Science Advances
AAAS

Supplementary Materials for

Parallel characterization of cis-regulatory elements for multiple genes using CRISPRpath

Xingjie Ren, Mengchi Wang, Bingkun Li, Kirsty Jamieson, Lina Zheng, Ian R. Jones, Bin Li, Maya Asami Takagi, Jerry Lee, Lenka Maliskova, Tsz Wai Tam, Miao Yu, Rong Hu, Lindsay Lee, Armen Abnousi, Gang Li, Yun Li, Ming Hu, Bing Ren, Wei Wang, Yin Shen*

*Corresponding author. Email: yin.shen@ucsf.edu

Published 15 September 2021, *Sci. Adv.* **7**, eabi4360 (2021)
DOI: 10.1126/sciadv.abi4360

The PDF file includes:

Figs. S1 to S10
Legends for tables S1 to S8

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S8

Fig. S1

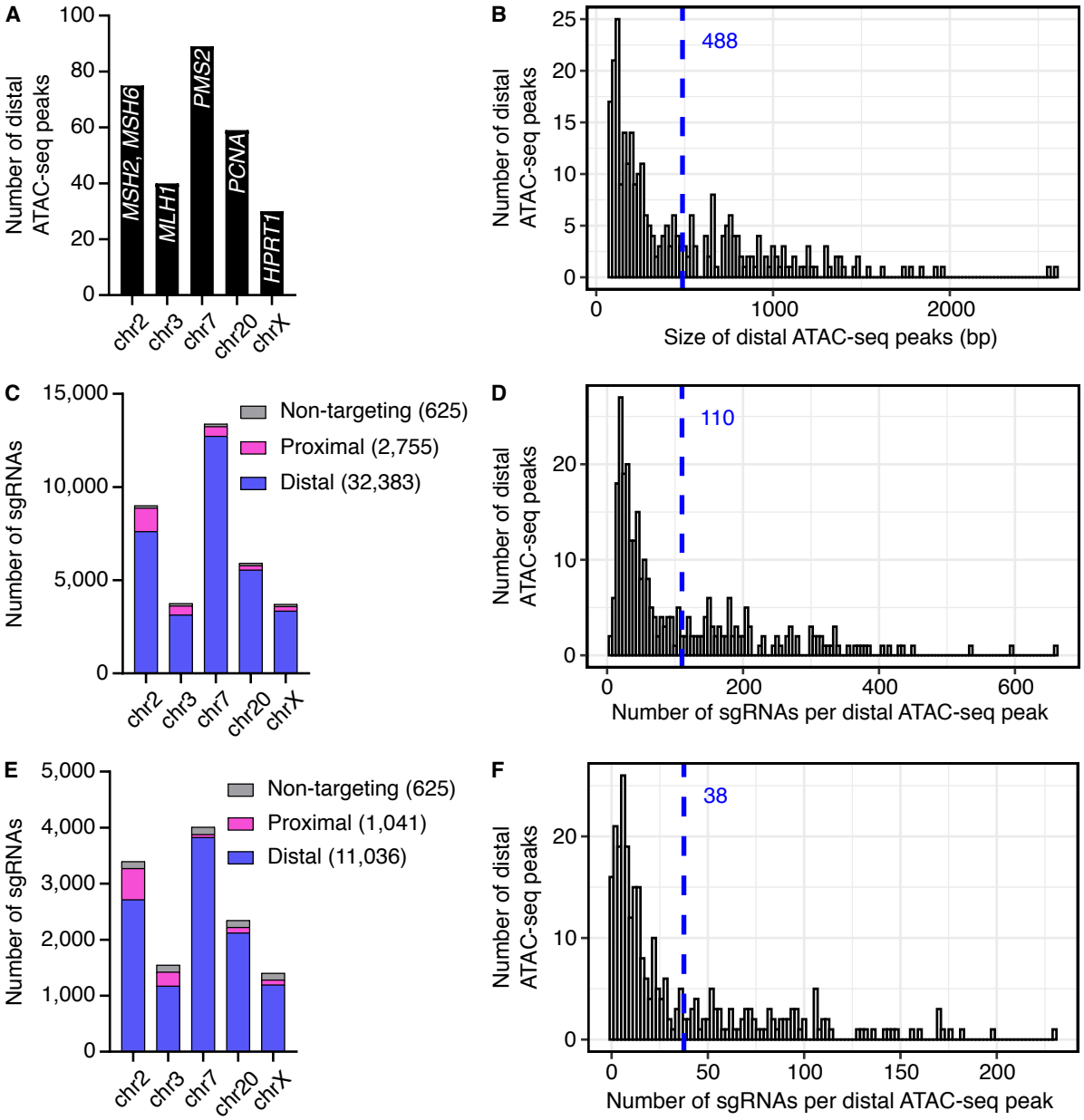

Fig. S1. Features of sgRNA library for CRISPRpath screen.

(**A**) Bar graph shows the number of distal ATAC-seq peaks used as candidate CREs for six MMR genes. (**B**) Histogram shows size distribution of distal ATAC-seq peaks. The average size is 488 bp (blue dash line). (**C**, **E**) The composition of the sgRNA library. In total, 35,763 sgRNAs were included in the library (C), and 12,702 sgRNAs are high quality sgRNAs (E). (**D**, **F**) Distribution of the number of total sgRNAs (D) or high quality sgRNAs (F) per distal ATAC-seq peak. Average numbers of sgRNA per ATAC-seq peak are indicated with blue dash lines.

Fig. S2

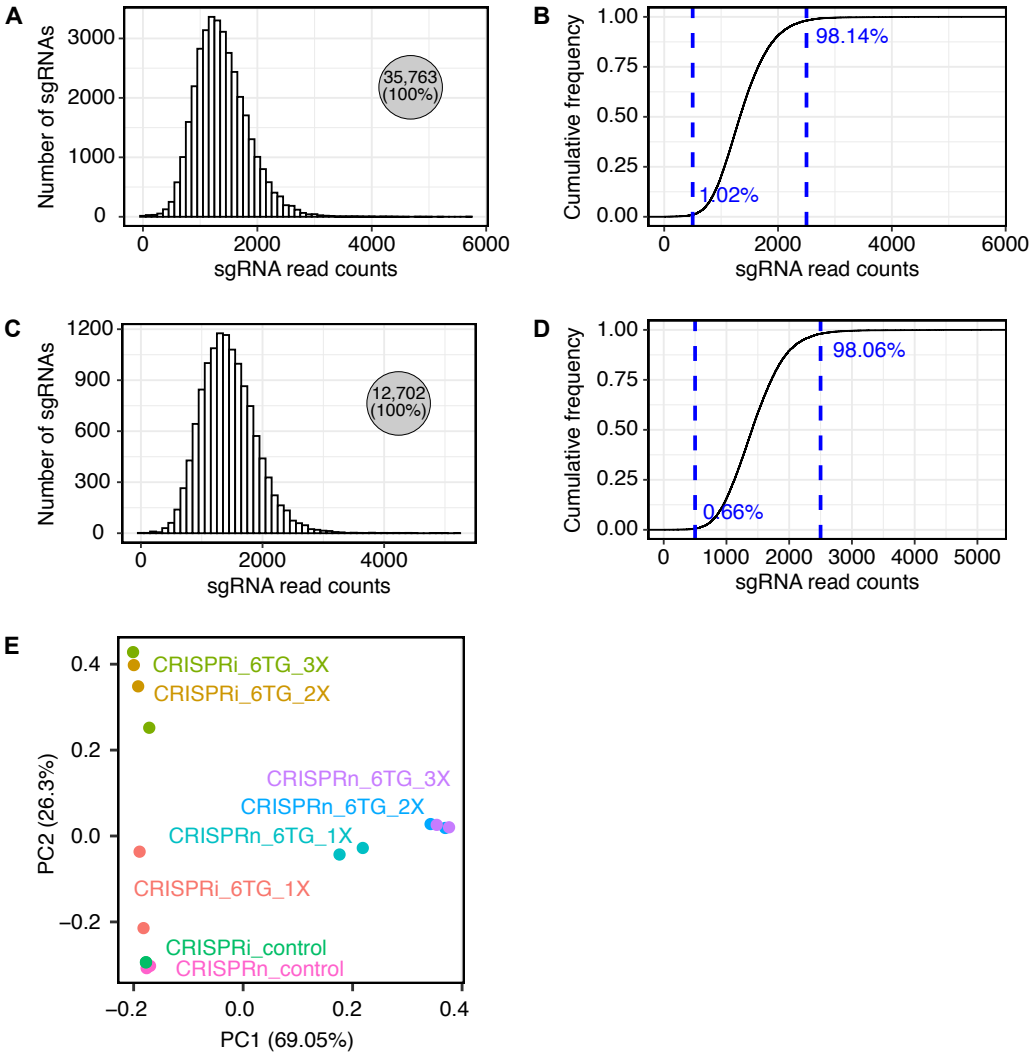

Fig. S2. Quality of the sgRNA library and CRISPRpath screen libraries.

(**A**, **C**) Distribution of sgRNA read counts for all the sgRNAs (A) and high quality sgRNAs (C) in the constructed sgRNA plasmid library. (**B**, **D**) Cumulative frequency of all the sgRNAs (B) and high quality sgRNAs (D) in the constructed sgRNA plasmid library. The constructed sgRNA plasmid library recovered all the designed sgRNAs (A, C) with the copy number difference less than five-fold for at least 97% of the designed sgRNAs (B, D). (**E**) PCA analysis shows the high reproducibility of the CRISPRpath screen libraries between biological replicates.

Fig. S3

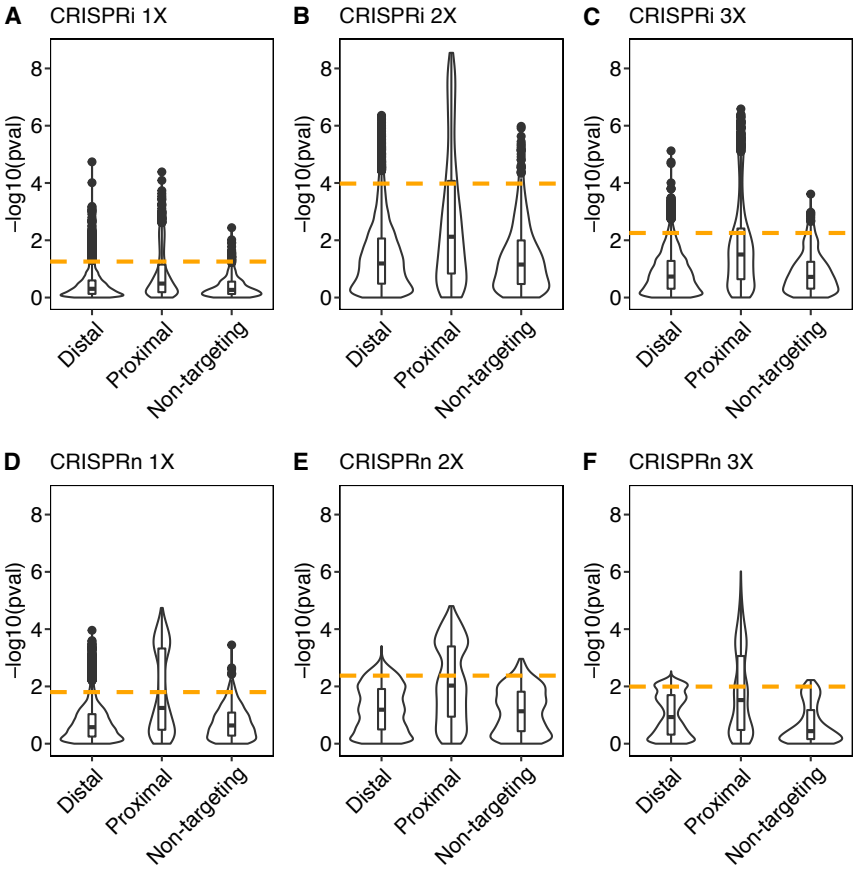

Fig. S3. *P* value cutoff used for identifying enriched sgRNAs from each screens.

(**A-F**) Distribution of *P* value for tested distal, proximal and non-targeting control sgRNA groups. Orange dash lines indicate 5% percentile of the *P* values from non-targeting control sgRNAs to achieve a false discovery rate of 5%.

Fig. S4

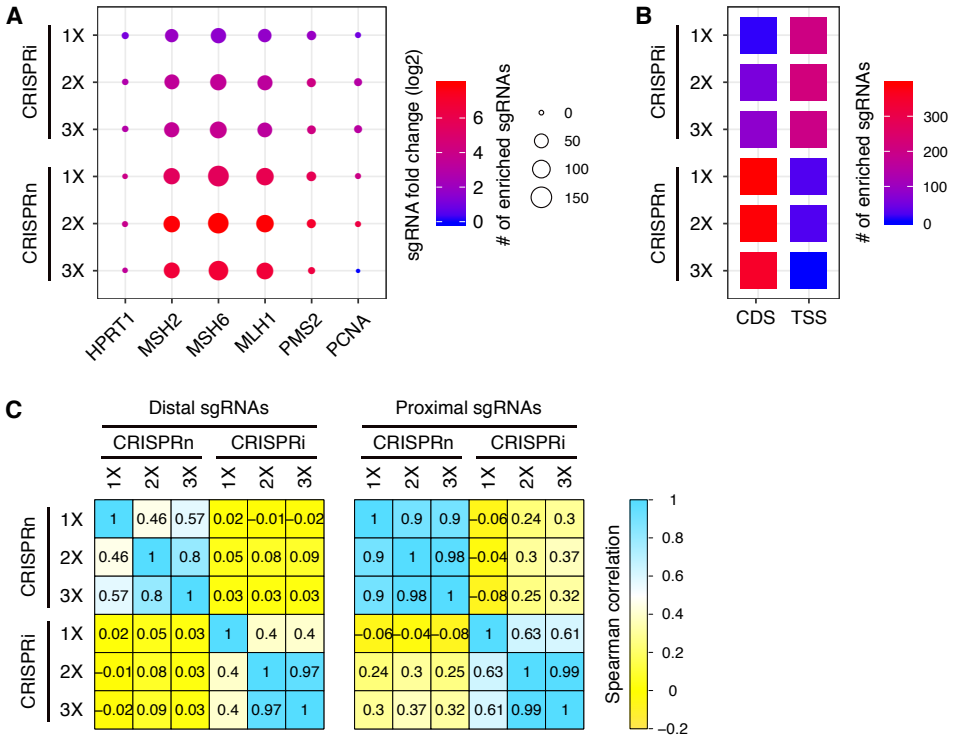

Fig. S4. Enriched proximal sgRNAs and sgRNA ranking analysis.

(**A**) Number and fold changes of the enriched proximal sgRNAs for the six target genes from CRISPRi and CRISPRn screens. The color indicates fold changes, and the size of circle indicates the number of enriched sgRNAs. (**B**) Enrichment analysis shows the enriched proximal sgRNAs bias towards to the TSS region for CRISPRi screens and the protein coding region (CDS) for CRISPRn screens. Color represents the number of enriched sgRNAs. (**C**) Spearman correlation analysis of the distal and proximal sgRNAs ranking shows proximal sgRNAs exhibiting higher correlation between each screen compared to distal sgRNAs.

Fig. S5

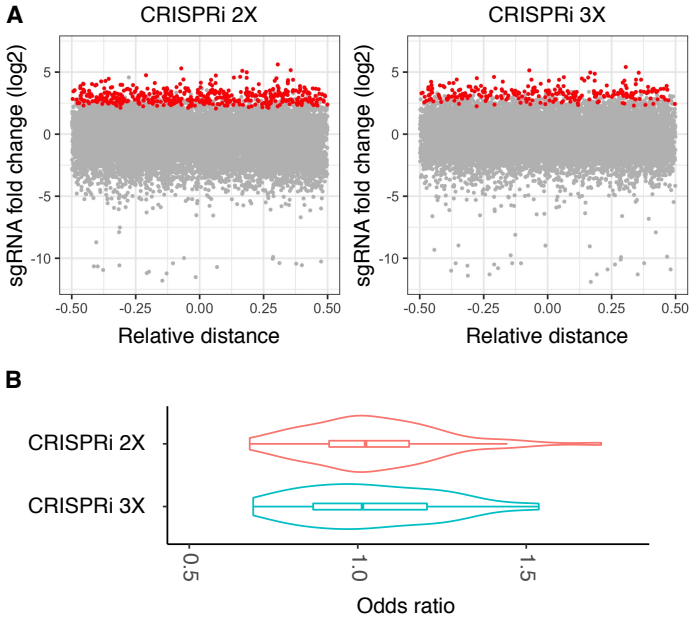

Fig. S5. Enriched sgRNAs identified from CRISPRi screens exhibit no position and strand preference.

(**A**) Enriched sgRNAs from CRISPRi 2X (red dots, n = 448) and CRISPRi 3X (red dots, n = 260) screens showed similar distributions across candidate CREs. (**B**) Odds ratio analysis of the fold changes of enriched sgRNAs shows enriched sgRNAs have no strand preference. Enhancers with enriched sgRNAs only targeting one strand were exclude for the analysis. Odds ratio was calculated for each element with the equation of ave($\log_2$(fold change of sgRNA targeting plus strand)) / ave($\log_2$(fold change of sgRNA targeting minus strand)). Violin plots show the distributions of odds ratio values within each screen, and boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR.

Fig. S6

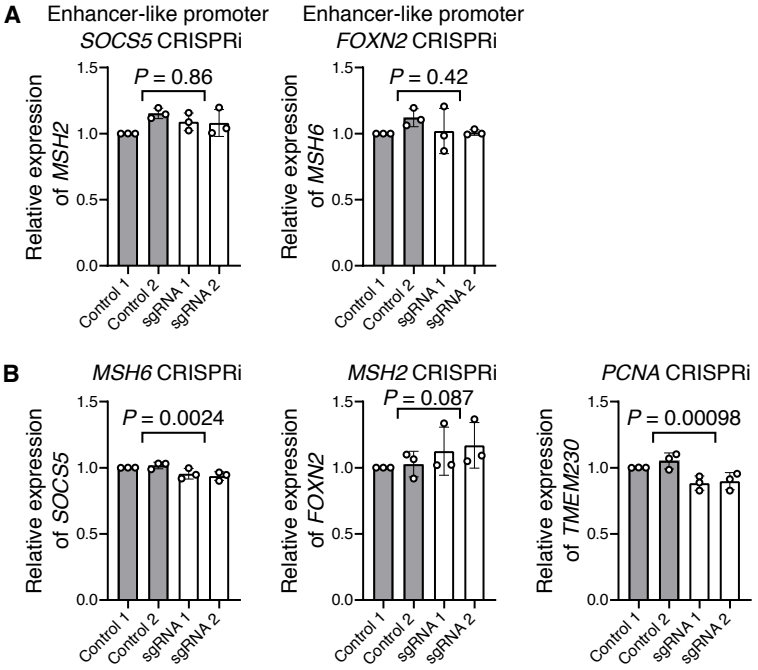

Fig. S6. CRISPRi experiments for enhancer-like promoters and their target genes.

(**A**) CRISPRi silencing of *SOCS5* and *FOXN2* promoters do not change *MSH2* and *MSH6* expression, respectively. (**B**) CRISPRi-mediated downregulation of *MSH6* and *PCNA* lead to subtle but significant changes in *SOCS5* and *TMEM230* expression, respectively. CRISPRi-mediated downregulation of *MSH2* does not lead to significant change in *FOXN2* expression. Three independent replicates per condition and two independent sgRNAs per replicate were used for each experiment. *P* values are from two-tailed two-sample *t*-test. Error bars represent the s.d.

Fig. S7

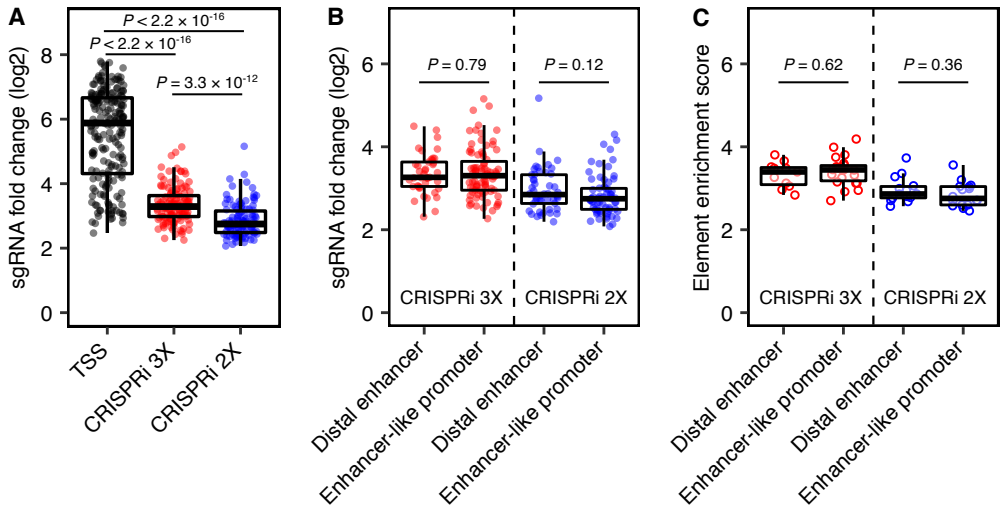

Fig. S7. sgRNA fold changes and element enrichment scores comparison for the identified enhancers.

(**A**) Box plots show the sgRNA fold changes of the enriched sgRNAs for the tested elements. The enriched sgRNAs in TSS regions of MMR genes show highest fold changes. Enriched sgRNAs of the enhancers uniquely identified from the lower selection pressure (CRISPRi 2X) exhibit lower fold changes compared to enriched sgRNAs of the enhancers identified from the higher selection pressure (CRISPRi 3X). Boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. The dots represent individual sgRNA. *P* values are from Wilcoxon test. (**B**) Box plots show the sgRNA fold changes of the enriched sgRNAs for different type enhancers. For the enhancers identified from CRISPRi 3X screen and the enhancers uniquely identified from CRISPRi 2X screen, sgRNA fold changes of the enriched sgRNAs did not show significant difference between distal enhancers and enhancer-like promoters. Boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. The dots represent individual sgRNA. *P* values are from Wilcoxon test. (**C**) Box plots show the element enrichment scores for enhancers. For the enhancers identified from CRISPRi 3X screen and the enhancers uniquely identified from CRISPRi 2X screen, element enrichment score did not show significant difference between distal enhancers and enhancer-like promoters. Boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. The dots represent individual sgRNA. *P* values are from Wilcoxon test.

Fig. S8

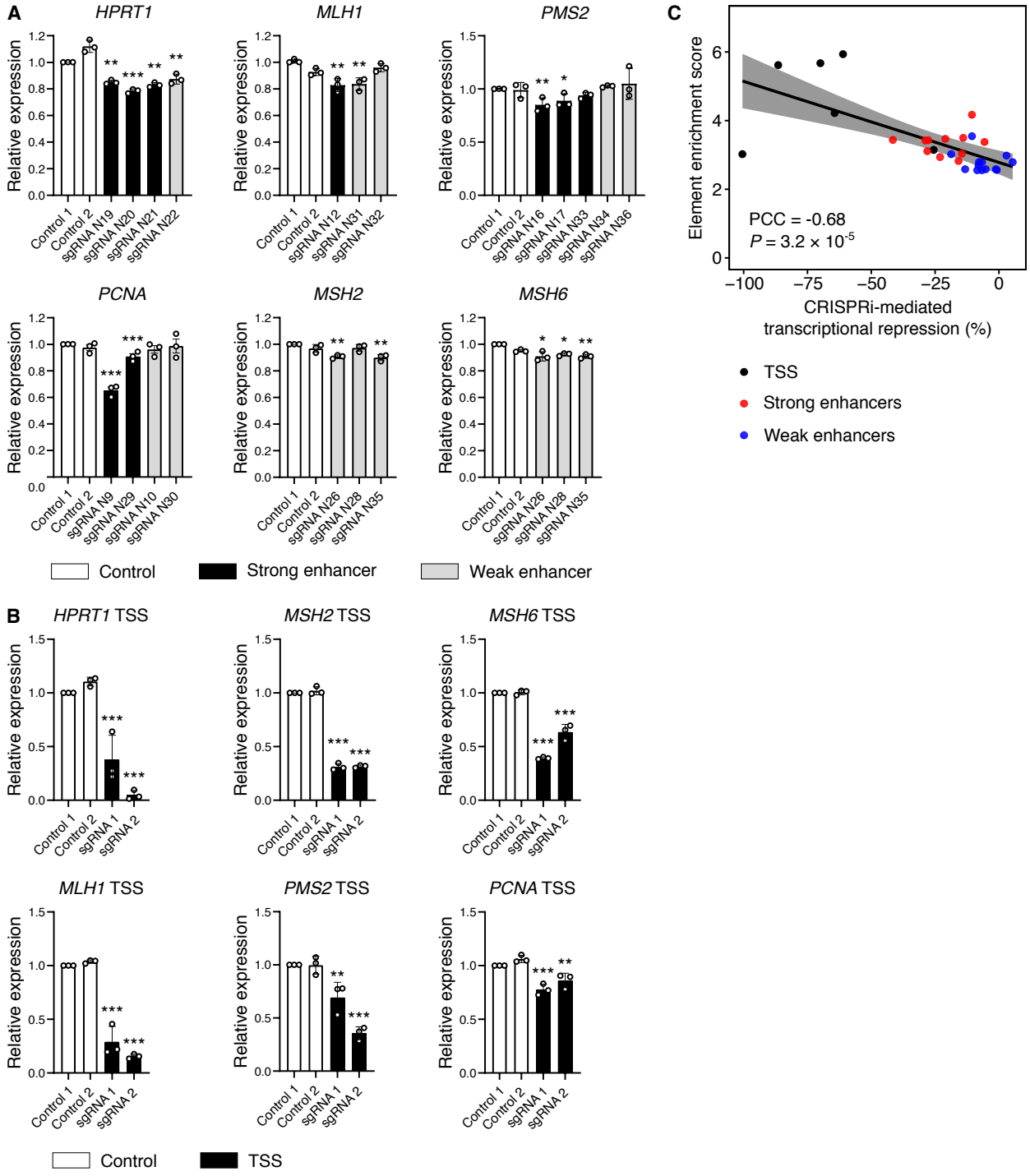

Fig. S8. Validation of CIRSPRpath identified enhancers.

(**A**) Validation of the strong (black) and weak (grey) enhancers with CRISPRi followed by RT-qPCR. Three independent replicates per condition. The significance was calculated with two-tailed two-sample *t*-test. Data are mean and s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. (**B**) CRISPRi-mediated transcriptional repression of six target genes by targeting TSS of each gene. Three independent replicates per condition. The significance was calculated with two-tailed two-sample *t*-test. Data are mean and s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. (**C**) Pearson correlation analysis reveals element enrichment scores from CRISPRpath screens correlate with element effect sizes on transcription from CRISPRi (Pearson correlation, PCC = -0.68, $P = 3.2 \times 10^{-5}$).

Fig. S9

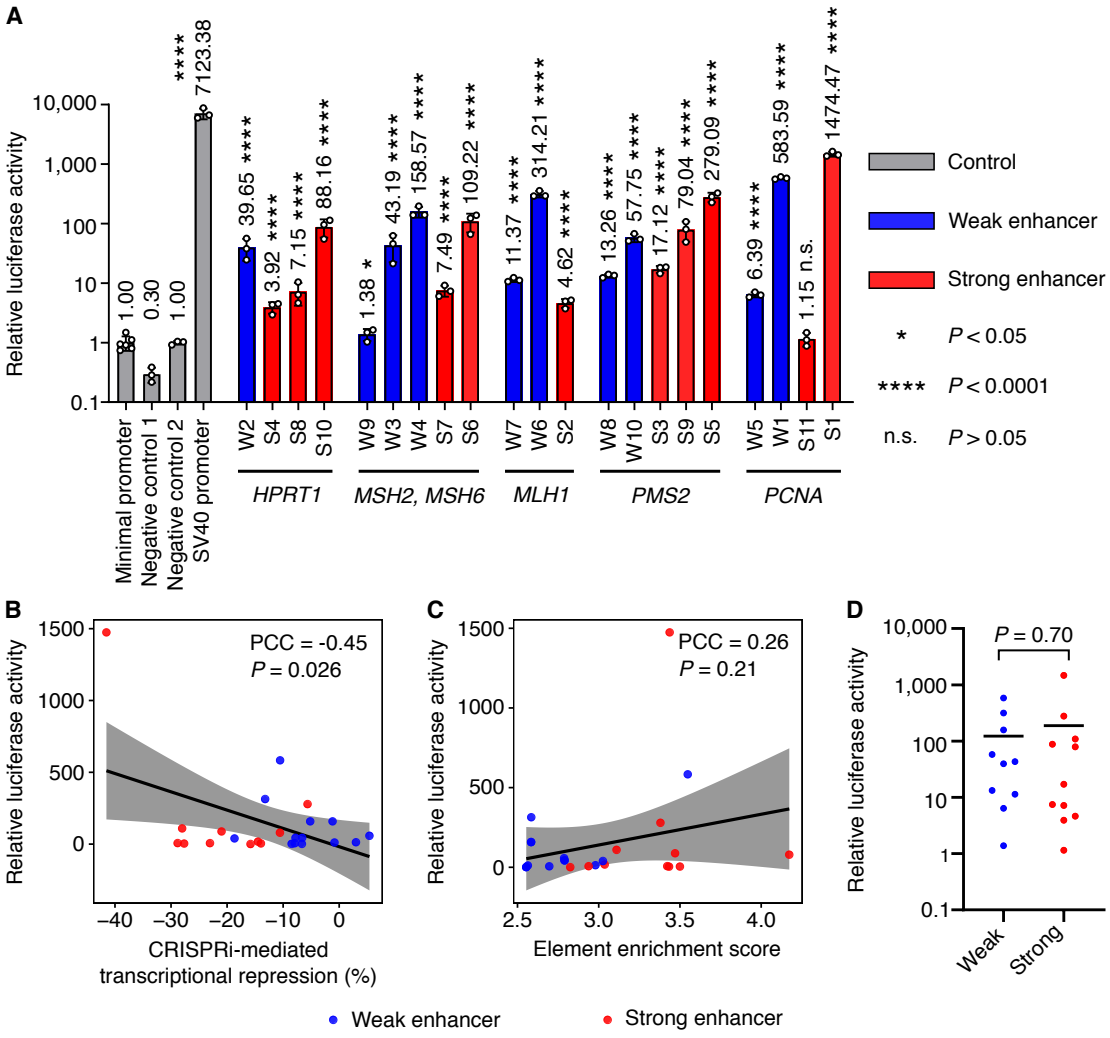

Fig. S9. Characterization of enhancers with dual-luciferase reporter assay.

(**A**) Relative luciferase reporter activities for the weak and strong enhancers. The relative luciferase activities were normalized to the average of the luciferase activity of minimal promoter. The significance was compared to minimal promoter and two negative control elements, and calculated with two-tailed two sample *t*-test. Data are mean and s.d. (**B**) Pearson correlation analysis reveals weak correlation between relative luciferase activities and element effect sizes on transcription from CRISPRi (Pearson correlation, PCC = -0.45, *P* = 0.026). (**C**) Pearson correlation analysis reveals weak correlation between relative luciferase activities and element enrichment scores from CRISPRpath screens (Pearson correlation, PCC = 0.26, *P* = 0.21). (**D**) Dotplot shows the distribution of relative luciferase activities of weak and strong enhancers. Each dot represents one tested enhancer. The black line labels the average of relative luciferase activity. The significance was calculated with two-tailed two sample *t*-test.

Fig. S10

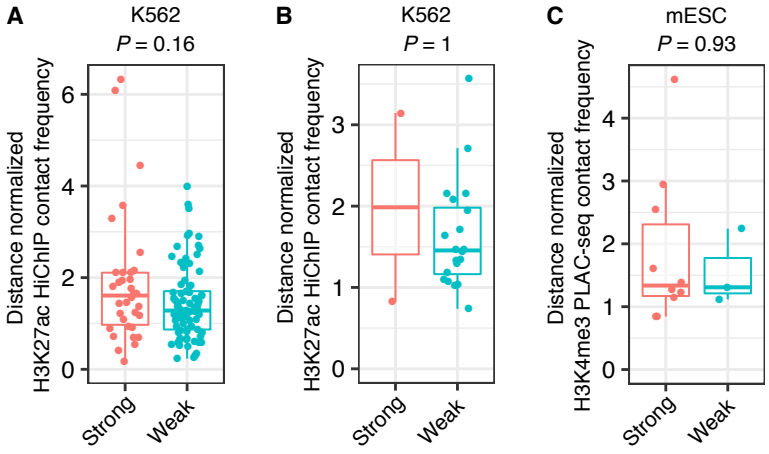

Fig. S10. Chromatin contract frequency analysis for the enhancers in K562 cells and mESCs.

(**A**) Distance normalized H3K27ac HiChIP contact frequency for strong (n = 34) and weak (n = 82) enhancers identified with crisprQTL mapping in K562 cells. (**B**) Distance normalized H3K27ac HiChIP contact frequency for strong (n = 2) and weak (n = 20) enhancers identified with CRISPRi-FlowFISH screen in K562 cells. (**C**) Distance normalized H3K4me3 PLAC-seq contact frequency for strong (n = 10) and weak (n = 3) enhancers identified in mouse embryonic stem cells. Boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. *P* values are from Wilcoxon test.

Supplementary Tables:

Table S1. ATAC-seq peak regions used to design the sgRNA library.

Table S2. List of sgRNA sequences used for CRISPRpath screen.

Table S3. Enhancers identified from CRISPRn 2X, CRISPRi 2X and CRISPRi 3X screens.

Table S4. List of primers used for RT-qPCR.

Table S5. List of shRNA sequences used for RNA interference experiments.

Table S6. List of sgRNA sequences used for CRISPRi-mediated enhancer validation experiments.

Table S7. Pairwise comparisons of data in Figure 4C.

Table S8. List of primers used for dual-luciferase reporter assay.