Parallel Characterization of *cis*-Regulatory Elements for Multiple Genes Using CRISPRpath

Authors: Xingjie Ren[1], Mengchi Wang[2], Bingkun Li[1], Kirsty Jamieson[1], Lina Zheng[2], Ian R. Jones[1], Bin Li[3], Maya Asami Takagi[1], Jerry Lee[1], Lenka Maliskova[1], Tsz Wai Tam[1], Miao Yu[3], Rong Hu[3], Lindsay Lee[4], Armen Abnousi[4], Gang Li[5], Yun Li[6,7,8], Ming Hu[4], Bing Ren[3,9,10], Wei Wang[2,11], Yin Shen[1,12,#]

Affiliation:

[1]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA.

[2]Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, USA.

[3]Ludwig Institute for Cancer Research, La Jolla, California, USA.

[4]Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH, USA.

[5]Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC, USA.

[6]Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA.

[7]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA.

[8]Department of Computer Science, University of North Carolina, Chapel Hill, NC, USA.

[9]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA.

[10]Moores Cancer Center, University of California, San Diego, La Jolla, California, USA.

[11]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA.

[12]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA.

[#]Corresponding Author: yin.shen@ucsf.edu

Abstract

Current pooled CRISPR screens for *cis*-regulatory elements (CREs) can only accommodate one gene based on its expression level. Here, we describe CRISPRpath, a scalable screening strategy for parallelly characterizing CREs of genes linked to the same biological pathway and converging phenotypes. We demonstrate the ability of CRISPRpath for simultaneously identifying functional enhancers of six genes in the 6-thioguanine-induced DNA mismatch repair pathway using both CRISPR interference (CRISPRi) and CRISPR nuclease (CRISPRn) approaches. 60% of the identified enhancers are known promoters with distinct epigenomic features compared to other active promoters, including increased chromatin accessibility and interactivity. Furthermore, by imposing different levels of selection pressure, CRISPRpath can distinguish enhancers exerting strong impact on gene expression from those exerting weak impact. Our results offer a nuanced view of *cis*-regulation and demonstrate that CRISPRpath can be leveraged for understanding the complex gene regulatory program beyond transcriptional output at scale.

Main

Cis-regulatory elements (CREs) are key regulators for spatial-temporal control of gene expression. Mutations in CREs can contribute to complex diseases by modulating gene expression over long genomic distances[1-3]. Thus, functionally characterizing CREs can provide important insight into gene regulation mechanisms and enable us to better interpret non-coding genetic variants associated with diseases. Despite the fact that tremendous numbers of candidate CREs have been mapped by biochemical signature[4], our knowledge of whether, how, and how much these putative CREs are functional on gene expression remain scarce in the human genome. Pooled CRISPR screens have been developed for testing CREs in their native chromatin context by monitoring the transcriptional levels for the gene of interest[5-11]. Although results from these studies have made significant contributions to the annotation of functional DNA elements, challenges remain in pooled CRISPR screens of CREs. First, CRISPR screens for enhancers based on gene expression levels largely depend on generating reporter knock-in cell lines[7] or using FlowFISH signals[8]. These procedures, involving generation of reporter lines and selection of cells with positive hits by flow cytometry, are time-consuming and difficult to scale up to multiple genes in the same experiment. Second, the approaches of using gene

61 expression as the screening phenotype[9, 10] fail to connect the functions of DNA elements from
62 transcriptional regulation at the molecular level to interpretable cellular and physiological
63 functions. Third, in cases of CRE screens using phenotypes such as cell proliferation and
64 survival[11, 12], they fail to quantify the effect sizes of enhancers on transcriptional output.
65
66 To address these limitations, we developed CRISPRpath, a pooled CRISPR screening
67 approach to simultaneously characterizing CREs for multiple target genes involved in the same
68 biological pathway. CRISPRpath allows us to screen functional DNA elements based on
69 phenotypes associated with well-defined biological pathways. We demonstrate the capacity of
70 CRISPRpath by performing CRISPR interference (CRISPRi) and nuclease (CRISPRn) screens
71 for six genes in human induced pluripotent stem cells (iPSCs), and reveal different strengths of
72 enhancer functions by imposing varying levels of selection pressure on the cells.
73
74 **Results**
75 **Leveraging CRISPRpath for parallel characterization of CREs for multiple genes in**
76 **iPSCs.**
77 To characterize candidate *cis*-regulatory elements (CREs) for multiple genes within the same
78 pooled CRISPR screening, we designed and applied CRISPRpath to six genomic loci
79 containing six genes (*HPRT1, MSH2, MSH6, MLH1, PMS2, PCNA*) involved in the 6-
80 thioguanine (6TG)-induced mismatch repair (MMR) (**Fig. 1a**). The MMR pathway is highly
81 conserved and essential for the maintenance of genome stability[13]. The MMR pathway
82 recognizes DNA mismatches caused by 6TG treatment and induces cell apoptosis[14, 15]. On the
83 other hand, cells with a malfunctioning MMR pathway, due to aberrant expression levels of 6TG
84 metabolism genes or MMR genes, may survive during 6TG treatment. Employing the properties
85 of the MMR pathway, we used cell survival for selecting cells with the reduced expression of
86 MMR genes due to defects in enhancer activities (**Fig. 1b**). To design the screening library, we
87 first identified open chromatin regions by performing Assay for Transposase Accessible
88 Chromatin using sequencing (ATAC-Seq) in WTC11 iPSCs. We included all open chromatin
89 regions defined by ATAC-seq peaks located 1Mb upstream and 1Mb downstream of each of
90 the six genes (spanning a total of 10.6 Mb genomic regions) as candidate CREs for functional
91 characterization (**Supplementary Fig. 1a, b, Supplementary Table 1**). We then designed a

92 sgRNA library with 32,383 distal sgRNAs targeting 294 distal ATAC-seq peaks, 2,755 proximal
93 sgRNAs targeting 81 ATAC-seq peaks overlapped with transcription start site (TSS) and coding
94 regions of the six genes, and 625 non-targeting sgRNAs with genomic sequences in the same
95 genomic loci but are not followed by PAM sequences (**Supplementary Fig. 1c, Supplementary**
96 **Table 2**). In total, we included 35,763 sgRNAs in the library with an average of 110 sgRNA per
97 ATAC-seq peak (**Supplementary Fig. 1d, Supplementary Fig. 2a, b**). We generated a
98 lentiviral library expressing these sgRNAs and transduced this library into two engineered
99 WTC11 iPSC lines, one expressing doxycycline-inducible dCas9-KRAB (CRISPRi) and the
100 other doxycycline-inducible Cas9 (CRISPRn)[16], both at a multiplicity of infection (MOI) of 0.5
101 (**Fig. 1b**).
102
103 To carry out the screening, we pre-determined the minimal lethal concentration of 6TG at 80
104 ng/mL for CRISPRi and CRISPRn iPSC lines (see **Methods** for more details), and applied three
105 different 6TG concentrations (1X: 80 ng/mL, 2X: 160 ng/mL, 3X: 240 ng/mL) in both CRISPRi
106 and CRISPRn screens. We extracted and sequenced DNA samples from the survival cells
107 seven days after 6TG treatment to determine enriched sgRNAs by comparing the results to that
108 of the control cells taken after sgRNA library infection before the 6TG treatment (**Fig. 1b**). To
109 avoid confounding signals generated by off-target effects of low-quality sgRNAs[17], we only used
110 sgRNAs with high specificity (defined as specificity score > 0.2[18], and without any off-target sites
111 with sequence similarity of ≤2 mismatches) for data analysis. This led to the use of a total of
112 12,702 high-quality sgRNAs with an average of 38 sgRNAs per ATAC-seq peak for analysis
113 (**Supplementary Fig. 1e, f, Supplementary Fig. 2c, d**). We performed each screen in two
114 biological replicates with each pair of replicates exhibiting high reproducibility (**Supplementary**
115 **Fig. 2e**). We compared the abundance of each sgRNA between the 6TG treated population and
116 the control population using a negative binomial model, and computed the fold change and *P*
117 value to quantify the effect size and the significance of enrichement of each sgRNA. We used
118 the 5% percentile of the *P* values from non-targeting control sgRNAs as the empirical
119 significance threshold to achieve a false discovery rate of 5%. sgRNAs with *P* value less than
120 the empirical significance threshold and with fold change > 2 were defined as enriched.
121 (**Supplementary Fig. 3)**. As expected, sgRNA targeting TSS and coding region were identified
122 as positive hits from both CRISPRi and CRISPRn screens exhibiting greater fold change in

123 CRISPRn screens compared to the CRISPRi screens (**Fig. 1e, Supplementary Fig. 4a**). We
124 also observed enrichment of sgRNA bias towards coding regions over TSS regions for
125 CRISPRn screen (**Supplementary Fig. 4b**). These results are consistent with CRISPRi
126 functioning best near TSS by inhibiting transcription, and CRISPRn can disrupt gene function
127 by generating indels downstream of TSS[19, 20].
128
129 Further sgRNA fold-change ranking analysis revealed strong positive correlation between the
130 screens with 2X and 3X 6TG treatment for both CRISPRi and CRISPRn screen (Spearman
131 correlation, CRISPRi = 0.97, CRISPRn = 0.84) (**Fig. 1c**) with the correlations for proximal
132 sgRNAs being higher than for distal sgRNAs (**Supplementary Fig. 4c**). On the contrary, results
133 from the 1X screen correlated poorly with either 2X or 3X screens (**Fig. 1c**), suggesting more
134 substantial selection pressure (2X and 3X) can reduce background noise in CRISPRpath
135 screens. Thus, we used sgRNAs enriched from 2X and 3X screens data for identifying active
136 enhancers in the following section (**Fig. 1d**).
137
138 **CRISPRi is more efficient than CRISPRn in pooled CRISPR screens of CREs.**
139 Performing CRISPRpath with CRISPRi and CRISPRn in the same genetic background with an
140 identical sgRNA library offers a unique opportunity for comparing the efficacies of CRISPRi and
141 CRISPRn in pooled CRISPR screens of CREs. We noticed that CRISPRn screens recovered
142 fewer enriched distal sgRNAs than CRISPRi screens (**Fig. 1f**). This is possibly due to the fact
143 that CRISPRi-mediated heterochromatin formation can more effectively perturb CREs
144 compared to CRISPRn-mediated genetics perturbations. We then called an candidate element
145 as an enhancer if there are at least 3 enriched sgRNAs in that CRE. Based on this criterion, we
146 identified 62 and 33 enhancers from the 2X and 3X CRISPRi screen, respectively, and 19
147 enhancers from the 2X CRIPSRn screen. (**Fig. 1g, Supplementary Table 3**). However, no
148 enhancer was identified from the 3X CRISPRn screen, indicating either the CRISPRn induced
149 mutations did not lead to any strong effect on gene expression to make the cells survive the 3X
150 6TG treatment or there are insufficent numbers of sgRNAs exhibiting deleterious effects on the
151 tested DNA elements to satisfy our criterion of calling functional enhancers. In total, 66 unique
152 enhancers were identified for the six target genes with CRISPRpath under different 6TG
153 treatments (**Fig. 1g**). Together, we demonstrate CRISPRpath can simultaneously identify

154 enhancers for multiple target genes with CRISPRi outperforming CRISPRn. For the following
155 analysis, we focused on the 63 enhancers identified from the 2X and 3X CRISPRi screens (**Fig.**
156 **1g**).
157
158 **Genomic feature of CRISPRpath identified enhancers.**
159 To determine the genomic feature of the enhancers, we plotted all the tested elements by their
160 genomic locations and enrichment scores (average of $\log_2$(fold change) of enriched sgRNAs of
161 each element) (**Fig. 2a**). Not surprisingly, our data suggest that each gene can be regulated by
162 multiple enhancers with the identified functional enhancers having no position bias relative to
163 the TSS. The average distance between an enhancer and its paired TSS is about 530 Kb (**Fig.**
164 **2b**) with an average of 10 interval genes between an identified enhancer and its target gene
165 pairs (**Fig. 2c**). Interestingly, we observed a weak negative correlation between the enhancer
166 enrichment score and the distance between an enhancer and its paired TSS (**Fig. 2d**, Pearson
167 correlation, ρ = -0.36, P = 0.01), suggesting enhancers near to TSS tend to have higher
168 regulatory activity compared to enhancers further away from their target genes. It is worth
169 noting, the relative positions for the enriched sgRNAs exhibited no preference relative to ATAC-
170 seq peaks (**Fig. 2e, Supplementary Fig. 5a**) and no preference for the strand on which the
171 sgRNAs were designed (**Supplementary Fig. 5b**), consistent with our knowledge that CRISPRi
172 mediated heterochromatin spreads over hundreds of base pairs in distance[21].
173
174 Previous studies have revealed that promoters can function as enhancers[7, 22]. Indeed, 60% (38
175 out of 63) of the functional enhancers identified in CRISPRi screens overlapped with annotated
176 promoters, providing an excellent opportunity to further explore the genomic features of these
177 enhancer-like promoters. To validate whether these promoters function as bona fide enhancers,
178 we targeted three enhancer-like promoters with CRISPRi. We confirmed significant
179 downregulation of their target genes including *MSH6*, *MSH2*, and *PCNA* (**Fig. 3a-c**). In contrast,
180 shRNAs against the transcripts from these promoters (*SOC5*, *FOXN2*, and *TMEM230*) only led
181 to a significant downregulation of its own transcripts and did not affect their target gene
182 expression (**Fig. 3a-c**). These results confirm that these promoter sequences identified by
183 CRISPRpath can function as enhancers. Although it has been shown that enhancer-like
184 protomors are enriched with active chromatin marks and physically close to target genes[7], it is

185 not clear whether enhancer-like promoters have unqiue genome features that can differentiate
186 them from other regular active promoters. To this end, we compared chromatin accessibility,
187 occupancy of histone 3 lysine 4 trimethylation (H3K4me3), histone 3 lysine 27 acetylation
188 (H3K27ac) and CTCF, transcription, and chromatin interactivity levels between enhancer-like
189 promoters and all other active promoters that did not show enhancer activity in our CRISPRi
190 screens. We show that enhancer-like promoters exhibit higher chromatin accessibility, higher
191 level of transcription, stronger H3K4me3 and H3K27ac signals than those at other active
192 promoters (**Fig. 3d**). On the other hand, we did not observe a significant difference for CTCF
193 binding signals between enhancer-like promoters and control promoters (**Fig. 3d**). Furthermore,
194 by evaluating chromatin interaction data using H3K4me3 Proximity Ligation-Assisted ChIP-seq
195 (PLAC-seq), we show enhancer-like promoters have significantly more and stronger
196 interactions compared to control promoters (**Fig. 3e**).
197
198 **CRISPRpath is capable of distinguishing enhancers with distinct effect sizes.**
199 Gene expression often is a result of combinatorial regulatory effects from multiple *cis*-regulatory
200 elements[11, 23]. Understanding how individual enhancers contribute to gene expression in a
201 quantitative manner is an important first step in dissecting how enhancers orchestrate precise
202 transcriptional control. We seek a new strategy to differentiate enhancers based on their effect
203 sizes on gene expression using CRISPRpath. We hypothesized that cells with drastic down-
204 regulation of MMR genes have a fitness advantage under higher 6TG concentration than cells
205 with modest down-regulation of MMR genes. Consistent with this hypothesis, proximal sgRNAs
206 exhibit larger fold changes than distal sgRNAs (**Fig. 1e**) because perturbing proximal regions
207 has more profound effects on gene down-regulation than perturbing distal regulatory regions.
208 Based on these observations, we hypothesize that enhancers identified under different
209 selection pressure represent distinct regulatory strengths on transcriptional activation. We
210 noticed that enhancers identified under strong selection pressure (3X) have higher enrichment
211 scores compared to enhancers uniquely identified under weak selection pressure (2X), with the
212 TSS regions manifesting the highest enrichment scores (**Fig. 4a**). Thus, enhancers identified in
213 the 3X screen are strong enhancers (n=33), while enhancers uniquely identified in the 2X screen
214 are weak enhancers (n=30) (**Fig. 1g**). To confirm the quantitative effect of enhancers on target
215 gene expression, we tested 11 strong and 10 weak enhancers using CRISPRi followed by RT-

216 qPCR measurement of the corresponding target gene expression (**Fig. 4b, Supplementary**
217 **Fig. 6a**). We show perturbations of strong enhancers led to significantly more down-regulation
218 of target gene expression (mean down-regulation of target gene by 21%) than perturbations of
219 weak enhancers (mean down-regulation of target gene by 6%), with the perturbations of TSS
220 regions achieving the strongest down-regulation of target genes, by an average of 68%
221 reduction in gene expression (**Fig. 4b, Supplementary Fig. 6a, b**). These quantitative effects
222 on target gene expression are consistent with the enrichment scores from our CRISPRpath
223 screens (**Supplementary Fig. 6c**) and demonstrate the capacity of distinguishing enhancers
224 with different effect sizes by imposing different levels of selection pressures.
225
226 We further explored chromatin features of strong and weak enhancers by analyzing chromatin
227 accessibility, H3K4me3, H3K27ac, and CTCF binding signals in these regions. At individual
228 chromatin mark level, while CRISPRpath-identified enhancers were more accessible, and
229 enriched with active chromatin marks, such as H3K4me3 and H3K27ac, and CTCF binding
230 compared to negative elements or random elements (**Fig. 4c**), we did not observe significant
231 differences between strong and weak enhancers in the chromatin features we individually
232 examined. However, strong enhancers tend to have more active chromatin signatures than
233 weak enhancers (**Fig. 4d**), suggesting combined signatures of active chromatin can be a better
234 indicator of enhancer strength. Strong enhancers tend to have higher distance normalized
235 PLAC-seq contact frequencies with their target promoters than weak enhancers, though not
236 statistically significant, possibly due to the small sample size in this study (**Fig. 4e**). We obtained
237 similar results by expanding this analysis for characterized enhancers in K562 cells and mouse
238 embryonic stem cells[6, 8] (**Supplementary Fig. 7**), which reinforces the idea that enhancers with
239 larger effects on gene expression tend to have higher chromatin interactions with their cognate
240 promoters. To explore the possible mechanisms that drive enhancer activities in a quantitative
241 manner, we evaluated potential transcription factors (TFs) binding motifs in strong and weak
242 enhancer sequences. Both strong and weak enhancers are enriched with CTCF binding motif
243 (**Fig. 4f**). Indeed, most of strong and weak enhancers are bound by CTCF (**Fig. 4d**), consistent
244 with the notion that CTCF-mediated chromatin loops are essential for gene activation[24].
245 Furthermore, strong enhancers and weak enhancers have differential enrichment with TFs
246 binding motifs. For example, the binding motifs for SP/KLF family[25] and E2F family[26, 27] appear

247 more frequently in strong enhancers compared to weak enhancers, suggesting these strong
248 enhancers could be major docking sites for master regulators in iPSCs (**Fig. 4f**).
249
250 **Discussion**
251 CRISPR-mediated high throughput screening using bulk cells allows for the functional
252 characterization of regulatory elements in their native genomic context. However, current
253 approaches are limited to validating a small number of regulatory elements for a single gene[5,
254 7, 9, 12, 28, 29]. To overcome this bottleneck, we developed CRISPRpath, a strategy for functional
255 characterization of enhancers for multiple genes simultaneously by leveraging the genes
256 involved in the same biological pathway so that the effects can be measured via a define
257 phenotype. For example, alpha-toxin resistance phenotype can be used to identify CREs for 17
258 genes in glycosylphosphatidylinositol (GPI)-anchor synthesis pathway[30]. CRISPRpath can also
259 be leveraged to identify CREs for protein folding regulators that contribute to the endoplasmic
260 reticulum stress-response pathway[31] using UPRE reporter in mammalian cells. Compared to
261 the existing pooled CRISPR screens of CREs[5, 7, 8, 10-12, 28, 29, 32], CRISPRpath is scalable with
262 additional benefits of connecting DNA elements to cellular function, beyond the most standard
263 molecular phenotype of gene expression.
264
265 Promoters can function as enhancers more widespread than expected, with more than half of
266 the enhancers identified for MMR genes in our study being previously annotated promoters.
267 This is consistent with previous reports that enhancer-like promoters are more prevalent for
268 ubiquitously expressed genes[33]. Enhancer-like promoters are more accessible compared to
269 other promoters, possibly because these regions are required to be more open to accommodate
270 additional transcriptional machinery such as TF for activating target gene expression besides
271 their own transcription[34]. Enhancer-like promoters also exhibit significantly higher levels of
272 chromatin interactions with distal regions compared to other active promoters. This observation
273 can be explained by the fact that enhancer-like promoters will not only form chromatin loops
274 with their distal target genes, but also with CREs for controlling the expression of their own
275 genes.
276

Genomic studies of chromatin marks have revealed hundreds of thousands candidate CREs in the human genome but with very little quantitative information regarding how CREs contribute to gene regulation[35, 36]. Using CRISPRpath, we can systematically classify enhancers based on their effect sizes on transcription. Identifying and charaterizing the effect size for each individual enhancer is the critical first step to future studies of their combinatory effects on target gene expression. Interestingly, strong and weak enhancers can not be distinguished by individual epigenetic marks we examined. One possible explanation for this observation is that chromatin features only mark enhancer's identity but do not quantify enhancer activity. On the other hand, the strong and weak enhancers we identified may regulate other genes differently from regulating the MMR gene. Interestingly, strong enhancers tend to harbor more than one active chromatin signature, which indicates that enhancer activities are regulated by multiple epigenetic factors, for example, TF mediated transcriptional regulation. Differential TFs binding motifs observed within strong and weak enhancers suggest that enhancer strength is modulated by TF binding. Future studies that further integrating TF binding datasets with functional data of enhancers will shed light on the molecular mechanisms that drives enhancers' effect sizes on gene regulation.

Acknowledgements

This work was supported by the National Institutes of Health grants UM1HG009402 (to Y.S.).

Data availability

The CRISPRpath screen datasets used in this study are available at the ENCODE portal (www.encodeproject.org, accession number: ENCSR617AZY (sgRNA plasmid library), ENCSR427OPP (CRISPRi control), ENCSR900AXT, ENCSR254SJU (CRISPRi 2X), ENCSR793DSE (CRISPRi 3X), ENCSR250ZWC (CRISPRn control), ENCSR117YGQ (CRISPRn 1X), ENCSR071ZGB (CRISPRn 2X), ENCSR482PHH (CRISPRn 3X)). WTC11 iPSCs H3K4me3 PLAC-seq datasets are available at the 4DN data portal (data.4dnucleome.org, accession number: 4DNESIZ5TTHO and 4DNESDRL4ZKM). ATAC-seq, ChIP-seq, CUT&Tag, and RNA-seq datasets in WTC11 iPSCs are available at the Gene Expression Omnibus under the accession number GSE166839.

Data can be visualized on the WashU Epigenome Browser using the session at the following link (https://epigenomegateway.wustl.edu/browser/?genome=hg38&sessionFile=https://shen-xren.s3-us-west-1.amazonaws.com/CRISPRpath/eg-session-QRXJ0218-4d710b60-6ea7-11eb-8d8d-03c7189570c0.json). Tracks include ATAC-seq, H3K27ac, H3K4me3, and CTCF signals, and the identified enhancers from CRISPRi 2X and 3X screens. RefGene 38 genes are also displayed.

The plasmids generated in this study are available from Addgene (#167186, #167188, # 167189).

Code availability

The computer code used for analyzing CRISPRpath datasets is available at https://github.com/MichaelMW/crispy. A copy of the custom code used for data analysis and figure generation in this study is available upon request.

Competing interests

B.R. is co-founder and shareholder of Arima Genomics and Epigenome Technologies. The other authors declare that they have no competing interests.

Author contributions

X.R. and Y.S. designed the study. X.R. and B.L. designed the sgRNA library under the supervision of Y.S. and B.R. X.R., K.J., I.R.J., M.A.T., J.L., L.M., and T.W.T. performed the experiments. M.W. designed the CRISPY under the supervision of W.W. X.R., B.L., L.Z., G.L., and Y.L. performed data analysis. M.Y. and R.H. constructed the H3K4me3 PLAC-seq libraries. L.L., A.A. and M.H. analyzed PLAC-seq and HiChIP data. X.R. and Y.S. prepared the manuscript with input from all other authors.

335 **Figure 1. CRISPRpath for identifying enhancers of multiple genes.**
336 (**a**) Six genes (*HPRT1*, *MSH2*, *MSH6*, *MLH1*, *PMS2* and *PCNA*) in the 6TG-induced mismatch
337 repair process were used for CRISPRpath screen in this study. (**b**) Schematic of the
338 CRISPRpath screening strategy with 6TG treatment in iPSCs. Cell survival was used as readout
339 for the screen. (**c**) Spearman correlation analysis of sgRNA ranking based on fold change for
340 CRISPRpath screens with different 6TG concentrations (1X, 2X and 3X). (**d**) Venn diagram
341 shows the overlapping enriched sgRNAs identified from the screens with 2X and 3X 6TG
342 treatments. (**e**) Box plots show the fold change of the enriched distal and proximal sgRNAs from
343 2X, 3X CRISPRi and CRISPRn screens. Star indicates no enriched distal sgRNA was identified
344 from 3X CRISPRn screen. Boxplots indicate the median, interquartile range (IQR), Q1 − 1.5 ×
345 IQR and Q3 + 1.5 × IQR. (**f**) Bar plot shows the number of enriched distal and proximal sgRNAs
346 from 2X, 3X CRISPRi and CRISPRn screens. Star indicates no enriched distal sgRNA was
347 identified from 3X CRISPRn screen. (**g**) Venn diagram shows the identified enhancers from
348 each CRISPRpath screen.
349
350 **Figure 2. Genomic features of identified enhancers from CRISPRpath using CRISPRi.**
351 (**a**) Genomic locations of identified enhancers relative to TSS. Circles indicate enhancers
352 identified from the CRISPRi 3X screen (red), enhancers uniquely identified from the CRISPRi
353 2X screen (blue), tested CREs that are not identified as enhancers (grey). Purple lines label the
354 location of each target gene. (**b**) Histogram shows the distance distribution between identified
355 enhancers and their paired TSS. (**c**) Histogram shows the number of interval genes between
356 enhancers and their target gene TSS. Mean is indicated with an orange dashed line and only
357 the enhancers for *MLH1*, *PMS2*, *PCNA*, *HPRT1* are included in **b** and **c**. (**d**) A weak negative
358 correlation is observed between enrichment score and genomic distance between enhancers
359 and their target genes (Pearson correlation, r = -0.36, P = 0.01). Black circles indicate
360 promoters. The red and blue circles are enhancers showing in **a**. (**e**) Density plot shows no
361 significant difference (two-tailed two-sample Kolmogorov–Smirnov test) for the distribution of all
362 distal sgRNAs (gray), enriched distal sgRNAs from 2X (blue) and 3X screen (red) CRISPRi
363 screens.
364
365 **Figure 3. Enhancer-like promoters act as functional enhancers.**

366 (**a**, **b**, **c**) Three examples of promoters function as enhancers. CRISPRi silencing of the
367 promoter region of *SOCS5*, *FOXN2* and *TMEM230* results in significant downregulation of
368 *MSH6*, *MSH2* and *PCNA*, respectively. shRNA knockdown of *SOCS5*, *FOXN2* and *TMEM230*
369 can only downregulate *SOCS5*, *FOXN2* and *TMEM230* expression. Three independent
370 replicates per condition and two independent sgRNAs or shRNAs per replicate were used for
371 each experiment. *P* values are from two-tailed two-sample t-test. (**d**) Average signal enrichment
372 of ATAC-seq, gene transcription, H3K4me3, H3K27ac and CTCF binding for enhancer-like
373 promoters (n = 38) and control promoters (n = 47). *P* values are from Wilcoxon test. Boxplots
374 indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. (**e**) Number of H3K4me3
375 mediated chromatin interactions and cumulative interaction score for enhancer-like promoters
376 (n = 31) and control promoters (n = 43). Boxplots indicate median, IQR, Q1 − 1.5 × IQR and Q3
377 + 1.5 × IQR. *P* values are calculated from Wilcoxon test.
378
379 **Figure 4. CRISPRpath can distinguish weak and strong enhancers by imposing different**
380 **selection pressures.**
381 (**a**) Box plots show the enrichment score of the tested elements. TSS regions (black circles)
382 show highest enrichment scores. Enhancers uniquely identified from the lower selection
383 pressure (CRISPRi 2X, blue circles) exhibit lower enrichment scores compared to the
384 enhancers identified from the higher selection pressure (CRISPRi 3X, red circles). *P* values are
385 from Wilcoxon test. (**b**) Box plots show the CRISPRi perturbation at enhancers induced various
386 degrees of transcriptional repression of target genes measured with RT-qPCR. Each dot
387 represents the average value from three biological replicates. CRISPRi targeting TSS regions
388 (dark gray) achieved the highest transcriptional repression. CRISPRi targeting strong
389 enhancers (pink) leads a more substantial transcription silencing on target gene compared to
390 CRISPRi targeting weak enhancers (cyan). *P* values are from Wilcoxon test. (**c**) Enrichment
391 analysis of ATAC-seq, H3K27ac, H3K4me3, CTCF binding signals for strong (n = 33) and weak
392 (n = 30) enhancers. Boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. *P*
393 values for the difference between strong and weak enhancers are from Wilcoxon test; see
394 Supplementary Table 7 for *P* values of all pairwise comparisons. (**d**) Intersection of genomic
395 features for weak enhancers (blue bar) and strong enhancers (red bar). (**e**) Distance normalized
396 H3K4me3 PLAC-seq contact frequency for strong (n = 23) and weak (n = 21) enhancers. Only

397 the enhancers for *MLH1*, *PMS2*, *PCNA*, *HPRT1* are included (see Methods). Boxplots indicate
398 the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. *P* value is from Wilcoxon test. (**f**) Heatmap
399 shows the frequency of transcription factor motifs found in strong and weak enhancers.
400
401 **Supplementary Figure 1. Features of sgRNA library for CRISPRpath screen.**
402 (**a**) Bar graph shows the number of distal ATAC-seq peaks used as candidate CREs for six
403 target genes. (**b**) Histogram shows size distribution of distal ATAC-seq peaks. The average size
404 is 488 bp (blue dash line). (**c**, **e**) The composition of the sgRNA library. In total, 35,763 sgRNAs
405 were included in the library (**c**), and 12,702 sgRNAs are high quality sgRNAs (**e**). (**d**, **f**)
406 Distribution of the number of sgRNAs per distal ATAC-seq peak. Average numbers of sgRNA
407 per ATAC-seq peak are indicated with blue dash lines.
408
409 **Supplementary Figure 2. Quality of the sgRNA library and CRISPRpath screen libraries.**
410 (**a**) Distribution of sgRNA oligo read counts in the sgRNA library. (**b**) Cumulative frequency of
411 sgRNAs in the sgRNA library. (**c**) Distribution of high quality sgRNAs read counts in the sgRNA
412 library. (**d**) Cumulative frequency of high quality sgRNAs in the sgRNA library. The constructed
413 sgRNA plasmid library recorvered all the designed sgRNAs with the copy number difference
414 less than five fold for at least 97% designed sgRNAs. (**e**) PCA analysis shows the high
415 reproducibility of the CRISPRpath screen libraries between biological replicates.
416
417 **Supplementary Figure 3. *P* value cutoff used for identifying enriched sgRNAs from each**
418 **screens.**
419 (**a-f**) Distribution of *P* value for tested distal, proximal and non-targeting control sgRNA groups.
420 Orange dash lines indicate 5% percentile of the *P* values from non-targeting control sgRNAs to
421 achieve a false discovery rate of 5%.
422
423 **Supplementary Figure 4. Enriched proximal sgRNAs and sgRNA ranking analysis.**
424 (**a**) Number and fold change of the enriched proximal sgRNAs for the six target genes from
425 CRISPRi and CRISPRn screens. The color indicates fold changes, and the size of circle
426 indicates the number of enriched sgRNAs. (**b**) Enrichment analysis shows the enriched proximal
427 sgRNAs bias towards to the TSS region for CRISPRi screens and the protein coding region

428 (CDS) for CRISPRn screens. Color represents the number of enriched sgRNAs. (**c**) Spearman
429 correlation analysis of the distal and proximal sgRNAs ranking shows proximal sgRNAs
430 exhibiting higher correlation between each screen compared to distal sgRNAs.
431
432 **Supplementary Figure 5. Enriched sgRNAs identified from CRISPRi screens exhibit no**
433 **position and strand preference.**
434 (**a**) Enriched sgRNAs from CRISPRi 2X (red dots, n=448) and CRISPRi 3X (red dots, n=260)
435 screens showed similar distributions across candidate CREs. (**b**) Odds ratio analysis of the fold
436 change of enriched sgRNAs shows enriched sgRNAs have no strand preference. Enhancers
437 with enriched sgRNAs only targeting one strand were exclude for the analysis. Odds ratio was
438 calculated for each element with the equation of ave($\log_2$(fold change of sgRNA targeting plus
439 strand)) / ave($\log_2$(fold change of sgRNA targeting minus strand)). Violin plots show the
440 distributions of odds ratio values within each screen, and boxplots indicate the median, IQR,
441 Q1 − 1.5 × IQR and Q3 + 1.5 × IQR.
442
443 **Supplementary Figure 6. Validation of CIRSPRpath identified enhancers.**
444 (**a**) Validation of the strong (black) and weak (grey) enhancers with CRISPRi followed by RT-
445 qPCR. Three independent replicates per condition. The significance was calculated with two-
446 tailed two-sample *t*-test. Data are mean and s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. (**b**)
447 CRISPRi-mediated transcriptional repression of six target genes by targeting TSS of each gene.
448 Three independent replicates per condition. The significance was calculated with two-tailed two-
449 sample *t*-test. Data are mean and s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. (**c**) Pearson
450 correlation analysis reveals element enrichment score from CRISPRpath screens correlates
451 with element effect size on transcription from CRISPRi (Pearson correlation, PCC = -0.68, P =
452 3.2×10^{-5}).
453
454 **Supplementary Figure 7. Chromatin contract frequency analysis for the enhancers in**
455 **K562 cells and mESCs.**
456 (**a**) Distance normalized H3K27ac HiChIP contact frequency for strong (n = 34) and weak (n =
457 82) enhancers identified with crisprQTL mapping in K562 cells. (**b**) Distance normalized
458 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 Table 1. ATAC-seq peak regions used to design the sgRNA library.

Supplementary Table 2. List of sgRNA sequences used for CRISPRpath screen.

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

Supplementary Table 4. List of primers used for RT-qPCR.

Supplementary Table 5. List of shRNA sequences used for RNA interference experiments.

Supplementary Table 6. List of sgRNA sequences used for CRISPRi-mediated enhancer validation experiments.

Supplementary Table 7. Pairwise comparisons of data in Figure 4c.

482 **Methods**

483

484 **Cell culture**

485 Doxycycline inducible CRISPRi and CRISPRn WTC11 iPSC lines were purchased from
486 Gladstone Stem Cell Core. Both CRISPRi and CRISPRn WTC11 iPS cells were cultured on
487 Matrigel-coated (Corning, 354277) plates with Essential 8™ Medium (Life Technologies,
488 A1517001). iPSCs were passaged using Accutase (STEMCELL Technologies, 07922) and 10
489 μM ROCK inhibitor Y-27632 (STEMCELL Technologies, 72302). HEK293T cells were cultured
490 in Dulbecco's modified Eagle's medium (Gibco, 11995065) with 10% fetal bovine serum
491 (CPSSerum, FBS-500). HEK293T cells were passaged with Trypsin-EDTA (Gibco, 25200072).
492 All the cells were grown with 5% CO_2 at 37°C and verified mycoplasma free using the MycoAlert
493 Mycoplasma Detection Kit (Lonza, LT07-218).

494

495 **sgRNA library design**

496 CRISPRpath sgRNA library was designed to screen cis-regulatory elements for *HPRT1*, *MSH2*,
497 *MSH6*, *MLH1*, *PMS2* and *PCNA*. ATAC-seq peaks within the region of 1 Mb upstream and 1
498 Mb downstream of each target gene including TSS and coding regions were selected as
499 targeting regions for the sgRNA library design (**Supplementary Table 1**). We generated a
500 genome-wide sgRNA database containing all the available unique sgRNAs, each followed by a
501 'NGG' PAM sequence. All the designed unique sgRNAs in the target regions were added in the
502 sgRNA library, excluding sgRNAs containing AATAAA, AAAAA, TTTTT or TTTTTT sequences.
503 Unique 20-bp sequences in the target regions that were not followed by the 'NGG' or 'NAG'
504 PAM sequences were taken as non-targeting control sgRNAs, excluding non-targeting sgRNAs
505 containing TTT, TTNTT, AATAAA, AAAAA, TTTTT or TTTTTT sequences. Then, a guanine
506 nucleotide was added to all the sgRNAs if the sequence did not start with G to increase
507 efficiency of transcription from U6 promoter. Final sgRNA oligos adhered to the following
508 template: 5'-ATATCTTGTGGAAAGGACGAAACACC-[20- or 21-bp sgRNA sequence]-
509 GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC-3'. In total, 35,763 sgRNAs were
510 included in the library (**Supplementary Fig. 1 and Supplementary Table 2**). We retrieved
511 specificity score and off-target site for each sgRNA from GuideScan[18] (www.guidescan.com)
512 and assigned the specificity score of sgRNAs not existed in the GuideScan database to 0. The

513 high-quality sgRNAs were filtered with specificity score >0.2 and without perfectly matched or
514 1-2 mismatches off-target sites.
515
516 **Oligo synthesis and library cloning**
517 sgRNA library oligos were synthesized by TWIST BIOSCIENCE and amplified with the forward
518 primer 5'- TCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAC-3' and the
519 reverse primer
520 5'-AACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC-3'. We replaced the
521 Cas9 sequence in lentiCRISPR v2 plasmid (Addgene, 52961) with blasticidin S deaminase
522 sequence to construct the lentiCRISPR-v2-Blast-Puro plasmid (Addgene, 167186). The PCR
523 products were purified via gel excision and column purification (Promega, A9282), and then
524 inserted into the BsmBI-digested lentiCRISPR-v2-Blast-Puro vector by Gibson assembly (New
525 England Biolabs, E2621L). The assembled products were transformed into NEB 5-alpha
526 electrocompetent *E. coli* cells (New England Biolabs, C2989K) by electroporation. About 40
527 million independent bacterial colonies were cultured, and sgRNA library plasmids were
528 extracted with the Qiagen EndoFree Plasmid Mega Kit (Qiagen, 12381). The recovery rate and
529 distribution of the sgRNA library were checked with next generation sequencing
530 (**Supplementary Fig. 2a-d**).
531
532 **Lentivirus production and titration**
533 To make the lentiviral library, 5 μg of sgRNA plasmid library was co-transfected with 3 μg of
534 psPAX (Addgene, 12260) and 1 μg of pMD2.G (Addgene, 12259) lentivirus packaging plasmids
535 into 8 million HEK293T cells in a 10-cm dish with PolyJet (SignaGen Laboratories, SL100688).
536 For each individual sgRNA, 3.75 μg of sgRNA plasmid was co-transfected with 2.25 μg of
537 psPAX (Addgene, 12260) and 0.75 μg of pMD2.G (Addgene, 12259) plasmids into 4 million
538 HEK293T cells in a T25 flask with PolyJet (SignaGen Laboratories, SL100688). Media was
539 replaced 12 h after transfection, and harvested every 24 h for a total of three harvests.
540 Harvested media containing the desired virus were filtered through Millex-HV 0.45-μm PVDF
541 filters (Millipore, SLHV033RS) and further concentrated with 100,000 NMWL Amicon Ultra-15
542 centrifugal filter units (Amicon, UFC910008).
543

544 The titer of lentivirus was determined by transducing 500,000 cells with varying amount (0, 0.5,
545 1.0, 2.0, 4.0 and 8.0 μL) of concentrated virus and polybrene (Millipore, TR-1003-G, 8 μg/mL,).
546 Viral transduction was performed by centrifuging the lentivirus and cell combination at 1000
547 RCF for 90 min at 37°C. 3 to 4 h later, virus containing media was replaced with fresh media.
548 24 h after the transduction, transduced cells were dissociated with Accutase, and seeded as
549 duplicates. One replicate was treated with blasticidin (Gibco, A1113903, 4 μg/mL), and the other
550 replicate was not treated with blasticidin. Four days later, the blasticidin resistant cells and
551 control cells were counted to calculate the ratio of infected cells and the viral titer.
552
553 **Determining 6TG concentration via killing curve titration**
554 Both CRISPRi and CRISPRn WTC11 iPSCs were used to determine the minimal lethal
555 concentration of 6TG. Cells were seeded in 24 well plates. When the cells reached around 50%
556 confluence (Day 0), they were treated with 6TG concentrations of 0 (control), 20, 40, 60, 80,
557 100, 120, 140, and 160 ng/mL. Two wells were allocated for each condition. The cells were
558 examined daily and cultured for 7 days. The media was replaced daily with the specified 6TG
559 concentration. After 3 days, wells with 6TG concentration greater than or equal to 100 ng/mL
560 had no surviving cells. On Day 4 of treatment, the wells with 80 ng/mL 6TG treatment had no
561 surviving cells. On the last day of treatment, the wells with 40 and 60 ng/mL treatments had
562 very few surviving cells, while the 20 ng/mL treatment had many surviving cells. Based on these
563 results, we set 80 ng/mL as the minimal lethal concentration for 6TG.
564
565 **CRISPRpath screening and sequencing library preparation**
566 CRISPRpath screens were carried out with 72 million doxycycline inducible CRISPRi or
567 CRISPRn iPSCs in biological replicate. The cells for lentiviral transduction were seeded into 6
568 well plates with 1 million cells per well, and the lentiviral library (MOI = 0.5) was transduced into
569 the iPSCs with 8 μg/mL Polybrene (Millipore, TR-1003-G) and spun at 1000 RCF at 37°C for
570 90 min. The transduced cells were treated with doxycycline (Sigma, D9891, 2 μM) and
571 blasticidin (Gibco, A1113903, 4 μg/mL) for 4 days. After this doxycycline and blasticidin
572 treatment, 10 million cells were reserved as a control population, and 100 million cells were
573 used for CRISPRpath screen with doxycycline and 6TG (Sigma, A4660) treatment for 7 days.
574 Finally, survival cells were collected from the 6TG treated population.

575

576 The genomic DNA was extracted from each sample via cell lysis and digestion (100 mM Tris-
577 HCl pH8.5, 5 mM EDTA, 200 mM NaCl, 0.2% SDS, 100 μg/mL proteinase k), phenol:chloroform
578 (Thermo Scientific, 17908) extraction and isopropanol (Fisher Scientific, BP2618500)
579 precipitation. To amplify the sgRNA sequences from each sample, thirty-two 50 μl PCR
580 reactions were performed using 500 ng genomic DNA for each reaction and NEBNext® High-
581 Fidelity 2X PCR Master Mix (New England Biolabs, M0541S). The purified libraries were
582 sequenced on the NovaSeq 6000 with 150-bp paired-end sequencing. The detail protocol is
583 available on ENCODE portal (https://www.encodeproject.org/documents/2e6451a9-3b98-
584 4d95-922e-a3d8d2100ddf/).
585

586 **CRISPRpath data analysis**

587 The sequence files were down-sampled to the same amount of total reads, and then mapped
588 to the sgRNA library with the requirement of exact match of designed sgRNA sequences in the
589 following patten 5'-CCG-[N19 or N20]-GTT-3'. Only the highly specific sgRNAs (specificity score
590 >0.2, without perfectly matched or 1-2 mismatches off-target sites) were used for downstream
591 data analysis. The sgRNA enrichment for each screen was calculated by comparing 6TG
592 treated samples with the associated control samples with edgeR and TMM normalization. We
593 first used edgeR[37] to calculate *P* value based on negative binomial model for both targeting
594 sgRNAs and non-targeting control sgRNAs. To achieve empirical false discovery rate less than
595 5%, we then selected a *P* value cutoff corresponding to the 5% percentile of *P* values from non-
596 targeting control sgRNAs. Finally, we defined enriched sgRNAs with *P* value less than the
597 selected *P* value cutoff, and fold change >2. The ATAC-seq peaks were identified as functional
598 enhancers for the six MMR genes by having at least 3 significant enriched sgRNAs. Analysis
599 scripts are available at https://github.com/MichaelMW/crispy.
600

601 **Analysis of genomic feature and chromatin signature of identified enhancers**

602 Genomic distances between enhancer and TSS pairs were calculated based on the distance
603 from the center of enhancers to the transcription start sites of the target genes. The number of
604 interval genes is the number of all the RefSeq annotated genes between each enhancer and
605 paired target gene. The signal of chromatin signatures, including ATAC-seq, H3K27ac,

606 H3K4me3, CTCF binding and RNA-seq, were calculated by deeptools (v3.4.3)[38]. The enhancer-
607 like promoters are the enhancers overlap with the region 500 bp upstream and downstream of
608 a RefSeq annotated TSS.
609
610 **Validation of identified enhancers using CRISPRi**
611 We cloned lentiCRISPR-v2-HygR-EGFP (Addgene, 167188) and lentiCRISPR-v2-HygR-
612 mCherry (Addgene, 167189) vectors by replacing the Cas9 and puromycin N-acetyltransferase
613 sequences in lentiCRISPR v2 plasmid (Addgene, 52961) with hygromycin B
614 phosphotransferase and EGFP or mCherry sequences. To validate the identified enhancers,
615 individual sgRNAs targeting identified enhancers were cloned into the lentiCRISPR-v2-HygR-
616 GFP or lentiCRISPR-v2-HygR-mCherry vector. The doxycycline inducible CRISPRi WTC11
617 iPSCs were infected with the lentivirus expressing sgRNAs for three replicates per sgRNA. The
618 sgRNA infected cells were grown with hygromycin (Gibco, 10687010, 150 μg/mL) and
619 doxycycline (Sigma, D9891, 2 μM) containing media. Seven days later, the cells were collected
620 and total RNA was extracted from the cells using the Qiagen RNeasy® Plus Kit (Qiagen,
621 74134). One μg of RNA was then used to synthesize cDNA using the Bio-RAD iScript cDNA
622 Synthesis Kit (Bio-RAD, 1708840). qPCR reactions for targeted genes were performed with the
623 Luminaris HiGreen qPCR Master Mix (Thermo Scientific, K0993) on the Roche LightCycler 96
624 System. The qPCR primers are listed in **Supplementary Table 4** and the sgRNA sequences
625 are listed in **Supplementary Table 6**. For each tested element in **Fig. 3a-c** and **Supplementary**
626 **Fig. 6b**, we performed CRISPRi experiments with two independent sgRNAs and used the
627 results from the sgRNA with stronger transcriptional repression in **Fig. 4b**.
628
629 **shRNA mediated RNA interference**
630 shRNAs were designed by using DSIR tool (http://biodev.extra.cea.fr/DSIR/DSIR.html)
631 targeting *SOCS5*, *FOXN2* and *TMEM230*. The sequences of shRNAs are listed in
632 **Supplementary Table 5**. The shRNAs were cloned into lentiCRISPR-v2-HygR-mCherry vector
633 under the control of human U6 promoter and packaged into lentivirus for cell transduction. The
634 WTC11 iPSCs transduced with shRNA lentivirus were treated with hygromycin (Gibco,
635 10687010, 150 μg/mL) for 7 days and then collected for RNA extraction and qPCR.
636

637 **ATAC-seq**

638 ATAC-seq was carried out using the Nextera DNA Library Prep Kit (Illumina, FC-121-1030) as
639 previously described[39]. The detailed protocol is available on the ENCODE portal
640 (https://www.encodeproject.org/documents/0317894c-5a42-4f03-b865-c2a2d08708ef/).
641 Briefly, each library started with 100,000 fresh iPSCs, and the cells were incubated with ice cold
642 nuclei extraction buffer (10 mM TrisHCl pH 7.5, 10 mM NaCl, 3 mM MgCl2, 0.1% Igepal CA630,
643 and 1x protease inhibitor) for 5 min on ice, then centrifuged at 500 RCF for 5 min. 50,000
644 resulting nuclei were treated with tagmentation buffer (25 µL Buffer TD with 50,000 nuclei, 22.5
645 µL water, 2.5 µL TDE1) for 30 min at 37°C. The transposed DNA was purified using Qiagen
646 MinElute PCR purification kit (Qiagen, 28006), amplified using Nextera primers, then size-
647 selected for fragments between 150 and 1000 bp using SPRISelect beads (Beckman Coulter,
648 B23319). Libraries were sent for single-end sequencing on the HiSeq 4000 (50 bp single-end
649 reads). Reads were mapped to hg38/GRCh38 and processed using the ENCODE pipeline
650 (https://github.com/kundajelab/atac_dnase_pipelines, V1.8.0), which ran on the default
651 settings. The ATAC-seq peaks were filtered with FDR cutoff of 0.1%, and adjacent peaks were
652 merged if they are less than 1 kb apart.

653

654 **RNA-seq**

655 RNA was extracted from fresh cells using the RNeasy Plus Mini Kit (Qiagen, 74134).
656 Approximately 1000 ng of extracted RNA was used to prepare libraries for sequencing using
657 the TruSeq Stranded mRNA Library Prep Kit (Illumina, 20020594). Libraries were sent for
658 paired-end sequencing on the NovaSeq 6000 (100 bp paired-end reads). Reads were aligned
659 to hg38/GRCh38 using STAR 2.7.0f[40] with the standard ENCODE settings, and transcript
660 quantification was performed in a strand-specific manner using RSEM 1.3.1[41] with the
661 annotation from GENCODE v32. Only the first read was used, and all reads were trimmed to
662 51bp using TrimGalore 0.4.5 running the following options: -q 20 --length 20 -- stringency 3 --
663 trim-n. The edgeR package in R (3.20.9)[37] was used to calculate TMM-normalized FPKM values
664 for each gene based on the expected counts and gene lengths for each library. The mean gene
665 expression across all replicates was used for analysis.

666

667 **ChIP-seq**

668 ChIP-seq libraries were constructed from 2 million WTC11 iPSCs. Cells were crosslinked in 1%
669 formaldehyde at room temperature for 20 min and then quenched with 2.5 M glycine at room
670 temperature for 5 min. Fixed cells were lysed and chromatin was sonicated by Covaris with the
671 following parameters: cycle 2, 105 W, 200 cycles per burst time for 30 min. Input chromatin was
672 removed and stored at -20°C for later processing. Magnetic beads (Invitrogen, Dynabeads
673 Protein A, 10001D) were preincubated with H3K27ac antibody (Active Motif, 39133, Lot
674 22618011) for 2 hours at 4°C before being added to sheared chromatin. Samples were
675 incubated overnight at 4°C. Beads were washed 3 times and chromatin was then eluted.
676 Samples were incubated at 65°C overnight to reverse the crosslinking. DNA was treated with
677 RNase A for 1 hr at 37°C and Proteinase K (New England Biolabs, 8107) for 1 hr at 55°C. DNA
678 was purified by phenol-chloroform extraction and ethanol precipitation. Libraries were prepared
679 using Tru-seq adapters and size-selected using SPRIselect beads prior to amplification and
680 paired-end sequencing. Libraries were sent for paired-end sequencing on the NovaSeq 6000
681 (150 bp paired-end reads). Sequencing reads were trimmed to 50 bp and mapped to hg38 using
682 bowtie2 with the following options: --local --very-sensitive-local --no-unal --no-mixed --no-
683 discordant --phred33 -I 10 -X 700. Picard Tools was used to remove blacklisted regions and
684 duplicate reads and MACS2 was used to call peaks on merged replicates at an FDR cutoff of
685 1%.
686
687 **CUT&Tag**
688 CUT&Tag libraries were constructed from 150,000 WTC11 iPSCs according to previously
689 described methods[42]. Cells were lysed in nuclei extraction buffer (20 mM HEPES-KOH pH 7.9,
690 10 mM MgCl2, 0.1% Triton X-100, 20% glycerol and 1x protease inhibitor) on ice for 10 min.
691 The samples were spun and resuspended in 100 µl nuclei extraction buffer. Meanwhile, 10 µl
692 of BioMag Plus Concanavalin A (Bangs Laboratories, BP531) were equilibrated in binding buffer
693 (1x PBS, 1 mM $CaCl_2$, 1 mM $MgCl_2$ and 1 mM $MnCl_2$). The equilibrated beads were added to
694 the samples and incubated with rotation for 15 min at 4°C. Nuclei-bound beads were washed
695 with Buffer 1 (20 mM HEPES-KOH pH 7.9, 150 mM NaCl, 2 mM EDTA, 0.5 mM spermidine,
696 0.1% BSA and 1x protease inhibitor) and Buffer 2 (20 mM HEPES-KOH pH 7.9, 150 mM NaCl,
697 0.5 mM spermidine, 0.1% BSA and 1x protease inhibitor). After washing nuclei-bound beads
698 were resuspended in 50 µl Buffer 2 with 0.5 µl antibody (H3K4me3 from Millipore, 04-745, Lot

699 3543820 and CTCF from Millipore, 07-729, Lot 3059608) and incubated with rotation overnight
700 at 4°C. Samples were washed twice with Buffer 2 and resuspended in 50 µl Buffer 2 with
701 antibody (antibodies-online Inc., Guinea Pig anti-Rabbit IgG, ABIN101961, Lot 42323) and
702 incubated for 1 hr at room temperature with rotation. Samples were washed again with Buffer 2
703 and resuspended in 100 µl Buffer 3 (20 mM HEPES-KOH pH 7.9, 300 mM NaCl, 0.5 mM
704 Spermidine, 0.1% BSA and 1x proteinase inhibitor) containing 0.04 µM pA-Tn5. Samples were
705 incubated for 1 hr at room temperature, washed three times with Buffer 3 and resuspended in
706 tagmentation buffer (20 mM HEPES-KOH pH 7.9, 300 mM NaCl, 0.5 mM Spermidine, 10 mM
707 $MgCl_2$, 0.1% BSA and 1x proteinase inhibitor). Samples were incubated for 1 hr at 37°C.
708 Samples were treated with Proteinase K (New England Biolabs, 8107) for 1 hr at 50°C. DNA
709 was purified by phenol-chloroform extraction and ethanol precipitation. Libraries were prepared
710 using Tru-seq adapters and size-selected using SPRIselect beads prior to amplification and
711 paired-end sequencing. Libraries were sent for paired-end sequencing on the Mini-seq (37 bp
712 paired-end reads, H3K4me3 libraries) or NovaSeq 6000 (150 bp paired-end reads, CTCF
713 libraries). Sequencing reads (CTCF libraries were trimmed to 50 bp) were mapped to hg38
714 using bowtie2 with the following options: --local --very-sensitive-local --no-unal --no-mixed --no-
715 discordant --phred33 -I 10 -X 700. Picard Tools was used to remove blacklisted regions and
716 duplicate reads and SEACR[43] was used to call peaks on merged replicates.
717
718 **H3K4me3 PLAC-seq**
719 H3K4me3 PLAC-seq data in WTC11 cells were generated as previously described[44] in
720 biological replicates (clone 6 and clone 28) (https://data.4dnucleome.org/experiment-set-
721 replicates/4DNESDRL4ZKM/ and https://data.4dnucleome.org/experiment-set-
722 replicates/4DNESIZ5TTHO/). We combined the two biological replicates, and applied the MAPS
723 pipeline[45] to identify significant long-range chromatin interactions at 5 kb bin resolution for the
724 genomic distance 10 kb ~ 1 Mb. The reference genome is GRCh38. In addition, for each 5 kb
725 bin pair anchored at H3K4me3 peaks, the MAPS pipeline outputs the normalized contact
726 frequency, which adjusts for the biases from effective fragment length, GC content, sequence
727 mappability, H3K4me3 enrichment level and 1D genomic distance effect.
728

729 **Comparison between strong enhancers and weak enhancers using H3K4me3 PLAC-seq**
730 **data**
731 For 4 genes *HPRT1*, *MLH1*, *PMS2* and *PCNA*, there are 23 enhancer-promoter pairs between
732 strong enhancers and their target genes, and 21 enhancer-promoter pairs between weak
733 enhancers and their target genes. We mapped each enhancer and promoter of target gene into
734 5 kb bins, and obtained the distance normalized H3K4me3 PLAC-seq contact frequency for 5
735 kb bin pairs containing the enhancer-promoter pairs. Since *MSH2* and *MSH6* are located within
736 407 kb linear genomic distance with each other and we can't assign enhancers to either gene
737 reliably, enhancers identified near *MSH2* and *MSH6* were excluded from this analysis.
738
739 **Comparison between enhancer-like promoters and control promoters using H3K4me3**
740 **PLAC-seq data**
741 For this analysis, control promoters are active promoter regions with annotated ATAC-seq
742 peaks and tested negative as enhancers for the for MMR genes. We mapped each promoter
743 into a 5 kb bin that was used in the PLAC-seq analysis. We only choose the bins with one
744 annotated active promoter, which gave us 31 enhancer-like promoters and 43 control promoters
745 in this analysis. We counted the number of significant H3K4me3 PLAC-seq interactions
746 anchored at the 5 kb bins with these promoter sequences. In addition, as described in our
747 previous study[23], for promoters with at least one significant interaction, we calculated the
748 summation of –log10 FDR of significant interactions, which is a measure of the overall
749 interaction strength.
750
751 **Chromatin contact frequency comparison between strong enhancers and weak**
752 **enhancers in K562 cells and mESCs**
753 For the chromatin contact frequency comparison of enhancers in K562 cells and mouse
754 embryonic stem cells (mESCs), we downloaded the identified enhancers from each publication[6,
755 8], and defined strong enhancer with cutoff of 50% ≤ transcriptional contribution ≤ 100%, weak
756 enhancer with cutoff of 0% < transcriptional contribution ≤ 20%. H3K27ac HiChIP data in K562
757 cells[46] and H3K4me3 PLAC-seq data in mESCs[45] were used for comparison. The comparisons
758 were performed in 10 kb resolution.
759

760 **Motif scan and Transcription factor identification**

761 The fasta files were first generated in the hg38 genome for the identified strong enhancers and
762 weak enhancers separately. For each strong enhancer and weak enhancer, the FIMO software
763 (version 5.1.0)[47] with human motif database HOCOMOCO (v11 FULL)[48] was used to scan the
764 motifs. All the FIMO motif scans were in default settings. We then filtered the transcription
765 factors (TFs) in each strong and weak enhancer loci by FDR cutoff of 0.05 and p-value cutoff of
766 0.0001 and gene expressions cutoff of FPKM > 1. By taking the TFs with TF motif appeared in
767 more than 80% enhancers, 47 TFs were considered as commonly appearing in the strong
768 enhancers, and 35 TFs were in the weak enhancers.

769

References

1. Soldner, F. et al. Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. *Nature* **533**, 95-99 (2016).
2. Visel, A., Rubin, E.M. & Pennacchio, L.A. Genomic views of distant-acting enhancers. *Nature* **461**, 199-205 (2009).
3. Long, H.K. et al. Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. *Cell Stem Cell* **27**, 765-783 e714 (2020).
4. Consortium, E.P. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020).
5. Sanjana, N.E. et al. High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545-1549 (2016).
6. Gasperini, M. et al. A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **176**, 377-390 e319 (2019).
7. Diao, Y. et al. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat Methods* **14**, 629-635 (2017).
8. Fulco, C.P. et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* **51**, 1664-1669 (2019).
9. Simeonov, D.R. et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111-115 (2017).
10. Klann, T.S. et al. CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol* **35**, 561-568 (2017).
11. Fulco, C.P. et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769-773 (2016).
12. Gasperini, M. et al. CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. *Am J Hum Genet* **101**, 192-205 (2017).
13. Pecina-Slaus, N., Kafka, A., Salamon, I. & Bukovac, A. Mismatch Repair Pathway, Genome Stability and Cancer. *Front Mol Biosci* **7**, 122 (2020).
14. Yan, T., Berry, S.E., Desai, A.B. & Kinsella, T.J. DNA mismatch repair (MMR) mediates 6-thioguanine genotoxicity by introducing single-strand breaks to signal a G2-M arrest in MMR-proficient RKO cells. *Clin Cancer Res* **9**, 2327-2334 (2003).
15. Li, G.M. Mechanisms and functions of DNA mismatch repair. *Cell Res* **18**, 85-98 (2008).
16. Mandegar, M.A. et al. CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell Stem Cell* **18**, 541-553 (2016).
17. Tycko, J. et al. Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nat Commun* **10**, 4063 (2019).
18. Perez, A.R. et al. GuideScan software for improved single and paired CRISPR guide RNA design. *Nat Biotechnol* **35**, 347-349 (2017).
19. Radzisheuskaya, A., Shlyueva, D., Muller, I. & Helin, K. Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. *Nucleic Acids Res* **44**, e141 (2016).
20. Rosenbluh, J. et al. Complementary information derived from CRISPR Cas9 mediated gene deletion and suppression. *Nat Commun* **8**, 15403 (2017).
21. Li, K. et al. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat Commun* **11**, 485 (2020).

22. Engreitz, J.M. et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452-455 (2016).
23. Song, M. et al. Cell-type-specific 3D epigenomes in the developing human cortex. *Nature* **587**, 644-649 (2020).
24. Kubo, N. et al. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat Struct Mol Biol* (2021).
25. Fernandez-Zapico, M.E. et al. A functional family-wide screening of SP/KLF proteins identifies a subset of suppressors of KRAS-mediated cell growth. *Biochem J* **435**, 529-537 (2011).
26. Ren, B. et al. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev* **16**, 245-256 (2002).
27. Polager, S., Kalma, Y., Berkovich, E. & Ginsberg, D. E2Fs up-regulate expression of genes involved in DNA replication, DNA repair and mitosis. *Oncogene* **21**, 437-446 (2002).
28. Canver, M.C. et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192-197 (2015).
29. Rajagopal, N. et al. High-throughput mapping of regulatory DNA. *Nat Biotechnol* **34**, 167-174 (2016).
30. Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* **32**, 267-273 (2014).
31. Adamson, B. et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882 e1821 (2016).
32. Diao, Y. et al. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res* **26**, 397-405 (2016).
33. Zabidi, M.A. et al. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556-559 (2015).
34. Dao, L.T.M. & Spicuglia, S. Transcriptional regulation by promoters with enhancer function. *Transcription* **9**, 307-314 (2018).
35. Mikhaylichenko, O. et al. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev* **32**, 42-57 (2018).
36. Thomas, H.F. et al. Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Mol Cell* (2021).
37. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
38. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).
39. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
40. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
41. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

42. Kaya-Okur, H.S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930 (2019).
43. Meers, M.P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).
44. Fang, R. et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res* **26**, 1345-1348 (2016).
45. Juric, I. et al. MAPS: Model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. *PLoS Comput Biol* **15**, e1006982 (2019).
46. Mumbach, M.R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet* **49**, 1602-1612 (2017).
47. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018 (2011).
48. Kulakovskiy, I.V. et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* **46**, D252-D259 (2018).

Figure 1

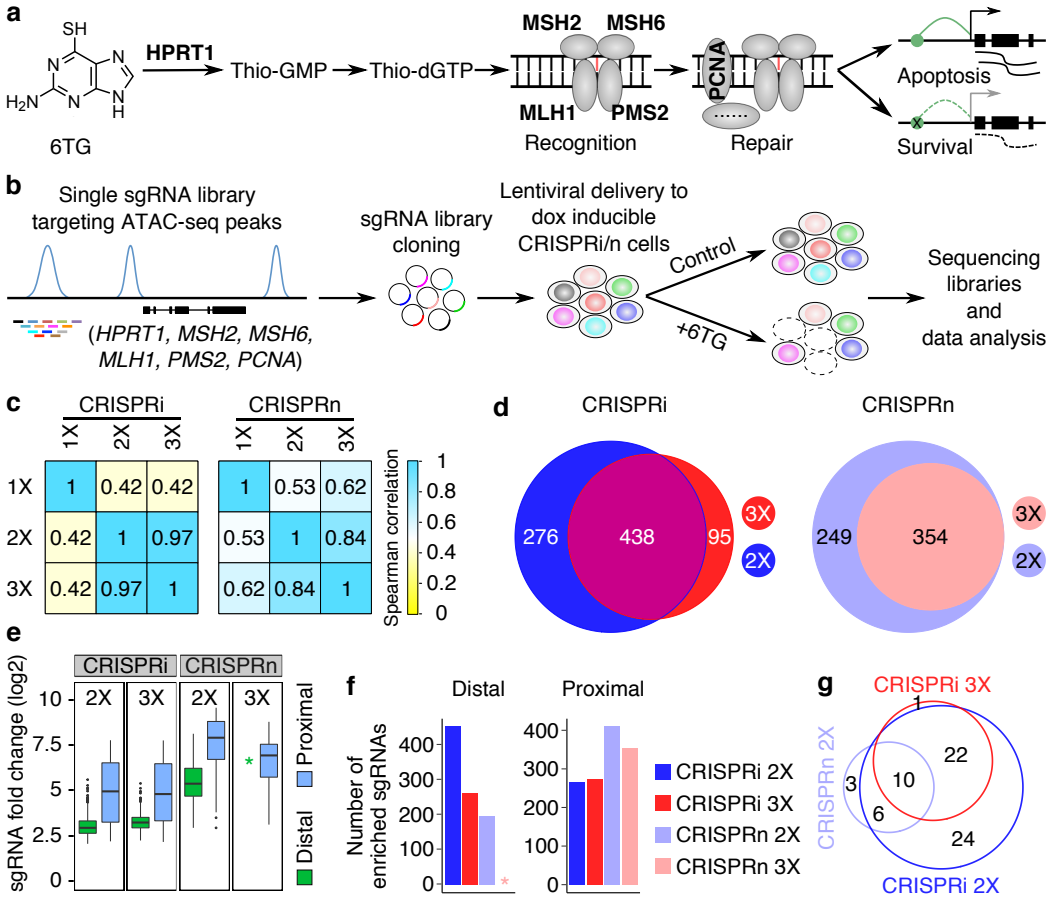


Figure 1. CRISPRpath for identifying enhancers of multiple genes.
(**a**) Six genes (*HPRT1*, *MSH2*, *MSH6*, *MLH1*, *PMS2* and *PCNA*) in the 6TG-induced mismatch repair process were used for CRISPRpath screen in this study. (**b**) Schematic of the CRISPRpath screening strategy with 6TG treatment in iPSCs. Cell survival was used as readout for the screen. (**c**) Spearman correlation analysis of sgRNA ranking based on fold change for CRISPRpath screens with different 6TG concentrations (1X, 2X and 3X). (**d**) Venn diagram shows the overlapping enriched sgRNAs identified from the screens with 2X and 3X 6TG treatments. (**e**) Box plots show the fold change of the enriched distal and proximal sgRNAs from 2X, 3X CRISPRi and CRISPRn screens. Star indicates no enriched distal sgRNA was identified from 3X CRISPRn screen. Boxplots indicate the median, interquartile range (IQR), Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. (**f**) Bar plot shows the number of enriched distal and proximal sgRNAs from 2X, 3X CRISPRi and CRISPRn screens. Star indicates no enriched distal sgRNA was identified from 3X CRISPRn screen. (**g**) Venn diagram shows the identified enhancers from each CRISPRpath screen.

Figure 2

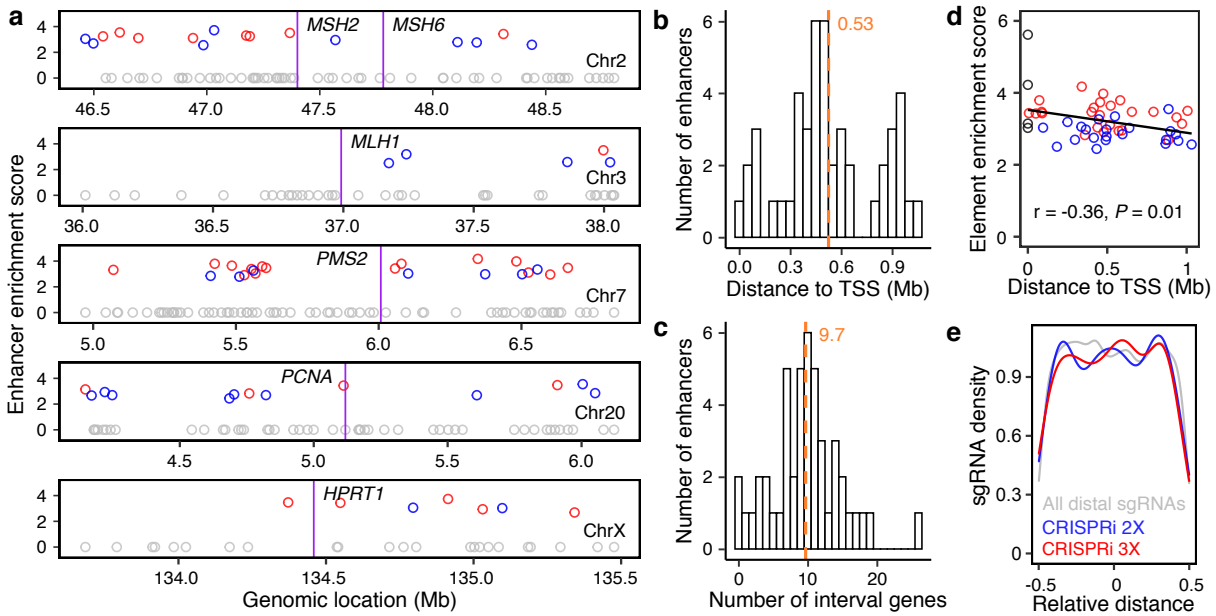


Figure 2. Genomic features of identified enhancers from CRISPRpath using CRISPRi.
(**a**) Genomic locations of identified enhancers relative to TSS. Circles indicate enhancers identified from the CRISPRi 3X screen (red), enhancers uniquely identified from the CRISPRi 2X screen (blue), tested CREs that are not identified as enhancers (grey). Purple lines label the location of each target gene. (b) Histogram shows the distance distribution between identified enhancers and their paired TSS. (**c**) Histogram shows the number of interval genes between enhancers and their target gene TSS. Mean is indicated with an orange dashed line and only the enhancers for *MLH1*, *PMS2*, *PCNA*, *HPRT1* are included in **b** and c. (**d**) A weak negative correlation is observed between enrichment score and genomic distance between enhancers and their target genes (Pearson correlation, r = -0.36, *P* = 0.01). Black circles indicate promoters. The red and blue circles are enhancers showing in **a**. (**e**) Density plot shows no significant difference (two-tailed two-sample Kolmogorov–Smirnov test) for the distribution of all distal sgRNAs (gray), enriched distal sgRNAs from 2X (blue) and 3X screen (red) CRISPRi screens.

Figure 3

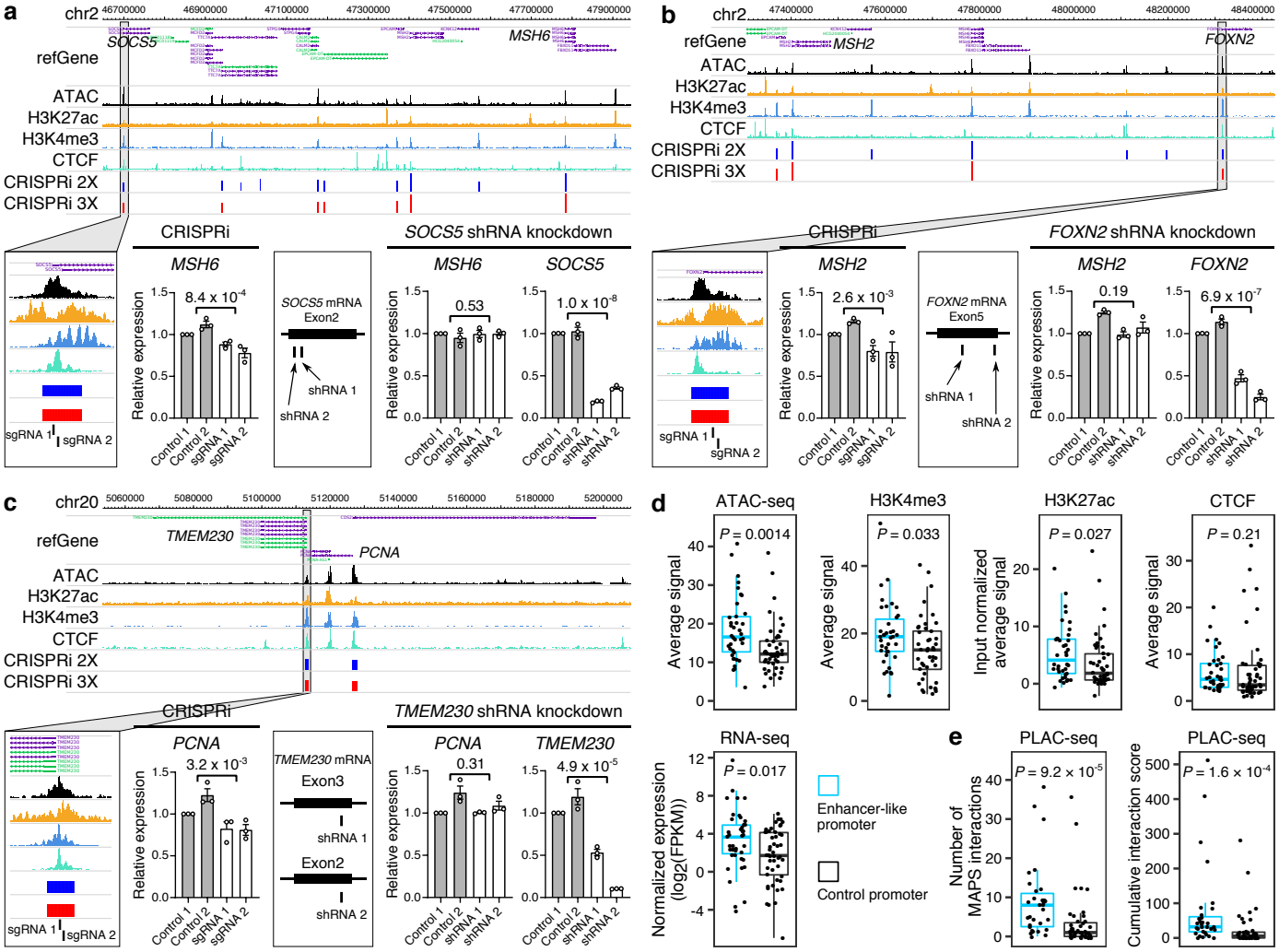


Figure 3. Enhancer-like promoters act as functional enhancers.
(**a, b, c**) Three examples of promoters function as enhancers. CRISPRi silencing of the promoter region of *SOCS5*, *FOXN2* and *TMEM230* results in significant downregulation of *MSH6*, *MSH2* and *PCNA*, respectively. shRNA knockdown of *SOCS5*, *FOXN2* and *TMEM230* can only downregulate *SOCS5*, *FOXN2* and *TMEM230* expression. Three independent replicates per condition and two independent sgRNAs or shRNAs per replicate were used for each experiment. *P* values are from two-tailed two-sample t-test. (**d**) Average signal enrichment of ATAC-seq, gene transcription, H3K4me3, H3K27ac and CTCF binding for enhancer-like promoters (n = 38) and control promoters (n = 47). *P* values are from Wilcoxon test. Boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. (**e**) Number of H3K4me3 mediated chromatin interactions and cumulative interaction score for enhancer-like promoters (n = 31) and control promoters (n = 43). Boxplots indicate median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. *P* values are calculated from Wilcoxon test.

Figure 4

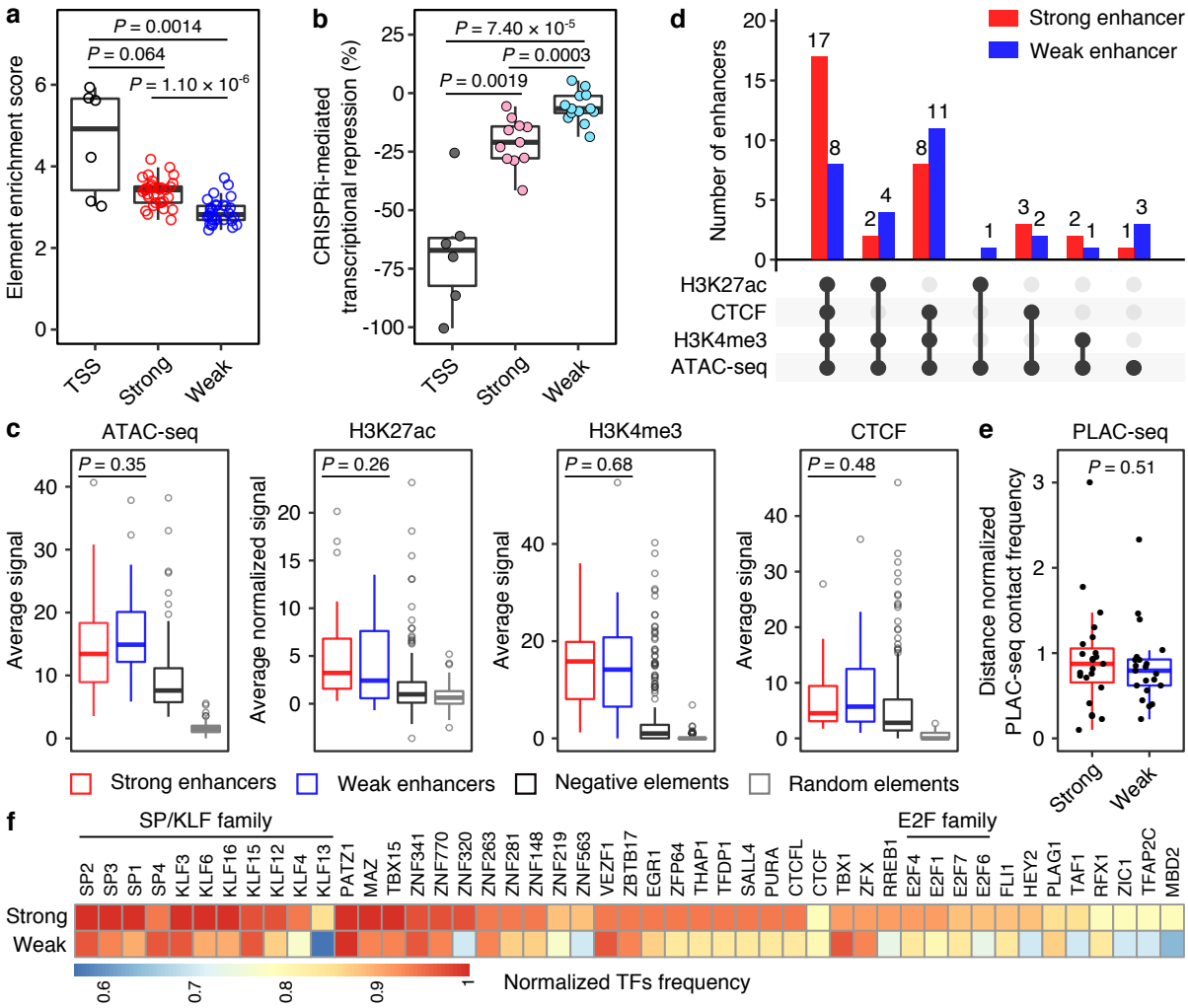
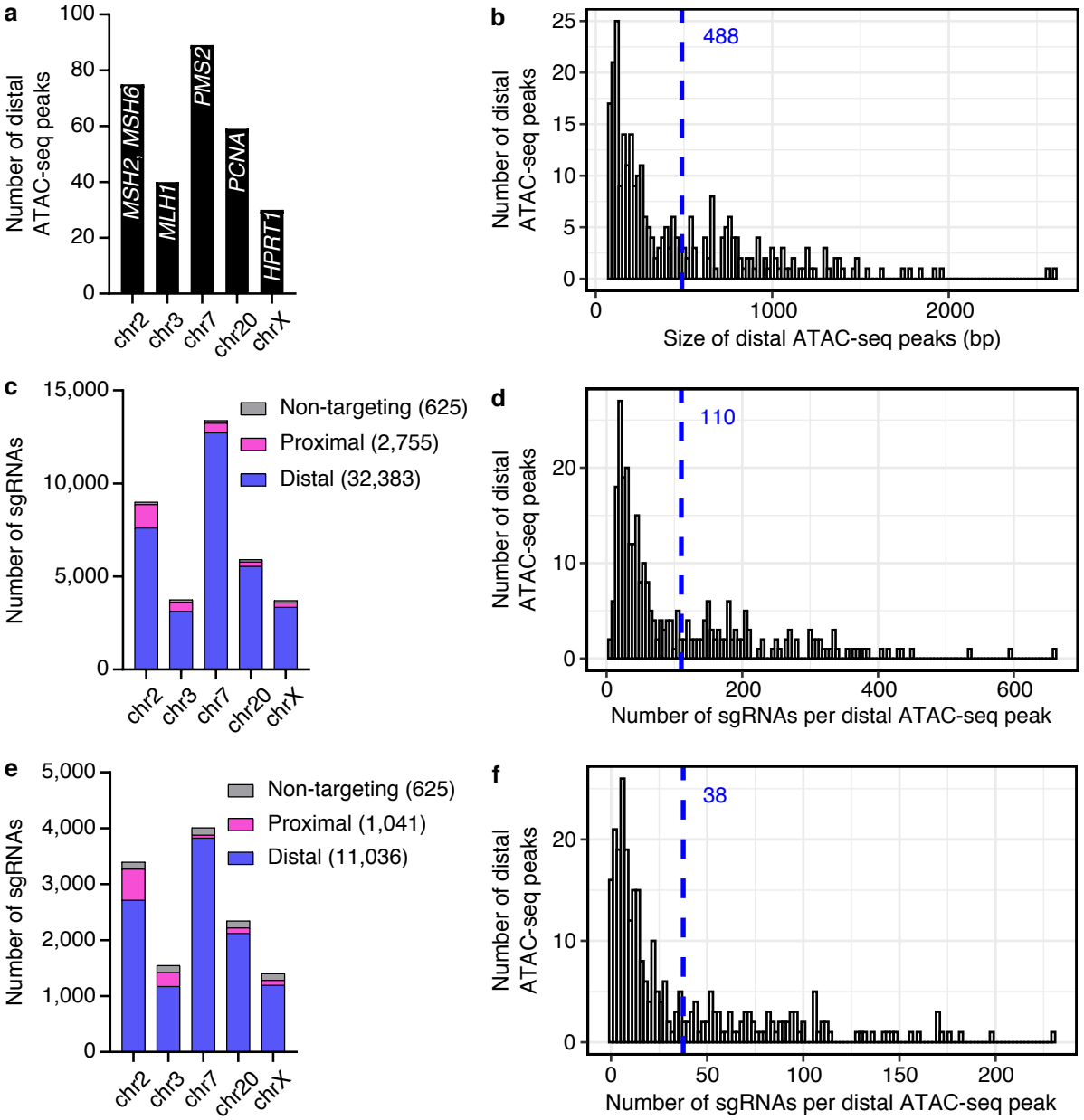


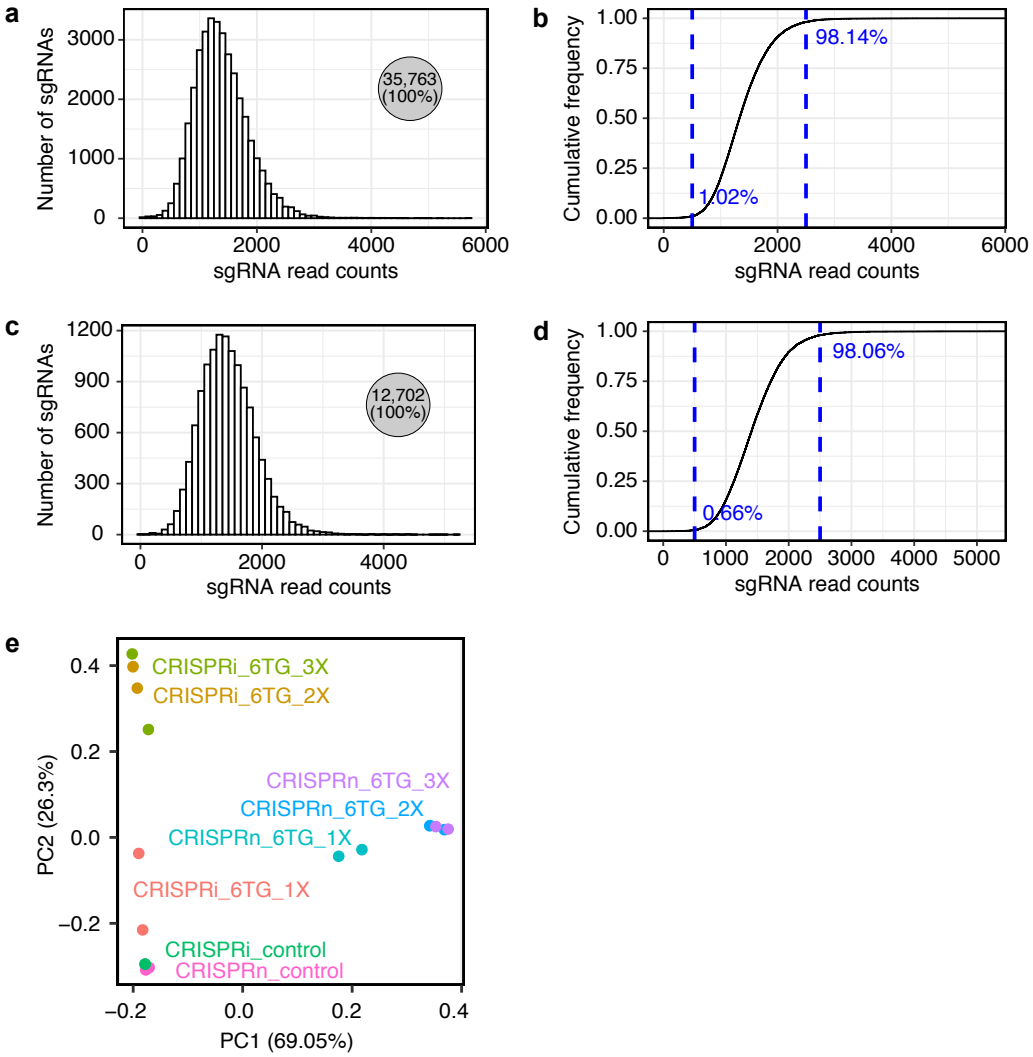
Figure 4. CRISPRpath can distinguish weak and strong enhancers by imposing different selection pressures. (**a**) Box plots show the enrichment score of the tested elements. TSS regions (black circles) show highest enrichment scores. Enhancers uniquely identified from the lower selection pressure (CRISPRi 2X, blue circles) exhibit lower enrichment scores compared to the enhancers identified from the higher selection pressure (CRISPRi 3X, red circles). *P* values are from Wilcoxon test. (**b**) Box plots show the CRISPRi perturbation at enhancers induced various degrees of transcriptional repression of target genes measured with RT-qPCR. Each dot represents the average value from three biological replicates. CRISPRi targeting TSS regions (dark gray) achieved the highest transcriptional repression. CRISPRi targeting strong enhancers (pink) leads a more substantial transcription silencing on target gene compared to CRISPRi targeting weak enhancers (cyan). *P* values are from Wilcoxon test. (**c**) Enrichment analysis of ATAC-seq, H3K27ac, H3K4me3, CTCF binding signals for strong (n = 33) and weak (n = 30) enhancers. Boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. *P* values for the difference between strong and weak enhancers are from Wilcoxon test; see Supplementary Table 7 for *P* values of all pairwise comparisons. (**d**) Intersection of genomic features for weak enhancers (blue bar) and strong enhancers (red bar). (**e**) Distance normalized H3K4me3 PLAC-seq contact frequency for strong (n = 23) and weak (n = 21) enhancers. Only the enhancers for *MLH1*, *PMS2*, *PCNA*, *HPRT1* are included (see Methods). Boxplots indicate the median, IQR, Q1 − 1.5 × IQR and Q3 + 1.5 × IQR. *P* value is from Wilcoxon test. (**f**) Heatmap shows the frequency of transcription factor motifs found in strong and weak enhancers.

Supplementary Figure 1



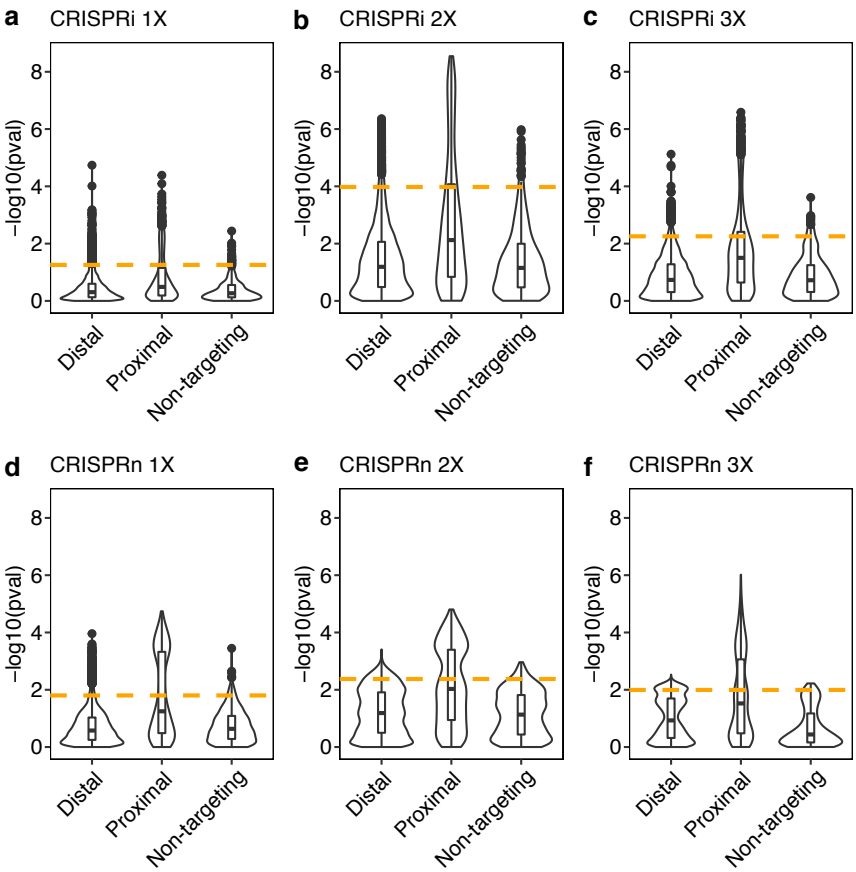
Supplementary Figure 1. Features of sgRNA library for CRISPRpath screen.
(**a**) Bar graph shows the number of distal ATAC-seq peaks used as candidate CREs for six target genes. (**b**) Histogram shows size distribution of distal ATAC-seq peaks. The average size is 488 bp (blue dash line). (**c, e**) 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 sgRNAs per distal ATAC-seq peak. Average numbers of sgRNA per ATAC-seq peak are indicated with blue dash lines.

Supplementary Figure 2



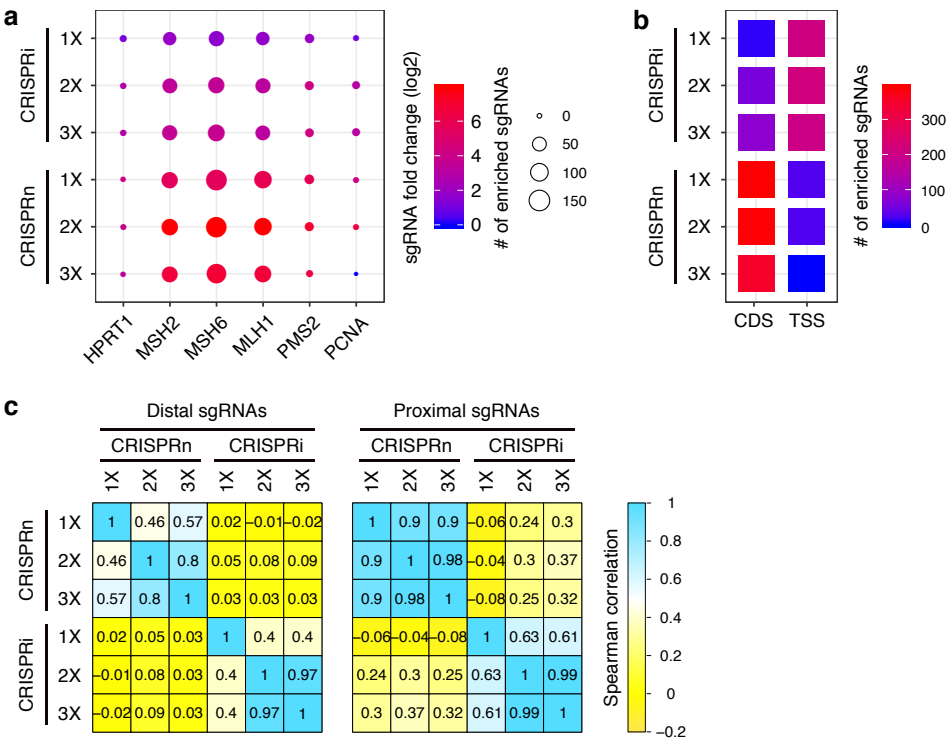
Supplementary Figure 2. Quality of the sgRNA library and CRISPRpath screen libraries.
(**a**) Distribution of sgRNA oligo read counts in the sgRNA library. (**b**) Cumulative frequency of sgRNAs in the sgRNA library. (**c**) Distribution of high quality sgRNAs read counts in the sgRNA library. (**d**) Cumulative frequency of high quality sgRNAs in the sgRNA library. The constructed sgRNA plasmid library recorvered all the designed sgRNAs with the copy number difference less than five fold for at least 97% designed sgRNAs. (**e**) PCA analysis shows the high reproducibility of the CRISPRpath screen libraries between biological replicates.

Supplementary Figure 3



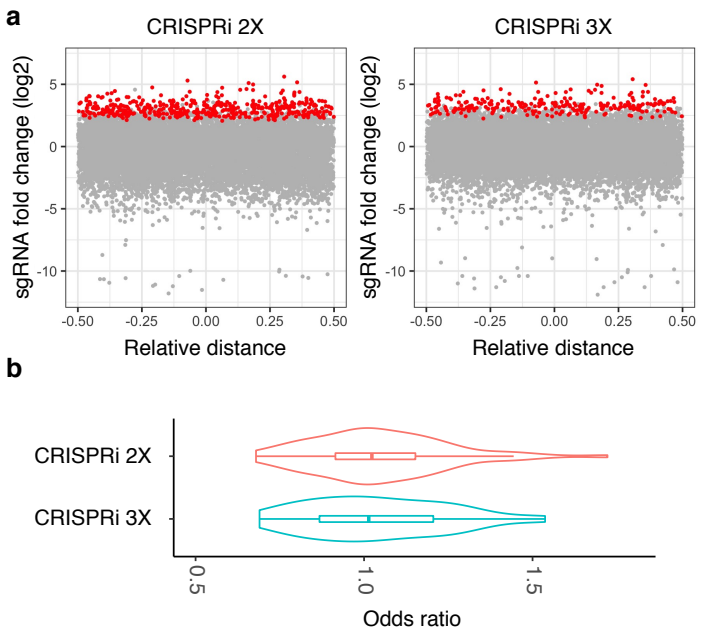
Supplementary Figure 3. P value cutoff used for identifying enriched sgRNAs from each screens. (**a-f**) Distribution of P value for tested distal, proximal and non-targeting control sgRNA groups. Orange dash lines indicate 5% percentile of the *P* values from non-targeting control sgRNAs to achieve a false discovery rate of 5%.

Supplementary Figure 4



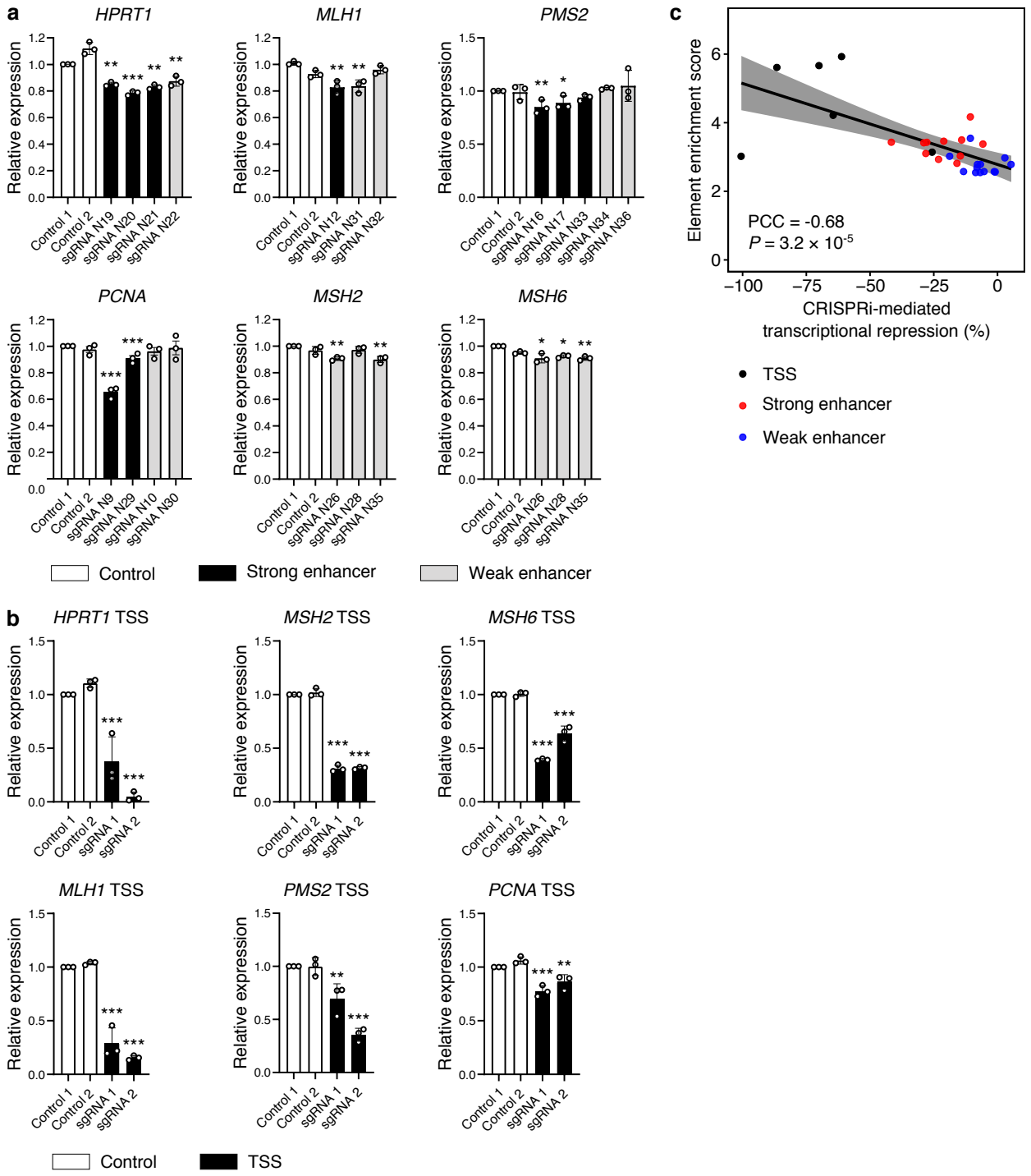
Supplementary Figure 4. Enriched proximal sgRNAs and sgRNA ranking analysis.
(**a**) Number and fold change of the enriched proximal sgRNAs for the six target genes from CRISPRi and CRISPRn screens. The color indicates fold changes, and the size of circle indicates the number of enriched sgRNAs. (**b**) 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.

Supplementary Figure 5



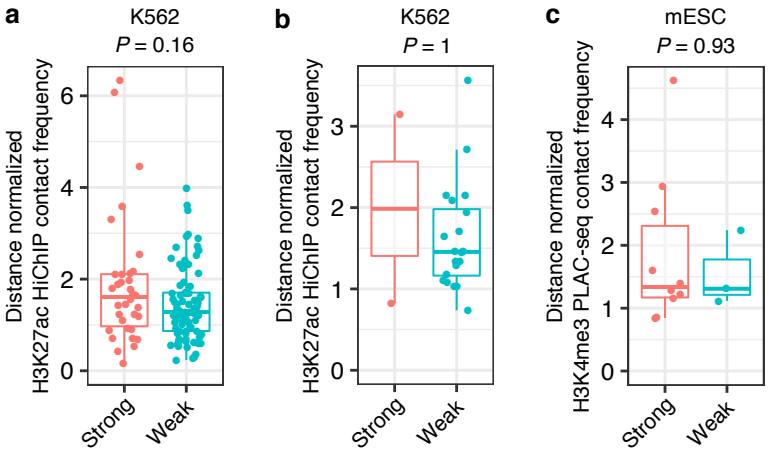
Supplementary Figure 5. Enriched sgRNAs identified from CRISPRi screens exhibit no position and strand preference.
(**a**) Enriched sgRNAs from CRISPRi 2X (red dots, n=448) and CRISPRi 3X (red dots, n=260) screens showed similar distributions across candidate CREs. (**b**) Odds ratio analysis of the fold change 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(log2(fold change of sgRNA targeting plus strand)) / ave(log2(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.

Supplementary Figure 6



Supplementary Figure 6. 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 score from CRISPRpath screens correlates with element effect size on transcription from CRISPRi (Pearson correlation, PCC = -0.68, $P = 3.2 \times 10\text{-}5$).

Supplementary Figure 7



Supplementary Figure 7. 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.