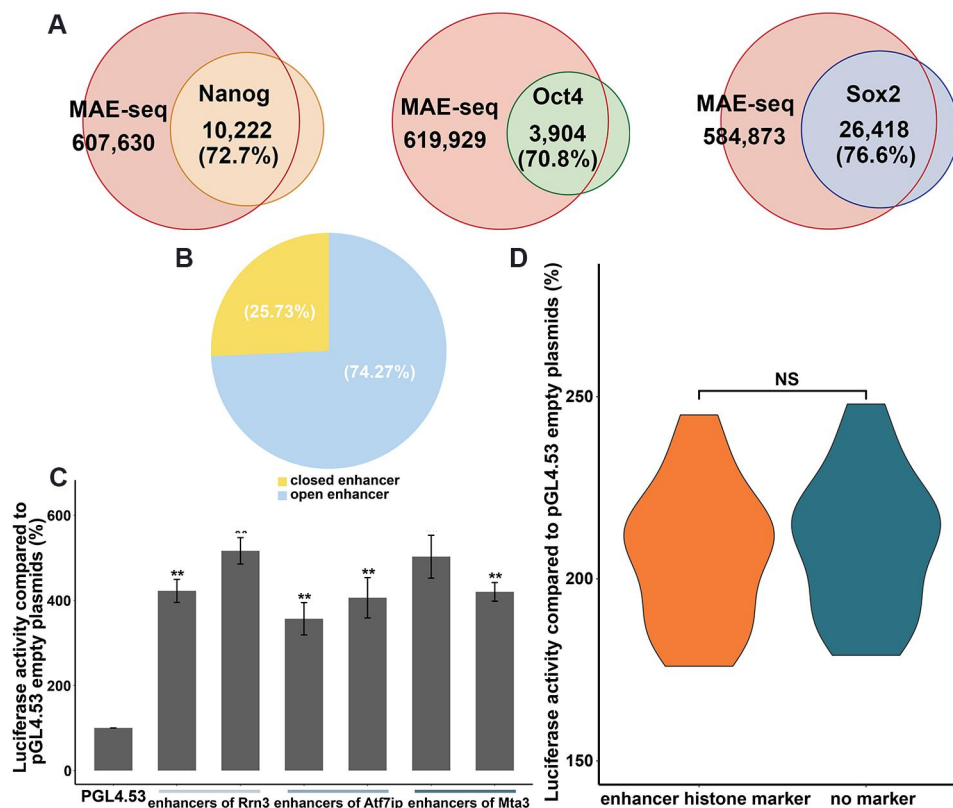
(G) MAE-seq sites were randomly selected in HEK293T cell for fluorescence expression verification. These sites were constructed upstream of the mini promoter of mCherry and transfected into HEK293T cells, and the fluorescence expression was observed 48 hours later. NC is negative control.

(H) The black curve is the product of normalized overlap signal/non-overlap signal(y) and normalized residual rate of overlap signal(y). And the formula is as follow:

$$F(x, y) = \frac{x - 1.75}{1.9 - 1.75} \times 2^y F(x, y) = \frac{x - 1.75}{1.9 - 1.75} \times 2^y$$

The grey curve is residual rate of overlap signal.

137
138



139

140 **Fig. S2. Epigenetic characteristics and chromatin accessibility of MAE-seq loci.**

141 (A) The percentage of different ES-specific TF covered by MAE-seq sites.

142 (B) The percentage of enhancers involved in closed chromatin and open chromatin.

143 (C) Activity of enhancers in heterochromatin regions interacting with housekeeping

144 genes. empty pGL4.53 plasmid was used as the control for baseline luciferase activity.

145 The y axis represents the percentage of luciferase activity compared to pGL4.53

146 empty plasmids in the respective cells (n = 3 biological independent samples; bars

147 show mean value $\pm$ s.e.m.; ns: non-significant, **:p<0.01, calculated using t-test).

148 (D) The relative Luciferase activity distribution of 25bp MAE-seq enhancers with

149 enhancer histone markers and 25bp MAE-seq enhancers with no enhancer histone

markers. T-test was used for difference analysis.

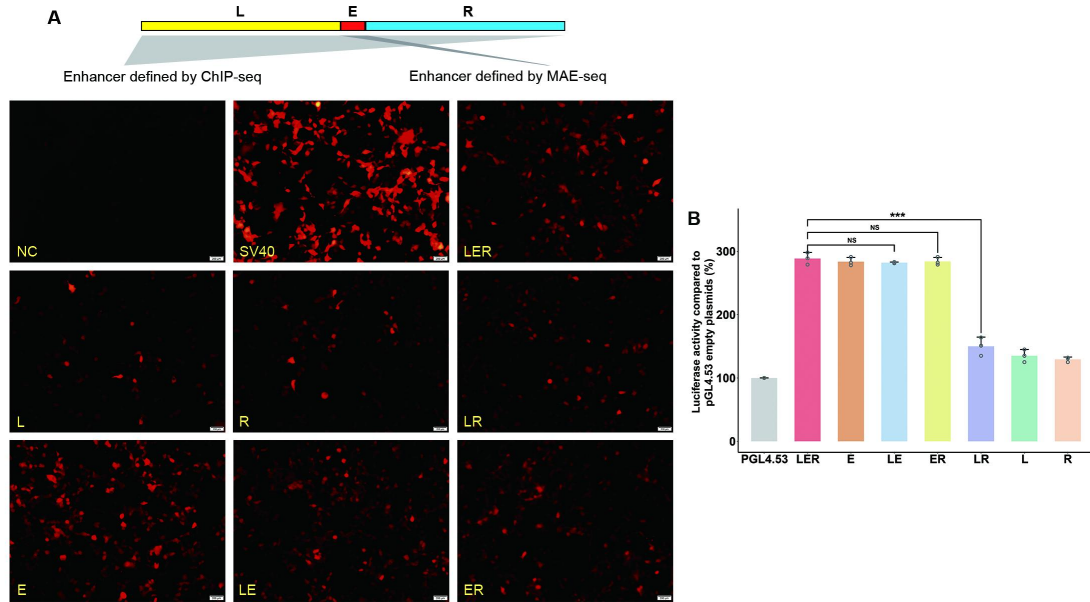


Fig.S3. Activity validation of MAE-seq enhancers with or without left or right flanking sequences in HEK 293T cell.

(A) Enhancer activity of different fragments in "LER" experiments after transfection of mESCs. Different fragments were constructed into the reporter vector with mCherry gene and transfected into cells, and the fluorescence expression was observed after 48 hours. NC, empty reporter vector; SV40, SV40 enhancer; L, left flanking sequence; R, right flanking sequence; E, 25bp enhancer.

(B) Differential analysis of enhancer activity of different fragments in "LER" experiment by luciferase. L, left flanking sequence; R, right flanking sequence; E, 25bp enhancer. empty pGL4.53 plasmid was used as the control for baseline luciferase activity. The y axis represents the percentage of luciferase activity compared to pGL4.53 empty plasmids in the respective cells (n = 3 biological independent samples; bars show mean value $\pm$ s.e.m.; ns: non-significant, ***:

167 $p < 0.001$, calculated using t-test).

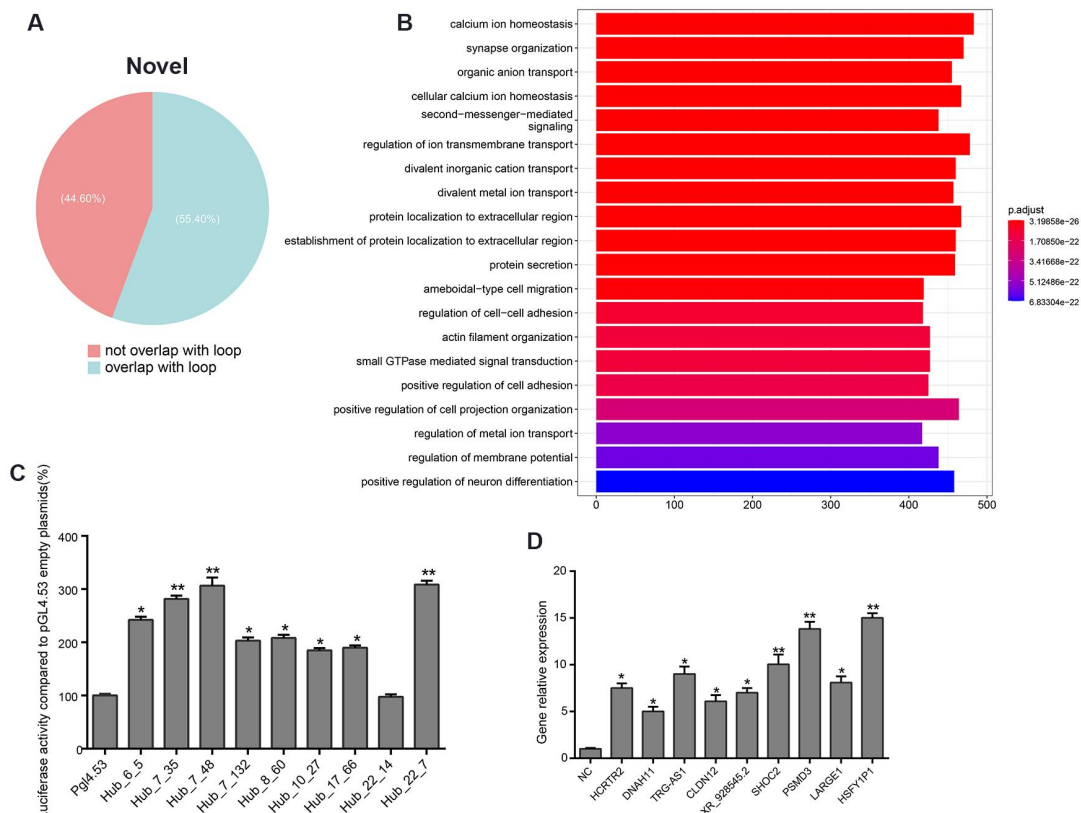


Fig.S4. Participate in the ratio of chromatin loop novel enhancers and the GO analysis of known enhancers.

(A) The percentage of novel enhancer overlapped with chromatin loop anchor.

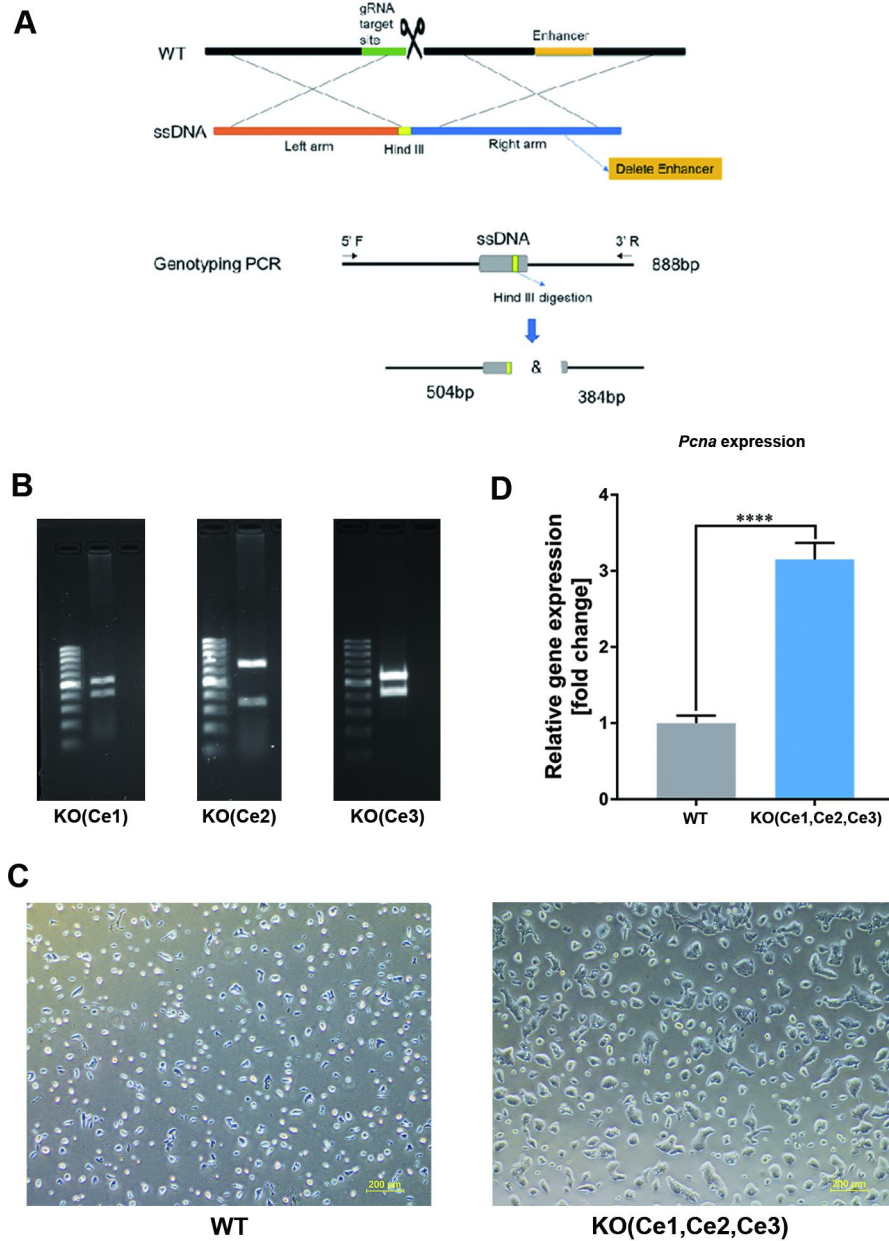
(B) GO enrichment analysis of nearest genes related to known enhancers.

(C) Double luciferase verification results of 9 sites in Hub region. The y axis represents the percentage of luciferase activity compared to pGL4.53 empty plasmids in the respective cells (n = 3 biological independent samples; bars show mean value $\pm$ s.e.m.; ns: non-significant, *: $p < 0.05$, **: $p < 0.01$, calculated using t-test).

(D) After the dCas9-VP64 system activated the 9 regions in (C), QPCR was used to detect the expression level of the interacting target genes. (*: $p < 0.05$; **: $p < 0.01$, calculated using t-test)

183

184



185

186 **Fig.S5. Ce1 to Ce3 knockout results supplement to figure 5.**

187 (A) Experimental procedure of enhancer knockout based on CRSPR/Cas9 and

188 ssDNA. Three MAE-seq enhancers (Ce1, Ce2 and Ce3) targeted to *Cdh1* are verified

189 in Hi-C data. This procedure takes Ce1 as example. WT, wild type locus; ssDNA,

donor single strand DNA without enhancer; Hind III, introduced restriction sites in ssDNA; 5'F, 3'R, forward and reverse primer for genotyping PCR.

(B) Genotyping PCR of Ce1 to Ce3 knockout in mESCs.

(C) Proliferation phenotype of triple knockout cell.

(D) Expression of *Pcna* gene of triple knockout cell. ****: $p < 0.0001$, calculated using t-test.

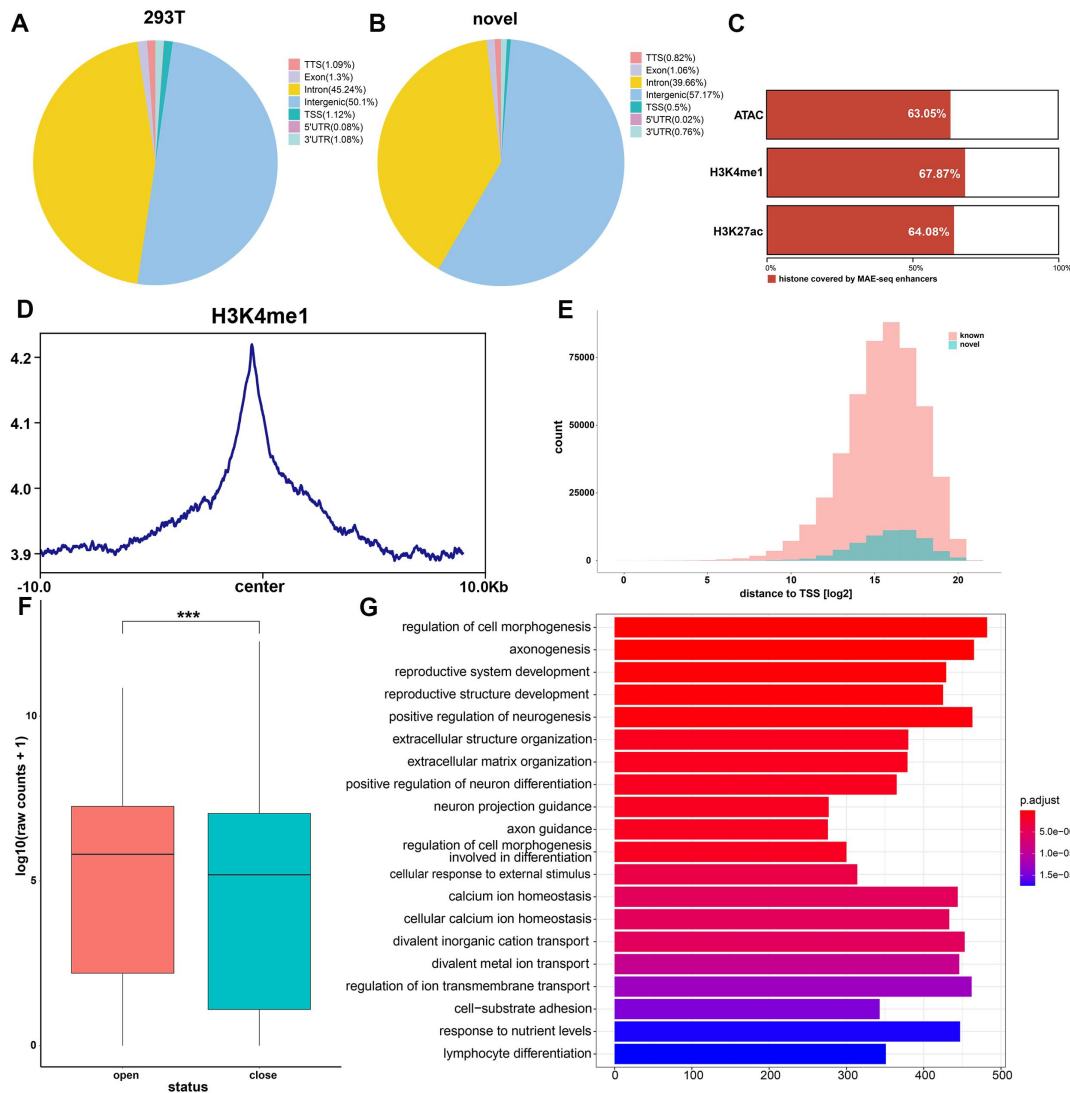


Fig. S6. Epigenetic characterization and taxonomic analysis of HEK 293T.

200 (A) Genomic distribution of HEK 293T MAE-seq enhancers.

201 (B) Genomic distribution of HEK 293T MAE-seq novel enhancers.

202 (C) The percentage of different enhancer epigenetic markers covered by HEK 293T
 203 MAE-seq sites.

204 (D) The enrichment of H3K4me1 marks around the HEK 293T MAE-seq loci.

205 (E) Statistics of distance to TSS of HEK 293T novel enhancers. The two types of
 206 enhancers did not differ significantly from the site of TSS. Red column represents
 207 known mouse enhancers; the blue column represents novel enhancers.

208 (F) Expression of nearest genes related to HEK 293T MAE-seq enhancers which are
 209 in open and closed chromatin regions. Open, represent ATAC; closed, represent none
 210 of ATAC. ***, $p < 0.001$.

211 (G) GO enrichment analysis of nearest genes related to HEK 293T enhancers.

212

213

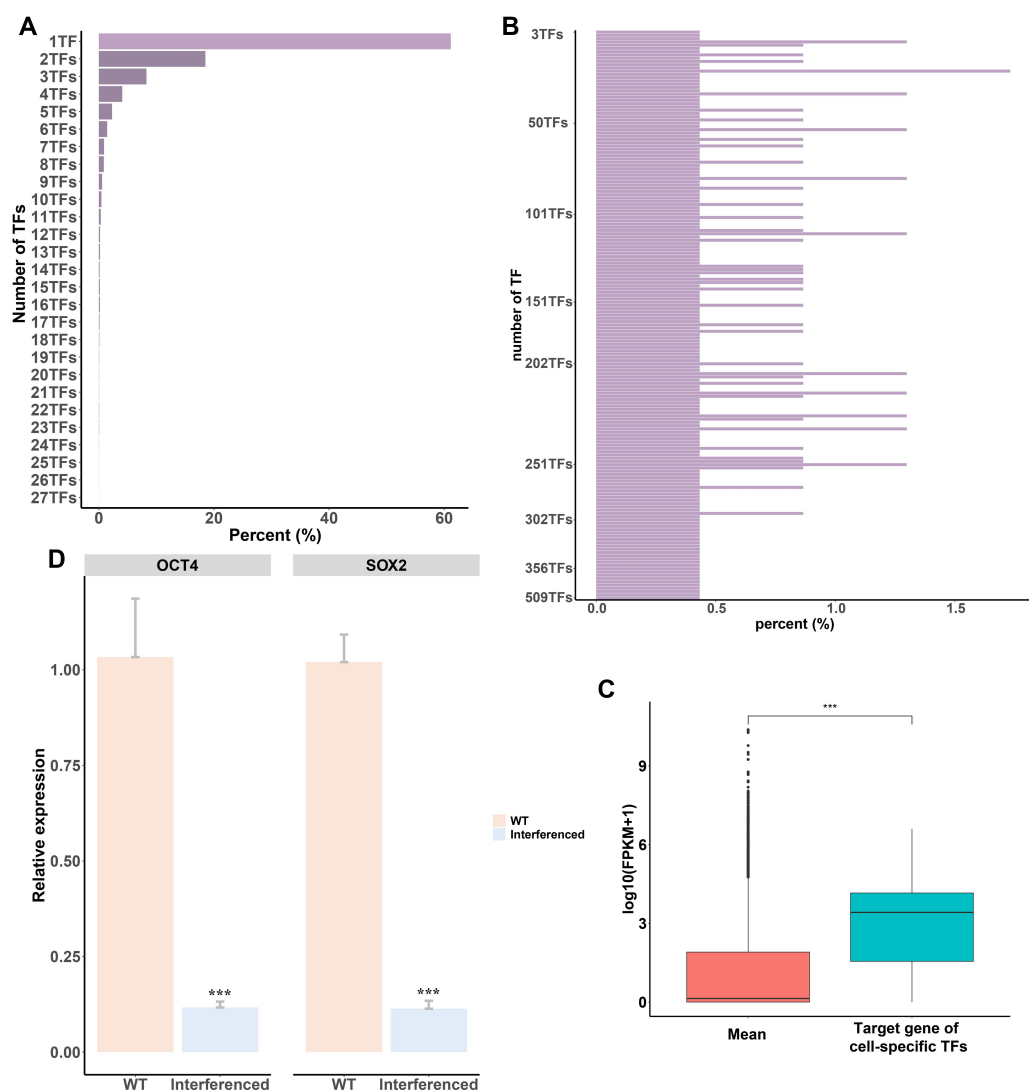


Fig.S7 The number of TFs bound to long sequences and the expression of target cofactors after interference with cells.

(A) The percentage statistics of the number of combined TFs on the MAE-seq enhancer sequence in mESCs.

(B) The percentage statistics of the number of combined TFs on the Super-Enhancers enhancer sequence in mESCs.

(C) The average level of target genes corresponding to single cell-specific TF enhancers was compared with that of genomic genes.

(D) QPCR was used to detect the change of target gene *OCT4* and *CDH1* expression

224 after Cas13d interference. WT, wildtype cell ; interferenced, Cas13d RNAi cell, ***p

225 < 0.001.

226