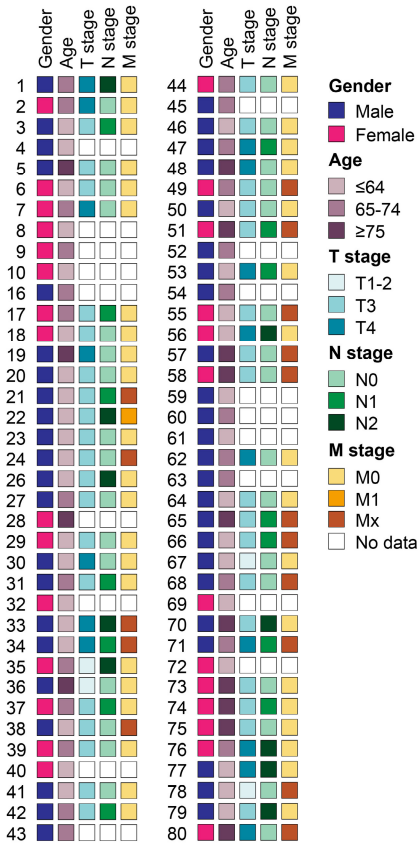
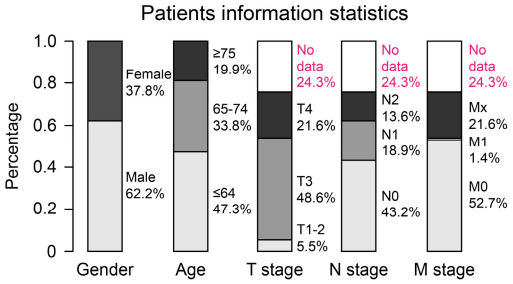
Extended Data Fig. S1

A



B



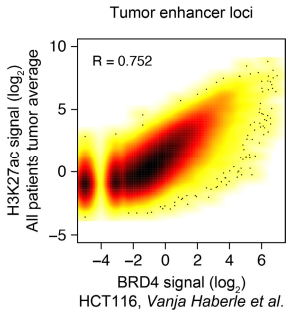
Extended Data Fig. S1 Information of collected CRC tissues. (A) The clinical information of all 74 CRC patients participated in our study. **(B)** Barplot showing the clinical information statistics of CRC patients.

Extended Data Fig. S2

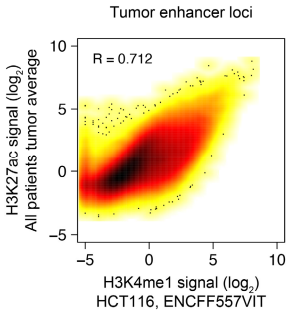
A

Genomic element	Tissue	Total number
Promoter	Native	9896
Promoter	Tumor	10663
Enhancer	Native	27156
Enhancer	Tumor	39207

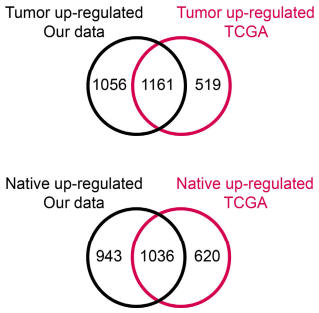
B



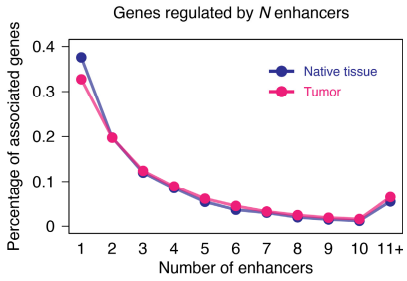
C



D

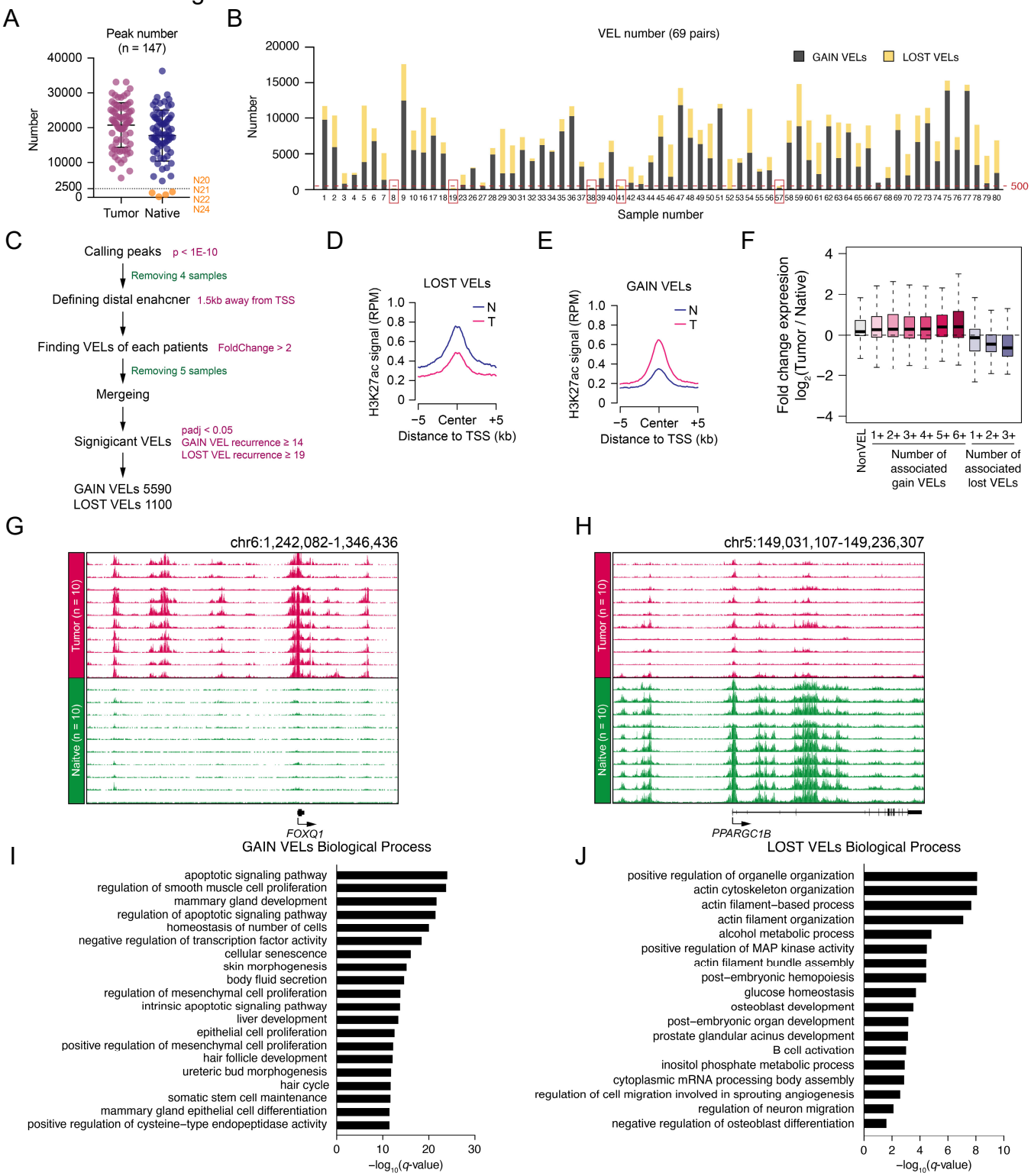


E



Extended Data Fig. S2 The enhancer analysis of CRC samples. (A) The number of significant promoters and enhancers in tumor and native tissues. **(B-C)** Mean H3K27ac level in CRC tumors versus BRD4 (B) and H3K4me1 (C) levels in HCT116. **(D)** Overlap of tumor (top) and native (bottom) higher-expressed genes in our data with those in TCGA COAD data. **(E)** Proportion of enhancer/gene assignments to assigned enhancer number in tumor and native tissues.

Extended Data Fig. S3



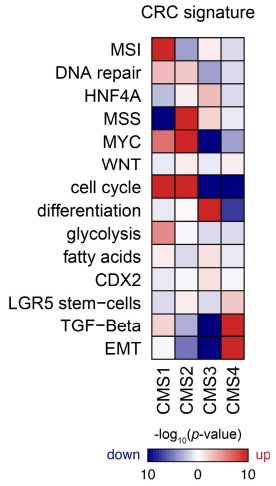
Extended Data Fig. S3 Functional analysis of gain and lost enhancers in CRC. (A) The number of significant H3K27ac peaks in each sample. The samples whose peak number less than 2500 was highlighted in yellow. **(B)** VEL (including gain and lost VELs) number of each pair of samples. The patient whose VEL number less than 500 was highlighted in red box. **(C)** Flow chart of VEL definition. **(D&E)** The average H3K27ac signal (RPM) in the region of lost (D) and gain (E) VELs. **(F)** Fold change of gene expression associated with gain (red) and lost (purple) VELs, and not associated with VELs (nonVEL; grey). **(G)** Representative H3K27ac tracks of gain VEL in *FOXO1* gene loci. **(H)** Representative H3K27ac tracks of lost VEL in *PPARGC1B* gene loci. **(I&J)** The biological processes in which the genes associated with gain (G) and lost (H) VELs were enriched, were analyzed by GREAT (version 3.0.0).

Extended Data Fig. S4

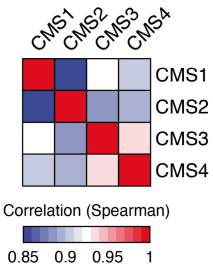
A

Subgroup	Number	Patient identifier
CMS1	10	1, 2, 18, 30, 33, 34, 46, 48, 77, 79
CMS2	19	5, 6, 17, 27, 31, 36, 37, 40, 42, 43, 50, 52, 56, 62, 63, 64, 65, 69, 76
CMS3	13	9, 10, 16, 23, 26, 28, 44, 51, 53, 54, 59, 66, 70
CMS4	15	4, 7, 29, 39, 45, 49, 55, 58, 60, 61, 67, 74, 75, 78, 80

B

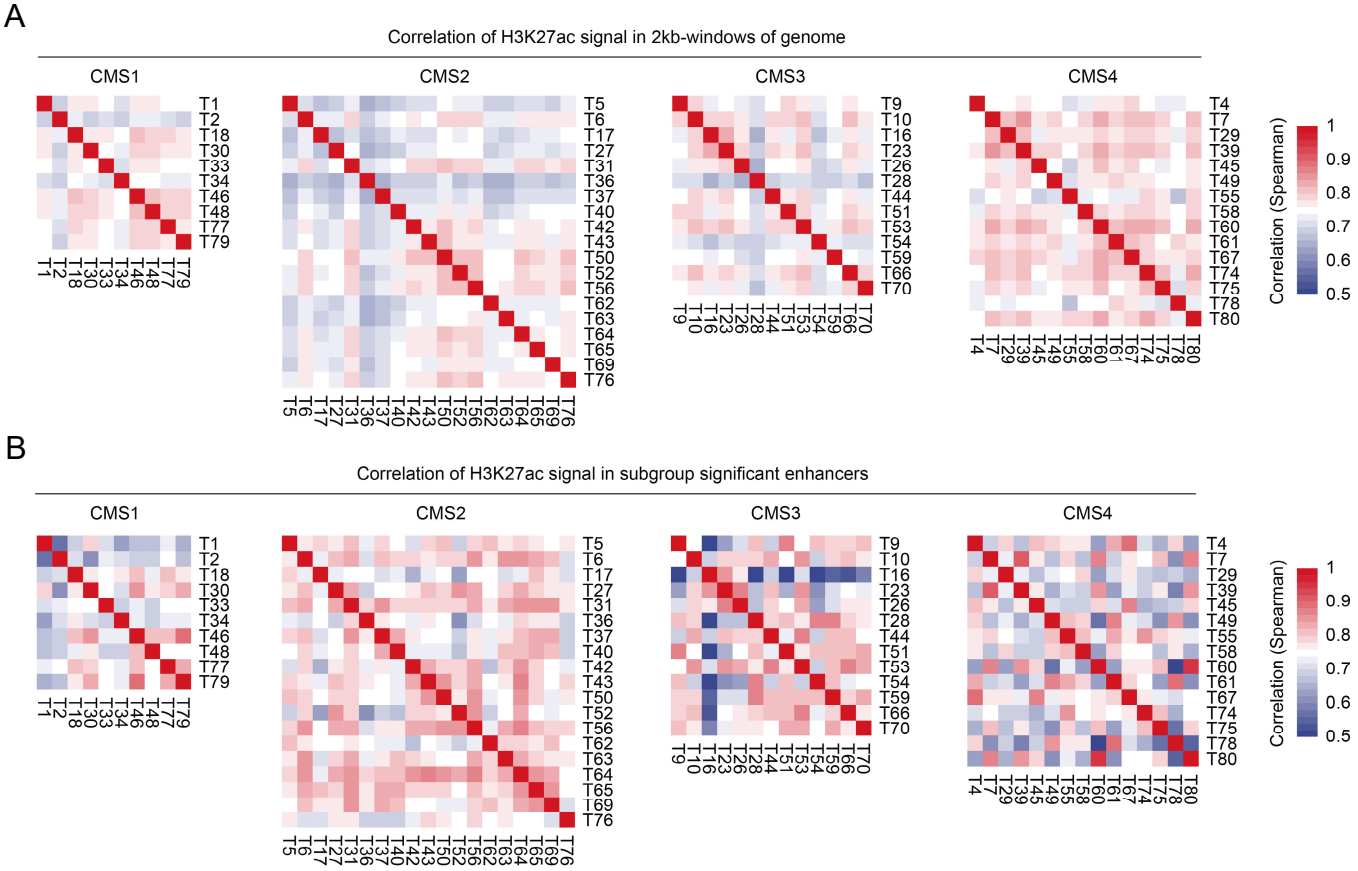


C



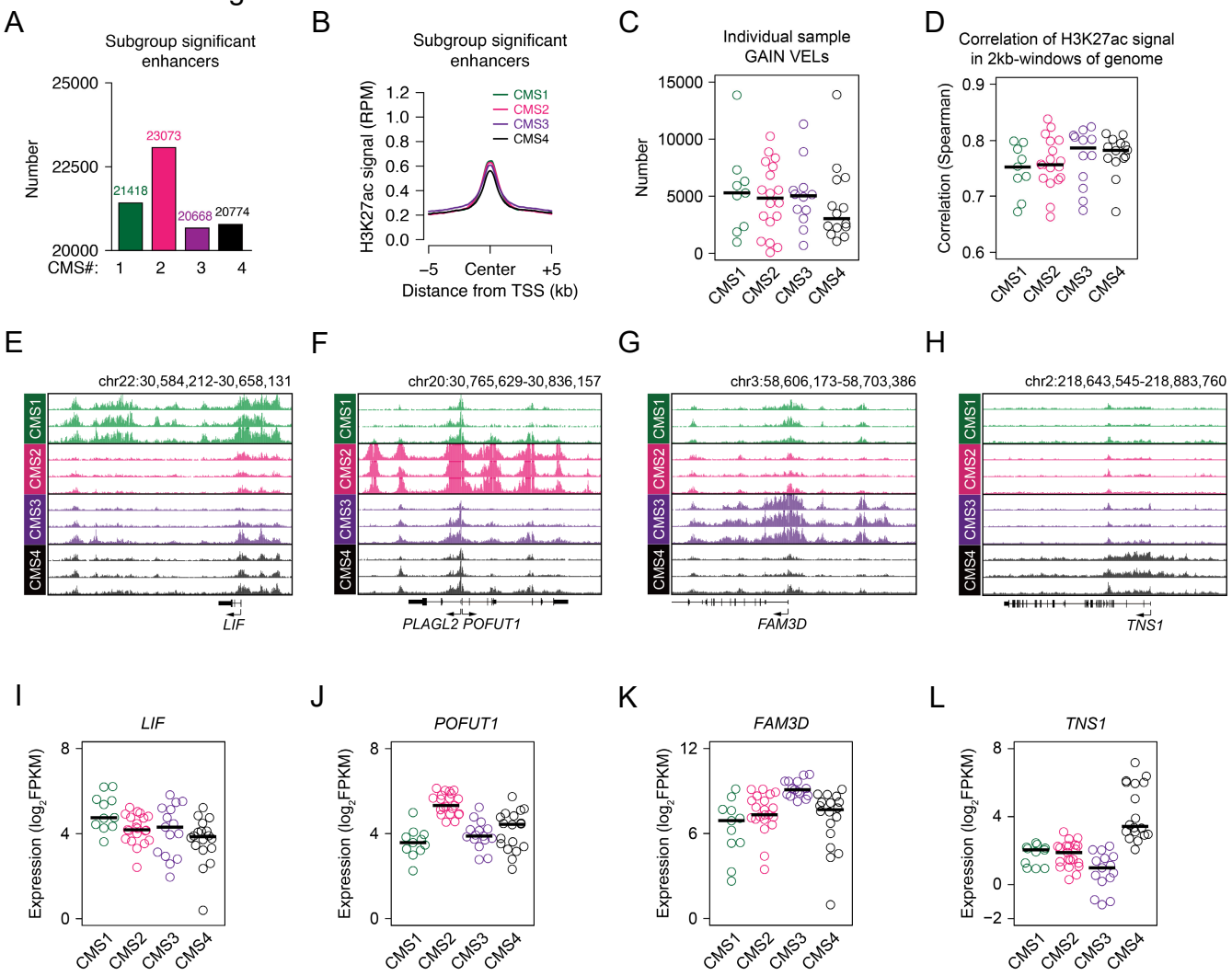
Extended Data Fig. S4 The consensus molecular subtypes (CMS) classification of CRC samples. (A) The patient identifier of members in four CMS subgroups. **(B)** The significance of CRC signature for four CMS subgroups identified by CMScaller. **(C)** Correlation of mean H3K27ac on the regions of gain VELs in four CMS subgroups. Correlations were calculated by Spearman correlation coefficient.

Extended Data Fig. S5



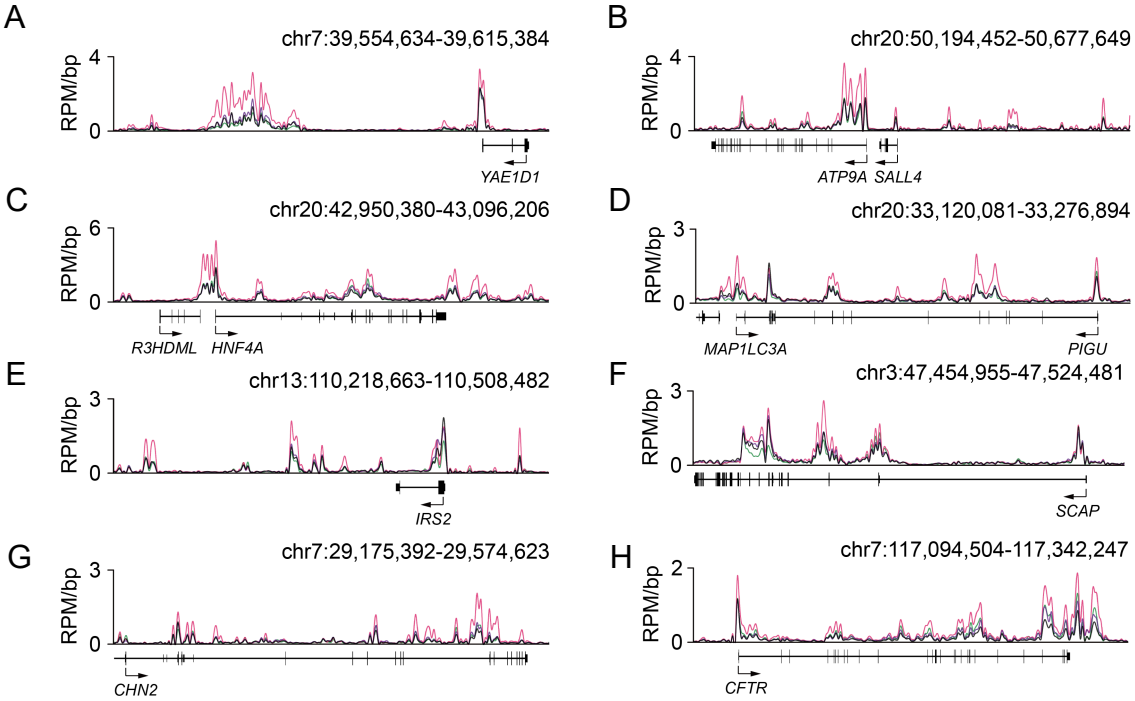
Extended Data Fig. S5 Correlation of H3K27ac in four CMS subgroups. (A) H3K27ac RPM values were generated in each 2-kb window for the entire genome and compared between different samples for correlation analysis (Spearman). Correlation of H3K27ac signal at 2kb-windows of genome in all tumor samples from CMS1-4 subgroup. **(B)** Heatmap showing the correlation of H3K27ac signal at CMS-specific gain VELs in tumor samples of each subgroup. Correlations were calculated by Spearman correlation coefficient.

Extended Data Fig. S6



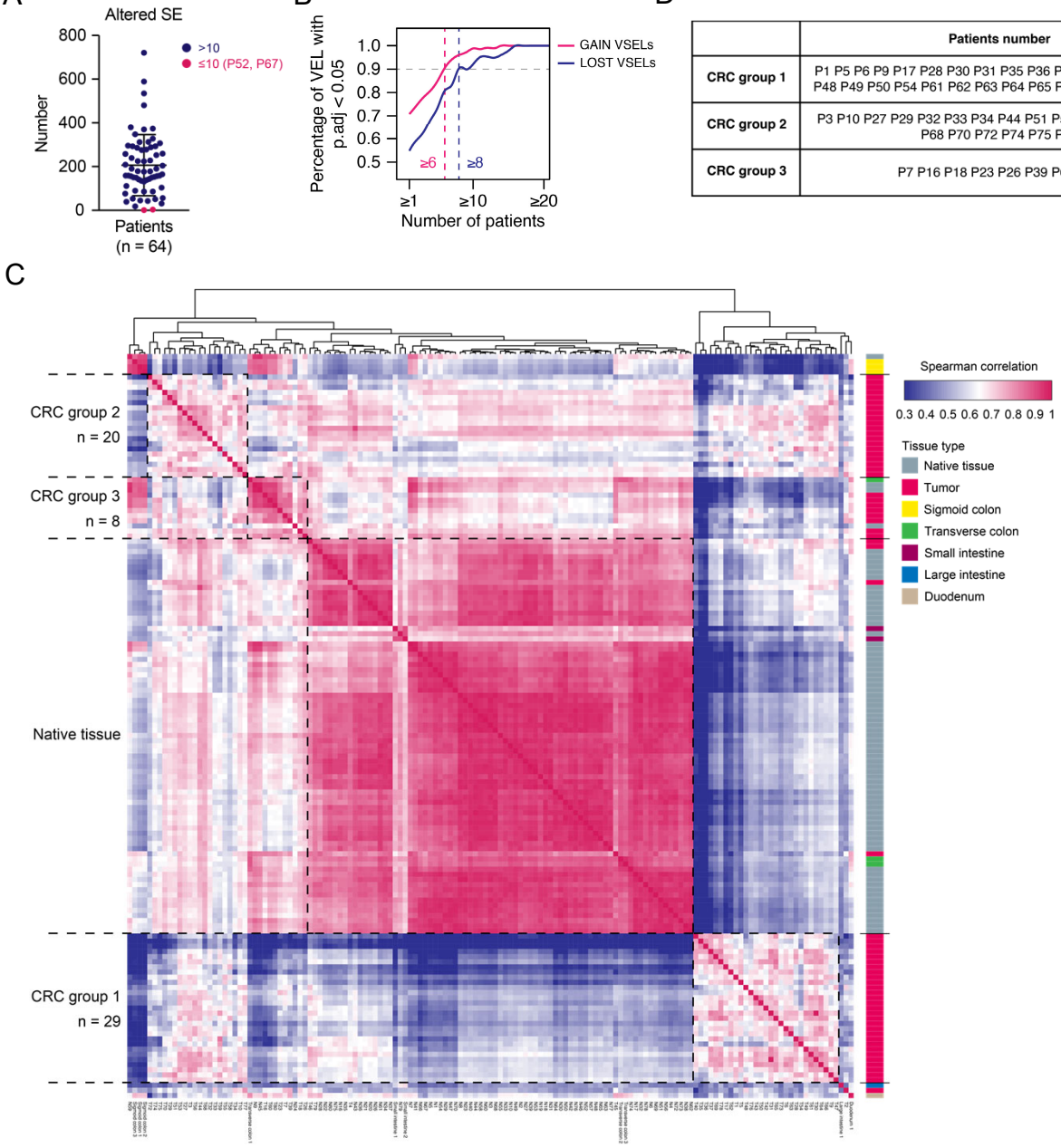
Extended Data Fig. S6 The enhancer analysis of four CMS groups. (A) The number of significant enhancers in four CMS subgroups. **(B)** The average H3K27ac signal (RPM) at the regions of specific significant enhancers in four CMS subgroups. **(C)** The number of gain VELs of individual samples of each subgroup. Black lines indicate the median number. **(D)** Correlation of H3K27ac signal at 2kb-windows of genome between paired tumor and native tissues in four CMS subgroups. Correlations were calculated by Spearman correlation coefficient. **(E-H)** Normalized H3K27ac tracks for representative gain VEL samples of four CMS subgroups. **(I-L)** Expression of the above four genes in S6E-H in tumor samples of four subgroups. Black lines indicate the median.

Extended Data Fig. S7



Extended Data Fig. S7 H3K27ac tracks of representative genes in four CMS groups. (A-H) Meta normalized H3K27ac tracks on *YAE1D1* (A), *ATP9A* (B), *HNF4A* (C), *PIGU* (D), *IRS2* (E), *SCAP* (F), *CHN2* (G) and *CFTR* (H) loci in four CMS subgroups.

Extended Data Fig. S8

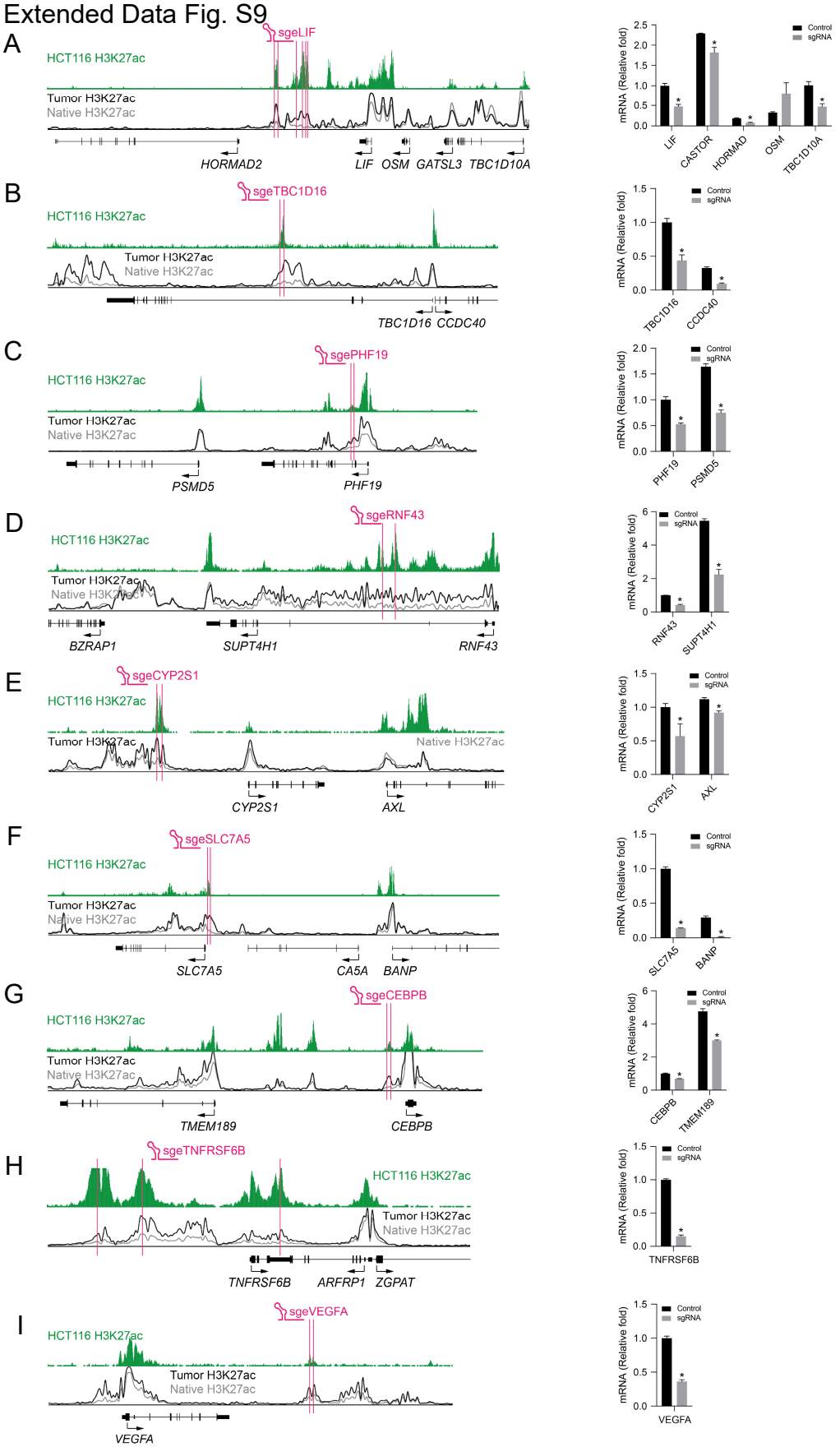


D

	Patients number
CRC group 1	P1 P5 P6 P9 P17 P28 P30 P31 P35 P36 P37 P40 P42 P43 P47 P48 P49 P50 P54 P61 P62 P63 P64 P65 P66 P69 P73 P76 P78
CRC group 2	P3 P10 P27 P29 P32 P33 P34 P44 P51 P53 P55 P56 P58 P59 P68 P70 P72 P74 P75 P77
CRC group 3	P7 P16 P18 P23 P26 P39 P60 P80

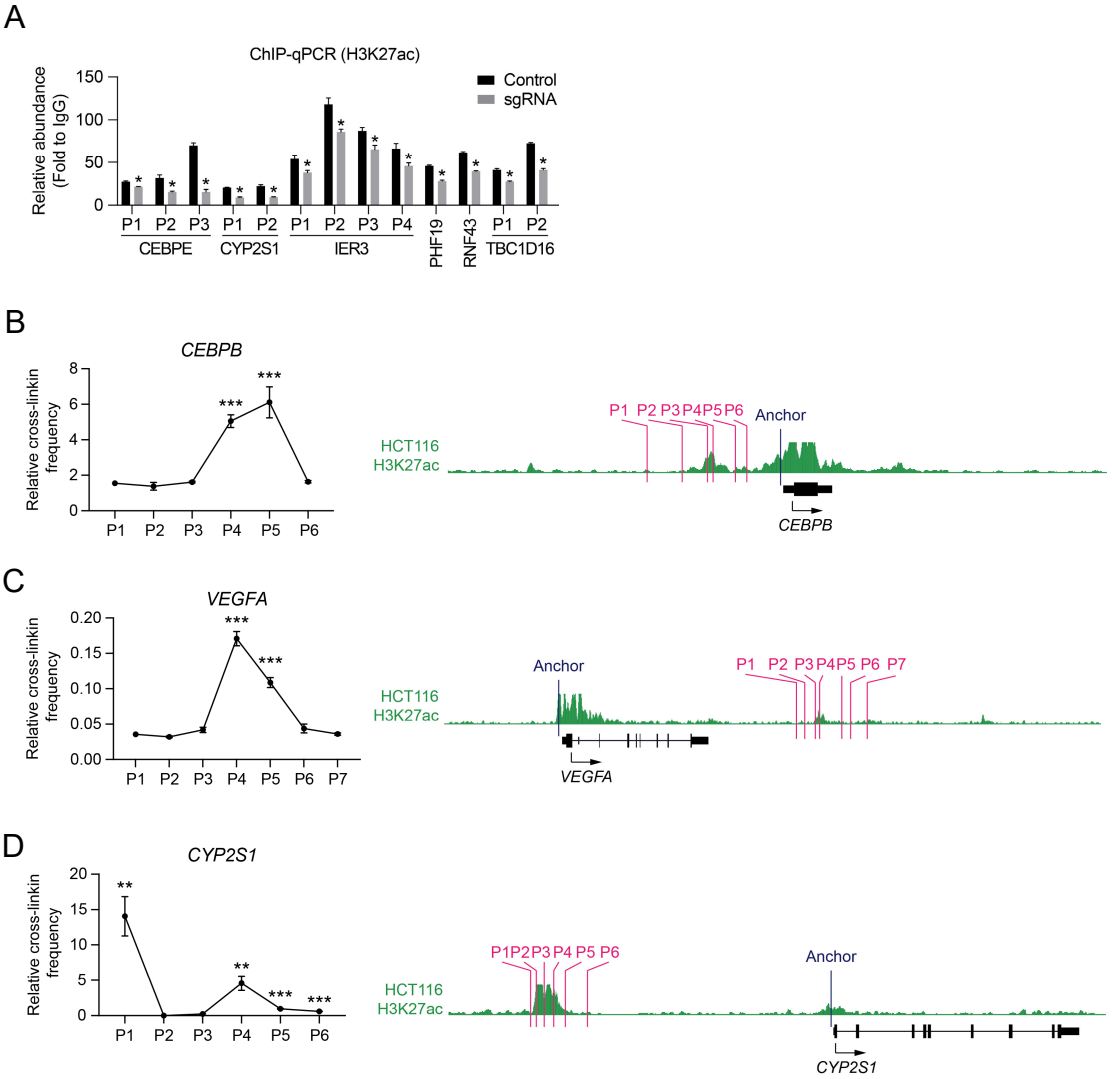
Extended Data Fig. S8 Analysis of super enhancers (SE) in CRC. (A) The number of significant VSELs (FC > 2) in each sample. The pairs whose VSEL number less than 10 were highlighted in red. **(B)** The recurrence requirements for VSELs to meet statistical significance (q-value < 0.05). The two vertical dashed lines highlight the recurrence of gain (red) and lost (blue) VSELs when achieving the cut-off (0.9, black horizontal dashed line) of significant percentage. **(C)** Heatmap to visualize the unsupervised clustering of pairwise correlations of H3K27ac signal (RPM) on VSELs (334 gain VSELs and 121 lost VSELs) for all. H3K27ac ChIP-seq data of five normal intestinal tissues downloaded from ENCODE were used, sigmoid colon (ENCFF611MTD, ENCFF860KII and ENCFF250TIL), transverse colon (ENCFF639EGR, ENCFF538APS and ENCFF485QGB), small intestine (ENCFF195YMX and ENCFF943RFS), large intestine (ENCFF805JFJ) and duodenum (ENCFF630DOD). **(D)** The members of CRC patient groups indicated in Extended Data 4C.

Extended Data Fig. S9



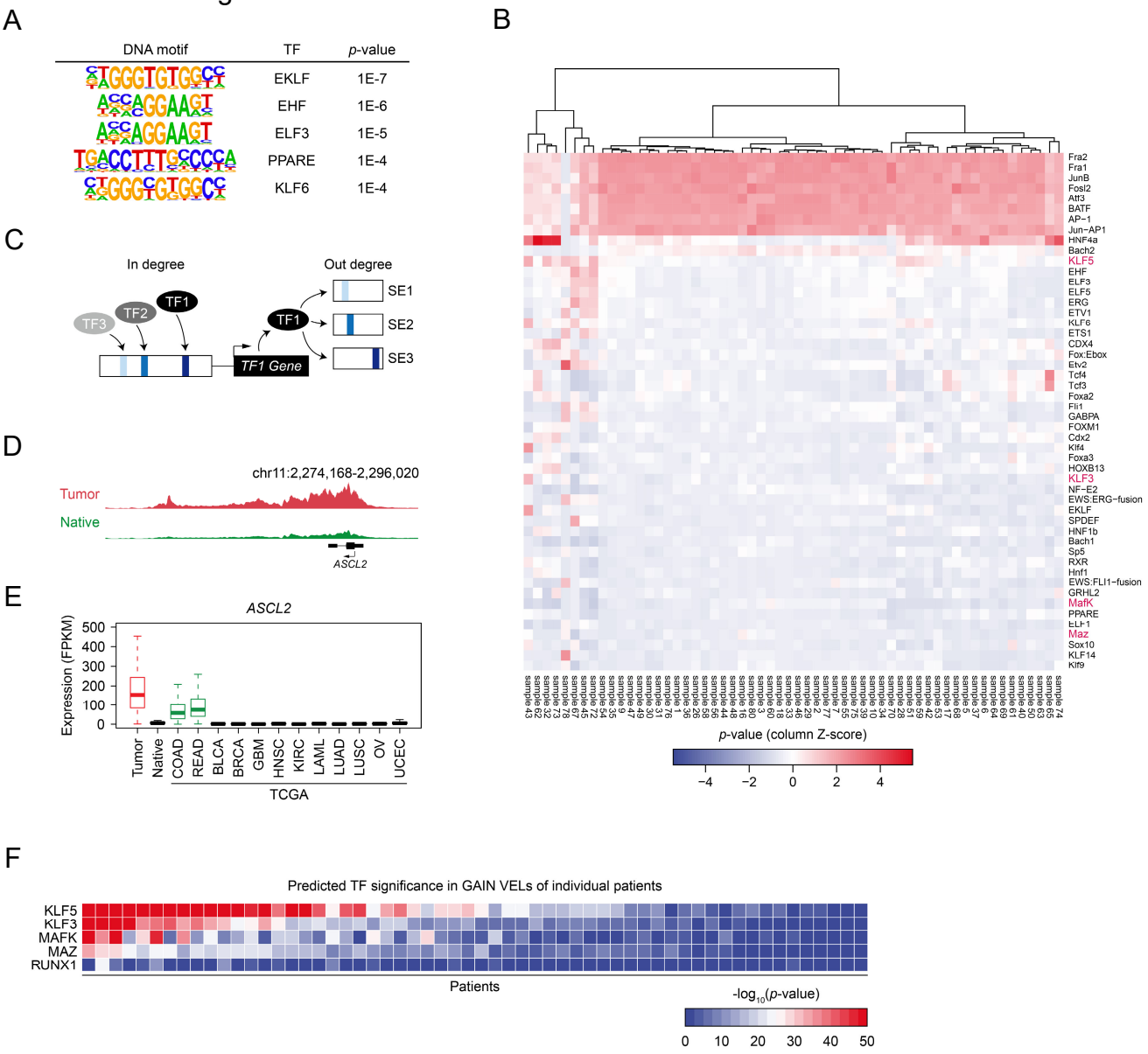
Extended Data Fig. S9 Experimentally verification of the effects on gene expression for CRC specific super enhancers. (A-I) The design of sgRNA targeting SEs were shown on the left, and their effects on gene transcription determined by quantitative PCR were shown on the right. * means p value < 0.05.

Extended Data Fig. S10



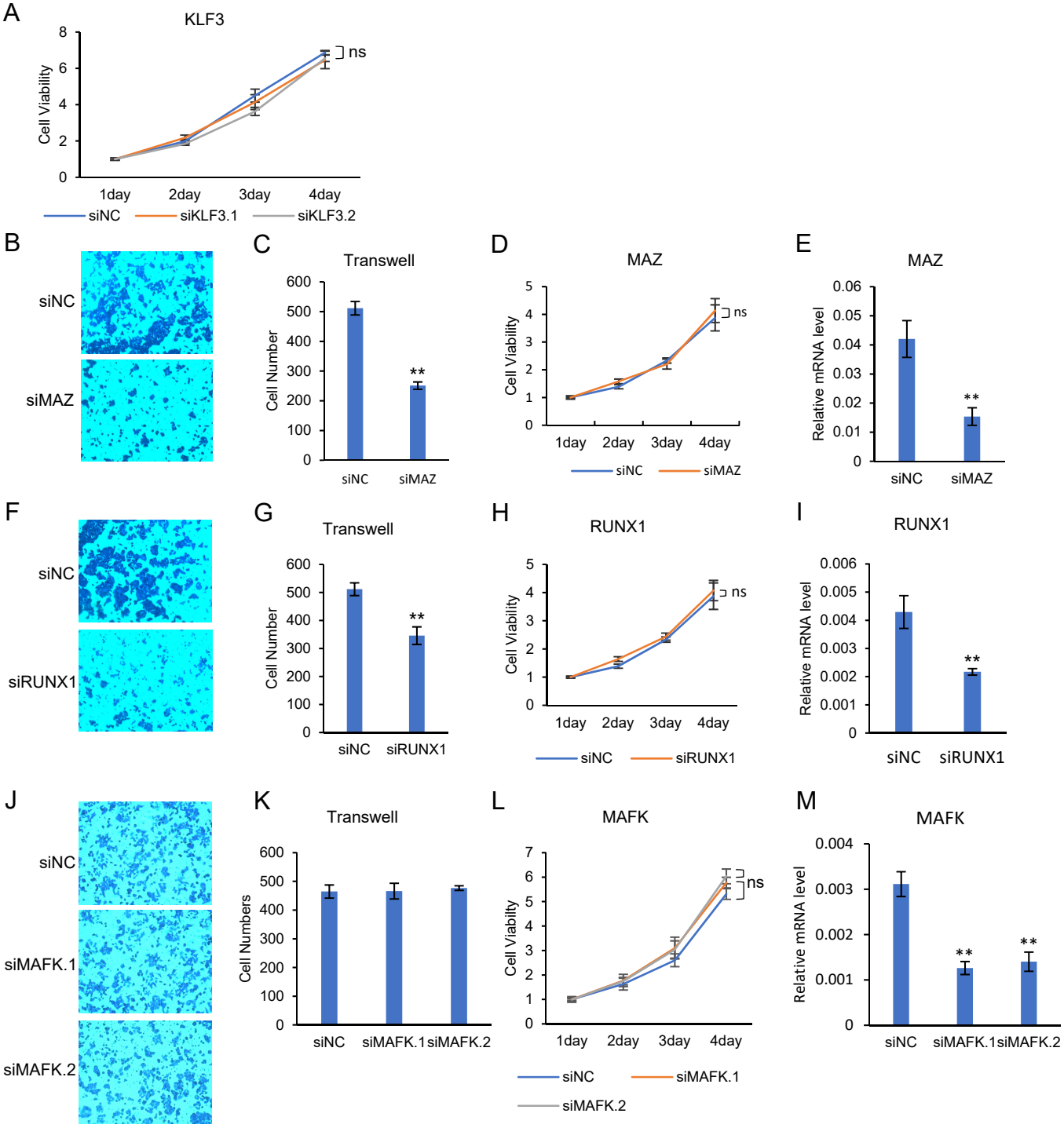
Extended Data Fig. S10 ChIP-qPCR of specific super enhancers in CRC. (A) ChIP-qPCR showing the H3K27ac level at CEBPE, CYP2S1, IER3, PHF19, RNF43 and TBC1D16 enhancer loci in control and sgRNA group. (B-D) 3C assays to check the chromatin interaction for *CEBPB* (B), *VEGFA* (C) and *CYP2S1* (D). * means p value < 0.05, ** means p value < 0.01, *** means p value < 0.001 .

Extended Data Fig. S11



Extended Data Fig. S11 Analysis of potential functional TFs in CRC. (A) DNA motifs enriched within nucleosome-free regions (NFRs) of lost VELs determined by HOMER motif analysis. **(B)** Heatmap of top 50 transcription factors ranked by motif p-value calculated from gain VEL NFRs of CRC patients. **(C)** Methodology for inferring the degree of core regulatory circuitry. **(D)** Meta normalized H3K27ac tracks at ASCL2 gene loci. **(E)** ASCL2 expression (FPKM) in patients of 12 cancer types. The group of our CRC tumor data was highlighted in red, and two intestinal cancer datasets from TCGA were highlighted in green. **(F)** Heatmap showing the significance (p-value) of transcription factors *KLF5*, *KLF3*, *MAFK*, *MAZ* and *RUNX1* in all patients.

Extended Data Fig. S12



Extended Data Fig. S12 Functional verification of predicted transcription factors . (A) Cell survival analysis of KLF3 knockdown cells. **(B-M)** Transwell and cell survival analysis of HCT116 cell with siRNAs of MAZ (B-E), RUNX1 (F-I), MAFK (J-M). ** means $p < 0.01$; ns means not significant.