

GENETICS

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

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

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 MAPK/AP1 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.

INTRODUCTION

Cancer cells undergo global transcriptional changes resulting from genetic and epigenetic alterations during tumorigenesis (1). While regulatory remodeling can arise from somatic noncoding 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 substantial source of enhancers that could shape cancer-specific gene regulation (10). Many cancers exhibit genome-wide transcriptional reactivation of TEs, which can directly affect 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 long terminal repeat 10 (LTR10) show enhancer activity in many epithelial tumors, and this activity is regulated by signaling pathways involving mitogen-activated protein kinase (MAPK) and activator protein 1 (AP1). 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. Last, 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 (TCGA) 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) (28). 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 (table S1). 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.

Copyright © 2024 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

¹BioFrontiers Institute and Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, CO, USA. ²Division of Medical Oncology, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. ³Division of Reproductive Sciences, Department of Obstetrics and Gynecology, University of Colorado School of Medicine, Aurora, CO, USA.

*Corresponding author. Email: edward.chuong@colorado.edu

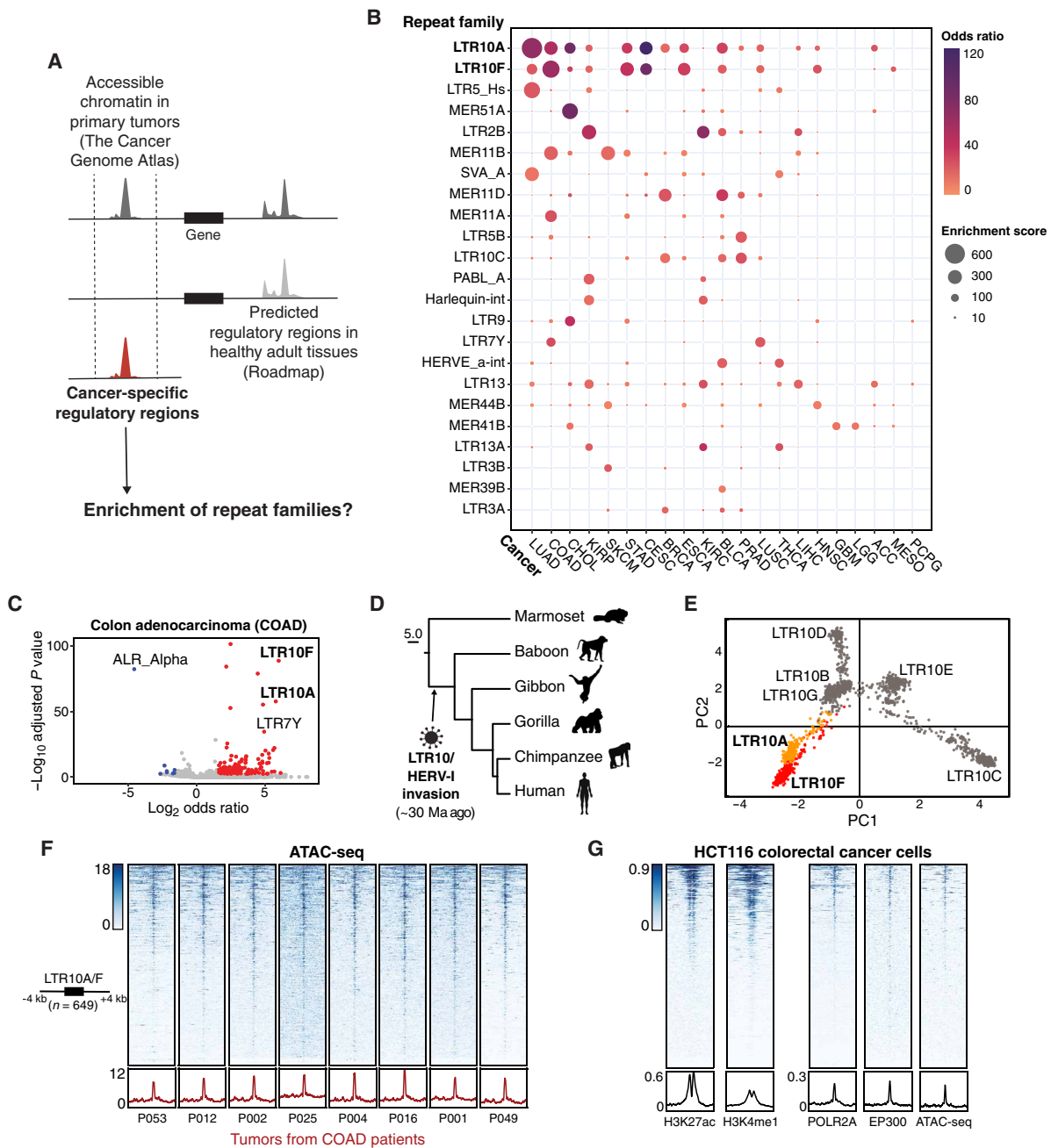


Fig. 1. Pan-cancer epigenomic analysis of TE activity. (A) Pipeline to estimate TE subfamily enrichment within cancer-specific regulatory regions. Aggregate ATAC-seq maps associated with TCGA tumor types were filtered to remove regulatory regions predicted in normal adult tissues. Cancer-specific accessible chromatin regions were tested for enrichment of 1315 repeat subfamilies. (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 on the basis of maximum enrichment score. (C) Enrichment of TE subfamilies within cancer-specific ATAC-seq associated with colon adenocarcinomas from TCGA. Every point represents a TE subfamily. 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 the genomic presence or absence. (E) Principal components analysis based on multiple sequence alignment of LTR10 sequences in the human genome. 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, and 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.

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 and fig. S1). LTR10 elements (including LTR10A-G, $n = 2331$) are derived from the LTR of the gammaretrovirus human endogenous retrovirus I (HERV-I), which integrated into the anthropoid genome 30 million years ago (Fig. 1, D and E) (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 and fig. S2). 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 histone H3 lysine 27 acetylation (H3K27ac) and histone H3 lysine 4 monomethylation (H3K4me1), the transcriptional coactivator p300, and RNA polymerase II occupancy (Fig. 1G). LTR10C elements have previously been identified as a source of p53 binding sites (22, 32). We did not see enhancer-associated chromatin marks at LTR10C elements in HCT116 cells (fig. S3), which are p53-wild-type, but did observe H3K27ac signal at LTR10C elements in several p53-mutant colorectal cancer cell lines (fig. S4 and table S2). Likewise, LTR10A and LTR10F enhancer signal varied across different colorectal cancer cell lines (fig. S4 and table S2), suggesting that LTR10 activation is highly cell line and tumor specific. While most LTR10A and LTR10F elements are not transcribed, some show evidence of transcription as promoters for full-length noncoding HERV-I insertions or cellular transcripts (fig. S5 and table S3). Together, 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 (fig. S6 and table S4). 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, ZNF671, TRIM28, and SETDB1 (Fig. 2, A and B, and tables S5 and S6), which are components of the Krüppel-associated box zinc finger protein (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 (fig. S7). 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. 2, C and D, and table S5) including the fos-like antigen 1 (FOSL1), jun D proto-oncogene (JUND), and activating transcription factor 3 (ATF3) transcription factors. The LTR10A and LTR10F consensus sequences harbor multiple predicted AP1 binding motifs (fig. S8), which most closely resemble binding sites for AP1 component FOSL1 (Fig. 2E), and are enriched within LTR10 elements marked by H3K27ac in HCT116 cells. These 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, although this correlation is largely driven by LTR10A enrichment in lung adenocarcinomas (labeled LUAD in fig. S9; removed as an outlier in fig. S10). In contrast, cancers without LTR10 enrichment show little to no enrichment of AP1 motifs in tumor-specific accessible chromatin (figs. S9 and S10). 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 and fig. S11). In contrast, LTR10 elements did not show global changes in H3K9me3 or H3K27me3 ChIP-seq signal in tumors compared to normal cells (fig. S12). 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.

We further assessed the transcriptional activity of LTR10 elements using matched tumor/normal RNA sequencing (RNA-seq) from 38 patients with colorectal adenocarcinomas from TCGA controlled access data (fig. S13) (27). 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, fig. S14, and table S7). Likewise, AP1 factor *FOSL1* showed a robust and significant increase in expression in tumor versus normal samples (Fig. 2G, fig. S15, and table S7), consistent with our hypothesis that the AP1 complex drives LTR10 transcriptional activity. In contrast, predicted LTR10 repressors such as KRAB-zinc finger *ZNF671* showed a significant decrease in expression in tumor versus normal samples (fig. S16 and table S7). Together, 15 of the 38 patients show a consistent increase in *FOSL1*, LTR10A, and LTR10F transcriptional activity in colorectal tumor cells (table S7).

LTR10 transcription marks tumor-specific epithelial cells

We next investigated LTR10 transcription at the single-cell level. We analyzed an independent cohort of 36 patients with colorectal cancer with publicly available single-cell RNA-seq (scRNA-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

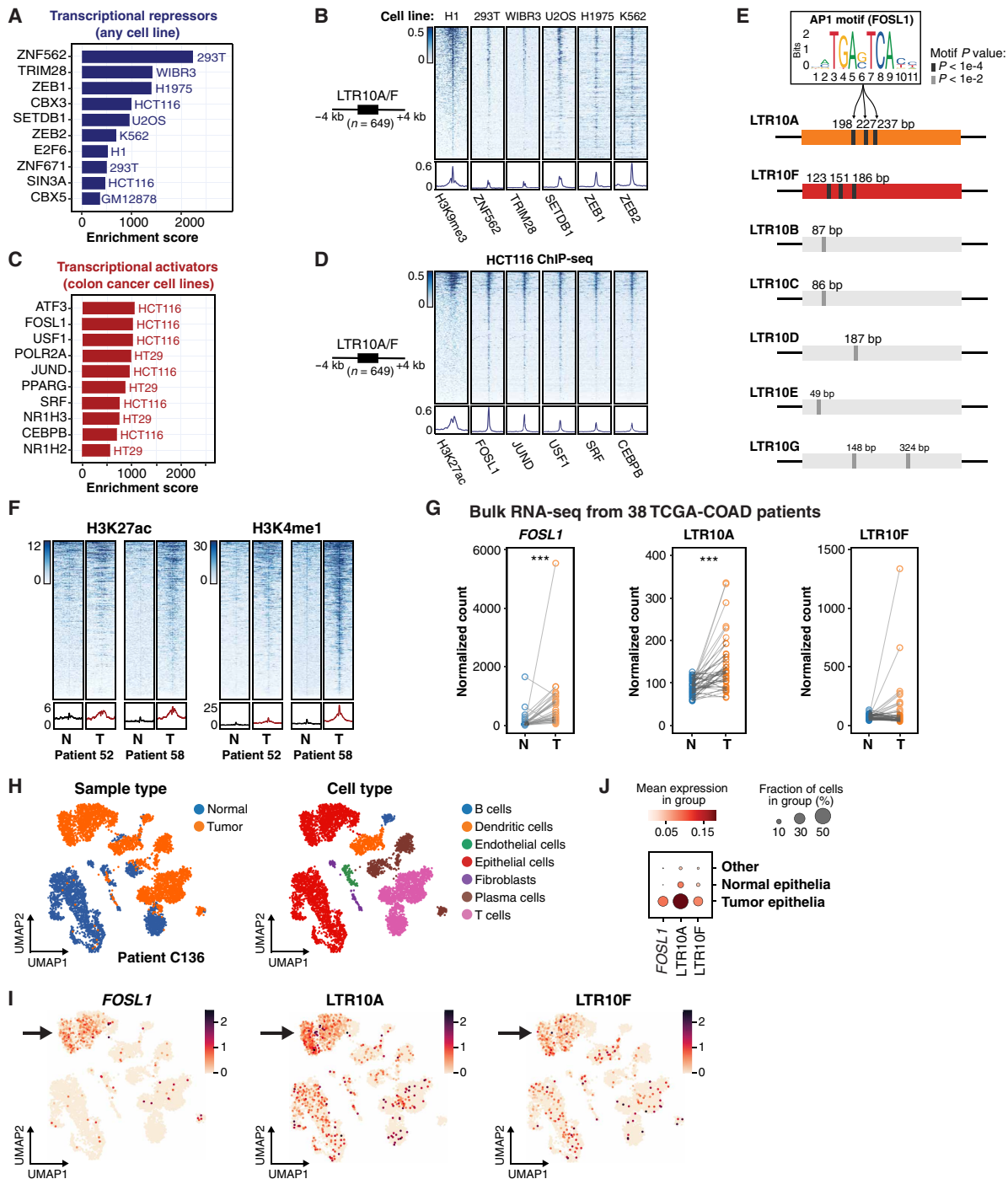


Fig. 2. Regulatory activity of LTR10 in tumor and normal cells. (A) Transcriptional repressors associated with LTR10A/F elements, ranked by enrichment score. (B) Heatmap of ChIP-seq signal from H3K9me3 and repressive factors, over LTR10A/F elements. From left to right: H3K9me3 ChIP-seq (GSE16256), ZNF562 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) and ChIP-seq for FOSL1, JUN, 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. (F) Heatmap of H3K27ac and H3K4me1 ChIP-seq signals from tumor (T) and normal (N) samples from patients AKCC52 and AKCC58 with colorectal cancer (39) 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 (40). UMAPs are colored by tissue type or cell type. (I) UMAP projections of the same patient, colored by the expression of *FOSL1*, LTR10A, or LTR10F. (J) Bubble plot of the same patient, showing the mean expression of *FOSL1*, LTR10A, and LTR10F in tumor epithelia versus normal epithelia.

transcription of LTR10 elements in tumor-specific epithelial cells for 12 of 36 patients (Fig. 2, H to J; figs. S17 to S20; and table S8). We observed coexpression of LTR10 and *FOSL1* in tumor-specific epithelial cells for 10 of these patients (table S8), consistent with a role for AP1 signaling in regulating LTR10 elements. As seen at the bulk level, the expression of LTR10-associated transcriptional repressors (e.g., *ZNF671*, *ZEB1*, and *ZEB2*) was negatively correlated with LTR10 activity (figs. S17 to S19). 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 (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 to 40% of patients with colorectal cancer 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 (fig. S21 and table S9), although we did not observe a significant difference in *FOSL1* expression (fig. S22 and table S9).

Next, we performed survival analyses based on the expression of LTR10 elements or proximal genes. Univariate Cox regression analysis identified the expression of endogenous LTR10 transcripts as a potential risk factor in colorectal tumors (fig. S23 and table S10). Other factors that significantly associated with survival outcome in this dataset of 358 tumors were tumor stage (fig. S24) and patient age (fig. S25). We extended our analysis to multivariate Cox regression models, integrating clinical variables such as patient age, tumor stage, and *KRAS* mutation status with endogenous LTR10 expression and gene set scores. Several multivariate models demonstrated significant predictive ability, with accuracies ranging from 74 to 81% (table S10). Last, we used the Gene Expression Profiling Interactive Analysis (GEPIA) platform (43) to assess the prognostic potential of LTR10-associated genes across a larger dataset of 7288 tumors spanning 21 epithelial cancers. This broader analysis highlighted *FOSL1* and several other genes proximal to LTR10 elements as robust predictors of low survival probability (figs. S26 and S27). Collectively, these findings suggest that LTR10 elements may influence cancer prognosis in epithelial cancers and underscore the need for larger colorectal tumor datasets to accurately determine survival associations.

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 (44). On the basis of our findings that LTR10 elements are bound by AP1, and LTR10 transcriptional activity is correlated with the expression of AP1 factor *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 hours with either tumor necrosis factor- α (TNF α) to stimulate signaling or cobimetinib [a MAPK kinase 1 (MEK1) inhibitor] to inhibit signaling. Consistent with regulation by AP1, cobimetinib treatment caused a decrease in LTR10-driven reporter activity, and TNF α 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 MAPK/AP1 signaling pathway in cancer cells.

To test the role of the AP1 complex in endogenous LTR10 regulation, we used CRISPRi to silence AP1 components *FOSL1* and *JUN*. Using HCT116 cells expressing dCas9-KRAB-MeCP2 (45), we transfected a guide RNA (gRNA) targeting the transcription start site (TSS) of either *FOSL1* or *JUN* and then used RNA-seq to compare gene and TE expression to control cells transfected with a non-targeting gRNA. For each experiment, we first confirmed silencing of the target gene (*FOSL1*: figs. S28 and S29; *JUN*: figs. S30 and S31) and then analyzed TE transcript expression. TE transcripts were summarized at the subfamily level to account for reads mapping to multiple insertions of the same TE (46). This analysis revealed that full-length LTR10/HERV-I elements were significantly down-regulated upon silencing *FOSL1* (Fig. 3B) or *JUN* (fig. S32), 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 both the RNA and chromatin level. We treated HCT116 cells with either cobimetinib or TNF α for 24 hours and profiled each response using RNA-seq and H3K27ac cleavage under targets and release using nuclease (CUT&RUN). Consistent with our reporter assay results, our RNA-seq analysis showed that full-length LTR10/HERV-I transcripts were significantly down-regulated upon cobimetinib treatment (Fig. 3C) and up-regulated upon TNF α treatment (Fig. 3D). Expression of *FOSL1* was likewise down-regulated by cobimetinib and up-regulated by TNF α (figs. S33 and S34). To confirm that LTR10 elements can be therapeutically silenced, we tested how LTR10 transcripts respond to treatment with a second MAPK inhibitor, trametinib. Cobimetinib and trametinib are both inhibitors of MEK1, a key protein kinase in the MAPK signaling pathway. Consistent with our cobimetinib results, our RNA-seq analysis showed that trametinib treatment results in down-regulation of *FOSL1* and LTR10 transcripts after 24 hours of treatment in HCT116 cells [publicly sourced RNA-seq from Gene Expression Omnibus (GEO) accession GSE78519; figs. S35 and S36 and table S11], as well as 21 days of treatment by oral gavage in HCT116 cell line-based xenografts (CDXs) in immunodeficient mice (figs. S37 and S38 and table S11). These results indicate that LTR10 elements are dependent on the MAPK signaling pathway and can be effectively silenced with the use of MEK1 inhibitors.

LTR10 elements showed similar responses based on H3K27ac CUT&RUN signal, exhibiting significant enrichment within the genome-wide set of predicted enhancers down-regulated by cobimetinib or up-regulated by TNF α (Fig. 3, E and F, and figs. S39 and S40). We also observed clear TNF α -induced H3K27ac signal over LTR10 elements in a published dataset of SW480 colorectal cancer

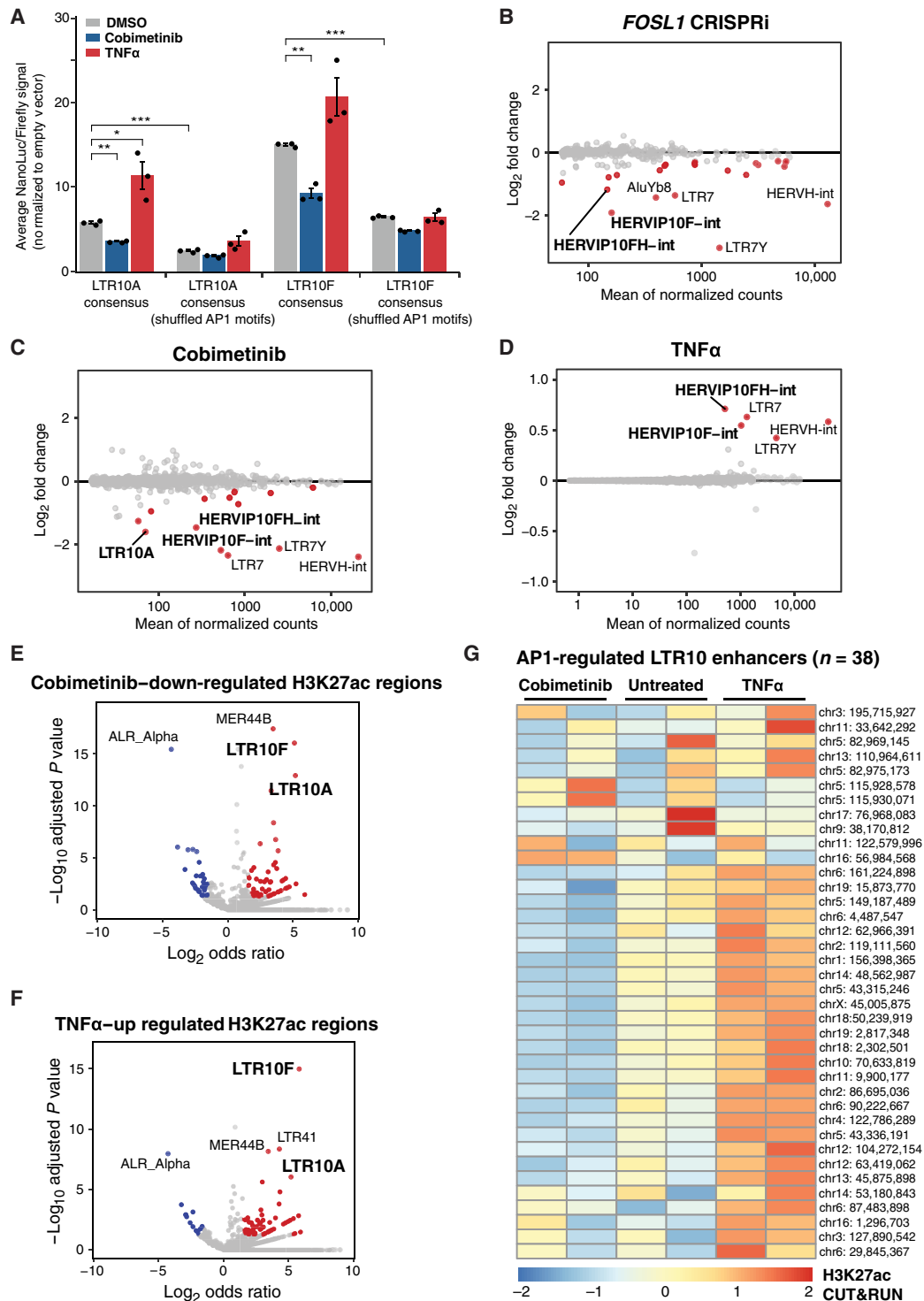


Fig. 3. Control of LTR10 activity by AP1/MAPK signaling. (A) Luciferase reporter assays of LTR10A/F consensus sequences, including sequence variants containing shuffled AP1 motifs. Reporter activity was measured in HCT116 cells treated with dimethyl sulfoxide (DMSO; $n = 3$), cobimetinib ($n = 3$), or TNF α ($n = 3$) for 24 hours. Values are normalized to firefly cotransfection controls and presented as fold change (FC) against the mean values from cells transfected with an empty minimal promoter pNL3.3 vector. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$, two-tailed Student's t test. Error bars denote SEM. (B to D) MA (also known as minus-average plots) plots of TE subfamilies showing significant differential expression in HCT116 cells subject to *FOSL1* silencing (B), 24-hour cobimetinib treatment (C), or 24-hour TNF α treatment (D), based on RNA-seq. Dots are colored in red if they are significant (adjusted $P < 0.05$, $\log_2FC < 0$ for *FOSL1*/cobimetinib and $\log_2FC > 0$ for TNF α). (E) Volcano plot showing TE subfamily enrichment in the set of H3K27ac regions significantly down-regulated by cobimetinib. (F) Volcano plot showing TE subfamily enrichment in the set of H3K27ac regions significantly up-regulated by TNF α . (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.

cells (fig. S41) (47). These results indicate that LTR10 elements represent a notable subset of genome-wide enhancers and transcripts in HCT116 cells that are directly modulated by AP1/MAPK signaling.

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. We first noted that zinc fingers predicted to repress LTR10, such as *ZNF671* and *ZEB2*, had barely detectable expression levels in HCT116 cells (fig. S42). This supports our results from the single-cell and bulk patient analyses and suggests that expression of these zinc fingers is protective (48, 49). We speculate that one of the ways in which loss of these zinc fingers contributes to cancer progression is by derepressing LTR10 enhancers and enabling the pathological activity of LTR10 target genes.

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 (50). Therefore, we defined potential AP1/MAPK-regulated genes using two approaches, based on our RNA-seq data from our *FOSL1* knockdown or TNF α /cobimetinib treatment. We first defined a set of 456 AP1-dependent genes based on being significantly down-regulated by our CRISPRi silencing of the AP1 component *FOSL1* (table S12). We identified LTR10 elements predicted to regulate these genes using the activity by contact model (51) 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 (table S12), including many with established roles in cancer pathophysiology.

In a secondary analysis, we defined 620 MAPK-dependent genes as genes that are both up-regulated by TNF α and down-regulated by cobimetinib, and found 57 LTR10-derived enhancers predicted to regulate 74 (11.9%) of these genes (fig. S43 and table S13). Collectively, we identified a total of 71 distinct LTR10 enhancers (table S14) 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 down-regulation 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* (Figs. 4 and 5; figs. S44 to S53; and tables S15 to S20). For three elements (LTR10.MEF2D, LTR10.MCPH1, and LTR10.KDM6A), the predicted target gene did not show significant expression changes, but we observed down-regulation of other AP1/

MAPK-dependent genes near the element (figs. S45, S49, and S51). Collectively, our characterization of six LTR10 elements verified that 21 genes are regulated by LTR10 enhancers; most (18 of 21) of which are regulated by AP1/MAPK signaling based on our RNA-seq data. These experiments demonstrate that multiple LTR10 enhancers mediate AP1/MAPK-dependent gene expression of nearby genes in HCT116 cells.

From a therapeutic perspective, it is important to know whether LTR10 elements can be broadly targeted. We designed six single guide RNAs (sgRNAs) to target all elements from the LTR10A and LTR10F subfamilies, picking the top sgRNAs based on the highest number of targeted elements, high on-target scores, and low off-targeting to other genomic regions (table S21). We found that six sgRNAs would target at most ~49% of intact LTR10 copies (356 of 734) or ~43% of predicted LTR10 enhancers and promoters (39 of 91) (table S21). This is a substantially lower proportion of elements than are usually targeted by multiplexed epigenome editing (52). It would likely be more effective to target specific LTR10 enhancers that contribute to tumor phenotypes or target the transcriptional activators of LTR10 elements instead (e.g., *FOSL1*).

We focused on two of the CRISPR-validated LTR10 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 down-regulation of *ATG12* as well as the neighboring gene *AP3S1* and eight other genes within 1.5 Mb (Fig. 4, B to D, and table S15). As a separate control, we used CRISPRi to silence the *ATG12* promoter and found highly specific silencing of *ATG12* (figs. S54 and S55 and table S22). 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 (fig. S56). Notably, we observed that multiple genes regulated by LTR10.ATG12 showed similar patterns of transcriptional down-regulation 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 (53–56). Expression of *ATG12* is associated with tumorigenesis and therapy resistance in colorectal and gastric cancer (57, 58), 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 (53).

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

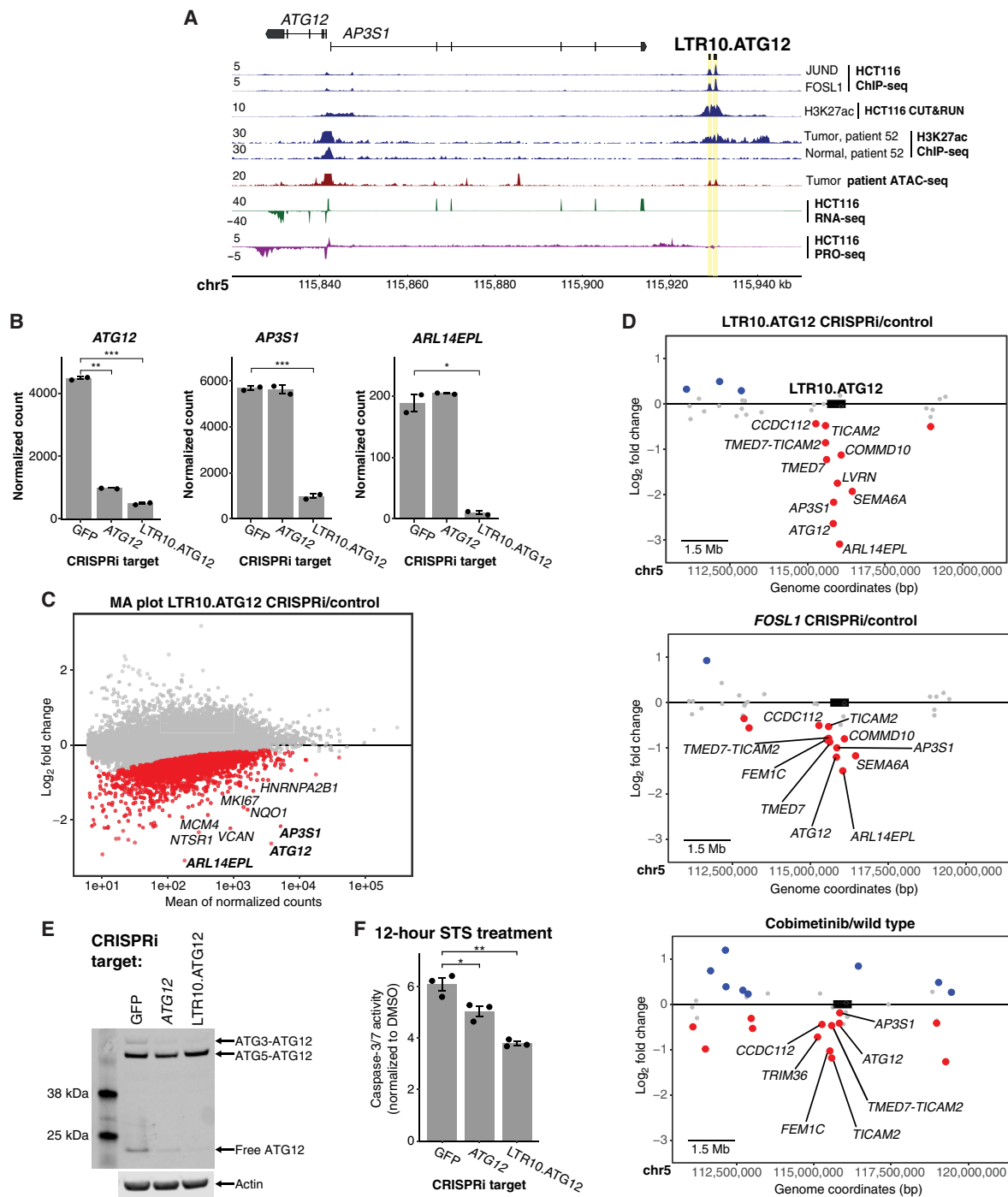


Fig. 4. Functional characterization of LTR10.ATG12 in HCT116 cells. (A) Genome browser screenshot of the *ATG12/AP3S1* locus with the LTR10.ATG12 enhancer labeled. From top to bottom: JUND and FOSL1 ChIP-seq (GSE32465), H3K27ac CUT&RUN (in-house), tumor/normal H3K27ac ChIP-seq from patient AKCC52 (39), tumor ATAC-seq from TCGA-COAD patient P022, HCT116 RNA-seq (in-house), and HCT116 PRO-seq (GSE129501). (B) Normalized RNA-seq expression values of *ATG12*, *AP3S1*, and *ARL14EPL* in dCas9-KRAB-MeCP2 HCT116 cells stably transfected with gRNAs targeting the *ATG12* TSS ($n = 2$), the LTR10.ATG12 element ($n = 2$), or nontargeting [green fluorescent protein (GFP)] control ($n = 2$). $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$, 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 down-regulated 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 down-regulated genes are shown in red; significantly up-regulated genes are shown in blue. Significantly down-regulated 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 hours 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. $*P < 0.05$ and $**P < 0.01$, Welch's t test. Error bars denote SEM.

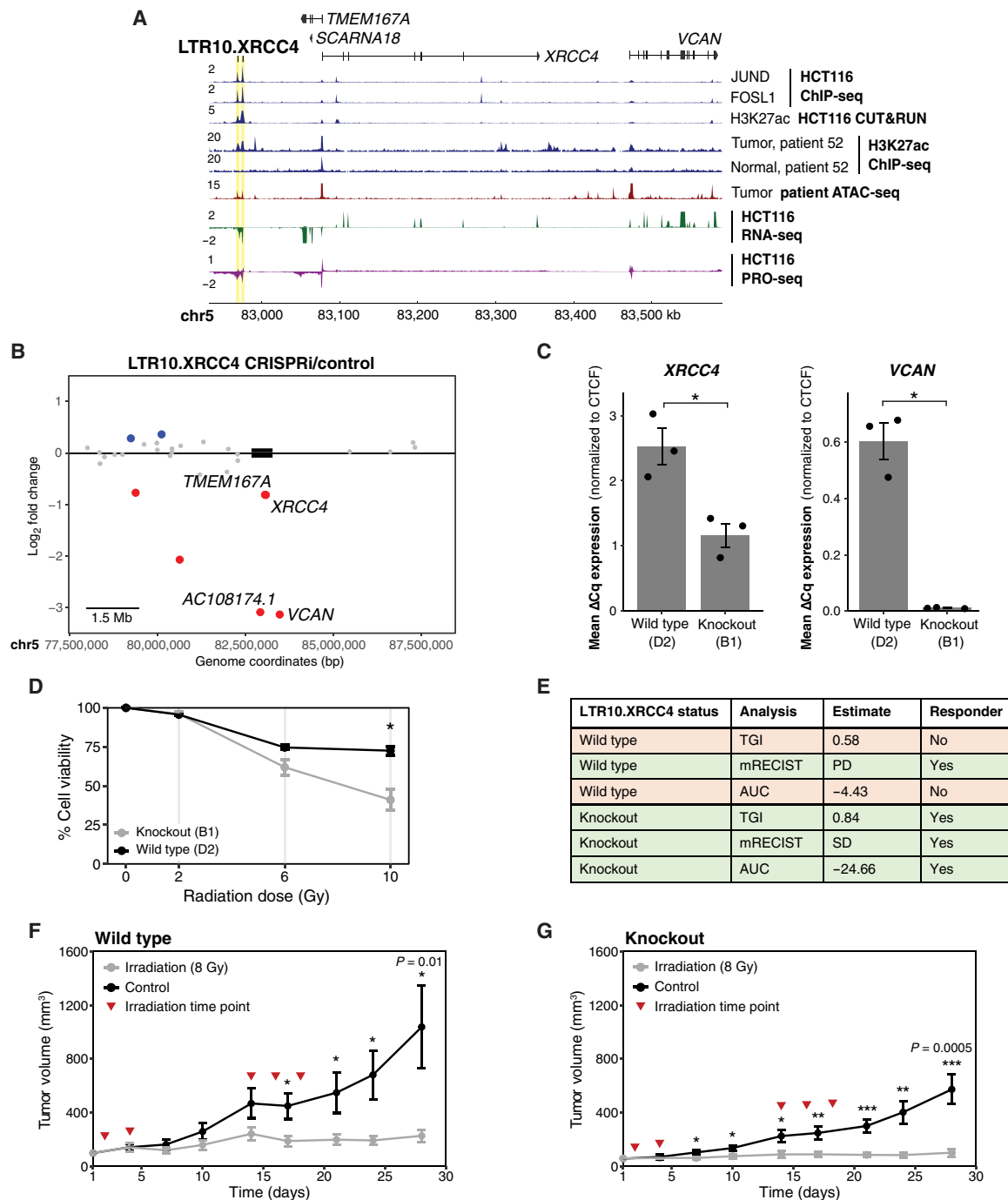


Fig. 5. Functional characterization of LTR10.XRCC4 in HCT116 cells and xenograft models. (A) Genome browser screenshot of the *XRCC4* locus with the LTR10.XRCC4 enhancer labeled. From top to bottom: JUND and FOSL1 ChIP-seq (GSE32465), H3K27ac CUT&RUN (in-house), H3K27ac ChIP-seq from patient AKCC52 (39), ATAC-seq from TCGA-COAD patient P022, HCT116 RNA-seq (in-house), and HCT116 PRO-seq (GSE129501). (B) Scatterplot of gene expression changes at the *XRCC4* locus after CRISPR silencing of the LTR10.XRCC4 enhancer. Significantly down-regulated genes are shown in red; significantly up-regulated genes are shown in blue. Significantly down-regulated genes located within 1.5 Mb of the targeted element are labeled. (C) Quantitative reverse transcription polymerase chain reaction expression values of *XRCC4* and *VCAN* in wild-type HCT116 cells ($n = 3$) and LTR10.XRCC4 knockout cells ($n = 3$). $*P < 0.05$, Welch's t test. Error bars denote SEM. CTCF, CCCTC-binding factor. (D) Dose-response curve showing cell viability in response to 0 to 10 Gy irradiation for LTR10.XRCC4 knockout and wild-type cells. $*P < 0.05$, paired Student's t test. Error bars denote SEM. (E) Classification of responder versus nonresponder for wild-type and LTR10.XRCC4 knockout cells, based on xenograft growth curves of untreated or irradiated mice. Three measures were calculated (100): tumor growth inhibition (TGI), modified response evaluation criteria in solid tumors (mRECIST), and area under the curve (AUC). PD, progressive disease; SD, stable disease. (F and G) Average growth curves for wild-type (F) versus LTR10.XRCC4 knockout (G) xenograft tumors, with and without irradiation, for 28 days. 8 Gy treatment time points (days 2, 4, 14, 16, and 18) are indicated by red triangles. $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$, two-sample t test assuming equal variances. Error bars denote SEM.

dependent on free ATG12 binding to B-cell lymphoma 2 (Bcl-2) (54). 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 (fig. S57), consistent with the lack of knockdown of the ATG5-ATG12 conjugate (56). Our experimental results from silencing both *ATG12* and the enhancer are concordant with previous studies directly silencing *ATG12* using small interfering RNAs in other cancer cell lines (53, 54). Together, these experiments demonstrate that the LTR10.ATG12 enhancer is functionally important for ATG12-dependent activity in HCT116 cells.

We next focused on the LTR10.XRCC4 enhancer, which regulates *XRCC4* and *VCAN* based on our CRISPRi silencing experiment (Fig. 5, A and B; figs. S58 and S59; and table S16). *XRCC4* is a DNA repair gene required for nonhomologous end joining and promotes resistance to chemotherapy and radiation therapy (59–63). *VCAN* is an extracellular matrix protein that promotes tumor metastasis, invasion, and growth (64–66). Both *VCAN* and *XRCC4* have been reported to be regulated by MAPK/AP1 signaling in tumor cells (60, 67), 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 (figs. S60 and S61) and confirmed that *XRCC4* and *VCAN* were significantly down-regulated 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 (59, 68, 69), including in HCT116 cells (70). 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). This is consistent with a previous study showing the role of *XRCC4* in tumor cell survival following irradiation (71). 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 (72); therefore, we tested whether reducing *XRCC4* expression by deleting LTR10.XRCC4 affects tumor growth inhibition by irradiation. 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. Our previous RNA-seq results showed that CRISPR silencing or deletion of the LTR10.XRCC4 enhancer silences both *XRCC4* and *VCAN* (Fig. 5, B and C, and figs. S58 and S59). By the end of the experiment, the specific growth rate of the non-irradiated tumors were similar for both the control and the knockout, allowing us to use all tumors in the study.

Tumors derived from both control and knockout cells showed growth inhibition in response to irradiation (Fig. 5, E to G; figs. S62 to S64; and table S23). However, LTR10.XRCC4 knockout tumors showed more significant overall tumor growth inhibition by irradiation (Fig. 5E), including at earlier time points (Fig. 5, F and G). No significant toxicities were seen in animal weights. We note that although we injected the same number of cells into mice for both the control and knockout (2.5 million), we observed differences in overall tumor sizes between wild-type and knockout tumors, particularly at the start of the experiment (fig. S65). This difference in initial tumor volume may be due to the fact that the LTR10.XRCC4

knockout also reduced expression of *VCAN*, a gene that is often associated with tumor growth and establishment (73, 74). While the difference seen in knockout tumors was modest, our results are consistent with previous studies showing how *XRCC4* knockdown in tumor cells leads to increased sensitivity to radiation or chemotherapy drugs such as cisplatin (59, 60). These findings support a role for the LTR10.XRCC4 enhancer in regulating a clinically relevant tumor phenotype.

LTR10 elements contain highly mutable VNTRs

Last, we investigated variation at LTR10 elements across 15,708 human genomes profiled by the Genome Aggregation database (gnomAD) (75). All LTR10 insertions are fixed, but we observed an unexpected enrichment of >10-bp 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- to 30-bp sequence that includes the AP1 motif (Fig. 6B and fig. S66). Individual LTR10 elements show a wide range of regulatory potential in HCT116 cells, as approximated by peak scores of H3K27ac CUT&RUN and FOSL1 ChIP-seq (fig. S67 and table S24) and demonstrated by the CRISPR-validated LTR10 enhancers. We speculate that the number of AP1 motifs within LTR10 elements may influence their regulatory potential. LTR10 elements annotated in the reference genome show extensive variation in tandem repeat length, with up to 33 copies of the AP1 motif (fig. S68 and table S25). The number of motifs strongly correlates with H3K27ac and FOSL1 binding activity in HCT116 cells (fig. S68), suggesting that tandem repeat length affects AP1-dependent regulation of individual elements. Across the human population, LTR10A and LTR10F elements harbor many rare and common indel structural variants of lengths that follow a 28- to 30-bp periodicity, and this pattern is absent in LTR10C elements which lack the tandem repeat region (Fig. 6, C and D). These elevated levels of polymorphism across copies and individuals are characteristic of unstable tandem repeat regions (76) and suggest that LTR10 VNTR regions may be a common source of genomic regulatory variation.

Accurately genotyping tandem repeat length polymorphisms remains a major challenge using short-read data; therefore, we validated the presence of LTR10 VNTR polymorphisms using structural variant calls generated from long-read whole-genome sequences from 15 individuals (77). We recovered indel structural variants within 24 distinct LTR10A and LTR10F elements, which also showed 28- to 30-bp periodicity (Fig. 6D and fig. S69). We confirmed the presence of additional LTR10 VNTR indels using a separate long-read dataset from 25 Asian individuals (Fig. 6E and figs. S70 to S75) (78). At the LTR10.ATG12 locus, we observed multiple indels supported by both short-read and long-read data that are predicted to affect AP1 motif copy number (Fig. 6E and fig. S71). At a genome-wide level, LTR10 elements were a significantly enriched source of long-read indels, despite being fixed in the population (Fig. 6F). Therefore, expansions or contractions within LTR10 VNTR regions are an underappreciated source of germline genetic variation that could underlie regulatory variation, consistent with polymorphisms recently reported in the VNTR region of SINE-VNTR-Alu (SVA) elements (79).

Last, we searched for evidence of tumor-specific somatic expansions within LTR10 VNTR regions. We analyzed a long-read

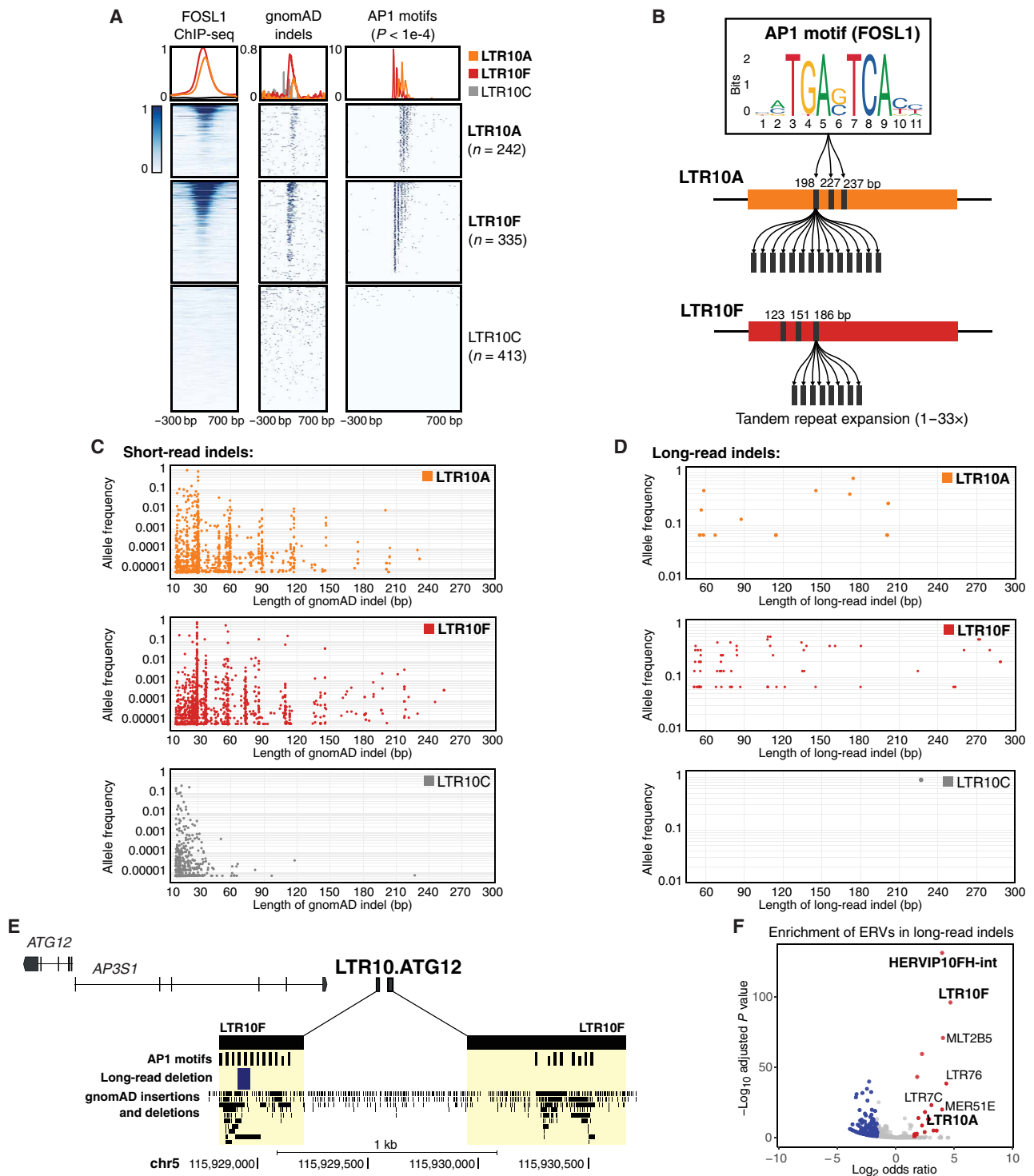


Fig. 6. LTR10 repeat instability and polymorphism. (A) Heatmap of FOSL1 ChIP-seq, gnomAD indels between 10 and 300 bp in length, and AP1 motif matches ($P < 1 \times 10^{-4}$) across LTR10A, LTR10F, and LTR10C elements. Overlapping elements were removed, retaining 990 LTR10 elements total across the three subfamilies. FOSL1 ChIP-seq was obtained from GSE32465. (B) Schematic of VNTR regions within LTR10A and LTR10F elements. (C) Scatterplot of high-confidence gnomAD indels between 10 and 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 [e.g., 58-bp deletion reported in (78)], 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.

whole-genome sequencing dataset generated from matched colorectal tumor and normal tissues from 20 patients (80), using Sniffles2 (81), 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 5 of 20 patients (figs. S76 to S80 and table S26). Three patients showed independent somatic expansions at the same LTR10A locus on chromosome 1 located near gene *GPR137B* (figs. S76, S77, and S80), suggesting that this locus is prone to interindividual variation at both the germline (fig. S69) and somatic levels. We also found evidence of tumor-specific mosaic VNTR deletions in four patients (figs. S81 to S84 and table S26). Our observation that tumors can exhibit both somatic deletions as well as expansions likely reflects the random nature of changes that occur in tumor cells. However, all of the tumor-specific VNTR deletions we identified were small variants affecting one or two AP1 motifs (figs. S81 to S84). In contrast, the predicted tumor-specific expansions were sometimes thousands of base pairs in length, introducing hundreds of additional AP1 motifs [e.g., see structural variant (SV) lengths for patients C553, C568, and C597 in table S26]. 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 (fig. S76 and table S26), as well as two deletions at different LTR10F VNTR loci (figs. S82 and S83). While a larger cohort would be necessary to determine whether 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. Given that AP1 motif copy number has been shown to positively correlate with expression levels (82), we suspect that these VNTR expansions may affect 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, 83). 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 scRNA-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 largely show tumor-specific enhancer activity, a recent study reported that some LTR10 elements also show enhancer activity in the developing human placenta (84), consistent

with the hypothesis that reactivation of placental-specific gene regulatory networks may contribute to cancer pathogenesis (85–87).

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 such as *ATG12*, *XRCC4*, and *VCAN* (57, 59, 63–65, 88–90). 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 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 (44, 91). 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 (92, 93) including colorectal cancer (94, 95), 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 MAPK inhibitors 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.

Last, 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 (79). 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 (96).

Our study has several limitations. First, all epigenomic analyses were performed using datasets derived from bulk tumors. While we were able to infer LTR10 activity using scRNA-seq datasets, analysis of single-cell ATAC-seq (scATAC-seq) datasets would be critical for identifying patient-relevant LTR10 regulatory elements and better understanding their activity in the context of the tumor microenvironment. Second, while we observed functional consequences from silencing or deleting LTR10 elements in HCT116 cells, these findings do not necessarily indicate that the specific enhancers are functionally significant in patient tumors. Further investigation using patient-derived organoid models and immunocompetent mice will be necessary. Third, while we found evidence of germline and

somatic variation of AP1 motif copy number at LTR10 VNTR regions, we did not experimentally demonstrate their effect on gene regulation. There is strong evidence that tandem repeat expansions can affect gene expression in disease (97, 98), and it is well established that having multiple copies of a motif within a regulatory element leads to increased regulatory activity (82, 99), which is consistent with our observed global correlation between AP1 motif copy number and ChIP-seq signal strength (fig. S68).

Despite these limitations, our work uncovers LTR10 elements as an important source of MAPK/AP1-mediated transcriptional dysregulation in colorectal cancer. 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.

MATERIALS AND METHODS

Cell culture

The HCT116 cell line was purchased from American Type Culture Collection (ATCC) and cultured in McCoy's 5A medium supplemented with 10% fetal bovine serum 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 hours with 1 μM cobimetinib, TNFα (100 ng/ml), or dimethyl sulfoxide (DMSO).

CRISPR-mediated silencing and knockout of LTR10s

For CRISPR-mediated silencing (e.g., CRISPRi) of select LTR10 elements and gene 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 cotransfected with the Super PiggyBac transposase expression vector (SPBT) into HCT116 cells. The pB-CAGGS-dCas9-KRAB-MeCP2 construct was a gift from Alejandro Chavez and George Church (Addgene plasmid #110824). Twenty-four hours posttransfection, 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 precomputed CRISPR target guides available on the UCSC Genome Browser hg38 assembly, and complementary oligos were synthesized by Integrated DNA Technologies (IDT). Complementary oligos were designed to generate Bst XI and Bln I overhangs for cloning into PB-CRISPRia, a custom PiggyBac CRISPR gRNA expression plasmid based on the lentiviral construct pCRISPRia (a gift from J. Weissman, Addgene plasmid #84832). Complementary gRNA-containing oligos were hybridized and phosphorylated in a single reaction and then ligated into a PB-CRISPRia expression plasmid linearized with Bst XI and Bln I (New England Biolabs). Chemically competent stable *Escherichia coli* (New England Biolabs) was transformed with 2 μl 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 cotransfected with the PiggyBac transposase vector into the HCT116 dCas9-KRAB-MeCP2 polyclonal stable line. The following number of uniquely mapping gRNA plasmids was designed per target based on the precomputed UCSC hg38 CRISPR target track: green fluorescent protein (GFP) (1), ATG12 (1), FOSL1 (1), JUN (1), LTR10.ATG12 (4), LTR10.FGF2 (2), LTR10.MCPH1 (3), LTR10.MEF2D (2), and LTR10.XRCC4 (2). The same total amount of gRNA plasmid was used for transfections involving one or multiple gRNAs. Twenty-four hours posttransfection, cells were treated with puromycin to select for integration of the sgRNA expression cassette(s). Selection was maintained for 5 days before transcriptional analyses.

For CRISPR-mediated knockout of LTR10.KDM6A, two gRNAs (one specific to each flank of the element) were identified and synthesized as sgRNAs by IDT. For CRISPR-mediated knockout of LTR10.XRCC4, four 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 (Thermo Fisher 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 polymerase chain reaction (PCR) using both flanking and internal primer pairs at the expected deletion site. gRNAs and PCR primers for each candidate are provided in table S27.

Cell autophagy and apoptosis assays

For assaying mitochondrial apoptosis, HCT116 CRISPRi cell lines were treated for 12 hours with STS at 0.5 μM or DMSO (vehicle) followed by measurement of caspase activity via the Caspase-Glo 3/7 assay (Promega). Results are representative of at least three independent experiments. For assaying autophagy, HCT116 CRISPRi cell lines were untreated or treated with bafilomycin A at 10 or 100 nM for 6 and 18 hours, followed by tubule-associated protein 1 light chain 3 beta (LC3B) Western blotting. Results are representative of at least three independent experiments.

Western blots

For ATG12 Western blots, cell lysates were prepared with mammalian protein extraction reagent (MPER) buffer (Thermo Fisher Scientific). For LC3B Western blots, cell lysates were prepared with radioimmunoprecipitation assay buffer. All cell lysates were resuspended in 4× NuPAGE LDS Sample buffer containing a reducing agent (Thermo Fisher 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 μg of protein was loaded per lane. For LC3B Western blots, 2 μg of total protein was loaded per lane. Antibodies used were as follows: ATG12 (catalog no. 4180T, Cell Signaling Technologies), β-actin (catalog no. 3700T, Cell Signaling Technologies), and LC3B (catalog no. NB100-2220, Novus Biologicals). Results are representative of at least three 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 sub-families LTR10A and LTR10F were downloaded from Dfam (v2.0). AP1 motifs within LTR10A and LTR10F were shuffled as follows: LTR10A (first two motifs): cctgagtacc to cagccccgta; LTR10A (third motif): cttagtacc to cagttacc; LTR10F (all three motifs): cctgactcatt to cgtatcctac. Sequences are provided in table S27. Because of their high repeat content, consensus sequences were synthesized as multiple fragments (IDT, 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. Twenty-four hours after transfection, media was replaced with media containing 1 μ M cobimetinib, TNF α (100 ng/ml), or DMSO (vehicle). Twenty-four hours following treatment, luminescence was measured using the NanoGlo Dual Luciferase Reporter Assay System (Promega). All experiments were performed with three treatment replicates per condition in a 96-well plate format. Luminescence readings were first normalized to firefly cotransfection controls and 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 three independent experiments. Barplots are presented as mean \pm SD.

Irradiation experiment

HCT116 control or knockout cells were irradiated using a Faxitron irradiator (model RX-650) at 0, 2, 6, or 10 Gy and 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 to 16 weeks at the start of the study) were purchased from Envigo (Indianapolis, IN) and implanted subcutaneously on the hind flanks with 2.5 million cells in 100 μ l of either HCT116 wild-type or LTR10.XRCC4 CRISPR knockout cells under isoflurane anesthesia with a 23-gauge, 1/2-inch needle. The cell solution injected consisted of 1:1 ratio of RPMI media and cultrex (Cultrex Basement Membrane Extract, PathClear, type 3 from Bio-Techne). We injected wild-type 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 and 100 mm³. Tumor volume was calculated by [(width²) \times length] \times 0.52. Eight days after injection, wild-type tumors had an average tumor volume of 97.88 mm³ (all 20 tumors combined), and LTR10.XRCC4 knockout tumors had an average tumor volume of 56.99 mm³ (all 19 tumors combined; one mouse had to be euthanized). These were then split into four groups: wild-type non-irradiated (average tumor volume, 97.6 mm³), wild-type irradiated (98.1 mm³), knockout non-irradiated (56.1 mm³), and knockout irradiated (57.7 mm³). To reduce variables, all groups were started on irradiation on the same day. Each following irradiation treatment

was also performed on the same day for wild-type and knockout groups. Irradiation treatment consisted of 8 Gy \times 3 fractions on days 2, 4, 14, 16, and 18. Tumor measurements were taken twice weekly using digital calipers, toxicity was monitored by measuring body weight twice weekly, and the study ended at 28 days. Tumor growth inhibition was measured using KulGaP (100).

Trametinib-treated xenograft experiment

HCT116 CDXs were used to compare trametinib treatment to vehicle controls. CDX samples were established from HCT116 cells that were engrafted into two immunodeficient mice. Trametinib was purchased from Selleck Chemicals (Houston, TX). Trametinib was dissolved in 10% Cremophor ethoxylated castor oil (EL) + 5% PEG-400 (polyethylene glycol, molecular weight 400) in water. The drug was dosed at 0.125 mg/kg by oral gavage daily for 21 days. Two replicates, labeled “0L trametinib” and “0R trametinib,” refer to CDXs from mouse number 0, treated with trametinib, with “L” and “R” indicating the left and right flank engraftment sites. Two more replicates, labeled “95L vehicle” and “95R vehicle,” denote the control CDXs from mouse number 95, treated with a vehicle, corresponding to the left and right flanks.

RNA sequencing

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 were 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 5 \times 10⁵ viable cells were used for each CUT&RUN reaction, and pull-downs were generated following the protocol from (101). All buffers were prepared according to the “high-Ca²⁺/low-salt” method using digitonin at a final concentration of 0.05%. The following antibodies were used at the noted dilutions: rabbit anti-mouse immunoglobulin G (1:100) and rabbit anti-H3K27ac (1:100). Protein A-micrococcal nuclease (pA-MNase; gift from S. Henikoff) was added to each sample following primary antibody incubation at a final concentration of 700 ng/ml. Chromatin digestion, release, and extraction were 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 0.8 \times /1.0 \times cleanup using KAPA BioSystems Pure Beads. Purified, adapter-ligated libraries were amplified using the following PCR cycling conditions: 45 s at 98 $^{\circ}$ C, 14 \times (15 s at 98 $^{\circ}$ C, 10 s at 60 $^{\circ}$ C), 60 s at 72 $^{\circ}$ C. Amplified libraries underwent two 1 \times 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) and 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 the Burrows-Wheeler Aligner (BWA) (v0.7.15) and filtered for uniquely mapping reads [mapping quality (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 precision run-on sequencing (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 counts per million (CPM) normalization (ignoring chrX and chrM) and a bin size of 1 bp. Some datasets from TCGA, Encyclopedia of DNA Elements (ENCODE), Cistrome database, and the The Canadian Centre for Epigenome Mapping Technologies (CEMT) Epigenomes Project were downloaded as postprocessed peaks and bigwig files.

TE colocalization analysis

To determine TE subfamily enrichment within regulatory regions, we used GIGGLE (v0.6.3) (102) 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 $-\log_{10}(P \text{ value})$ and $\log_2(\text{odds ratio})$. Enriched TE subfamilies were defined as those with at least 25 overlaps between TE copies and a set of peak regions, odds ratio of more than 10, and a GIGGLE score of more than 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 TCGA (103), 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 and 7_Enh. We did not include fetal tissues (e.g., placental tissues, embryonic stem cells, and trophoblast stem cells) in our set of Roadmap healthy regulatory regions, due to the high levels of basal ERV regulatory activity in these tissues. Last, 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 (104). 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 ($n = 5$). Each discovered motif was searched for similarity to known motifs using the JASPAR 2018 nonredundant 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 1×10^{-4} .

Motif analysis of cancer-specific ATAC-seq peaks from 21 TCGA cancer types was likewise performed using the MEME suite (v5.1.0) (104). 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 tumor type was as follows: adrenocortical carcinoma (ACC; $n = 8123$), bladder urothelial carcinoma (BLCA; $n = 13,737$), breast invasive carcinoma (BRCA; $n = 30,494$), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC; $n = 2449$), cholangiocarcinoma (CHOL; $n = 3012$), colon adenocarcinoma (COAD; $n = 9370$), esophageal carcinoma (ESCA; $n = 12,538$), glioblastoma multiforme (GBM; $n = 4114$), head and neck squamous cell carcinoma (HNSC; $n = 9441$), kidney renal clear cell carcinoma (KIRC; $n = 4807$), kidney renal papillary cell carcinoma (KIRP; $n = 12,315$), brain lower grade glioma (LGG; $n = 3673$), liver hepatocellular carcinoma (LIHC; $n = 8469$), lung adenocarcinoma (LUAD; $n = 16,862$), lung squamous cell carcinoma (LUSC; $n = 15,143$), mesothelioma (MESO; $n = 5275$), pheochromocytoma and paraganglioma (PCPG; $n = 7891$), prostate adenocarcinoma (PRAD; $n = 12,130$), skin cutaneous melanoma (SKCM; $n = 13,710$), stomach adenocarcinoma (STAD; $n = 11,222$), and thyroid carcinoma (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 nonredundant 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 and 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, and -g gene_id to specify the attribute type).

To quantify TE expression at the subfamily level, RNA-seq samples were first realigned to hg38 using hisat2 with `-k 100` to allow multimapping reads and `--no-softclip` to disable soft clipping of reads. TETranscripts (v2.1.4) was then used in multimapping 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 overrepresented TE subfamilies.

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

BAM files of matched tumor/normal RNA-seq datasets from 38 deidentified patients with colon adenocarcinomas were downloaded from TCGA-COAD using the Genomic Data Commons (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 multifactor DESeq2 design of `~patient ID + condition`, where condition was either primary tumor or solid normal tissue. Potential outliers were identified using principal components analysis based on gene counts (e.g., see fig. S13), 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, 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.

Kaplan-Meier survival curves and Cox proportional hazards regression analyses were conducted using normalized expression counts for genes and TEs, along with all available clinical metadata for each tumor. Multiple testing corrections were implemented using both the Bonferroni and Benjamini-Hochberg methods to control for false discovery rates. Multivariate Cox regression models (detailed in table S10) were developed to integrate clinical variables such as patient age, tumor stage, KRAS mutation status, and endogenous LTR10 expression with gene set scores. Gene set scores were derived from average expression or gene set variation analysis of FOSL1-regulated genes ($n = 456$; table S12), MAPK-regulated genes ($n = 620$; table S13), and LTR10 target genes ($n = 120$; table S14).

Pan-cancer survival analysis of LTR10-associated genes

The prognostic potential of LTR10-associated genes was further evaluated across a broader dataset using the GEPIA platform (43). This analysis included 7288 tumors from 21 epithelial cancers identified by TCGA tumor abbreviations [ACC, BLCA, BRCA, CESC, CHOL, COAD, ESCA, HNSC, KIRC, KIRP, LIHC, LUAD, LUSC, ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), PRAD, rectum adenocarcinoma (READ), STAD, THCA, uterine corpus endometrial carcinoma (UCEC), and uterine carcinosarcoma (UCS)]. For each gene of interest, patients were stratified into high- and low-expression groups based on the upper and lower quartiles of gene expression. The survival rates of these groups were then compared to assess the prognostic significance of these genes.

Reanalysis of patient-derived scRNA-seq tumor/normal colon datasets

scRNA-seq datasets of matched tumor/normal colon from 36 deidentified patients with colon adenocarcinomas from (40) 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 and 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 unique molecular identifier (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: (i) fewer than 1200 reads, (ii) fewer than 750 genes, and (iii) 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 five 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 versus N) information.

For each patient, the filtered and merged data were normalized to 10,000 reads per cell, log-transformed, and then clustered. Dimensionality reduction was performed using principal components analysis (`log = True`, `n_pcs = 40`), t-distributed stochastic neighbor embedding (tSNE; `perplexity = 30`, `learning_rate = 1000`, `random_state = 0`, `n_pcs = 40`), and uniform manifold approximation and projection (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 (105) of gene expression markers, with manual verification.

Identification of potential LTR10 promoters

LTR10A/F element coordinates in the hg38 genome were searched against the GeneHancer Regulatory Elements and Gene Interactions database (106) using the UCSC Table Browser (107). GeneHancer

regulatory elements that overlapped an LTR10 element were filtered for element type “Promoter/Enhancer” or “Promoter.” Each predicted promoter was matched to its corresponding gene by filtering for the highest element-gene association score. LTR10-derived promoter candidates were then manually checked for promoter-like epigenomic signatures using the UCSC genome browser. Promoter candidates were categorized on the basis of their orientation with respect to the associated gene (i.e., sense versus antisense) and ranked from highest to lowest GeneHancer element-gene association score (table S3).

Identification 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 (51). Publicly available HCT116 ATAC-seq (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 of more than 0.001 were intersected with H3K27ac-marked LTR10A/F elements. Putative LTR10 enhancers were then checked for proximity (e.g., within 1.5 Mb) to *FOSL1*-regulated genes (i.e., genes that were significantly down-regulated by *FOSL1* knockdown) or MAPK-regulated genes (i.e., genes that were significantly affected by MAPK treatments cobimetinib and TNF α , based on in-house RNA-seq).

sgRNA prediction for targeting all LTR10 copies in the genome

To determine whether LTR10 elements could be broadly targeted, we used a combination of CRISPR-TE (108) and CRISPOR 5.01 (109) to design six sgRNAs for targeting LTR10A/F elements. For CRISPR-TE, we selected “targeting TE subfamily” as the design strategy, “LTR10A” or “LTR10F” as the targeting TE subfamily, and “GRCh38/hg38” as the genome assembly. For CRISPOR, we first retrieved the LTR10A and LTR10F consensus sequences from the Dfam database (v2.0) and then submitted these to CRISPOR with the option “20 bp-NGG-SpCas9.” sgRNA target sites were predicted using CRISPR-TE and verified using Cas-OFFinder allowing a maximum of three mismatches. The top six sgRNAs were picked on the basis of the following criteria: (i) sgRNAs predicted to target the largest number of LTR10A/F copies, (ii) sgRNAs with no off-targeting to other TE families and minimal off-targeting to other genomic regions, (iii) sgRNAs with high on-target scores (primarily the Doench’16 score and the Moreno-Mateos score from CRISPOR), and (iv) sgRNAs targeting LTR10A/F copies predicted to be enhancer/promoter elements. The final six sgRNA sequences and all sgRNA target sites are available in table S21.

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 200 bp for LTR10 sequences (-sortbylength -minseqlength 200), before alignment. MUSCLE (v3.8.1551) was used to align the remaining sequences. Jalview (v2.11.1.4) was used to perform a

principal components 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 to 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 (110), and the HERV-I insertion estimate was confirmed on the basis of the presence or absence of LTR10 sequences across mammals, using BLAST (v2.7.1+) (111).

VNTR identification

gnomAD (v3.1) variant call format (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 subset by size to retain insertions or deletions between 10 and 300 bp 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 to 300 bp for short reads; 50 to 300 bp for long reads, since the minimum length reported by long-read SV callers is 50 bp). 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 (112), 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 adenocarcinomas (80). For each sample, we used minimap2 (113) (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 used 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) (81) 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 variants with SUPP_VEC = 01, which signifies the absence in the normal sample and the presence in the tumor sample). Last, 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.

Statistical analysis

All statistical analyses were performed using R (v3.6.0) and are detailed above.

Supplementary Materials

This PDF file includes:

Figs. S1 to S84

Legends for tables S1 to S27

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S27

REFERENCES AND NOTES

- J. S. You, P. A. Jones, Cancer genetics and epigenetics: Two sides of the same coin? *Cancer Cell* **22**, 9–20 (2012).
- E. Rheinbay, M. M. Nielsen, F. Abascal, J. A. Wala, O. Shapira, G. Tiao, H. Hornshøj, J. M. Hess, R. I. Juul, Z. Lin, L. Feuerbach, R. Sabarinathan, T. Madsen, J. Kim, L. Mularoni, S. Shuai, A. Lanzós, C. Herrmann, Y. E. Maruvka, C. Shen, S. B. Amin, P. Bandopadhyay, J. Bertl, K. A. Borojevich, J. Busanovich, J. Carlevaro-Fita, D. Chakravarty, C. W. Y. Chan, D. Craft, P. Dhingra, K. Diamanti, N. A. Fonseca, A. Gonzalez-Perez, Q. Guo, M. P. Hamilton, N. J. Haradivala, C. Hong, K. Isaev, T. A. Johnson, M. Juul, A. Kahles, A. Kahraman, Y. Kim, J. Komorowski, K. Kumar, S. Kumar, D. Lee, K.-V. Lehmann, Y. Li, E. M. Liu, L. Lochovsky, K. Park, O. Pich, N. D. Roberts, G. Saksena, S. E. Schumacher, N. Sidiropoulos, L. Sieverling, N. Sinnott-Armstrong, C. Stewart, D. Tamborero, J. M. C. Tubio, H. M. Umer, L. Uusküla-Reimand, C. Wadelius, L. Wadi, X. Yao, C.-Z. Zhang, J. Zhang, J. E. Haber, A. Hobolth, M. Imielinski, M. Kellis, M. S. Lawrence, C. von Mering, H. Nakagawa, B. J. Raphael, M. A. Rubin, C. Sander, L. D. Stein, J. M. Stuart, T. Tsunoda, D. A. Wheeler, R. Johnson, J. Reimand, M. Gerstein, E. Khurana, P. J. Campbell, N. López-Bigas; PCAWG Drivers and Functional Interpretation Working Group; PCAWG Structural Variation Working Group, J. Weischenfeldt, R. Beroukhim, I. Martincorena, J. S. Pedersen, G. Getz; PCAWG Consortium, Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578**, 102–111 (2020).
- A. J. Cohen, A. Saiakhova, O. Corradin, J. M. Luppino, K. Lovrenet, C. F. Bartels, J. J. Morrow, S. C. Mack, G. Dhillon, L. Beard, L. Myeroff, M. F. Kalady, J. Willis, J. E. Bradner, R. A. Kerl, N. A. Berger, S. M. Pruett-Miller, S. D. Markowitz, P. C. Scacheri, Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat. Commun.* **8**, 14400 (2017).
- B. Chapuy, M. R. McKeown, C. Y. Lin, S. Monti, M. G. M. Roemer, J. Qi, P. B. Rahl, H. H. Sun, K. T. Yeda, J. G. Doench, E. Reichert, A. L. Kung, S. J. Rodig, R. A. Young, M. A. Shipp, J. E. Bradner, Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790 (2013).
- J.-S. Roe, C.-I. Hwang, T. D. D. Somerville, J. P. Milazzo, E. J. Lee, B. Da Silva, L. Maiorino, H. Tiriác, C. M. Young, K. Miyabayashi, D. Filippini, B. Creighton, R. A. Burkhardt, J. M. Buscaglia, E. J. Kim, J. L. Grem, A. J. Lazenby, J. A. Grunkemeyer, M. A. Hollingsworth, P. M. Grandgenett, M. Egeblad, Y. Park, D. A. Tuveson, C. R. Vakoc, Enhancer reprogramming promotes pancreatic cancer metastasis. *Cell* **170**, 875–888.e20 (2017).
- D. Hnisz, J. Schuijers, C. Y. Lin, A. S. Weintraub, B. J. Abraham, T. I. Lee, J. E. Bradner, R. A. Young, Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell* **58**, 362–370 (2015).
- J. E. Bradner, D. Hnisz, R. A. Young, Transcriptional addiction in cancer. *Cell* **168**, 629–643 (2017).
- I. Sur, J. Taipale, The role of enhancers in cancer. *Nat. Rev. Cancer* **16**, 483–493 (2016).
- W. A. Flavahan, E. Gaskell, B. E. Bernstein, Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017).
- N. Jansz, G. J. Faulkner, Endogenous retroviruses in the origins and treatment of cancer. *Genome Biol.* **22**, 147 (2021).
- Y. Kong, C. M. Rose, A. A. Cass, A. G. Williams, M. Darwish, S. Lianoglou, P. M. Haverty, A.-J. Tong, C. Blanchette, M. L. Albert, I. Mellman, R. Bourgon, J. Greally, S. Jhunjunhwal, H. Chen-Harris, Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nat. Commun.* **10**, 5228 (2019).
- R. Shukla, K. R. Upton, M. Muñoz-Lopez, D. J. Gerhardt, M. E. Fisher, T. Nguyen, P. M. Brennan, J. K. Baillie, A. Collino, S. Ghisletti, S. Sinha, F. Iannelli, E. Radaelli, A. Dos Santos, D. Rapoud, C. Guettier, D. Samuel, G. Natoli, P. Carninci, F. D. Ciccarelli, J. L. Garcia-Perez, J. Faivre, G. J. Faulkner, Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153**, 101–111 (2013).
- B. Rodriguez-Martin, E. G. Alvarez, A. Baez-Ortega, J. Zamora, F. Supek, J. Demeulemeester, M. Santamarina, Y. S. Ju, J. Temes, D. Garcia-Souto, H. Detering, Y. Li, J. Rodriguez-Castro, A. Dueso-Barroso, A. L. Bruzos, S. C. Dentro, M. G. Blanco, G. Contino, D. Ardeljan, M. Tojo, N. D. Roberts, S. Zumalave, P. A. W. Edwards, J. Weischenfeldt, M. Puiggròs, Z. Chong, K. Chen, E. A. Lee, J. A. Wala, K. Raine, A. Butler, S. M. Waszak, F. C. P. Navarro, S. E. Schumacher, J. Monlong, F. Maura, N. Bolli, G. Bourque, M. Gerstein, P. J. Park, D. C. Wedge, R. Beroukhim, D. Torrents, J. O. Korbel, I. Martincorena, R. C. Fitzgerald, P. Van Loo, H. H. Kazazian, K. H. Burns; PCAWG Structural Variation Working Group, P. J. Campbell, J. M. C. Tubio; PCAWG Consortium, Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat. Genet.* **52**, 306–319 (2020).
- N. Rodić, R. Sharma, R. Sharma, J. Zampella, L. Dai, M. S. Taylor, R. H. Hruban, C. A. Iacobuzio-Donahue, A. Maitra, M. S. Torbenson, M. Goggins, I.-M. Shih, A. S. Duffield, E. A. Montgomery, E. Gabrielson, G. J. Netto, T. L. Lotan, A. M. De Marzo, W. Westra, Z. A. Binder, B. A. Orr, G. L. Gallia, C. G. Eberhart, J. D. Boeke, C. R. Harris, K. H. Burns, Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.* **184**, 1280–1286 (2014).
- A. Babaian, D. L. Mager, Endogenous retroviral promoter exaptation in human cancer. *Mob. DNA* **7**, 24 (2016).
- C. Yu, X. Lei, F. Chen, S. Mao, L. Lv, H. Liu, X. Hu, R. Wang, L. Shen, N. Zhang, Y. Meng, Y. Shen, J. Chen, P. Li, S. Huang, C. Lin, Z. Zhang, K. Yuan, ARID1A loss derepresses a group of human endogenous retrovirus-H loci to modulate BRD4-dependent transcription. *Nat. Commun.* **13**, 3501 (2022).
- A. Babaian, M. T. Romanish, L. Gagnier, L. Y. Kuo, M. M. Karimi, C. Steidl, D. L. Mager, Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. *Oncogene* **35**, 2542–2546 (2016).
- B. Lamprecht, K. Walter, S. Kreher, R. Kumar, M. Hummel, D. Lenze, K. Köchert, M. A. Boulhel, J. Richter, E. Soler, R. Stadhouers, K. Jöhrens, K. D. Würster, D. F. Callen, M. F. Harte, M. Giefing, R. Barlow, H. Stein, I. Anagnostopoulos, M. Janz, P. N. Cockerill, R. Siebert, B. Dörken, C. Bonifer, S. Mathas, Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.* **16**, 571–579 (2010).
- B. Edginton-White, P. Cauchy, S. A. Assi, S. Hartmann, A. G. Riggs, S. Mathas, P. N. Cockerill, C. Bonifer, Global long terminal repeat activation participates in establishing the unique gene expression programme of classical Hodgkin lymphoma. *Leukemia* **33**, 1463–1474 (2019).
- H. S. Jang, N. M. Shah, A. Y. Du, Z. Z. Dailey, E. C. Pehrsson, P. M. Godoy, D. Zhang, D. Li, X. Xing, S. Kim, D. O'Donnell, J. I. Gordon, T. Wang, Transposable elements drive widespread expression of oncogenes in human cancers. *Nat. Genet.* **51**, 611–617 (2019).
- J. Attig, J. Pape, L. Doglio, A. Kazachenka, E. Ottina, G. R. Young, K. S. Enfield, I. V. Aramburu, K. W. Ng, N. Faulkner, W. Bolland, V. Papayannopoulos, C. Swanton, G. Kassiotis, Human endogenous retrovirus onco-exaptation counters cancer cell senescence through Calbindin. *J. Clin. Invest.* **133**, e164397 (2023).
- T. Wang, J. Zeng, C. B. Lowe, R. G. Sellers, S. R. Salama, M. Yang, S. M. Burgess, R. K. Brachmann, D. Haussler, 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).
- V. Sundaram, Y. Cheng, Z. Ma, D. Li, X. Xing, P. Edge, M. P. Snyder, T. Wang, Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **24**, 1963–1976 (2014).
- P.-E. Jacques, J. Jeyakani, G. Bourque, The majority of primate-specific regulatory sequences are derived from transposable elements. *PLOS Genet.* **9**, e1003504 (2013).
- Ö. Deniz, M. Ahmed, C. D. Todd, A. Rio-Machin, M. A. Dawson, M. R. Branco, Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. *Nat. Commun.* **11**, 3506 (2020).
- G. Grillo, T. Keshavarzian, S. Linder, C. Arlidge, L. Mout, A. Nand, M. Teng, A. Qamra, S. Zhou, K. J. Kron, A. Murison, J. R. Hawley, M. Fraser, T. H. van der Kwast, G. V. Raj, H. H. He, W. Zwart, M. Lupien, Transposable elements are co-opted as oncogenic regulatory elements by lineage-specific transcription factors in prostate cancer. *Cancer Discov.* **13**, 2470–2487 (2023).

27. M. R. Corces, J. M. Granja, S. Shams, B. H. Louie, J. A. Seoane, W. Zhou, T. C. Silva, C. Groeneveld, C. K. Wong, S. W. Cho, A. T. Satpathy, M. R. Mumbach, K. A. Hoadley, A. G. Robertson, N. C. Sheffield, I. Felau, M. A. A. Castro, B. P. Berman, L. M. Staudt, J. C. Zenklusen, P. W. Laird, C. Curtis; Cancer Genome Atlas Analysis Network, W. J. Greenleaf, H. Y. Chang, The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
28. Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfennig, X. Wang, M. Clausnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shores, C. B. Epstein, E. Gjonneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthal, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
29. B. Akhtar-Zaidi, R. Cowper-Sal-lari, O. Corradin, A. Saiakhova, C. F. Bartels, D. Balasubramanian, L. Myeroff, J. Lutterbaugh, A. Jarrar, M. F. Kalady, J. Willis, J. H. Moore, P. J. Tesar, T. Laframboise, S. Markowitz, M. Lupien, P. C. Scacheri, Epigenomic enhancer profiling defines a signature of colon cancer. *Science* **336**, 736–739 (2012).
30. L. Baranello, D. Wojtowicz, K. Cui, B. N. Devaiah, H.-J. Chung, K. Y. Chan-Salis, R. Guha, K. Wilson, X. Zhang, H. Zhang, J. Piotrowski, C. J. Thomas, D. S. Singer, B. F. Pugh, Y. Pommier, T. M. Przytycka, F. Kouzine, B. A. Lewis, K. Zhao, D. Levens, RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. *Cell* **165**, 357–371 (2016).
31. R. Zheng, C. Wan, S. Mei, Q. Qin, Q. Wu, H. Sun, C.-H. Chen, M. Brown, X. Zhang, C. A. Meyer, X. S. Liu, Cistrome data browser: Expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* **47**, D729–D735 (2019).
32. E. J. Grow, B. D. Weaver, C. M. Smith, J. Guo, P. Stein, S. C. Shadle, P. G. Hendrickson, N. E. Johnson, R. J. Butterfield, R. Menafra, S. L. Kloet, S. M. van der Maarel, C. J. Williams, B. R. Cairns, p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. *Nat. Genet.* **53**, 1207–1220 (2021).
33. M. Imbeault, P.-Y. Helleboid, D. Trono, KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550–554 (2017).
34. B. E. Bernstein, J. A. Stamatoyannopoulos, J. F. Costello, B. Ren, A. Milosavljevic, A. Meissner, M. Kellis, M. A. Marra, A. L. Beaudet, J. R. Ecker, P. J. Farnham, M. Hirst, E. S. Lander, T. S. Mikkelsen, J. A. Thomson, 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. R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
37. M. Bruno, M. Mahgoub, T. S. Macfarlan, The arms race between KRAB–Zinc finger proteins and endogenous retroelements and its impact on mammals. *Annu. Rev. Genet.* **53**, 393–416 (2019).
38. E. Orouji, A. T. Raman, A. K. Singh, A. Sorokin, E. Arslan, A. K. Ghosh, J. Schulz, C. Terranova, S. Jiang, M. Tang, M. Maitiuheti, S. C. Callahan, P. Barrodia, K. Tomczak, Y. Jiang, Z. Jiang, J. S. Davis, S. Ghosh, H. M. Lee, L. Reyes-Urbe, C. Chang, Y. Liu, H. Chen, A. Azhdarinia, J. Morris, E. Vilar, K. S. Carmon, S. E. Kopetz, K. Rai, Chromatin state dynamics confers specific therapeutic strategies in enhancer subtypes of colorectal cancer. *Gut* **71**, 938–949 (2022).
39. D. Bujold, D. A. de Lima Morais, C. Gauthier, C. Côté, M. Caron, T. Kwan, K. C. Chen, J. Laperle, A. N. Markovits, T. Pastinen, B. Caron, A. Veilleux, P.-É. Jacques, G. Bourque, The International Human Epigenome Consortium Data Portal. *Cell Syst.* **3**, 496–499.e2 (2016).
40. K. Pelka, M. Hofree, J. H. Chen, S. Sarkizova, J. D. Pirl, V. Jorgji, A. Bejnood, D. Dionne, W. H. Ge, K. H. Xu, S. X. Chao, D. R. Zollinger, D. J. Lieb, J. W. Reeves, C. A. Fuhrman, M. L. Hoang, T. Delorey, L. T. Nguyen, J. Waldman, M. Klapholz, I. Wakiro, O. Cohen, J. Albers, C. S. Millie, M. S. Cuoco, J. Wu, M.-J. Su, J. Yeung, B. Vijaykumar, A. M. Magnuson, N. Asinovski, T. Moll, M. N. Goder-Reiser, A. S. Applebaum, L. K. Kraus, L. K. DelloStritto, S. L. Denning, S. T. Phillips, E. K. Hill, J. K. Meehan, D. T. Frederick, T. Sharova, A. Kanodia, E. Z. Todres, J. Jané-Valbuena, M. Biton, B. Izar, C. D. Lambden, T. E. Clancy, R. Bleday, N. Melnitchouk, J. Irani, H. Kunitake, D. L. Berger, A. Srivastava, J. L. Hornick, S. Ogino, A. Rotem, S. Vignneau, B. E. Johnson, R. B. Corcoran, A. H. Sharpe, V. K. Kuchroo, K. Ng, M. Giannakis, L. T. Nieman, G. M. Boland, A. J. Aguirre, A. C. Anderson, O. Rozenblatt-Rosen, A. Regev, N. Hacohen, Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* **184**, 4734–4752.e20 (2021).
41. J. He, I. A. Babarinde, L. Sun, S. Xu, R. Chen, J. Shi, Y. Wei, Y. Li, G. Ma, Q. Zhuang, A. P. Hutchins, J. Chen, 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. G. Zhu, L. Pei, H. Xia, Q. Tang, F. Bi, Role of oncogenic KRAS in the prognosis, diagnosis and treatment of colorectal cancer. *Mol. Cancer* **20**, 143 (2021).
43. Z. Tang, C. Li, B. Kang, G. Gao, C. Li, Z. Zhang, GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**, W98–W102 (2017).
44. E. F. Wagner, A. R. Nebreda, Signal integration by JNK and p38 MAPK pathways in cancer development. *Nat. Rev. Cancer* **9**, 537–549 (2009).
45. N. C. Yeo, A. Chavez, A. Lance-Byrne, Y. Chan, D. Menn, D. Milanova, C.-C. Kuo, X. Guo, S. Sharma, A. Tung, R. J. Cecchi, M. Tuttle, S. Pradhan, E. T. Lim, N. Davidsohn, M. R. Ebrahimkhani, J. J. Collins, N. E. Lewis, S. Kiani, G. M. Church, An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods* **15**, 611–616 (2018).
46. Y. Jin, O. H. Tam, E. Paniagua, M. Hammell, Tetrascripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599 (2015).
47. H. Rahnamoun, H. Lu, S. H. Duttke, C. Benner, C. K. Glass, S. M. Lauberth, Mutant p53 shapes the enhancer landscape of cancer cells in response to chronic immune signaling. *Nat. Commun.* **8**, 754 (2017).
48. H. Xie, Z. Wu, Z. Li, Y. Huang, J. Zou, H. Zhou, Significance of ZEB2 in the immune microenvironment of colon cancer. *Front. Genet.* **13**, 995333 (2022).
49. J. Zhang, J. Luo, H. Jiang, T. Xie, J. Zheng, Y. Tian, R. Li, B. Wang, J. Lin, A. Xu, X. Huang, Y. Yuan, The tumor suppressor role of zinc finger protein 671 (ZNF671) in multiple tumors based on cancer single-cell sequencing. *Front. Oncol.* **9**, 1214 (2019).
50. H. K. Koul, M. Pal, S. Koul, Role of p38 MAP kinase signal transduction in solid tumors. *Genes Cancer* **4**, 342–359 (2013).
51. C. P. Fulco, J. Nasser, T. R. Jones, G. Munson, D. T. Bergman, V. Subramanian, S. R. Grossman, R. Anyoha, B. R. Doughty, T. A. Patwardhan, T. H. Nguyen, M. Kane, E. M. Perez, N. C. Durand, C. A. Lareau, E. K. Stamenova, E. L. Aiden, E. S. Lander, J. M. Engreitz, Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
52. J. Yang, L. Cook, Z. Chen, Systematic evaluation of retroviral LTRs as cis-regulatory elements in mouse embryos. *Cell Rep.* **43**, 113775 (2024).
53. M. Haller, A. K. Hock, E. Giampazolias, A. Oberst, D. R. Green, J. Debnath, K. M. Ryan, K. H. Vousden, S. W. G. Tait, Ubiquitination and proteasomal degradation of ATG12 regulates its proapoptotic activity. *Autophagy* **10**, 2269–2278 (2014).
54. A. D. Rubinstein, M. Eisenstein, Y. Ber, S. Bialik, A. Kimchi, The autophagy protein Atg12 associates with antiapoptotic Bcl-2 family members to promote mitochondrial apoptosis. *Mol. Cell* **44**, 698–709 (2011).
55. L. Radosheвич, L. Murrow, N. Chen, E. Fernandez, S. Roy, C. Fung, J. Debnath, ATG12 conjugation to ATG3 regulates mitochondrial homeostasis and cell death. *Cell* **142**, 590–600 (2010).
56. T. Hanada, N. N. Noda, Y. Satomi, Y. Ichimura, Y. Fujioka, T. Takao, F. Inagaki, Y. Ohsumi, The Atg12-Atg5 conjugate has a novel E3-like activity for protein lipidation in autophagy. *J. Biol. Chem.* **282**, 37298–37302 (2007).
57. J. L. Hu, G. Y. He, X. L. Lan, Z. C. Zeng, J. Guan, Y. Ding, X. L. Qian, W. T. Liao, Y. Q. Ding, L. Liang, Inhibition of ATG12-mediated autophagy by miR-214 enhances radiosensitivity in colorectal cancer. *Oncogenesis* **7**, 16 (2018).
58. H. YiRen, Y. YingCong, Y. Sunwu, L. Keqin, T. Xiaochun, C. Senrui, C. Ende, L. XiZhou, C. Yanfan, Long noncoding RNA MALAT1 regulates autophagy associated chemoresistance via miR-23b-3p sequestration in gastric cancer. *Mol. Cancer* **16**, 174 (2017).
59. Z. Zheng, W. L. Ng, X. Zhang, J. J. Olson, C. Hao, W. J. Curran, Y. Wang, RNAi-mediated targeting of noncoding and coding sequences in DNA repair gene messages efficiently radiosensitizes human tumor cells. *Cancer Res.* **72**, 1221–1228 (2012).
60. M. Xu, X. Huang, C. Zheng, J. Long, Q. Dai, Y. Chen, J. Lu, C. Pan, S. Yao, J. Li, Platinum-resistant ovarian cancer is vulnerable to the cJUN-XRCC4 pathway inhibition. *Cancer* **14**, 6068 (2022).
61. A. Zeng, Z. Wei, W. Yan, J. Yin, X. Huang, X. Zhou, R. Li, F. Shen, W. Wu, X. Wang, Y. You, Exosomal transfer of miR-151a enhances chemosensitivity to temozolomide in drug-resistant glioblastoma. *Cancer Lett.* **436**, 10–21 (2018).
62. M. Goldstein, M. B. Kastan, The DNA damage response: Implications for tumor responses to radiation and chemotherapy. *Annu. Rev. Med.* **66**, 129–143 (2015).
63. Y. Gao, D. O. Ferguson, W. Xie, J. P. Manis, J. Sekiguchi, K. M. Frank, J. Chaudhuri, J. Horner, R. A. DePinho, F. W. Alt, Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. *Nature* **404**, 897–900 (2000).

64. T.-L. Yeung, C. S. Leung, K.-K. Wong, G. Samimi, M. S. Thompson, J. Liu, T. M. Zaid, S. Ghosh, M. J. Birrer, S. C. Mok, TGF- β modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. *Cancer Res.* **73**, 5016–5028 (2013).
65. A. Papadas, G. Arauz, A. Cicala, J. Wiesner, F. Asimakopoulos, Versican and versican-matrikines in cancer progression, inflammation, and immunity. *J. Histochem. Cytochem.* **68**, 871–885 (2020).
66. Y. Zhang, X. Zou, W. Qian, X. Weng, L. Zhang, L. Zhang, S. Wang, X. Cao, L. Ma, G. Wei, Y. Wu, Z. Hou, Enhanced PAPSS2/VCAN sulfation axis is essential for Snail-mediated breast cancer cell migration and metastasis. *Cell Death Differ.* **26**, 565–579 (2019).
67. C. Domenzain-Reyna, D. Hernández, L. Miquel-Serra, M. J. Docampo, C. Badenas, A. Fabra, A. Bassols, 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).
68. Z. Li, T. Otevrel, Y. Gao, H. L. Cheng, B. Seed, T. D. Stamato, G. E. Taccioli, F. W. Alt, The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **83**, 1079–1089 (1995).
69. S. E. Critchlow, R. P. Bowater, S. P. Jackson, Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr. Biol.* **7**, 588–598 (1997).
70. T. Katsube, M. Mori, H. Tsuji, T. Shiomu, N. Shiomu, M. Onoda, Differences in sensitivity to DNA-damaging agents between XRCC4- and artemis-deficient human cells. *J. Radiat. Res.* **52**, 415–424 (2011).
71. Y. Wen, G. Dai, L. Wang, K. Fu, S. Zuo, Silencing of XRCC4 increases radiosensitivity of triple-negative breast cancer cells. *Biosci. Rep.* **39**, BSR20180893 (2019).
72. J.-D. Leu, B.-S. Wang, S.-J. Chiu, C.-Y. Chang, C.-C. Chen, F.-D. Chen, S. Avirmed, Y.-J. Lee, Combining fisetin and ionizing radiation suppresses the growth of mammalian colorectal cancers in xenograft tumor models. *Oncol. Lett.* **12**, 4975–4982 (2016).
73. K. Asano, C. M. Nelson, S. Nandadasa, N. Aramaki-Hattori, D. J. Lindner, T. Alban, J. Inagaki, T. Ohtsuki, T. Ohashi, S. S. Apte, S. Hirohata, Stromal versican regulates tumor growth by promoting angiogenesis. *Sci. Rep.* **7**, 17225 (2017).
74. A. G. Pappas, S. Magkouta, I. S. Pateras, I. Skianis, C. Moschos, M. E. Vazakidou, K. Psarra, V. G. Gorgoulis, I. Kalomenidis, Versican modulates tumor-associated macrophage properties to stimulate mesothelioma growth. *Oncoimmunology* **8**, e1537427 (2019).
75. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alfoldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferreria, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski; Genome Aggregation Database Consortium, B. M. Neale, M. J. Daly, D. G. MacArthur, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
76. M. Gymrek, A genomic view of short tandem repeats. *Curr. Opin. Genet. Dev.* **44**, 9–16 (2017).
77. P. A. Audano, A. Sulovari, T. A. Graves-Lindsay, S. Cantsilieris, M. Sorensen, A. E. Welch, M. L. Dougherty, B. J. Nelson, A. Shah, S. K. Dutcher, W. C. Warren, V. Magrini, S. D. McGrath, Y. I. Li, R. K. Wilson, E. E. Eichler, Characterizing the major structural variant alleles of the human genome. *Cell* **176**, 663–675.e19 (2019).
78. C. Quan, Y. Li, X. Liu, Y. Wang, J. Ping, Y. Lu, G. Zhou, Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression. *Genome Biol.* **22**, 159 (2021).
79. E. J. van Bree, R. L. F. P. Guimarães, M. Lundberg, E. R. Blujdea, J. L. Rosenkrantz, F. T. G. White, J. Poppinga, P. Ferrer-Raventós, A.-F. E. Schneider, I. Clayton, D. Haussler, M. J. T. Reinders, H. Holstege, A. D. Ewing, C. Moses, F. M. J. Jacobs, A hidden layer of structural variation in transposable elements reveals potential genetic modifiers in human disease-risk loci. *Genome Res.* **32**, 656–670 (2022).
80. L. Xu, X. Wang, X. Lu, F. Liang, Z. Liu, H. Zhang, X. Li, S. Tian, L. Wang, Z. Wang, Long-read sequencing identifies novel structural variations in colorectal cancer. *PLOS Genet.* **19**, e1010514 (2023).
81. M. Smolka, L. F. Paulin, C. M. Grochowski, M. Mahmoud, S. Behera, M. Gandhi, K. Hong, D. Pehlivan, S. W. Scholz, C. M. B. Carvalho, C. Proukakis, F. J. Sedlazeck, Comprehensive structural variant detection: From mosaic to population-level. bioRxiv 487055 [Preprint] (2022). <https://doi.org/10.1101/2022.04.04.487055>.
82. I. Georgakopoulos-Soares, C. Deng, Y. Agarwal, C. S. Y. Chan, J. Zhao, F. Inoue, N. Ahituv, Transcription factor binding site orientation and order are major drivers of gene regulatory activity. *Nat. Commun.* **14**, 2333 (2023).
83. K. Karttunen, D. Patel, J. Xia, L. Fei, K. Palin, L. Aaltonen, B. Sahu, Transposable elements as tissue-specific enhancers in cancers of endodermal lineage. *Nat. Commun.* **14**, 5313 (2023).
84. J. M. Frost, S. M. Amante, H. Okae, E. M. Jones, B. Ashley, R. M. Lewis, J. K. Cleal, M. P. Caley, T. Arima, T. Maffucci, M. R. Branco, Regulation of human trophoblast gene expression by endogenous retroviruses. *Nat. Struct. Mol. Biol.* **30**, 527–538 (2023).
85. C. F. Lynch-Sutherland, A. Chatterjee, P. A. Stockwell, M. R. Eccles, E. C. Macaulay, Reawakening the developmental origins of cancer through transposable elements. *Front. Oncol.* **10**, 468 (2020).
86. V. Costanzo, A. Bardelli, S. Siena, S. Abrignani, Exploring the links between cancer and placenta development. *Open Biol.* **8**, 180081 (2018).
87. G. P. Wagner, Kshitiz, A. Dighe, A. Levchenko, The coevolution of placentation and cancer. *Annu. Rev. Anim. Biosci.* **10**, 259–279 (2022).
88. D. C. Rubinsztein, P. Codogno, B. Levine, Autophagy modulation as a potential therapeutic target for diverse diseases. *Nat. Rev. Drug Discov.* **11**, 709–730 (2012).
89. D.-T. Bau, M.-D. Yang, Y.-A. Tsou, S.-S. Lin, C.-N. Wu, H.-H. Hsieh, R.-F. Wang, C.-W. Tsai, W.-S. Chang, H.-M. Hsieh, S.-S. Sun, R.-Y. Tsai, Colorectal cancer and genetic polymorphism of DNA double-strand break repair gene XRCC4 in Taiwan. *Anticancer Res.* **30**, 2727–2730 (2010).
90. P. Chatterjee, G. S. Choudhary, T. Alswillah, X. Xiong, W. D. Heston, C. Magi-Galluzzi, J. Zhang, E. A. Klein, A. Almasan, 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).
91. J. Y. Fang, B. C. Richardson, The MAPK signalling pathways and colorectal cancer. *Lancet Oncol.* **6**, 322–327 (2005).
92. J. M. Olson, A. R. Hallahan, p38 MAP kinase: A convergence point in cancer therapy. *Trends Mol. Med.* **10**, 125–129 (2004).
93. L. Santarpia, S. M. Lippman, A. K. El-Naggar, Targeting the MAPK-RAS-RAF signaling pathway in cancer therapy. *Expert Opin. Ther. Targets* **16**, 103–119 (2012).
94. M. Pashirzad, R. Khorasani, M. M. Fard, M.-H. Arjmand, H. Langari, M. Khazaei, S. Soleimanpour, M. Rezayi, G. A. Ferns, S. M. Hassanian, A. Avan, The therapeutic potential of MAPK/ERK inhibitors in the treatment of colorectal cancer. *Curr. Cancer Drug Targets* **21**, 932–943 (2021).
95. V. Grossi, A. Peserico, T. Tezil, C. Simone, p38 α MAPK pathway: A key factor in colorectal cancer therapy and chemoresistance. *World J. Gastroenterol.* **20**, 9744–9758 (2014).
96. S. Hung, A. Saiakhova, Z. J. Faber, C. F. Bartels, D. Neu, I. Bayles, E. Ojo, E. S. Hong, W. D. Pontius, A. R. Morton, R. Liu, M. F. Kalady, D. N. Wald, S. Markowitz, P. C. Scacheri, Mismatch repair-signature mutations activate gene enhancers across human colorectal cancer epigenomes. *eLife* **8**, e40760 (2019).
97. A. Mitina, M. Khan, R. Lesurf, Y. Yin, W. Engchuan, O. Hamdan, G. Pellecchia, B. Trost, I. Backstrom, K. Guo, L. M. Palotto, P. H. Lam Doong, Z. Wang, T. Nalpathamkalam, B. Thiruvahindrapuram, T. Papaz, C. E. Pearson, J. Ragoussis, P. Subbarao, M. B. Azad, S. E. Turvey, P. Mandhane, T. J. Moraes, E. Simons, S. W. Scherer, J. Loughheed, T. Mondal, J. Smythe, L. Altamirano-Diaz, E. Oechslin, S. Mital, R. K. C. Yuen, Genome-wide enhancer-associated tandem repeats are expanded in cardiomyopathy. *EBioMedicine* **101**, 105027 (2024).
98. C. Depienne, J.-L. Mandel, 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am. J. Hum. Genet.* **108**, 764–785 (2021).
99. D. X. Quang, M. R. Erdos, S. C. J. Parker, F. S. Collins, Motif signatures in stretch enhancers are enriched for disease-associated genetic variants. *Epigenetics Chromatin* **8**, 23 (2015).
100. J. Ortmann, L. Rampásek, E. Tai, A. S. Mer, R. Shi, E. L. Stewart, C. Mascaux, A. Fares, N.-A. Pham, G. Beri, C. Eeles, D. Tkachuk, C. Ho, S. Sakashita, J. Weiss, X. Jiang, G. Liu, D. W. Cescon, C. A. O'Brien, S. Guo, M.-S. Tsao, B. Haibe-Kains, A. Goldenberg, Assessing therapy response in patient-derived xenografts. *Sci. Transl. Med.* **13**, eabf4969 (2021).
101. M. P. Meers, T. D. Bryson, J. G. Henikoff, S. Henikoff, Improved CUT&RUN chromatin profiling tools. *eLife* **8**, e46314 (2019).
102. R. M. Layer, B. S. Pedersen, T. DiSera, G. T. Marth, J. Gertz, A. R. Quinlan, GIGGLE: A search engine for large-scale integrated genome analysis. *Nat. Methods* **15**, 123–126 (2018).
103. The Cancer Genome Atlas Research Network, J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. M. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J. M. Stuart, The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
104. T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, W. S. Noble, MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
105. O. Franzén, L.-M. Gan, J. L. M. Björkegren, PanglaoDB: A web server for exploration of mouse and human single-cell RNA sequencing data. *Database* **2019**, baz046 (2019).
106. S. Fishilevich, R. Nudel, N. Rappaport, R. Hadar, I. Plaschkes, T. Iny Stein, N. Rosen, A. Kohn, M. Twik, M. Safran, D. Lancet, D. Cohen, GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017**, bax028 (2017).
107. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
108. Y. Guo, Z. Xue, M. Gong, S. Jin, X. Wu, W. Liu, CRISPR-TE: A web-based tool to generate single guide RNAs targeting transposable elements. *Mob. DNA* **15**, 3 (2024).
109. J.-P. Concordet, M. Haussler, CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).

110. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
111. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
112. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
113. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

Acknowledgments: We thank the University of Colorado Genomics Shared Resource and BioFrontiers Computing core for technical support during this study. We thank B. Nebenfuhr and N. Arnould for assistance with the cell irradiation experiments and M. Jackson, C. Binns, S. Smoots, and A. Dominguez for assistance with the animal studies. **Funding:** This work was supported by National Cancer Center fellowship (A.I.), National Institutes of Health grant 1R35GM128822 (E.B.C.), Alfred P. Sloan Foundation grant (E.B.C.), David and Lucile Packard Foundation grant (E.B.C.), Boettcher Foundation grant (E.B.C.), and American Cancer Society grant DBG-23-1155983-01-DMC (E.B.C.). **Author contributions:** Writing—original draft: A.I. and E.B.C. Conceptualization: A.I., D.M.S., T.M.P., and E.B.C. Investigation: A.I., D.M.S., O.M.J., S.M.B., L.L.N., and T.M.P. Writing—review and editing: A.I., O.M.J., B.G.B., T.M.P., and E.B.C. Methodology:

A.I., D.M.S., S.M.B., T.M.P., and E.B.C. Resources: A.I., D.M.S., O.M.J., S.M.B., L.L.N., B.G.B., T.M.P., and E.B.C. Funding acquisition: A.I., B.G.B., and E.B.C. Data curation: A.I. Validation: A.I., D.M.S., L.L.N., T.M.P., and E.B.C. Supervision: A.I. and E.B.C. Formal analysis: A.I. and S.M.B. Software: A.I. Project administration: A.I., D.M.S., B.G.B., and E.B.C. Visualization: A.I., D.M.S., S.M.B., and T.M.P.

Competing interests: The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. High-throughput sequencing data (RNA-seq and CUT&RUN) has been deposited in the 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 database (downloaded February 2019), Roadmap/ENCODE (downloaded February 2019), TCGA (downloaded September 2019), CEMT Canadian Epigenome Project (downloaded July 2020), Dfam 2.0, and gnomAD v3.1. Source code and analysis scripts are available on Zenodo (<https://doi.org/10.5281/zenodo.10996183>) and GitHub (https://github.com/atmaivancevic/ERV_cancer_enhancers).

Submitted 17 January 2024

Accepted 13 June 2024

Published 17 July 2024

10.1126/sciadv.ado1218