TITLE PAGE

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

Authors: Atma Ivancevic[1], David M. Simpson[1], Olivia M. Joyner[1], Stacey M. Bagby[2], Lily L. Nguyen[1,3], Ben G. Bitler[3], Todd M. Pitts[2], Edward B. Chuong[1]

Corresponding Author:
Edward B. Chuong
edward.chuong@colorado.edu
596 UCB
Boulder, CO 80309
303-735-8573

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

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

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

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

Running title: ERV-derived enhancers in colorectal cancer

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

Main Text:

INTRODUCTION

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

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

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

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

RESULTS

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

LTR10 elements exhibit cancer-specific regulatory activity

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

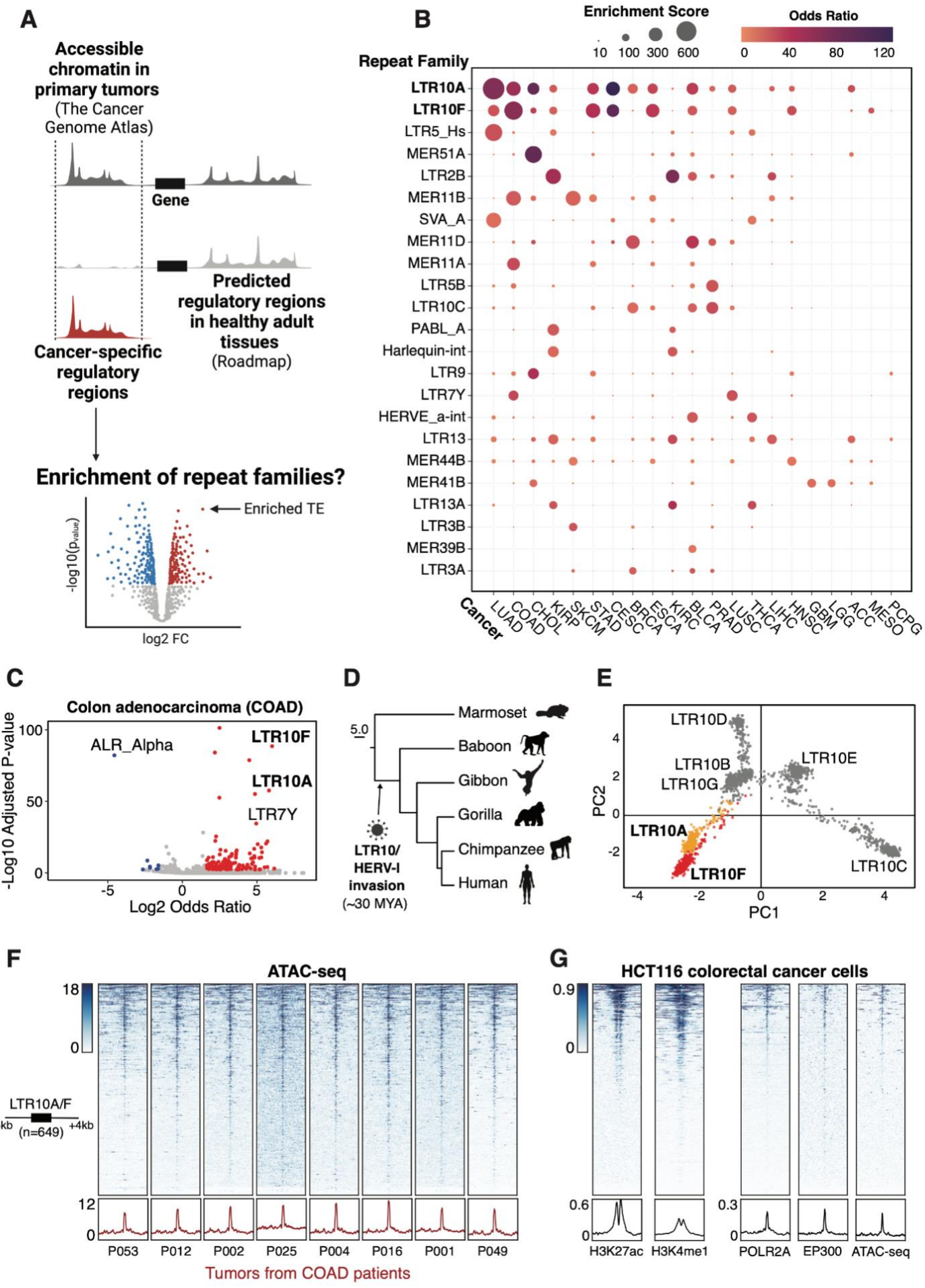


Figure 1: Pan-cancer epigenomic analysis of TE activity. (A) Pipeline to estimate TE subfamily enrichment within cancer-specific regulatory regions. Aggregate ATAC-seq maps associated with each TCGA tumor type were filtered to remove regulatory regions predicted in any normal adult tissues by Roadmap.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

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

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

specific TEs [37], LTR10 elements are normally subject to H3K9me3-mediated epigenetic silencing in somatic tissues.

LTR10 elements are bound by the AP1 transcription factor complex

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

LTR10 epigenetic and transcriptional activity is elevated in patient tumor cells

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

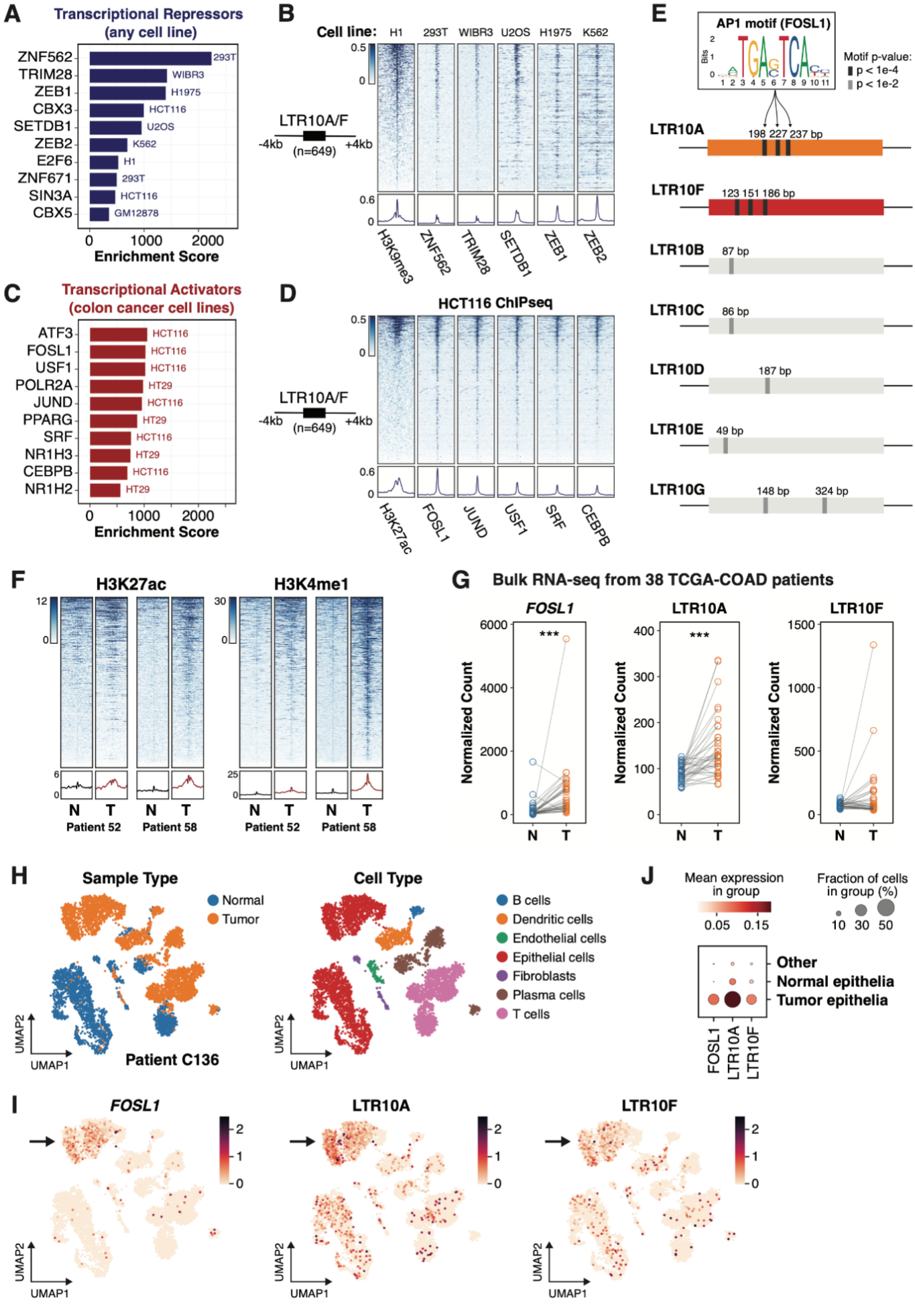


Figure 2: Regulatory activity of LTR10 in tumor and normal cells. **(A)** Transcriptional repressors associated with LTR10A/F elements, ranked by enrichment score. **(B)** Heatmap of ChIP-seq signal from H3K9me3 and repressive factors, over LTR10A/F elements. From left to right: H3K9me3 ChIP-seq (GSE16256),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

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

LTR10 transcription marks tumor-specific epithelial cells

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

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

LTR10 transcription is associated with dysregulated MAPK signaling

Our initial analyses of patient cohorts suggest that LTR10 elements become transcriptionally activated in about 30% of colorectal tumors. To determine which tumor molecular subtypes are most likely to drive LTR10 activation, we performed correlative studies between LTR10 activity and tumor mutations, patient survival rates, and clinical outcomes. For this purpose, we obtained and analyzed RNA-seq from 358 primary tumor samples derived from TCGA patients with colon adenocarcinomas or adenomas [27]. We first focused on correlating LTR10 transcriptional activity with KRAS mutation status. KRAS is one of the most frequently mutated oncogenes in cancer: approximately 30-40% of colorectal cancer patients harbor missense mutations in KRAS, and KRAS mutations have long been associated with increased tumor aggressiveness, resistance to treatment, and poor patient outcomes [42]. We found that LTR10A transcripts, in particular, are significantly elevated in tumors that harbor a KRAS mutation (Supp Fig S2K, Supp Table 7), although we did not observe a noticeable difference in *FOSL1* expression (Supp Fig S2L, Supp Table 7). Survival analyses based on the expression of LTR10 transcripts or AP1 factors did not reveal any significant correlations (Supp Fig S2M, S2N). However, we note that our analyses might have been limited by the fact that subfamily-level LTR10 transcription by RNA-seq is an imperfect proxy for enhancer activity, and only a small subset of colorectal cancer samples from TCGA had associated survival data, limiting our statistical power.

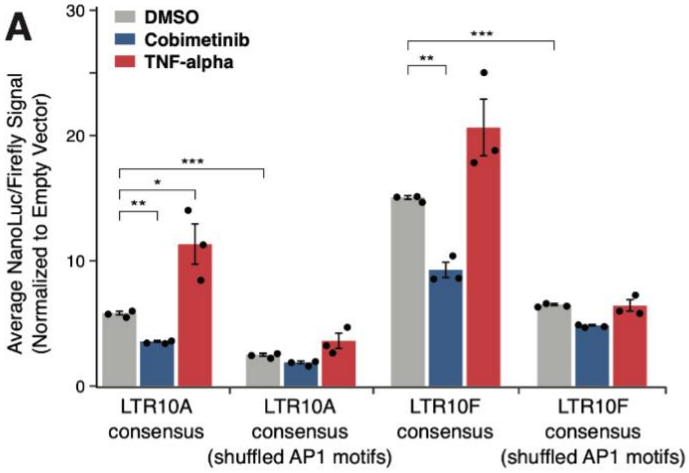
AP1 signaling is required for LTR10 enhancer activity

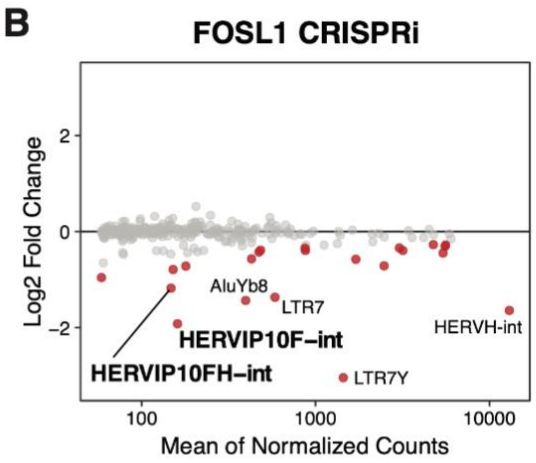
Dysregulation of AP1 signaling occurs in many cancers, driven by mutations that cause oncogenic activation of the MAPK signaling pathway [43]. Based on our findings that LTR10 elements are bound by AP1, and LTR10 transcriptional activity is correlated with the expression of AP1 component *FOSL1*, we tested whether LTR10 regulatory activity is affected by modulation of the AP1/MAPK signaling pathway using luciferase reporter assays. We synthesized the LTR10A and LTR10F consensus sequences as well as variants where the AP1 motifs were disrupted, and cloned the sequences into an enhancer reporter

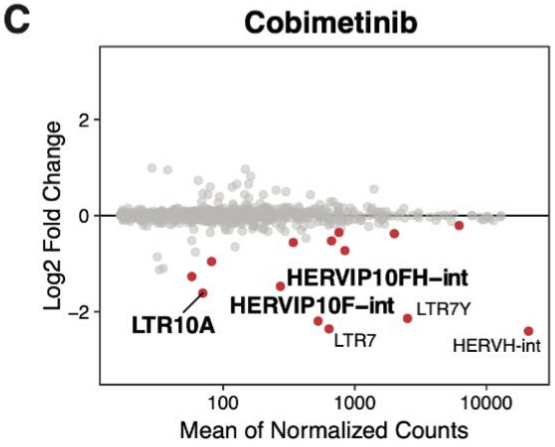
construct. We measured reporter activity in HCT116 cells that were treated for 24 hrs with either TNFa to stimulate signaling or cobimetinib (a MEK1 inhibitor) to inhibit signaling. Consistent with regulation by AP1, cobimetinib treatment caused a decrease in LTR10-driven reporter activity, and TNFa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

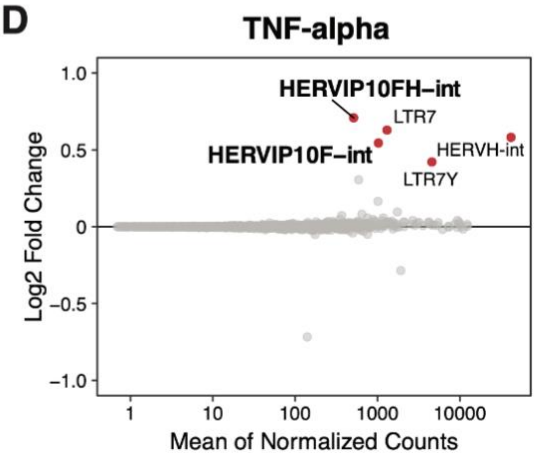
To test the role of the AP1 complex in endogenous LTR10 regulation, we silenced the AP1 component *FOSL1* using CRISPRi to determine the impact on LTR10 transcriptional activity. Using HCT116 cells expressing dCas9-KRAB-MeCP2 [44], we transfected a guide RNA (gRNA) targeting the *FOSL1* transcription start site (TSS) and used RNA-seq to compare gene and TE expression to control cells transfected with a non-targeting gRNA. We first confirmed silencing of *FOSL1* (Supp Fig S3A, S3B), then analyzed TE transcript expression summarized at the subfamily level to account for reads mapping to multiple insertions of the same TE [45]. This analysis revealed that full-length LTR10/HERV-I elements were significantly downregulated upon silencing *FOSL1* (Fig 3B), supporting a direct role for the AP1 complex in regulating LTR10 activity.

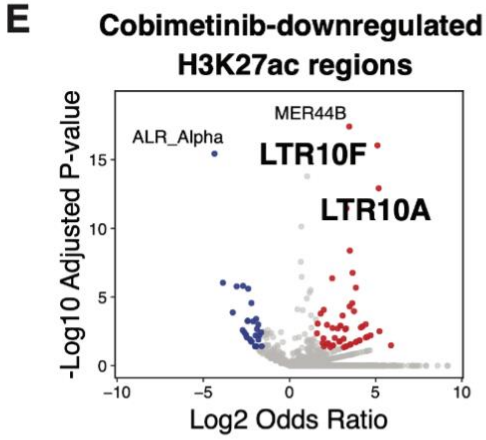
Next, we investigated how endogenous LTR10 elements respond to modulation of MAPK/AP1 signaling at the chromatin level. We treated HCT116 cells with either cobimetinib or TNFa for 24 hrs and profiled each response using RNA-seq and H3K27ac CUT&RUN (Supp Fig S3C, S3D, S3E, S3F). Consistent with our reporter assay results, our RNA-seq analysis showed that full-length LTR10/HERV-I transcripts were significantly downregulated upon cobimetinib treatment (Fig 3C), and upregulated upon TNFa treatment (Fig 3D). LTR10 elements showed similar responses based on H3K27ac CUT&RUN signal, exhibiting significant enrichment within the genome-wide set of predicted enhancers downregulated by cobimetinib or upregulated by TNFa (Fig 3E, 3F). We also observed clear TNFa-induced H3K27ac signal over LTR10 elements in a published dataset of SW480 colorectal cancer cells [46] (Supp Fig S3G). These results indicate that LTR10 elements represent a significant subset of genome-wide enhancers and transcripts in HCT116 cells that are directly modulated by MAPK/AP1 signaling.

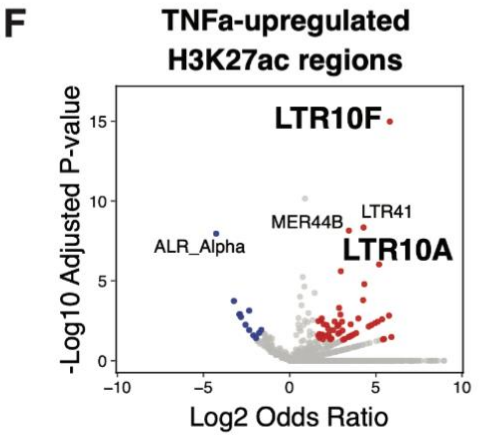












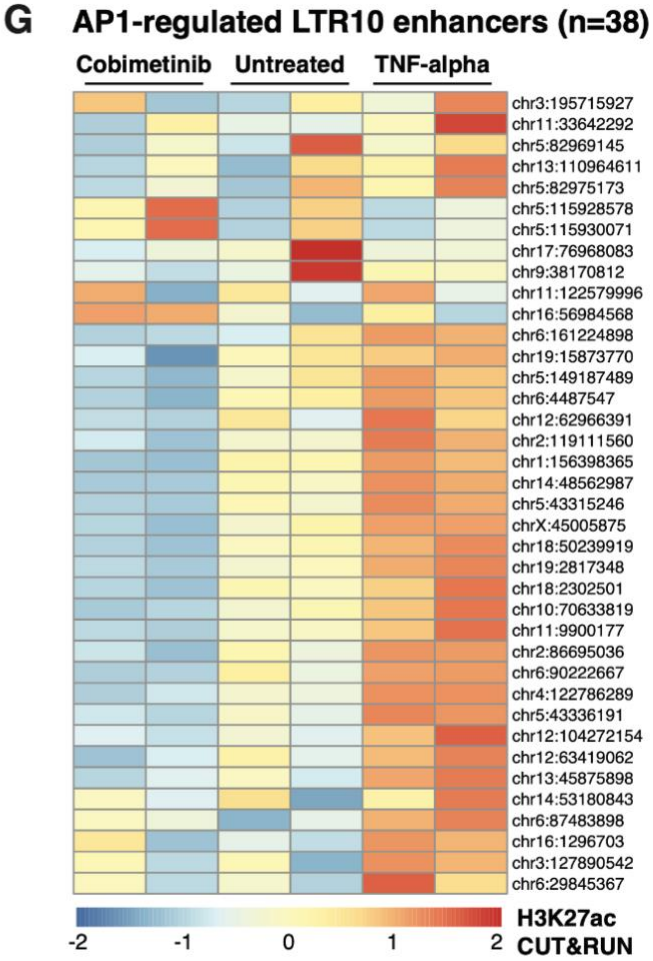


Figure 3: Control of LTR10 activity by AP1/MAPK signaling. (A) Luciferase reporter assays of LTR10A/F consensus sequences, including sequence variants containing shuffled AP1 motifs. Reporter activity was measured in HCT116 cells treated with DMSO (n=3), cobimetinib (n=3), or TNF-alpha (n=3) for 24 hrs.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

LTR10 elements regulate cancer-specific pathological gene expression

To determine whether any LTR10-derived enhancers have a functional effect on AP1/MAPK-dependent gene expression in colorectal cancer cells, we used our RNA-seq and CUT&RUN data from HCT116 cells to identify elements predicted to have gene regulatory activity. While we found that the AP1 component FOSL1 is required for LTR10 regulatory activity, oncogenic MAPK signaling can mediate transcriptional dysregulation through additional pathways beyond FOSL1 and AP1 signaling [47]. Therefore, we defined potential AP1/MAPK-regulated genes using two approaches, based on our RNA-seq data from our FOSL1 knockdown or TNFa/cobimetinib treatment. We first defined a set of 456 AP1-dependent genes based on being significantly downregulated by our CRISPRi silencing of the AP1 component *FOSL1* (Supp Table 8). We identified LTR10 elements predicted to regulate these genes using the activity by contact model [48] to assign enhancer-gene targets based on LTR10 element H3K27ac signal and chromatin interaction data. This identified 38 LTR10-derived enhancers (Fig 3G) predicted to regulate 56 (12.2%) of the 456 AP1-dependent genes (Supp Table 8), including many with established roles in cancer pathophysiology. In a secondary analysis, we defined 620 MAPK-dependent genes as genes that are both upregulated by TNFa and downregulated by cobimetinib, and found 57 LTR10-derived enhancers predicted to regulate 74 (11.9%) of these genes (Supp Fig S3H, Supp Table 9). Collectively, we identified a total of 71 distinct LTR10 enhancers (Supp Table 10) predicted to contribute to the regulation of roughly 12% of genes with AP1 or MAPK-dependent gene expression in HCT116 cells, supporting an important role in mediating global transcriptional rewiring in cancer.

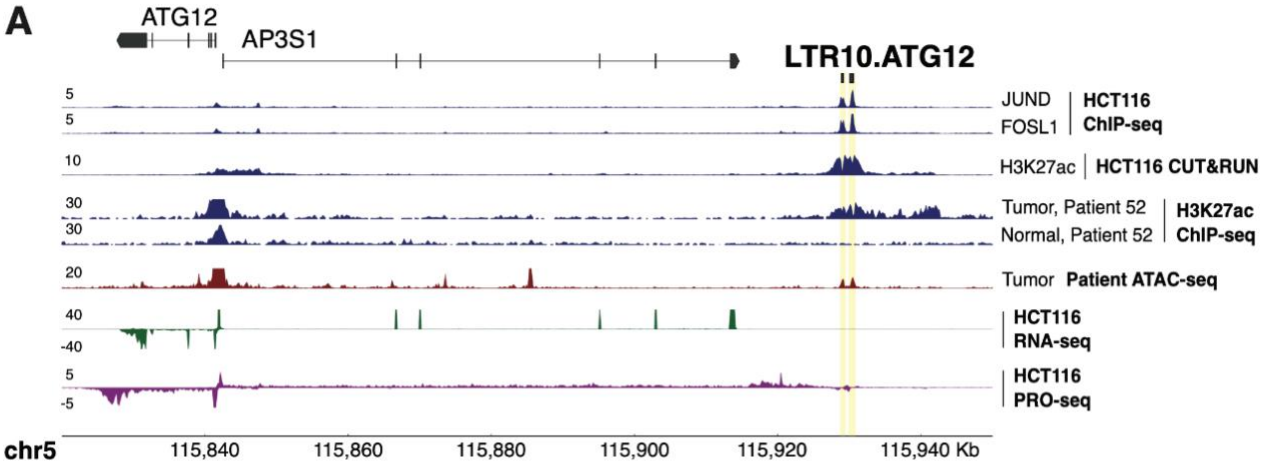
We tested the regulatory activity of six predicted LTR10 enhancers using CRISPR to knock down or knock out individual elements in HCT116 cells. We prioritized the elements based on epigenomic evidence of tumor-specific enhancer activity and having predicted target genes with reported relevance to tumor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

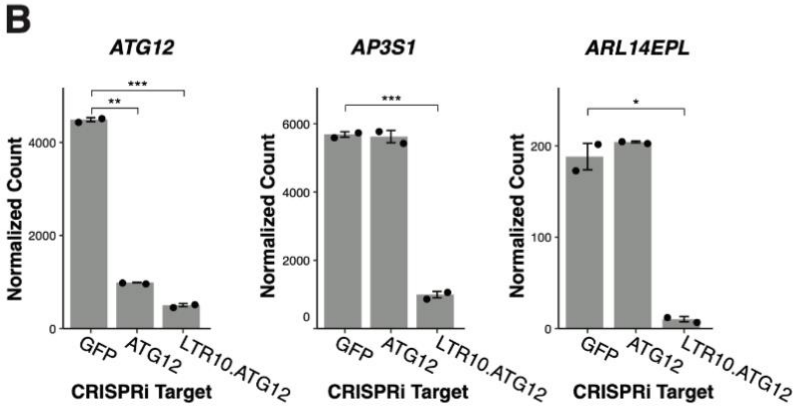
We focused on two of these LTR10-derived enhancers to explore their functional impact on tumor cells. We first investigated an enhancer that regulates *ATG12* (LTR10.ATG12), formed by two LTR10F elements on chromosome 5, located 87 kb from predicted target genes *ATG12* and *AP3S1* (Fig 4A). Silencing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

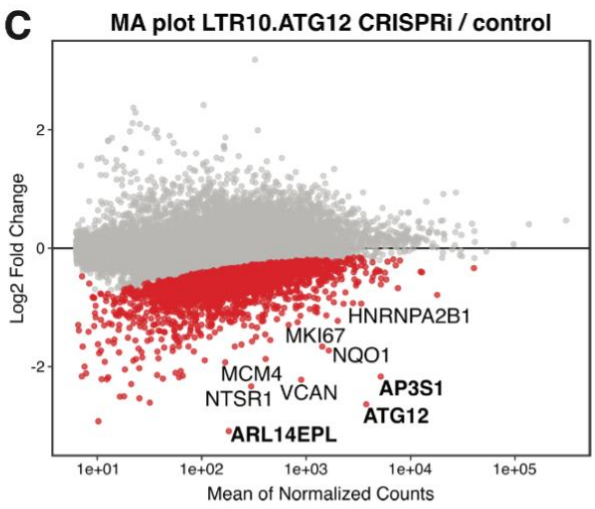
LTR10.ATG12 acts as an enhancer that controls AP1-dependent transcriptional activation of multiple genes in the *ATG12/AP3S1* locus in HCT116 cells.

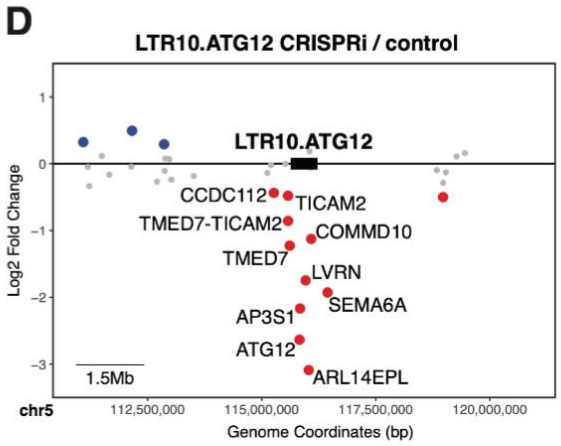
The *ATG12* gene encodes a ubiquitin-like modifier required for macroautophagy as well as mitochondrial homeostasis and apoptosis [49–52]. Expression of *ATG12* is associated with tumorigenesis and therapy resistance in colorectal and gastric cancer [53,54], but the mechanism of cancer-specific regulation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

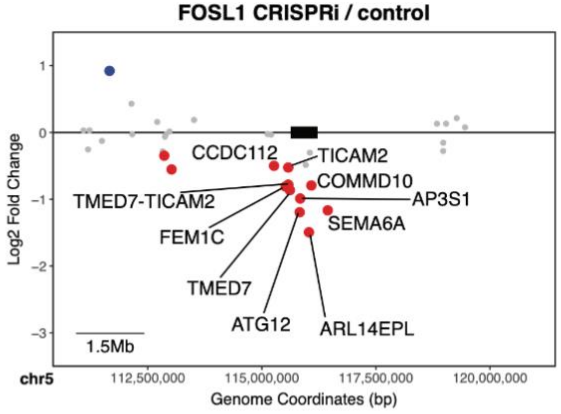
We tested whether ATG12-dependent functions require the activity of the LTR10.ATG12 enhancer. We treated each cell line with staurosporine (STS) to trigger mitochondrial apoptosis, which is dependent on free ATG12 binding to Bcl-2 [50]. In cells where either *ATG12* or the enhancer was silenced, we observed significantly reduced caspase 3/7 activity, indicating defective mitochondrial apoptosis (Fig 4F). We did not detect differences in macroautophagy in cells treated with bafilomycin (Supp Fig S4N), consistent with the lack of knockdown of the ATG5-ATG12 conjugate [52]. Our experimental results from silencing both *ATG12* and the enhancer are concordant with previous studies directly silencing *ATG12* using siRNAs in other cancer cell lines [49,50]. Together, these experiments demonstrate that the LTR10.ATG12 enhancer is functionally important for ATG12-dependent activity in HCT116 cells.

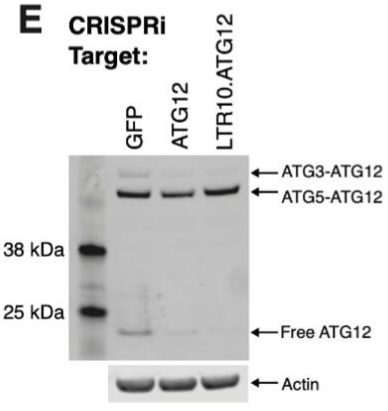


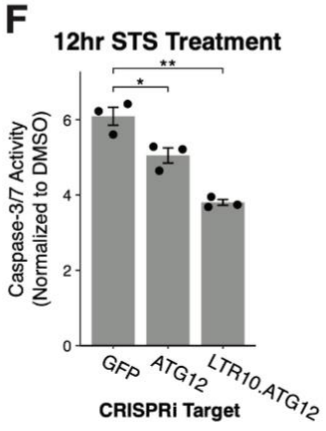












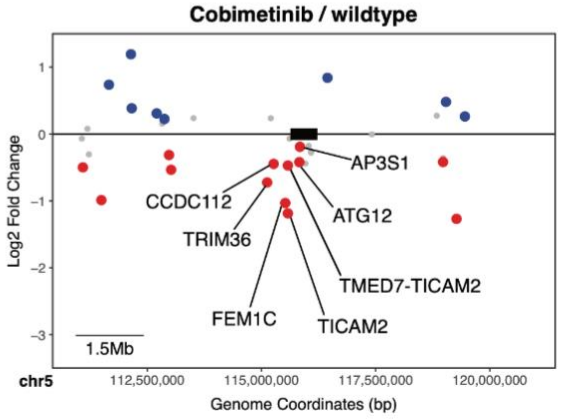
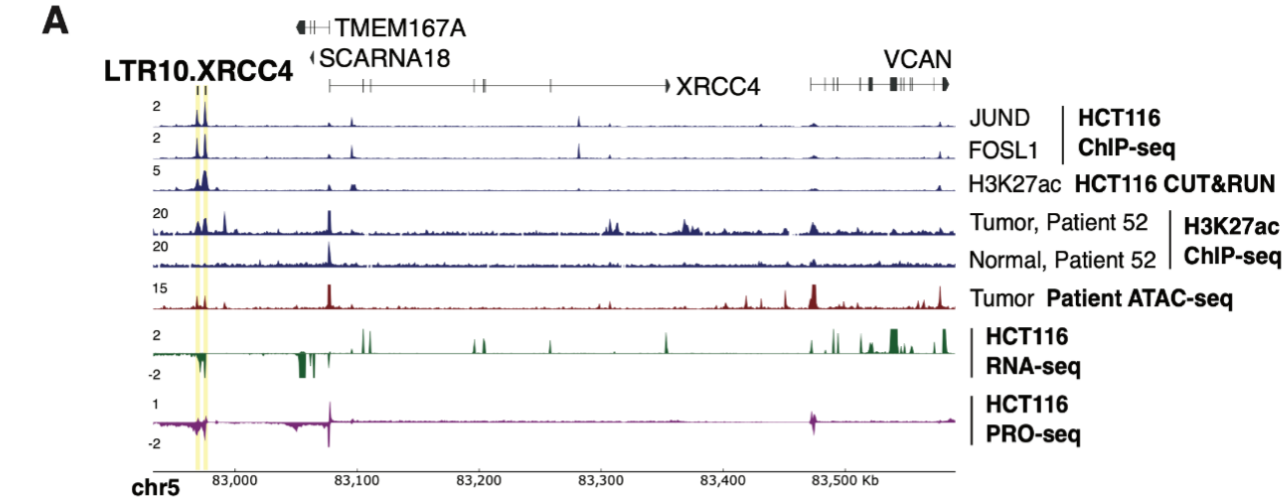
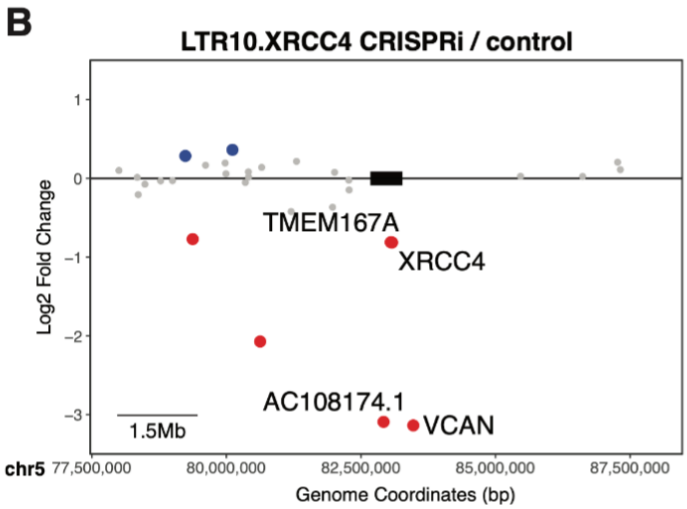


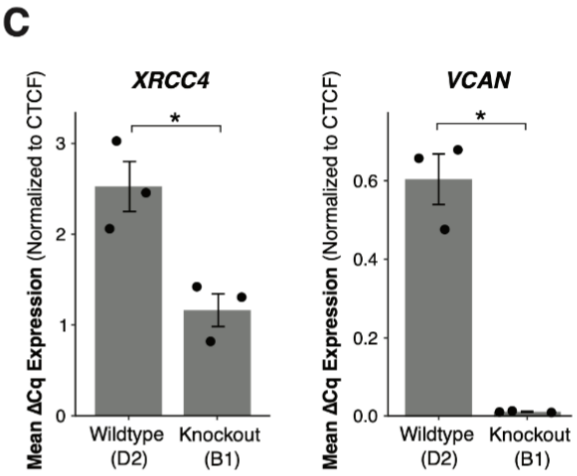
Figure 4: Functional characterization of LTR10.ATG12 in HCT116 cells. (A) Genome browser screenshot of the *ATG12/AP3S1* locus with the LTR10.ATG12 enhancer labeled. From top to bottom: JUND and FOSL1 ChIP-seq (GSE32465), H3K27ac CUT&RUN (in-house), H3K27ac ChIP-seq from matched tumor/normal samples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

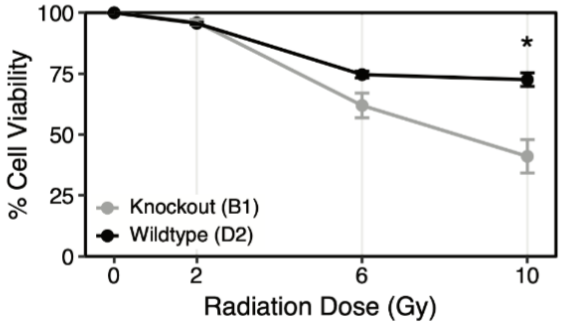
We next focused on the LTR10.XRCC4 enhancer, which regulates *XRCC4* and *VCAN* based on our CRISPRi silencing experiment (Fig 5A, 5B, Supp Fig S5A, S5B, Supp Table 12). *XRCC4* is a DNA repair gene required for non-homologous end joining and promotes resistance to chemotherapy and radiation therapy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

Previous studies have demonstrated that silencing or knocking out *XRCC4* directly causes increased sensitivity to DNA-damaging agents such as irradiation [55,64,65], including in HCT116 cells [66]. To test whether the LTR10.XRCC4 enhancer regulates *XRCC4* function in cancer, we subjected control and knockout cells to 10 Gy irradiation, and found that knockout cells showed reduced viability following irradiation (Fig 5D), consistent with a previous study showing the role of XRCC4 in tumor cell survival following irradiation [67]. We next tested how the deletion of LTR10.XRCC4 affects tumor response to irradiation in a mouse xenograft model. Irradiation inhibits the growth of tumors derived from HCT116 cells [68], therefore we tested whether reducing XRCC4 expression by deleting LTR10.XRCC4 affects tumor growth inhibition by irradiation.



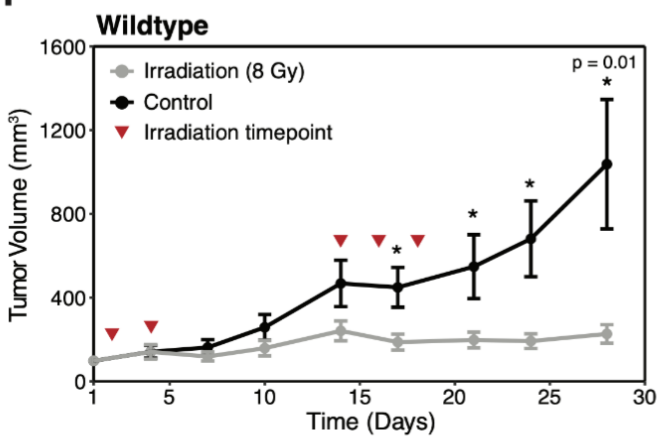






E

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



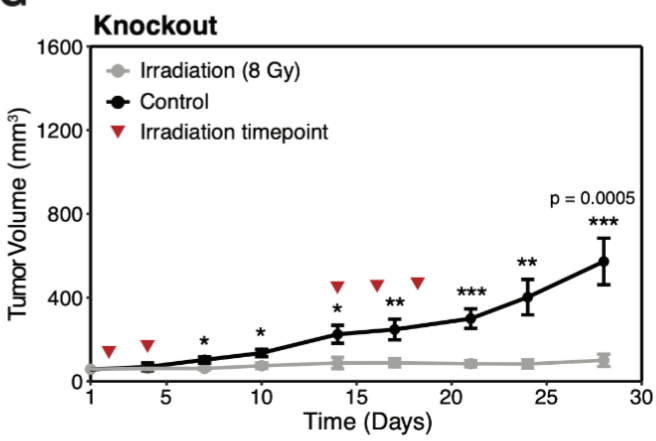


Figure 5: Functional characterization of LTR10.XRCC4 in HCT116 cells and xenograft models. (A) Genome browser screenshot of the *XRCC4* locus with the LTR10.XRCC4 enhancer labeled. From top to bottom: JUND and FOSL1 ChIP-seq (GSE32465), H3K27ac CUT&RUN (in-house), H3K27ac ChIP-seq from matched tumor/normal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

We transplanted either control HCT116 cells or cells harboring a homozygous deletion of LTR10.XRCC4 into athymic nude mice, and subjected the mice to 8 Gy irradiation or mock irradiation. Tumors derived from both control and knockout cells showed growth inhibition in response to irradiation (Fig 5E-G, Supp Fig S5E-G, Supp Table 18). However, LTR10.XRCC4 knockout tumors showed more significant overall tumor growth inhibition by irradiation (Fig 5E), including at earlier timepoints (Fig 5F, 5G). No significant toxicities were seen in animal weights. While the difference seen in knockout tumors was modest, these results are consistent with previous studies in other cancers showing increased sensitivity to radiation in tumor xenografts when XRCC4 is silenced [55], and implicate a role for the LTR10.XRCC4 enhancer in regulating a clinically relevant tumor phenotype.

LTR10 elements contain highly mutable VNTRs

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

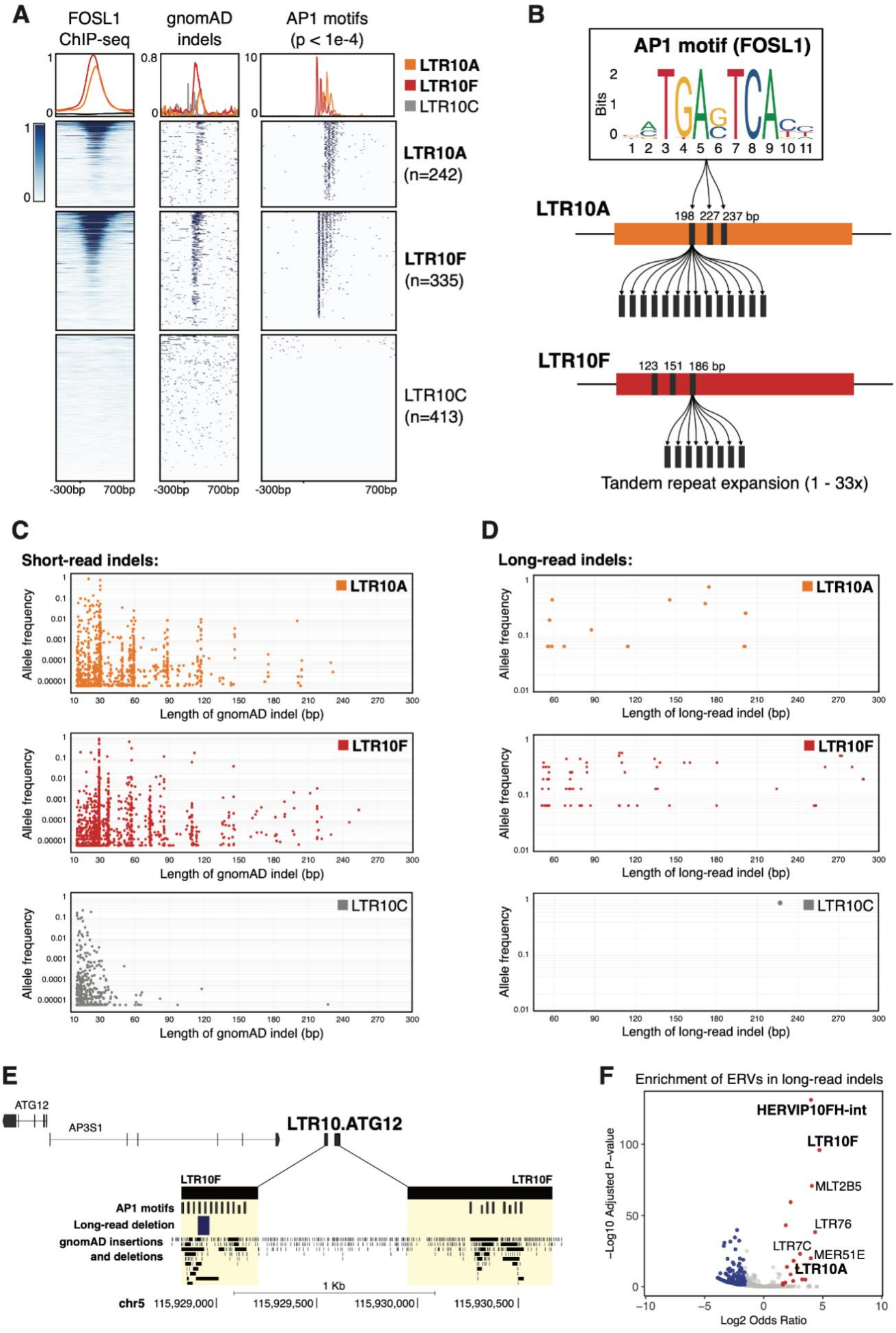


Figure 6: LTR10 repeat instability and polymorphism. (A) Heatmap of FOSL1 ChIP-Seq, gnomAD indels between 10-300bp in length, and AP1 motif matches (p<1e-4) across LTR10A, LTR10F, and LTR10C elements. Overlapping elements were removed, retaining 990 LTR10 elements total across the three subfamilies.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

Finally, we searched for evidence of tumor-specific somatic expansions within LTR10 VNTR regions. We analyzed a long-read whole-genome sequencing dataset generated from matched colorectal tumor and normal tissues from 20 patients [75], using Sniffles2 [76] to identify tumor-specific repeat expansions within LTR10 VNTR regions. After manually inspecting reads at each locus, we found evidence for tumor-specific VNTR expansions at H3K27ac-marked LTR10 elements in five out of 20 patients (Supp Table 21). Three patients showed independent somatic expansions at the same LTR10A locus on chromosome 1 located near gene *GPR137B* (Supp Fig S6K, S6L, S6O), suggesting that this locus is prone to inter-individual variation at both the germline (Supp Fig S6D) and somatic level. We also found evidence of tumor-specific mosaic VNTR deletions in four patients (Supp Table 21). In fact, one patient with high microsatellite instability showed evidence of multiple tumor-specific LTR10 variants: a predicted LTR10A VNTR expansion over 11,600 bp in length (Supp Fig S6K, Supp Table 21), as well as two tumor-specific deletions at different LTR10F VNTR loci (Supp Fig S6Q, S6R). While a larger cohort would be necessary to determine if these expansions are enriched within tumors with microsatellite instability, these analyses provide evidence that LTR10 VNTRs are subject to tumor-specific somatic expansions and contractions, which can alter tumor-specific gene regulatory activity.

DISCUSSION

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

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

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

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

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

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

METHODS

Cell culture

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

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

CRISPR-mediated silencing and knockout of LTR10s

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

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

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

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

Cell autophagy and apoptosis assays

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

Western blots

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

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

Luciferase assay

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

Irradiation experiment

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

Mouse xenograft experiment

All experiments were approved by the Institutional Animal Care and Use Committee of the University of Colorado Anschutz Medical Campus and conducted in accordance with the National Institutes of Health Guidelines for the Care and Use of Laboratory Animals. Female athymic nude mice (aged 15-16 weeks at start of study) were purchased from Envigo (Indianapolis, IN) and implanted subcutaneously on the hind flanks with 2.5 million cells in 100 ul of either HCT116 wildtype or LTR10.XRCC4 CRISPR knockout cells under isoflurane anesthesia with a 23 ga ½ needle. The cell solution injected consisted of 1:1 ratio of RMPI media and cultrex (Cultrex PathClear BME, Type 3 from Bio-Techne). We injected wildtype or knockout cells into 40 mice (20 each, one side per mouse), then mice were randomized into treatment groups (20 irradiated, 20 mock) and treatments were initiated when the average tumor volume reached between 50-100 mm^3. Tumor volume was calculated by {($width^2$) x length} x 0.52. Irradiation treatment consisted of 8 Gy x 3 fractions on days 2, 4, 14, 16, and 18. Tumor measurements were taken twice weekly using digital calipers, and toxicity was monitored by measuring body weight twice weekly and the study ended at 28 days. Tumor growth inhibition was measured using KuLGaP [69].

RNA-seq

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

CUT&RUN

Libraries were prepared from treatment replicates. Approximately $5x10^5$ viable cells were used for each CUT&RUN reaction, and pulldowns were generated following the protocol from [91]. All buffers were prepared according to the "High Ca^{2+}/Low Salt" method using

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

Processing of sequencing data

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

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

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

TE colocalization analysis

To determine TE subfamily enrichment within regulatory regions, we used GIGGLE (v0.6.3) [92] to generate a genomic interval index of all TE subfamilies in the hg38 human genome, based on Dfam v2.0 repeat annotation (n=1315 TE subfamilies). Regulatory regions (e.g., ATAC, ChIP-Seq, or CUT&RUN peaks) were queried against the TE interval index using the GIGGLE search function (-g 3209286105 -s). Results were ranked by GIGGLE enrichment score, which is a composite of the product of −log10(P value) and log2(odds ratio). Significantly enriched TE subfamilies were defined as those with at least 25 overlaps between TE copies and a set of peak regions, odds ratio over 10, and GIGGLE score over 100 in at least one cancer type.

Defining cancer-specific regulatory elements

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

Transcription factor motif analyses

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

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

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

Differential analysis using DESeq2

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

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

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

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

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

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

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

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

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

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

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

component analysis (log=True, n_pcs=40), tSNE (perplexity=30, learning_rate=1000, random_state=0, n_pcs=40), and UMAP (n_neighbors=30, n_pcs=40, min_dist=0.8, spread=1, random_state=0, maxiter=100). Leiden clustering (resolution=0.75) was used to assign cells to clusters, and cell clusters with less than 20 cells were excluded from final UMAP visualizations. Cell types were annotated using the PanglaoDB database [95] of gene expression markers, with manual verification.

Prediction of LTR10 enhancer gene targets

LTR10 elements were initially prioritized for CRISPR silencing or deletion based on enhancer predictions from the Activity-by-Contact (ABC) model [48]. Publicly available HCT116 ATACseq (GEO accession GSM3593802) and in-house HCT116 H3K27ac CUT&RUN were used as input to the ABC pipeline, as well as the provided averaged human cell line HiC file. Predicted enhancer regions with an ABC interaction score over 0.001 were intersected with H3K27ac-marked LTR10A/F elements. Putative LTR10 enhancers were then checked for proximity (e.g. within 1.5Mb) to FOSL1-regulated genes (i.e. genes that were significantly downregulated by FOSL1 knockdown), or MAPK-regulated genes (i.e. genes that were significantly affected by MAPK treatments Cobimetinib and TNF-alpha, based on inhouse RNAseq).

Evolutionary analysis of LTR10 sequences

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

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

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

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

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

VNTR identification

gnomAD (v3.1) VCF files for each hg38 chromosome were filtered for high-confidence indels (FILTER=PASS) using the query function of bcftools (v1.8) with format parameter -f'%CHROM\t%POS0\t%END\t%ID\t%REF\t%ALT\t%AF\t%TYPE\tFILTER=%FILTER\n'. The remaining indels were then subsetted by size to retain insertions or deletions between 10 to 300bp in length. Chromosome VCFs were concatenated into one whole genome BED file. Bedtools (v2.28.0) was used to intersect the indel BED file with LTR10 elements from each subfamily, based on Dfam (v2.0) repeat annotation.

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

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

minimap2 [99] (v2.22-r1101) to align reads to the hg38 reference genome, with parameters -a to generate output in SAM format, -x map-ont to specify nanopore input reads, -t 4 to set the number of threads to 4, and -Y to use soft clipping for supplementary alignments. We then samtools (v1.10) to generate sorted BAM files, with commands samtools view -bS to convert from SAM to BAM format, samtools sort (default parameters) to sort reads by coordinate, and samtools index (default parameters) to generate a BAM index file for each BAM. We then used Sniffles2 (v2.0.7) [76] to identify tumor-specific SVs within LTR10 VNTR regions. For each tumor/normal pair, we called SVs using both the default parameters (optimized for germline variants), and then again using the --non-germline parameter for the tumor sample only (optimized for detecting low frequency or mosaic variants). The reference genome was set to hg38 and --tandem-repeats were annotated using the Sniffles-provided human_GRCh38_no_alt_analysis_set.trf.bed file. Sniffles was run with the -snf option to save candidate SVs to the SNF binary file, per sample. For each patient, tumor and normal SNF files were then merged using the Sniffles population verge with --vcf to specify VCF output format. All VCF output files were intersected with LTR10 VNTR regions using bedtools (v2.28.0). For each patient, tumor-specific variants were extracted using the SUPP_VEC tag in the INFO field of the output VCFs (i.e. by extracting all SVs with SUPP_VEC=01, which signifies absence in the normal sample and presence in the tumor sample). Finally, for each called insertion or deletion, we manually inspected aligned reads using the UCSC genome browser to confirm differences between the tumor and normal samples.

DATA AVAILABILITY

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

CODE AVAILABILITY

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

COMPETING INTEREST STATEMENT

The authors declare no competing interests.

REFERENCES

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

8. Sur, I. & Taipale, J. The role of enhancers in cancer. *Nat. Rev. Cancer* **16,** 483–493 (2016).

9. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357,** (2017).

10. Jansz, N. & Faulkner, G. J. Endogenous retroviruses in the origins and treatment of cancer. *Genome Biol.* **22,** 147 (2021).

11. Kong, Y., Rose, C. M., Cass, A. A., Williams, A. G., Darwish, M., Lianoglou, S., Haverty, P. M., Tong, A.-J., Blanchette, C., Albert, M. L., Mellman, I., Bourgon, R., Greally, J., Jhunjhunwala, S. & Chen-Harris, H. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nat. Commun.* **10,** 5228 (2019).

12. Shukla, R., Upton, K. R., Muñoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T., Brennan, P. M., Baillie, J. K., Collino, A., Ghisletti, S., Sinha, S., Iannelli, F., Radaelli, E., Dos Santos, A., Rapoud, D., Guettier, C., Samuel, D., Natoli, G., Carninci, P., Ciccarelli, F. D., Garcia-Perez, J. L., Faivre, J. & Faulkner, G. J. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153,** 101–111 (2013).

13. Rodriguez-Martin, B., Alvarez, E. G., Baez-Ortega, A., Zamora, J., Supek, F., Demeulemeester, J., Santamarina, M., Ju, Y. S., Temes, J., Garcia-Souto, D., Detering, H., Li, Y., Rodriguez-Castro, J., Dueso-Barroso, A., Bruzos, A. L., Dentro, S. C., Blanco, M. G., Contino, G., Ardeljan, D., Tojo, M., Roberts, N. D., Zumalave, S., Edwards, P. A. W., Weischenfeldt, J., Puiggròs, M., Chong, Z., Chen, K., Lee, E. A., Wala, J. A., Raine, K., Butler, A., Waszak, S. M., Navarro, F. C. P., Schumacher, S. E., Monlong, J., Maura, F., Bolli, N., Bourque, G., Gerstein, M., Park, P. J., Wedge, D.

C., Beroukhim, R., Torrents, D., Korbel, J. O., Martincorena, I., Fitzgerald, R. C., Van Loo, P., Kazazian, H. H., Burns, K. H., PCAWG Structural Variation Working Group, Campbell, P. J., Tubio, J. M. C. & PCAWG Consortium. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat. Genet.* **52,** 306–319 (2020).

14. Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M. S., Hruban, R. H., Iacobuzio-Donahue, C. A., Maitra, A., Torbenson, M. S., Goggins, M., Shih, I.-M., Duffield, A. S., Montgomery, E. A., Gabrielson, E., Netto, G. J., Lotan, T. L., De Marzo, A. M., Westra, W., Binder, Z. A., Orr, B. A., Gallia, G. L., Eberhart, C. G., Boeke, J. D., Harris, C. R. & Burns, K. H. Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.* **184,** 1280–1286 (2014).
15. Babaian, A. & Mager, D. L. Endogenous retroviral promoter exaptation in human cancer. *Mob. DNA* **7,** 24 (2016).
16. Yu, C., Lei, X., Chen, F., Mao, S., Lv, L., Liu, H., Hu, X., Wang, R., Shen, L., Zhang, N., Meng, Y., Shen, Y., Chen, J., Li, P., Huang, S., Lin, C., Zhang, Z. & Yuan, K. ARID1A loss derepresses a group of human endogenous retrovirus-H loci to modulate BRD4-dependent transcription. *Nat. Commun.* **13,** 3501 (2022).
17. Babaian, A., Romanish, M. T., Gagnier, L., Kuo, L. Y., Karimi, M. M., Steidl, C. & Mager, D. L. Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. *Oncogene* **35,** 2542–2546 (2016).
18. Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K., Bouhlel, M. A., Richter, J., Soler, E., Stadhouders, R., Jöhrens, K., Wurster, K. D., Callen, D. F., Harte, M. F., Giefing, M., Barlow, R., Stein, H., Anagnostopoulos, I.,

Janz, M., Cockerill, P. N., Siebert, R., Dörken, B., Bonifer, C. & Mathas, S. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.* **16,** 571–579 (2010).

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

24. Jacques, P.-É., Jeyakani, J. & Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **9,** e1003504 (2013).

25. Deniz, Ö., Ahmed, M., Todd, C. D., Rio-Machin, A., Dawson, M. A. & Branco, M. R. Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. *Nat. Commun.* **11,** 3506 (2020).

26. Grillo, G., Keshavarzian, T., Linder, S., Arlidge, C., Mout, L., Nand, A., Teng, M., Qamra, A., Zhou, S., Kron, K. J., Murison, A., Hawley, J. R., Fraser, M., van der Kwast, T. H., Raj, G. V., He, H. H., Zwart, W. & Lupien, M. Transposable elements are co-opted as oncogenic regulatory elements by lineage-specific transcription factors in prostate cancer. *Cancer Discov.* (2023). doi:10.1158/2159-8290.CD-23-0331

27. Corces, M. R., Granja, J. M., Shams, S., Louie, B. H., Seoane, J. A., Zhou, W., Silva, T. C., Groeneveld, C., Wong, C. K., Cho, S. W., Satpathy, A. T., Mumbach, M. R., Hoadley, K. A., Robertson, A. G., Sheffield, N. C., Felau, I., Castro, M. A. A., Berman, B. P., Staudt, L. M., Zenklusen, J. C., Laird, P. W., Curtis, C., Cancer Genome Atlas Analysis Network, Greenleaf, W. J. & Chang, H. Y. The chromatin accessibility landscape of primary human cancers. *Science* **362,** (2018).

28. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., Wu, Y.-C., Pfenning, A. R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R. A., Shoresh, N., Epstein, C. B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R. D., Lister, R., Hong, C., Gascard, P., Mungall, A. J., Moore, R., Chuah, E., Tam, A., Canfield, T. K., Hansen, R. S., Kaul, R., Sabo, P. J., Bansal,

M. S., Carles, A., Dixon, J. R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T. R., Neph, S. J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R. C., Siebenthall, K. T., Sinnott-Armstrong, N. A., Stevens, M., Thurman, R. E., Wu, J., Zhang, B., Zhou, X., Beaudet, A. E., Boyer, L. A., De Jager, P. L., Farnham, P. J., Fisher, S. J., Haussler, D., Jones, S. J. M., Li, W., Marra, M. A., McManus, M. T., Sunyaev, S., Thomson, J. A., Tlsty, T. D., Tsai, L.-H., Wang, W., Waterland, R. A., Zhang, M. Q., Chadwick, L. H., Bernstein, B. E., Costello, J. F., Ecker, J. R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J. A., Wang, T. & Kellis, M. Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).

29. Akhtar-Zaidi, B., Cowper-Sal-lari, R., Corradin, O., Saiakhova, A., Bartels, C. F., Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M. F., Willis, J., Moore, J. H., Tesar, P. J., Laframboise, T., Markowitz, S., Lupien, M. & Scacheri, P. C. Epigenomic enhancer profiling defines a signature of colon cancer. *Science* **336,** 736–739 (2012).

30. Baranello, L., Wojtowicz, D., Cui, K., Devaiah, B. N., Chung, H.-J., Chan-Salis, K. Y., Guha, R., Wilson, K., Zhang, X., Zhang, H., Piotrowski, J., Thomas, C. J., Singer, D. S., Pugh, B. F., Pommier, Y., Przytycka, T. M., Kouzine, F., Lewis, B. A., Zhao, K. & Levens, D. RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* **165,** 357–371 (2016).

31. Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.-H., Brown, M., Zhang, X., Meyer, C. A. & Liu, X. S. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* **47,** D729–D735 (2019).

32. Grow, E. J., Weaver, B. D., Smith, C. M., Guo, J., Stein, P., Shadle, S. C.,

Hendrickson, P. G., Johnson, N. E., Butterfield, R. J., Menafra, R., Kloet, S. L., van der Maarel, S. M., Williams, C. J. & Cairns, B. R. p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. *Nat. Genet.* **53,** 1207–1220 (2021).

33. Imbeault, M., Helleboid, P.-Y. & Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543,** 550–554 (2017).

34. Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., Farnham, P. J., Hirst, M., Lander, E. S., Mikkelsen, T. S. & Thomson, J. A. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28,** 1045–1048 (2010).

35. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

36. Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B. & Ecker, J. R. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).

37. Bruno, M., Mahgoub, M. & Macfarlan, T. S. The Arms Race Between KRAB–Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. (2019). doi:10.1146/annurev-genet-112618-043717

38. Orouji, E., Raman, A. T., Singh, A. K., Sorokin, A., Arslan, E., Ghosh, A. K., Schulz, J., Terranova, C., Jiang, S., Tang, M., Maitituoheti, M., Callahan, S. C., Barrodia, P., Tomczak, K., Jiang, Y., Jiang, Z., Davis, J. S., Ghosh, S., Lee, H. M., Reyes-Uribe, L., Chang, K., Liu, Y., Chen, H., Azhdarinia, A., Morris, J., Vilar, E., Carmon, K. S.,

Kopetz, S. E. & Rai, K. Chromatin state dynamics confers specific therapeutic strategies in enhancer subtypes of colorectal cancer. *Gut* (2021). doi:10.1136/gutjnl-2020-322835

39. Bujold, D., Morais, D. A. de L., Gauthier, C., Côté, C., Caron, M., Kwan, T., Chen, K. C., Laperle, J., Markovits, A. N., Pastinen, T., Caron, B., Veilleux, A., Jacques, P.-É. & Bourque, G. The International Human Epigenome Consortium Data Portal. *Cell Syst* **3,** 496–499.e2 (2016).

40. Pelka, K., Hofree, M., Chen, J. H., Sarkizova, S., Pirl, J. D., Jorgji, V., Bejnood, A., Dionne, D., Ge, W. H., Xu, K. H., Chao, S. X., Zollinger, D. R., Lieb, D. J., Reeves, J. W., Fuhrman, C. A., Hoang, M. L., Delorey, T., Nguyen, L. T., Waldman, J., Klapholz, M., Wakiro, I., Cohen, O., Albers, J., Smillie, C. S., Cuoco, M. S., Wu, J., Su, M.-J., Yeung, J., Vijaykumar, B., Magnuson, A. M., Asinovski, N., Moll, T., Goder-Reiser, M. N., Applebaum, A. S., Brais, L. K., DelloStritto, L. K., Denning, S. L., Phillips, S. T., Hill, E. K., Meehan, J. K., Frederick, D. T., Sharova, T., Kanodia, A., Todres, E. Z., Jané-Valbuena, J., Biton, M., Izar, B., Lambden, C. D., Clancy, T. E., Bleday, R., Melnitchouk, N., Irani, J., Kunitake, H., Berger, D. L., Srivastava, A., Hornick, J. L., Ogino, S., Rotem, A., Vigneau, S., Johnson, B. E., Corcoran, R. B., Sharpe, A. H., Kuchroo, V. K., Ng, K., Giannakis, M., Nieman, L. T., Boland, G. M., Aguirre, A. J., Anderson, A. C., Rozenblatt-Rosen, O., Regev, A. & Hacohen, N. Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* **184,** 4734–4752.e20 (2021).

41. He, J., Babarinde, I. A., Sun, L., Xu, S., Chen, R., Shi, J., Wei, Y., Li, Y., Ma, G., Zhuang, Q., Hutchins, A. P. & Chen, J. Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a

processing pipeline scTE. *Nat. Commun.* **12,** 1456 (2021).

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

49. Haller, M., Hock, A. K., Giampazolias, E., Oberst, A., Green, D. R., Debnath, J., Ryan, K. M., Vousden, K. H. & Tait, S. W. G. Ubiquitination and proteasomal degradation of ATG12 regulates its proapoptotic activity. *Autophagy* **10,** 2269–2278 (2014).

50. Rubinstein, A. D., Eisenstein, M., Ber, Y., Bialik, S. & Kimchi, A. The autophagy protein Atg12 associates with antiapoptotic Bcl-2 family members to promote mitochondrial apoptosis. *Mol. Cell* **44,** 698–709 (2011).

51. Radoshevich, L., Murrow, L., Chen, N., Fernandez, E., Roy, S., Fung, C. & Debnath, J. ATG12 conjugation to ATG3 regulates mitochondrial homeostasis and cell death. *Cell* **142,** 590–600 (2010).

52. Hanada, T., Noda, N. N., Satomi, Y., Ichimura, Y., Fujioka, Y., Takao, T., Inagaki, F. & Ohsumi, Y. The Atg12-Atg5 Conjugate Has a Novel E3-like Activity for Protein Lipidation in Autophagy *. *J. Biol. Chem.* **282,** 37298–37302 (2007).

53. Hu, J. L., He, G. Y., Lan, X. L., Zeng, Z. C., Guan, J., Ding, Y., Qian, X. L., Liao, W. T., Ding, Y. Q. & Liang, L. Inhibition of ATG12-mediated autophagy by miR-214 enhances radiosensitivity in colorectal cancer. *Oncogenesis* **7,** 16 (2018).

54. YiRen, H., YingCong, Y., Sunwu, Y., Keqin, L., Xiaochun, T., Senrui, C., Ende, C., XiZhou, L. & Yanfan, C. Long noncoding RNA MALAT1 regulates autophagy associated chemoresistance via miR-23b-3p sequestration in gastric cancer. *Mol. Cancer* **16,** 174 (2017).

55. Zheng, Z., Ng, W. L., Zhang, X., Olson, J. J., Hao, C., Curran, W. J. & Wang, Y. RNAi-mediated targeting of noncoding and coding sequences in DNA repair gene messages efficiently radiosensitizes human tumor cells. *Cancer Res.* **72,** 1221–1228 (2012).

56. Xu, M., Huang, X., Zheng, C., Long, J., Dai, Q., Chen, Y., Lu, J., Pan, C., Yao, S. & Li, J. Platinum-Resistant Ovarian Cancer Is Vulnerable to the cJUN-XRCC4 Pathway

Inhibition. *Cancers* **14,** (2022).

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

human melanoma cells. *J. Biol. Chem.* **284,** 12306–12317 (2009).

64. Li, Z., Otevrel, T., Gao, Y., Cheng, H. L., Seed, B., Stamato, T. D., Taccioli, G. E. & Alt, F. W. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **83,** 1079–1089 (1995).

65. Critchlow, S. E., Bowater, R. P. & Jackson, S. P. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr. Biol.* **7,** 588–598 (1997).

66. Katsube, T., Mori, M., Tsuji, H., Shiomi, T., Shiomi, N. & Onoda, M. Differences in sensitivity to DNA-damaging Agents between XRCC4- and Artemis-deficient human cells. *J. Radiat. Res.* **52,** 415–424 (2011).

67. Wen, Y., Dai, G., Wang, L., Fu, K. & Zuo, S. Silencing of XRCC4 increases radiosensitivity of triple-negative breast cancer cells. *Biosci. Rep.* **39,** (2019).

68. Leu, J.-D., Wang, B.-S., Chiu, S.-J., Chang, C.-Y., Chen, C.-C., Chen, F.-D., Avirmed, S. & Lee, Y.-J. Combining fisetin and ionizing radiation suppresses the growth of mammalian colorectal cancers in xenograft tumor models. *Oncol. Lett.* **12,** 4975–4982 (2016).

69. Ortmann, J., Rampášek, L., Tai, E., Mer, A. S., Shi, R., Stewart, E. L., Mascaux, C., Fares, A., Pham, N.-A., Beri, G., Eeles, C., Tkachuk, D., Ho, C., Sakashita, S., Weiss, J., Jiang, X., Liu, G., Cescon, D. W., O'Brien, C. A., Guo, S., Tsao, M.-S., Haibe-Kains, B. & Goldenberg, A. Assessing therapy response in patient-derived xenografts. *Sci. Transl. Med.* **13,** eabf4969 (2021).

70. Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E.

G., Kosmicki, J. A., Walters, R. K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J. X., Samocha, K. E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A. H., Minikel, E. V., Weisburd, B., Lek, M., Ware, J. S., Vittal, C., Armean, I. M., Bergelson, L., Cibulskis, K., Connolly, K. M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M. E., Genome Aggregation Database Consortium, Neale, B. M., Daly, M. J. & MacArthur, D. G. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581,** 434–443 (2020).

71. Gymrek, M. A genomic view of short tandem repeats. *Curr. Opin. Genet. Dev.* **44,** 9–16 (2017).

72. Audano, P. A., Sulovari, A., Graves-Lindsay, T. A., Cantsilieris, S., Sorensen, M., Welch, A. E., Dougherty, M. L., Nelson, B. J., Shah, A., Dutcher, S. K., Warren, W. C., Magrini, V., McGrath, S. D., Li, Y. I., Wilson, R. K. & Eichler, E. E. Characterizing the Major Structural Variant Alleles of the Human Genome. *Cell* **176,** 663–675.e19 (2019).

73. Quan, C., Li, Y., Liu, X., Wang, Y., Ping, J., Lu, Y. & Zhou, G. Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression. *Genome Biol.* **22,** 159 (2021).

74. van Bree, E. J., Guimarães, R. L. F. P., Lundberg, M., Blujdea, E. R., Rosenkrantz, J. L., White, F. T. G., Poppinga, J., Ferrer-Raventós, P., Schneider, A.-F. E., Clayton, I., Haussler, D., Reinders, M. J. T., Holstege, H., Ewing, A. D., Moses, C. & Jacobs, F. M. J. A hidden layer of structural variation in transposable elements reveals potential genetic modifiers in human disease-risk loci. *Genome Res.* **32,** 656–670 (2022).

75. Xu, L., Wang, X., Lu, X., Liang, F., Liu, Z., Zhang, H., Li, X., Tian, S., Wang, L. & Wang, Z. Long-read sequencing identifies novel structural variations in colorectal cancer. *PLoS Genet.* **19,** e1010514 (2023).

76. Smolka, M., Paulin, L. F., Grochowski, C. M., Horner, D. W., Mahmoud, M., Behera, S., Kalef-Ezra, E., Gandhi, M., Hong, K., Pehlivan, D., Scholz, S. W., Carvalho, C. M. B., Proukakis, C. & Sedlazeck, F. J. Comprehensive Structural Variant Detection: From Mosaic to Population-Level. *bioRxiv* 2022.04.04.487055 (2023). doi:10.1101/2022.04.04.487055

77. Karttunen, K., Patel, D., Xia, J., Fei, L., Palin, K., Aaltonen, L. & Sahu, B. Transposable elements as tissue-specific enhancers in cancers of endodermal lineage. *Nat. Commun.* **14,** 5313 (2023).

78. Frost, J. M., Amante, S. M., Okae, H., Jones, E. M., Ashley, B., Lewis, R. M., Cleal, J. K., Caley, M. P., Arima, T., Maffucci, T. & Branco, M. R. Regulation of human trophoblast gene expression by endogenous retroviruses. *Nat. Struct. Mol. Biol.* **30,** 527–538 (2023).

79. Lynch-Sutherland, C. F., Chatterjee, A., Stockwell, P. A., Eccles, M. R. & Macaulay, E. C. Reawakening the Developmental Origins of Cancer Through Transposable Elements. *Front. Oncol.* **10,** 468 (2020).

80. Costanzo, V., Bardelli, A., Siena, S. & Abrignani, S. Exploring the links between cancer and placenta development. *Open Biol.* **8,** (2018).

81. Wagner, G. P., Kshitiz, Dighe, A. & Levchenko, A. The Coevolution of Placentation and Cancer. *Annu Rev Anim Biosci* **10,** 259–279 (2022).

82. Rubinsztein, D. C., Codogno, P. & Levine, B. Autophagy modulation as a potential therapeutic target for diverse diseases. *Nat. Rev. Drug Discov.* **11,** 709–730 (2012).

83. Bau, D.-T., Yang, M.-D., Tsou, Y.-A., Lin, S.-S., Wu, C.-N., Hsieh, H.-H., Wang, R.-F., Tsai, C.-W., Chang, W.-S., Hsieh, H.-M., Sun, S.-S. & Tsai, R.-Y. Colorectal cancer and genetic polymorphism of DNA double-strand break repair gene XRCC4 in Taiwan. *Anticancer Res.* **30,** 2727–2730 (2010).

84. Chatterjee, P., Choudhary, G. S., Alswillah, T., Xiong, X., Heston, W. D., Magi-Galluzzi, C., Zhang, J., Klein, E. A. & Almasan, A. The TMPRSS2-ERG Gene Fusion Blocks XRCC4-Mediated Nonhomologous End-Joining Repair and Radiosensitizes Prostate Cancer Cells to PARP Inhibition. *Mol. Cancer Ther.* **14,** 1896–1906 (2015).

85. Fang, J. Y. & Richardson, B. C. The MAPK signalling pathways and colorectal cancer. *Lancet Oncol.* **6,** 322–327 (2005).

86. Olson, J. M. & Hallahan, A. R. p38 MAP kinase: a convergence point in cancer therapy. *Trends Mol. Med.* **10,** 125–129 (2004).

87. Santarpia, L., Lippman, S. M. & El-Naggar, A. K. Targeting the MAPK-RAS-RAF signaling pathway in cancer therapy. *Expert Opin. Ther. Targets* **16,** 103–119 (2012).

88. Pashirzad, M., Khorasanian, R., Fard, M. M., Arjmand, M.-H., Langari, H., Khazaei, M., Soleimanpour, S., Rezayi, M., Ferns, G. A., Hassanian, S. M. & Avan, A. The therapeutic potential of MAPK/ERK inhibitors in the treatment of colorectal cancer. *Curr. Cancer Drug Targets* **21,** 932–943 (2021).

89. Grossi, V., Peserico, A., Tezil, T. & Simone, C. p38α MAPK pathway: a key factor in colorectal cancer therapy and chemoresistance. *World J. Gastroenterol.* **20,** 9744–9758 (2014).

90. Hung, S., Saiakhova, A., Faber, Z. J., Bartels, C. F., Neu, D., Bayles, I., Ojo, E., Hong, E. S., Pontius, W. D., Morton, A. R., Liu, R., Kalady, M. F., Wald, D. N., Markowitz, S. & Scacheri, P. C. Mismatch repair-signature mutations activate gene enhancers

across human colorectal cancer epigenomes. *Elife* **8,** (2019).

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

3094–3100 (2018).

ACKNOWLEDGEMENTS

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

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