

1 **TITLE PAGE**

2

3 **Title:** Endogenous retroviruses mediate transcriptional rewiring in response to oncogenic
4 signaling in colorectal cancer

5

6 **Authors:** Atma Ivancevic¹, David M. Simpson¹, Olivia M. Joyner¹, Stacey M. Bagby², Lily L.
7 Nguyen^{1,3}, Ben G. Bitler³, Todd M. Pitts², Edward B. Chuong¹

8

9 **Corresponding Author:**

10 Edward B. Chuong

11 edward.chuong@colorado.edu

12 596 UCB

13 Boulder, CO 80309

14 303-735-8573

15

16 **Affiliations:**

17 1. BioFrontiers Institute and Department of Molecular, Cellular & Developmental Biology,
18 University of Colorado Boulder, Boulder, CO, United States

19

20 2. Division of Medical Oncology, Department of Medicine, University of Colorado Anschutz
21 Medical Campus, Aurora, CO, United States

22

23 3. Division of Reproductive Sciences, Department of Obstetrics and Gynecology, University
24 of Colorado School of Medicine, Aurora, CO, United States

25

26 **Keywords:** Endogenous retroviruses, epigenomics, colorectal cancer, AP1 signaling

27

28 **Running title:** ERV-derived enhancers in colorectal cancer

29 **Abstract:** Cancer cells exhibit rewired transcriptional regulatory networks that promote
30 tumor growth and survival. However, the mechanisms underlying the formation of these
31 pathological networks remain poorly understood. Through a pan-cancer epigenomic
32 analysis, we found that primate-specific endogenous retroviruses (ERVs) are a rich source
33 of enhancers displaying cancer-specific activity. In colorectal cancer and other epithelial
34 tumors, oncogenic AP1/MAPK signaling drives the activation of enhancers derived from the
35 primate-specific ERV family LTR10. Functional studies in colorectal cancer cells revealed
36 that LTR10 elements regulate tumor-specific expression of multiple genes associated with
37 tumorigenesis, such as *ATG12* and *XRCC4*. Within the human population, individual LTR10
38 elements exhibit germline and somatic structural variation resulting from a highly mutable
39 internal tandem repeat region, which affects AP1 binding activity. Our findings reveal that
40 ERV-derived enhancers contribute to transcriptional dysregulation in response to oncogenic
41 signaling and shape the evolution of cancer-specific regulatory networks.

42 **Main Text:**

43

44 **INTRODUCTION**

45 Cancer cells undergo global transcriptional changes resulting from genetic and epigenetic
46 alterations during tumorigenesis ¹. While regulatory remodeling can arise from somatic non-
47 coding mutations ², epigenomic studies have revealed that transformation is associated with
48 aberrant epigenetic activation of enhancer sequences that are typically silenced in normal
49 tissues ³⁻⁵. Pathological enhancer activity is an established mechanism underlying
50 tumorigenesis and therapy resistance, and therapeutic modulation of enhancer activity is an
51 active area of investigation ⁶⁻⁹. However, we have a limited understanding of the molecular
52 processes that shape and establish the enhancer landscapes of cancer cells.

53

54 Transposable elements (TEs) including endogenous retroviruses (ERVs) represent a
55 potentially significant source of enhancers that could shape cancer-specific gene regulation
56 ¹⁰. Many cancers exhibit genome-wide transcriptional reactivation of TEs, which can directly
57 impact cells by promoting oncogenic mutations and stimulating immune signaling ¹¹⁻¹⁴. In
58 addition, the reactivation of TEs is increasingly recognized to have gene regulatory
59 consequences in cancer cells ^{15,16}. Several transcriptomic studies have uncovered TEs as
60 a source of cancer-specific alternative promoters across many types of cancer, with some
61 examples shown to drive oncogene expression ¹⁷⁻²¹. TEs also show chromatin signatures
62 of enhancer activity in cancer cell lines ²²⁻²⁴, yet their functional relevance in patient tumors
63 has remained largely unexplored. Recent studies have characterized TE-derived enhancers
64 with oncogenic effects in acute myeloid leukemia ²⁵ and prostate cancer ²⁶ but the
65 prevalence and mechanisms of TE-derived enhancer activity are unknown for most cancer
66 types.

67

68 Here, we analyzed published cancer epigenome datasets to understand how TEs influence
69 enhancer landscapes and gene regulation across cancer types. Our pan-cancer analysis
70 revealed that elements from a primate-specific ERV named LTR10 show enhancer activity
71 in many epithelial tumors, and this activity is regulated by MAPK/AP1 signaling. We
72 conducted functional studies in HCT116 colorectal cancer cells, and found that LTR10
73 elements regulate AP1-dependent gene expression at multiple loci that include genes with

74 established roles in tumorigenesis. Finally, we discovered that LTR10 elements contain
75 highly mutable sequences that potentially contribute genomic variation affecting cancer-
76 specific gene expression. Our work implicates ERVs as a source of pathological regulatory
77 variants that facilitate transcriptional rewiring in cancer.

78

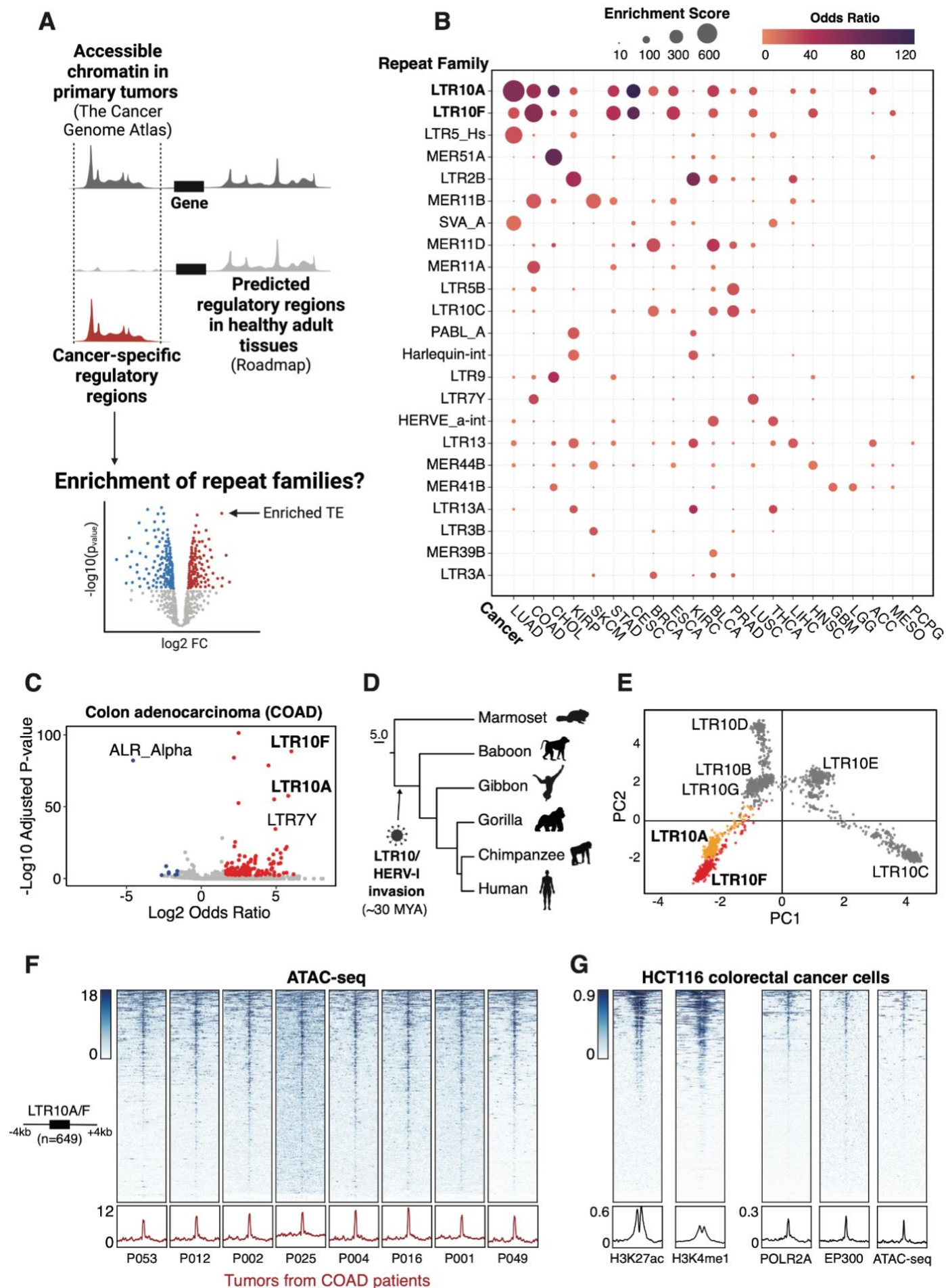
79 **RESULTS**

80 To assess the contribution of TEs to cancer cell epigenomes, we analyzed aggregate
81 chromatin accessibility maps from 21 human cancers generated by The Cancer Genome
82 Atlas project ²⁷. We defined cancer-specific subsets of accessible regions by subtracting
83 regions that show evidence of regulatory activity in any healthy adult tissue profiled by the
84 Roadmap Consortium (Fig 1A, Methods) ²⁸. Out of 1315 total repeat subfamilies annotated
85 in the human genome, we found 23 subfamilies that showed significant enrichment within
86 the accessible chromatin in at least one cancer type (Fig 1B), of which 19 correspond to
87 long terminal repeat (LTR) regions of primate-specific ERVs (Supp Table 1). These
88 observations from chromatin accessibility data generated from primary tumors confirm
89 previous reports of LTR-derived regulatory activity in cancer cell lines ^{22,24,25}, and support a
90 role for ERVs in shaping patient tumor epigenomes.

91

92 *LTR10 elements exhibit cancer-specific regulatory activity*

93 To further investigate the cancer-specific regulatory activity of ERVs, we focused on LTR10
94 elements, which were enriched within cancer-specific accessible chromatin for several types
95 of epithelial tumors including colorectal, stomach, prostate, and lung tumors (Fig 1C, Supp
96 Fig S1A). LTR10 elements (including LTR10A-G, n=2331) are derived from the LTR of the
97 gammaretrovirus HERV-I, which integrated into the anthropoid genome 30 million years ago
98 (Fig 1D, 1E) ²². As our initial TCGA analysis was conducted using aggregate data for each
99 tumor type, we first confirmed that LTR10 elements showed recurrent chromatin accessibility
100 across colorectal tumors from multiple individual patients (Fig 1F, Supp Fig S1B). We then
101 analyzed epigenomic datasets from the HCT116 colorectal cancer cell line ^{3,29-31} and found
102 that LTR10A and LTR10F elements exhibit canonical chromatin hallmarks of enhancer
103 activity, including enrichment of histone modifications H3K27ac and H3K4me1, the
104 transcriptional coactivator p300, and RNA Polymerase II occupancy (Fig 1G).



106 **Figure 1: Pan-cancer epigenomic analysis of TE activity.** (A) Pipeline to estimate TE subfamily enrichment
107 within cancer-specific regulatory regions. Aggregate ATAC-seq maps associated with each TCGA tumor type
108 were filtered to remove regulatory regions predicted in any normal adult tissues by Roadmap. Cancer-specific
109 accessible chromatin regions were tested for enrichment of 1315 repeat subfamilies using GIGGLE. (B) Bubble
110 chart summarizing TE subfamily enrichment within cancer-specific ATAC-seq regions across 21 cancer types
111 profiled by TCGA (acronyms shown on the x-axis; full names provided at [https://gdc.cancer.gov/resources-](https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations)
112 [tcga-users/tcga-code-tables/tcga-study-abbreviations](https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations)). TE subfamilies and cancer types are sorted based on
113 maximum enrichment score. (C) Enrichment of TE subfamilies within cancer-specific ATAC-seq associated
114 with colon adenocarcinomas (COAD) from TCGA. Every point represents a TE subfamily. Significantly
115 enriched TEs are shown in red; depleted TEs are shown in blue. (D) Estimated origin of HERV-I elements on
116 the primate phylogeny based on genomic presence or absence. (E) Principal component analysis based on
117 multiple sequence alignment of all LTR10 sequences over 200 bp in length in the human genome (n=1806).
118 Every point represents an individual LTR10 sequence. LTR10A and LTR10F sequences are colored orange
119 and red, respectively. (F) Heatmap of representative patient tumor ATAC-seq signals (TCGA patients COAD
120 P053, P012, P002, P025, P004, P016, P001, P049) over the merged set of 649 LTR10A/F elements. Bottom
121 metaprofiles represent average normalized ATAC signal across elements. (G) Heatmap of enhancer-
122 associated chromatin marks from HCT116 cells over the merged set of 649 LTR10A/F elements. From left to
123 right: H3K27ac ChIP-seq (GSE97527), H3K4me1 ChIP-seq (GSE101646), POLR2A ChIP-seq (GSE32465),
124 EP300 ChIP-seq (GSE51176), and HCT116 ATAC-seq (GSE126215). Bottom metaprofiles represent the
125 normalized signal across elements.

126

127 We did not observe enhancer-like profiles at LTR10C elements, which have previously been
128 identified as a source of p53 binding sites ^{22,32} (Supp Fig S1C). While most LTR10A and
129 LTR10F elements are not transcribed, some show evidence of transcription as promoters
130 for full-length non-coding HERV-I insertions or cellular transcripts (Supp Fig S1D).
131 Therefore, elements derived from the LTR10A and LTR10F subfamilies (hereafter referred
132 to as LTR10 elements) show robust epigenomic signatures associated with enhancer
133 activity in colorectal cancer cells.

134

135 We expanded our analysis to include epigenomic states from all adult tissues ²⁸. We found
136 no evidence for LTR10 enhancer activity in normal tissues, but instead observed general
137 enrichment of H3K9me3-associated heterochromatin marks (Supp Fig S2A, Supp Table 2).
138 To identify factors that directly bind to and potentially repress LTR10 elements, we analyzed
139 the Cistrome database ³¹ of published human ChIP-Seq datasets to identify transcriptional
140 repressors with evidence for enriched binding within LTR10 elements. Considering all cell
141 types, we found that LTR10 elements are significantly enriched for binding by ZNF562,
142 TRIM28, and SETDB1 (Fig 2A, 2B, Supp Table 3, Supp Table 4), which are components of
143 the KRAB-ZNF transposon silencing pathway ³³. In additional datasets generated from
144 healthy colorectal tissue samples ^{3,34–36}, LTR10 elements do not show any evidence of
145 enhancer activity (Supp Fig S2B). Our analysis suggests that, as expected for most primate-

146 specific TEs³⁷, LTR10 elements are normally subject to H3K9me3-mediated epigenetic
147 silencing in somatic tissues.

148

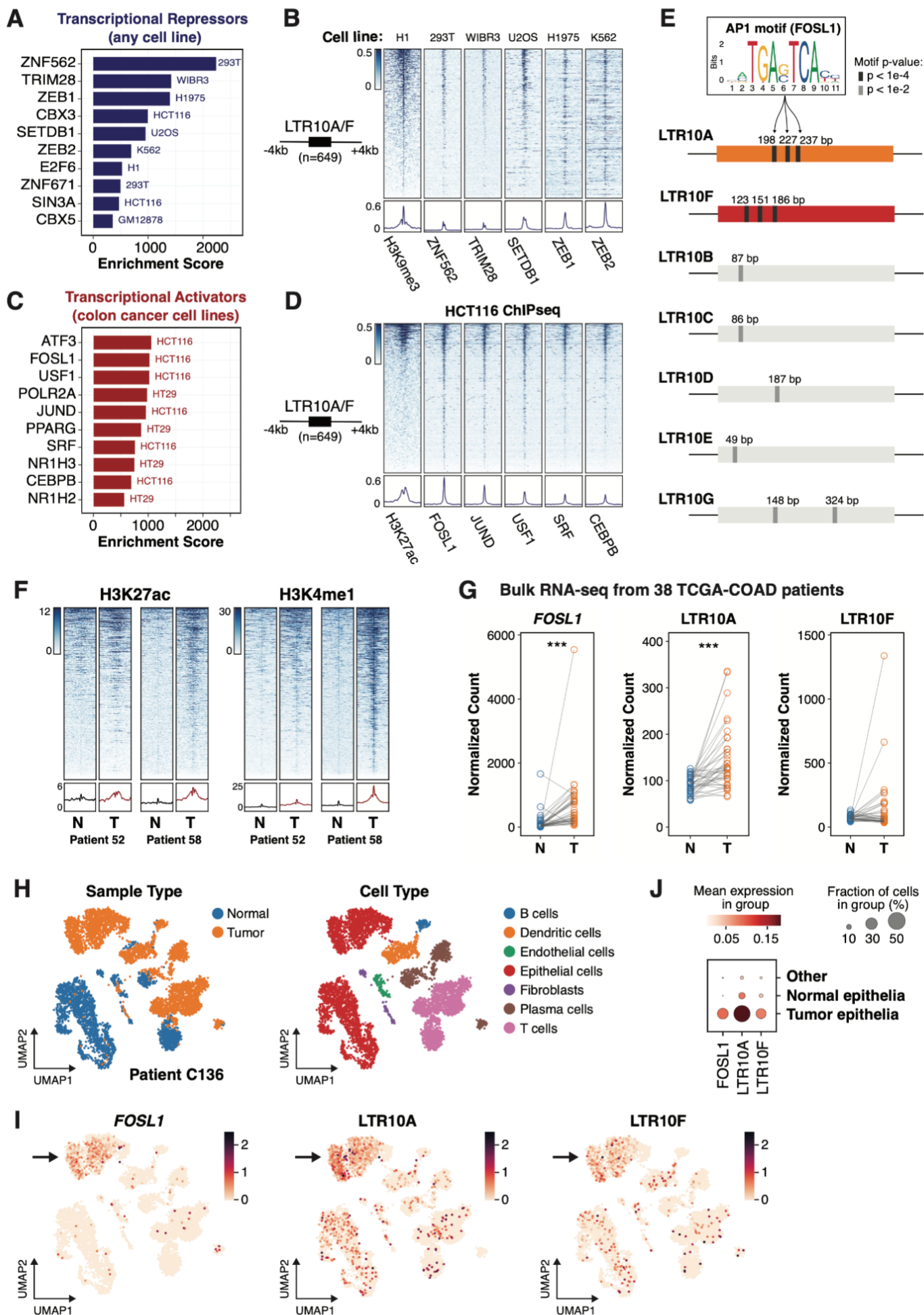
149 *LTR10 elements are bound by the AP1 transcription factor complex*

150 To identify which pathways are responsible for cancer-specific reactivation of LTR10
151 elements, we focused our Cistrome enrichment analysis on activating transcription factors
152 in colorectal cancer cell lines. LTR10 elements were significantly enriched for binding by
153 AP1 complex members (Fig 2C, 2D, Supp Table 3) including the FOSL1, JUND, and ATF3
154 transcription factors. The LTR10A and LTR10F consensus sequences harbor multiple
155 predicted AP1 binding motifs, which are enriched within LTR10 elements marked by
156 H3K27ac in HCT116 cells. Moreover, the AP1 motifs are largely absent in other LTR10
157 subfamilies (Fig 2E). Expanding our motif analysis to tumor-specific accessible chromatin
158 from 21 different cancer types, we found that AP1 motif enrichment generally correlates with
159 LTR10 enrichment, especially for LTR10A (Supp Fig S2C). In contrast, cancers without
160 LTR10 enrichment show little to no enrichment of AP1 motifs in tumor-specific accessible
161 chromatin (Supp Fig S2C). These analyses indicate that the cancer-specific enhancer
162 activity of LTR10 elements is likely driven by sequence-specific recruitment of the AP1
163 complex.

164

165 *LTR10 epigenetic and transcriptional activity is elevated in patient tumor cells*

166 We next compared the epigenetic status of LTR10 elements between patient-derived
167 colorectal cancer cells and normal cells. In multiple patient-matched epigenomic datasets
168^{38,39}, LTR10 elements show globally increased levels of enhancer-associated histone
169 modifications H3K27ac and H3K4me1 in tumor samples compared to adjacent normal
170 colorectal tissues (Fig 2F, Supp Fig S2D). In contrast, LTR10 elements did not show global
171 changes in H3K9me3 or H3K27me3 ChIP-seq signal in tumors compared to normal cells
172 (Supp Fig S2E). These observations suggest that removal of repressive histone marks may
173 not be required for LTR10 enhancer activity, however, single-cell epigenomic profiling would
174 be necessary to determine whether LTR10 elements are marked by both active and
175 repressive marks in the same cells.



177 **Figure 2: Regulatory activity of LTR10 in tumor and normal cells. (A)** Transcriptional repressors
178 associated with LTR10A/F elements, ranked by enrichment score. **(B)** Heatmap of ChIP-seq signal from
179 H3K9me3 and repressive factors, over LTR10A/F elements. From left to right: H3K9me3 ChIP-seq
180 (GSE16256), ZNF562 ChIP-seq (GSE78099), TRIM28 ChIP-seq (GSE84382), SETDB1 ChIP-seq
181 (GSE31477), ZEB1 (GSE106896), and ZEB2 ChIP-seq (GSE91406). **(C)** Transcriptional activators associated
182 with LTR10A/F elements, ranked by enrichment score. **(D)** Heatmap of ChIP-seq signal from H3K27ac and
183 activating transcription factors in HCT116 cells, over LTR10A/F elements. From left to right: H3K27ac ChIP-
184 seq (GSE96299), as well as ChIP-seq for FOSL1, JUND, USF1, SRF and CEBPB (all from GSE32465). **(E)**
185 Schematic of AP1 motif locations for LTR10 consensus sequences from each subfamily. Sequence logo for
186 AP1 motif FOSL1 (MA0477.1 from JASPAR) is shown, and predicted motif locations are marked on each
187 consensus. **(F)** Heatmap of H3K27ac and H3K4me1 ChIP-seq signals from tumor (T) and normal (N) samples
188 from colorectal cancer patients (CEMT patients AKCC52 and AKCC58) over LTR10A/F elements. Bottom
189 metaprofiles represent average normalized ChIP signal. **(G)** Dot plots of normalized counts for *FOSL1*,
190 LTR10A and LTR10F from bulk RNA-seq derived from a cohort of 38 TCGA patients with colorectal
191 adenocarcinomas. Each patient has one tumor (T) sample and one normal (N) colon sample. ***: $p < 0.001$,
192 paired sample Wilcoxon test. **(H)** UMAP projections of the single cell transcriptome of patient C136 from Pelka
193 et al (2021). UMAPs are colored according to tissue type or cell type. **(I)** UMAP projections of the same patient,
194 colored according to the expression of *FOSL1*, LTR10A, or LTR10F. In each case, the arrow points to the
195 tumor-specific cluster of epithelial cells. **(J)** Bubble plot of the same patient, showing the mean expression of
196 *FOSL1*, LTR10A and LTR10F in tumor epithelia vs normal epithelia.
197

198 We further assessed the transcriptional activity of LTR10 elements using matched
199 tumor/normal RNA-seq from 38 patients with colorectal adenocarcinomas from TCGA
200 controlled access data ²⁷ (Supp Fig S2F). Our RNA-seq analysis of the patient cohort
201 suggests that LTR10 transcripts are generally increased in tumor versus normal samples,
202 particularly at LTR10A elements (Fig 2G, Supp Fig S2G, Supp Table 5). Likewise, AP1 factor
203 *FOSL1* showed a robust and significant increase in expression in tumor versus normal
204 samples (Fig 2G, Supp Fig S2H, Supp Table 5), consistent with our hypothesis that the AP1
205 complex drives LTR10 transcriptional activity. Altogether, 15 of the 38 patients show a
206 consistent increase in *FOSL1*, LTR10A and LTR10F transcriptional activity in colorectal
207 tumor cells (Supp Table 5).

208 *LTR10* transcription marks tumor-specific epithelial cells

210 We next investigated LTR10 transcription at the single-cell level. We analyzed an
211 independent cohort of 36 colorectal cancer patients with publicly available single-cell RNA-
212 seq from matched tumor and normal samples for each patient ⁴⁰. We used scTE ⁴¹ to
213 reprocess the datasets and measure cell population-specific expression of TE subfamilies.
214 In line with our previous results from bulk RNA-seq, we found significant and recurrent
215 transcription of LTR10 elements in tumor-specific epithelial cells for 12 out of 36 patients
216 (Fig 2H-J, Supp Fig S2I & S2J, Supp Table 6). We observed co-expression of LTR10 and
217 *FOSL1* in tumor-specific epithelial cells for 10 of these patients (Supp Table 6), consistent

218 with a role for AP1 signaling in regulating LTR10 elements. Thus, our single-cell analysis
219 indicates that a subset of patients show robust LTR10 transcriptional activity specifically in
220 tumor-specific epithelial cells.

221
222 *LTR10 transcription is associated with dysregulated MAPK signaling*

223 Our initial analyses of patient cohorts suggest that LTR10 elements become transcriptionally
224 activated in about 30% of colorectal tumors. To determine which tumor molecular subtypes
225 are most likely to drive LTR10 activation, we performed correlative studies between LTR10
226 activity and tumor mutations, patient survival rates, and clinical outcomes. For this purpose,
227 we obtained and analyzed RNA-seq from 358 primary tumor samples derived from TCGA
228 patients with colon adenocarcinomas or adenomas²⁷. We first focused on correlating LTR10
229 transcriptional activity with KRAS mutation status. KRAS is one of the most frequently
230 mutated oncogenes in cancer: approximately 30-40% of colorectal cancer patients harbor
231 missense mutations in KRAS, and KRAS mutations have long been associated with
232 increased tumor aggressiveness, resistance to treatment, and poor patient outcomes⁴². We
233 found that LTR10A transcripts, in particular, are significantly elevated in tumors that harbor
234 a KRAS mutation (Supp Fig S2K, Supp Table 7), although we did not observe a noticeable
235 difference in *FOSL1* expression (Supp Fig S2L, Supp Table 7). Survival analyses based on
236 the expression of LTR10 transcripts or AP1 factors did not reveal any significant correlations
237 (Supp Fig S2M, S2N). However, we note that our analyses might have been limited by the
238 fact that subfamily-level LTR10 transcription by RNA-seq is an imperfect proxy for enhancer
239 activity, and only a small subset of colorectal cancer samples from TCGA had associated
240 survival data, limiting our statistical power.

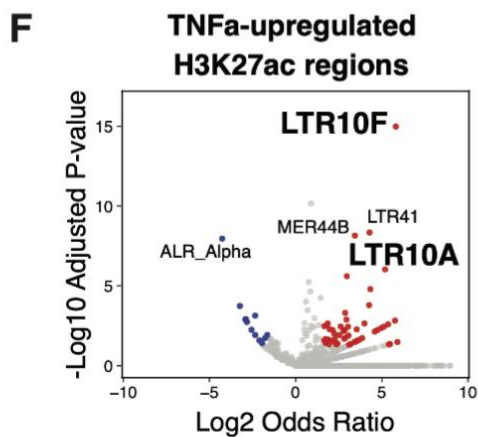
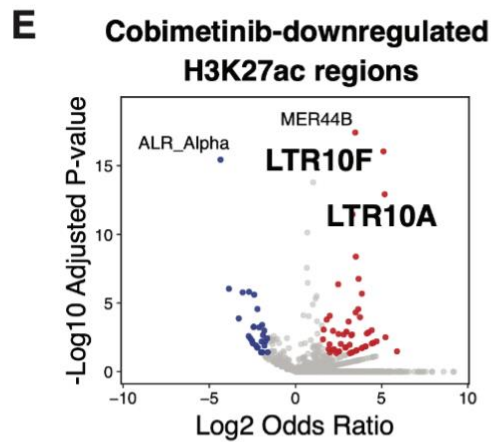
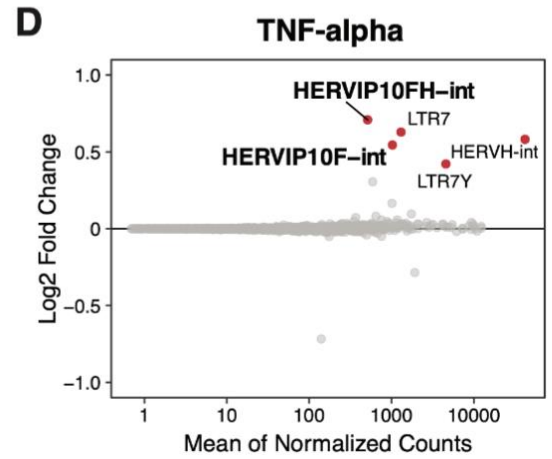
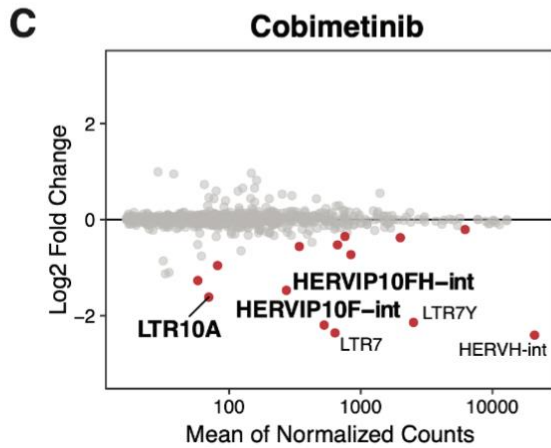
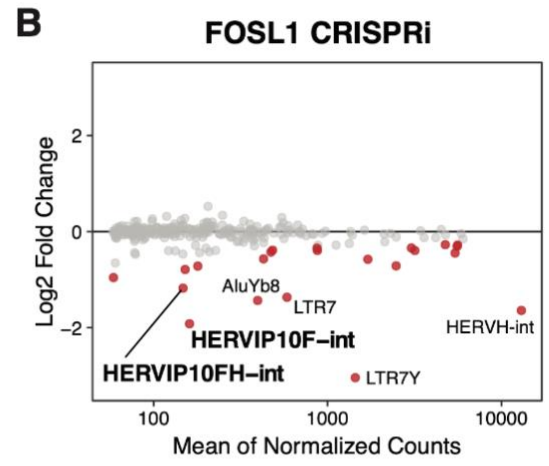
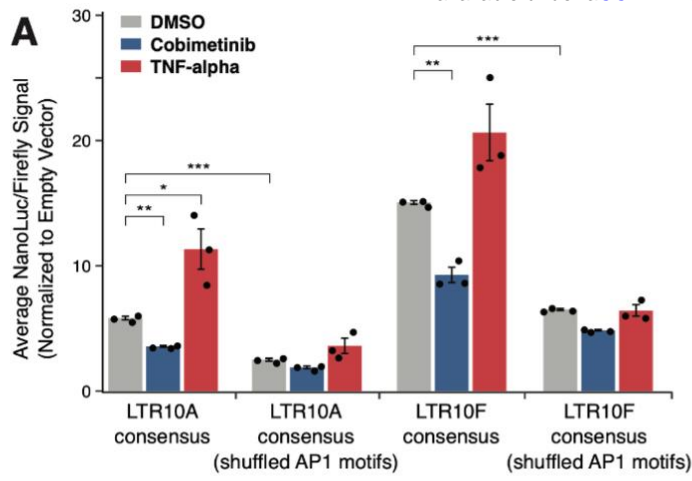
241
242 *AP1 signaling is required for LTR10 enhancer activity*

243 Dysregulation of AP1 signaling occurs in many cancers, driven by mutations that cause
244 oncogenic activation of the MAPK signaling pathway⁴³. Based on our findings that LTR10
245 elements are bound by AP1, and LTR10 transcriptional activity is correlated with the
246 expression of AP1 component *FOSL1*, we tested whether LTR10 regulatory activity is
247 affected by modulation of the AP1/MAPK signaling pathway using luciferase reporter
248 assays. We synthesized the LTR10A and LTR10F consensus sequences as well as variants
249 where the AP1 motifs were disrupted, and cloned the sequences into an enhancer reporter

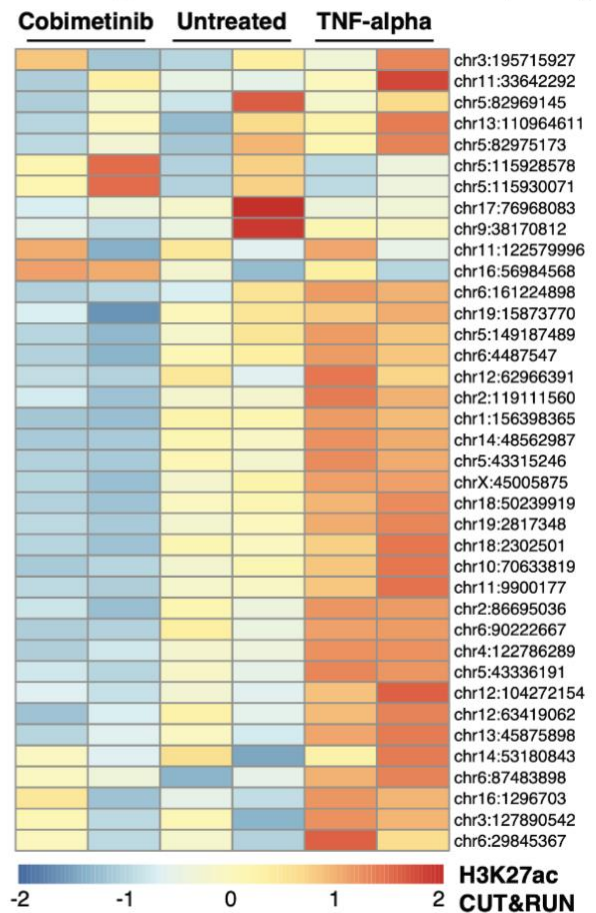
250 construct. We measured reporter activity in HCT116 cells that were treated for 24 hrs with
251 either TNF α to stimulate signaling or cobimetinib (a MEK1 inhibitor) to inhibit signaling.
252 Consistent with regulation by AP1, cobimetinib treatment caused a decrease in LTR10-
253 driven reporter activity, and TNF α caused an increase (Fig 3A). Overall regulatory activity
254 was greatly reduced in sequences where the AP1 motif was disrupted (Fig 3A). These
255 results show that LTR10 enhancer activity can be directly regulated by modulation of the
256 AP1/MAPK signaling pathway in cancer cells.

257
258 To test the role of the AP1 complex in endogenous LTR10 regulation, we silenced the AP1
259 component *FOSL1* using CRISPRi to determine the impact on LTR10 transcriptional activity.
260 Using HCT116 cells expressing dCas9-KRAB-MeCP2⁴⁴, we transfected a guide RNA
261 (gRNA) targeting the *FOSL1* transcription start site (TSS) and used RNA-seq to compare
262 gene and TE expression to control cells transfected with a non-targeting gRNA. We first
263 confirmed silencing of *FOSL1* (Supp Fig S3A, S3B), then analyzed TE transcript expression
264 summarized at the subfamily level to account for reads mapping to multiple insertions of the
265 same TE⁴⁵. This analysis revealed that full-length LTR10/HERV-I elements were
266 significantly downregulated upon silencing *FOSL1* (Fig 3B), supporting a direct role for the
267 AP1 complex in regulating LTR10 activity.

268
269 Next, we investigated how endogenous LTR10 elements respond to modulation of
270 MAPK/AP1 signaling at the chromatin level. We treated HCT116 cells with either cobimetinib
271 or TNF α for 24 hrs and profiled each response using RNA-seq and H3K27ac CUT&RUN
272 (Supp Fig S3C, S3D, S3E, S3F). Consistent with our reporter assay results, our RNA-seq
273 analysis showed that full-length LTR10/HERV-I transcripts were significantly downregulated
274 upon cobimetinib treatment (Fig 3C), and upregulated upon TNF α treatment (Fig 3D). LTR10
275 elements showed similar responses based on H3K27ac CUT&RUN signal, exhibiting
276 significant enrichment within the genome-wide set of predicted enhancers downregulated
277 by cobimetinib or upregulated by TNF α (Fig 3E, 3F). We also observed clear TNF α -induced
278 H3K27ac signal over LTR10 elements in a published dataset of SW480 colorectal cancer
279 cells⁴⁶ (Supp Fig S3G). These results indicate that LTR10 elements represent a significant
280 subset of genome-wide enhancers and transcripts in HCT116 cells that are directly
281 modulated by MAPK/AP1 signaling.



G **AP1-regulated LTR10 enhancers (n=38)**



283 **Figure 3: Control of LTR10 activity by AP1/MAPK signaling. (A)** Luciferase reporter assays of LTR10A/F
284 consensus sequences, including sequence variants containing shuffled AP1 motifs. Reporter activity was
285 measured in HCT116 cells treated with DMSO (n=3), cobimetinib (n=3), or TNF-alpha (n=3) for 24 hrs. Values
286 are normalized to firefly co-transfection controls, and presented as fold-change against the mean values from
287 cells transfected with a empty minimal promoter pNL3.3 vector. *: p < 0.05, **: p < 0.01, ***: p < 0.001, two-
288 tailed student's t-test. Error bars denote SEM. **(B-D)** MA plots of TE subfamilies showing significant differential
289 expression in HCT116 cells subject to FOSL1 silencing **(B)**, 24hr cobimetinib treatment **(C)**, or 24hr TNF-alpha
290 treatment **(D)**, based on RNA-seq. Dots are colored in red if they are significant (adjusted p<0.05, log2FC<0
291 for FOSL1/cobimetinib and log2FC>0 for TNFa). **(E)** Volcano plot showing TE subfamily enrichment in the set
292 of H3K27ac regions significantly downregulated by cobimetinib. **(F)** Volcano plot showing TE subfamily
293 enrichment in the set of H3K27ac regions significantly upregulated by TNF-alpha. **(G)** Heatmap of normalized
294 H3K27ac CUT&RUN signal for 38 LTR10 elements predicted to function as enhancers regulating AP1 target
295 genes for each treatment replicate.
296

297 *LTR10 elements regulate cancer-specific pathological gene expression*

298 To determine whether any LTR10-derived enhancers have a functional effect on
299 AP1/MAPK-dependent gene expression in colorectal cancer cells, we used our RNA-seq
300 and CUT&RUN data from HCT116 cells to identify elements predicted to have gene
301 regulatory activity. While we found that the AP1 component FOSL1 is required for LTR10
302 regulatory activity, oncogenic MAPK signaling can mediate transcriptional dysregulation
303 through additional pathways beyond FOSL1 and AP1 signaling⁴⁷. Therefore, we defined
304 potential AP1/MAPK-regulated genes using two approaches, based on our RNA-seq data
305 from our FOSL1 knockdown or TNFa/cobimetinib treatment. We first defined a set of 456
306 AP1-dependent genes based on being significantly downregulated by our CRISPRi silencing
307 of the AP1 component *FOSL1* (Supp Table 8). We identified LTR10 elements predicted to
308 regulate these genes using the activity by contact model⁴⁸ to assign enhancer-gene targets
309 based on LTR10 element H3K27ac signal and chromatin interaction data. This identified 38
310 LTR10-derived enhancers (Fig 3G) predicted to regulate 56 (12.2%) of the 456 AP1-
311 dependent genes (Supp Table 8), including many with established roles in cancer
312 pathophysiology. In a secondary analysis, we defined 620 MAPK-dependent genes as
313 genes that are both upregulated by TNFa and downregulated by cobimetinib, and found 57
314 LTR10-derived enhancers predicted to regulate 74 (11.9%) of these genes (Supp Fig S3H,
315 Supp Table 9). Collectively, we identified a total of 71 distinct LTR10 enhancers (Supp Table
316 10) predicted to contribute to the regulation of roughly 12% of genes with AP1 or MAPK-
317 dependent gene expression in HCT116 cells, supporting an important role in mediating
318 global transcriptional rewiring in cancer.

319

320 We tested the regulatory activity of six predicted LTR10 enhancers using CRISPR to knock
321 down or knock out individual elements in HCT116 cells. We prioritized the elements based
322 on epigenomic evidence of tumor-specific enhancer activity and having predicted target
323 genes with reported relevance to tumor development or therapy resistance. We separately
324 silenced each LTR10 element using CRISPRi and selected one element (LTR10.KDM6A)
325 to delete using CRISPR/Cas9, due to its intronic location. We used RNA-seq to determine
326 the transcriptional consequences of perturbing each element. For each LTR10 tested, we
327 observed local downregulation of multiple genes within 1.5 MB of the targeted element,
328 confirming their activity as functional enhancers in HCT116 cells. These included *ATG12*,
329 *XRCC4*, *TMEM167A*, *VCAN*, *NES*, *FGF2*, *AGPAT5*, *MAOB*, and *MIR222HG* (Fig 4, Fig 5,
330 Supp Fig S4A-S4J, Supp Tables 11-16). For three elements (LTR10.MEF2D,
331 LTR10.MCPH1 and LTR10.KDM6A), the predicted target gene did not show significant
332 expression changes, but we observed downregulation of other AP1/MAPK-dependent
333 genes near the element (Supp Fig S4B, S4F, S4H). Collectively, our characterization of six
334 LTR10 elements verified that 21 genes are regulated by LTR10 elements; most (18/21) of
335 which are regulated by AP1/MAPK signaling based on our RNA-seq data. These
336 experiments demonstrate that multiple LTR10-derived enhancers mediate AP1/MAPK-
337 dependent gene expression of nearby genes in HCT116 cells.

338
339 We focused on two of these LTR10-derived enhancers to explore their functional impact on
340 tumor cells. We first investigated an enhancer that regulates *ATG12* (LTR10.ATG12),
341 formed by two LTR10F elements on chromosome 5, located 87 kb from predicted target
342 genes *ATG12* and *AP3S1* (Fig 4A). Silencing the LTR10.ATG12 enhancer resulted in
343 downregulation of *ATG12* as well the neighboring gene *AP3S1* and eight other genes within
344 1.5 Mb (Fig 4B, 4C, 4D, Supp Table 11). As a separate control, we used CRISPRi to silence
345 the *ATG12* promoter and found highly specific silencing of *ATG12* (Supp Fig S4K, S4L, Supp
346 Table 17). These results indicate that the LTR10.ATG12 element functions as an enhancer
347 that affects multiple genes in the locus. Genome-wide, we observed differential regulation
348 of other genes, possibly due to indirect effects from target gene knockdown or off-target
349 silencing of other LTR10 elements (Supp Fig 4M). Notably, we observed that multiple genes
350 regulated by LTR10.ATG12 showed similar patterns of transcriptional downregulation in
351 response to *FOSL1* silencing and cobimetinib treatment (Fig 4D). These results indicate that

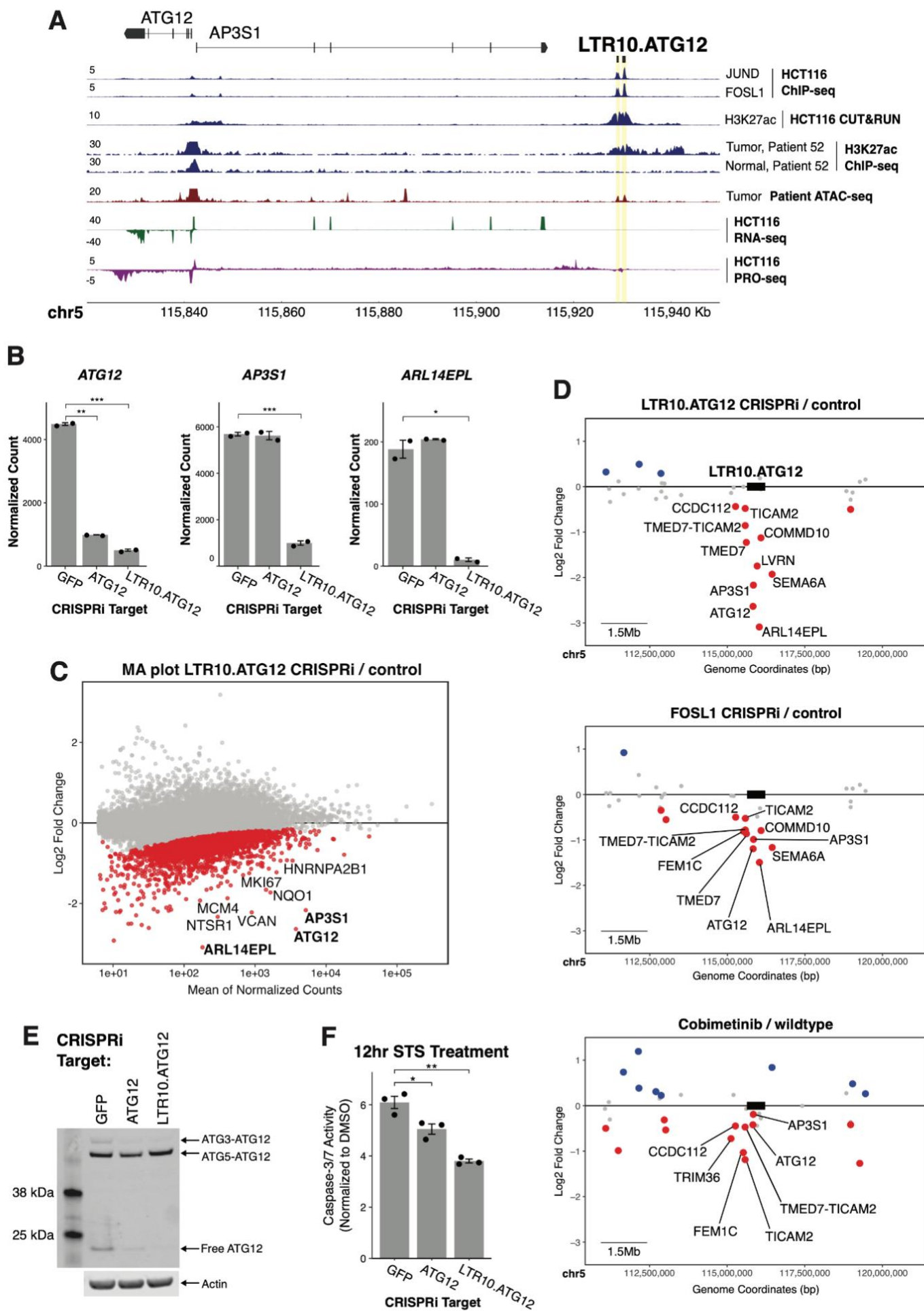
352 LTR10.ATG12 acts as an enhancer that controls AP1-dependent transcriptional activation
353 of multiple genes in the *ATG12/AP3S1* locus in HCT116 cells.

354

355 The *ATG12* gene encodes a ubiquitin-like modifier required for macroautophagy as well as
356 mitochondrial homeostasis and apoptosis^{49–52}. Expression of *ATG12* is associated with
357 tumorigenesis and therapy resistance in colorectal and gastric cancer^{53,54}, but the
358 mechanism of cancer-specific regulation of *ATG12* has not been characterized. Therefore,
359 we aimed to determine whether the LTR10.ATG12 enhancer was responsible for regulating
360 *ATG12* expression and activity in HCT116 cells. First, we validated that silencing the
361 enhancer resulted in decreased ATG12 protein levels by immunoblotting (Fig 4E). In cells
362 where either *ATG12* or the enhancer was silenced, there was a clear reduction in protein
363 levels of both free ATG12 and the ATG3-ATG12 conjugate. There was minimal knockdown
364 effect on the levels of the ATG5-ATG12 conjugate, which has previously been observed in
365 *ATG12* silencing experiments and is due to the high stability of the ATG5-ATG12 complex
366⁴⁹.

367

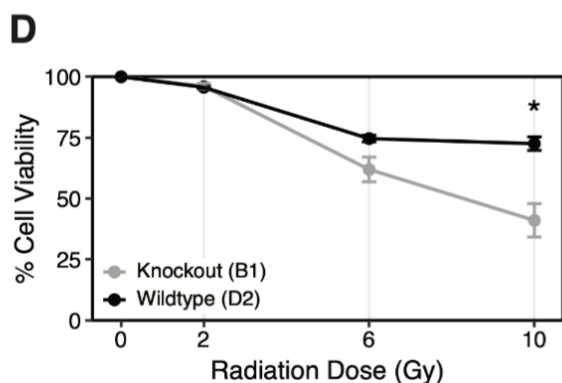
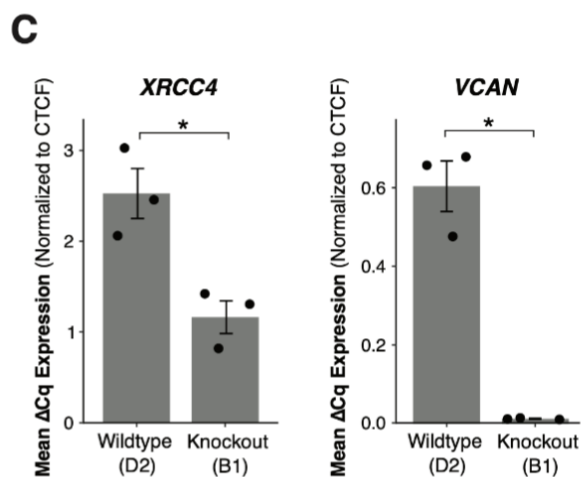
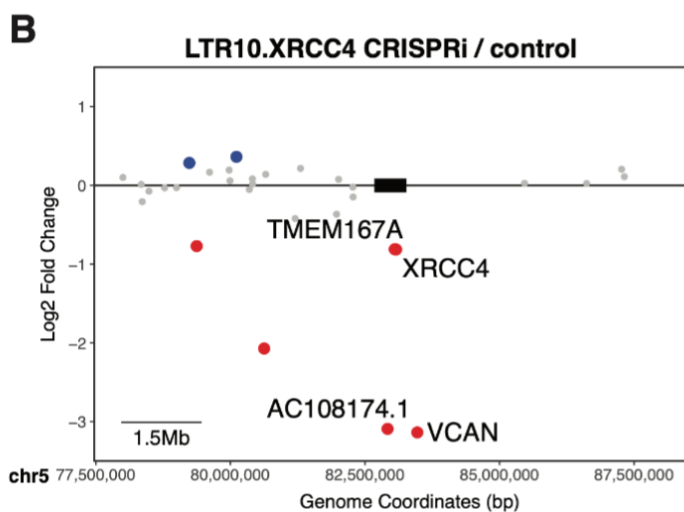
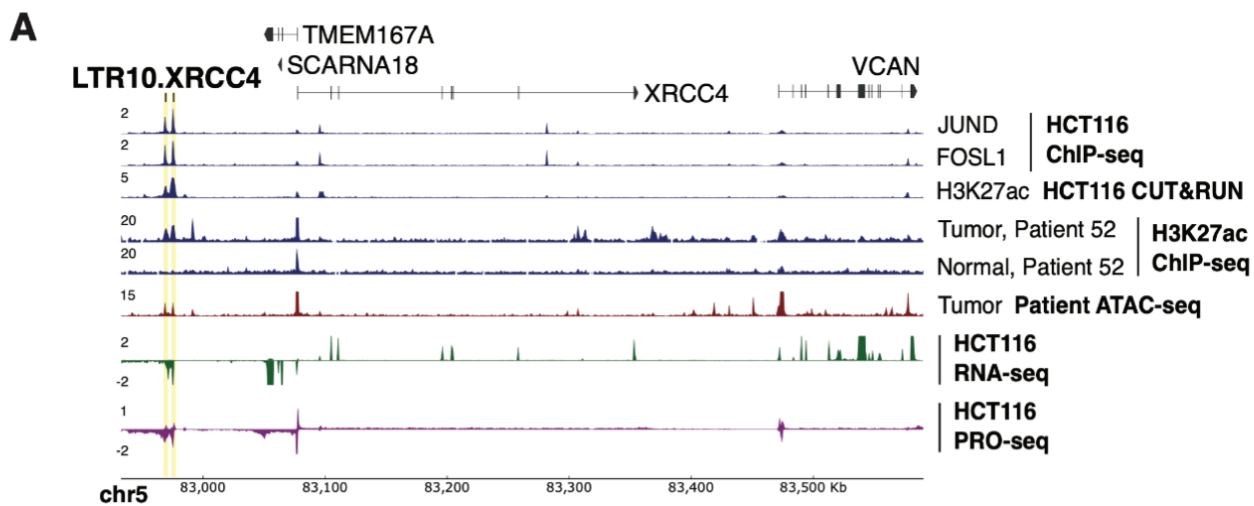
368 We tested whether ATG12-dependent functions require the activity of the LTR10.ATG12
369 enhancer. We treated each cell line with staurosporine (STS) to trigger mitochondrial
370 apoptosis, which is dependent on free ATG12 binding to Bcl-2⁵⁰. In cells where either
371 *ATG12* or the enhancer was silenced, we observed significantly reduced caspase 3/7
372 activity, indicating defective mitochondrial apoptosis (Fig 4F). We did not detect differences
373 in macroautophagy in cells treated with bafilomycin (Supp Fig S4N), consistent with the lack
374 of knockdown of the ATG5-ATG12 conjugate⁵². Our experimental results from silencing
375 both *ATG12* and the enhancer are concordant with previous studies directly silencing *ATG12*
376 using siRNAs in other cancer cell lines^{49,50}. Together, these experiments demonstrate that
377 the LTR10.ATG12 enhancer is functionally important for ATG12-dependent activity in
378 HCT116 cells.



380 **Figure 4: Functional characterization of LTR10.ATG12 in HCT116 cells. (A)** Genome browser screenshot
381 of the *ATG12/AP3S1* locus with the LTR10.ATG12 enhancer labeled. From top to bottom: JUND and FOSL1
382 ChIP-seq (GSE32465), H3K27ac CUT&RUN (in-house), H3K27ac ChIP-seq from matched tumor/normal
383 samples from the CEMT Canadian Epigenome Project (patient AKCC52), tumor ATAC-seq from TCGA (patient
384 COAD P022), HCT116 RNA-seq (in-house), and HCT116 PRO-seq (GSE129501). Axis numbers represent
385 the upper limit of the range; the lower limit is always zero. **(B)** Normalized RNA-seq expression values of
386 *ATG12*, *AP3S1*, and *ARL14EPL* in dCas9-KRAB-MeCP2 HCT116 cells stably transfected with gRNAs
387 targeting the *ATG12* transcription start site (n=2), the LTR10.ATG12 element (n=2), or non-targeting (GFP)
388 control (n=2). *: p < 0.05, **: p < 0.01, ***: p < 0.001, Welch's t-test. Error bars denote SEM. **(C)** MA plot
389 showing global gene expression changes in cells in response to silencing LTR10.ATG12. Significantly
390 downregulated genes are shown in red. **(D)** Scatterplot of gene expression changes in the locus containing
391 the LTR10.ATG12 element, associated with i) silencing LTR10.ATG12, ii) silencing *FOSL1*, or iii) cobimetinib
392 treatment. Significantly downregulated genes are shown in red; significantly upregulated genes are shown in
393 blue. Significantly downregulated genes located within 1.5 MB of the targeted element are labeled (element
394 box not drawn to scale). **(E)** Immunoblot of endogenous ATG12 in each CRISPRi cell line. Different ATG12
395 conjugate forms are labeled. **(F)** Caspase-3/7 activity after 12 hrs staurosporine (STS) treatment, measured
396 by the Caspase-Glo 3/7 assay. Treatments were performed in triplicate and signal for each cell line was
397 normalized to signal from DMSO-treatment. *: p < 0.05, **: p < 0.01, Welch's t-test. Error bars denote SEM.
398

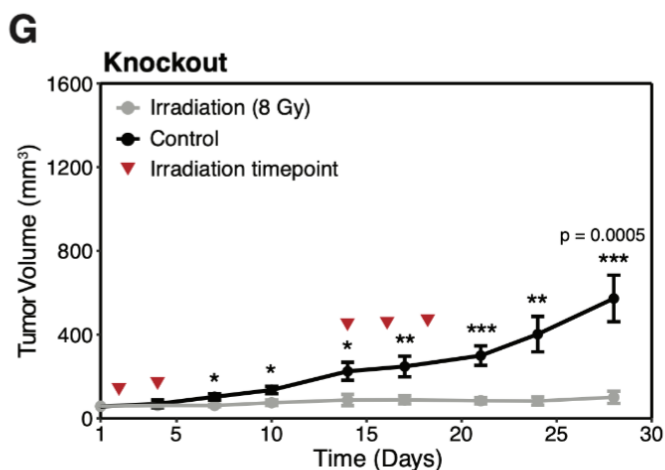
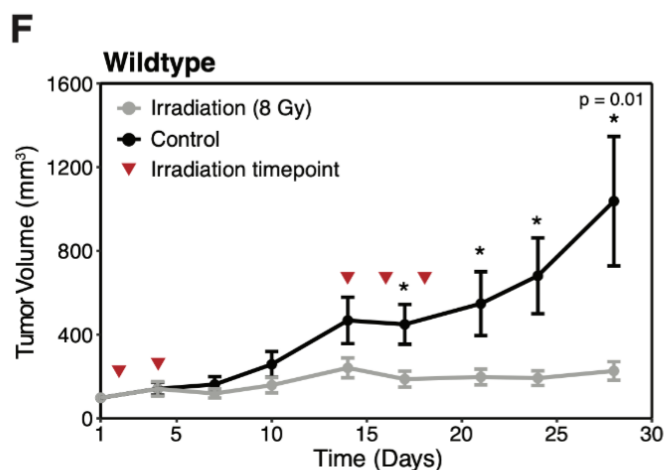
399 We next focused on the LTR10.XRCC4 enhancer, which regulates *XRCC4* and *VCAN*
400 based on our CRISPRi silencing experiment (Fig 5A, 5B, Supp Fig S5A, S5B, Supp Table
401 12). *XRCC4* is a DNA repair gene required for non-homologous end joining and promotes
402 resistance to chemotherapy and radiation therapy^{55–59}. *VCAN* is an extracellular matrix
403 protein that promotes tumor metastasis, invasion, and growth^{60–62}. Both *VCAN* and *XRCC4*
404 have been reported to be regulated by MAPK/AP1 signaling in tumor cells^{56,63}, but the
405 specific regulatory elements driving tumor-specific expression of these genes are unknown.
406 We validated the enhancer activity of this element by generating cells harboring
407 homozygous deletions using CRISPR (Supp Fig S5C, S5D) and confirmed that *XRCC4* and
408 *VCAN* were significantly downregulated in edited cells (Fig 5C).

409
410 Previous studies have demonstrated that silencing or knocking out *XRCC4* directly causes
411 increased sensitivity to DNA-damaging agents such as irradiation^{55,64,65}, including in
412 HCT116 cells⁶⁶. To test whether the LTR10.XRCC4 enhancer regulates *XRCC4* function in
413 cancer, we subjected control and knockout cells to 10 Gy irradiation, and found that knockout
414 cells showed reduced viability following irradiation (Fig 5D), consistent with a previous study
415 showing the role of *XRCC4* in tumor cell survival following irradiation⁶⁷. We next tested how
416 the deletion of LTR10.XRCC4 affects tumor response to irradiation in a mouse xenograft
417 model. Irradiation inhibits the growth of tumors derived from HCT116 cells⁶⁸, therefore we
418 tested whether reducing *XRCC4* expression by deleting LTR10.XRCC4 affects tumor growth
419 inhibition by irradiation.



E

LTR10.XRCC4 Status	Analysis	Estimate	Responder
Wildtype	TGI	0.58	No
Wildtype	mRECIST	PD	Yes
Wildtype	AUC	-4.43	No
Knockout	TGI	0.84	Yes
Knockout	mRECIST	SD	Yes
Knockout	AUC	-24.66	Yes



421 **Figure 5: Functional characterization of LTR10.XRCC4 in HCT116 cells and xenograft models. (A)**
422 Genome browser screenshot of the *XRCC4* locus with the LTR10.XRCC4 enhancer labeled. From top to
423 bottom: JUND and FOSL1 ChIP-seq (GSE32465), H3K27ac CUT&RUN (in-house), H3K27ac ChIP-seq from
424 matched tumor/normal samples from the CEMT Canadian Epigenome Project (patient AKCC52), tumor ATAC-
425 seq from TCGA (patient COAD P022), HCT116 RNA-seq (in-house), and HCT116 PRO-seq (GSE129501).
426 Axis numbers represent the upper limit of the range; the lower limit is always zero. **(B)** Scatterplot of gene
427 expression changes at the *XRCC4* locus after CRISPR silencing of the LTR10.XRCC4 enhancer. Significantly
428 downregulated genes are shown in red; significantly upregulated genes are shown in blue. Significantly
429 downregulated genes located within 1.5 MB of the targeted element are labeled (element box not drawn to
430 scale). **(C)** RT-qPCR expression values of *XRCC4* and *VCAN* in wildtype HCT116 cells (n=3) and
431 LTR10.XRCC4 knockout cells (n=3). *: $p < 0.05$, Welch's t-test. Error bars denote SEM. **(D)** Dose-response
432 curve showing cell viability in response to 0-10 Gy irradiation for LTR10.XRCC4 knockout and wildtype cells.
433 Cell viability was measured by CellTiterGlo after 5 days and each replicate (n=2) was normalized to non-
434 irradiated cells (n=2, averaged). *: $p < 0.05$, paired student's t-test. Error bars denote SEM. **(E)** Classification
435 of responder versus non-responder for both wildtype and LTR10.XRCC4 knockout cells, based on xenograft
436 growth curves of untreated or irradiated mice. Three measures were calculated using Ortmann et al. (2021)⁶⁹:
437 tumor growth inhibition (TGI), mRECIST, and area under the curve (AUC). **(F-G)** Average growth curves across
438 replicates (n=9-10) for wildtype **(F)** versus LTR10.XRCC4 knockout **(G)** HCT116 xenograft tumors, with and
439 without irradiation, for 28 days. 8 Gy treatment timepoints (days 2, 4, 14, 16, and 18) are indicated by red
440 triangles. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, two sample t-test assuming equal variances. Error bars denote
441 SEM. Individual growth curves are shown in Supp Fig S5E and S5F.
442

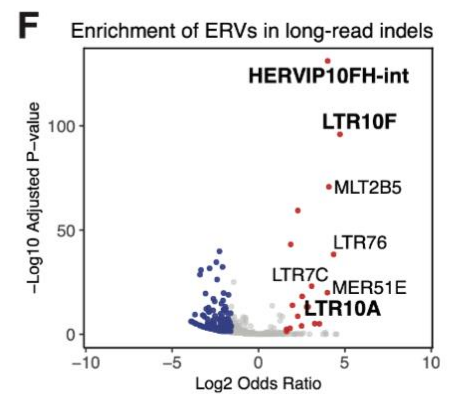
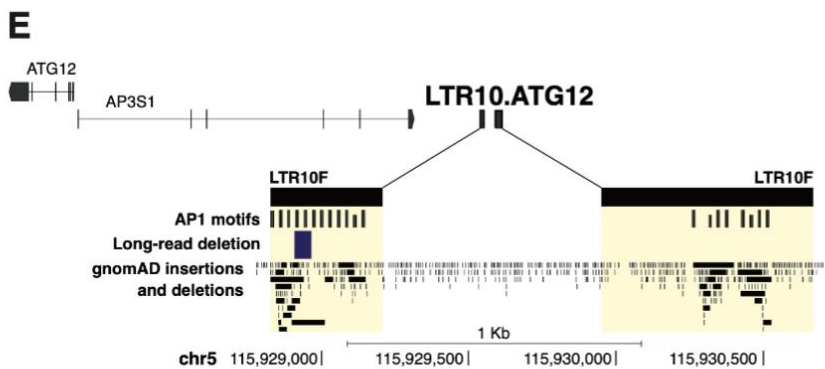
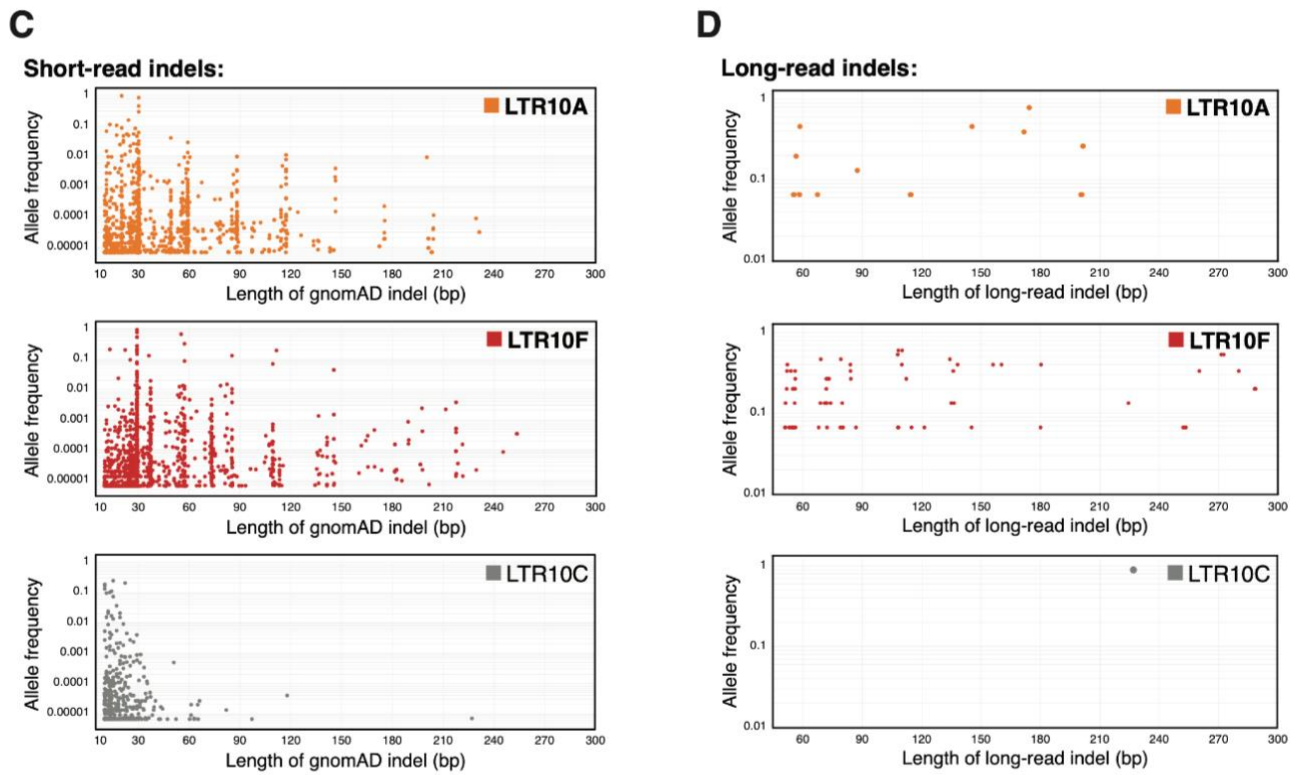
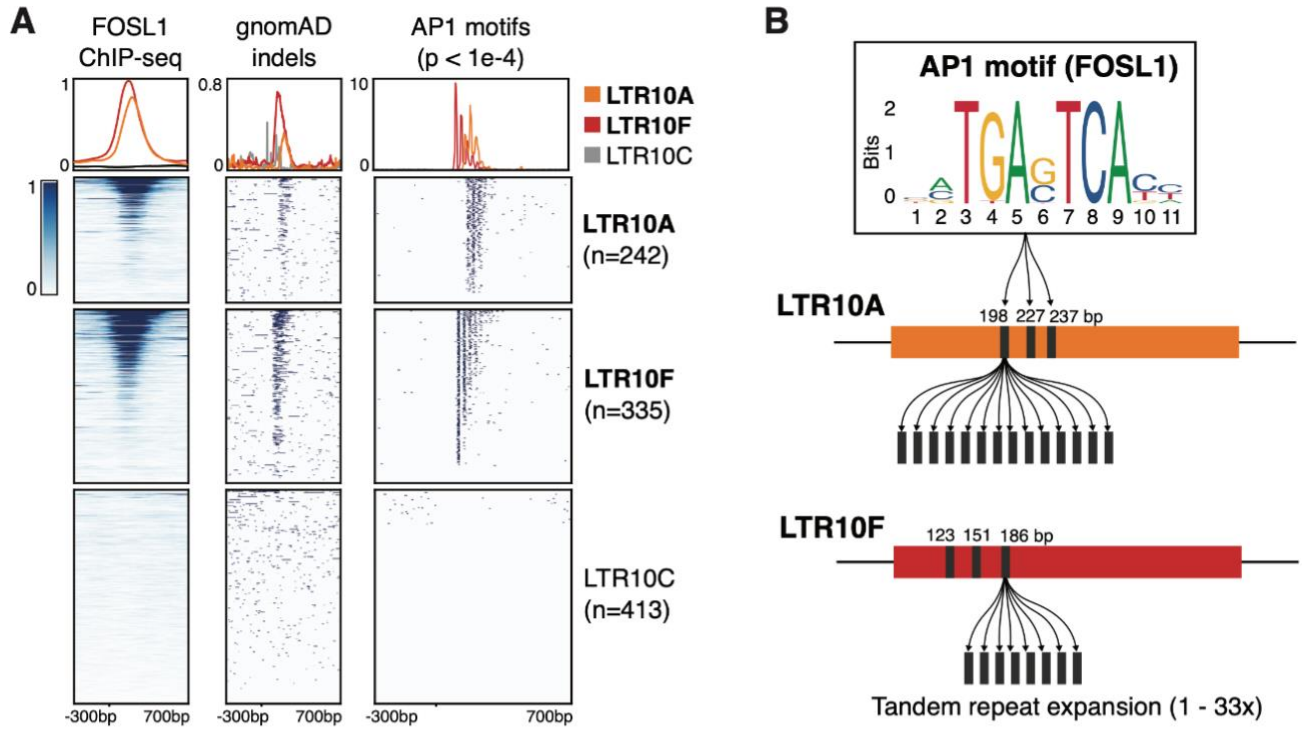
443 We transplanted either control HCT116 cells or cells harboring a homozygous deletion of
444 LTR10.XRCC4 into athymic nude mice, and subjected the mice to 8 Gy irradiation or mock
445 irradiation. Tumors derived from both control and knockout cells showed growth inhibition in
446 response to irradiation (Fig 5E-G, Supp Fig S5E-G, Supp Table 18). However,
447 LTR10.XRCC4 knockout tumors showed more significant overall tumor growth inhibition by
448 irradiation (Fig 5E), including at earlier timepoints (Fig 5F, 5G). No significant toxicities were
449 seen in animal weights. While the difference seen in knockout tumors was modest, these
450 results are consistent with previous studies in other cancers showing increased sensitivity
451 to radiation in tumor xenografts when *XRCC4* is silenced⁵⁵, and implicate a role for the
452 LTR10.XRCC4 enhancer in regulating a clinically relevant tumor phenotype.

453 454 *LTR10 elements contain highly mutable VNTRs*

455 Lastly, we investigated variation at LTR10 elements across 15,708 human genomes profiled
456 by the Genome Aggregation database (gnomAD)⁷⁰. All LTR10 insertions are fixed, but we
457 observed an unexpected enrichment of >10bp indel structural variants affecting the AP1
458 motif region specific to LTR10A and LTR10F, but not other LTR10 subfamilies such as
459 LTR10C (Fig 6A). Further sequence inspection revealed that LTR10A and LTR10F elements
460 contain an internal variable number of tandem repeats (VNTR) region, composed of a 28-
461 30 bp sequence that includes the AP1 motif (Fig 6B, Supp Fig S6A). Individual LTR10

462 elements show a wide range of regulatory potential in HCT116 cells, as approximated by
463 peak scores of H3K27ac CUT&RUN and FOSL1 ChIPseq (Supp Fig S6B, Supp Table 19),
464 and demonstrated by the CRISPR-validated LTR10 enhancers. We speculate that the
465 number of AP1 motifs within LTR10 elements may influence their regulatory potential.
466 LTR10 elements annotated in the reference genome show extensive variation in tandem
467 repeat length, with up to 33 copies of the AP1 motif (Supp Fig S6C, Supp Table 20). The
468 number of motifs strongly correlates with H3K27ac and FOSL1 binding activity in HCT116
469 cells (Supp Fig S6C), suggesting that tandem repeat length affects AP1-dependent
470 regulation of individual elements. Across the human population, LTR10A and LTR10F
471 elements harbor many rare and common indel structural variants of lengths that follow a 28-
472 30 bp periodicity, and this pattern is absent in LTR10C elements which lack the tandem
473 repeat region (Fig 6C, 6D). These elevated levels of polymorphism across copies and
474 individuals are characteristic of unstable tandem repeat regions ⁷¹, and suggest that LTR10
475 VNTR regions may be a common source of genomic regulatory variation.

476
477 Accurately genotyping tandem repeat length polymorphisms remains a major challenge
478 using short-read data, therefore we validated the presence of LTR10 VNTR polymorphisms
479 using structural variant calls generated from long-read whole-genome sequences from 15
480 individuals ⁷². We recovered indel structural variants within 24 distinct LTR10A and LTR10F
481 elements, which also showed 28-30 bp periodicity (Fig 6D, Supp Fig S6D). We confirmed
482 the presence of additional LTR10 VNTR indels using a separate long-read dataset from 25
483 Asian individuals (Fig 6E, Supp Fig S6E-J) ⁷³. At the LTR10.ATG12 locus, we observed
484 multiple indels supported by both short-read and long-read data that are predicted to affect
485 AP1 motif copy number (Fig 6E, Supp Fig S6F). At a genome-wide level, LTR10 elements
486 were a significantly enriched source of long-read indels, despite being fixed in the population
487 (Fig 6F). Therefore, expansions or contractions within LTR10 VNTR regions are an
488 underappreciated source of germline genetic variation that could underlie regulatory
489 variation, consistent with polymorphisms recently reported in the VNTR region of SINE-
490 VNTR-Alu (SVA) elements ⁷⁴.



492 **Figure 6: LTR10 repeat instability and polymorphism.** (A) Heatmap of FOSL1 ChIP-Seq, gnomAD indels
493 between 10-300bp in length, and AP1 motif matches ($p < 1e-4$) across LTR10A, LTR10F, and LTR10C
494 elements. Overlapping elements were removed, retaining 990 LTR10 elements total across the three
495 subfamilies. FOSL1 ChIP-seq was obtained from GSE32465. (B) Schematic of variable number tandem repeat
496 (VNTR) region within LTR10A and LTR10F elements. (C) Scatterplot of high-confidence gnomAD indels
497 between 10-300 bp in length detected in LTR10A, LTR10F, or LTR10C subfamilies. Each indel is plotted by
498 its length and allele frequency. (D) As in (C) but using long-read supported data. (E) Genome browser
499 screenshot of LTR10.ATG12 showing AP1 motifs, long-read indels (58bp deletion reported by Quan et al.,
500 2021), and gnomAD indels. (F) GIGGLE enrichment of ERVs within long-read indels. Significantly enriched
501 ERVs are shown in red; significantly depleted ERVs are shown in blue.

502

503 Finally, we searched for evidence of tumor-specific somatic expansions within LTR10 VNTR
504 regions. We analyzed a long-read whole-genome sequencing dataset generated from
505 matched colorectal tumor and normal tissues from 20 patients⁷⁵, using Sniffles2⁷⁶ to identify
506 tumor-specific repeat expansions within LTR10 VNTR regions. After manually inspecting
507 reads at each locus, we found evidence for tumor-specific VNTR expansions at H3K27ac-
508 marked LTR10 elements in five out of 20 patients (Supp Table 21). Three patients showed
509 independent somatic expansions at the same LTR10A locus on chromosome 1 located near
510 gene *GPR137B* (Supp Fig S6K, S6L, S6O), suggesting that this locus is prone to inter-
511 individual variation at both the germline (Supp Fig S6D) and somatic level. We also found
512 evidence of tumor-specific mosaic VNTR deletions in four patients (Supp Table 21). In fact,
513 one patient with high microsatellite instability showed evidence of multiple tumor-specific
514 LTR10 variants: a predicted LTR10A VNTR expansion over 11,600 bp in length (Supp Fig
515 S6K, Supp Table 21), as well as two tumor-specific deletions at different LTR10F VNTR loci
516 (Supp Fig S6Q, S6R). While a larger cohort would be necessary to determine if these
517 expansions are enriched within tumors with microsatellite instability, these analyses provide
518 evidence that LTR10 VNTRs are subject to tumor-specific somatic expansions and
519 contractions, which can alter tumor-specific gene regulatory activity.

520

521 **DISCUSSION**

522 Our study demonstrates that oncogenic MAPK/AP1 signaling drives global epigenetic and
523 transcriptional activation of LTR10 elements in colorectal cancer and other epithelial
524 cancers. A subset of these elements act as enhancers that facilitate pathological AP1-
525 dependent transcriptional rewiring at multiple loci in cancer cells. Collectively, our data have

526 several key implications for understanding how TEs shape cancer-specific regulatory
527 networks.

528

529 First, our pan-cancer epigenomic analysis revealed multiple primate-specific ERV families
530 that are enriched within tumor-specific accessible chromatin across all 21 solid tumor types
531 profiled by TCGA ²⁷. This implicates ERVs as a pervasive source of regulatory elements that
532 shape gene regulation across most tumor types, expanding on recent studies that
533 characterized TE-derived enhancers in prostate cancer ²⁶ and acute myeloid leukemia ²⁵ as
534 well as other genomic studies profiling tumor-specific TE-derived enhancer activity in
535 different cancers ^{17,18,20,21,77}. We focused on LTR10 elements as a case example, which
536 showed recurrent epigenomic signatures of enhancer activity in epithelial cancers including
537 colorectal cancer. Both bulk and single-cell RNA-seq analysis of patient tumors revealed
538 that LTR10 elements display tumor-specific transcriptional activation in a substantial fraction
539 (~30%) of cases. While our study found that LTR10 elements are normally repressed in adult
540 somatic tissues and show largely tumor-specific enhancer activity, a recent study reported
541 that many LTR10 elements also show enhancer activity in the developing human placenta
542 ⁷⁸, consistent with the hypothesis that reactivation of placental-specific gene regulatory
543 networks may contribute to cancer pathogenesis ^{79–81}.

544

545 Using CRISPR to silence or knock out individual elements in HCT116 colorectal cancer cells,
546 we found that LTR10-derived enhancers causally drive AP1-dependent gene expression at
547 multiple loci, including genes with established roles in tumorigenesis and therapy resistance
548 including *ATG12*, *XRCC4*, and *VCAN* ^{53,55,59–61,82–84}. While we focused on LTR10 elements
549 predicted to regulate genes with established relevance to cancer, we also uncovered many
550 elements that did not have predicted gene regulatory or functional consequences, indicating
551 that LTR10 enhancer activity is not intrinsically pathological. Moreover, the regulatory activity
552 of different LTR10-derived enhancers across the genome is likely to vary across individual
553 tumors depending on the genetic and epigenetic background of the tumor and individual.
554 Nevertheless, our findings support a model where LTR10-derived enhancers are important
555 contributors to tumor-specific transcriptional dysregulation, which in some cases can
556 significantly influence tumorigenesis and therapy resistance.

557

558 Second, our work shows that ERV-derived enhancers link oncogenic AP1/MAPK signaling
559 to pathological transcriptional rewiring in colorectal cancer. Components of the MAPK
560 pathway are frequently mutated in cancers, leading to oncogenic hyperactivation of MAPK
561 signaling which promotes pathological gene expression and tumor cell proliferation ^{43,85}.
562 However, this process is poorly defined at the genomic level, and the specific regulatory
563 elements that drive AP1-dependent transcriptional dysregulation have remained uncharted.
564 Furthermore, inhibition of MAPK signaling is a common therapeutic strategy for many
565 cancers ^{86,87} including colorectal cancer ^{88,89}, but we have an incomplete understanding of
566 how MAPK inhibition alters cancer epigenomes to achieve a therapeutic effect. Our study
567 shows that oncogenic AP1/MAPK signaling results in activation of LTR10 enhancers, and
568 treatment with a MAPK inhibitor effectively silences LTR10 regulatory activity in cancer cells.
569 Therefore, the silencing of LTR10 ERV regulatory activity is an important but
570 underappreciated mechanism underlying therapeutic MAPK inhibition.

571

572 Finally, we discovered that LTR10 elements are frequently affected by tandem repeat
573 expansions that could influence their regulatory activity. Although all LTR10 insertions are
574 fixed in the human population, they contain internal tandem repeats that show high levels of
575 length polymorphism associated with repeat instability, consistent with a recent report of
576 variable-length SVA elements which also contain internal tandem repeats ⁷⁴. Germline or
577 somatic variation in AP1 motif copy number within these elements may alter cancer-specific
578 enhancer landscapes, and we found evidence that LTR10 VNTRs can be subject to somatic
579 expansions or contractions in cancer cells with microsatellite instability ⁹⁰. Our study of
580 LTR10 highlights how TEs that are normally silenced can become reactivated in cancer and
581 cause aberrant gene expression. For elements that promote pathogenesis, their restricted
582 activity in age-associated diseases like cancer may result in reduced or nearly neutral fitness
583 consequences. Therefore, the accumulation of TEs subject to epigenetic silencing may be
584 a fundamental process that shapes cancer-specific gene regulatory networks.

585

586 **METHODS**

587 Cell culture

588 The HCT116 cell line was purchased from ATCC and cultured in McCoy's 5A medium
589 supplemented with 10% FBS and 1% penicillin/streptomycin (Gibco). Cells were cultured at

590 37°C in 5% carbon dioxide. Transfections were performed using FuGENE (Promega). For
591 treatments modulating MAPK signaling, HCT116 cells were untreated or treated for 24 hrs
592 with 1 uM Cobimetinib, 100 ng/mL TNF alpha, or DMSO.

593

594 CRISPR-mediated silencing and knockout of LTR10s

595 For CRISPR-mediated silencing (e.g., CRISPRi) of select LTR10 elements and gene
596 transcription start sites (TSS), a HCT116 dCas9-KRAB-MeCP2 stable line was first
597 generated using the PiggyBac system (System Bioscience). The PiggyBac Donor plasmid,
598 PB-CAGGS-dCas9-KRAB-MeCP2 was co-transfected with the Super PiggyBac
599 transposase expression vector (SPBT) into HCT116 cells. The pB-CAGGS-dCas9-KRAB-
600 MeCP2 construct was a gift from Alejandro Chavez & George Church (Addgene plasmid #
601 110824). 24 hours post-transfection, cells were treated with Blasticidin to select for
602 integration of the dCas9 expression cassette, and selection was maintained for 10 days.
603 CRISPR gRNAs specific to the DNA elements of interest (i.e., 0 predicted off target
604 sequences) were selected using pre-computed CRISPR target guides available on the
605 UCSC Genome Browser hg38 assembly, and complementary oligos were synthesized by
606 Integrated DNA Technologies. Complementary oligos were designed to generate BstXI and
607 BlnI overhangs for cloning into PB-CRISPRia, a custom PiggyBac CRISPR gRNA
608 expression plasmid based on the lentiviral construct pCRISPRia (a gift from Jonathan
609 Weissman, Addgene plasmid # 84832). Complementary gRNA-containing oligos were
610 hybridized and phosphorylated in a single reaction, then ligated into a PB-CRISPRia
611 expression plasmid linearized with BstXI and BlnI (New England Biolabs). Chemically
612 competent Stable E. Coli (New England Biolabs) was transformed with 2 uL of each ligation
613 reaction and resulting colonies were selected for plasmid DNA isolation using the ZymoPure
614 Plasmid miniprep kit (Zymo Research). Each cloned gRNA sequence-containing PB-
615 CRISPRia plasmid was verified by Sanger sequencing (Quintara Bio).

616

617 To generate CRISPRi stable lines, PB-CRISPRia gRNA plasmids were co-transfected with
618 the PiggyBac transposase vector into the HCT116 dCas9-KRAB-MeCP2 polyclonal stable
619 line. The following number of uniquely-mapping gRNA plasmids were designed per target
620 based on the pre-computed UCSC hg38 CRISPR target track: ATG12 (1), GFP (1), FOSL1
621 (1), LTR10.ATG12 (4), LTR10.FGF2 (2), LTR10.MCPH1 (3), LTR10.MEF2D (2),

622 LTR10.XRCC4 (2). The same total amount of gRNA plasmid was used for transfections
623 involving one or multiple gRNAs. 24 hours post-transfection, cells were treated with
624 Puromycin to select for integration of the sgRNA expression cassette(s). Selection was
625 maintained for 5 days prior to transcriptional analyses.

626
627 For CRISPR-mediated knockout of LTR10.KDM6A, 2 gRNAs (1 specific to each flank of the
628 element) were identified and synthesized as sgRNAs by IDT. For CRISPR-mediated
629 knockout of LTR10.XRCC4, 4 gRNAs (two specific to each flank of the element) were
630 identified and synthesized as sgRNAs by IDT. Using IDT's AltR technology, RNP complexes
631 were generated in vitro, and electroporated into HCT116 cells using the Neon system
632 (ThermoFisher Scientific). Clonal lines were isolated using the array dilution method in a 96-
633 well plate format, and single clones were identified and screened for homozygous deletions
634 by PCR using both flanking and internal primer pairs at the expected deletion site. gRNAs
635 and PCR primers for each candidate are provided in Supp Table 22.

636
637 Cell autophagy and apoptosis assays

638 For assaying mitochondrial apoptosis, HCT116 CRISPRi cell lines were treated for 12 hours
639 with Staurosporine (STS) at 0.5 μ M or DMSO (vehicle) followed by measurement of
640 Caspase activity via the Caspase-Glo 3/7 assay (Promega). Results are representative of
641 at least 3 independent experiments. For assaying autophagy, HCT116 CRISPRi cell lines
642 were untreated or treated with Bafilomycin A at 10 nM or 100 nM for 6 hrs and 18 hrs,
643 followed by LC3B Western blotting. Results are representative of at least 3 independent
644 experiments.

645
646 Western blots

647 For ATG12 Western blots, cell lysates were prepared with MPER buffer (ThermoFisher
648 Scientific). For LC3B Western blots, cell lysates were prepared with RIPA buffer. All cell
649 lysates were resuspended in 4x NuPage LDS Sample buffer containing reducing agent
650 (ThermoFisher Scientific). For ATG12 Western blots, total protein was concentrated and
651 size-selected by passing through an Amicon Ultra 10K column (Millipore), retaining the high
652 molecular weight fraction, and 40 μ g of protein was loaded per lane. For LC3B Western
653 blots, 2 μ g total protein was loaded per lane. Antibodies used were as follows: ATG12 (cat

654 #4180T, Cell Signaling Technologies), Beta-Actin (cat # 3700T, Cell Signaling
655 Technologies), LC3B (cat # NB100-2220, Novus Biologicals). Results are representative of
656 at least 3 independent experiments.

657

658 Luciferase assay

659 Reporter assays were conducted in HCT116 cells using the secreted NanoLuc enhancer
660 activity reporter pNL3.3 (Promega) and normalized against a constitutively active firefly
661 luciferase reporter vector, pGL4.50 (Promega). LTR10 consensus sequences for
662 subfamilies LTR10A and LTR10F were downloaded from Dfam. AP1 motifs within LTR10A
663 and LTR10F were shuffled as follows: LTR10A (first two motifs): cctgagtcacc to cagccccgtta;
664 LTR10A (third motif): cttagtcacc to cagtttacc; LTR10F (all three motifs): cctgactcatt to
665 cgtatcctac. Sequences are provided in Supp Table 22. Due to their high repeat content,
666 consensus sequences were synthesized as multiple fragments (Integrated DNA
667 Technologies, Twist BioScience) and then assembled into pNL3.3 enhancer reporter
668 plasmids using Gibson Assembly (New England Biolabs). Each cloned reporter plasmid was
669 verified by Sanger sequencing (Quintara Bio). To assay reporter activity, HCT116 cells were
670 transfected with a reporter construct as well as the pGL4.50 construct constitutively
671 expressing firefly luciferase. 24 hrs after transfection, media was replaced with media
672 containing 1 μ M Cobimetinib, 100 ng/mL TNF alpha, or DMSO (vehicle). 24 hours following
673 treatment, luminescence was measured using the NanoGlo Dual Luciferase Reporter Assay
674 System (Promega). All experiments were performed with 3 treatment replicates per condition
675 in a 96-well plate format. Luminescence readings were first normalized to firefly co-
676 transfection controls, then presented as fold-change against cells transfected with an empty
677 minimal promoter pNL3.3 vector as a negative control. Results are representative of at least
678 3 independent experiments. Barplots are presented as mean \pm s.d.

679

680 Irradiation experiment

681 HCT116 control or knockout cells were irradiated using a Faxitron irradiator (Model RX-650)
682 at 0, 2, 6 or 10 Gy then left to recover for up to 5 days. Cell viability was measured by
683 CellTiter-Glo luminescence assay (Promega). Two replicates (each based on the average
684 of three CellTiter-Glo readings) were normalized to unirradiated (0 Gy) as a control.

685

686 Mouse xenograft experiment

687 All experiments were approved by the Institutional Animal Care and Use Committee of the
688 University of Colorado Anschutz Medical Campus and conducted in accordance with the
689 National Institutes of Health Guidelines for the Care and Use of Laboratory Animals. Female
690 athymic nude mice (aged 15-16 weeks at start of study) were purchased from Envigo
691 (Indianapolis, IN) and implanted subcutaneously on the hind flanks with 2.5 million cells in
692 100 ul of either HCT116 wildtype or LTR10.XRCC4 CRISPR knockout cells under isoflurane
693 anesthesia with a 23 ga ½ needle. The cell solution injected consisted of 1:1 ratio of RMPI
694 media and cultrex (Cultrex PathClear BME, Type 3 from Bio-Techne). We injected wildtype
695 or knockout cells into 40 mice (20 each, one side per mouse), then mice were randomized
696 into treatment groups (20 irradiated, 20 mock) and treatments were initiated when the
697 average tumor volume reached between 50-100 mm³. Tumor volume was calculated by
698 $\{(\text{width}^2) \times \text{length}\} \times 0.52$. Irradiation treatment consisted of 8 Gy x 3 fractions on days 2, 4,
699 14, 16, and 18. Tumor measurements were taken twice weekly using digital calipers, and
700 toxicity was monitored by measuring body weight twice weekly and the study ended at 28
701 days. Tumor growth inhibition was measured using KuLGaP⁶⁹.

702

703 RNA-seq

704 Sequencing libraries were prepared from RNA harvested from treatment or transfection
705 replicates. Total RNA was extracted using the Quick-RNA Miniprep Plus Kit (Zymo
706 Research). PolyA enrichment and library preparation was performed using the KAPA
707 BioSystems mRNA HyperPrep Kit according to the manufacturer's protocols. Briefly, 500 ng
708 of RNA was used as input, and KAPA BioSystems single-index or unique dual-index
709 adapters were added at a final concentration of 7 nM. Purified, adapter-ligated library was
710 amplified for a total of 11 cycles following the manufacturer's protocol. The final libraries
711 were pooled and sequenced on an Illumina NovaSeq 6000 (University of Colorado
712 Genomics Core) as 150 bp paired-end reads.

713

714 CUT&RUN

715 Libraries were prepared from treatment replicates. Approximately 5×10^5 viable cells were
716 used for each CUT&RUN reaction, and pulldowns were generated following the protocol
717 from⁹¹. All buffers were prepared according to the "High Ca²⁺/Low Salt" method using

718 digitonin at a final concentration of 0.05%. The following antibodies were used at the noted
719 dilutions: rabbit anti-mouse IgG (1:100), rabbit anti-H3K27ac (1:100). pA-MNase (gift from
720 Steven Henikoff) was added to each sample following primary antibody incubation at a final
721 concentration of 700 ng/mL. Chromatin digestion, release, and extraction was carried out
722 according to the standard protocol. Sequencing libraries were generated using the KAPA
723 BioSystems HyperPrep Kit according to the manufacturer's protocol with the following
724 modifications: Freshly diluted KAPA BioSystems single-index adapters were added to each
725 library at a final concentration of 9 nM. Adapter-ligated libraries underwent a double-sided
726 0.8X/1.0X cleanup using KAPA BioSystems Pure Beads. Purified, adapter-ligated libraries
727 were amplified using the following PCR cycling conditions: 45 s at 98°C, 14x(15 s at 98°C,
728 10 s at 60°C), 60 s at 72°C. Amplified libraries underwent two 1X cleanups using Pure
729 Beads. The final libraries were quantified using Qubit dsDNA High Sensitivity and
730 TapeStation 4200 HSD5000. Libraries were pooled and sequenced on an Illumina NovaSeq
731 6000 (University of Colorado Genomics Core) as 150 bp paired-end reads.

732

733 Processing of sequencing data

734 Reads obtained from our own datasets and from published studies were reprocessed using
735 a uniform analysis pipeline. FASTQ reads were evaluated using FastQC (v0.11.8) and
736 MultiQC (v1.7), then trimmed using BBDuk/BBMap (v38.05). For ATAC-seq, ChIP-seq, and
737 CUT&RUN datasets, reads were aligned to the hg38 human genome using BWA (v0.7.15)
738 and filtered for uniquely mapping reads (MAPQ > 10) with samtools (v1.10). ChIP-Seq and
739 ATAC-seq peak calls and normalized signal coverage bigwig plots were generated using
740 MACS2 (v2.1.1) (with setting --SPMR). CUT&RUN peak calls were generated using MACS2
741 in paired-end mode using a relaxed p-value threshold without background normalization (--
742 format BAMPE --pvalue 0.01 --SPMR -B --call-summits). MACS2 was also run in single-
743 end mode with additional parameters --shift -75 and --extsize 150, and Bedtools (v2.28.0)
744 was used to merge peaks across the two modes of peak calling for each sample (bedtools
745 merge with options -c 5 -o max).

746

747 RNA-seq and PRO-seq reads were aligned to hg38 using hisat2 (v2.1.0) with option --no-
748 softclip and filtered for uniquely mapping reads with samtools for MAPQ > 10. Bigwig tracks
749 were generated using the bamCoverage function of deepTools (v3.0.1), with CPM

750 normalization (ignoring chrX and chrM) and bin size 1bp. Some datasets from TCGA,
751 ENCODE, Cistrome DB and the CEMT Canadian Epigenomes Project were downloaded as
752 post-processed peaks and bigwig files.

753

754 TE colocalization analysis

755 To determine TE subfamily enrichment within regulatory regions, we used GIGGLE (v0.6.3)
756 ⁹² to generate a genomic interval index of all TE subfamilies in the hg38 human genome,
757 based on Dfam v2.0 repeat annotation (n=1315 TE subfamilies). Regulatory regions (e.g.,
758 ATAC, ChIP-Seq, or CUT&RUN peaks) were queried against the TE interval index using the
759 GIGGLE search function (-g 3209286105 -s). Results were ranked by GIGGLE enrichment
760 score, which is a composite of the product of $-\log_{10}(\text{P value})$ and $\log_2(\text{odds ratio})$.
761 Significantly enriched TE subfamilies were defined as those with at least 25 overlaps
762 between TE copies and a set of peak regions, odds ratio over 10, and GIGGLE score over
763 100 in at least one cancer type.

764

765 Defining cancer-specific regulatory elements

766 To define cancer-specific regulatory elements, we first obtained aggregate ATAC-seq
767 regions associated with each tumor type profiled by The Cancer Genome Atlas ⁹³, which
768 represent a union of recurrent ATAC-seq regions associated with each tumor type. Next, we
769 identified regulatory regions in healthy adult tissues based on chromHMM regulatory regions
770 defined by the Roadmap project. We used healthy adult tissues from categories 1_TssA,
771 6_EnhG & 7_Enh. We did not include fetal tissues (e.g. placental tissues, embryonic stem
772 cells, trophoblast stem cells) in our set of Roadmap healthy regulatory regions, due to the
773 high levels of basal ERV regulatory activity in these tissues. Finally, cancer-specific
774 regulatory regions were defined using the subtract function of bedtools (option -A) to subtract
775 Roadmap “healthy adult” regulatory regions from each cancer peak set.

776

777 Transcription factor motif analyses

778 Motif analysis of LTR10 elements was performed using the MEME suite (v5.1.0) in
779 differential enrichment mode ⁹⁴. Entire LTR10 sequences were used for the motif analysis.
780 HCT116 CUT&RUN H3K27ac-marked LTR10A/F sequences (n=144) were used as input
781 against a background set of unmarked LTR10A/F sequences (n=561), with default settings

782 other than the number of motif repetitions (Any) and the number of motifs to find (5). Each
783 discovered motif was searched for similarity to known motifs using the JASPAR 2018 non-
784 redundant DNA database with TomTom (v5.1.0). FIMO (v5.1.0) was then used to extract
785 motif frequency and hg38 genomic coordinates, with p-value threshold set to 1e-4.

786
787 Motif analysis of cancer-specific ATAC-seq peaks from 21 TCGA cancer types was likewise
788 performed using the MEME suite (v5.1.0)⁹⁴. Cancer-specific peaks for each cancer were
789 defined by subtracting away Roadmap regulatory regions from each cancer peak set, as
790 described in the previous section. The number of cancer-specific peaks for each cancer
791 were as follows: ACC (n=8123), BLCA (n=13737), BRCA (n=30494), CESC (n=2449),
792 CHOL (n=3012), COAD (n=9370), ESCA (n=12538), GBM (n=4114), HNSC (n=9441), KIRC
793 (n=4807), KIRP (n=12315), LGG (n=3673), LIHC (n=8469), LUAD (n=16862), LUSC
794 (n=15143), MESO (n=5275), PCPG (n=7891), PRAD (n=12130), SKCM (n=13710), STAD
795 (n=11222), THCA (n=9991). Bedtools (v2.28.0) getfasta was used to convert the BED format
796 peak files to FASTA format, and all nucleotides were converted to uppercase letters. MEME-
797 ChIP (v5.1.0) was then run on each cancer-specific FASTA file, with settings -ccut 100
798 (maximum size of a sequence before it is cut down to a centered section), -order 1 (to set
799 the order of the Markov background model that is generated from the sequences), -meme-
800 mod anr (to allow any number of motif repetitions), -meme-minw 6 (minimum motif width), -
801 meme-maxw 20 (maximum motif width), -meme-nmotifs 10 (maximum number of motifs to
802 find), and the JASPAR 2018 non-redundant motif database. The output from CentriMo was
803 used to obtain the AP1 motif p-value for each cancer type (i.e. adjusted p-value for motif id
804 MA0477.1, alt id FOSL1).

805

806 Differential analysis using DESeq2

807 For RNA-seq samples, gene count tables were generated using featureCounts from the
808 subread (v1.6.2) package with the GENCODE v34 annotation gtf to estimate counts at the
809 gene level, over each exon (including -p to count fragments instead of reads for paired-end
810 reads; -O to assign reads to their overlapping meta-features; -s 2 to specify reverse-
811 strandedness; -t exon to specify the feature type; -g gene_id to specify the attribute type).

812

813 To quantify TE expression at the subfamily level, RNA-seq samples were first re-aligned to
814 hg38 using hisat2 with -k 100 to allow multi-mapping reads and --no-softclip to disable soft-
815 clipping of reads. Tetranscripts (v2.1.4) was then used in multi-mapping mode with the
816 GENCODE v34 annotation gtf and hg38 GENCODE TE gtf to assign count values to both
817 genes and TE elements.

818
819 For H3K27ac CUT&RUN samples, bedtools multicov was used to generate a count table of
820 the number of aligned reads that overlap MACS2-defined peak regions. The top 20,000
821 peaks were extracted from each sample and merged (using bedtools merge with -d 100) to
822 produce the peak file used as input to bedtools multicov.

823
824 All count tables were processed with DEseq2 (v1.32.0). Normalized count values were
825 calculated using the default DEseq2 transformation. R packages ggplot2 (v3.3.2), ggrepel
826 (v0.8.2) and apegglm (v1.8.0) were used to visualize differentially expressed genes and TEs.
827 The same DEseq2 analyses were used to identify differentially enriched peak regions
828 between H3K27ac CUT&RUN samples (e.g. in response to MAPK treatment). Significantly
829 differentially enriched regions were queried against the GIGGLE index of human repeats to
830 identify over-represented TE subfamilies.

831
832 Reanalysis of patient-derived bulk RNA-seq tumor/normal colon datasets

833 BAM files of matched tumor/normal RNA-seq datasets from 38 de-identified patients with
834 colon adenocarcinomas were downloaded from TCGA-COAD using the GDC Data Transfer
835 Client with a restricted access token. Each patient had one normal colon sample and one
836 colorectal tumor sample. Gene and TE counts were assigned using Tetranscripts (v2.1.4)
837 in multi-mapping mode, as above, with the GENCODE v34 annotation gtf and hg38
838 GENCODE TE gtf. Count tables were processed using DEseq2 (v1.32.0) and normalized
839 count values were calculated using the multi-factor DEseq2 design of
840 ~Patient.ID+Condition, where Condition was either Primary Tumor or Solid Normal Tissue.
841 Potential outliers were identified using principal component analysis based on gene counts
842 (e.g. see Supp Fig S2F), but all samples were retained for downstream analysis. R packages
843 ggplot2 (v3.3.2), ggrepel (v0.8.2) and apegglm (v1.8.0) were used to visualize differentially
844 expressed genes and TEs.

845

846 Similarly, in order to perform correlative studies between LTR10 activity and tumor mutations
847 or patient survival rates, RNA-seq BAM files from 358 patient-derived tumor samples were
848 obtained from TCGA-COAD controlled access data. The steps above were repeated for
849 each tumor sample to quantify transcription of LTR10 subfamilies. KRAS mutation status
850 and survival status for each patient were derived from the TCGA-COAD patient metadata.

851

852 Reanalysis of patient-derived single cell RNA-seq tumor/normal colon datasets

853 Single cell RNA-seq datasets of matched tumor/normal colon from 36 de-identified patients
854 with colon adenocarcinomas from Pelka et al (2021) were downloaded using dbGaP
855 controlled access (phs002407.v1.p1). Only patients with both tumor and adjacent normal
856 tissue were analyzed (n=36). Raw FASTQ files for each sample were renamed according to
857 the required Cell Ranger format, then processed with Cell Ranger (v7.0.0) count function
858 using default parameters and the Cell Ranger transcriptome for the human reference
859 genome (refdata-gex-GRCh38-2020-A). The resulting BAM files were filtered to remove
860 lines without cell barcodes using samtools (v1.10). scTE (v1.0) was used to remap reads to
861 both genes and TEs, using the provided hg38 index and default parameters except for -p 8
862 (number of threads to use), --hdf5 True (to save the output as a .h5ad formatted file), and -
863 CB CB -UMI UB (to specify that the BAM file was generated by Cell Ranger, with cell
864 barcodes and UMI integrated into the read 'CB:Z' or 'UB:Z' tag).

865

866 Output h5ad files were processed using Scanpy (v1.9.1) in a customized scRNA-seq
867 workflow. Each patient was processed separately. Cell barcodes were excluded if they
868 satisfied any of the following criteria: (1) fewer than 1200 reads, (2) fewer than 750 genes,
869 (3) more than 25% of UMIs mapping to the mitochondrial genome. Genes and TEs were
870 excluded if their expression level was deemed “undetectable”, i.e. at least two cells had to
871 contain at least 5 reads from the gene/TE. Tumor and normal samples from the same patient
872 were merged after filtering and quality control, retaining the tissue of origin (T vs N)
873 information.

874

875 For each patient, the filtered and merged data was normalized to 10,000 reads per cell, log-
876 transformed, and then clustered. Dimensionality reduction was performed using principal

877 component analysis (log=True, n_pcs=40), tSNE (perplexity=30, learning_rate=1000,
878 random_state=0, n_pcs=40), and UMAP (n_neighbors=30, n_pcs=40, min_dist=0.8,
879 spread=1, random_state=0, maxiter=100). Leiden clustering (resolution=0.75) was used to
880 assign cells to clusters, and cell clusters with less than 20 cells were excluded from final
881 UMAP visualizations. Cell types were annotated using the PanglaoDB database⁹⁵ of gene
882 expression markers, with manual verification.

883

884 Prediction of LTR10 enhancer gene targets

885 LTR10 elements were initially prioritized for CRISPR silencing or deletion based on
886 enhancer predictions from the Activity-by-Contact (ABC) model⁴⁸. Publicly available
887 HCT116 ATACseq (GEO accession GSM3593802) and in-house HCT116 H3K27ac
888 CUT&RUN were used as input to the ABC pipeline, as well as the provided averaged human
889 cell line HiC file. Predicted enhancer regions with an ABC interaction score over 0.001 were
890 intersected with H3K27ac-marked LTR10A/F elements. Putative LTR10 enhancers were
891 then checked for proximity (e.g. within 1.5Mb) to FOSL1-regulated genes (i.e. genes that
892 were significantly downregulated by FOSL1 knockdown), or MAPK-regulated genes (i.e.
893 genes that were significantly affected by MAPK treatments Cobimetinib and TNF-alpha,
894 based on inhouse RNAseq).

895

896 Evolutionary analysis of LTR10 sequences

897 Genomic coordinates of LTR10 elements in the hg38 human genome were obtained from
898 Dfam v2.0, based on RepeatMasker v4.0.6 repeat annotation. The nucleotide sequence of
899 each LTR10 element was extracted using the getfasta function from bedtools (using -name+
900 to include coordinates in the header and -s for strand specificity). VSEARCH (v2.14.1) was
901 used to set a minimum length threshold of 200bp for LTR10 sequences (-sortbylength -
902 minseqlength 200), prior to alignment. MUSCLE (v3.8.1551) was used to align the remaining
903 sequences. Jalview (v2.11.1.4) was used to perform a principal component analysis on
904 pairwise similarity scores derived from the multiple sequence alignment.

905

906 To confirm that LTR10 elements can be uniquely mapped, all individual LTR10A/F
907 sequences were clustered at 99% identity (-qmask none -id 0.99) with VSEARCH (v2.14.1).

908 No clusters contained more than one sequence, indicating that no identical LTR10A/F copies
909 exist within the human genome.

910

911 LTR10 consensus sequences representing each LTR10 subfamily (A-G) were downloaded
912 from Dfam v2.0. Sequences were concatenated into one FASTA file and aligned using
913 MUSCLE. FastTree was used to infer a maximum likelihood phylogeny representing the
914 LTR10 subfamily relationships.

915

916 The phylogeny of known primate relationships was obtained from TimeTree ⁹⁶ and the
917 HERV-I insertion estimate was confirmed based on the presence or absence of LTR10
918 sequences across mammals, using BLAST (v2.7.1+) ⁹⁷.

919

920 VNTR identification

921 gnomAD (v3.1) VCF files for each hg38 chromosome were filtered for high-confidence indels
922 (FILTER=PASS) using the query function of bcftools (v1.8) with format parameter -
923 'f%CHROM\t%POS0\t%END\t%ID\t%REF\t%ALT\t%AF\t%TYPE\tFILTER=%FILTER\n'.

924 The remaining indels were then subsetted by size to retain insertions or deletions between
925 10 to 300bp in length. Chromosome VCFs were concatenated into one whole genome BED
926 file. Bedtools (v2.28.0) was used to intersect the indel BED file with LTR10 elements from
927 each subfamily, based on Dfam (v2.0) repeat annotation.

928

929 Indels from additional short- and long-read datasets were likewise filtered by variant type
930 (INS or DEL) and indel length (10-300bp for short reads; 50-300bp for long reads, since the
931 minimum length reported by long-read SV callers is 50bp). Filtered VCFs were then
932 intersected with LTR10 elements using bedtools (v2.28.0). Deletion length versus allele
933 frequency was plotted for each subfamily, for each separate dataset. VNTR regions within
934 LTR10 elements were also intersected with GTEx v8 fine-mapped CAVIAR and DAP-G cis-
935 eQTL files ⁹⁸, again using bedtools (v2.28.0).

936

937 To identify tumor-specific VNTR expansions or contractions, we downloaded a long-read
938 whole-genome nanopore sequencing dataset generated from matched tumor/normal tissues
939 from 20 patients with advanced colorectal adenocarcinomas ⁷⁵. For each sample, we used

940 minimap2⁹⁹ (v2.22-r1101) to align reads to the hg38 reference genome, with parameters -
941 a to generate output in SAM format, -x map-ont to specify nanopore input reads, -t 4 to set
942 the number of threads to 4, and -Y to use soft clipping for supplementary alignments. We
943 then samtools (v1.10) to generate sorted BAM files, with commands samtools view -bS to
944 convert from SAM to BAM format, samtools sort (default parameters) to sort reads by
945 coordinate, and samtools index (default parameters) to generate a BAM index file for each
946 BAM. We then used Sniffles2 (v2.0.7)⁷⁶ to identify tumor-specific SVs within LTR10 VNTR
947 regions. For each tumor/normal pair, we called SVs using both the default parameters
948 (optimized for germline variants), and then again using the --non-germline parameter for the
949 tumor sample only (optimized for detecting low frequency or mosaic variants). The reference
950 genome was set to hg38 and --tandem-repeats were annotated using the Sniffles-provided
951 human_GRCh38_no_alt_analysis_set.trf.bed file. Sniffles was run with the -snf option to
952 save candidate SVs to the SNF binary file, per sample. For each patient, tumor and normal
953 SNF files were then merged using the Sniffles population verge with --vcf to specify VCF
954 output format. All VCF output files were intersected with LTR10 VNTR regions using
955 bedtools (v2.28.0). For each patient, tumor-specific variants were extracted using the
956 SUPP_VEC tag in the INFO field of the output VCFs (i.e. by extracting all SVs with
957 SUPP_VEC=01, which signifies absence in the normal sample and presence in the tumor
958 sample). Finally, for each called insertion or deletion, we manually inspected aligned reads
959 using the UCSC genome browser to confirm differences between the tumor and normal
960 samples.

961

962 **DATA AVAILABILITY**

963 High-throughput sequencing data (RNA-seq, CUT&RUN) has been deposited in the Gene
964 Expression Omnibus (GEO) with the accession code GSE186619. GSE IDs of public
965 datasets used in this study are listed in the figure legends and GitHub repository. The
966 following databases were also used: Cistrome DB (downloaded Feb 2019),
967 Roadmap/ENCODE (downloaded Feb 2019), The Cancer Genome Atlas (downloaded Sept
968 2019), CEMT Canadian Epigenome Project (downloaded July 2020), Dfam 2.0 and gnomAD
969 v3.1.

970

971 **CODE AVAILABILITY**

972 Source code and workflows are available on GitHub:
973 https://github.com/atmaivancevic/ERV_cancer_enhancers

974

975 **COMPETING INTEREST STATEMENT**

976 The authors declare no competing interests.

977

978 **REFERENCES**

- 979 1. You, J. S. & Jones, P. A. Cancer genetics and epigenetics: two sides of the same
980 coin? *Cancer Cell* **22**, 9–20 (2012).
- 981 2. Rheinbay, E., Nielsen, M. M., Abascal, F., Wala, J. A., Shapira, O., Tiao, G.,
982 Hornshøj, H., Hess, J. M., Juul, R. I., Lin, Z., Feuerbach, L., Sabarinathan, R.,
983 Madsen, T., Kim, J., Mularoni, L., Shuai, S., Lanzós, A., Herrmann, C., Maruvka, Y.
984 E., Shen, C., Amin, S. B., Bandopadhyay, P., Bertl, J., Boroevich, K. A., Busanovich,
985 J., Carlevaro-Fita, J., Chakravarty, D., Chan, C. W. Y., Craft, D., Dhingra, P.,
986 Diamanti, K., Fonseca, N. A., Gonzalez-Perez, A., Guo, Q., Hamilton, M. P.,
987 Haradhvala, N. J., Hong, C., Isaev, K., Johnson, T. A., Juul, M., Kahles, A.,
988 Kahraman, A., Kim, Y., Komorowski, J., Kumar, K., Kumar, S., Lee, D., Lehmann, K.-
989 V., Li, Y., Liu, E. M., Lochovsky, L., Park, K., Pich, O., Roberts, N. D., Saksena, G.,
990 Schumacher, S. E., Sidiropoulos, N., Sieverling, L., Sinnott-Armstrong, N., Stewart,
991 C., Tamborero, D., Tubio, J. M. C., Umer, H. M., Uusküla-Reimand, L., Wadelius, C.,
992 Wadi, L., Yao, X., Zhang, C.-Z., Zhang, J., Haber, J. E., Hobolth, A., Imielinski, M.,
993 Kellis, M., Lawrence, M. S., von Mering, C., Nakagawa, H., Raphael, B. J., Rubin, M.
994 A., Sander, C., Stein, L. D., Stuart, J. M., Tsunoda, T., Wheeler, D. A., Johnson, R.,
995 Reimand, J., Gerstein, M., Khurana, E., Campbell, P. J., López-Bigas, N., PCAWG
996 Drivers and Functional Interpretation Working Group, PCAWG Structural Variation

- 997 Working Group, Weischenfeldt, J., Beroukhim, R., Martincorena, I., Pedersen, J. S.,
998 Getz, G. & PCAWG Consortium. Analyses of non-coding somatic drivers in 2,658
999 cancer whole genomes. *Nature* **578**, 102–111 (2020).
- 1000 3. Cohen, A. J., Saiakhova, A., Corradin, O., Luppino, J. M., Lovrenert, K., Bartels, C. F.,
1001 Morrow, J. J., Mack, S. C., Dhillon, G., Beard, L., Myeroff, L., Kalady, M. F., Willis, J.,
1002 Bradner, J. E., Keri, R. A., Berger, N. A., Pruett-Miller, S. M., Markowitz, S. D. &
1003 Scacheri, P. C. Hotspots of aberrant enhancer activity punctuate the colorectal cancer
1004 epigenome. *Nat. Commun.* **8**, 14400 (2017).
- 1005 4. Chapuy, B., McKeown, M. R., Lin, C. Y., Monti, S., Roemer, M. G. M., Qi, J., Rahl, P.
1006 B., Sun, H. H., Yeda, K. T., Doench, J. G., Reichert, E., Kung, A. L., Rodig, S. J.,
1007 Young, R. A., Shipp, M. A. & Bradner, J. E. Discovery and characterization of super-
1008 enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**,
1009 777–790 (2013).
- 1010 5. Roe, J.-S., Hwang, C.-I., Somerville, T. D. D., Milazzo, J. P., Lee, E. J., Da Silva, B.,
1011 Maiorino, L., Tiriach, H., Young, C. M., Miyabayashi, K., Filippini, D., Creighton, B.,
1012 Burkhart, R. A., Buscaglia, J. M., Kim, E. J., Grem, J. L., Lazenby, A. J.,
1013 Grunkemeyer, J. A., Hollingsworth, M. A., Grandgenett, P. M., Egeblad, M., Park, Y.,
1014 Tuveson, D. A. & Vakoc, C. R. Enhancer Reprogramming Promotes Pancreatic
1015 Cancer Metastasis. *Cell* **170**, 875–888.e20 (2017).
- 1016 6. Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I.,
1017 Bradner, J. E. & Young, R. A. Convergence of developmental and oncogenic signaling
1018 pathways at transcriptional super-enhancers. *Mol. Cell* **58**, 362–370 (2015).
- 1019 7. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell* **168**,
1020 629–643 (2017).

- 1021 8. Sur, I. & Taipale, J. The role of enhancers in cancer. *Nat. Rev. Cancer* **16**, 483–493
1022 (2016).
- 1023 9. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks
1024 of cancer. *Science* **357**, (2017).
- 1025 10. Jansz, N. & Faulkner, G. J. Endogenous retroviruses in the origins and treatment of
1026 cancer. *Genome Biol.* **22**, 147 (2021).
- 1027 11. Kong, Y., Rose, C. M., Cass, A. A., Williams, A. G., Darwish, M., Lianoglou, S.,
1028 Haverty, P. M., Tong, A.-J., Blanchette, C., Albert, M. L., Mellman, I., Bourgon, R.,
1029 Greally, J., Jhunjhunwala, S. & Chen-Harris, H. Transposable element expression in
1030 tumors is associated with immune infiltration and increased antigenicity. *Nat.*
1031 *Commun.* **10**, 5228 (2019).
- 1032 12. Shukla, R., Upton, K. R., Muñoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T.,
1033 Brennan, P. M., Baillie, J. K., Collino, A., Ghisletti, S., Sinha, S., Iannelli, F., Radaelli,
1034 E., Dos Santos, A., Rapoud, D., Guettier, C., Samuel, D., Natoli, G., Carninci, P.,
1035 Ciccarelli, F. D., Garcia-Perez, J. L., Faivre, J. & Faulkner, G. J. Endogenous
1036 retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell*
1037 **153**, 101–111 (2013).
- 1038 13. Rodriguez-Martin, B., Alvarez, E. G., Baez-Ortega, A., Zamora, J., Supek, F.,
1039 Demeulemeester, J., Santamarina, M., Ju, Y. S., Temes, J., Garcia-Souto, D.,
1040 Detering, H., Li, Y., Rodriguez-Castro, J., Dueso-Barroso, A., Bruzos, A. L., Dentro, S.
1041 C., Blanco, M. G., Contino, G., Ardeljan, D., Tojo, M., Roberts, N. D., Zumalave, S.,
1042 Edwards, P. A. W., Weischenfeldt, J., Puiggròs, M., Chong, Z., Chen, K., Lee, E. A.,
1043 Wala, J. A., Raine, K., Butler, A., Waszak, S. M., Navarro, F. C. P., Schumacher, S.
1044 E., Monlong, J., Maura, F., Bolli, N., Bourque, G., Gerstein, M., Park, P. J., Wedge, D.

- 1045 C., Beroukhir, R., Torrents, D., Korbel, J. O., Martincorena, I., Fitzgerald, R. C., Van
1046 Loo, P., Kazazian, H. H., Burns, K. H., PCAWG Structural Variation Working Group,
1047 Campbell, P. J., Tubio, J. M. C. & PCAWG Consortium. Pan-cancer analysis of whole
1048 genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat.*
1049 *Genet.* **52**, 306–319 (2020).
- 1050 14. Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M. S., Hruban, R. H.,
1051 Iacobuzio-Donahue, C. A., Maitra, A., Torbenson, M. S., Goggins, M., Shih, I.-M.,
1052 Duffield, A. S., Montgomery, E. A., Gabrielson, E., Netto, G. J., Lotan, T. L., De
1053 Marzo, A. M., Westra, W., Binder, Z. A., Orr, B. A., Gallia, G. L., Eberhart, C. G.,
1054 Boeke, J. D., Harris, C. R. & Burns, K. H. Long interspersed element-1 protein
1055 expression is a hallmark of many human cancers. *Am. J. Pathol.* **184**, 1280–1286
1056 (2014).
- 1057 15. Babaian, A. & Mager, D. L. Endogenous retroviral promoter exaptation in human
1058 cancer. *Mob. DNA* **7**, 24 (2016).
- 1059 16. Yu, C., Lei, X., Chen, F., Mao, S., Lv, L., Liu, H., Hu, X., Wang, R., Shen, L., Zhang,
1060 N., Meng, Y., Shen, Y., Chen, J., Li, P., Huang, S., Lin, C., Zhang, Z. & Yuan, K.
1061 ARID1A loss derepresses a group of human endogenous retrovirus-H loci to modulate
1062 BRD4-dependent transcription. *Nat. Commun.* **13**, 3501 (2022).
- 1063 17. Babaian, A., Romanish, M. T., Gagnier, L., Kuo, L. Y., Karimi, M. M., Steidl, C. &
1064 Mager, D. L. Onco-exaptation of an endogenous retroviral LTR drives IRF5
1065 expression in Hodgkin lymphoma. *Oncogene* **35**, 2542–2546 (2016).
- 1066 18. Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K.,
1067 Bouhlel, M. A., Richter, J., Soler, E., Stadhouders, R., Jöhrens, K., Wurster, K. D.,
1068 Callen, D. F., Harte, M. F., Giefing, M., Barlow, R., Stein, H., Anagnostopoulos, I.,

- 1069 Janz, M., Cockerill, P. N., Siebert, R., Dörken, B., Bonifer, C. & Mathas, S.
1070 Derepression of an endogenous long terminal repeat activates the CSF1R proto-
1071 oncogene in human lymphoma. *Nat. Med.* **16**, 571–579 (2010).
- 1072 19. Edginton-White, B., Cauchy, P., Assi, S. A., Hartmann, S., Riggs, A. G., Mathas, S.,
1073 Cockerill, P. N. & Bonifer, C. Global long terminal repeat activation participates in
1074 establishing the unique gene expression programme of classical Hodgkin lymphoma.
1075 *Leukemia* **33**, 1463–1474 (2019).
- 1076 20. Jang, H. S., Shah, N. M., Du, A. Y., Dailey, Z. Z., Pehrsson, E. C., Godoy, P. M.,
1077 Zhang, D., Li, D., Xing, X., Kim, S., O'Donnell, D., Gordon, J. I. & Wang, T.
1078 Transposable elements drive widespread expression of oncogenes in human cancers.
1079 *Nature Genetics* **51**, 611–617 Preprint at <https://doi.org/10.1038/s41588-019-0373-3>
1080 (2019)
- 1081 21. Attig, J., Pape, J., Doglio, L., Kazachenka, A., Ottina, E., Young, G. R., Enfield, K. S.,
1082 Aramburu, I. V., Ng, K. W., Faulkner, N., Bolland, W., Papayannopoulos, V., Swanton,
1083 C. & Kassiotis, G. Human endogenous retrovirus onco-exaptation counters cancer cell
1084 senescence through Calbindin. *J. Clin. Invest.* (2023). doi:10.1172/JCI164397
- 1085 22. Wang, T., Zeng, J., Lowe, C. B., Sellers, R. G., Salama, S. R., Yang, M., Burgess, S.
1086 M., Brachmann, R. K. & Haussler, D. Species-specific endogenous retroviruses shape
1087 the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl.*
1088 *Acad. Sci. U. S. A.* **104**, 18613–18618 (2007).
- 1089 23. Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M. P. & Wang, T.
1090 Widespread contribution of transposable elements to the innovation of gene regulatory
1091 networks. *Genome Research* **24**, 1963–1976 Preprint at
1092 <https://doi.org/10.1101/gr.168872.113> (2014)

- 1093 24. Jacques, P.-É., Jeyakani, J. & Bourque, G. The majority of primate-specific regulatory
1094 sequences are derived from transposable elements. *PLoS Genet.* **9**, e1003504
1095 (2013).
- 1096 25. Deniz, Ö., Ahmed, M., Todd, C. D., Rio-Machin, A., Dawson, M. A. & Branco, M. R.
1097 Endogenous retroviruses are a source of enhancers with oncogenic potential in acute
1098 myeloid leukaemia. *Nat. Commun.* **11**, 3506 (2020).
- 1099 26. Grillo, G., Keshavarzian, T., Linder, S., Arlidge, C., Mout, L., Nand, A., Teng, M.,
1100 Qamra, A., Zhou, S., Kron, K. J., Murison, A., Hawley, J. R., Fraser, M., van der
1101 Kwast, T. H., Raj, G. V., He, H. H., Zwart, W. & Lupien, M. Transposable elements are
1102 co-opted as oncogenic regulatory elements by lineage-specific transcription factors in
1103 prostate cancer. *Cancer Discov.* (2023). doi:10.1158/2159-8290.CD-23-0331
- 1104 27. Corces, M. R., Granja, J. M., Shams, S., Louie, B. H., Seoane, J. A., Zhou, W., Silva,
1105 T. C., Groeneveld, C., Wong, C. K., Cho, S. W., Satpathy, A. T., Mumbach, M. R.,
1106 Hoadley, K. A., Robertson, A. G., Sheffield, N. C., Felau, I., Castro, M. A. A., Berman,
1107 B. P., Staudt, L. M., Zenklusen, J. C., Laird, P. W., Curtis, C., Cancer Genome Atlas
1108 Analysis Network, Greenleaf, W. J. & Chang, H. Y. The chromatin accessibility
1109 landscape of primary human cancers. *Science* **362**, (2018).
- 1110 28. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky,
1111 M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J.,
1112 Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G.,
1113 Sandstrom, R. S., Eaton, M. L., Wu, Y.-C., Pfening, A. R., Wang, X., Claussnitzer,
1114 M., Liu, Y., Coarfa, C., Harris, R. A., Shores, N., Epstein, C. B., Gjoneska, E., Leung,
1115 D., Xie, W., Hawkins, R. D., Lister, R., Hong, C., Gascard, P., Mungall, A. J., Moore,
1116 R., Chuah, E., Tam, A., Canfield, T. K., Hansen, R. S., Kaul, R., Sabo, P. J., Bansal,

- 1117 M. S., Carles, A., Dixon, J. R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni,
1118 A., Li, D., Lowdon, R., Elliott, G., Mercer, T. R., Neph, S. J., Onuchic, V., Polak, P.,
1119 Rajagopal, N., Ray, P., Sallari, R. C., Siebenthal, K. T., Sinnott-Armstrong, N. A.,
1120 Stevens, M., Thurman, R. E., Wu, J., Zhang, B., Zhou, X., Beaudet, A. E., Boyer, L.
1121 A., De Jager, P. L., Farnham, P. J., Fisher, S. J., Haussler, D., Jones, S. J. M., Li, W.,
1122 Marra, M. A., McManus, M. T., Sunyaev, S., Thomson, J. A., Tlsty, T. D., Tsai, L.-H.,
1123 Wang, W., Waterland, R. A., Zhang, M. Q., Chadwick, L. H., Bernstein, B. E., Costello,
1124 J. F., Ecker, J. R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B.,
1125 Stamatoyannopoulos, J. A., Wang, T. & Kellis, M. Integrative analysis of 111
1126 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- 1127 29. Akhtar-Zaidi, B., Cowper-Sal-lari, R., Corradin, O., Saiakhova, A., Bartels, C. F.,
1128 Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M. F., Willis, J.,
1129 Moore, J. H., Tesar, P. J., Laframboise, T., Markowitz, S., Lupien, M. & Scacheri, P.
1130 C. Epigenomic enhancer profiling defines a signature of colon cancer. *Science* **336**,
1131 736–739 (2012).
- 1132 30. Baranello, L., Wojtowicz, D., Cui, K., Devaiah, B. N., Chung, H.-J., Chan-Salis, K. Y.,
1133 Guha, R., Wilson, K., Zhang, X., Zhang, H., Piotrowski, J., Thomas, C. J., Singer, D.
1134 S., Pugh, B. F., Pommier, Y., Przytycka, T. M., Kouzine, F., Lewis, B. A., Zhao, K. &
1135 Levens, D. RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient
1136 Transcription. *Cell* **165**, 357–371 (2016).
- 1137 31. Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.-H., Brown, M., Zhang,
1138 X., Meyer, C. A. & Liu, X. S. Cistrome Data Browser: expanded datasets and new
1139 tools for gene regulatory analysis. *Nucleic Acids Res.* **47**, D729–D735 (2019).
- 1140 32. Grow, E. J., Weaver, B. D., Smith, C. M., Guo, J., Stein, P., Shadle, S. C.,

- 1141 Hendrickson, P. G., Johnson, N. E., Butterfield, R. J., Menafra, R., Kloet, S. L., van
1142 der Maarel, S. M., Williams, C. J. & Cairns, B. R. p53 convergently activates
1143 Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy
1144 cell models. *Nat. Genet.* **53**, 1207–1220 (2021).
- 1145 33. Imbeault, M., Helleboid, P.-Y. & Trono, D. KRAB zinc-finger proteins contribute to the
1146 evolution of gene regulatory networks. *Nature* **543**, 550–554 (2017).
- 1147 34. Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A.,
1148 Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., Farnham, P. J.,
1149 Hirst, M., Lander, E. S., Mikkelsen, T. S. & Thomson, J. A. The NIH Roadmap
1150 Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
- 1151 35. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the
1152 human genome. *Nature* **489**, 57–74 (2012).
- 1153 36. Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J.,
1154 Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart,
1155 R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B. & Ecker, J. R. Human DNA
1156 methylomes at base resolution show widespread epigenomic differences. *Nature* **462**,
1157 315–322 (2009).
- 1158 37. Bruno, M., Mahgoub, M. & Macfarlan, T. S. The Arms Race Between KRAB–Zinc
1159 Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. (2019).
1160 doi:10.1146/annurev-genet-112618-043717
- 1161 38. Orouji, E., Raman, A. T., Singh, A. K., Sorokin, A., Arslan, E., Ghosh, A. K., Schulz, J.,
1162 Terranova, C., Jiang, S., Tang, M., Maitituoheti, M., Callahan, S. C., Barrodia, P.,
1163 Tomczak, K., Jiang, Y., Jiang, Z., Davis, J. S., Ghosh, S., Lee, H. M., Reyes-Uribe, L.,
1164 Chang, K., Liu, Y., Chen, H., Azhdarinia, A., Morris, J., Vilar, E., Carmon, K. S.,

- 1165 Kopetz, S. E. & Rai, K. Chromatin state dynamics confers specific therapeutic
1166 strategies in enhancer subtypes of colorectal cancer. *Gut* (2021). doi:10.1136/gutjnl-
1167 2020-322835
- 1168 39. Bujold, D., Morais, D. A. de L., Gauthier, C., Côté, C., Caron, M., Kwan, T., Chen, K.
1169 C., Laperle, J., Markovits, A. N., Pastinen, T., Caron, B., Veilleux, A., Jacques, P.-É. &
1170 Bourque, G. The International Human Epigenome Consortium Data Portal. *Cell Syst*
1171 **3**, 496–499.e2 (2016).
- 1172 40. Pelka, K., Hofree, M., Chen, J. H., Sarkizova, S., Pirl, J. D., Jorgji, V., Bejnood, A.,
1173 Dionne, D., Ge, W. H., Xu, K. H., Chao, S. X., Zollinger, D. R., Lieb, D. J., Reeves, J.
1174 W., Fuhrman, C. A., Hoang, M. L., Delorey, T., Nguyen, L. T., Waldman, J., Klapholz,
1175 M., Wakiro, I., Cohen, O., Albers, J., Smillie, C. S., Cuoco, M. S., Wu, J., Su, M.-J.,
1176 Yeung, J., Vijaykumar, B., Magnuson, A. M., Asinovski, N., Moll, T., Goder-Reiser, M.
1177 N., Applebaum, A. S., Brais, L. K., DelloStritto, L. K., Denning, S. L., Phillips, S. T.,
1178 Hill, E. K., Meehan, J. K., Frederick, D. T., Sharova, T., Kanodia, A., Todres, E. Z.,
1179 Jané-Valbuena, J., Biton, M., Izar, B., Lambden, C. D., Clancy, T. E., Bleday, R.,
1180 Melnitchouk, N., Irani, J., Kunitake, H., Berger, D. L., Srivastava, A., Hornick, J. L.,
1181 Ogino, S., Rotem, A., Vigneau, S., Johnson, B. E., Corcoran, R. B., Sharpe, A. H.,
1182 Kuchroo, V. K., Ng, K., Giannakis, M., Nieman, L. T., Boland, G. M., Aguirre, A. J.,
1183 Anderson, A. C., Rozenblatt-Rosen, O., Regev, A. & Hacohen, N. Spatially organized
1184 multicellular immune hubs in human colorectal cancer. *Cell* **184**, 4734–4752.e20
1185 (2021).
- 1186 41. He, J., Babarinde, I. A., Sun, L., Xu, S., Chen, R., Shi, J., Wei, Y., Li, Y., Ma, G.,
1187 Zhuang, Q., Hutchins, A. P. & Chen, J. Identifying transposable element expression
1188 dynamics and heterogeneity during development at the single-cell level with a

- 1189 processing pipeline scTE. *Nat. Commun.* **12**, 1456 (2021).
- 1190 42. Zhu, G., Pei, L., Xia, H., Tang, Q. & Bi, F. Role of oncogenic KRAS in the prognosis,
1191 diagnosis and treatment of colorectal cancer. *Mol. Cancer* **20**, 143 (2021).
- 1192 43. Wagner, E. F. & Nebreda, A. R. Signal integration by JNK and p38 MAPK pathways in
1193 cancer development. *Nat. Rev. Cancer* **9**, 537–549 (2009).
- 1194 44. Yeo, N. C., Chavez, A., Lance-Byrne, A., Chan, Y., Menn, D., Milanova, D., Kuo, C.-
1195 C., Guo, X., Sharma, S., Tung, A., Cecchi, R. J., Tuttle, M., Pradhan, S., Lim, E. T.,
1196 Davidsohn, N., Ebrahimkhani, M. R., Collins, J. J., Lewis, N. E., Kiani, S. & Church, G.
1197 M. An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat.*
1198 *Methods* **15**, 611–616 (2018).
- 1199 45. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. Tetranscripts: a package for
1200 including transposable elements in differential expression analysis of RNA-seq
1201 datasets. *Bioinformatics* **31**, 3593–3599 (2015).
- 1202 46. Rahnamoun, H., Lu, H., Duttke, S. H., Benner, C., Glass, C. K. & Lauberth, S. M.
1203 Mutant p53 shapes the enhancer landscape of cancer cells in response to chronic
1204 immune signaling. *Nat. Commun.* **8**, 754 (2017).
- 1205 47. Koul, H. K., Pal, M. & Koul, S. Role of p38 MAP Kinase Signal Transduction in Solid
1206 Tumors. *Genes Cancer* (2013). doi:10.1177/1947601913507951
- 1207 48. Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V.,
1208 Grossman, S. R., Anyoha, R., Doughty, B. R., Patwardhan, T. A., Nguyen, T. H.,
1209 Kane, M., Perez, E. M., Durand, N. C., Lareau, C. A., Stamenova, E. K., Aiden, E. L.,
1210 Lander, E. S. & Engreitz, J. M. Activity-by-contact model of enhancer-promoter
1211 regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669
1212 (2019).

- 1213 49. Haller, M., Hock, A. K., Giampazolias, E., Oberst, A., Green, D. R., Debnath, J., Ryan,
1214 K. M., Vousden, K. H. & Tait, S. W. G. Ubiquitination and proteasomal degradation of
1215 ATG12 regulates its proapoptotic activity. *Autophagy* **10**, 2269–2278 (2014).
- 1216 50. Rubinstein, A. D., Eisenstein, M., Ber, Y., Bialik, S. & Kimchi, A. The autophagy
1217 protein Atg12 associates with antiapoptotic Bcl-2 family members to promote
1218 mitochondrial apoptosis. *Mol. Cell* **44**, 698–709 (2011).
- 1219 51. Radoshevich, L., Murrow, L., Chen, N., Fernandez, E., Roy, S., Fung, C. & Debnath,
1220 J. ATG12 conjugation to ATG3 regulates mitochondrial homeostasis and cell death.
1221 *Cell* **142**, 590–600 (2010).
- 1222 52. Hanada, T., Noda, N. N., Satomi, Y., Ichimura, Y., Fujioka, Y., Takao, T., Inagaki, F. &
1223 Ohsumi, Y. The Atg12-Atg5 Conjugate Has a Novel E3-like Activity for Protein
1224 Lipidation in Autophagy *. *J. Biol. Chem.* **282**, 37298–37302 (2007).
- 1225 53. Hu, J. L., He, G. Y., Lan, X. L., Zeng, Z. C., Guan, J., Ding, Y., Qian, X. L., Liao, W. T.,
1226 Ding, Y. Q. & Liang, L. Inhibition of ATG12-mediated autophagy by miR-214 enhances
1227 radiosensitivity in colorectal cancer. *Oncogenesis* **7**, 16 (2018).
- 1228 54. YiRen, H., YingCong, Y., Sunwu, Y., Keqin, L., Xiaochun, T., Senrui, C., Ende, C.,
1229 XiZhou, L. & Yanfan, C. Long noncoding RNA MALAT1 regulates autophagy
1230 associated chemoresistance via miR-23b-3p sequestration in gastric cancer. *Mol.*
1231 *Cancer* **16**, 174 (2017).
- 1232 55. Zheng, Z., Ng, W. L., Zhang, X., Olson, J. J., Hao, C., Curran, W. J. & Wang, Y. RNAi-
1233 mediated targeting of noncoding and coding sequences in DNA repair gene messages
1234 efficiently radiosensitizes human tumor cells. *Cancer Res.* **72**, 1221–1228 (2012).
- 1235 56. Xu, M., Huang, X., Zheng, C., Long, J., Dai, Q., Chen, Y., Lu, J., Pan, C., Yao, S. & Li,
1236 J. Platinum-Resistant Ovarian Cancer Is Vulnerable to the cJUN-XRCC4 Pathway

- 1237 Inhibition. *Cancers* **14**, (2022).
- 1238 57. Zeng, A., Wei, Z., Yan, W., Yin, J., Huang, X., Zhou, X., Li, R., Shen, F., Wu, W.,
1239 Wang, X. & You, Y. Exosomal transfer of miR-151a enhances chemosensitivity to
1240 temozolomide in drug-resistant glioblastoma. *Cancer Lett.* **436**, 10–21 (2018).
- 1241 58. Goldstein, M. & Kastan, M. B. The DNA damage response: implications for tumor
1242 responses to radiation and chemotherapy. *Annu. Rev. Med.* **66**, 129–143 (2015).
- 1243 59. Gao, Y., Ferguson, D. O., Xie, W., Manis, J. P., Sekiguchi, J., Frank, K. M.,
1244 Chaudhuri, J., Horner, J., DePinho, R. A. & Alt, F. W. Interplay of p53 and DNA-repair
1245 protein XRCC4 in tumorigenesis, genomic stability and development. *Nature* **404**,
1246 897–900 (2000).
- 1247 60. Yeung, T.-L., Leung, C. S., Wong, K.-K., Samimi, G., Thompson, M. S., Liu, J., Zaid,
1248 T. M., Ghosh, S., Birrer, M. J. & Mok, S. C. TGF- β modulates ovarian cancer invasion
1249 by upregulating CAF-derived versican in the tumor microenvironment. *Cancer Res.*
1250 **73**, 5016–5028 (2013).
- 1251 61. Papadas, A., Arauz, G., Cicala, A., Wiesner, J. & Asimakopoulos, F. Versican and
1252 Versican-matrikines in Cancer Progression, Inflammation, and Immunity. *J.*
1253 *Histochem. Cytochem.* **68**, 871–885 (2020).
- 1254 62. Zhang, Y., Zou, X., Qian, W., Weng, X., Zhang, L., Zhang, L., Wang, S., Cao, X., Ma,
1255 L., Wei, G., Wu, Y. & Hou, Z. Enhanced PAPSS2/VCAN sulfation axis is essential for
1256 Snail-mediated breast cancer cell migration and metastasis. *Cell Death Differ.* **26**,
1257 565–579 (2019).
- 1258 63. Domenzain-Reyna, C., Hernández, D., Miquel-Serra, L., Docampo, M. J., Badenas,
1259 C., Fabra, A. & Bassols, A. Structure and regulation of the versican promoter: the
1260 versican promoter is regulated by AP-1 and TCF transcription factors in invasive

- 1261 human melanoma cells. *J. Biol. Chem.* **284**, 12306–12317 (2009).
- 1262 64. Li, Z., Otevrel, T., Gao, Y., Cheng, H. L., Seed, B., Stamato, T. D., Taccioli, G. E. &
1263 Alt, F. W. The XRCC4 gene encodes a novel protein involved in DNA double-strand
1264 break repair and V(D)J recombination. *Cell* **83**, 1079–1089 (1995).
- 1265 65. Critchlow, S. E., Bowater, R. P. & Jackson, S. P. Mammalian DNA double-strand
1266 break repair protein XRCC4 interacts with DNA ligase IV. *Curr. Biol.* **7**, 588–598
1267 (1997).
- 1268 66. Katsube, T., Mori, M., Tsuji, H., Shiomi, T., Shiomi, N. & Onoda, M. Differences in
1269 sensitivity to DNA-damaging Agents between XRCC4- and Artemis-deficient human
1270 cells. *J. Radiat. Res.* **52**, 415–424 (2011).
- 1271 67. Wen, Y., Dai, G., Wang, L., Fu, K. & Zuo, S. Silencing of XRCC4 increases
1272 radiosensitivity of triple-negative breast cancer cells. *Biosci. Rep.* **39**, (2019).
- 1273 68. Leu, J.-D., Wang, B.-S., Chiu, S.-J., Chang, C.-Y., Chen, C.-C., Chen, F.-D., Avirmed,
1274 S. & Lee, Y.-J. Combining fisetin and ionizing radiation suppresses the growth of
1275 mammalian colorectal cancers in xenograft tumor models. *Oncol. Lett.* **12**, 4975–4982
1276 (2016).
- 1277 69. Ortmann, J., Rampášek, L., Tai, E., Mer, A. S., Shi, R., Stewart, E. L., Mascaux, C.,
1278 Fares, A., Pham, N.-A., Beri, G., Eeles, C., Tkachuk, D., Ho, C., Sakashita, S., Weiss,
1279 J., Jiang, X., Liu, G., Cescon, D. W., O'Brien, C. A., Guo, S., Tsao, M.-S., Haibe-
1280 Kains, B. & Goldenberg, A. Assessing therapy response in patient-derived xenografts.
1281 *Sci. Transl. Med.* **13**, eabf4969 (2021).
- 1282 70. Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q.,
1283 Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H.,
1284 Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E.

- 1285 G., Kosmicki, J. A., Walters, R. K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T.,
1286 Wang, A., Seed, C., Whiffin, N., Chong, J. X., Samocha, K. E., Pierce-Hoffman, E.,
1287 Zappala, Z., O'Donnell-Luria, A. H., Minikel, E. V., Weisburd, B., Lek, M., Ware, J. S.,
1288 Vittal, C., Armean, I. M., Bergelson, L., Cibulskis, K., Connolly, K. M., Covarrubias, M.,
1289 Donnelly, S., Ferreira, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D.,
1290 Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V.,
1291 Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski,
1292 M. E., Genome Aggregation Database Consortium, Neale, B. M., Daly, M. J. &
1293 MacArthur, D. G. The mutational constraint spectrum quantified from variation in
1294 141,456 humans. *Nature* **581**, 434–443 (2020).
- 1295 71. Gymrek, M. A genomic view of short tandem repeats. *Curr. Opin. Genet. Dev.* **44**, 9–
1296 16 (2017).
- 1297 72. Audano, P. A., Sulovari, A., Graves-Lindsay, T. A., Cantsilieris, S., Sorensen, M.,
1298 Welch, A. E., Dougherty, M. L., Nelson, B. J., Shah, A., Dutcher, S. K., Warren, W. C.,
1299 Magrini, V., McGrath, S. D., Li, Y. I., Wilson, R. K. & Eichler, E. E. Characterizing the
1300 Major Structural Variant Alleles of the Human Genome. *Cell* **176**, 663–675.e19 (2019).
- 1301 73. Quan, C., Li, Y., Liu, X., Wang, Y., Ping, J., Lu, Y. & Zhou, G. Characterization of
1302 structural variation in Tibetans reveals new evidence of high-altitude adaptation and
1303 introgression. *Genome Biol.* **22**, 159 (2021).
- 1304 74. van Bree, E. J., Guimarães, R. L. F. P., Lundberg, M., Blujdea, E. R., Rosenkrantz, J.
1305 L., White, F. T. G., Poppinga, J., Ferrer-Raventós, P., Schneider, A.-F. E., Clayton, I.,
1306 Haussler, D., Reinders, M. J. T., Holstege, H., Ewing, A. D., Moses, C. & Jacobs, F.
1307 M. J. A hidden layer of structural variation in transposable elements reveals potential
1308 genetic modifiers in human disease-risk loci. *Genome Res.* **32**, 656–670 (2022).

- 1309 75. Xu, L., Wang, X., Lu, X., Liang, F., Liu, Z., Zhang, H., Li, X., Tian, S., Wang, L. &
1310 Wang, Z. Long-read sequencing identifies novel structural variations in colorectal
1311 cancer. *PLoS Genet.* **19**, e1010514 (2023).
- 1312 76. Smolka, M., Paulin, L. F., Grochowski, C. M., Horner, D. W., Mahmoud, M., Behera,
1313 S., Kalef-Ezra, E., Gandhi, M., Hong, K., Pehlivan, D., Scholz, S. W., Carvalho, C. M.
1314 B., Proukakis, C. & Sedlazeck, F. J. Comprehensive Structural Variant Detection:
1315 From Mosaic to Population-Level. *bioRxiv* 2022.04.04.487055 (2023).
1316 doi:10.1101/2022.04.04.487055
- 1317 77. Karttunen, K., Patel, D., Xia, J., Fei, L., Palin, K., Aaltonen, L. & Sahu, B.
1318 Transposable elements as tissue-specific enhancers in cancers of endodermal
1319 lineage. *Nat. Commun.* **14**, 5313 (2023).
- 1320 78. Frost, J. M., Amante, S. M., Okae, H., Jones, E. M., Ashley, B., Lewis, R. M., Cleal, J.
1321 K., Caley, M. P., Arima, T., Maffucci, T. & Branco, M. R. Regulation of human
1322 trophoblast gene expression by endogenous retroviruses. *Nat. Struct. Mol. Biol.* **30**,
1323 527–538 (2023).
- 1324 79. Lynch-Sutherland, C. F., Chatterjee, A., Stockwell, P. A., Eccles, M. R. & Macaulay, E.
1325 C. Reawakening the Developmental Origins of Cancer Through Transposable
1326 Elements. *Front. Oncol.* **10**, 468 (2020).
- 1327 80. Costanzo, V., Bardelli, A., Siena, S. & Abrignani, S. Exploring the links between
1328 cancer and placenta development. *Open Biol.* **8**, (2018).
- 1329 81. Wagner, G. P., Kshitiz, Dighe, A. & Levchenko, A. The Coevolution of Placentation
1330 and Cancer. *Annu Rev Anim Biosci* **10**, 259–279 (2022).
- 1331 82. Rubinsztein, D. C., Codogno, P. & Levine, B. Autophagy modulation as a potential
1332 therapeutic target for diverse diseases. *Nat. Rev. Drug Discov.* **11**, 709–730 (2012).

- 1333 83. Bau, D.-T., Yang, M.-D., Tsou, Y.-A., Lin, S.-S., Wu, C.-N., Hsieh, H.-H., Wang, R.-F.,
1334 Tsai, C.-W., Chang, W.-S., Hsieh, H.-M., Sun, S.-S. & Tsai, R.-Y. Colorectal cancer
1335 and genetic polymorphism of DNA double-strand break repair gene XRCC4 in Taiwan.
1336 *Anticancer Res.* **30**, 2727–2730 (2010).
- 1337 84. Chatterjee, P., Choudhary, G. S., Alswillah, T., Xiong, X., Heston, W. D., Magi-
1338 Galluzzi, C., Zhang, J., Klein, E. A. & Almasan, A. The TMPRSS2-ERG Gene Fusion
1339 Blocks XRCC4-Mediated Nonhomologous End-Joining Repair and Radiosensitizes
1340 Prostate Cancer Cells to PARP Inhibition. *Mol. Cancer Ther.* **14**, 1896–1906 (2015).
- 1341 85. Fang, J. Y. & Richardson, B. C. The MAPK signalling pathways and colorectal cancer.
1342 *Lancet Oncol.* **6**, 322–327 (2005).
- 1343 86. Olson, J. M. & Hallahan, A. R. p38 MAP kinase: a convergence point in cancer
1344 therapy. *Trends Mol. Med.* **10**, 125–129 (2004).
- 1345 87. Santarpia, L., Lippman, S. M. & El-Naggar, A. K. Targeting the MAPK-RAS-RAF
1346 signaling pathway in cancer therapy. *Expert Opin. Ther. Targets* **16**, 103–119 (2012).
- 1347 88. Pashirzad, M., Khorasanian, R., Fard, M. M., Arjmand, M.-H., Langari, H., Khazaei,
1348 M., Soleimanpour, S., Rezayi, M., Ferns, G. A., Hassanian, S. M. & Avan, A. The
1349 therapeutic potential of MAPK/ERK inhibitors in the treatment of colorectal cancer.
1350 *Curr. Cancer Drug Targets* **21**, 932–943 (2021).
- 1351 89. Grossi, V., Peserico, A., Tezil, T. & Simone, C. p38 α MAPK pathway: a key factor in
1352 colorectal cancer therapy and chemoresistance. *World J. Gastroenterol.* **20**, 9744–
1353 9758 (2014).
- 1354 90. Hung, S., Saiakhova, A., Faber, Z. J., Bartels, C. F., Neu, D., Bayles, I., Ojo, E., Hong,
1355 E. S., Pontius, W. D., Morton, A. R., Liu, R., Kalady, M. F., Wald, D. N., Markowitz, S.
1356 & Scacheri, P. C. Mismatch repair-signature mutations activate gene enhancers

- 1357 across human colorectal cancer epigenomes. *Elife* **8**, (2019).
- 1358 91. Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN
1359 chromatin profiling tools. *Elife* **8**, (2019).
- 1360 92. Layer, R. M., Pedersen, B. S., DiSera, T., Marth, G. T., Gertz, J. & Quinlan, A. R.
1361 GIGGLE: a search engine for large-scale integrated genome analysis. *Nat. Methods*
1362 **15**, 123–126 (2018).
- 1363 93. Weinstein, J. N., The Cancer Genome Atlas Research Network, Collisson, E. A., Mills,
1364 G. B., Mills Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C. &
1365 Stuart, J. M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*
1366 **45**, 1113–1120 Preprint at <https://doi.org/10.1038/ng.2764> (2013)
- 1367 94. Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li,
1368 W. W. & Noble, W. S. MEME SUITE: tools for motif discovery and searching. *Nucleic*
1369 *Acids Res.* **37**, W202–8 (2009).
- 1370 95. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: a web server for
1371 exploration of mouse and human single-cell RNA sequencing data. *Database* **2019**,
1372 (2019).
- 1373 96. Kumar, S., Stecher, G., Suleski, M. & Blair Hedges, S. TimeTree: A Resource for
1374 Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution* **34**,
1375 1812–1819 Preprint at <https://doi.org/10.1093/molbev/msx116> (2017)
- 1376 97. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local
1377 alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- 1378 98. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across
1379 human tissues. *Science* **369**, 1318–1330 (2020).
- 1380 99. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**,

1381 3094–3100 (2018).

1382 **ACKNOWLEDGEMENTS**

1383 We thank the University of Colorado Genomics Shared Resource and BioFrontiers
1384 Computing core for technical support during this study. We thank Ben Nebenfuehr and
1385 Nausica Arnoult for assistance with the cell irradiation experiments, and Marilyn Jackson,
1386 Cameron Binns, Stephen Smoots and Adrian Dominguez for assistance with the animal
1387 studies. A.I. was supported by the National Cancer Center. E.B.C. was supported by the
1388 National Institutes of Health (1R35GM128822), the Alfred P. Sloan Foundation, the David
1389 and Lucile Packard Foundation, and the Boettcher foundation.

1390

1391 **Author contributions:** A.I. designed and conducted the bioinformatics analyses. D.M.S.
1392 and O.J. designed and conducted the CRISPR and cell irradiation experiments. S.M.B and
1393 T.M.P designed and conducted the mouse xenograft experiments. L.L.N. performed RNA
1394 extractions from colorectal cancer cell lines. A.I. and E.B.C. conceived the study, analyzed
1395 the data, and wrote the paper. B.G.B supervised the study and helped write the paper.