

CANCER

The landscape of cryptic antisense transcription in human cancers reveals an oncogenic noncoding RNA in lung cancer

Zhaozhao Zhao^{1,2†}, Yu Chen^{1†}, Xiaomeng Cheng¹, Leihuan Huang¹, Haimei Wen¹, Qiushi Xu¹, Xiaolan Zhou¹, Xiaoyang Zhang^{3‡}, Jing Chen^{4*}, Ting Ni^{1,5*}

Cryptic transcription initiation has been previously linked to activation of oncogenic transcripts. However, the prevalence and impact of cryptic antisense transcription from the opposite strand of protein-coding genes were mostly unknown in cancer. Applying a robust computational pipeline to publicly available transcriptome and epigenome datasets, we identified hundreds of previously unannotated cryptic antisense polyadenylated transcripts (CAPTs) that were enriched in tumor samples. We showed that the activation of cryptic antisense transcription was associated with increased chromatin accessibility and active histone marks. Accordingly, we found that many of the antisense transcripts were inducible by treatment of epigenetic drugs. Moreover, CRISPR-mediated epigenetic editing assays revealed that transcription of a noncoding RNA *LRRK1*-CAPT promoted LUSC cell proliferation, suggesting its oncogenic role. Our findings largely expand our understanding of cancer-associated transcription events, which may facilitate the development of novel strategies for cancer diagnosis and treatment.

INTRODUCTION

Previous studies estimated that up to 85% of the human genome can be transcribed by RNA polymerase II, resulting in large amounts of RNA products (1–3). The application of high-throughput RNA sequencing (RNA-seq) and bioinformatic methods has revealed the diversity and complexity of RNA transcription and processing in human cells (4–7). Cancer is a complex and heterogeneous disease involving both genetic and epigenetic alterations that lead to a complicated transcriptome (6–11). Recently, a wave of discoveries has demonstrated that the broad role of transcriptome alteration in various oncogenic properties underlying the hallmarks of cancer (8, 9). For instance, de novo alternative transcription initiation in *ALK* intron 19 produces a novel *ALK* transcript, which encodes three oncogenic truncated proteins (8). Cryptic promoters derived from activated transposable elements (TEs) have also been associated with widespread expression of oncogenes (9). However, our understanding of transcripts that are derived from the opposite strand of protein-coding genes remains limited.

Natural antisense transcripts (NATs) are RNA molecules transcribed from the opposite strands of DNA to that partially or

completely overlap with a sense transcript of either protein-coding or noncoding genes (12). NATs may arise from independent transcriptional units containing cryptic promoters situated within genes, typically in intronic regions, or near transcriptional start sites of neighboring genes (12). Previous studies have identified individual NATs that are dysregulated in various types of cancer, and NATs have been regarded as pivotal regulators of the hallmarks of cancer (13, 14). However, considering that the cancer transcriptome was complicated and heterogeneous, whether there exists aberrant activation of novel antisense transcripts that were unannotated by reference transcriptome in cancer samples is an intriguing question while remains unknown.

Here, we investigated the cryptic antisense transcription using publicly available RNA-seq data from The Cancer Genome Atlas (TCGA). We identified hundreds of previously unannotated cryptic antisense polyadenylated transcripts (CAPTs), which are enriched in tumors as compared to normal tissues. Through integrating multiomics datasets, we revealed that the emergence of cryptic antisense transcription is highly associated with activation of chromatin structures. Accordingly, inhibition of DNA methyltransferase and histone deacetylase promoted cryptic antisense transcription. We showed that a CAPT transcribed from the opposite strand of the protein-coding gene *LRRK1* drives proliferation of lung cancer cells. This study reveals a new type of transcripts and their upstream regulation and downstream effect in the cancer genome, extending our understanding regarding the diversity and complexity of cancer transcriptome.

RESULTS

Identification of cryptic antisense transcription across cancer types

To globally characterize cryptic antisense transcription, we canvassed RNA-seq data across 14 cancer types from the TCGA

Copyright © 2023 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 NonCommercial License 4.0 (CC BY-NC).

¹State Key Laboratory of Genetic Engineering, National Clinical Research Center for Aging and Medicine, Huashan Hospital, Collaborative Innovation Center of Genetics and Development, Human Phenome Institute, Shanghai Engineering Research Center of Industrial Microorganisms, School of Life Sciences, Fudan University, Shanghai 200438, China. ²MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200438, China. ³State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200438, China. ⁴National Clinical Research Center for Aging and Medicine, Huashan Hospital, Fudan University, Shanghai 200040, China. ⁵State key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, Institutes of Biomedical Sciences, College of Life Sciences, Inner Mongolia University, Hohhot 010070, China.

*Corresponding author. Email: chenjing1998@fudan.edu.cn (J.C.); tingni@fudan.edu.cn (T.N.)

†These authors contributed equally to this work.

‡Present address: Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112, USA.

Research Network (fig. S1A and table S1). We constructed a computational pipeline that identifies antisense transcripts that are unannotated and highly tumor-enriched (fig. S1B). Briefly, we assembled transcriptomes in each sample using StringTie (15) and retained the newly assembled transcripts that were unannotated and overlapped with known protein-coding genes. Considering that TCGA RNA-seq data did not preserve strand information of sequencing reads, we used splice-site orientation by recognizing splice junction reads to strengthen the inference of the transcript orientation (Fig. 1A and fig. S1, C and D). After taking into account tumor enrichment and expression frequency (see Materials and Methods for details), we identified 57 to 316 cryptic antisense genes for each cancer type and found a total of 644 nonredundant cryptic antisense genes across the 14 cancer types (Fig. 1B, fig. S2A, and table S2). In the upstream [−40 nucleotide (nt)] of termination sites of these antisense transcripts, canonical poly(A) signal AATAAA can be successfully identified by MEME motif enrichment analysis (16) (559 of these antisense genes have AATAAA

motif within 40 nt upstream of its transcription termination site) (fig. S2B), suggesting that these transcripts are polyadenylated and thus they could be detected by polyA⁺ RNA-seq libraries in TCGA. We refer to these transcripts as “CAPTs.” We detected at least one CAPT event in 85% of all tumors, with prevalence ranging from 51 to 95% across cancer types (fig. S2C). In addition, about 64% of these antisense genes could undergo alternative splicing and generate multiple isoforms, as exemplified by antisense transcripts overlapped with the gene *INPP4B* (fig. S2, D and E). Examining the distribution of CAPT expression showed strong heterogeneity: Some CAPTs such as *KCP*-CAPT were present across multiple cancer types, while *NCEH1*-CAPT was only expressed in four lung squamous cell carcinoma (LUSC) tumors (fig. S2, F and G, and table S3). We classified these CAPTs into three groups: 66 ubiquitous CAPTs expressed in ≥7 cancer types, 451 intermediately specific CAPTs that are expressed in 2 to 7 cancer types, and 127 cancer-type-specific CAPTs that are expressed in only 1 cancer

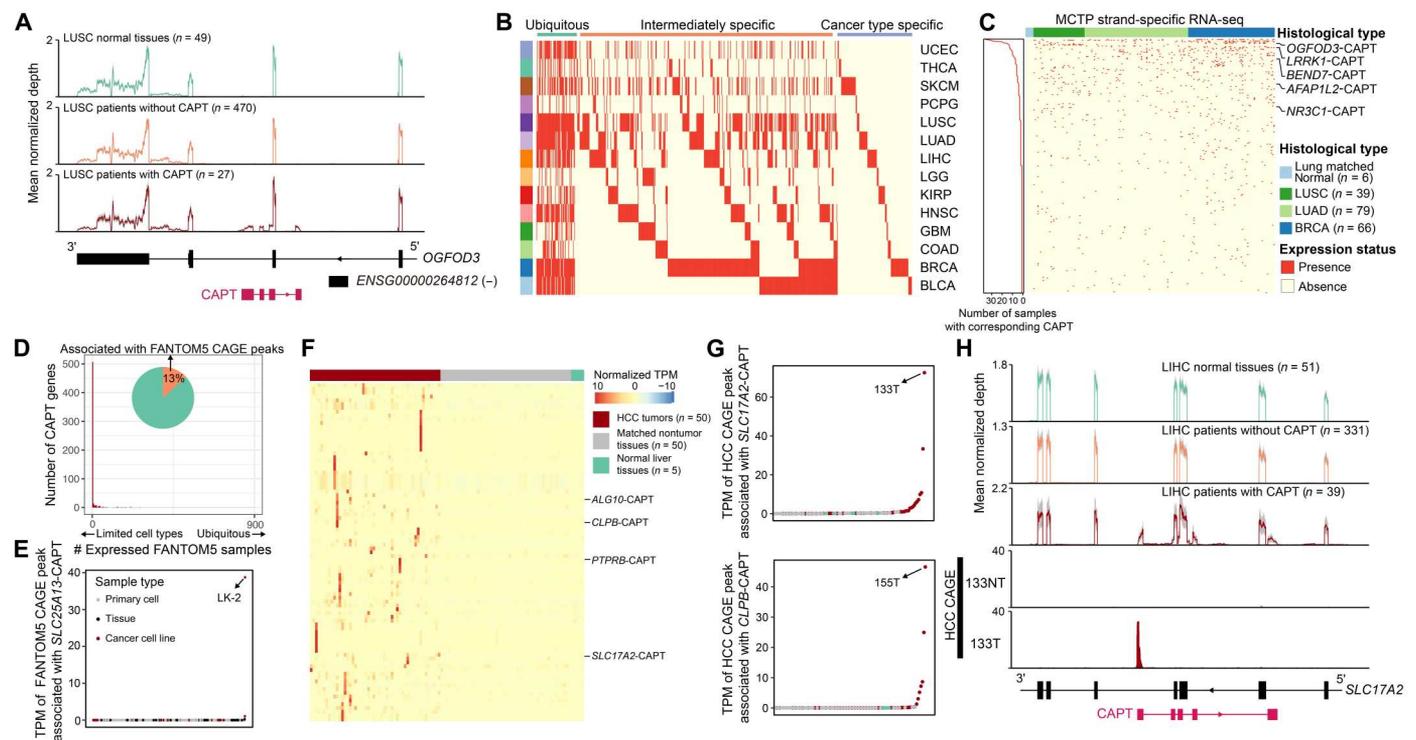


Fig. 1. The CAPT landscape across cancer types. (A) Track plots showing the mean normalized coverage distribution of *OGFD3*-CAPT expression across LUSC samples. Mean normalized coverage depth of each nucleotide position was shown as a solid line, and corresponding 95% confidence interval was shown in gray area. The total patient number of different sample types (normal tissues, tumors without CAPT, and tumors with CAPT) was indicated at corresponding plot. (B) Expression profile of CAPTs across human cancers. UCEC, uterine corpus endometrial carcinoma; THCA, thyroid carcinoma; SKCM, skin cutaneous melanoma; PCPG, pheochromocytoma and paraganglioma; LUSC, lung squamous cell carcinoma; LUAD, lung adenocarcinoma; LIHC, liver hepatocellular carcinoma; LGG, low-grade glioma; KIRP, kidney renal papillary cell carcinoma; HNSC, head and neck squamous cell carcinoma; GBM, glioblastoma multiforme; COAD, colon adenocarcinoma; BRCA, breast invasive carcinoma; BLCA, bladder urothelial carcinoma. (C) Heatmap showing the expression status of all 316 CAPT genes supported by Michigan Center for Translational Pathology (MCTP) strand-specific RNA-seq data in all analyzed MCTP samples. The left histogram showing the number of analyzed MCTP samples with corresponding CAPTs. Several CAPTs are highlighted. (D) Expression pattern of CAPTs across various human tissues and cell types from the FANTOM5 project. (E) Expression level distribution of CAGE peak associated with *SLC25A13*-CAPT in FANTOM5 samples. Sample types are indicated by different colors. (F) Heatmap showing the expression level of CAGE peaks associated with CAPTs in 50 hepatocellular carcinoma (HCC) tumor tissues, 50 matched nontumor tissues, and 5 morphologically normal liver tissues. Several peaks associated with certain CAPTs were indicated. TPM, tags per uniquely mapped million tags. (G) Scatter plots showing CAGE peaks associated with *SLC17A2*-CAPT (top) and *CLPB*-CAPT (bottom) were highly expressed in several HCC tumors. (H) Top three track plots showing the mean normalized coverage distribution of *SLC17A2*-CAPT across matched normal liver samples, LIHC tumors without corresponding CAPT expression, and LIHC tumors with corresponding CAPT expression. Bottom two normalized CAGE profiling track plots showed that the CAGE peak supported the tumor-specific expression of *SLC17A2*-CAPT.

type (Fig. 1B). These results provided a global profile of cryptic antisense transcription across 14 cancer types.

To support the reliability of CAPTs identified in this study, we first used additional strand-specific RNA-seq data of three cancer types [breast invasive carcinoma (BRCA), lung adenocarcinoma (LUAD), and LUSC] from the Michigan Center for Translational Pathology (MCTP) compendium to validate the authenticity of CAPTs (17). A total of 316 CAPTs (49%) identified in this study were supported by this batch of strand-specific RNA-seq data, and the proportion of CAPTs that were expressed in specific TCGA cancer type and supported by MCTP dataset simultaneously was increased to around 70% (74% for CAPTs expressed in TCGA-LUSC and 67% for CAPTs expressed in TCGA-BRCA) (Fig. 1C and fig. S3A). In addition, the expression pattern of CAPTs in all these MCTP cancer samples showed strong heterogeneity: Some CAPTs such as *OGFOD3*-CAPT were present in all these three cancer types, while some CAPTs were only expressed in specific cancer type (Fig. 1C and fig. S3, B to D). We found that CAPTs supported by MCTP RNA-seq data have significantly higher expression frequency in TCGA samples than those not supported by MCTP RNA-seq data (fig. S3E). For example, *NCEH1*-CAPT, only expressed in four TCGA-LUSC samples (fig. S2G), was not expressed in all these 79 MCTP-LUSC samples. We also performed de novo identification of cryptic antisense transcripts from strand-specific RNA-seq data of six pairs of hepatocellular carcinoma (HCC) and matched normal adjacent liver tissues (18) and further compared them with the analysis results from TCGA-liver hepatocellular carcinoma (LIHC). We identified 11 cryptic antisense genes in this batch of HCC tumors, where 10 among them were also identified in our TCGA-LIHC result (fig. S3F). Again, their expression levels in HCC tumors showed strong heterogeneity, as exemplified by *PTPRB*-CAPT, which was specifically highly expressed in tumor NT10T (fig. S3G). In conclusion, all above results not only support the authenticity of CAPTs identified in this study and robustness of our computational pipeline used in this study but also reveal the characteristics of CAPTs, which are specifically highly expressed in tumor samples, and their expression levels show strong heterogeneity.

To further estimate how widely these CAPTs are expressed in the human body, we used the functional annotation of the mammalian genome 5 (FANTOM5) atlas, which provides expression data measured by CAGE (Cap Analysis of Gene Expression) across a broad panel of primary cells, tissues, and cancer cell lines (19). We found that only 13% of CAPT loci were associated with any CAGE peak from FANTOM5 (Fig. 1D), where this ratio was significantly lower than annotated antisense genes (42%; $P < 2.2 \times 10^{-16}$, chi-square test; fig. S4, A and B). Some CAPTs supported by FANTOM5 CAGE peaks were only expressed at a fraction of samples, particularly highly expressed in cancer cell lines, as exemplified by *SLC25A13*-CAPT and *MBD5*-CAPT (Fig. 1E and fig. S4C). We also systematically analyzed additional CAGE profiling data from 50 HCC tumors, 50 matched nontumor tissues, and 5 morphologically normal liver tissues (6). We found that about 16% of CAPT loci were associated with these HCC CAGE peaks (fig. S4D). The fraction of CAPTs associated with HCC CAGE peaks increased to 45% when only checking CAPTs expressed in LIHC tumors (fig. S4D). These CAPT-associated CAGE peaks were specifically highly expressed in HCC tumors (Fig. 1F). Again, their expression levels in HCC tumor samples showed strong heterogeneity, as exemplified

by CAGE peak associated with *PTPRB*-CAPT, only stably expressed in several HCC tumors (fig. S4, E and F). In these HCC CAGE peaks-associated CAPTs, 48 were not associated with FANTOM5 CAGE peaks, as exemplified by *SLC17A2*-CAPT and *CLPB*-CAPT (Fig. 1, G and H, and fig. S4G). *SLC17A2*-CAPT-associated HCC CAGE peak was highly expressed in several HCC tumors, especially in sample 133T, but not expressed in matched nontumor tissues and normal liver tissues (Fig. 1, G and H). These results not only show that the FANTOM5 CAGE peaks remain far from saturation but also support the characteristics (tumor-enriched and strong heterogeneity) and authenticity of CAPTs identified in this study.

The characterization of CAPTs

The majority of CAPTs were transcribed from intronic regions of relevant sense genes, which was similar to previously annotated antisense transcripts (Fig. 2A) (20). However, transcription start site (TSS) of CAPTs had significantly higher fraction located in intergenic region than annotated antisense transcripts (Fig. 2A; $P = 1.34 \times 10^{-11}$, chi-square test). Consistent with previous reports (17, 18), the expression level of CAPTs was significantly lower than those of annotated protein coding genes and antisense genes (Fig. 2B). In addition, the expression level of CAPTs was significantly lower than that of their matched sense protein coding genes (Fig. 2C and fig. S5, A to C). In addition, CAPTs had significantly lower coding potential than annotated protein coding genes and the majority of them (about 90%) were predicted to produce noncoding RNAs (Fig. 2D) (21). Antisense transcripts could be classified as head to head, tail to tail, or embedded (one transcript is fully contained within the other) according to their orientation with respect to sense genes (fig. S5D) (12). We found that the majority of CAPTs were embedded with sense protein coding genes (fig. S5D). The length of CAPTs was comparable to annotated antisense transcripts (fig. S5E). Because it has been shown that TEs play significant roles in regulating gene networks through novel promoter (6, 8, 9), we examined which fraction of the TSSs of CAPTs overlapped with the major types of repetitive elements (long interspersed nuclear element (LINE), long terminal repeat (LTR), and short interspersed nuclear element (SINE)). About 26% of TSSs of the CAPTs overlapped with LTR elements in the sense direction, which was significantly higher than the annotated antisense transcripts (Fig. 2E; $P < 2.2 \times 10^{-16}$, chi-square test). These results imply that while CAPTs share certain common features with the annotated antisense transcripts, they also have distinct characteristics.

Expression of sense-antisense transcript pairs is often linked, either positively or negatively, much more than randomly distributed (12, 17). To measure the expression correlation between CAPTs and their matched sense genes, we examined whether the expression level of sense gene in tumors with corresponding CAPT was different from that in tumors without CAPT or normal samples. Consistent with previous reports about overall positive correlation between sense/antisense genes (12, 17), we found that more sense genes had higher expression in tumors with related CAPT than those that had lower expression in tumors (Fig. 2F and fig. S5, F to H). For example, *NAA30*-CAPT was transcribed from the intron region of the gene *NAA30*, and the expression level of *NAA30* in LUAD tumors with *NAA30*-CAPT was significantly higher than those in LUAD tumors without this CAPT (Fig. 2, G and H). These results suggest that there may exist common

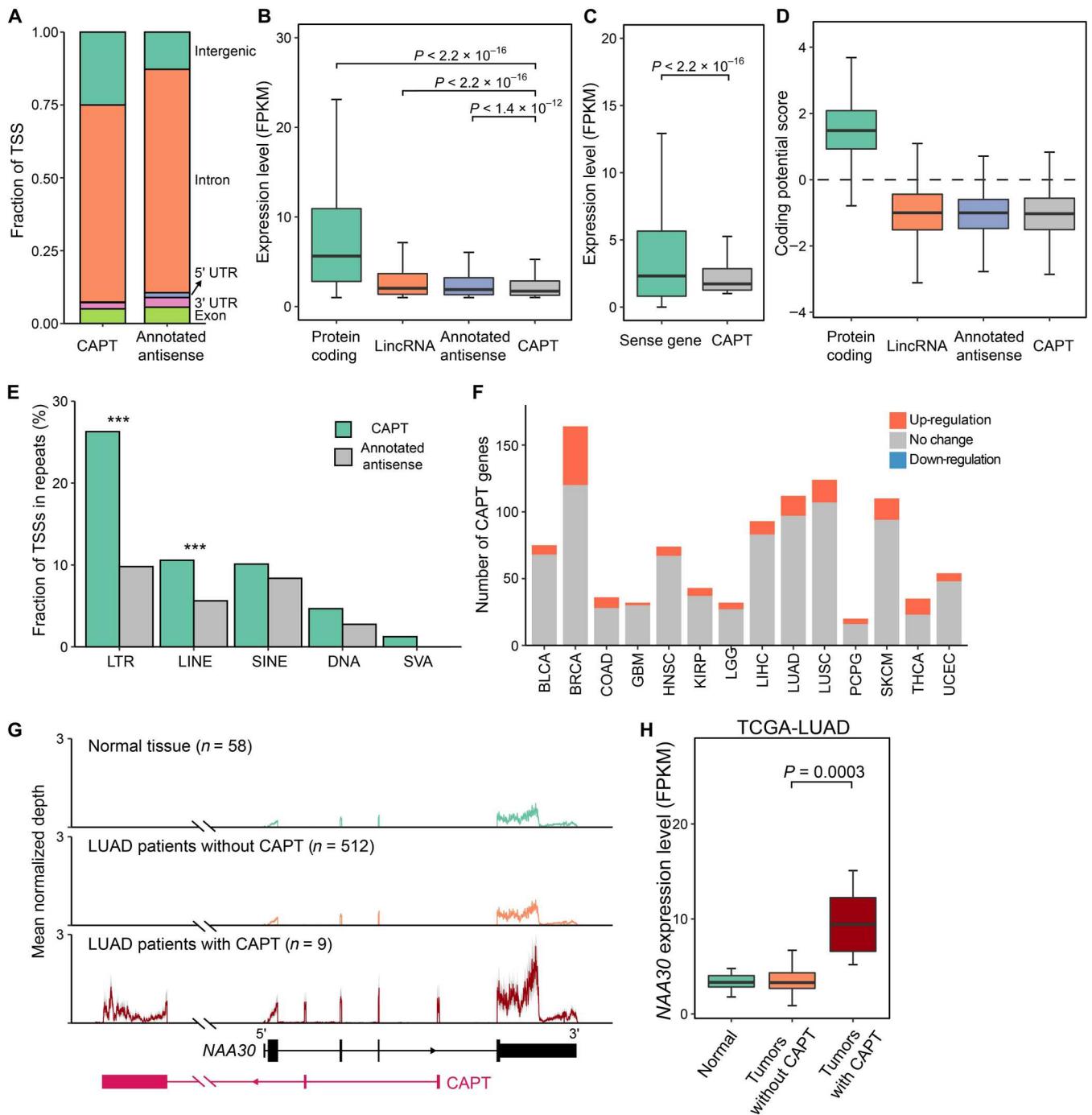


Fig. 2. Characterization of CAPT. (A) Genomic distribution of transcription start sites (TSSs) of CAPTs and annotated antisense transcripts. UTR, untranslated region. (B) Boxplot showing the comparison of expression level of annotated protein coding genes, long intergenic noncoding RNA (lincRNA), annotated antisense transcripts and CAPTs among TCGA samples. (C) Boxplot showing the comparison of expression level of CAPTs and their matched sense protein coding genes in TCGA tumors with the presence of CAPT. (D) The coding potential of annotated protein-coding genes, lincRNA, annotated antisense transcripts and CAPTs was assessed using PLEK (predictor of long non-coding RNAs and messenger RNAs based on an improved k-mer scheme) (21). Dashed line denotes the threshold for protein-coding transcripts. (E) Fraction of CAPT TSSs overlapping with repetitive elements in the sense direction. $***P \leq 0.001$ by chi-square test. (F) Numbers of sense-antisense genes pairs whose expression has correlation in different cancer types. (G) RNA-seq track plots showing the mean normalized coverage distribution of *NAA30*-CAPT across matched normal lung samples, LUAD tumors without *NAA30*-CAPT expression, and LUAD tumors with *NAA30*-CAPT. Mean normalized coverage depth of each nucleotide position was shown as a solid line, and corresponding 95% confidence interval was shown in gray area. (H) Boxplot showing that the gene *NAA30* in LUAD tumors with *NAA30*-CAPT has higher expression than those without *NAA30*-CAPT.

coordinated regulatory mechanisms controlling the sense-antisense transcription.

Changes in the chromatin landscape underlie the expression of CAPTs

To cast light on the mechanism of CAPT activation, we integrated epigenomic data including chromatin accessibility, DNA methylation, and covalent histone modifications to investigate the potential cancer-associated epigenetic reprogramming accompanying CAPT expression. The assay for transposase-accessible chromatin using sequencing (ATAC-seq) data from TCGA profiled genome-wide chromatin accessibility to identify potential regulatory elements in diverse types of human primary tumors (22). We analyzed matched RNA-seq and ATAC-seq data of 275 tumor samples from the above investigated 14 cancer types. We found that about 80% of TSSs from those CAPTs overlapped with ATAC-seq peaks (Fig. 3A). Then, we examined the correlation between ATAC-seq peak signal and their overlapped CAPT expression across all samples. Globally, we found that TSS regions of expressed CAPTs had significant increase in accessibility compared to those of other nonexpressed CAPTs (Fig. 3, B and C). This trend still remained in specific cancer type, such as kidney renal papillary cell carcinoma (KIRP) tumor (fig. S6, A and B). TSS of *ALG10*-CAPT, which have been shown specifically expressed in the liver tumors (figs. S2F and S4F), overlapped with an ATAC-seq peak named LHC_80799. The tumor sample expressed *ALG10*-CAPT (TCGA-BC-A10U) had higher accessibility in this peak region (Fig. 3, D and E). Peak regions overlapped with several CAPTs such as *MYT1L*-CAPT and *ATP8A2*-CAPT had overall higher accessibility in KIRP tumors compared with other cancer types (fig. S6, C to F). However, the KIRP tumors with the expression of related CAPTs still had higher accessibility in the related peak regions (highlighted in red, fig. S6, C to F). We speculated that some CAPTs were not supported by ATAC-seq peaks due to the fact that ATAC-seq peaks derived from these tumors remain far from saturation, as exemplified by *HTR1E*-CAPT, not associated with TCGA ATAC-seq peaks, but expressed in NCI-H838 cells and supported by ATAC peak and CAGE peak (fig. S6G). Given that promoters and TSS-proximal regions were always highly accessible for transcription factors (TFs) to initiate transcription (23, 24), these results support the potential regulation of CAPTs across cancer types.

DNA methylation alteration has been observed in various cancers, and global hypomethylation is frequently observed in highly and moderately repeated DNA sequences (25, 26). Combined with DNA methylation array data, we found that the TSS regions of some expressed CAPTs were hypomethylated in tumors with the presence of relevant CAPT (fig. S7, A to C). For example, array sites cg02775804 and cg15864074, located around TSS of *EPB41L5*-CAPT, were significantly hypomethylated in tumors with the presence of *EPB41L5*-CAPT (fig. S7B). We also examined the correlation between CAPT expression level and DNA methylation level of associated CpG array site in all tumor samples with matched RNA-seq and Illumina 450K methylation array data and found that 60.6% of CAPT-to-CpG site links showed significant negative correlation across cancer types, as exemplified by *WIPF1*-CAPT and CpG array site cg24932433 (fig. S7, D and E, and table S4).

Epigenetic marks on chromatin cooperate with TFs to regulate transcription (27). Epigenetic deregulation has emerged as a

paradigm of cancer biology that underlies the hallmarks of cancer cells (28). One early study collected 73 pairs of colorectal cancer tissues (tumor and adjacent native tissue from the same patients) and generated a series of high-quality sequencing data, including 147 H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq), 86 H3K4me3 ChIP-seq, and 144 RNA-seq (10). We used this batch of data to investigate the dynamics of histone modifications accompanying the activation of CAPT. We first checked the expression status of CAPTs in this batch of data and found that 30 CAPTs were expressed here and the majority of them were uniquely expressed in tumor samples (fig. S8A). We found that histone modifications H3K4me3 and H3K27ac, which were associated with transcriptional activation (29, 30), showed enrichment at TSS of CAPTs (Fig. 3, F and G, and fig. S6, B to D). For example, the expression of *EPB41L5*-CAPT was accompanied by co-occurrence of H3K27ac and H3K4me3 signals on its promoter in tumor tissues, while the matched adjacent native tissues had undetectable *EPB41L5*-CAPT expression and no enrichment of the two histone marks (Fig. 3F). The average H3K27ac signal and H3K4me3 signal of TSSs in tumors with expressed CAPT were also higher than those in tumors without CAPT or normal tissues (Fig. 3G).

To investigate the potential TFs regulating CAPTs, the DNA sequences of regions around TSSs of CAPTs were used for prediction by HOMER software (31). We identified several TFs such as KLF5 (Fig. 3H), which was a driver of diverse cancer-related phenotype (32). Together, these results reveal that the activation of CAPT is tightly regulated by chromatin modulation and related TFs.

Treatment of epigenetic drugs induced the expression of CAPTs

The correlation between epigenomic dynamics and expression of CAPTs leads us to hypothesize that genome-wide epigenetic changes following DNA methyltransferase and histone deacetylase inhibitors (DNMTi and HDACi, respectively) treatment could result in the activation of CAPTs (33). To this end, we systematically analyzed the CAGE data of NCI-H1299 cells treated with DNMTi (DAC), HDACi (SB939 or SAHA), or both DAC and SB939 (DAC + SB939). We found that epigenetic treatment activated about 40 CAPTs, with the combinatorial treatment (DAC + SB939) showing the strongest effects (Fig. 4, A and B, and fig. S9A). In line with previous finding of the synergistic effect on gene expression by combined demethylation and histone deacetylase inhibition (33), we found that multiple CAPTs were exclusively expressed after DAC + SB939 treatment (Fig. 4B). Majority of these activated CAPTs were not associated with FANTOM5 CAGE peaks, as exemplified by *CEP63*-CAPT (fig. S9, B to E). ChIP-seq revealed that TSS regions of activated CAPTs upon treatment with DAC or DAC + SB939 showed increased signal of H3K4 trimethylation (Fig. 4C). Furthermore, we experimentally validated that DNMTi and HDACi induced the up-regulation of several CAPTs in different cell lines by quantitative reverse transcription polymerase chain reaction (qRT-PCR) (Fig. 4, D and E, and fig. S9F). Together, these findings suggest that these CAPTs are authentic transcripts and regulated by chromatin modulation.

LRK1-CAPT as a regulator of LUSC cell proliferation and tumorigenesis

To investigate the potential role of CAPTs, we first evaluated global CAPT expression pattern in 186 LUSC cell lines from the Cancer

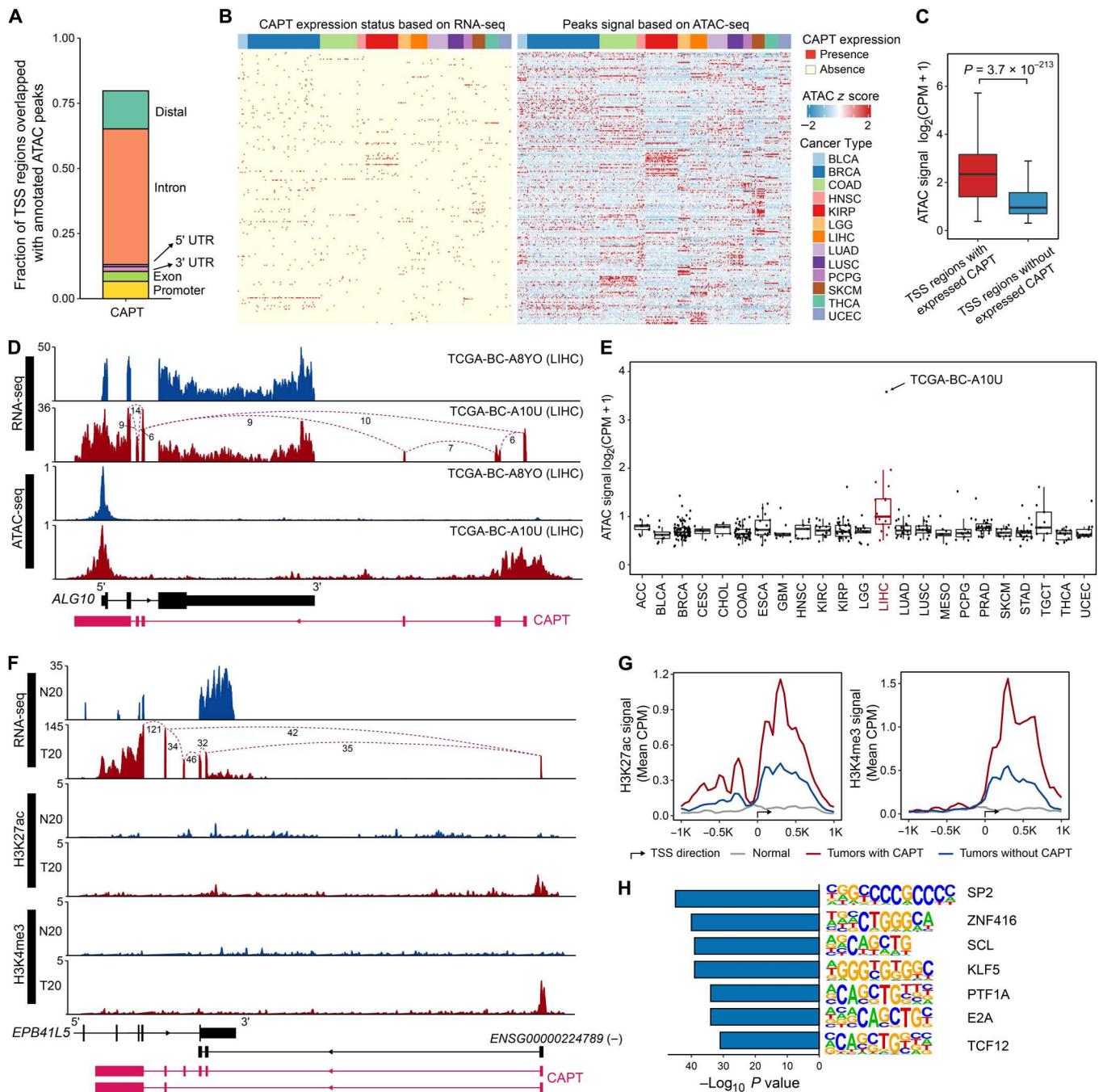


Fig. 3. The activation of CAPT was associated with epigenetics alteration. (A) Fraction of CAPT TSSs overlapped with pan-cancer peak set called from TCGA ATAC-seq data. Color indicates the type of genomic region overlapped by the ATAC-seq peak. (B) Heatmap representation of overlapped CAPT-to-peak links. Each column represents a tumor sample. Each row represents an individual link between one CAPT and its TSS overlapped ATAC-seq peak. Color represents the CAPT expression status (left) or relative ATAC-seq accessibility (right) for each link. (C) Boxplot showing the comparison of chromatin accessibility between ATAC peak-overlapped TSS regions with expressed CAPT and those without expressed CAPT in all tumor samples. The two-side Wilcoxon rank sum test was used to assess the statistical significance. CPM, counts per million mapped reads. (D) RNA-seq density plots and normalized ATAC-seq tracks showing that the promoter of *ALG10*-CAPT had increased accessibility in tumor sample TCGA-BC-A10U. Only the annotation of *ALG10*-CAPT transcript that had the most exons was shown. (E) Boxplots showing the normalized counts of ATAC-seq peak named LHC_80799 in different cancer types. Each dot represents an individual donor. Tumor with *ALG10*-CAPT shown in (D) was indicated. (F) RNA-seq and ChIP-seq tracks showing that the significant enrichment of histone H3K27ac and H3K4me3 near the TSS of *EPB41L5*-CAPT in colorectal cancer sample. T20 and N20 mean tumor and adjacent tissue of patient 20. ChIP-seq signal have been normalized. (G) The average normalized H3K27ac (left) and H3K4me3 (right) signal around TSS of *EPB41L5*-CAPT in tumors with *EPB41L5*-CAPT, tumors without *EPB41L5*-CAPT, and adjacent normal tissues. (H) Motif analysis of TSSs of all CAPTs for putative transcription factor (TF) binding sites. Regions (400 nt) centered on TSSs of all CAPTs were used for motif discovery.

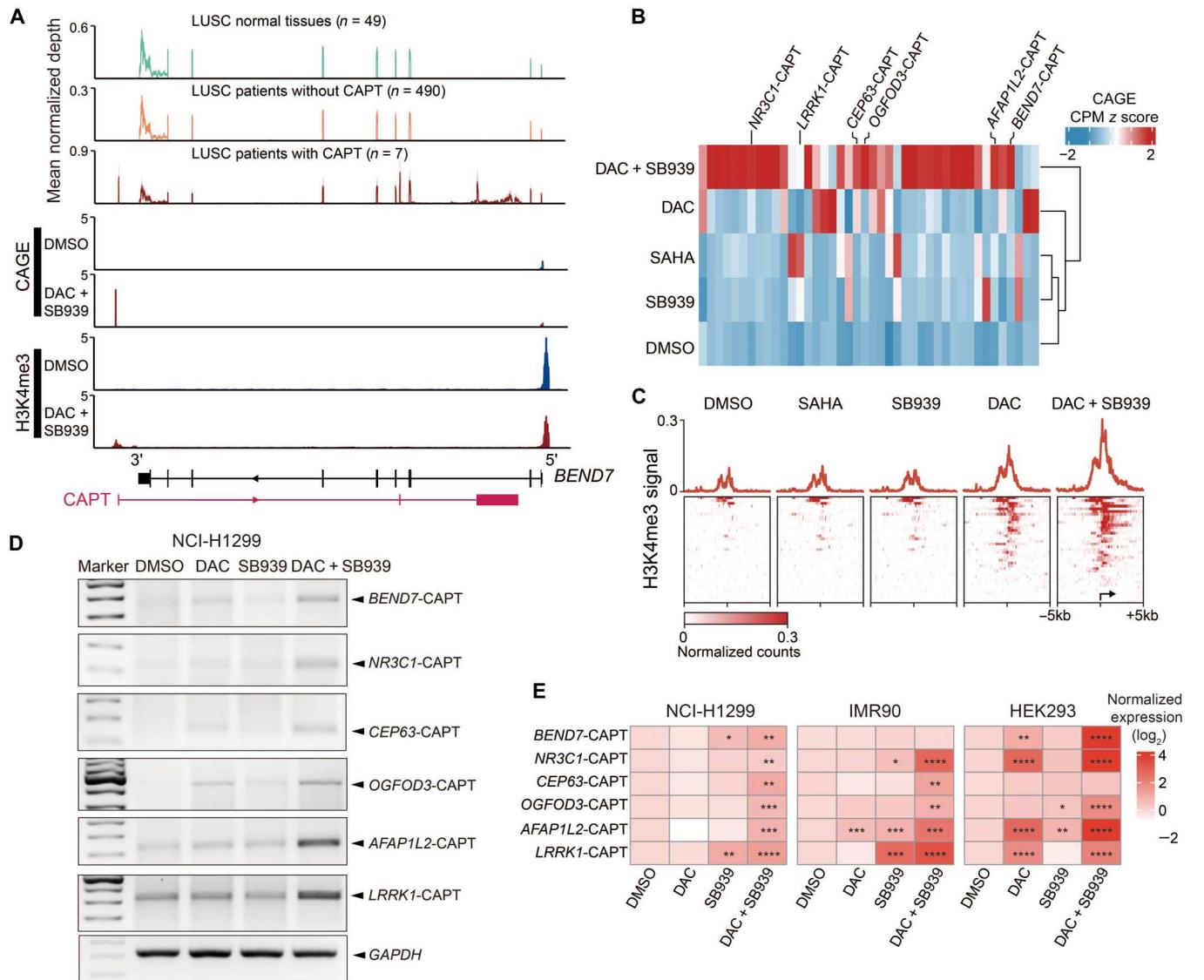


Fig. 4. DNMTi and HDACi induced the activation of a portion of CAPTs. (A) RNA-seq, CAGE, and ChIP-seq showing that combinatorial treatment (DAC + SB939) induced the activation of *BEND7*-CAPT. Top three mean normalized coverage plots showing that *BEND7*-CAPT was expressed in some LUSC samples. CAGE and H3K4me3 ChIP-seq signal have been normalized. (B) Heatmap showing that epigenetic treatment increased the expression of some CAPTs. Color represents the relative expression of CAPT (CPM) as a z score. (C) ChIP-seq occupancy plots showing the average H3K4me3 level centered around TSSs of all activated CAPT upon treatment. (D) Reverse transcription polymerase chain reaction (RT-PCR) analysis of several CAPTs upon drug treatment in NCI-H1299 lung cancer cells. (E) Heatmap showing the CAPT expression level upon drug treatment in NCI-H1299 (left), IMR90 (middle), and HEK2993 (right) cells. The \log_2 of the mean expression from three independent experiments relative to expression level of dimethyl sulfoxide (DMSO) condition is shown. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, and **** $P \leq 0.0001$ by a two-tailed Student's *t* test.

Cell Lines Encyclopedia (CCLE) (34) and found that 146 CAPT genes were expressed in these LUSC cell lines (fig. S10A). The expression of CAPTs also showed strong heterogeneity in cancer cell lines. *EPB41L5*-CAPT was frequently expressed in 18 cell lines, while *OGFOD3*-CAPT was uniquely expressed in NCI-H196 cells (fig. S10A). To verify the existence of CAPTs in relevant cell lines and determine their full-length sequence, we performed rapid amplification of complementary DNA (cDNA) ends (RACE) followed by Sanger sequencing for three CAPTs (*EPB41L5*-CAPT and *SLC25A13*-CAPT in NCI-H520 cells, and *LRRK1*-CAPT in NCI-H1299 cells) (Fig. 5A and fig. S10, B to D). These three CAPT

genes all had multiple isoforms generated by alternative splicing. To evaluate the potential roles of these three CAPTs, we performed functional knockdown experiments using small interfering RNA (siRNA) targeting common region of isoforms of each of these three CAPTs in the corresponding cell lines. We found that silencing of *LRRK1*-CAPT could significantly inhibit NCI-H1299 cell proliferation, as detected by Cell Counting Kit-8 (CCK-8) assay and related proliferation markers (Fig. 5, B and C). Silencing of *EPB41L5*-CAPT could significantly inhibit NCI-H520 cell proliferation, while silencing of *SLC25A13*-CAPT had no significant effect on NCI-H520 cell proliferation (fig. S10, E to I). Considering that

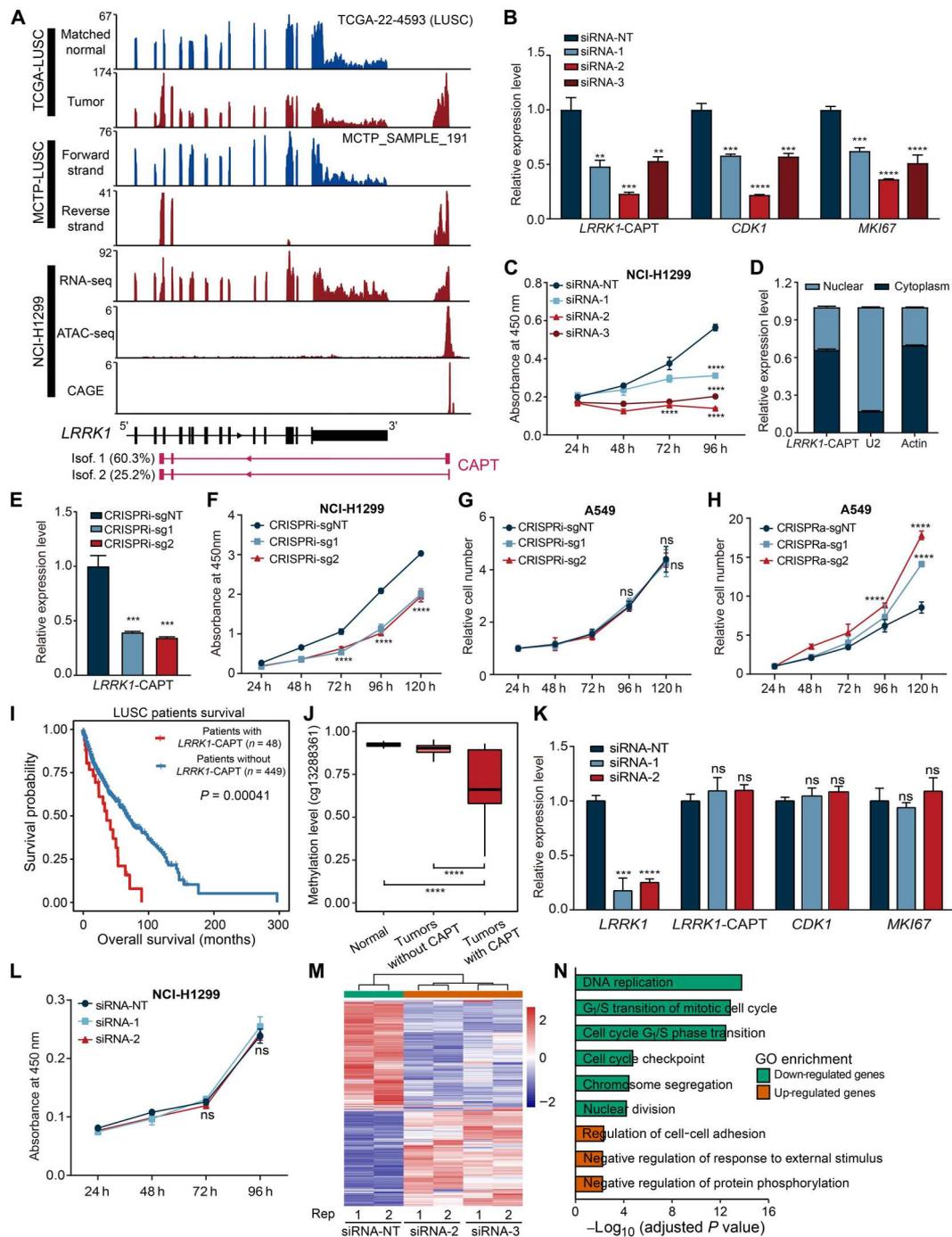


Fig. 5. *LRRK1-CAPT* was a regulator of LUSC cell proliferation and tumorigenesis. (A) RNA-seq density plots and normalized ATAC-seq and CAGE tracks showing that *LRRK1-CAPT* was expressed in LUSC samples and NCI-H1299 cells. (B) qPCR analysis for expression level of *LRRK1-CAPT* and proliferation-related marker genes *CDK1* and *MKI67* upon *LRRK1-CAPT* knockdown in NCI-H1299 cells. (C) Cell proliferation rate evaluated by Cell Counting Kit-8 (CCK-8) assay upon *LRRK1-CAPT* knockdown in NCI-H1299 cells. (D) qRT-PCR for *LRRK1-CAPT*, U2, and actin following nuclear and cytoplasmic fractionation of NCI-H1299 cell lysates. (E) qPCR analysis for expression level of *LRRK1-CAPT* in NCI-H1299 cells for nontargeting sgRNA or two independent sgRNA with dCas9-KRAB. (F) Cell proliferation rate of NCI-H1299 cells evaluated by CCK-8 assay upon *LRRK1-CAPT* knockdown. (G) Cell proliferation rate of A549 cells evaluated by CCK-8 assay upon *LRRK1-CAPT* knockdown. (H) Cell proliferation rate evaluated by CCK-8 assay upon *LRRK1-CAPT* overexpression in A549 cells. (I) Kaplan-Meier survival analysis of patients with LUSC with or without *LRRK1-CAPT* expression. P value was calculated by log-rank test. (J) Comparison of promoter methylation level of *LRRK1-CAPT* in normal tissues and in patients with LUSC with or without *LRRK1-CAPT* expression. (K) qPCR analysis for expression level of *LRRK1*, *LRRK1-CAPT*, *CDK1*, and *MKI67* upon *LRRK1* knockdown in NCI-H1299 cells. (L) Cell proliferation rate evaluated by CCK-8 assay upon *LRRK1* knockdown in NCI-H1299 cells. (M) Heatmap of expression level of the top 1000 most variable genes upon *LRRK1-CAPT* knockdown in NCI-H1299 cells. (N) Gene Ontology (GO) enrichment analysis for significantly down-regulated genes and significantly up-regulated genes upon *LRRK1-CAPT* knockdown in NCI-H1299 cells. For all panels, $**P \leq 0.01$, $***P \leq 0.001$, and $****P \leq 0.0001$ by a two-tailed Student's t test. ns, not significant.

LRRK1-CAPT had more profound effect than *EPB41L5*-CAPT on slowing the proliferation rate when knocking down (Fig. 5C and fig. S10F), we focused on *LRRK1*-CAPT to further explore its function in lung cancer.

Cellular localization of *LRRK1*-CAPT via qRT-PCR following cellular fractionation showed the predominant expression in cytoplasm (Fig. 5D). Multiple coding potential assessing tools all indicated that *LRRK1*-CAPT was a noncoding RNA (fig. S11A) (21, 35, 36). Transwell assays showed that *LRRK1*-CAPT knockdown significantly reduced the migration of NCI-H1299 cells (fig. S11B). We also suppressed *LRRK1*-CAPT via inhibiting its transcription by CRISPR interference (37), and both two distinct single-guide RNAs (sgRNAs) targeting *LRRK1*-CAPT significantly reduced proliferation capacity and colony formation of NCI-H1299 cells (Fig. 5, E and F, and fig. S11C). Moreover, we transfected these guide RNAs (gRNAs) into A549 cells that did not express this CAPT (fig. S10A), and the results showed that knockdown *LRRK1*-CAPT had no effect on proliferation capacity and colony formation of A549 cells, which validated the specificity of *LRRK1*-CAPT knockdown (Fig. 5G and fig. S11D). Overexpression of *LRRK1*-CAPT by CRISPR activation (37) exhibited a significant increase in proliferative capacity and colony formation of both NCI-H1299 cells and A549 cells (Fig. 5H and fig. S11, E to I). In addition, patients with LUSC that showed *LRRK1*-CAPT expression in tumor samples had significantly worse survival rates (Fig. 5I). DNA methylation array site cg13288361, located around TSS of *LRRK1*-CAPT, was significantly hypomethylated in tumors with *LRRK1*-CAPT (Fig. 5J). In addition, we found that epigenetic treatment could increase the expression of *LRRK1*-CAPT (Fig. 4, D and E).

We then examined whether *LRRK1*-CAPT played a role by regulating the expression of its sense protein coding gene *LRRK1*. *LRRK1*-CAPT knockdown had no effect on the expression of sense gene *LRRK1* (fig. S12A), consistent with that patients with LUSC with *LRRK1*-CAPT expression had comparable expression level of sense gene *LRRK1* to those without *LRRK1*-CAPT expression, suggesting that these two overlapping genes were independently regulated (fig. S12B). Moreover, the expression level of *LRRK1* in LUSC tumors had no significant change compared with matched normal tissues, and the survival of patients with LUSC was not associated with *LRRK1* expression level (fig. S12, C and D). We experimentally knocked down sense gene *LRRK1* by siRNA in NCI-H1299 cells and found that silencing of *LRRK1* had no significant effect on proliferation and migration of NCI-H1299 cells (Fig. 5, K and L, and fig. S12E), which was consistent with that *LRRK1* was not essential to proliferation of lung cancer cell lines revealed by CRISPR dependency data analysis (fig. S12F) (38). These lines of evidence were combined to show that *LRRK1*-CAPT played a role in regulating LUSC cell proliferation, not by affecting its sense gene *LRRK1*.

We also conducted RNA-seq in the NCI-H1299 cells transduced with either nontargeting siRNA or two distinct *LRRK1*-CAPT-specific siRNAs. We observed widespread gene expression change in NCI-H1299 cells upon *LRRK1*-CAPT knockdown and found a total of 124 and 88 genes that were significantly up- and down-regulated, respectively (Fig. 5M and fig. S12G). Gene Ontology (GO) enrichment analysis revealed that down-regulated genes were enriched in pathways including DNA replication and cell cycle (Fig. 5N), which was in accordance with our observation in the phenotype change described above. Moreover, gene set enrichment

analysis (GSEA) also identified several key cellular growth biological processes, such as E2F and Myc signaling pathway-dependent gene signatures, which were markedly down-regulated upon *LRRK1*-CAPT knockdown (fig. S12H). Together, these results indicated that *LRRK1*-CAPT acted as an oncogenic long noncoding RNA (lncRNA) in promoting the proliferation of LUSC cells.

DISCUSSION

In this study, we systematically characterized the global landscape of cryptic antisense transcription across diverse cancer types. De novo transcript assembly approaches provide an unbiased modality for transcript discovery and have been successful in pinpointing new cancer-associated lncRNAs (39, 40). However, given that the cancer transcriptomes have strong heterogeneity and the expression of assembled transcripts is often inconsistent across different cancer samples, it is still challenging to handle large amounts and different types of sequencing data. By integrating multiomics data including RNA-seq, CAGE profiling, ChIP-seq, and ATAC-seq, we discovered hundreds of cryptic antisense transcripts in human cancers and found that their expression was associated with epigenetic activation. Additional strand-specific RNA-seq data and CAGE profiling data supported the authenticity and characteristics of CAPTs identified in this study. The number of CAPT events per individual tumor was quite low, and the expression of CAPT showed strong heterogeneity across tumor samples. While 95% of BRCA tumors had at least one CAPT events, more than half of these tumors had only two to four CAPT events (fig. S2C). In addition, majority of CAPTs were specifically expressed in a small number of tumors, as exemplified by *NCEH1*-CAPT in LUSC tumors (fig. S2, F and G). We verified the existence of some CAPTs in relevant cell lines and further revealed that *LRRK1*-CAPT acted as an oncogenic lncRNA in promoting the proliferation of LUSC cells. Considering that *LRRK1*-CAPT was only expressed in certain cancer samples and not expressed in most normal tissues, *LRRK1*-CAPT could be a cancer-specific target for the potential personalized treatment. Here, we only used *LRRK1*-CAPT as an example to demonstrate the function of this type of new transcript, and we believe that, with the revealing of hundreds of such cryptic antisense transcripts, this study would provide more potential durable targets for the field and could be of great value in the future.

To dissect the mechanism by which *LRRK1*-CAPT promoted lung cancer cell proliferation, we summarized its following features: (i) The expression of *LRRK1*-CAPT and its sense gene *LRRK1* was independently regulated, and sense gene *LRRK1* was not essential to proliferation of lung cancer cell lines (Fig. 5, K and J, and fig. S12, A to F); (ii) genes associated with E2F1, DNA replication, and cell cycle pathways were enriched in down-regulated genes upon *LRRK1*-CAPT knockdown in NCI-H1299 cells (Fig. 5, M and N, and fig. S12, G and H); and (iii) cellular localization of *LRRK1*-CAPT showed the predominant expression in cytoplasm (Fig. 5D). Previous studies have reported that cytoplasmic lncRNAs could act as competing endogenous RNA or "RNA sponges," interacting with microRNAs in a manner that can sequester these molecules and reduce their regulatory effect on target mRNA (41). Cytoplasmic lncRNAs could also interact with proteins to modulate protein function, regulate protein-protein interactions, or direct localization within cellular compartments (42, 43). Considering that further elucidating the direct interactions between

LRRK1-CAPT and proteins (or RNA molecules) have significant challenges, we raised two potential models to explain the functional mechanism of *LRRK1*-CAPT based on all aforementioned experiment results: (i) *LRRK1*-CAPT may directly interact with and stabilize certain proteins, which could play important roles in regulating cell proliferation, such as TF E2F1. Knocking down *LRRK1*-CAPT would reduce the protein level of E2F1 and thus down-regulate the expression level of E2F1 pathway associated genes accordingly. (ii) *LRRK1*-CAPT may bind to microRNA in a manner that can sequester these molecules and reduce their regulatory effect on target mRNA of certain proteins such as E2F1. The interaction between released microRNA upon *LRRK1*-CAPT knockdown and mRNA of E2F1 would reduce the mRNA stability level (or translation efficiency) level of E2F1. Future in-depth mechanism investigation is required to fully address this question.

DNMTi and HDACi have been used for the treatment of several hematopoietic cancers and solid tumors (44, 45). Although these compounds have been in clinical use for several years, there is still a lack of knowledge regarding mechanisms of action of these drugs (46, 47). Recent study reported that treatment with hypomethylating agents led to demethylation and up-regulation of *SALL4*, a known oncogene that plays an important role in multiple types of cancer, which probably influenced the clinical progression of the disease (48). Here, we showed that DNMTi and HDACi could induce de novo transcription of cryptic antisense transcripts. We found that epigenetic drugs treatment could up-regulate oncogenic lncRNA *LRRK1*-CAPT, which provided a novel mechanism for the action of different classes of epigenetic inhibitors and may partially explain the side effect of these clinically used inhibitors.

Together, this study expands our understanding of complex antisense transcription in human cancers and provides a resource that will enable researchers to elucidate the mechanisms of sense/antisense regulation in cancer and provide potential new targets for cancer treatment.

MATERIALS AND METHODS

Data resource

We downloaded RNA-seq BAM files of 6492 tumor samples across 14 cancer types and their related 493 normal tissue samples from the Genomic Data Commons (GDC) Data Portal (<https://portal.gdc.cancer.gov/>) (table S1). Included were bladder urothelial carcinoma, BRCA, colon adenocarcinoma, glioblastoma multiforme, head and neck squamous cell carcinoma, KIRP, low-grade glioma (LGG), LIHC, LUAD, LUSC, pheochromocytoma and paraganglioma, skin cutaneous melanoma (SKCM), thyroid carcinoma, and uterine corpus endometrial carcinoma. For two cancer types (LGG and SKCM) that have no matched normal tissues in TCGA project, we included 200 brain-related tissue samples (involving cortex, hippocampus, hypothalamus, and substantia nigra) and 200 skin-related tissue samples from Genotype-Tissue Expression Project (GTEx) project [database of Genotypes and Phenotypes (dbGaP), phs000424.v8.p2] as normal tissue samples (table S1) (49). In addition, normalized gene expression data (HTSeq-FPKM-UQ) and methylation 450K array data were downloaded using the `gdc-client` command line utility. We downloaded paired-end RNA-seq BAM files of 186 LUSC cancer cell lines from CCLE from the NCI's GDC (<https://portal.gdc.cancer.gov/legacy-archive>) (34). We downloaded the TCGA ATAC-seq data

for tumors with matched RNA-seq data (<https://gdc.cancer.gov/about-data/publications/ATACseq-AWG>) (22). We downloaded some sequencing data from the National Center for Biotechnology Information (NCBI) dbGaP (www.ncbi.nlm.nih.gov/gap/): accession number phs000937.v1.p1 for strand-specific RNA-seq data of different types of cancer samples from the MCTP compendium and accession number phs000885.v1.p1 for human HCC CAGE profiling data (6, 17). Other sequencing data are available from the NCBI Gene Expression Omnibus database: GSE113946 (CAGE profiling and ATAC-seq from NCI-H838 cells) (9), GSE174338 (strand-specific RNA-seq data from six pairs of HCC and normal adjacent liver tissues) (18), GSE81322 (CAGE profiling and ChIP-seq from NCI-H1299 upon treatment of epigenetic drugs) (33), and GSE156614 (RNA-seq and ChIP-seq data from colorectal cancer tissues with paired adjacent tissues) (10).

Identification and quantification of cryptic antisense transcript

StringTie v.1.3.5 was used to assemble the BAM files for all the RNA-seq samples (15). Default parameters were used, and the GENCODE reference annotation (v38) was supplied to guide each assembly (20). All individual transcriptomes were then merged together using StringTie merge (-F 1 -m 200). From the merged transcript model, we extracted the transcripts that contained at least one exon overlapped with protein-coding genes annotated by GENCODE (v38) in the opposite strand and supported by at least 10 junction reads at least one sample, where canonical splice site motif (GT for donor site and AG for acceptor site) was used to strengthen the inference of the transcript orientation.

To perform the comparison between the newly assembled antisense transcripts with the reference annotation, we first combined the set of genes in GENCODE (v38) and "curated" RefSeq (v200) records (that is, including only "NM_" and "NR_" identifiers; see www.ncbi.nlm.nih.gov/refseq/about) into a joint transcriptome that we named annotated RNAs. We then used the CuffCompare (50) tool to compute the overlap of the set of newly assembled antisense transcripts with the annotated RNAs. Each antisense gene was considered to be annotated by reference transcriptome if at least one of its transcripts was assigned one of the class codes "=" or "c." The unannotated antisense genes were used for downstream analysis and were merged with the reference GENCODE v38 annotation file to create a merged transcriptome model.

To quantify the expression level of antisense transcripts, we used StringTie (-e -b) with the merged transcriptome model as the reference. For each sample, we labeled a candidate antisense transcript as being present if it met the following criteria: (i) the antisense transcript contained at least one exon overlapped with sense gene and supported by at least 10 junction reads; and (ii) the antisense gene had at least one FPKM expression. Next, we further filtered for tumor-enriched antisense genes: (i) showed minimal to no expression in normal tissues (maximum expression in all normal tissue samples was less than 1 FPKM); (ii) showed highly tumor-enriched (median expression level in all tumor samples with the presence of corresponding antisense gene was at least 10-fold higher than median expression level in all normal samples); and (iii) presented at least three tumor samples. Last, we obtained a master list of 644 tumor-enriched unannotated antisense genes.

Strand-specific RNA-seq data analysis

Considering the overlap between cancer types from TCGA analyzed in this study and cancer types from MCTP compendium dataset, we have chosen 190 RNA-seq data samples from three cancer types (66 for BRCA, 79 for LUAD, 39 for LUSC, and 6 lung matched normal samples) from MCTP dataset for downstream analysis. After filtering out low-quality reads, the remaining reads were aligned to hg38 reference genome sequence using TopHat2 (the option fr-first-strand was used while all other parameters were set to default values) (51). We used StringTie to assemble transcripts in a reference-guided manner (-G) and a strand-specific manner (--rf, which assumes a stranded library fr-firststrand). The reference and assembled transcript models were merged by StringTie merge to obtain the merged transcript model. We also estimated the expression levels (FPKM) for the merged transcript model by running StringTie in its expression abundance estimation mode (StringTie -e -b --rf). Then, we used CuffCompare to compare CAPT transcript model annotation identified from TCGA with the newly merged transcript model assembled in MCTP datasets. Each CAPT gene was considered to be supported by MCTP datasets if at least one of its transcripts was assigned class code "=" or class code "c," and the expression level of corresponding CAPT gene was more than 1 FPKM in at least one MCTP cancer sample. Last, 316 CAPTs identified in this study were supported by strand-specific RNA-seq data from MCTP compendium.

For strand-specific RNA-seq data from six pairs of HCC and normal adjacent liver tissues, we performed de novo identification of cryptic antisense transcripts and further compared them with the analysis results from TCGA-LIHC. Using the same pipeline of TCGA data analysis, we first used StringTie to de novo assemble transcripts and then merged them. In addition, we selected transcripts that contained at least one exon overlapped with sense protein coding genes in the opposite strand and supported by 10 splicing junction reads in at least one sample. Then, we kept 68 unannotated antisense genes. After quantifying expression level and selecting further for tumor-enriched genes, we lastly identified 11 cryptic antisense genes in this batch of HCC tumors.

Differentially expressed gene analysis

To measure the expression correlation between CAPTs and their matched sense genes, a cutoff of fold change > 2 and adjusted P value < 0.05 was applied to obtain sense genes whose expression level was significantly changed in tumors with the presence of related CAPT. Only CAPT genes expressed at least five tumor samples in the corresponding cancer type were used for comparison.

Differentially expressed gene analysis upon *LRRK1*-CAPT knockdown in NCI-H1299 cells was performed using edgeR (52), and a statistical cutoff of adjusted P value < 0.05 and fold change > 2 was applied to obtain significantly dysregulated genes. GO enrichment analysis and GSEA were performed by clusterProfiler (53).

Comparative analysis with FANTOM5 CAGE peak and HCC CAGE peak

FANTOM5 CAGE peak expression matrix across about 900 samples including tissues, primary cells, and cancer cell lines was obtained using the hg19.cage_peak_phase1and2combined_tpm_ann.osc.txt file provided in the FANTOM5 website (19). We first used liftOver to convert the hg19 coordinates of FANTOM5 CAGE peaks and

HCC CAGE peaks to hg38 version. If the TSS of CAPT had a distance of < 500 nt to the annotated CAGE peaks in the same strand, then we thought that the CAPT was associated with CAGE peak. A transcript was considered expressed in a given FANTOM5 CAGE sample if the CAGE sample contained an active CAGE peak (tags per million > 0) with a distance of < 500 nt to the TSS of corresponding transcript in the same strand.

DNA methylation analysis

To investigate the change of DNA methylation level accompanied with the emergence of antisense transcript, we first filtered positions on the Illumina 450K methylation array for those occurring within ± 500 -nt regions around TSS of CAPT genes. To this end, we only observe 406 of the 450K array sites around TSS regions of identified CAPTs. Then, we compared the methylation level of specific array site in patients with the presence of related CAPTs with those with the absence of related CAPTs or matched normal tissues. A cutoff of median methylation level difference > 0.1 and adjusted P value < 0.05 was applied to obtain CAPT genes whose expression was associated with significantly changed methylation level. We also computed the Spearman correlation between CAPT expression level [$\log_2(\text{FPKM} + 1)$] and DNA methylation of associated CpG assay site [$\log_2(\text{Beta value})$] in all tumor samples that have both RNA-seq and Illumina 450K methylation array data.

TCGA ATAC-seq data analysis

For the pan-cancer ATAC-seq datasets, the normalized count matrix for 562,709 peaks of all samples was directly obtained from the corresponding study. If the TSS of CAPT had a distance of < 500 nt to the annotated ATAC-seq peaks, then we considered the CAPT as supported by ATAC peaks. For matched ATAC-seq peak and CAPT, we compared the chromatin accessibility of peak region in patients with the presence of corresponding CAPT with those with the absence of corresponding CAPT.

ChIP-seq data analysis

Among the raw reads obtained from ChIP-seq experiments, low-quality reads were filtered out, followed by alignment to human reference genome sequence (hg38). For dataset of colorectal cancer patient tissues, we first detected expressed CAPT in tumor samples by analyzing RNA-seq data, and then, we compared the histone modification signal level (including H3K4me3 and H3K27ac) around TSS region of CAPT in tumor samples with the presence of related CAPT with those without related CAPT or normal tissues. For dataset of NCI-H1299 cells upon epigenetic drugs treatment, we first detected the up-regulated CAPT upon treatment, and then, we examined the change of histone modification around TSS of up-regulated antisense transcript before and after treatment.

Cell culture and treatments

HEK293 (human embryonic kidney 293 cells), A549, NCI-H1299, and NCI-H520 (human lung cancer cells) cells were purchased from National Collection of Authenticated Cell Cultures. IMR90 cells were purchased from the American Type Culture Collection (ATCC). IMR90 cells were cultured in ATCC-formulated Eagle's minimum essential medium supplemented with 10% fetal bovine serum (FBS; Gibco). HEK293 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen, 11960044)

supplemented with 10% FBS (Gibco), 1% GlutaMAX (Invitrogen, 35050061), 1% sodium pyruvate (Invitrogen, 11360070), streptomycin (100 µg/ml), and penicillin (100 U/ml). A549 cells were cultured in F-12 K (Invitrogen, 21127-022) supplemented with 10% FBS, streptomycin (100 µg/ml), and penicillin (100 U/ml). NCI-H1299 cells were cultured in RPMI 1640 (Invitrogen, 11875093) supplemented with 10% FBS, 1% GlutaMAX, 1% sodium pyruvate, streptomycin (100 µg/ml), and penicillin (100 U/ml). NCI-H520 cells were cultured in RPMI 1640 supplemented with 10% FBS, streptomycin (100 µg/ml), and penicillin (100 U/ml). All cell lines were maintained at 37°C in a humidified incubator with 5% CO₂. NCI-H1299, IMR90, and HEK293 were treated with 500 nM DAC, 500 nM SB939, or 500 nM DAC + 500 nM SB939 for 72, 18, or 72 + 18 hours, respectively, and compound-containing medium was refreshed every 24 hours.

RNA preparation and qRT-PCR

Total RNA was extracted with TRIzol reagent (Invitrogen, 15596018) according to the manufacturer's instruction. Then, the total RNA was reverse-transcribed into cDNA with reverse transcription (RT-PCR) (TIANGEN, KR116-02) with oligo dT18-S primer. qRT-PCR was performed using the Bio-Rad CFX96 Real-Time PCR operating instrument with the ChamQ Universal SYBR qPCR Master Mix (Vazyme, Q711-02). The relative expression of each RNA was determined using the 2^{-ΔΔCt} method. Each qRT-PCR analysis was performed in triplicate.

RT-PCR was performed using the 2× Phanta Flash Master Mix (Vazyme, P520), and the targeted locus was PCR-amplified with the corresponding primers. Glyceraldehyde-3-phosphate dehydrogenase was measured as a loading control. The RT-PCR products were separated by gel electrophoresis through a 2% agarose gel in 1× tris-acetate-EDTA buffer (Tsingke Biotechnology Co.). To confirm the sequence of each band, RT-PCR products were gel-purified using the Zymoclean Gel DNA Extraction kit (Zymo) and verified by Sanger sequencing. Primer sequences were listed in table S5.

siRNA knockdown and CRISPR-mediated gene modulation

siRNA was synthesized by Gene Pharma (Shanghai, China). The sequences used are shown in table S5. The cells were transfected at a final concentration of 20 nM with a Lipofectamine iMax transfection reagent (Thermo Fisher Scientific, 13778075) per the manufacturer's instructions, nontargeting siRNA was used as a control. Briefly, 30 pmol siRNA was diluted in 150 µl of Opti-MEM. For each siRNA, 7.5 µl of iMAX was diluted in 150 µl of Opti-MEM and incubated for 5 min at room temperature. Diluted siRNA and RNAiMAX were mixed and incubated for another 5 min at room temperature. Cells were seeded in six-well plates at a density of 60% replaced with 1.2 ml of fresh medium before adding the complexes. Cells were grown for a total of 72 hours before being collected.

For CRISPR-mediated gene interference and activation, lenti-CRISPRv2- KRAB-dSpCas9 (Addgene plasmid no. 139094) and lentiCRISPRv2- dSpCas9-VPR (Addgene plasmid no. 139090) were constructed to express nontargeting gRNA or gRNAs that target *LRRK1*-CAPT. Lentiviral particles were produced in HEK293 cells using psPAX2 and pMD2.G packaging vectors using Lipofectamine 2000. Two days after infection, cells were selected and maintained with puromycin (2 µg/ml; Sigma-Aldrich,

p8833). The interference and activation efficiency were examined by qRT-PCR.

Rapid amplification of cDNA ends

The full length of CAPT was identified and amplified from the total RNA of NCI-H1299 cells and NCI-H520 by 5'- and 3'-RACE using the SMARTer RACE cDNA Amplification Kit (Takara Bio, USA, catalog no. 634858) following the manufacturer's protocol. The PCR products were subsequently cloned to the PCE2 vector by the TOPO-Blunt Cloning Kit (Vazyme, C602) and were validated by Sanger sequencing. The primers used for 5'- and 3'-RACE were listed in table S5.

Cell proliferation assay

Cells were plated 1500 cells per well on 96-well plates and incubated according to specific experimental design with three-well replicates. Cell viability was examined by CCK-8 assay (Dojindo, Rockville, MD, USA) following the protocol of the manufacturer.

Colony formation assay

For the colony formation assay, cells were digested and resuspended and counted under a microscope. In addition, the cells were cultured in six-well plate at a density of 800 cells per well. The cells were cultured under normal culture conditions for 14 days. The supernatant was removed, the cells were fixed with 4% paraformaldehyde, and the cells were stained with 1% crystal violet (Sigma-Aldrich) for 15 min. Then, the plates were washed with phosphate-buffered saline (PBS) and photographed.

Transwell migration assay

We conducted the transwell migration assay using a 24-well transwell chamber. The cells suspended in nonserum DMEM were seeded in the top chamber of the transwell with a density of 1 × 10⁴ per well, and 300 µl of fresh complete DMEM (10% FBS) was added to the bottom chamber. After incubating for 48 hours, the cells in the top chamber were washed with PBS twice and fixed with 4% paraformaldehyde for 15 min, followed by staining with 1% crystal violet (Sigma-Aldrich) for 30 min. After washing and wiping of the nonmigratory cells in the inner side of the top chamber, the migratory cells adhering to the bottom surface of the membrane were observed and photographed. Migratory cells were counted by ImageJ software.

Supplementary Materials

This PDF file includes:

Figs. S1 to S12

Legends for tables S1 to S5

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S5

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakraborty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow,

- E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, T. R. Gingeras, Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
2. A. Jacquier, The complex eukaryotic transcriptome: Unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.* **10**, 833–844 (2009).
3. M. J. Hangauer, I. W. Vaughn, M. T. McManus, Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet.* **9**, e1003569 (2013).
4. Z. Zhao, Q. Xu, R. Wei, L. Huang, W. Wang, G. Wei, T. Ni, Comprehensive characterization of somatic variants associated with intronic polyadenylation in human cancers. *Nucleic Acids Res.* **49**, 10369–10381 (2021).
5. PCAWG Transcriptome Core Group, C. Calabrese, N. R. Davidson, D. Demircioğlu, N. A. Fonseca, Y. He, A. Kahles, K.-V. Lehmann, F. Liu, Y. Shiraishi, C. M. Soulette, L. Urban, L. Greger, S. Li, D. Liu, M. D. Perry, Q. Xiang, F. Zhang, J. Zhang, P. Bailey, S. Erkek, K. A. Hoadley, Y. Hou, M. R. Huska, H. Kilpinen, J. O. Korbel, M. G. Marin, J. Markowski, T. Nandi, Q. Pan-Hammarström, C. S. Peadarallu, R. Siebert, S. G. Stark, H. Su, P. Tan, S. M. Waszak, C. Yung, S. Zhu, P. Awadalla, C. J. Creighton, M. Meyerson, B. F. Ouellette, K. Wu, H. Yang; PCAWG Transcriptome Working Group, A. Brazma, A. N. Brooks, J. Göke, G. Rätsch, R. F. Schwarz, O. Stegle, Z. Zhang; PCAWG Consortium, Genomic basis for RNA alterations in cancer. *Nature* **578**, 129–136 (2020).
6. K. Hashimoto, A. M. Suzuki, A. dos Santos, C. Desterke, A. Collino, S. Ghisletti, E. Braun, A. Bonetti, A. Fort, X. Y. Qin, E. Radaelli, B. Kaczkowski, A. R. R. Forrester, S. Kojima, D. Samuel, G. Natoli, M. A. Buendia, J. Faivre, P. Carninci, CAGE profiling of ncRNAs in hepatocellular carcinoma reveals widespread activation of retroviral LTR promoters in virus-induced tumors. *Genome Res.* **25**, 1812–1824 (2015).
7. D. Demircioğlu, E. Cukuroglu, M. Kindermans, T. Nandi, C. Calabrese, N. A. Fonseca, A. Kahles, K.-V. Lehmann, O. Stegle, A. Brazma, A. N. Brooks, G. Rätsch, P. Tan, J. Göke, A pan-cancer transcriptome analysis reveals pervasive regulation through alternative promoters. *Cell* **178**, 1465–1477.e1417 (2019).
8. T. Wiesner, W. Lee, A. C. Obenauf, L. Ran, R. Murali, Q. F. Zhang, E. W. P. Wong, W. Hu, S. N. Scott, R. H. Shah, I. Landa, J. Button, N. Lailler, A. Sboner, D. Gao, D. A. Murphy, Z. Cao, S. Shukla, T. J. Hollmann, L. Wang, L. Borsu, T. Merghoub, G. K. Schwartz, M. A. Postow, C. E. Ariyan, J. A. Fagin, D. Zheng, M. Ladanyi, K. J. Busam, M. F. Berger, Y. Chen, P. Chi, Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature* **526**, 453–457 (2015).
9. 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).
10. Q. L. Li, X. Lin, Y. L. Yu, L. Chen, Q. X. Hu, M. Chen, N. Cao, C. Zhao, C. Y. Wang, C. W. Huang, L. Y. Li, M. Ye, M. Wu, Genome-wide profiling in colorectal cancer identifies PHF19 and TBC1D16 as oncogenic super enhancers. *Nat. Commun.* **12**, 6407 (2021).
11. Z. Zhao, Q. Xu, R. Wei, W. Wang, D. Ding, Y. Yang, J. Yao, L. Zhang, Y. Q. Hu, G. Wei, T. Ni, Cancer-associated dynamics and potential regulators of intronic polyadenylation revealed by IPAfinder using standard RNA-seq data. *Genome Res.* **31**, 2095–2106 (2021).
12. V. Pelechano, L. M. Steinmetz, Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14**, 880–893 (2013).
13. S. Zhao, X. Zhang, S. Chen, S. Zhang, Natural antisense transcripts in the biological hallmarks of cancer: Powerful regulators hidden in the dark. *J. Exp. Clin. Cancer Res.* **39**, 187 (2020).
14. R. A. Gupta, N. Shah, K. C. Wang, J. Kim, H. M. Horlings, D. J. Wong, M. C. Tsai, T. Hung, P. Argani, J. L. Rinn, Y. Wang, P. Brzoska, B. Kong, R. Li, R. B. West, M. J. van de Vijver, S. Sukumar, H. Y. Chang, Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
15. M. Perteza, G. M. Perteza, C. M. Antonescu, T. C. Chang, J. T. Mendell, S. L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
16. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
17. O. A. Balbin, R. Malik, S. M. Dhanasekaran, J. R. Prensner, X. Cao, Y. M. Wu, D. Robinson, R. Wang, G. Chen, D. G. Beer, A. I. Nesvizhskii, A. M. Chinnaiyan, The landscape of antisense gene expression in human cancers. *Genome Res.* **25**, 1068–1079 (2015).
18. F. Bellido Molias, A. Sim, K. W. Leong, O. An, Y. Song, V. H. E. Ng, M. W. J. Lim, C. Ying, J. X. J. Teo, J. Göke, L. Chen, Antisense RNAs influence promoter usage of their counterpart sense genes in cancer. *Cancer Res.* **81**, 5849–5861 (2021).
19. The FANTOM Consortium and the RIKEN PMI and CLST (DGT), A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
20. A. Frankish, M. Diekhans, I. Jungreis, J. Lagarde, J. E. Loveland, J. M. Mudge, C. Sisu, J. C. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, C. Boix, S. Carbonell Sala, F. Cunningham, T. di Domenico, S. Donaldson, I. T. Fiddes, C. García Girón, J. M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, K. L. Howe, T. Hunt, O. G. Izuogu, R. Johnson, F. J. Martin, L. Martínez, S. Mohanan, P. Muir, F. C. P. Navarro, A. Parker, B. Pei, F. Pozo, F. C. Riera, M. Ruffier, B. M. Schmitt, E. Stapleton, M. M. Suner, I. Sycheva, B. Uszczyńska-Ratajczak, M. Y. Wolf, J. Xu, Y. T. Yang, A. Yates, D. Zerbinio, Y. Zhang, J. S. Choudhary, M. Gerstein, R. Guigó, T. J. P. Hubbard, M. Kellis, B. Paten, M. L. Tress, P. Flicek, GENCODE 2021. *Nucleic Acids Res.* **49**, D916–D923 (2021).
21. A. Li, J. Zhang, Z. Zhou, PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics* **15**, 311 (2014).
22. 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).
23. F. Spitz, E. E. Furlong, Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
24. Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenyk, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. M. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shores, C. B. Epstein, E. Gjoneska, 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.-H. Kim, A. Kulkarni, D. Li, R. Lowdon, G. N. 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).
25. G. C. Hon, R. D. Hawkins, O. L. Caballero, C. Lo, R. Lister, M. Pelizzola, A. Valsesia, Z. Ye, S. Kuan, L. E. Edsall, A. A. Camargo, B. J. Stevenson, J. R. Ecker, V. Bafna, R. L. Strausberg, A. J. Simpson, B. Ren, Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* **22**, 246–258 (2012).
26. A. Blattler, L. Yao, H. Witt, Y. Guo, C. M. Nicolet, B. P. Berman, P. J. Farnham, Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol.* **15**, 469 (2014).
27. E. Calo, J. Wsocka, Modification of enhancer chromatin: What, how, and why? *Mol. Cell* **49**, 825–837 (2013).
28. W. A. Flavahan, E. Gaskell, B. E. Bernstein, Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017).
29. B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, E. S. Lander, A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
30. Z. Wang, A. G. Chivu, L. A. Choate, E. J. Rice, D. C. Miller, T. Chu, S. P. Chou, N. B. Kingsley, J. L. Petersen, C. J. Finno, R. R. Bellone, D. F. Antczak, J. T. Lis, C. G. Danko, Prediction of histone post-translational modification patterns based on nascent transcription data. *Nat. Genet.* **54**, 295–305 (2022).
31. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
32. X. Zhang, P. S. Choi, J. M. Francis, G. F. Gao, J. D. Campbell, A. Ramachandran, Y. Mitsuishi, G. Ha, J. Shih, F. Vazquez, A. Tsherniak, A. M. Taylor, J. Zhou, Z. Wu, A. C. Berger, M. Giannakis, W. C. Hahn, A. D. Cherniack, M. Meyerson, Somatic superenhancer duplications and hotspot mutations lead to oncogenic activation of the KLF5 transcription factor. *Cancer Discov.* **8**, 108–125 (2018).
33. D. Brocks, C. R. Schmidt, M. Daskalakis, H. S. Jang, N. M. Shah, D. Li, J. Li, B. Zhang, Y. Hou, S. Laudato, D. B. Lipka, J. Schott, H. Bierhoff, Y. Assenov, M. Helf, A. Resnerova, M. S. Islam, A. M. Lindroth, S. Haas, M. Essers, C. D. Imbusch, B. Brors, I. Oehme, O. Witt, M. Lübbert, J. P. Mallm, K. Rippe, R. Will, D. Weichenhan, G. Stoeklin, C. Gerhäuser, C. C. Oakes, T. Wang, C. Plass, DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat. Genet.* **49**, 1052–1060 (2017).

34. M. Ghandi, F. W. Huang, J. Jané-Valbuena, G. V. Kryukov, C. C. Lo, E. R. McDonald III, J. Barretina, E. T. Gelfand, C. M. Bielski, H. Li, K. Hu, A. Y. Andreev-Drakhlin, J. Kim, J. M. Hess, B. J. Haas, F. Aguet, B. A. Weir, M. V. Rothberg, B. R. Paolella, M. S. Lawrence, R. Akbani, Y. Lu, H. L. Tiv, P. C. Gokhale, A. de Weck, A. A. Mansour, C. Oh, J. Shih, K. Hadi, Y. Rosen, J. Bistline, K. Venkatesan, A. Reddy, D. Sonkin, M. Liu, J. Lehar, J. M. Korn, D. A. Porter, M. D. Jones, J. Golji, G. Caponigro, J. E. Taylor, C. M. Dunning, A. L. Creech, A. C. Warren, J. M. McFarland, M. Zamanighomi, A. Kauffmann, N. Stransky, M. Imielinski, Y. E. Maruvka, A. D. Cherniack, A. Tsherniak, F. Vazquez, J. D. Jaffe, A. A. Lane, D. M. Weinstock, C. M. Johannessen, M. P. Morrissey, F. Stegmeier, R. Schlegel, W. C. Hahn, G. Getz, G. B. Mills, J. S. Boehm, T. R. Golub, L. A. Garraway, W. R. Sellers, Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).
35. L. Wang, H. J. Park, S. Dasari, S. Wang, J. P. Kocher, W. Li, CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41**, e74 (2013).
36. Y. J. Kang, D. C. Yang, L. Kong, M. Hou, Y. Q. Meng, L. Wei, G. Gao, CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017).
37. L. A. Gilbert, M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann, J. S. Weissman, Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
38. R. M. Meyers, J. G. Bryan, J. M. McFarland, B. A. Weir, A. E. Sizemore, H. Xu, N. V. Dharia, P. G. Montgomery, G. S. Cowley, S. Pantel, A. Goodale, Y. Lee, L. D. Ali, G. Jiang, R. Lubonja, W. F. Harrington, M. Strickland, T. Wu, D. C. Hawes, V. A. Zhivich, M. R. Wyatt, Z. Kalani, J. J. Chang, M. Okamoto, K. Stegmaier, T. R. Golub, J. S. Boehm, F. Vazquez, D. E. Root, W. C. Hahn, A. Tsherniak, Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
39. M. K. Iyer, Y. S. Niknafs, R. Malik, U. Singhal, A. Sahu, Y. Hosono, T. R. Barrette, J. R. Prensner, J. R. Evans, S. Zhao, A. Poliakov, X. Cao, S. M. Dhanasekaran, Y. M. Wu, D. R. Robinson, D. G. Beer, F. Y. Feng, H. K. Iyer, A. M. Chinnaiyan, The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208 (2015).
40. L. Lorenzi, H. S. Chiu, F. Avila Cobos, S. Gross, P. J. Volders, R. Cannoodt, J. Nuytens, K. Vanderheyden, J. Anckaert, S. Lefever, A. P. Tay, E. J. de Bony, W. Trypsteen, F. Gysens, M. Vromman, T. Goovaerts, T. B. Hansen, S. Kuersten, N. Nijs, T. Taghon, K. Vermaelen, K. R. Bracke, Y. Saeys, T. de Meyer, N. P. Deshpande, G. Anande, T. W. Chen, M. R. Wilkins, A. Unnikrishnan, K. de Preter, J. Kjems, J. Koster, G. P. Schroth, J. Vandesompele, P. Sumazin, P. Mestdagh, The RNA Atlas expands the catalog of human non-coding RNAs. *Nat. Biotechnol.* **39**, 1453–1465 (2021).
41. Y. Tay, J. Rinn, P. P. Pandolfi, The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505**, 344–352 (2014).
42. J. A. West, C. P. Davis, H. Sunwoo, M. D. Simon, R. I. Sadreyev, P. I. Wang, M. Y. Tolstorukov, R. E. Kingston, The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* **55**, 791–802 (2014).
43. M. Li, H. Gou, B. K. Tripathi, J. Huang, S. Jiang, W. Dubois, T. Waybright, M. Lei, J. Shi, M. Zhou, J. Huang, An Apela RNA-containing negative feedback loop regulates p53-mediated apoptosis in embryonic stem cells. *Cell Stem Cell* **16**, 669–683 (2015).
44. A. C. West, R. W. Johnstone, New and emerging HDAC inhibitors for cancer treatment. *J. Clin. Invest.* **124**, 30–39 (2014).
45. S. C. Navada, J. Steinmann, M. Lübbert, L. R. Silverman, Clinical development of demethylating agents in hematology. *J. Clin. Invest.* **124**, 40–46 (2014).
46. K. B. Chiappinelli, P. L. Strissel, A. Desrichard, H. Li, C. Henke, B. Akman, A. Hein, N. S. Rote, L. M. Cope, A. Snyder, V. Makarov, S. Buhu, D. J. Slamon, J. D. Wolchok, D. M. Pardoll, M. W. Beckmann, C. A. Zahnow, T. Merghoub, T. A. Chan, S. B. Baylin, R. Strick, Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* **162**, 974–986 (2015).
47. D. Roulois, H. Loo Yau, R. Singhanian, Y. Wang, A. Danesh, S. Y. Shen, H. Han, G. Liang, P. A. Jones, T. J. Pugh, C. O'Brien, D. D. de Carvalho, DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* **162**, 961–973 (2015).
48. Y. C. Liu, J. Kwon, E. Fabiani, Z. Xiao, Y. V. Liu, M. Y. Follo, J. Liu, H. Huang, C. Gao, J. Liu, G. Falconi, L. Valentini, C. Gurnari, C. Finelli, L. Cocco, J. H. Liu, A. I. Jones, J. Yang, H. Yang, J. A. I. Thoms, A. Unnikrishnan, J. E. Pimanda, R. Pan, M. A. Bassal, M. T. Voso, D. G. Tenen, L. Chai, Demethylation and up-regulation of an oncogene after hypomethylating therapy. *N. Engl. J. Med.* **386**, 1998–2010 (2022).
49. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
50. C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, L. Pachter, Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
51. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
52. D. J. McCarthy, Y. Chen, G. K. Smyth, Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
53. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

Acknowledgments

Funding: This work was financially supported by the National Key R&D Program of China (2021YFA0909300 and 2020YFC2005000) and the National Natural Science Foundation of China (92249302 and 91949107). **Author contributions:** Z.Z. and Y.C. designed the study and analyzed the data. L.H., H.W., Q.X., X.Zhou, and X.C. provided part of the data. Z.Z., Y.C., J.C., and T.N. wrote the manuscript. T.N., J.C., and X.Zhang supervised the study and revised the manuscript. **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. The sequencing data generated in this study have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE220952 in a preprocessed format.

Submitted 14 October 2022

Accepted 7 March 2023

Published 5 April 2023

10.1126/sciadv.adf3264