

Proximity of pre-mRNA 3' end processing and transcription termination is associated with enhanced gene expression

Agata Stępień ^{1,2}✉, Deepshika Pulimamidi ^{1,2,8}, Martyna Plens-Gałąska^{1,2,9}, Maxime Lalonde ^{3,9}, Magda Kopczyńska ^{1,2,9}, Klaudia Ostatek ^{1,2}, Mariia Kapeliukha^{1,2}, Hiroshi Kimura ⁴, Shazia Irshad ⁵, Michał R Gdula ^{1,6}, Stephan Hamperl ³ & Kinga Kamieniarz-Gdula ^{1,2,7}✉

Abstract

Two key factors required for pre-mRNA 3' end processing, CPSF73 and PCF11, are linked to poor cancer prognosis, yet their oncogenic roles remain unclear due to their dual function in transcription termination. Using a colorectal cancer (CRC) progression model, we show that primary tumor cells depend on high CPSF73 and PCF11 levels, whereas metastatic cells are less sensitive. This dependency is not explained by alternative polyadenylation (APA) or transcription-replication conflicts. Instead, we identify an uncoupling of 3' cleavage and transcription termination during CRC progression. Primary tumor cells exhibit a global shift toward proximal termination alongside distal APA usage, while metastatic cells partially revert to normal termination patterns and favor proximal APA. Notably, the distance between 3' cleavage and termination is nearly halved in primary tumor cells compared with normal cells and becomes particularly short in oncogenes. Reduced cleavage-termination distance correlates with increased gene expression both within and across cell states. These findings identify a previously unrecognized link between transcription termination architecture and gene expression regulation.

Subject Categories Chromatin, Transcription & Genomics; RNA Biology
<https://doi.org/10.1038/s44318-026-00928-4>

Received 19 March 2026; Revised 24 August 2026;

Accepted 7 September 2026

Published online: 14 October 2026

Introduction

The annotated gene end is defined by the site of its pre-mRNA 3' end processing. It is catalyzed by the cleavage and polyadenylation (CPA) complex and can occur at different alternative

polyadenylation sites (PAS) within a single gene. The levels of CPA factors are frequently deregulated in various cancers. Many CPA factors, such as CPSF1, CPSF3, CPSF4, and CSTF2, are upregulated in different cancer types (Xia et al, 2014; Chen et al, 2013, 2021, 2022; Alahmari et al, 2024), while CFIm25 is downregulated in liver, breast, and cervical cancers (Tan et al, 2018; Tamaddon et al, 2020; Xing et al, 2021). In line with the changes in CPA factor levels, a global 3' untranslated region (UTR) shortening and increased usage of intronic PAS are often observed in cancer cells (Xia et al, 2014; Mayr and Bartel, 2009; Lin et al, 2012; Burri and Zavolan, 2021). However, in all cancer types tested, 3' UTR changes are not unidirectional, and particularly abundant 3' UTR lengthening events have been reported in colorectal and breast cancers (Fischl et al, 2019; Qiao et al, 2024). Altogether, these findings highlight the complexity of cleavage and polyadenylation dysregulation in cancer.

The second crucial event at gene ends is RNA polymerase II (RNAPII) transcription termination. Termination occurs when RNAPII stops RNA synthesis and dissociates from the DNA template. It is necessary to separate adjacent RNAPII transcription units, both in tandem and convergent orientation, for co-regulating the expression of multiple genes (Whitelaw and Proudfoot, 1986; Greger, 1998; Prescott and Proudfoot, 2002). Indeed, defective termination results in transcriptional interference on neighboring genes, which can lead to a severe reduction in mRNA levels. Collided RNAPII are often removed from the DNA template through ubiquitylation-directed proteolysis (Hobson et al, 2012). Thus, it is not surprising that the production of readthrough transcripts in healthy human tissues is restricted in genomic regions with high gene activity (Caldas et al, 2024).

Transcription termination of RNAPII-transcribed units is triggered by a whole plethora of complexes and protein factors. Some species of non-coding transcripts are terminated by Integrator and Microprocessor complexes (Baillat et al, 2005;

¹Center for Advanced Technologies, Adam Mickiewicz University, 61-614 Poznań, Poland. ²Department of Molecular and Cellular Biology, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, 61-614 Poznań, Poland. ³Chromosome Dynamics and Genome Stability, Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München, 81377 Munich, Germany. ⁴Cell Biology Centre, Institute of Integrated Research, Institute of Science Tokyo, 226-8503 Yokohama, Japan. ⁵Nuffield Department of Clinical Laboratory Sciences, Radcliffe Department of Medicine, University of Oxford, OX3 9DU Oxford, UK. ⁶Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, 61-614 Poznań, Poland. ⁷X: @STOPlabPI ⁸These authors contributed equally as first authors: Agata Stępień, Deepshika Pulimamidi. ⁹These authors contributed equally: Martyna Plens-Gałąska, Maxime Lalonde, Magda Kopczyńska.

✉E-mail: agata.stepien@amu.edu.pl; kinga.kamieniarz-gdula@amu.edu.pl

Lykke-Andersen et al, 2021; Wagner et al, 2023; Dhir et al, 2015). The Restrictor limits the expression of promoter upstream transcripts (PROMPTs) and early transcription (Austena et al, 2021; Estell et al, 2021; Rouvière et al, 2023). Additionally, DNA-guided termination on T-tracts can function in RNAPII termination, as recently shown in yeast and humans (Han et al, 2023; Davidson et al, 2024). This diversity of (partially redundant) mechanisms and complexes involved in RNAPII transcription termination underscores both the critical importance and the complexity of the process.

Transcription termination of protein-coding genes is mainly triggered by CPA factors and the XRN2 5'-3' exonuclease (Eaton et al, 2018; Rodríguez-Molina et al, 2023). In humans, this usually occurs several thousand nucleotides downstream of PAS (Kopczyńska et al, 2024; Schlackow et al, 2017; Kamieniarz-Gdula et al, 2019). Mechanistically, there are two key aspects of transcription termination on protein-coding genes, unified in a single model (Logan et al, 1987; Connelly & Manley, 1988; Proudfoot, 1989; Eaton et al, 2020). First, transcription of the polyadenylation signal induces an allosteric change in RNAPII and makes it competent for termination (Carminati et al, 2023). Second, pre-mRNA 3' end cleavage exposes the 5' end of nascent RNA, allowing XRN2 to degrade it and help to displace RNAPII (Kim et al, 2004; West et al, 2004; Luo et al, 2006; Cortazar et al, 2022). Transcription termination is facilitated and preceded by RNAPII pausing, and terminally paused RNAPII can be detected by a specific phospho-mark on the C-terminal domain (CTD) of RNAPII's largest subunit (RNAPII CTD-T4ph) (Kopczyńska et al, 2024).

Two components of the CPA complex, CPSF73 and PCF11, are particularly important for transcription termination of protein-coding genes. CPSF73 is the endoribonuclease that cleaves the 3' end of pre-mRNA (Mandel et al, 2006). Notably, rapid depletion of CPSF73 leads to extensive transcription readthrough, a hallmark of impaired termination (Eaton et al, 2020). PCF11 is essential for pre-mRNA 3' cleavage both in yeast and humans (Gross and Moore, 2001; Schmidt et al, 2022; Boreikaite et al, 2022). It is also the only CPA factor shown to directly trigger transcription termination by binding to the RNAPII CTD and slowing it down (Kamieniarz-Gdula et al, 2019; Birse et al, 1998; Sadowski, 2003; Zhang et al, 2005; West & Proudfoot, 2007; Grzechnik et al, 2015; Liu et al, 2025a). Besides stimulating normal termination on protein-coding gene ends globally, PCF11 also triggers premature termination within the introns of specific genes (Kamieniarz-Gdula et al, 2019; Wang et al, 2019; Kamieniarz-Gdula and Proudfoot, 2019).

Given their biological importance, it is not surprising that the levels and activity of PCF11 and CPSF73 are critical in cancer biology, and their manipulation has emerged as a promising therapeutic strategy (Pagani and Gandellini, 2024). Low PCF11 levels are associated with higher survival rates in neuroblastoma patients and spontaneous tumor regression (Ogorodnikov et al, 2018), as well as a better prognosis in hematologic malignancies (Salifu and Doughan, 2022). Additionally, single-nucleotide variants in the promoter of *PCF11* are recurrent in different cancer types (Hornshøj et al, 2018). Similarly, pancreatic tumor patients with high *CPSF73* expression had significantly worse overall survival than patients with low *CPSF73* expression (Alahmari et al, 2024). These findings justify the targeting of CPSF73 and PCF11 in cancer therapies.

For instance, JTE-607 is a direct chemical inhibitor of CPSF73 activity, which was initially believed to inhibit cytokine production in human cells (Kakutani et al, 1999; Kakegawa et al, 2019; Ross

et al, 2020). In a leukemia model, JTE-607 administration significantly prolonged mice survival (Uesato et al, 2006; Tajima et al, 2010). The compound has also been reported to attenuate glioblastoma growth in vivo as well as breast and pancreatic cancer cell proliferation in vitro (Alahmari et al, 2024; Lin et al, 2024; Liu et al, 2022). While CPA and termination factors CPSF73 and PCF11 emerge as novel targetable vulnerabilities for human cancers, it is unclear whether their roles in pre-mRNA 3' processing or transcription termination are crucial for oncogenesis. Furthermore, the mechanisms linking their increased expression with poor patient prognosis remain obscure.

Colorectal cancer (CRC) is the third most diagnosed malignant tumor and the second leading cause of cancer deaths (Ferlay et al, 2019; Bray et al, 2024). It is an etiologically heterogeneous disease that develops through three distinctive pathways: adenoma-carcinoma sequence, serrated pathway, and inflammatory pathway (Keum and Giovannucci, 2019). Although several critical genes have been reported to contribute to CRC development, ~60–65% of CRC cases arise sporadically (Keum and Giovannucci, 2019). Due to CRC's high prevalence, accompanied by poor tumor-selective drug delivery and low treatment efficacy in patients, there is a critical need to develop new therapies for CRCs (Maeda and Khatami, 2018).

Here, we employ a cell line-based model of colorectal cancer progression that exhibits elevated expression of CPA and termination factors, consistent with patient biopsies. We first show that the clonogenic potential of CRC cells critically depends on PCF11 and CPSF73, with cells from primary tumors displaying particularly high addiction to those factors. We then demonstrate that delaying transcription termination, either by genetic manipulation of termination factors or by chemical inhibition, does not increase the frequency of transcription-replication collisions, contrary to previous assumptions. Further on, we find no evidence supporting a role for alternative polyadenylation in CRC progression. Instead, by systematically characterizing the genomic landscape of RNAPII transcription termination and pre-mRNA 3' end cleavage site usage in our model, we uncover an unexpected uncoupling between changes in transcription termination and in pre-mRNA 3' end processing during CRC progression. To quantify this uncoupling, we introduce a new measure—distance between the cleavage site and the location in which RNAPII enters terminal pausing. We find that CRC cell sensitivity to CPA/termination factor depletion is associated with a shortened distance between pre-mRNA 3' cleavage and transcription termination - particularly short in the case of oncogenes. Finally, we show that shortening of the cleavage-termination distance (i.e., increased proximity of cleavage and termination) on a gene is associated with its enhanced 3' end formation and increased gene expression.

Results

Primary and metastatic tumor cells show different sensitivities to RNA 3' cleavage and transcription termination abrogation

To study the role of CPSF73 and PCF11 in oncogenesis, we employed a cellular model of colorectal cancer (CRC) progression: 1CT, non-transformed human colon epithelial cells (Roig et al,

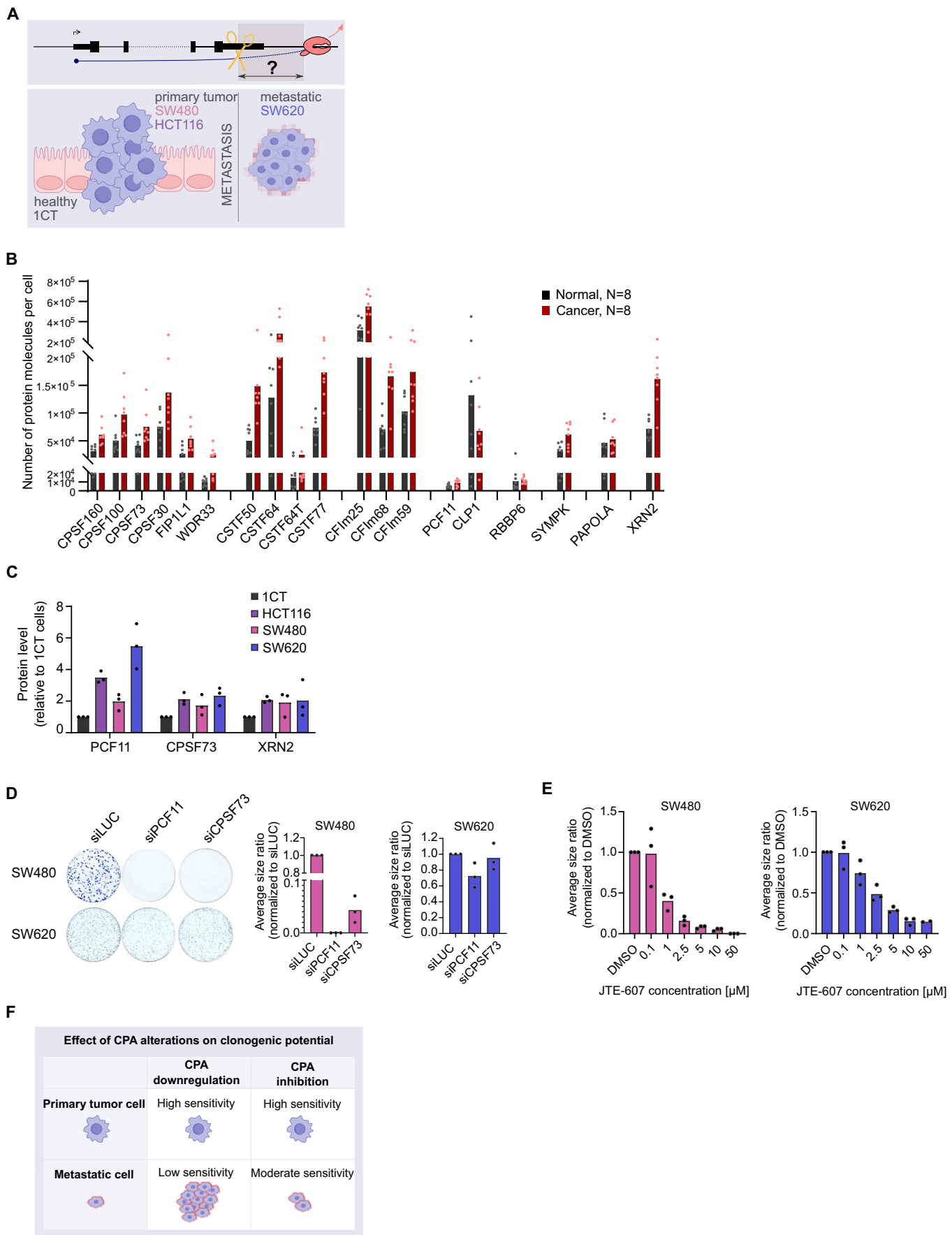


Figure 1. Primary CRC cells are more sensitive than metastatic cells to 3' pre-mRNA cleavage and transcription termination abrogation.

(A) Design of this study. The interplay between pre-mRNA 3' cleavage and transcription termination was studied in a cell culture model of CRC progression. 1CT, non-transformed human colon epithelial cells; HCT116 and SW480, cells derived from colon primary tumor; SW620, cells isolated from metastasis in the SW480 donor patient. (B) Bar graph showing the number of protein molecules per cell of CPA and termination factors (data from patient biopsies (Wiśniewski et al, 2015)). (C) Bar graph showing the quantification of Western blots for PCF11, CPSF73, and XRN2 in the indicated cell lines. (D, E) Colony-formation assays in SW480 and SW620 cells. Cells were treated with siPCF11 or siCPSF73 (D), or with CPSF73 inhibitor JTE-607 (E). Here and in all other figures, bar graphs represent the mean, while dots indicate individual biological replicates. (F) Schematic representation of the sensitivity of CRC cells to PCF11 or CPSF73 manipulation. Source data are available online for this figure.

2010); HCT116, a widely used cell line derived from a primary tumor of the colon (Brattain et al, 1981); and two cell lines derived from the same patient: SW480 (from the primary tumor of the colon) and SW620 (derived one year later from metastasis to the lymph nodes) (Leibovitz et al) (Fig. 1A). This model recapitulates changes in the levels of the main players in pre-mRNA 3' end processing and transcription termination of protein-coding genes: PCF11, CPSF73, and XRN2, as observed in patients. Proteomic data from biopsies of healthy colon and colorectal carcinoma (Wiśniewski et al, 2015) indicated that the levels of all analyzed cleavage and polyadenylation (CPA) and termination factors, with the exception of CLP1, increased in biopsies from colon cancer patients compared to normal colon tissue. This increase was observed in both absolute values as well as normalized to RNAPII (Figs. 1B and EV1A). The increase of PCF11 and CPSF73 protein levels in cancer tissue was further corroborated in Clinical Proteomic Tumor Analysis Consortium data (Fig. EV1B).

In line with the data obtained from patient samples, western blot analysis in our cellular model revealed a gradual increase of PCF11 protein levels concomitant with colorectal cancer progression (Figs. 1C and EV1C). The levels of both CPSF73 and XRN2 were also elevated in CRC cells compared to 1CT cells, but did not increase from primary to metastatic tumor cells (Figs. 1C and EV1C). We conclude that protein levels of CPA and transcription termination factors are generally increased in CRC cells compared to normal colonic epithelial cells—both in our cellular model and in patient samples.

We then checked if and how the cancerous properties of primary tumor and metastatic CRC cells are affected by the loss of CPA and termination factor levels or activity. To address this, we first used SW480 and SW620 cell lines, as they originate from the same patient (Fig. 1A). Due to different cell and colony size and different baseline clonogenic potential (Jana et al, 2020; Abdulrehman et al, 2018), we normalized each biological replicate to the control sample to analyze the response to the treatment only. Knockdown of CPSF73 resulted in a strong inhibition of colony formation in SW480 cells, and PCF11 knockdown blocked colony formation altogether (Figs. 1D and EV1D). In contrast, the clonogenic potential of SW620 cells was unaltered by CPSF73 downregulation and only slightly decreased upon PCF11 knockdown (Figs. 1D and EV1D). Inhibition of CPSF73 endo-nucleolytic activity by JTE-607 at high concentrations abolished the ability of SW480 single cells to form colonies and decreased this ability for SW620 cells (Fig. 1E; Appendix Fig. S1A). This differential sensitivity of cells derived from primary and metastatic tumors in response to JTE-607 was consistent across a wide range of inhibitor concentrations (Fig. 1E; Appendix Fig. S1A). We then tested the sensitivity to JTE-607 in two further, unrelated CRC cell lines. Primary tumor-derived HCT116 cells again showed a greater

reduction in colony-forming ability than metastatic T84 cells (Fig. EV1E,F; Appendix Fig. S1B).

These experiments suggest that the clonogenic potential of the tested primary tumor-derived CRC cells is highly sensitive to changes in CPA factor levels and is readily impaired by their manipulation. In contrast, tested cells derived from CRC metastasis show a higher tolerance to CPA factor alterations, maintaining their ability to form colonies (Fig. 1F).

Induction of transcriptional readthrough by alteration of CPA factor levels or activity in CRC cells does not affect the frequency of transcription-replication collision

We next sought to understand the molecular background underlying the different sensitivities of primary tumor and metastatic CRC cell lines to varying levels of CPA/termination factors. Depletion of the elongation factor NELF-C has been shown to delay termination and increase the probability of collisions between RNAPII and replication machinery (Nakayama et al, 2026). In line with this possibility, impaired pre-mRNA 3' end cleavage and transcription termination were shown to cause reduction of replication fork speed and excessive origin firing (Teloni et al, 2019). This could lead to a higher frequency of transcription-replication collisions (TRCs) and cell death.

Before testing the hypothesis that CPSF73 and/or PCF11 downregulation increases the number of TRCs, we first assessed the number of TRCs in normal, primary, and metastatic tumor-derived cells. For this, we used the proximity ligation assay (PLA), which measures contacts between the transcription and replication machineries (Hamperl et al, 2017). PCNA and T4ph RNAPII were used as markers for replication and transcription termination, respectively (Fig. 2A). Intriguingly, we observed a gradual decrease in TRCs in our cancer progression model, with the highest frequency in normal cells and the lowest in metastatic cells (Figs. 2A, right and EV2A,B). However, when we manipulated CPA/termination factor levels or RNA 3' cleavage activity, we did not observe the increase in TRC frequency predicted by our hypothesis. This result was consistent across multiple conditions: knock down of PCF11 and CPSF73 in SW620 cells (Fig. 2B, left side), CPSF73 endo-nucleolytic activity inhibition by JTE-607 in SW480 and SW620 (Fig. 2B, right side), or CPSF73 depletion by an IAA-induced degron system in HCT116 cells (Eaton et al, 2020) (Figs. 2C and EV2C). The only manipulation that resulted in a mildly elevated frequency of TRCs was depletion of XRN2 in HCT116 cells for 4 h (but not after 8 h) (Eaton et al, 2018) (Fig. EV2D). The percentage of replicating cells (i.e., stained with EdU) after PCF11 or CPSF73 KD in SW620 cells also remained constant (Fig. EV2E).

To rule out potential bias from using the T4ph RNAPII-specific antibody in this assay, we additionally measured TRCs associated with serine 2 phosphorylated (S2ph) elongating RNAPII and

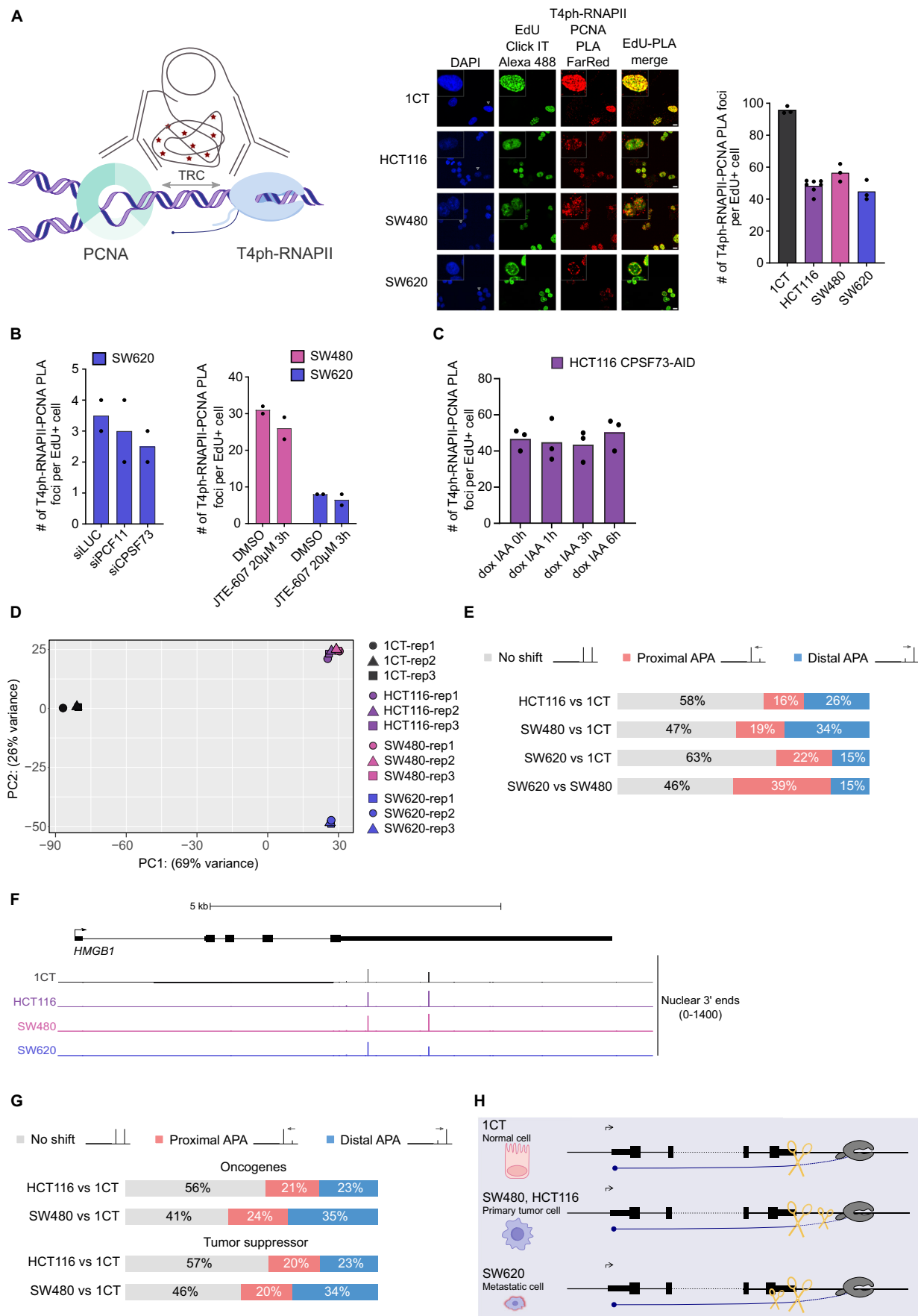


Figure 2. The sensitivity of primary CRC cells is driven neither by the transcription-replication collisions nor by alternative polyadenylation.

(A) Left: schematic representation of transcription-replication collisions (TRCs) quantification using proximity ligation assay (PLA). Right: representative pictures and TRC frequencies in the CRC model. Scale bar = 10 μ m. (A–C) Bar graphs represent average values, each dot represents one biological replicate ($N = 3–6$), calculated as the mean of 2–3 technical replicates (100–1500 analyzed cells per technical replicate). The exact numbers are provided in the Appendix Supplementary Methods. (B) Frequency of TRCs in control conditions (siLUC, DMSO) and after knockdown of PCF11 (siPCF11) or CPSF73 (siCPSF73), or JTE-607 treatment in SW480 and SW620 cells. (C) Frequency of TRCs after CPSF73 depletion (dox IAA 0–6 h) in HCT116 CPSF73-AID cells. (D) Principal component analysis (PCA) plot of 3' mRNA-seq biological replicates ($N = 3$ biological replicates). (E–G) The analysis was performed on the merged three biological replicates. (E) Bar graph representing the alternative polyadenylation (APA) analysis performed for separated active protein-coding genes ($n = 12,469$) in the CRC progression model. (F) Genomic profile of *HMGBl*, showing nuclear 3' mRNA-seq signals. (G) Bar graph representing the APA analysis in the CRC progression model for separated oncogenes ($n = 533$) and tumor suppressor genes ($n = 764$) in primary tumor cells. (H) Schematic representation of PAS selection in the CRC progression model. The small scissors indicate APA direction. Source data are available online for this figure.

PCNA, as previously described (Hamperl et al, 2017; Narain et al, 2021; Bayona-Feliu et al, 2021). Consistent with the previous results, neither knockdown of PCF11 or CPSF73 in SW620 cells nor treatment with JTE-607 in SW480 and SW620 cells led to changes in S2ph-linked TRC frequencies (Fig. EV2F). Finally, we combined extended PCF11 knockdown (72 h) with high-concentration JTE-607 treatment (50 μ M) in SW620 cells, which still did not alter TRC frequency (Fig. EV2G).

We conclude that PCF11/CPSF73 depletion and/or pre-mRNA 3' cleavage inhibition, which were previously demonstrated to delay termination and induce transcriptional readthrough (Kamieniarz-Gdula et al, 2019; Liu et al, 2023; Cui et al, 2023; Podsiwiałow-Bartnicka et al, 2025), do not affect the number of transcription-replication collisions in CRC cells. Therefore, TRCs do not appear to underlie the sensitivity of CRC cells to PCF11/CPSF73 modulation in colony-formation assays (Fig. 1D–F).

Upregulation of CPA factor levels in CRC cells is associated with complex APA events instead of gradual mRNA shortening

A previous study investigating the effects of the JTE-607 inhibitor suggested that cells with higher CPA factor expression levels may undergo mRNA shortening, rendering them more sensitive to JTE-607 (Cui et al, 2023). To test this hypothesis in our model, we performed 3' mRNA sequencing of nuclear RNA. Principal component analysis (PCA) revealed not only high reproducibility between biological replicates, but also a clear clustering and overlap of data from the two cell lines derived from primary tumors of different patients. This remarkable level of similarity in their PAS usage suggests a strong transcriptional memory of their tissue of origin (Fig. 2D).

We then conducted alternative polyadenylation (APA) analysis to determine the direction of changes in 3' end usage of CRC cells relative to 1CT (Figs. 2E and EV2H; Dataset EV1; see also Methods). 3422 analyzed genes had a single PAS. Among genes undergoing APA, metastatic SW620 cells showed more proximal (22%, 2010 genes) than distal (15%, 1374 genes) APA events relative to 1CT. Strikingly, both primary tumor cell lines (HCT116 and SW480) displayed a clear opposite trend with a preference for distal (26%, 2393 genes and 34%, 3090 genes, respectively) relative to proximal (16%, 1455 genes and 19%, 1750 genes) APA events (Figs. 2E and EV2H; Dataset EV1). Inspection of individual genes confirmed this analysis (Figs. 2F and EV2I; Dataset EV1).

We further checked whether genes involved in carcinogenesis show any tendency towards APA changes in the CRC model. We performed APA analysis for oncogenes and tumor suppressor

genes. The results mirrored the trend observed for separated active protein-coding genes (see Methods), with no increase in APA incidence or directionality for cancer-related genes in CRC (Figs. 2E,G and EV2H,J). Moreover, we did not observe a consistent correlation between APA and nuclear RNA level (Fig. EV2K). Together, these findings show that upregulation of CPA factor levels is not associated with gradual mRNA shortening, but rather with complex, bi-directional APA events. They also suggest that in our model, APA showed no major global contribution to carcinogenesis or to the sensitivity of primary tumor cells to CPA factor alterations (Fig. 2H).

Changes in PAS selection and transcription termination are uncoupled during CRC progression

Since CPSF73 and PCF11 are required not only for pre-mRNA 3' cleavage but also for terminating transcription on protein-coding genes, we next investigated nascent transcription and transcription termination changes in our colorectal cancer model. We first globally examined nascent transcription profiles using polymerase-intact nascent transcript sequencing (POINT-seq) (Sousa-Luis et al, 2021). For a positive control and also to mirror the conditions of our colony formation and transcription-replication collisions assays (Figs. 1D–F and 2A,B), cells were treated with the CPSF73 inhibitor, JTE-607. JTE-607 treatment caused a strong decrease in nascent transcription from the TSS throughout the gene body, as measured by POINT-seq, and strong transcriptional readthrough downstream of the PAS in both SW480 and SW620 cell lines (Figs. 3A,B and EV3A). Surprisingly, while SW480 cells were more sensitive to JTE-607 in colony-formation assays, the termination defect upon JTE-607 treatment was much stronger in SW620 cells (Figs. 3A,B and EV3A). Comparison between the CRC cell lines in control conditions showed the same trend—SW620 cells had markedly delayed termination, as compared to SW480 cells (Figs. 3A,B and EV3A).

To analyze the nascent transcription profile with a higher (single nucleotide) resolution and in a RNAPII pausing-sensitive way, we then used mammalian native elongating transcript sequencing (mNET-seq) with an antibody recognizing total RNAPII (Nojima et al, 2015), in all four cell lines of our CRC model. This revealed a proximal shift of nascent transcripts associated with total RNAPII (earlier termination) in PAS regions of cells derived from primary CRC (SW480 and HCT116), suggesting accelerated termination (Fig. 3C,D).

To obtain an even more sensitive and specific view of transcription termination, we next performed mNET-seq with an antibody against RNAPII T4ph - which detects nascent RNA in the

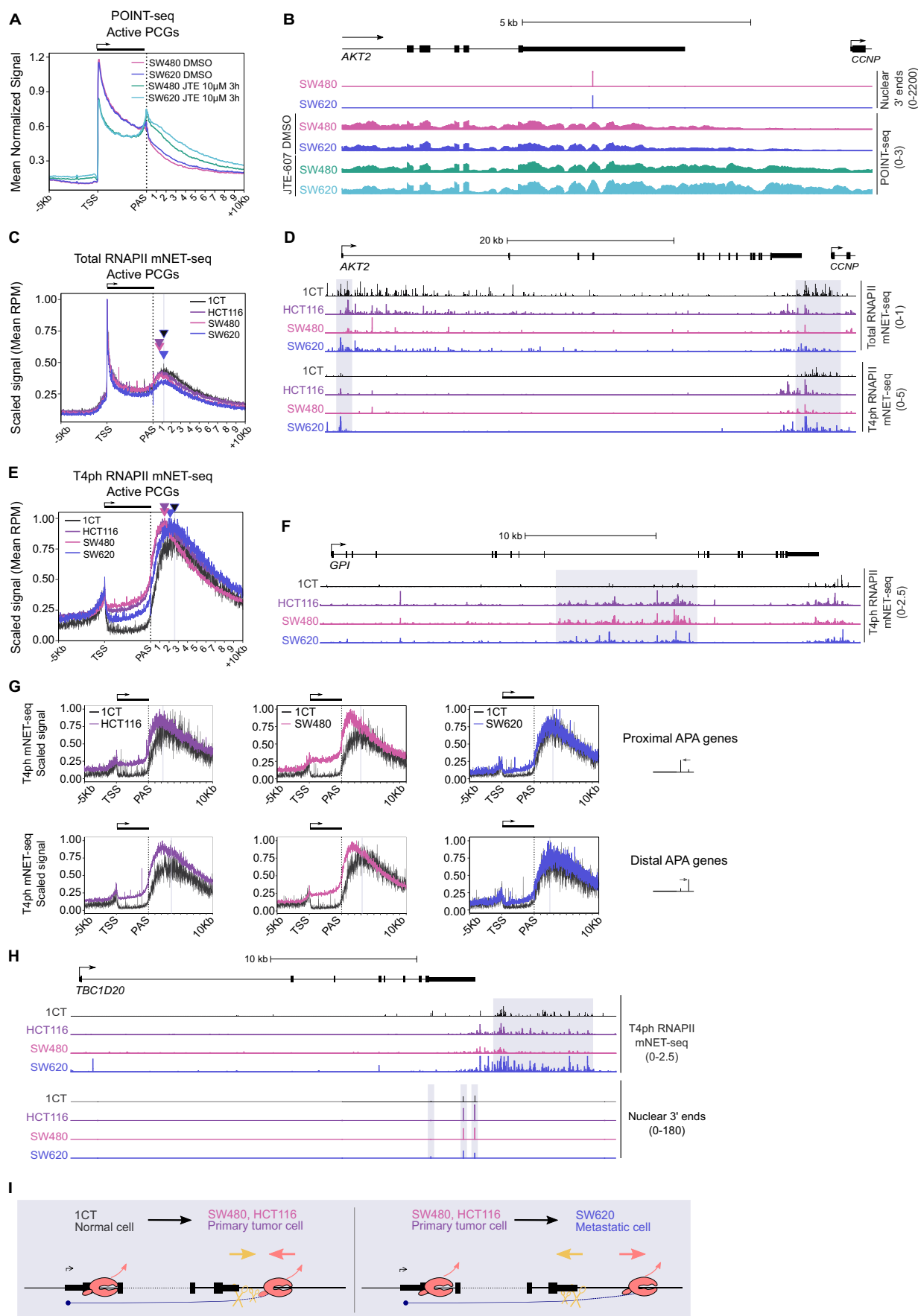


Figure 3. Changes in PAS selection and transcription termination are uncoupled during CRC progression.

(A) Metagene profiles of POINT-seq signal along active protein-coding genes (PCGs) ($n = 13,577$). The dotted line indicates the PAS position. Based on two merged biological replicates. (B) Genomic profiles of *AKT2*, showing nuclear 3' mRNA-seq and POINT-seq signals. (C, E) Metagene profiles of Total RNAPII (C) and T4ph RNAPII (E) mNET-seq signal along active protein-coding genes (PCGs) ($n = 13,577$). The dotted line indicates the PAS position, and the gray highlight indicates the signal maximum downstream of PAS in 1CT cells. Triangles depict the signal maxima downstream of PAS. (C–H) Total RNAPII mNET-seq, $N = 2$ biological replicates; T4ph mNET-seq, $N = 2$ (1CT, HCT116, SW480) or 3 (SW620) biological replicates. The replicates were merged. (D, F) Genomic profiles of *AKT2* (D) and *GPI* (F), showing Total and T4ph RNAPII mNET-seq signal. The gray highlight indicates RNAPII pausing in 1CT cells (D) or RNAPII terminal pausing region in CRC cells (F). (G) Metagene profile of T4ph-mNET-seq signal on genes undergoing proximal or distal APA in the CRC progression model. (H) Genomic profile of *TBC1D20*, showing T4ph-mNET-seq and nuclear 3' mRNA-seq signals. The gray highlight indicates T4ph RNAPII pausing in 1CT cells or PAS usage. (I) Schematic representation of uncoupling of changes in PAS usage (yellow) and RNAPII termination (red) in the CRC progression model. Left: transition from normal cell to primary tumor. Right: transition from primary to metastatic tumor. The arrows indicate the direction of change.

active site of RNAPII shortly before termination, i.e., undergoing terminal pausing (Kopczyńska et al, 2024). The T4ph RNAPII mark and T4ph RNAPII mNET-seq is highly enriched at gene ends and only present in the gene body in regions undergoing premature termination. Total RNAPII mNET-seq signal within gene bodies is instead dominated by S2ph-modified elongating form of RNAPII, which is also substantially more abundant than T4ph RNAPII in human cells (Schüller et al, 2016). Consequently, total RNAPII mNET-seq does not efficiently detect premature (gene-body) transcription termination.

T4ph mNET-seq confirmed that transcription termination downstream of the PAS (full-length termination) was indeed shifted proximally in CRC cells, most prominently in primary tumor cells, but also detectable in SW620 cells (Figs. 3D,E,H and EV3F,G). Interestingly, full-length termination in metastatic cells shared characteristics of both normal and primary tumor cells (Figs. 3D,E,H and EV3F,G), suggesting a partial reversal of termination patterns to those observed in normal cells. In addition to shifted gene-end termination, we detected more T4ph-marked RNAPII within gene bodies, suggesting an increase in the incidence of premature transcription termination in CRC cell lines (especially from primary tumors) compared to 1CT cells (Figs. 3E,F and EV3B,C,G). This interpretation is strengthened by the observation of overlaps of gene-body RNAPII terminal pausing (T4ph mNET-seq) and a decrease of the nascent transcript signal (POINT-seq) for several genes (Fig. EV3D; Appendix Fig. S2), and in line with a previously published analysis performed globally (Kopczyńska et al, 2024).

Taken together, CRC progression in our model is associated with increased RNAPII terminal pausing in the gene body (a proxy of premature termination) and a proximal shift (acceleration) of full-length termination in cells from primary tumors compared to normal cells. Strikingly, cells from metastatic tumor display a partial reversal of both premature transcription termination and the onset of full-length termination, with a termination landscape in between those observed in normal and primary tumor-derived cells. At the same time, metastasis-derived SW620 cells show transcriptional readthrough further downstream than cells from primary colorectal epithelium or primary CRC-derived cells.

The proximal shift in transcription termination in cell lines from the primary tumor was particularly surprising—because the direction of changes was opposite to the tendency toward distal APA in those cell lines (Fig. 2D–H). To corroborate this and exclude the possibility that the proximal termination shift and distal APA occur mutually exclusively (i.e., on different subsets of genes), we checked transcription termination profiles on genes

across all APA categories. This revealed accelerated termination (T4ph mNET-seq) in CRC cells compared to 1CT cells, regardless of presence/absence or even the direction of APA in the analyzed gene category (Figs. 3G and EV3E). This accelerated transcription was visible as both increased gene-body terminal pausing (premature termination) and proximally shifted full-length termination. Inspection of individual genes confirmed this analysis—earlier full-length termination, yet distal PAS usage in cells derived from primary CRC, can occur on the same gene (Figs. 3H and EV3F). Admittedly, we also found examples where earlier transcription termination in primary CRC cells was accompanied by proximal PAS selection (Fig. EV3G). Collectively, the above results suggest that the direction of APA and changes in transcription termination are globally uncoupled during colorectal cancer progression (Fig. 3I).

Proximity of pre-mRNA 3' end cleavage and transcription termination is associated with upregulation of processed nuclear transcripts

The unexpected finding that changes in PAS selection and localization of RNAPII terminal pausing occur in the opposite direction during CRC progression prompted us to investigate their relationship more systematically. We quantified the distances between the genomic coordinate of the major PAS used for a given gene in each cell line, and the start coordinate of the T4ph RNAPII enrichment region (termination window) in each cell line (Fig. 4A). For simplicity, we refer to this measure as the cleavage-termination distance.

Our analysis revealed that the median cleavage-termination distance in primary tumor-derived cells dropped sharply, to roughly half of the distance observed in normal cells, and partially recovered in CRC metastasis-derived cells (Fig. 4A; Dataset EV1). We observed the same trend when calculating the distance between the cleavage site and the peak of the T4ph RNAPII signal (the highest signal in the termination window) (Appendix Fig. S3A), as well as in gene-wise comparison (Appendix Fig. S3B). We also noticed that the cleavage-termination distance tends to be particularly short for oncogenes, as compared to both active protein-coding genes and tumor suppressor genes, for all cell lines of our model (Figs. 4B and EV4A,B). This prompted us to identify the biological function of genes undergoing cleavage-termination distance shortening in CRC cells. Within the top ten enriched gene ontology terms, we found processes related to DNA replication, DNA damage and repair, and RNA processing (Fig. 4C; Dataset EV1).

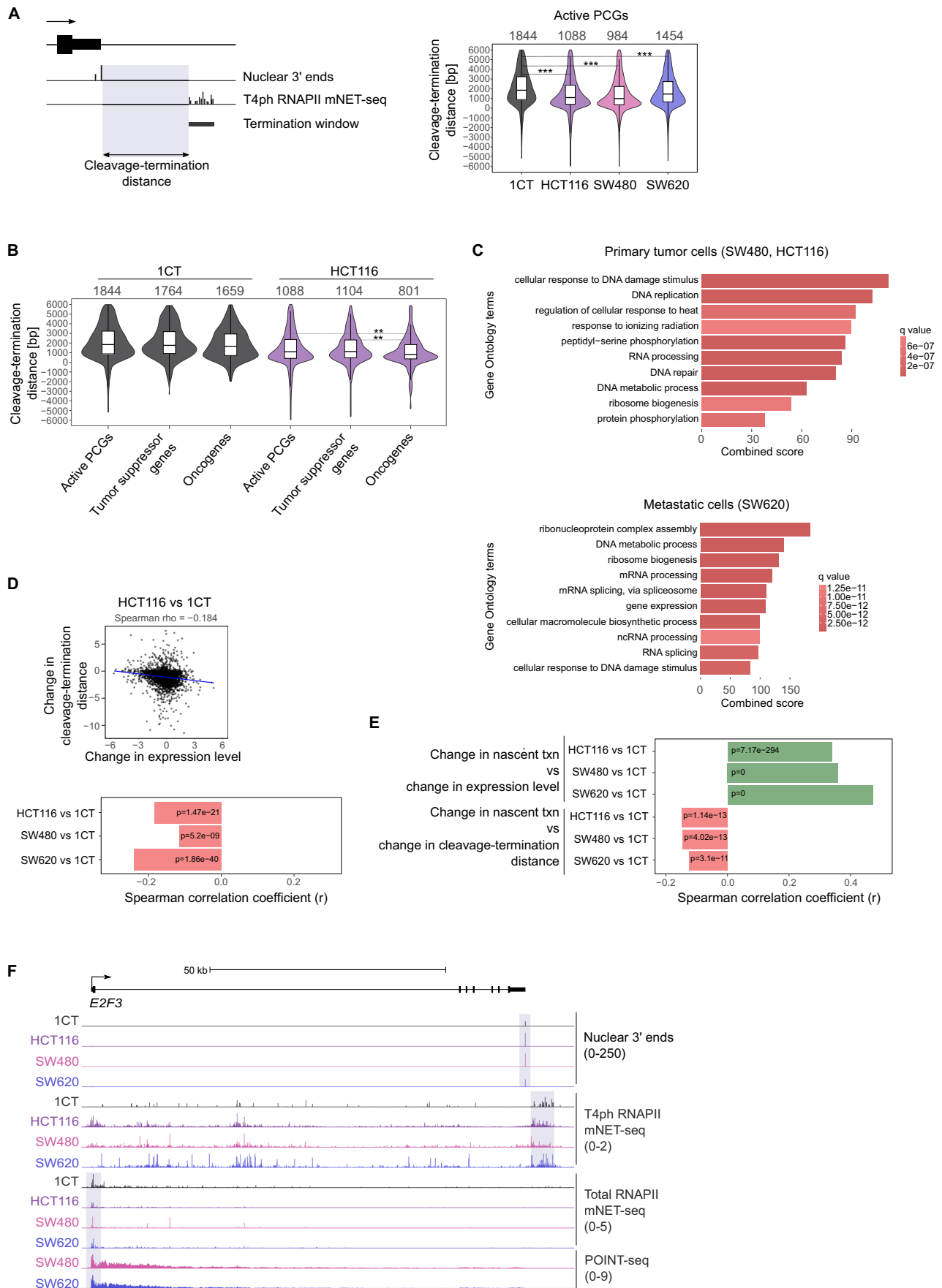


Figure 4. Closer proximity of PAS and transcription termination (onset of terminal pausing) is associated with higher gene expression.

(A) Schematic representation of the cleavage-termination distance measure. Violin plot representation of the cleavage-termination distance for active PCGs in CRC cells. (A–F) Based on merged biological replicates. $N = 3$ (nuclear 3' ends), 2–3 (T4ph mNET-seq), 2 (POINT-seq) biological replicates. (A, B) *** $P_{adj} < 0.001$; ** < 0.01 ; * < 0.05 . The pairwise Wilcoxon test exact p values are provided in Dataset EV1. Non-indicated differences are not significant. The numbers above the plots represent the median. The boxplots display the median (center line), first and third quartiles (lower and upper bounds of the box), and whiskers extending to the most extreme data points within $1.5 \times IQR$ of the lower and upper quartiles. (B) Violin plot representation of the cleavage-termination distance for active PCGs, tumor suppressor genes, and oncogenes, in 1CT and HCT116 cells. (C) Gene Ontology enrichment analysis of genes with shortened cleavage-termination distance in CRC tumor cells compared to 1CT cells. Bar plots show the top ten enriched GO Biological Process (BP) terms. The x-axis indicates the combined score, whereas the y-axis shows enriched GO-BP terms. Color represents the q value and the bar length indicates the combined score. Upper panel: primary tumor-associated genes (overlapping genes between HCT116 and SW480), lower: metastatic genes (SW620). (D) Pairwise correlations between changes in gene expression and the cleavage-termination distance in cancer cells compared to 1CT cells. The scatter plot comparing HCT116 and 1CT cells (upper panel) and the r and p values for all the comparisons performed (lower panel). Positive values indicate an increase in expression and cleavage-termination distance; negative values decrease. Each bar represents a pairwise comparison; its width indicates the Spearman correlation coefficient. Pink bars indicate negative correlation. (E) Pairwise correlations between changes in nascent transcription (txn), gene expression, and cleavage-termination distance in cancer cells compared to 1CT cells. Each bar represents a pairwise comparison; its width indicates the Spearman correlation coefficient. Pink bars indicate negative correlation, green bars indicate positive correlation. (F) Genomic profile of *E2F3*, showing nuclear 3' mRNA-seq, T4ph-mNET-seq, Total-mNET-seq, and POINT-seq signals. The gray highlight indicates PAS usage and RNAPII pausing in 1CT cells.

We wondered whether the cleavage-termination distance might influence gene expression. For the purpose of our study, we defined gene expression as the level of correctly processed transcripts measured by nuclear 3'mRNA sequencing (see Methods). All cell lines from our model displayed a slight but consistent tendency for lowered gene expression with increased cleavage-termination distance (Fig. EV4C, darker violin plots). A small subset of active genes (7–11% per cell line) showed a negative cleavage-termination distance—terminal RNAPII pausing could be observed on those genes already proximally to the main cleavage site (Fig. EV4C). We found negative cleavage-termination distance to be associated with lower gene expression (Fig. EV4C, lighter violin plots), but did not study this small subset further.

We then checked how changes in the cleavage-termination distance relate to changes in the expression of the same gene in our CRC model. For each of the CRC cell lines, an increase in cleavage-termination distance of a gene during progression from normal to cancer cell significantly anticorrelated with an increase in its expression (Figs. 4D and EV4D; Dataset EV1). In simple terms, shortening of the cleavage-termination distance was associated with increased expression of the gene.

We next wanted to understand the mechanism behind the increased expression of genes with shorter cleavage-termination distances. We hypothesized it could be driven by increased nascent transcription, more efficient RNA processing, or both. First, we checked how changes in nascent transcription (total RNAPII mNET-seq signal) of the gene impacted changes in gene expression (defined by nuclear 3'mRNA sequencing) in our CRC model, relative to 1CT. As expected, those two measures showed a highly significant ($p = 0$) correlation (Fig. 4E, top panel; Dataset EV1). However, the correlation coefficient was below 0.5, indicating that transcriptional changes were not solely or dominantly responsible for the changes in levels of correctly processed nuclear transcripts.

In line with this result, changes in nascent transcription of the gene significantly anticorrelated with changes in its cleavage-termination distance (Figs. 4E, bottom panel and EV4E; Dataset EV1), although with a smaller correlation coefficient compared to changes in gene expression (Fig. 4D; Dataset EV1). This implied that the increase in gene expression associated with shorter cleavage-termination distance is partially due to increased nascent transcription but also enhanced RNA processing. This interpretation is supported by the observation of individual oncogenes (*E2F3*, *CDK1*, and *EZH2*) in CRC compared to normal cells. They showed

shortened cleavage-termination distance and increased expression, yet without an increase in nascent transcription signal (Figs. 4F and EV4F,G), implying increased efficiency of RNA processing.

CRC progression is associated with a whole plethora of events, and hence, gene expression or cleavage-termination distance changes during CRC progression cannot be attributed to any one factor. To evaluate the relationship between cleavage-termination distance and gene expression (Fig. 5A) under more controlled conditions, we used a system which we previously characterized in depth. We triggered a downstream shift of the transcription termination window (termination delay) by depletion of PCF11 in the cervical cancer cell line HeLa (Kamieniarz-Gdula et al, 2019). Reassuringly, both in control and knockdown conditions, we observed the same trend as in CRC: the shorter the cleavage-termination distance, the higher the expression level (Fig. EV5A). Further on, PCF11 knockdown resulted in increased cleavage-termination distance (Fig. 5B; Dataset EV1). This lengthening of the cleavage-termination distance of a gene by PCF11 knockdown correlated significantly with its decreased expression level (Fig. 5C), and the correlation coefficient was similar to that observed in CRC progression (Fig. 4D; Dataset EV1). This finding was also corroborated on individual genes (Fig. 5D).

We noticed that the sensitivity of primary CRC cell lines to JTE-607 is associated with an unusually short cleavage-termination distance. To test whether the cleavage-termination distance might be predictive of responsiveness of a cell line to JTE-607, we used two pancreatic cancer cell lines previously shown to differ in their response to JTE-607 in a colony-formation assay (Alahmari et al, 2024). Specifically, Panc1 cells were more sensitive to JTE-607 than MIA PaCa-2 cells. We performed T4ph mNET-seq and nuclear 3' mRNA-seq, which allowed us to calculate their cleavage-termination distance. Higher JTE-607 sensitivity of Panc1 cells was associated with a proximal shift of full-length termination and shorter cleavage-termination distance (Figs. 5E–G and EV5B; Dataset EV1). This is consistent with our hypothesis, and thus encourages further investigation. Additionally, within each cell line, increased cleavage-termination distance was linked with lower gene expression (Fig. EV5C), similarly to the CRC model.

In summary, our data reveal that the distance between the site of 3' pre-mRNA cleavage and RNAPII entering into terminal pausing mode is halved in cells from primary CRC, compared to normal cells, and partially restored in cells from metastasis. Shorter cleavage-termination distance is associated with elevated levels of

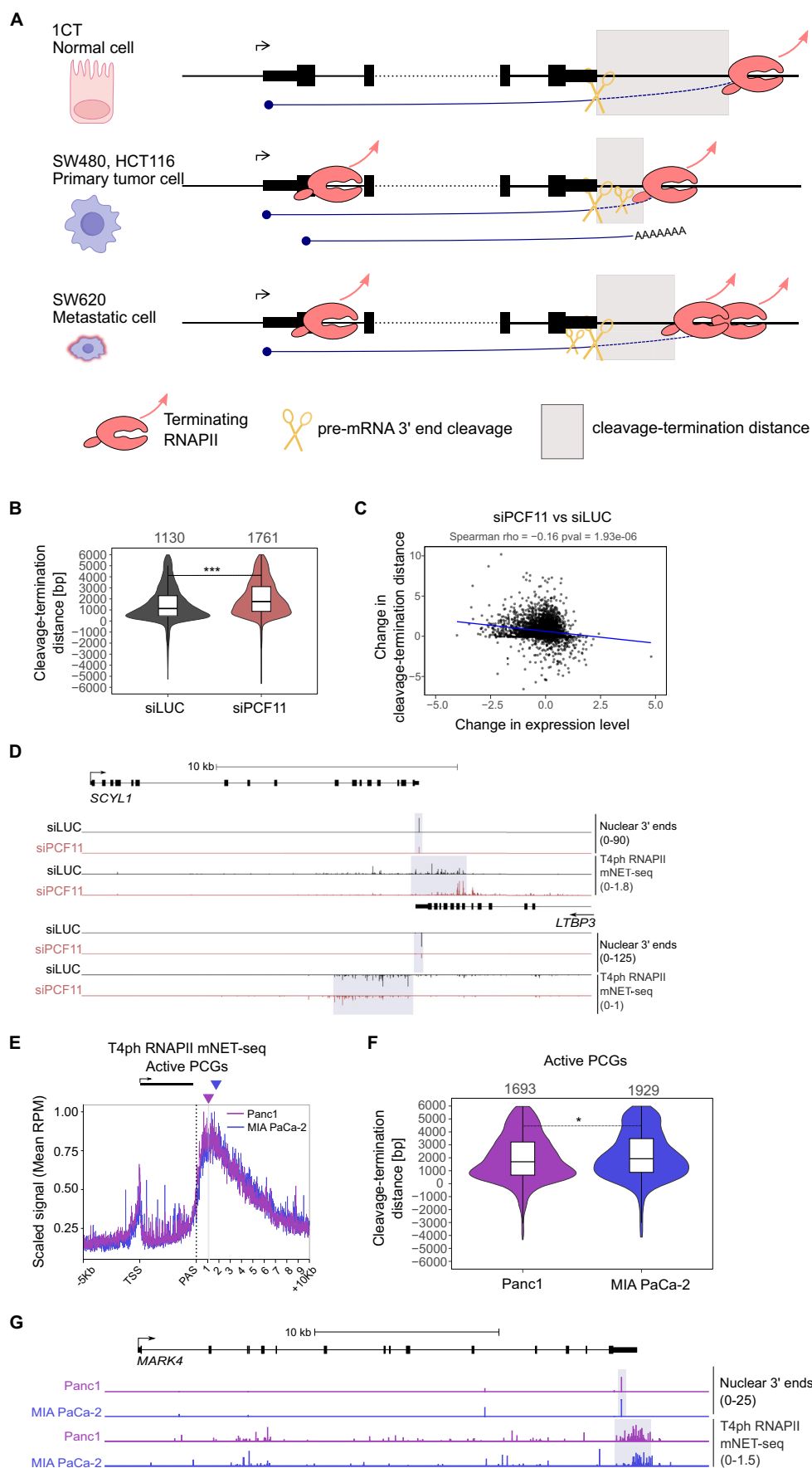


Figure 5. The relationship between cleavage-termination distance and gene expression extends beyond the CRC model.

(A) CRC progression is associated with more premature and proximal full-length transcription termination in primary tumors compared to normal cells. In metastatic cells, premature termination is still increased, but transcription termination downstream of PAS is partially reversed to the location in normal cells. Smaller scissors indicate APA direction: distal in primary tumor cells and proximal in metastatic cells. Primary tumor cells are characterized by a shorter cleavage-termination distance that is associated with a higher expression level. (B) Violin plot representation of the cleavage-termination distance for active PCGs in HeLa cells in control conditions (siLUC) or upon PCF11 knockdown (siPCF11). (B–D) $N = 3$ (T4ph mNETseq) or 4 (nuclear 3' ends) biological replicates. (B, F) $***P_{adj} < 0.001$; $** < 0.01$; $* < 0.05$. The pairwise Wilcoxon test exact p values are provided in Dataset EV1. The numbers above the plots represent the median. The boxplots display the median (center line), first and third quartiles (lower and upper bounds of the box), and whiskers extending to the most extreme data points within $1.5 \times IQR$ of the lower and upper quartiles. (C) Pairwise correlation between changes in gene expression and the cleavage-termination distance upon PCF11 knockdown in HeLa cells. Positive values indicate an increase in expression and cleavage-termination distance; negative values decrease. (D) Genomic profile of *SCYL1* and *LTBP3*, showing T4ph-mNET-seq and nuclear 3' mRNA-seq signals. The gray highlight indicates T4ph RNAPII pausing in control conditions (siLUC) cells or PAS usage. (E) Metagene profiles of T4ph RNAPII mNET-seq signal along active protein-coding genes (PCGs) ($n = 13,793$). The dotted line indicates the PAS position, and the gray highlight indicates the signal maximum downstream of PAS in Panc1 cells. Triangles depict the signal maxima downstream of PAS. (E–G) $N = 2$ (T4ph mNET-seq) or 3 (nuclear 3' ends) biological replicates. The replicates were merged. (F) Violin plot representation of the cleavage-termination distance for active PCGs in Panc1 and MIA PaCa-2 cells. (G) Genomic profile of *MARK4*, showing T4ph-mNET-seq and nuclear 3' mRNA-seq signals. The gray highlight indicates T4ph RNAPII pausing in Panc1 cells or PAS usage.

the corresponding nuclear mRNA, partially due to elevated nascent transcription, and in part to enhanced RNA processing. We speculate that spatial (and possibly also temporal) proximity of cleavage and termination could underline increased sensitivity of cancer cells to manipulations of CPA levels with respect to their colony-formation ability (Fig. 1).

Discussion

CPA and termination factors PCF11 and CPSF73 emerged recently as having oncogenic properties, with their high levels being predictors of poor outcome in several cancer types (Ogorodnikov et al, 2018; Salifu and Doughan, 2022; Alahmari et al, 2024). However, the molecular mechanisms involved are unclear. Our study revealed that CRC cells derived from the primary tumor or its metastasis differ in their sensitivity to alterations in PCF11 and CPSF73 levels and/or activity. While those alterations block colony formation from primary tumor cells, metastatic cells were less sensitive to CPA factor depletion (Figs. 1D and EV1D) and RNA 3' cleavage inhibition by JTE-607 (Fig. 1E). Since we observed similar knockdown efficiency for PCF11 in both cell lines (Fig. EV1D), the higher sensitivity of primary tumor-derived cells to PCF11 knockdown might be dictated by a lower number of PCF11 molecules per cell left after the treatment. However, the result of higher sensitivity of SW480 cells was obtained not only for PCF11 knockdown, but also for CPSF73 knockdown (Figs. 1D and EV1D), for which the baseline levels were similar and knockdown was equally efficient in both cell lines, as well as for CPSF73 inhibition (Fig. 1E).

A previous study found a strong correlation of gene expression changes in response to JTE-607 treatment with PCF11 knockdown, pointing to a common transcriptomic signature (Cui et al, 2023). The same study suggested that increased CPA factor levels lead to proximal PAS usage, which could sensitize cells to JTE-607 treatment. However, this hypothesis was based on the comparison of two unrelated cell lines (liver- and cervix-derived). Here, we analyzed four cell lines originating from the same tissue (colon) with a comprehensive set of genomic assays. Two of the analyzed cell lines—SW480 (from the primary tumor) and SW620 (from its metastasis)—were even isolated from the same patient. The level of CPA and termination factors increased in our colorectal progression model (Figs. 1C and EV1C). In contrast, changes in APA were

not progressive: while we observed more distal APA events in cells from the primary tumor compared to normal cells, metastatic cells were characterized by more proximal PAS events (Figs. 2D–H and EV2H–K). Since cells from primary tumors, with a tendency to distal APA, were more sensitive to JTE-607 and CPA downregulation, this contradicts the previous hypothesis (Cui et al, 2023) and demonstrates that shortening of mRNA does not by itself sensitize cells to CPA inhibition.

We additionally found that the clonogenic potential of CRC cells was more dependent on PCF11 than CPSF73, in both SW480 and SW620 cell lines (Figs. 1D and EV1D). This is consistent with PCF11 being a limiting and regulatory factor within the CPA complex (Kamieniarz-Gdula et al, 2019). Coincidentally, it is also the only factor known to be able to directly trigger termination (Zhang et al, 2005).

Our results further refuted the hypothesis that the high sensitivity of primary tumor cells to PCF11 or CPSF73 knockdown results from increased frequency of transcription-replication collisions (TRCs) (Figs. 2A–C and EV2A–G). We performed PCF11 KD, CPSF73 KD, acute CPSF73 depletion with degron, and CPSF73 inhibition, which all result in global transcription readthrough (Eaton et al, 2018; Kamieniarz-Gdula et al, 2019; Liu et al, 2023; Cui et al, 2023) but did not increase TRC frequency (Figs. 2B,C and EV2C–G). The only readthrough inducing manipulation which led to a modest increase in TRC frequency was 4 h of acute XRN2 depletion (Fig. EV2D). However, besides its role in termination, XRN2 resolves R-loops, which may block replication fork progression (Morales et al, 2016; Gómez-González and Aguilera, 2019).

Two studies reported replication stress upon transcription termination crisis. Depletion of elongation factor SPT6 led to increased levels of elongating RNAPII in intergenic regions and accumulation of R-loops and DNA damage presented by gamma-H2AX foci (Nojima et al, 2018). Knockdown of CPA factor WDR33 resulted in Ser2-RNAPII readthrough, transcription-dependent reduced replication fork speed, and more replication initiation events (Teloni et al, 2019). However, the influence of RNAPII readthrough, triggered by transcription termination factors loss, on TRC frequency was not measured before. Our data indicate that inducing transcription readthrough by altering CPA and transcription termination factor levels does not affect the number of TRCs in CRC cells. The unchanged TRC frequency in the presence of elongation-competent but termination-deficient RNAPII might be

connected with the ability of RNAPII to redistribute pre-replication complexes to transcriptionally inactive regions before the onset of replication (Liu et al, 2025b; Marchal et al, 2019).

It is worth noticing that we observed more TRC foci in SW480 compared to SW620 cells, in all conditions tested (Figs. 2A,B and EV2A,B,F). Since these cell lines were isolated from the same patient (Leibovitz et al), we speculate that in this patient, CRC cells evolved during metastasis to deal with DNA damage by limiting TRC frequency. Changes in TRC frequency upon cancer progression have not been measured before, so it would be interesting to check in future studies if downregulation of TRCs occurs during oncogenesis also in other types of cancers.

One of the most surprising findings in this study is that changes in pre-mRNA 3' processing and transcription termination in CRC progression were uncoupled and even occurred in the opposite direction (Fig. 3G). In comparison to normal cells, primary tumor cells had accelerated (proximal) transcription termination (Figs. 3A–H and EV3A,E–G). At the same time, distal cleavage sites were used more often, even on the same gene (Figs. 2E–G, 3H and EV3F,G). Then, during the progression of CRC from primary to metastatic tumor, the opposite was observed: termination was delayed, while APA changes tended to shift in the proximal direction.

The second key finding of our study is that changes in the cleavage-termination distance on a given gene correlated negatively with changes in its expression, measured by the level of its processed transcript in the nucleus (Figs. 4D and EV4D). Thus, shortening of the cleavage-termination distance is associated with increased gene expression. While correlation does not imply causality, it is important to note that this relationship was true not only in the CRC progression model (Figs. 4D and EV4D), but also upon a relatively short genetic manipulation (48 h knockdown of PCF11), which primarily delays transcription termination (Figs. 5B and EV5A). The increase in gene expression is partially due to an increase in nascent transcription (Fig. 4E, upper).

We think that increased nascent transcription could be due to accelerated recycling of RNAPII for subsequent rounds of transcription. Admittedly, as we used total-mNET-seq results as a measure of nascent transcription levels, these results might be influenced by RNAPII pausing levels. Changes that cannot be attributed to transcription, we reason, are due to more efficient pre-mRNA 3' processing. On some genes, enhanced processing efficiency likely plays a dominant role, as exemplified by the oncogenes shown in Figs. 4F and EV4F,G, where no increase in nascent transcription is visible.

In addition, the POINT-seq metagene analysis for SW480 and SW620 cells revealed a similar pattern of changes in the nascent transcription profile, i.e., reduced signal at transcription start site (TSS) and delayed termination in metastatic cells, resembling the nascent transcription profile observed in samples with reduced pre-mRNA cleavage efficiency (Figs. 3A,B and EV3A). This suggests a possible link between the cleavage-termination distance and cleavage efficiency and highlights an interesting direction for future studies.

The median cleavage-termination distance in cells from the primary tumor was shortened to about half in comparison with cells from healthy colon epithelium (Fig. 4A). We speculate that such a shortened opportunity window for cleavage and

transcription termination in primary tumor cells makes these cells more vulnerable to PCF11 or CPSF73 manipulation (Fig. 1D–F). Only half of genes containing multiple alternative PASs (and a third of all active genes), underwent any APA between 1CT and SW480 cells (Fig. 2E), yet termination shifted proximally in SW480 cells even if there was no APA (Fig. EV3E). This implies that the high sensitivity of primary tumor cells to CPA alterations is more likely to be caused by earlier termination (a robust and unidirectional shift), rather than by APA (which happens only on a subset of genes, and in two opposite directions, with just a slight tendency towards distal sites).

We would like to acknowledge the following limitations of our work. Firstly, we used a simplified cellular model of CRC progression, involving only four cell lines. Therefore, our conclusions cannot be extrapolated to CRC cell lines, which were not studied here, or directly to patients. Secondly, our study focused entirely on nuclear events, including nascent transcription and nuclear RNA processing, but provides no data regarding the cytoplasmic fate of transcripts, e.g., the effects of APA on mRNA stability or translation efficiency. Thirdly, our nascent transcriptomic analysis was based primarily on the mNET-seq methodology, which is sensitive to RNAPII pausing, which might potentially lead to biases.

Overall, our work demonstrates that the interplay between pre-mRNA 3' end cleavage and transcription termination is more complex than previously appreciated. It also suggests that the effect of altered transcription termination on cancer biology might be underestimated. So far, multiple studies have measured the extent of readthrough, which can be done with any nascent transcriptomics technique. Here, the use of termination-specific T4ph-mNET-seq (Kopczyńska et al, 2024) has put us in the unique position of being able to check the distance between the PAS and the onset of RNAPII terminal pausing. Interestingly, this distance negatively correlates with the expression of the gene and becomes critically short for oncogenes in primary CRC cells, which opens a new area of investigation. It can help to uncover the biogenesis and function of mRNA 3' ends and the role of transcription termination in cancer progression. We hope such molecular studies will aid in exploiting cancer's vulnerability to CPA and termination factor levels and developing new oncological drugs that will be used precisely in the cancer types and stages where they can be most effective.

Methods

Reagents and tools table

Reagent/resource	Reference or source	Identifier or catalog number
Experimental models		
1CT (<i>H. sapiens</i>)	Shazia Irshad's Lab	n/a
HCT116 (<i>H. sapiens</i>)	The Maria Skłodowska-Curie Greater Poland Cancer Centre	n/a
SW480 (<i>H. sapiens</i>)	Shazia Irshad's Lab	n/a
SW620 (<i>H. sapiens</i>)	Shazia Irshad's Lab	n/a

Reagent/resource	Reference or source	Identifier or catalog number
HCT116 CPSF73-AID (<i>H. sapiens</i>)	Steve West's Lab	n/a
T84 (<i>H. sapiens</i>)	Shazia Irshad's Lab	n/a
HeLa (<i>H. sapiens</i>)	Nick Proudfoot's Lab	n/a
Panc1	Argyris Papantonis's Lab	n/a
MiaPaCa2	Argyris Papantonis's Lab	n/a
BxPC3	Argyris Papantonis's Lab	n/a
Recombinant DNA		
n/a		
Antibodies		
Rabbit anti-PCF11	Abcam	Cat# ab134391; RRID: AB_2783786
Rabbit anti-CPSF73	Abcam	Cat# ab72295, RRID:AB_1268249
Rabbit anti-XRN2	Thermo Fisher Scientific	Cat# A301-103A, RRID:AB_2218876
Mouse monoclonal anti-RNAPII total	Stasevich et al. 2014	Cat# CMA601; RRID: AB_2827956
Rat monoclonal anti-RNAPII T4ph	Active Motif	Cat# 61361, RRID:AB_2750848
Rabbit monoclonal anti-RNAPII CTD Ser2	Abcam	Cat# ab193468; RRID:AB_2905557
Mouse anti-PCNA	Santa Cruz Biotechnology	Cat# sc-56, RRID:AB_628110
Goat anti-rabbit IgG –Peroxidase	MERCK	Cat# A0545; RRID:AB_257896
Goat anti-rabbit IgG – LI-COR IRDye 800CW	LI-COR	Cat# 926-32211; RRID:AB_621843
Oligonucleotides and other sequence-based reagents		
siLUC	MERCK	Sequence (5'-3') Sense: GAUUAUGUCCGGUUAU GUAUU Antisense: [phos] UACAUAACCGGA CA UAAUCUU
siPCF11 (human) ON-TARGETplus SMARTpool	Dharmacon	Cat# L-015381-01
siCPSF73 (human) ON-TARGETplus SMARTpool	Dharmacon	Cat# L-006365-01
Chemicals, enzymes and other reagents		
JTE-607	MERCK	Cat# SML2833
Human EGF	Thermo Fisher Scientific	Cat# PHG6045
Hydrocortisone	MERCK	Cat# H0888-1G
Apo-Transferrin	MERCK	Cat# T2036-100MG
Lipofectamine™ RNAiMAX Transfection Reagent	Thermo Fisher Scientific	Cat# 13778150

Reagent/resource	Reference or source	Identifier or catalog number
7.5% and 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels	Bio-Rad	Cat# 4561026, 4561086
Novex 6% TBE gel, 12-well	Thermo Fisher Scientific	Cat# EC62652BOX
Novex 6% TBE-Urea (TBU) gel, 12-well	Thermo Fisher Scientific	Cat# EC68652BOX
T4 polynucleotide kinase (PNK), 3' phosphatase minus	NEB	Cat# M02365
TRI Reagent™ Solution	Thermo Fisher Scientific	Cat# AM9738
MNase	NEB	Cat# M02475
TURBO Dnase	Thermo Fisher Scientific	Cat# AM2238
Empigen BB	MERCK	Cat# 30326
RiboLock	Thermo Fisher Scientific	Cat# EO0382
NP-40	Thermo Fisher Scientific	Cat# 28324
PhosSTOP	MERCK	Cat# 4906837001
Protease inhibitors	MERCK	Cat# 04693159001
Sheep anti-mouse IgG Dynabeads	Thermo Fisher Scientific	Cat# 11202D
Protein G Dynabeads	Thermo Fisher Scientific	Cat# 10004D
Software		
FastQC (v0.11.9)	Simon Andrews https://github.com/sandrews/FastQC	
TrimGalore (v0.6.6)	Felix Krueger https://github.com/FelixKrueger/TrimGalore	
STAR (v2.7.6a)	https://github.com/alexdobin/STAR Dobin et al, 2013	
Sambamba (v1.0.1)	https://github.com/biod/sambamba Tarasov et al, 2015	
SAMtools (v1.10)	https://github.com/samtools/samtools Danecek et al 2021	
deepTools (v3.5.1)	https://github.com/deeptools/deepTools Ramirez et al 2016	
MACS2 (v2.2.7.1)	https://github.com/macs3-project/MACS Zhang et al, 2008	
HOMER (v4.9.1.6)	http://homer.ucsd.edu/homer/ Heinz et al, 2010	
ggplot2 (v3.4.4)	https://cran.rproject.org/web/packages/ggplot2/index.html ; https://ggplot2.tidyverse.org/ Wickham, 2016	
DESeq2 (v1.40.2)		

Reagent/resource	Reference or source	Identifier or catalog number
	https://bioconductor.org/packages/release/bioc/html/DESeq2.html Love et al, 2014	
DEXseq (v1.46.0)	https://bioconductor.org/packages/DEXSeq Anders et al, 2012	
GENCODE release 43	https://www.gencodegenes.org/ Frankish et al, 2023	
Bioconductor (v3.17)	https://www.bioconductor.org/ Gentleman et al, 2004	
bedtools (v2.27.1)	https://bedtools.readthedocs.io/en/latest/	
bcl2fastq (v2.20.0.422)	https://emea.support.illumina.com/sequencing/sequencing_software/bcl2fastqconversion-software.html	
kentUtilis (v2.9)	https://github.com/ENCODE-DCC/kentUtils	
All the above software and packages were used on Linux (Ubuntu 20.04.6) mostly within the R (4.4.2) environment.		
Other		
NEBNext Small RNA Library Prep Set for Illumina	NEB	Cat# E7330
NEBNext® Ultra™ II RNA Library Prep Kit for Illumina	NEB	Cat# E7770
QuantSeq 3' mRNA-Seq Library Prep Kit REV	Lexogen	Cat# 016.96
Direct-zol RNA Micro prep Kit	Zymo Research	Cat# R2062
NextSeq High-Output Kit v2.5, 75 cycles	Illumina	Cat# 20024906
Duolink In Situ Detection Reagents FarRed kit	MERCK	Cat# DUO92013

Cell culture

1CT cells were cultured in DMEM High Glucose + Medium 199 (4:1), 2% FBS, EGF (25 µg/mL, Thermo Fisher Scientific, PHG6045), hydrocortisone (1 µg/mL, MERCK, H0888), insulin (10 µg/mL, MERCK, I3536), transferrin (2 µg/mL, MERCK, T2036), sodium selenite (5 nM, MERCK, S5261), gentamicin sulfate (50 µg/mL, MERCK, G1397) on Corning Primaria 100 mm dish at 37 °C, 7% CO₂, 3% O₂.

HCT116, SW480, SW620, MIA PaCa-2, Panc1, and HeLa cells were cultured in DMEM High Glucose, 10% FBS, penicillin/streptomycin in 5% CO₂ at 37 °C.

T84 cells were cultured in DMEM-F12 GLUTAMAX, 5% FBS, penicillin/streptomycin in 5% CO₂ at 37 °C.

All cell lines were negative for mycoplasma.

Colony-formation assay

Cells were split into six-well dishes, 1500 cells (SW480, SW620) or 1000 cells (HCT116), or 10,000 cells (T84) per well. For analysis of knockdown of PCF11 or CPSF73, cells were transfected with Lipofectamine™ RNAiMAX Transfection Reagent (Thermo Fisher Scientific, 13778150) and siRNAs (siLUC, siPCF11, siCPSF73, Dharmacon) at 15 nM concentration. For JTE-607 treatment, cells were split at 1500 cells per well and ~24 h later were treated with either DMSO or various concentrations of JTE-607 (MERCK, SML2833). Then, the cell culture continued for 10 days (SW480, SW620) or 7 days (HCT116, T84). After that, the medium was discarded, and wells were rinsed once with PBS, stained with 500 µL of Crystal Violet (MERCK; C0775) solution per well, and incubated for 2 min at room temperature. Then, Crystal Violet was aspirated, and wells were washed four times with 2 mL of water. Once the residual liquid dried, pictures were taken with the Odyssey® M Imaging System. Pictures were analyzed with ImageJ. Three biological replicates were performed.

Immunoblotting

10–40 µg of proteins were resolved by electrophoresis using 7.5% or 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad) and transferred onto nitrocellulose membranes. Blots were blocked with 4% non-fat milk in PBST (PBS, 0.1% Tween-20), incubated with the following primary antibodies: anti-PCF11 (Abcam, ab134391, 1:200; RRID: AB_2783786); anti-CPSF73 (Abcam, ab72295, 1:200; RRID:AB_1268249); anti-XRN2 (Thermo Fisher Scientific, A301-103A, 1:200; RRID:AB_2218876), and anti-rabbit IgG—IgG—IRDye 800CW (LI-COR, 926-32211; RRID:AB_621843) or anti-rabbit IgG—Peroxidase (MERCK, A0545; RRID:AB_257896). Three biological replicates were performed.

mNET-seq

100 µL of sheep anti-mouse IgG Dynabeads (Thermo Fisher Scientific, 11202D) for total RNAPII (Stasevich et al. 2014), CMA601; RRID: AB_2827956) or 50 µL protein G Dynabeads (Thermo Fisher Scientific, 10004D) for T4ph RNAPII (Acitve Motif, 61361; RRID:AB_2750848) per one sample were washed 2x with ice-cold NET2+Empigen buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.05% NP-40, 1% Empigen BB), mixed with 5 µg of the antibody, incubated O/N in cold room, washed 2x with ice-cold NET2+Empigen buffer and put on ice.

1CT cells on 14 Primaria 100 mm dishes or cancer cells on one 150 mm dish were washed with ice-cold PBS, harvested by scraping on ice, moved to a 10 mL NUNC tube, and centrifuged at 500×g for 5 min at 4 °C. The cell pellets were then resuspended in 4 mL of ice-cold HLB + N buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 2.5 mM MgCl₂, and 0.5% NP-40) and lysed for 6 min on ice. The cell lysates were underlaid with 1 mL ice-cold HLB + NS buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 2.5 mM MgCl₂, 0.5% NP-40, and 10% Sucrose) and centrifuged at 800×g for 5 min at 4 °C. The nuclear pellets were resuspended in 120 µL of ice-cold NUN1 buffer

(20 mM Tris, pH 7.9, 75 mM NaCl, 0.5 mM EDTA, and 50% Glycerol). The nuclear lysates were transferred to new tubes, mixed with 1.2 mL of ice-cold NUN2 (20 mM HEPES-KOH, pH 7.6, 300 mM NaCl, 0.2 mM EDTA, 7.5 mM MgCl₂, 1% NP-40, and 1 M Urea, PhosSTOP (MERCK, 4906837001), protease inhibitors (MERCK, 04693159001)), vortexed at maximum speed and incubated on ice for 10 min (sample vortexed every 3–4 min). The chromatin was then centrifuged at 4000×g for 1 min at 4 °C, washed with 500 µL of ice-cold PBS, 100 µL of ice-cold 1x MNase buffer (NEB, M0247) and digested in 100 µL of MNase (40 units/µL) reaction buffer for 2–10 min at 37 °C in a thermomixer (1200 rpm). The digestion reaction was stopped by the addition of 12.5 µL of 0.2 M EGTA and centrifuged at 16,000×g for 5 min at 4 °C. The 100 µL of supernatant was diluted 10x with 900 µL of the ice-cold NET2+Empigen buffer, and RNAPII conjugated beads were added and incubated for 1 h in a cold room. The beads were then washed six times with 1 mL of the ice-cold NET2+Empigen buffer and 50 µL of the ice-cold PNKT buffer (1 x T4 PNK buffer, NEB, M0236S, and 0.1% Tween-20). Washed beads were incubated in 50 µL PNK reaction mix (1 x PNKT, 1 mM ATP, 0.05 U/ml T4 PNK 3' phosphatase minus (NEB, M0236S) in Thermomixer (1200 rpm) at 37 °C for 5 min, and then washed with the NET2+Empigen buffer. Next, RNA was isolated from beads by the addition of 1 ml of TRI Reagent™ Solution (Thermo Fisher, AM9738), chloroform purification, and precipitated with isopropanol. The RNA was resuspended in 10 µL of Urea dye (7 M Urea, 0.05% Xylene cyanol, and 0.05% Bromophenol blue), denatured and resolved on 6% TBE-Urea gel (Thermo Fisher Scientific, EC68652BOX) at 180 V for 15–18 min. In order to size-select 25–100 nt RNA fragments, a gel fragment was cut between BPB and XC dye markers. A 0.5 mL tube was prepared with four to five small holes made with a 25G needle and placed in a 1.5 mL tube. Gel fragments were placed in the layered tube and broken down by centrifugation at 16,000×g for 1 min. The RNA fragments were eluted from the gel using the RNA elution buffer (1 M NaOAc and 1 mM EDTA) at 25 °C for 1 h in a Thermomixer (1200 rpm). The slurry was put into a SpinX column (Costar, Corning, 8160) with two glass filters (Whatman, WHA1823010) and the flow-through RNA was ethanol-precipitated. mNET-seq libraries were prepared using NEBNext Small RNA Library Prep Set for Illumina (NEB, E7330L). About 15–18 cycles of PCR were used to amplify the libraries. Before sequencing, the libraries were size-selected on a 6% TBE gel (Thermo Fisher Scientific, EC62652BOX), selecting only the 150–230 bp PCR product to exclude primer dimers. Gel elution was performed as described above. The libraries were sequenced on NextSeq500 using NextSeq High-Output Kit, 75 cycles (Illumina) or NovaSeq 6000, or Novaseq X platform (Novogene). Two to three biological replicates were performed for each cell line.

3' mRNA seq

The nuclei isolation was performed as for mNETseq experiments. The nuclear pellets were washed with 1 ml of ice-cold PBS and suspended in 500 µL of TRI Reagent™ Solution (Thermo Fisher Scientific, AM9738), and RNA was purified and ethanol-precipitated according to the manufacturer's instructions. DNA was removed by incubation with 4U of TURBO Dnase (Thermo Fisher Scientific, AM2238) at 37 °C for 30 min and another round

of TRI Reagent™ Solution-based purification. The RNA integrity was monitored by Agilent TapeStation. The libraries were prepared using Lexogen QuantSeq 3' mRNA-Seq Library Prep Kit REV according to the manufacturer's protocol. 500 ng of RNA was used as an input. About 12–13 cycles of PCR were used to amplify the libraries. The library size was controlled by Agilent TapeStation. Indexed libraries were quantified, pooled for sequencing, and sequenced on Illumina NovaSeq 6000 or on AVITI High Output (Lexogen). Three biological replicates were performed.

POINT-seq

POINT-seq was performed as previously described (Sousa-Luís et al, 2021) with minor modifications. 200 µL of sheep anti-mouse IgG Dynabeads (Thermo Fisher Scientific, 11202D) for total RNAPII antibody (Hiroshi Kimura, CMA601; RRID: AB_2827956) per one sample were washed 2x with ice-cold NET2+Empigen 2 buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 0.05% NP-40, and 3% Empigen BB), mixed with 10 µg of the antibody, incubated O/N in a cold room, washed 2x with ice-cold NET2 buffer and put on ice.

Twenty million cells were washed with ice-cold PBS, harvested by scraping on ice, moved to a 10 mL NUNC tube, and centrifuged at 500×g for 5 min at 4 °C. The cell pellets were then resuspended in 4 mL of ice-cold HLB + N buffer (10 mM Tris pH 7.5, 10 mM NaCl, 2.5 mM MgCl₂, and 0.5% NP-40) and lysed for 6 min on ice. The cell lysates were underlaid with 1 mL ice-cold HLB + NS buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 2.5 mM MgCl₂, 0.5% NP-40, and 10% Sucrose) and centrifuged at 420×g for 5 min at 4 °C. The nuclear pellets were resuspended in 200 µL of ice-cold NUN1 buffer (20 mM Tris, pH 7.9, 75 mM NaCl, 0.5 mM EDTA, and 50% Glycerol) and mixed with 2 mL of ice-cold NUN2E (20 mM HEPES-KOH, pH 7.6, 300 mM NaCl, 0.2 mM EDTA, 7.5 mM MgCl₂, 1% NP-40, 1 M urea, 3% Empigen BB, PhosSTOP (MERCK, 4906837001), protease inhibitors (MERCK, 04693159001); Empigen BB, PhosSTOP, and protease inhibitors were added freshly each time). Immediately after adding NUN2E, the samples were mixed by inverting and incubated in ice for 5 min. The chromatin string was washed with 10 mL PBS and digested in DNase mixture (10 mM Tris-HCl, pH 7.5, 400 mM NaCl, 10 mM MnCl₂, 2 units/µL RiboLock, and 0.2 units/µL Turbo DNase) at 37 °C for 18 min at 1200 rpm in the Thermomixer (Eppendorf). The supernatant was collected by high-speed centrifugation (14,000×g, 10 min at 4 °C). A fivefold volume of ice-cold NET-2E buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.05% NP-40, and 3% Empigen BB) was used for the dilution of the supernatant, and RNAPII conjugated beads were added and incubated for 1.5 h in a cold room. After the incubation, beads were washed with 1 mL of ice-cold NET-2E buffer six times. Next, RNA was isolated from beads by the addition of 1 ml of TRI Reagent™ Solution (Thermo Fisher, AM9738), chloroform purification, precipitated with isopropanol, treated with Turbo DNase once at 37 °C for 15 min and purified using Direct-zol Micro prep kit following the manufacturer's protocol. The POINTs were fragmented to 150–300 nucleotides for sequencing according to the protocol of the NEBNext Ultra II Directional RNA library prep kit. The POINT-seq libraries were sequenced in the Illumina PE150 NovaSeq X platform (Novogene).

Transcription-replication collisions

All PLA reactions were performed as described in (Lalonde et al, 2024) with a few modifications. Briefly, cells were seeded, 30,000 per well, in 96-well culture plates (Ibidi, 89626), grown as described previously and treated the following day.

For the knock-down experiment, the cells were transfected according to a reversed transfection protocol. Final concentration of siRNA in the transfection reaction reached 15 nM and 5 μ L of Lipofectamine RNAiMAX (Thermo Fisher Scientific, 13778150).

To inhibit CPSF73, cells were treated with JTE-607 20 or 50 μ M for 3 or 7 h, respectively.

For CPSF73 depletion, cells were pretreated with doxycycline (1 μ g/ml, MERCK, D9891) for 18 h and then treated with 1 μ M IAA (MERCK, SML3574; multiple timepoints). For XRN2 depletion, cells were treated with 1 μ M IAA (multiple timepoints).

Before fixation, the media was supplemented with 10 μ M EdU for 30 min for S phase identification. Immediately after, cells were washed 2x with PBS and preextracted for 2 min at RT using PBS + 0.5% Triton X-100. 1CT cells were preextracted for 10 min. After pre-extraction, the cells were fixed with 4% PFA for 15 min at RT, washed 3x with PBS, permeabilized for 1 h with PBS + 0.5% Triton X-100, washed 3x with PBS, and then blocked with a PBS + 5% BSA solution for overnight at 4 °C.

For S phase identification, EdU was labeled with AlexaFluor 488 Azide using a Click-it chemistry reaction as described in (Lalonde et al, 2024). Then, cells were washed 2x with PBS and blocked for 1 h in PBS + 5% BSA at RT. PLA reactions were performed following the manufacturer's protocol (Merck) using the Duolink In Situ Detection Reagents FarRed kit (MERCK, DUO92013) and a mouse anti-PCNA antibody (Santa Cruz Biotechnology, sc-56,1200; RRID:AB_628110) in combination with either a rabbit anti-(phospho-Ser2) RNA polymerase II antibody (Abcam, ab193468, 1:500; RRID:AB_2905557) or a rabbit anti-(phospho-Thr4) RNA polymerase II antibody (Novus biologicals, NBP1-49546, 1:100; RRID:AB_10011602) as primary antibodies. Lastly, a DAPI staining was performed by incubating the cells in a PBS + 5 μ g/mL DAPI for 1 h at RT. Cells were then washed 3x with PBS, and 400 μ L of PBS was left in the wells.

134 Z-stack images (11 μ m, 1 image per μ m) per well were acquired using a Nikon T2 inverted microscope with an Andor Dragonfly spinning disk, combined with a Plan Apochromat 40X air objective with a numerical aperture of 0.95, a 2.0X magnification lens, and an iXon Life 888 EMCCD camera. Images were analyzed using ImageJ and an automated image analysis pipeline as described in (Lalonde et al, 2024). Briefly, the Z-plane, being the most centered on nuclei, is selected from the stack by identifying the highest DAPI variance plane. Then, the camera background signal is subtracted, nuclei are segmented using the DAPI signal, the Stardist Plugin, and a trained segmentation model (Schmidt et al, 2018; Weigert et al, 2018), and finally, fluorescence signals are measured, and PLA foci are counted using the ImageJ Maxima Finder tool. Two to six biological replicates were performed for each experiment.

Because cell confluency and local cell density can influence PLA efficiency by altering antibody and enzyme accessibility to nuclear targets, only conditions cultured within the same 96-well plate, displaying comparable confluences, and processed simultaneously were directly compared, thereby minimizing technical bias. PLA

foci were quantified using a standardized nuclear segmentation and intensity-based spot detection pipeline restricted to the nuclear compartment. Quantification parameters were adapted to the high-density PLA signal and visually validated across conditions. As assay controls, PLA foci counts in EdU-positive cells were compared with EdU-negative cells within the same fields of view, confirming enrichment of PLA signal in replicating cells and demonstrating that PCNA-associated signal was above background levels (Fig. EV2A). In addition, total nuclear PLA signal intensity per cell was used as an orthogonal internal control for foci quantification consistency (Fig. EV2B).

Data analysis

Genomic annotation and analyzed gene sets

Hg38/GRCh38 was used as a reference genome for mNET, POINT and 3'mRNA-Seq analyses with basic gene annotation based on GENCODE release 43 (Frankish et al, 2023). For most analyses, we selected protein-coding genes (PCGs) with at least three reads in the 1CT 3'mRNA data (see section "3' mRNA-Seq") and named them "active PCGs" ($n = 13,577$, Dataset EV1). Using these genes, we customized annotations by assigning the TSS of the gene based on GENCODE coordinates, while the end coordinate corresponded to the experimentally detected PAS in 1CT 3'mRNA data. For genes with multiple PAS, we chose the end coordinate as the major PAS, i.e., the PAS with the highest read count. This annotation was further used to extract subsets and generate metagene profiles.

For APA analysis, genes from the active PCGs gene set that did not overlap with another annotated gene on the same strand, and had a 3' end isolated by at least 6 kb from the downstream annotated gene on the same strand were selected. This resulted in 12,649 PCGs, referred to as "separated active PCGs" (Dataset EV1).

Carcinogenesis-related genes, including general oncogenes and tumor suppressor genes, were obtained from the oncogene (ONGene) (<https://bioinfo-minzhao.org/ongene/>) (Liu et al, 2017) and tumor suppressor gene (TSGene) databases (<https://bioinfo.uth.edu/TSGene/>) (Zhao et al, 2013, 2016), respectively, yielding 679 oncogenes and 986 tumor suppressor PCGs. Further filtering for APA analysis within the active separated PCGs set identified 533 onco- and 764 tumor suppressor PCGs (Fig. 3D; Figs. S3C and S5A,B).

The active PCGs and separated active PCG gene sets were similarly generated using the MIA PaCa-2 pancreatic 3'mRNA data and were subsequently used for pancreatic-specific data analyses.

3' mRNA-seq

Colorectal 3' mRNA data sequenced on NovaSeq 6000 were converted to FastQ files using Illumina software bcl2fastq (v2.20.0.422), whereas pancreatic 3'mRNA demultiplexed data generated on Aviti were provided directly. The files were then processed based on the specifications provided by Lexogen (https://github.com/Lexogen-Tools/quantseqpool_analysis). Aligned BAM files were then separated based on strand using SAMtools (v1.10). Further downstream analysis (below) was adapted from the previously published workflow for 3' mRNA-seq analysis (Kamieniarz-Gdula et al, 2019).

For PAS calls and APA analysis (below), we have adapted previously published workflows (Derti et al, 2012; Fontes et al, 2017; Rot et al, 2017). The re-analyzed data listed in the Table 1.

Table 1. List of sample accession IDs corresponding to the re-analyzed data.

Sample	Sample accession Id	GEO accession Id	Condition
T4ph_mNETseq_siLUC_rep1	GSM3495849	GSE123104	Control
T4ph_mNETseq_siLUC_rep2	GSM3495850	GSE123104	Control
T4ph_mNETseq_siLUC_rep3	GSM3495851	GSE123104	Control
T4ph_mNETseq_siPCF11_rep1	GSM3495852	GSE123104	KnockDown
T4ph_mNETseq_siPCF11_rep2	GSM3495853	GSE123104	KnockDown
T4ph_mNETseq_siPCF11_rep3	GSM3495854	GSE123104	KnockDown
3'mRNAseq_siLUC_rep1	GSM3536477	GSE123105	Control
3'mRNAseq_siLUC_rep2	GSM3536478	GSE123105	Control
3'mRNAseq_siLUC_rep3	GSM3536479	GSE123105	Control
3'mRNAseq_siLUC_rep4	GSM3536480	GSE123105	Control
3'mRNAseq_siPCF11_rep1	GSM3536481	GSE123105	KnockDown
3'mRNAseq_siPCF11_rep2	GSM3536482	GSE123105	KnockDown
3'mRNAseq_siPCF11_rep3	GSM3536483	GSE123105	KnockDown
3'mRNAseq_siPCF11_rep4	GSM3536484	GSE123105	KnockDown

In addition to the datasets generated in this study, we also re-analyzed previously deposited data. HeLa_PCF11 knockdown T4ph mNET-Seq and 3'mRNA-Seq are part of the superseries [GSE123105](#). The accession IDs for these samples are provided in Table 1.

Filtering internal priming events. To minimize false positives caused by internal priming of the QuantSeq assay on A/T-rich genomic regions, we generated a basic genomic mask. This mask flags the areas in the genome with sequences containing six or more consecutive A's (forward strand) or T's (reverse strand), as well as any 10-nucleotide windows with more than six A's or T's. Next, to ensure that genuine polyadenylation sites (PAS) falling in A/T-rich regions were not included in the mask, we refined the mask by unmasking a 20-nucleotide window around/centered at each gene's 3' end (GENCODE v43) as well as experimentally validated PAS sites from previous studies (Derti et al, 2012). Aligned 3' mRNA-seq reads that overlap with the strand-specific refined masks were filtered out, and the remaining reads from the most distal nucleotide were retained for downstream analysis.

Polyadenylation site (PAS) detection and quantification. To identify polyadenylation sites (PAS) accurately, we used the filtered 3' mRNA-seq reads (from the step described above) to create strand-specific, depth-normalized density profiles across the genome. These profiles were generated by summing up coverage signals from all 3' mRNA-seq samples separately for each strand. Based on these profiles, PAS events were detected as follows: First, all signals within a 30-nucleotide window were summed up, and the windows with values above 30 were considered further. Within each window, the peak with the strongest signal was identified, and a new 30-nucleotide region was created around it. This step was repeated to eliminate overlapping intervals, yielding a set of reliable polyadenylation sites that were used to retrieve polyadenylation sites for each sample and to analyze the usage of alternative polyadenylation sites (APA).

Alternative polyadenylation (APA) analysis. For APA analysis, DEXSeq (Anders et al, 2012; Reyes et al, 2013) was employed along

with the previously published workflow for APA analysis (Kamieniarz-Gdula et al, 2019). Protein-coding genes that were considered active in 1CT 3'mRNA data, non-overlapping with any other gene on the same strand ($n = 12,649$; see "genomic annotation and analyzed gene sets", Dataset EV1) were subject to differential PAS usage quantification with DEXSeq. Gene coordinates were extended by 6 kb downstream of the 3' end (1CT major PAS) to allow for the detection of distal APA beyond gene ends (1CT major PAS). DEXSeq returns the fold change ($\log_2$) of the experimental sample over the control and adjusted p value for each PAS. Genes where no PAS reaches significance ($p_{\text{adj}} \geq 0.05$) were categorized as APA no shift. To assign the direction of APA shift in genes with multiple PAS, the two most statistically differentially used PAS were retrieved ($p_{\text{adj}} < 0.05$). If the ratio of the distal to proximal site usage was higher in the experimental than in the control cells, the shift in the experimental sample was classified as distal, in the opposite situation as proximal.

For pancreatic cells, MIA PaCa-2 was used as a reference dataset against which differential PAS usage was assessed.

Differential expression analysis. Gene expression was quantified by summing nuclear 3' mRNA-seq reads mapping to polyadenylation sites (PAS) within each gene's 3'UTR, extended 6 kb downstream (to capture potential readthrough transcripts). These counts were then subjected to differential expression analysis with DESeq2 (Love et al, 2014). DESeq2 outputs normalized counts for both experimental and control samples, along with the fold change ($\log_2$) of the experimental sample over the control and an adjusted p value for each gene. These values were further used for correlation analysis.

mNET-seq

For mNET-seq analysis, both T4ph and total paired-end reads in fastq files were first subjected to quality control (FastQC v0.11.9), and then the adapter sequences were trimmed using TrimGalore (v0.6.6). Trimmed reads were aligned to the human genome using STAR (Dobin et al, 2013) (v2.7.6a), only allowing for one alignment to the reference. Strand-specific BAMs were generated using SAMtools (v1.10), and the replicates were merged into a single BAM using the same. To obtain single-nucleotide resolution of strand-specific BAMs, the last transcribed nucleotide position from each aligned read was retrieved, and RPM normalized using an in-house R script based on Bioconductor libraries. The re-analyzed data listed in the Table 1.

Cleavage-termination distance analysis

To quantify cleavage-termination distances, we first identified termination windows—genomic regions enriched for T4ph-specific mNET-seq signal—using the MACS2 (v2.2.7.1) bdgbroadcall function on strand-specific bedGraph files with default parameters. These windows were then used to calculate the distance between each gene's major PAS (gene end) and the start of the termination signal—cleavage-termination distance. We excluded the windows that began within the gene body and retained only those windows starting at the 3'UTR and within 6 kb of a gene's major PAS. This resulted in two categories of distances (1) Negative—when the termination window begins upstream of a gene's major PAS; (2) Positive—when the termination window begins downstream of the gene's major PAS. For each cell line its corresponding major PAS was used and the analysis was performed for both active PCGs and

carcinogenesis-related genes. Distances were then visualized in R using the `ggplot2` (v3.4.4) function.

POINT-seq

POINT-Seq analysis was performed similarly to mNET-Seq. Paired-end reads from fastq files were first subjected to quality control (FastQC v0.11.9) followed by adapter trimming using TrimGalore (v0.6.6). Trimmed reads were then aligned to the human genome (hg38) using STAR (Dobin et al, 2013) (v2.7.6a), allowing for only one alignment to the reference. Strand-specific BAMs were generated using SAMtools (v1.10), and the replicates were merged into a single BAM using the same software.

Metagene profiles

Metagene profiles were plotted on the active PCGs and different colorectal APA category gene sets (Dataset EV1). The profiles were generated using `computeMatrix` and `plotProfile` functions from the `deepTools` suite (v3.5.1).

To scale the metagene profiles to 1, for each profile, a tab file consisting of numerical data underlying the plot was retrieved by specifying an option “`--outFileNameData`” while plotting using `plotProfile`. This tab file was used to extract the maximum value for each sample on the plot. Then, we used this maximum value to calculate a scaling factor = $1/\text{max}$. The scaling factor of each sample was multiplied by its corresponding unscaled dataset (matrix). The profiles were then plotted again using the scaled dataset as an input for `plotProfile` and customized in R using the `ggplot2` (v3.4.4) function.

Correlation analysis

Correlation analysis was performed across gene expression, nascent expression, APA and cleavage-termination distance. Nascent transcription levels were quantified by extracting the total RNAPII signal from the transcription start site (TSS) to the 1CT major PAS for each gene using metagene profiles. For each gene, $\log_2$ -relative values were computed for colorectal cancer (CRC) cells (HCT116, SW480, and SW620) relative to 1CT, for HeLa siPCF11 cells relative to siLUC, and for pancreatic Panc1 cells relative to MIAPaCa2. These $\log_2$ -relative values were then used to compute pairwise Spearman correlations among the colorectal model cells, HeLa cells, and pancreatic cells.

Relative values are defined as below: gene expression was quantified as the $\log_2$ fold change of CRC vs 1CT (DESeq2), with negative values indicating downregulation and positive values as upregulation; APA was represented as the $\log_2(\text{distal/proximal})$ difference between CRC and 1CT (DEXSeq) ($\log_2[\text{CRC_distal/proximal}] - \log_2[\text{1CT_distal/proximal}]$), where negative values indicate a shift towards proximal PAS usage and positive towards distal PAS usage; cleavage-termination distance: $\log_2$ difference in distance between CRC and 1CT ($\log_2[\text{CRC_distance}] - \log_2[\text{1CT_distance}]$), with negative values representing shorter and positive as longer distances; and nascent transcription: $\log_2$ difference in total RNAPII signal between CRC and 1CT ($\log_2[\text{CRC_signal}] - \log_2[\text{1CT_signal}]$), where negative values indicate lower and positive as higher nascent transcription signal/levels (Dataset EV1). The same calculations were applied to the other two cells.

Gene ontology (GO) enrichment analysis

Gene ontology enrichment analysis was performed on genes showing a shortened cleavage-termination distance and increased nuclear gene expression in primary tumor cells (HCT116, SW480) and metastatic cells (SW620) compared with healthy cells (1CT).

To define a consensus primary tumor-associated signature, genes overlapping between HCT116 and SW480 (~72%) were chosen for enrichment analysis. GO biological process (BP) was performed using Enrichr-KG (<https://maayanlab.cloud/enrichr-kg>) (Evangelista et al, 2023), with a q value <0.05 considered statistically significant (Dataset EV1).

Proteomic expression analysis

Proteomic data for colon adenocarcinoma (COAD) were obtained from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) (Edwards et al, 2015), from the PNNL Proteome TMT analytical pipeline (March 2017 release) (Vasaikar et al, 2019). Unshared log-ratio protein expression values were extracted for matched tumor and normal tissue samples. The datasets were cross-referenced to identify patients with paired tumor and normal samples, retaining only those with the expression values for the proteins of interest, CPSF3 and PCF11, in both sample types. This filtering resulted in a final cohort of 96 matched tumor–normal pairs for CPSF3 and 60 matched pairs for PCF11. The expression values were used to generate plots.

Data availability

The raw sequencing data of all NGS datasets (mNET-seq, POINT and 3'mRNA-Seq) generated in this paper are deposited in GEO (Gene Expression Omnibus) under the superseries accession number [GSE291826](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE291826) (Colorectal cells - mNET-Seq: [GSE291825](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE291825), POINT-seq: [GSE338427](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE338427) and 3'mRNA-Seq: [GSE291823](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE291823); Pancreatic cells - 3'mRNA-Seq: [GSE338428](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE338428) and T4ph mNET-seq: [GSE338429](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE338429)). The scripts used in this work are publicly available at <https://github.com/STOP-lab/Genomic-analysis-of-transcription-termination-and-3-pre-mRNA-cleavage-in-colorectal-carcinogenesis>. Required material can be obtained by contacting the lead contact, Kinga Kamieniarz-Gdula (kinga.kamieniarz-gdula@amu.edu.pl).

The source data of this paper are collected in the following database record: [biostudies:S-SCDT-10_1038-S44318-026-00928-4](https://www.ebi.ac.uk/biostudies/studies/S-SCDT-10_1038-S44318-026-00928-4).

Expanded view data, supplementary information, appendices are available for this paper at <https://doi.org/10.1038/s44318-026-00928-4>.

Peer review information

A peer review file is available at <https://doi.org/10.1038/s44318-026-00928-4>

References

- Abdulrehman G, Xv K, Li Y, Kang L (2018) Effects of meta-tetrahydroxyphenylchlorin photodynamic therapy on isogenic colorectal cancer SW480 and SW620 cells with different metastatic potentials. *Lasers Med Sci* 33:1581–1590
- Alahmari AA, Chaubey AH, Jonnakuti VS, Tisdale AA, Schwarz CD, Cornwell AC, Maraszek KE, Paterson EJ, Kim M, Venkat S et al (2024) CPSF3 inhibition blocks pancreatic cancer cell proliferation through disruption of core histone mRNA processing. *RNA* 30:281–297
- Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res* 22:2008–2017

- Austena LMI, Piccolo V, Russo M, Prosperini E, Polletti S, Polizzese D, Ghisletti S, Barozzi I, Diaferia GR, Natoli G (2021) A first exon termination checkpoint preferentially suppresses extragenic transcription. *Nat Struct Mol Biol* 28:337–346
- Baillat D, Hakimi M-A, Nääre AM, Shilatfard A, Cooch N, Shiekhattar R (2005) Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 123:265–276
- Bayona-Feliu A, Barroso S, Muñoz S, Aguilera A (2021) The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated transcription-replication conflicts. *Nat Genet* 53:1050–1063
- Birre CE, Minvielle-Sebastia L, Lee BA, Keller W, Proudfoot NJ (1998) Coupling termination of transcription to messenger RNA maturation in yeast. *Science* 280:298–301
- Boreikaite V, Elliott TS, Chin JW, Passmore LA (2022) RBBP6 activates the pre-mRNA 3' end processing machinery in humans. *Genes Dev* 36:210–224
- Brattain MG, Fine WD, Khaled FM, Thompson J, Brattain DE (1981) Heterogeneity of malignant cells from a human colonic carcinoma. *Cancer Res* 41:1751–1756
- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 74:229–263
- Burri D, Zavolan M (2021) Shortening of 3' UTRs in most cell types composing tumor tissues implicates alternative polyadenylation in protein metabolism. *RNA* 27:1459–1470
- Caldas P, Luz M, Baseggio S, Andrade R, Sobral D, Grosso AR (2024) Transcription readthrough is prevalent in healthy human tissues and associated with inherent genomic features. *Commun Biol* 7:100
- Carminati M, Rodríguez-Molina JB, Manav MC, Bellini D, Passmore LA (2023) A direct interaction between CPF and RNA Pol II links RNA 3' end processing to transcription. *Mol Cell* 83:4461–4478.e13
- Chen S, Zhu Z, Yang X, Liu L, He Y, Yang M, Guan X, Wang X, Yun J (2021) Cleavage and polyadenylation specific factor 1 promotes tumor progression via alternative polyadenylation and splicing in hepatocellular carcinoma. *Front Cell Dev Biol* 9:616835
- Chen W, Guo W, Li M, Shi D, Tian Y, Li Z, Wang J, Fu L, Xiao X, Liu QQ et al (2013) Upregulation of cleavage and polyadenylation specific factor 4 in lung adenocarcinoma and its critical role for cancer cell survival and proliferation. *PLoS One* 8:e82728
- Chen Z, Hao W, Tang J, Gao W-Q, Xu H (2022) CSTF2 Promotes hepatocarcinogenesis and hepatocellular carcinoma progression via aerobic glycolysis. *Front Oncol* 12:897804
- Connelly S, Manley JL (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* 2:440–452
- Cortazar MA, Erickson B, Fong N, Pradhan SJ, Ntini E, Bentley DL (2022) Xrn2 substrate mapping identifies torpedo loading sites and extensive premature termination of RNA pol II transcription. *Genes Dev* 36:1062–1078
- Cui Y, Wang L, Ding Q, Shin J, Cassel J, Liu Q, Salvino JM, Tian B (2023) Elevated pre-mRNA 3' end processing activity in cancer cells renders vulnerability to inhibition of cleavage and polyadenylation. *Nat Commun* 14:4480
- Davidson L, Rouvière JO, Sousa-Luis R, Nojima T, Proudfoot NJ, Jensen TH, West S (2024) DNA-directed termination of mammalian RNA polymerase II. *Genes Dev* 38:998–1019
- Derti A, Garrett-Engele P, MacIsaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T (2012) A quantitative atlas of polyadenylation in five mammals. *Genome Res* 22:1173–1183
- Dhir A, Dhir S, Proudfoot NJ, Jopling CL (2015) Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. *Nat Struct Mol Biol* 22:319–327
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21
- Eaton JD, Davidson L, Bauer DLV, Natsume T, Kanemaki MT, West S (2018) Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev* 32:127–139
- Eaton JD, Francis L, Davidson L, West S (2020) A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes Dev* 34:132–145
- Edwards NJ, Oberti M, Thangudu RR, Cai S, McGarvey PB, Jacob S, Madhavan S, Ketchum KA (2015) The CPTAC data portal: a resource for cancer proteomics research. *J Proteome Res* 14:2707–2713
- Estell C, Davidson L, Steketee PC, Monier A, West S (2021) ZC3H4 restricts non-coding transcription in human cells. *eLife* 10:e67305
- Evangelista JE, Xie Z, Marino GB, Nguyen N, Clarke DJB, Ma'ayan A (2023) Enrichr-KG: bridging enrichment analysis across multiple libraries. *Nucleic Acids Res* 51:W168–W179
- Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, Znaor A, Bray F (2019) Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 144:1941–1953
- Fischl H, Neve J, Wang Z, Patel R, Louey A, Tian B, Furger A (2019) hnRNPC regulates cancer-specific alternative cleavage and polyadenylation profiles. *Nucleic Acids Res* 47:7580–7591
- Fontes MM, Guvenek A, Kawaguchi R, Zheng D, Huang A, Ho VM, Chen PB, Liu X, O'Dell TJ, Coppola G et al (2017) Activity-dependent regulation of alternative cleavage and polyadenylation during hippocampal long-term potentiation. *Sci Rep* 7:17377
- Frankish A, Carbonell-Sala S, Diekhans M, Jungreis I, Loveland JE, Mudge JM, Sisu C, Wright JC, Arnan C, Barnes I et al (2023) GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Res* 51:D942–D949
- Gómez-González B, Aguilera A (2019) Transcription-mediated replication hindrance: a major driver of genome instability. *Genes Dev* 33:1008–1026
- Greger IH (1998) Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of *Saccharomyces cerevisiae*. *EMBO J* 17:4771–4779
- Gross S, Moore C (2001) Five subunits are required for reconstitution of the cleavage and polyadenylation activities of *Saccharomyces cerevisiae* cleavage factor I. *Proc Natl Acad Sci USA* 98:6080–6085
- Grzechnik P, Gdula MR, Proudfoot NJ (2015) Pcf1 orchestrates transcription termination pathways in yeast. *Genes Dev* 29:849–861
- Hamperl S, Bocek MJ, Saldivar JC, Swigut T, Cimprich KA (2017) Transcription-replication conflict orientation modulates R-Loop levels and activates distinct DNA damage responses. *Cell* 170:774–786.e19
- Han Z, Moore GA, Mitter R, Lopez Martinez D, Wan L, Dirac Svejstrup AB, Rueda DS, Svejstrup JQ (2023) DNA-directed termination of RNA polymerase II transcription. *Mol Cell* 83:3253–3267.e7
- Hobson DJ, Wei W, Steinmetz LM, Svejstrup JQ (2012) RNA polymerase II collision interrupts convergent transcription. *Mol Cell* 48:365–374
- Hornshøj H, Nielsen MM, Sinnott-Armstrong NA, Świtnicki MP, Juul M, Madsen T, Sallari R, Kellis M, Ørntoft T, Hobolth A et al (2018) Pan-cancer screen for mutations in non-coding elements with conservation and cancer specificity reveals correlations with expression and survival. *npj Genomic Med* 3:1
- Jana A, Das A, Krett NL, Guzman G, Thomas A, Mancinelli G, Bauer J, Ushio-Fukai M, Fukai T, Jung B (2020) Nuclear translocation of Atox1 potentiates activin A-induced cell migration and colony formation in colon cancer. *PLoS One* 15:e0227916
- Kakegawa J, Sakane N, Suzuki K, Yoshida T (2019) JTE-607, a multiple cytokine production inhibitor, targets CPSF3 and inhibits pre-mRNA processing. *Biochem Biophys Res Commun* 518:32–37

- Kakutani M, Takeuchi K, Waga I, Iwamura H, Wakitani K (1999) JTE-607, a novel inflammatory cytokine synthesis inhibitor without immunosuppression, protects from endotoxin shock in mice. *Inflamm Res* 48:461-468
- Kamieniarz-Gdula K, Gdula MR, Panser K, Nojima T, Monks J, Wiśniewski JR, Riepsaame J, Brockdorff N, Pauli A, Proudfoot NJ (2019) Selective roles of vertebrate PCF11 in premature and full-length transcript termination. *Mol Cell* 74:158-172.e9
- Kamieniarz-Gdula K, Proudfoot NJ (2019) Transcriptional control by premature termination: a forgotten mechanism. *Trends Genet* 35:553-564
- Keum N, Giovannucci E (2019) Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol* 16:713-732
- Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432:517-522
- Kopczyńska M, Saha U, Romanenko A, Nojima T, Gdula MR, Kamieniarz-Gdula K (2024) Defining gene ends: RNA polymerase II CTD threonine 4 phosphorylation marks transcription termination regions genome-wide. *Nucleic Acids Res* 53:gkae1240
- Lalonde M, Ummethum H, Trauner M, Ettinger A, Hamperl S (2024) An automated image analysis pipeline to quantify the coordination and overlap of transcription and replication activity in mammalian genomes. *Methods Cell Biol* 182:199-219
- Leibovitz A, Stinson JC, McCombs WB, McCoy CE, Mazur KC, Mabry ND (2019) Classification of human colorectal adenocarcinoma cell lines. *Cancer Res* 79:4562-4569
- Lin P, Chen W, Long Z, Yu J, Yang J, Xia Z, Wu Q, Min X, Tang J, Cui Y et al (2024) RBBP6 maintains glioblastoma stem cells through CPSF3-dependent alternative polyadenylation. *Cell Discov* 10:32
- Lin Y, Li Z, Ozsolak F, Kim SW, Arango-Argoty G, Liu TT, Tenenbaum SA, Bailey T, Monaghan AP, Milos PM et al (2012) An in-depth map of polyadenylation sites in cancer. *Nucleic Acids Res* 40:8460-8471
- Liu H, Heller-Trulli D, Moore CL (2022) Targeting the mRNA endonuclease CPSF73 inhibits breast cancer cell migration, invasion, and self-renewal. *iScience* 25:104804
- Liu L, Yu AM, Wang X, Soles LV, Teng X, Chen Y, Yoon Y, Sarkan KSK, Valdez MC, Linder J et al (2023) The anticancer compound JTE-607 reveals hidden sequence specificity of the mRNA 3' processing machinery. *Nat Struct Mol Biol* 30:1947-1957
- Liu W, Deng L, Wang M, Liu X, Ouyang X, Wang Y, Miao N, Luo X, Wu X, Lu X et al (2025a) Pcf11/Spt5 condensates stall RNA polymerase II to facilitate termination and piRNA-guided heterochromatin formation. *Mol Cell* 85:929-947.e10
- Liu Y, Sun J, Zhao M (2017) ONGene: a literature-based database for human oncogenes. *J Genet Genomics* 44:119-121
- Liu Y, Zhangding Z, Liu X, Hu J (2025b) Chromatin-centric insights into DNA replication. *Trends Genet* 41:412-424
- Logan J, Falck-Pedersen E, Darnell JE, Shenk T (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci USA* 84:8306-8310
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550
- Luo W, Johnson AW, Bentley DL (2006) The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev* 20:954-965
- Lykke-Andersen S, Žumer K, Molska EŠ, Rouvière JO, Wu G, Demel C, Schwalb B, Schmid M, Cramer P, Jensen TH (2021) Integrator is a genome-wide attenuator of non-productive transcription. *Mol Cell* 81:514-529.e6
- Maeda H, Khatami M (2018) Analyses of repeated failures in cancer therapy for solid tumors: poor tumor-selective drug delivery, low therapeutic efficacy and unsustainable costs. *Clin Trans Med* 7:e11
- Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444:953-956
- Marchal C, Sima J, Gilbert DM (2019) Control of DNA replication timing in the 3D genome. *Nat Rev Mol Cell Biol* 20:721-737
- Mayr C, Bartel DP (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138:673-684
- Morales JC, Richard P, Patidar PL, Motea EA, Dang TT, Manley JL, Boothman DA (2016) XRN2 links transcription termination to DNA damage and replication stress. *PLoS Genet* 12:e1006107
- Nakayama C, Fang Q, Daigaku Y, Aoi Y, Ito S, Takahashi M, Shimatani R, Minamisawa T, Ozturk Y, Kimura H et al (2026) NELF prevents transcriptional readthrough into DNA replication zones in cancer cells. *EMBO Rep* 27:1720-1758
- Narain A, Bhandare P, Adhikari B, Backes S, Eilers M, Dölken L, Schlosser A, Erhard F, Baluapuri A, Wolf E (2021) Targeted protein degradation reveals a direct role of SPT6 in RNAPII elongation and termination. *Mol Cell* 81:3110-3127.e14
- Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ (2015) Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* 161:526-540
- Nojima T, Tellier M, Foxwell J, Ribeiro De Almeida C, Tan-Wong SM, Dhir S, Dujardin G, Dhir A, Murphy S, Proudfoot NJ (2018) Deregulated expression of mammalian lncRNA through loss of SPT6 induces R-loop formation, replication stress, and cellular senescence. *Mol Cell* 72:970-984.e7
- Ogorodnikov A, Levin M, Tattikotla S, Tokalov S, Hoque M, Scherzinger D, Marini F, Poetsch A, Binder H, Macher-Göppinger S et al (2018) Transcriptome 3' end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. *Nat Commun* 9:5331
- Pagani G, Gandellini P (2024) Cleavage and polyadenylation machinery as a novel targetable vulnerability for human cancer. *Cancer Gene Ther* 31:957-960
- Podsiywałow-Bartnicka P, Shine M, Lin J, Neugebauer KM (2026) Transcriptional readthrough precedes alternative splicing programs triggered in CML cells by imatinib. *Sci Adv* 20:eaea2475 <https://doi.org/10.1126/sciadv.aea2475>
- Prescott EM, Proudfoot NJ (2002) Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci USA* 99:8796-8801
- Proudfoot NJ (1989) How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* 14:105-110
- Qiao P, Zhang C, Shi Y, Du H (2024) The role of alternative polyadenylation in breast cancer. *Front Genet* 15:1377275
- Reyes A, Anders S, Weatheritt RJ, Gibson TJ, Steinmetz LM, Huber W (2013) Drift and conservation of differential exon usage across tissues in primate species. *Proc Natl Acad Sci USA* 110:15377-15382
- Rodríguez-Molina JB, West S, Passmore LA (2023) Knowing when to stop: transcription termination on protein-coding genes by eukaryotic RNAPII. *Mol Cell* 83:404-415
- Roig AI, Eskicak U, Hight SK, Kim SB, Delgado O, Souza RF, Spechler SJ, Wright WE, Shay JW (2010) Immortalized epithelial cells derived from human colon biopsies express stem cell markers and differentiate in vitro. *Gastroenterology* 138:1012-1021.e5
- Ross NT, Lohmann F, Carbonneau S, Fazal A, Weihofen WA, Gleim S, Salcius M, Sigoillot F, Henault M, Carl SH et al (2020) CPSF3-dependent pre-mRNA processing as a druggable node in AML and Ewing's sarcoma. *Nat Chem Biol* 16:50-59

- Rot G, Wang Z, Huppertz I, Modic M, Lenče T, Hallegger M, Haberman N, Curk T, Von Mering C, Ule J (2017) High-resolution RNA maps suggest common principles of splicing and polyadenylation regulation by TDP-43. *Cell Rep* 19:1056–1067
- Rouvière JO, Salerno-Kochan A, Lykke-Andersen S, Garland W, Dou Y, Rathore O, Molska EŠ, Wu G, Schmid M, Bugai A et al (2023) ARS2 instructs early transcription termination-coupled RNA decay by recruiting ZC3H4 to nascent transcripts. *Mol Cell* 83:2240–2257.e6
- Sadowski M (2003) Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *EMBO J* 22:2167–2177
- Salifu SP, Doughan A (2022) New clues to prognostic biomarkers of four hematological malignancies. *J Cancer* 13:2490–2503
- Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ (2017) Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol Cell* 65:25–38
- Schmidt M, Kluge F, Sandmeir F, Kühn U, Schäfer P, Tüting C, Ihling C, Conti E, Wahle E (2022) Reconstitution of 3' end processing of mammalian pre-mRNA reveals a central role of RBBP6. *Genes Dev* 36:195–209
- Schmidt U, Weigert M, Broaddus C, Myers G (2018) Cell detection with star-convex polygons. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G (eds) *Medical image computing and computer assisted intervention-MICCAI 2018*. Springer, 265–273
- Schüller R, Forné I, Straub T, Schreieck A, Texier Y, Shah N, Decker T-M, Cramer P, Imhof A, Eick D (2016) Heptad-specific phosphorylation of RNA polymerase II CTD. *Mol Cell* 61:305–314
- Sousa-Luis R, Dujardin G, Zukher I, Kimura H, Weldon C, Carmo-Fonseca M, Proudfoot NJ, Nojima T (2021) POINT technology illuminates the processing of polymerase-associated intact nascent transcripts. *Mol Cell* 81:1935–1950.e6
- Tajima N, Fukui K, Uesato N, Maruhashi J, Yoshida T, Watanabe Y, Tojo A (2010) JTE-607, a multiple cytokine production inhibitor, induces apoptosis accompanied by an increase in p21^{waf1/cip1} in acute myelogenous leukemia cells. *Cancer Sci* 101:774–781
- Tamaddon M, Shokri G, Hosseini Rad SMA, Rad I, Emami Razavi A, Kouhkan F (2020) Involved microRNAs in alternative polyadenylation intervene in breast cancer via regulation of cleavage factor “CFlm25”. *Sci Rep* 10:11608
- Tan S, Li H, Zhang W, Shao Y, Liu Y, Guan H, Wu J, Kang Y, Zhao J, Yu Q et al (2018) NUDT21 negatively regulates PSMB2 and CXXC5 by alternative polyadenylation and contributes to hepatocellular carcinoma suppression. *Oncogene* 37:4887–4900
- Teloni F, Michelena J, Lezaja A, Kilic S, Ambrosi C, Menon S, Dobrovolna J, Imhof R, Janscak P, Baubec T et al (2019) Efficient pre-mRNA cleavage prevents replication-stress-associated genome instability. *Mol Cell* 73:670–683.e12
- Uesato N, Fukui K, Maruhashi J, Tojo A, Tajima N (2006) JTE-607, a multiple cytokine production inhibitor, ameliorates disease in a SCID mouse xenograft acute myeloid leukemia model. *Exp Hematol* 34:1385–1392
- Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, Dou Y, Zhang Y, Shi Z, Arshad OA et al (2019) Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* 177:1035–1049.e19
- Wagner EJ, Tong L, Adelman K (2023) Integrator is a global promoter-proximal termination complex. *Mol Cell* 83:416–427
- Wang R, Zheng D, Wei L, Ding Q, Tian B (2019) Regulation of intronic polyadenylation by PCF11 impacts mRNA expression of long genes. *Cell Rep* 26:2766–2778.e6
- Weigert M, Schmidt U, Boothe T, Müller A, Dibrov A, Jain A, Wilhelm B, Schmidt D, Broaddus C, Culley S et al (2018) Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat Methods* 15:1090–1097
- West S, Gromak N, Proudfoot NJ (2004) Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432:522–525
- West S, Proudfoot NJ (2007) Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. *Nucleic Acids Res* 36:905–914
- Whitelaw E, Proudfoot N (1986) Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *EMBO J* 5:2915–2922
- Wiśniewski JR, Duś-Szachniewicz K, Ostasiewicz P, Ziolkowski P, Rakus D, Mann M (2015) Absolute proteome analysis of colorectal mucosa, adenoma, and cancer reveals drastic changes in fatty acid metabolism and plasma membrane transporters. *J Proteome Res* 14:4005–4018
- Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, Li W (2014) Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun* 5:5274
- Xing Y, Chen L, Gu H, Yang C, Zhao J, Chen Z, Xiong M, Kazobinka G, Liu Y, Hou T (2021) Downregulation of NUDT21 contributes to cervical cancer progression through alternative polyadenylation. *Oncogene* 40:2051–2064
- Zhang Z, Fu J, Gilmour DS (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev* 19:1572–1580
- Zhao M, Kim P, Mitra R, Zhao J, Zhao Z (2016) TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res* 44:D1023–D1031
- Zhao M, Sun J, Zhao Z (2013) TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res* 41:D970–D976
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools Abstract GigaScience 10: <https://doi.org/10.1093/gigascience/giab008>
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T (2016) deepTools2: a next generation web server for deep-sequencing data analysis *Nucleic Acids Research* 44:W160–W165 <https://doi.org/10.1093/nar/gkw257>
- Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P (2015) Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31:2032–2034 <https://doi.org/10.1093/bioinformatics/btv098>
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137 <https://doi.org/10.1186/gb-2008-9-9-r137>
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* 38:576–589 <https://doi.org/10.1016/j.molcel.2010.05.004>
- Wickham H, Wickham H (2016) ggplot2 Data Analysis. Springer International Publishing Cham 189–201 <https://doi.org/10.1007/978-3-319-24277-4>
- Robert C, Gentleman Vincent J, Carey Douglas M, Bates Ben, Bolstad Marcel, Dettling Sandrine, Dudoit Byron, Ellis Laurent, Gautier Yongchao, Ge Jeff, Gentry Kurt, Hornik Torsten, Hothorn Wolfgang, Huber Stefano, Iacus Rafael, Irizarry Friedrich, Leisch Cheng, Li Martin, Maechler Anthony J, Rossini Gunther, Sawitzki Colin, Smith Gordon, Smyth Luke, Tierney Jean YH, Yang Jianhua, Zhang (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5:R80 <https://doi.org/10.1186/gb-2004-5-10-r80>
- Stasevich TJ, Hayashi-Takanaka Y, Sato Y, Maehara K, Ohkawa Y, Sakata-Sogawa K, Tokunaga M, Nagase T, Nozaki N, McNally JG, Kimura H (2014)

Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* 516(7530) 272–275 <https://doi.org/10.1038/nature13714>

Acknowledgements

We thank Nick Proudfoot and Joanna Kwiatkowska for their critical reading of the manuscript. This work was funded by the European Union, Horizon Europe, ERC grant AlternativeEnds project number 101042642, National Science Centre, Poland, project number 2018/30/E/NZ1/00073, Polish National Agency for Academic Exchange project PPN/PPO/2018/1/00053, and European Molecular Biology Organization EMBO IG 4728-2020. MRG was supported by the National Science Centre, Poland, project number 2018/31/B/NZ2/04004. SH was supported by the Helmholtz Association and the European Research Council (ERC Starting Grant 852798 ConflictResolution). HK was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI(KP24H02325, JP26H02362).

Author contributions

Agata Stepień: Conceptualization; Data curation; Formal analysis; Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing. **Deepshika Pulimamidi:** Data curation; Formal analysis; Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing. **Martyna Plens-Galaska:** Formal analysis; Investigation; Visualization; Methodology; Writing—review and editing. **Maxime Lalonde:** Formal analysis; Investigation; Visualization; Methodology; Writing—review and editing. **Magda Kopczyńska:** Formal analysis; Investigation; Visualization; Methodology; Writing—review and editing. **Klaudia Ostatek:** Formal analysis; Investigation; Visualization. **Mariia Kapeliukha:** Formal analysis; Investigation; Methodology; Writing—review and editing. **Hiroshi Kimura:** Resources; Writing—review and editing. **Shazia Irshad:** Resources; Writing—review and editing. **Michał R Gdula:** Methodology; Writing—review and editing. **Stephan Hamperl:** Funding acquisition; Investigation; Writing—review and editing. **Kinga**

Kamieniarz-Gdula: Conceptualization; Data curation; Supervision; Funding acquisition; Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing.

Source data underlying figure panels in this paper may have individual authorship assigned. Where available, figure panel/source data authorship is listed in the following database record: [biostudies:S-SCDT-10_1038-S44318-026-00928-4](https://biostudies.org/studies/S-SCDT-10_1038-S44318-026-00928-4).

Disclosure and competing interests statement

The authors declare no competing interests.

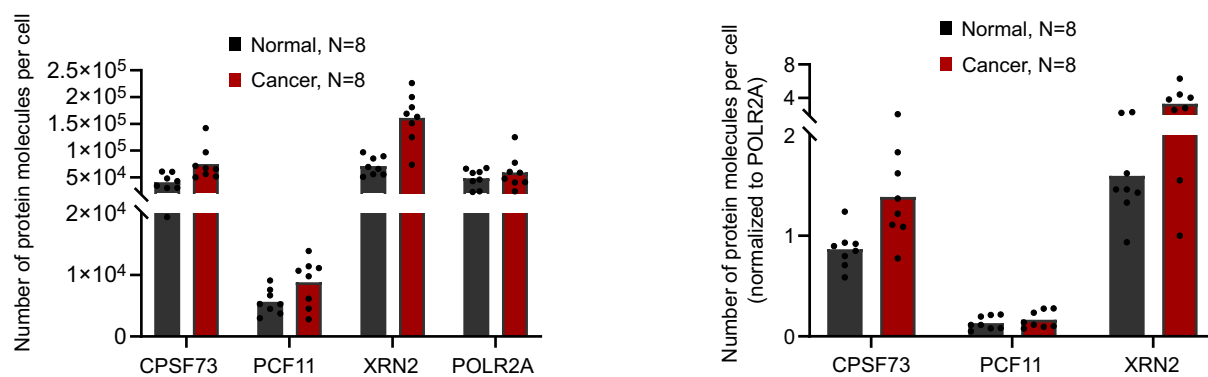
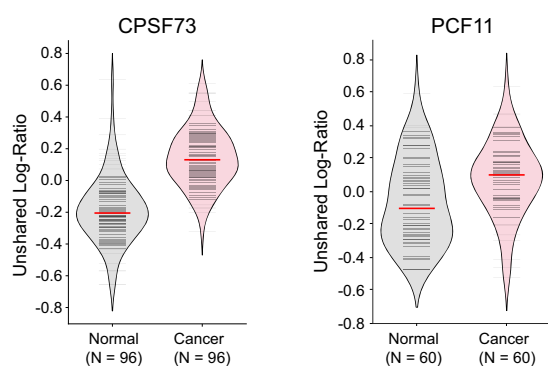
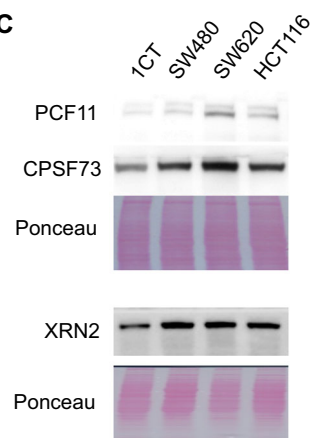
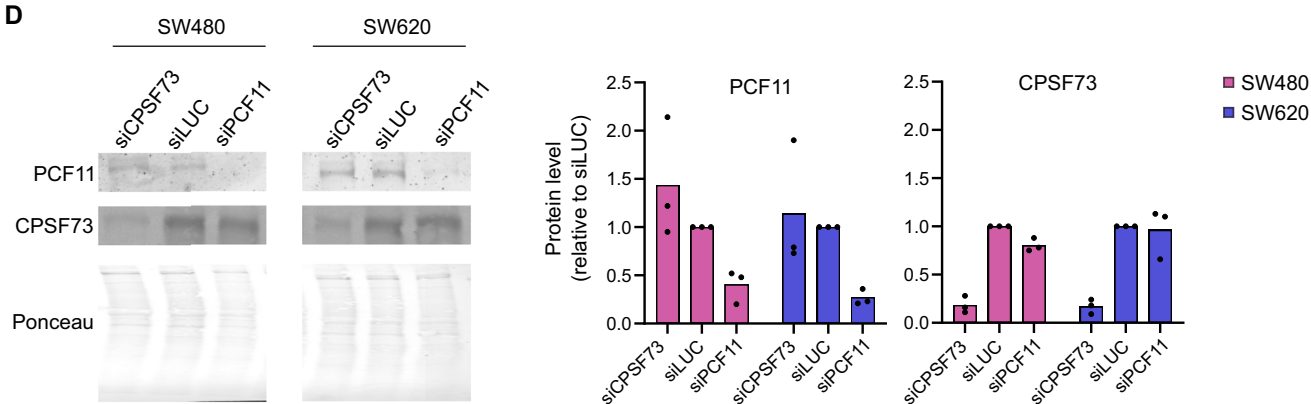
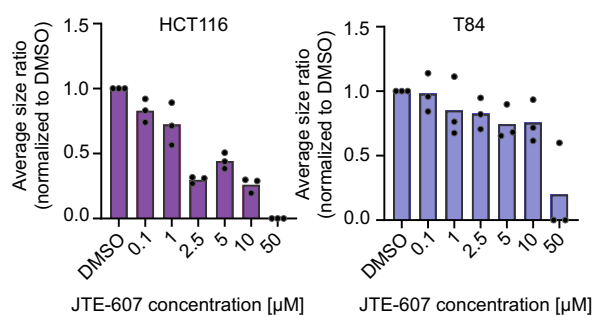
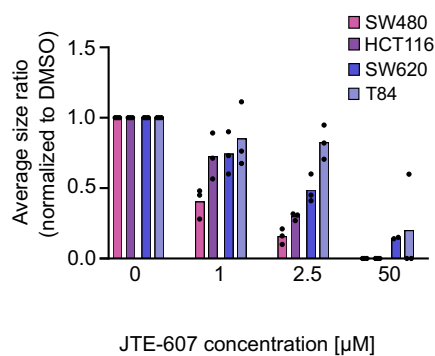
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the data associated with this article, unless otherwise stated in a credit line to the data, but does not extend to the graphical or creative elements of illustrations, charts, or figures. This waiver removes legal barriers to the re-use and mining of research data. According to standard scholarly practice, it is recommended to provide appropriate citation and attribution whenever technically possible.

© The Author(s) 2026

Expanded View Figures

Figure EV1. Protein levels of PCF11, CPSF73, and XRN2 in CRC and the effect of their manipulations on the clonogenic potential of CRC cells.

Related to Fig. 1. (A) Bar graph showing the absolute (left panel) or normalized (to the largest subunit of RNAPII; right panel) number of protein molecules per cell based on global quantitative proteomics from patient biopsies (Wiśniewski et al, 2015). (A, C, E, F) Bar graphs represent the mean, while dots indicate individual biological replicates. (B) Violin plots showing the levels of the PCF11 and CPSF73 proteins in CRC tissue versus matched normal tissue in data from the Clinical Proteomic Tumor Analysis Consortium dataset (CPTAC 2019) ($N = 60$ pairs for PCF11; $N = 96$ pairs for CPSF73). Protein level is quantified as the “unshared log-ratio,” a $\log^2$ -transformed abundance ratio relative to a common reference standard, with quantification restricted to unique peptides mapping to the target gene. The red horizontal bar denotes the median for each cohort. (C) Western blot showing the levels of PCF11, CPSF73, and XRN2 in the cell culture-based CRC progression model. (D) Western blot showing the levels of PCF11 and CPSF73 after PCF11 or CPSF73 knockdown (siPCF11 or siCPSF73) in SW480 and SW620 cells, with the calculations from three biological replicates. (E, F) Bar graphs showing the quantification of the colony-formation assay.

A**B****C****D****E****F**

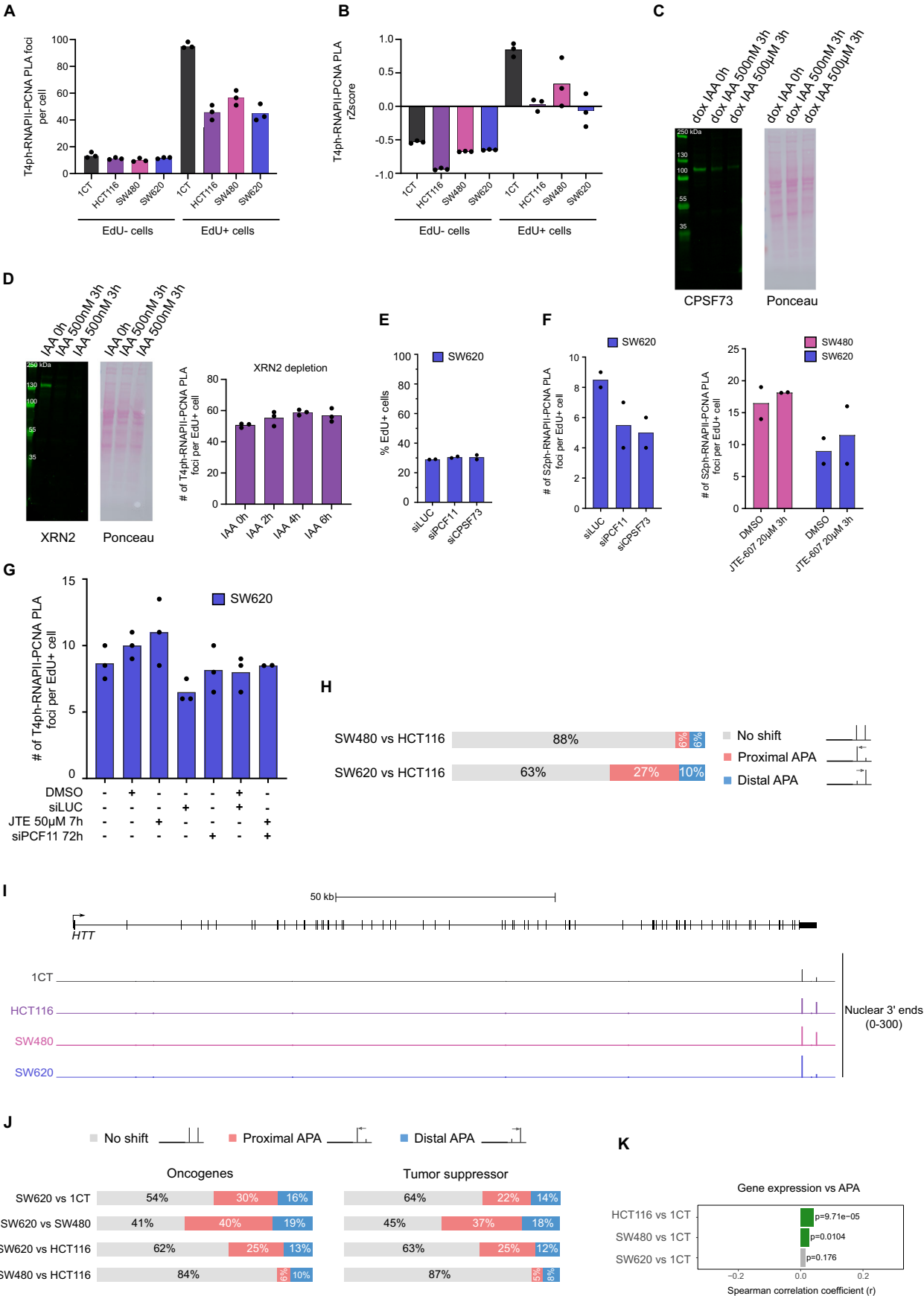
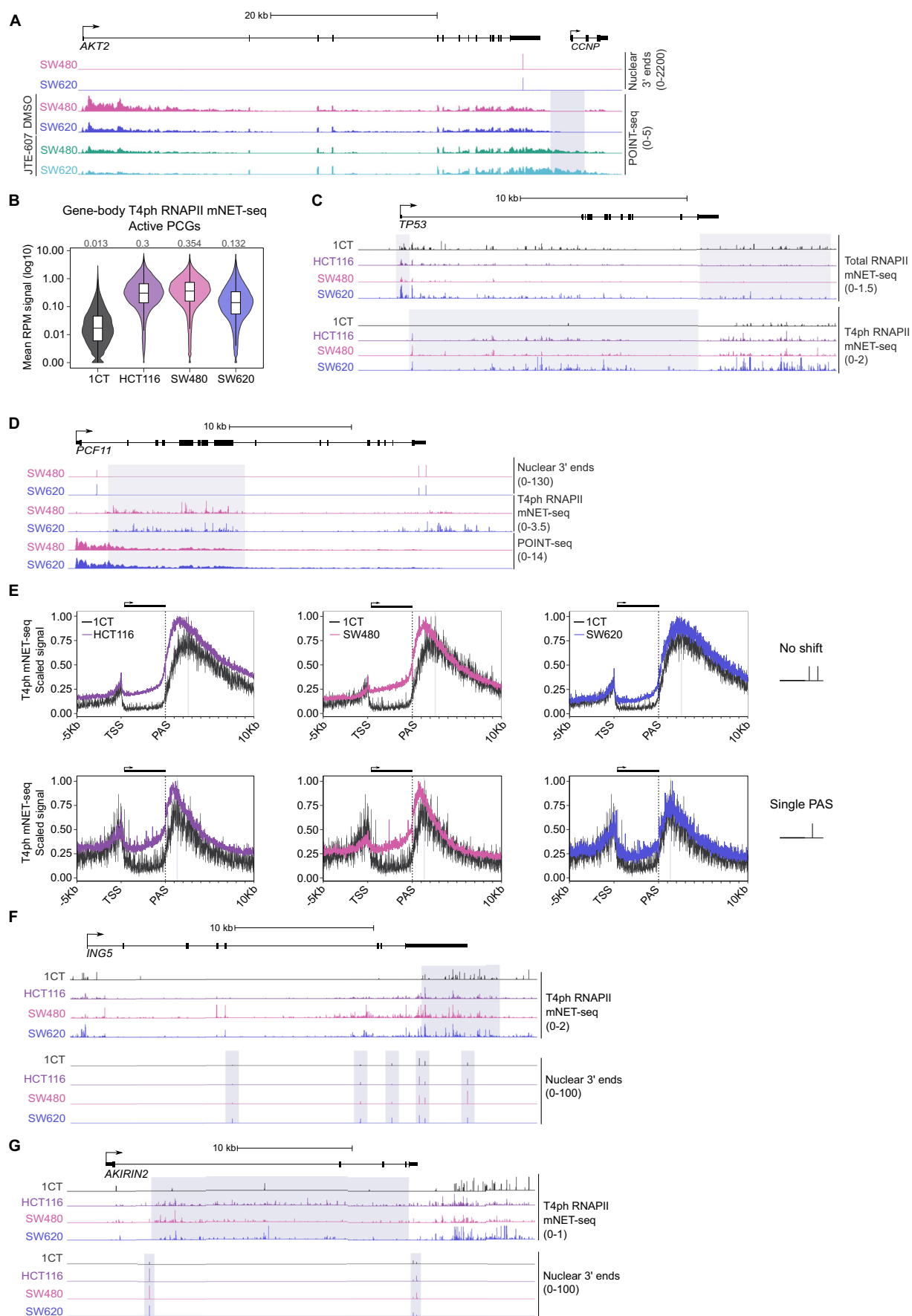


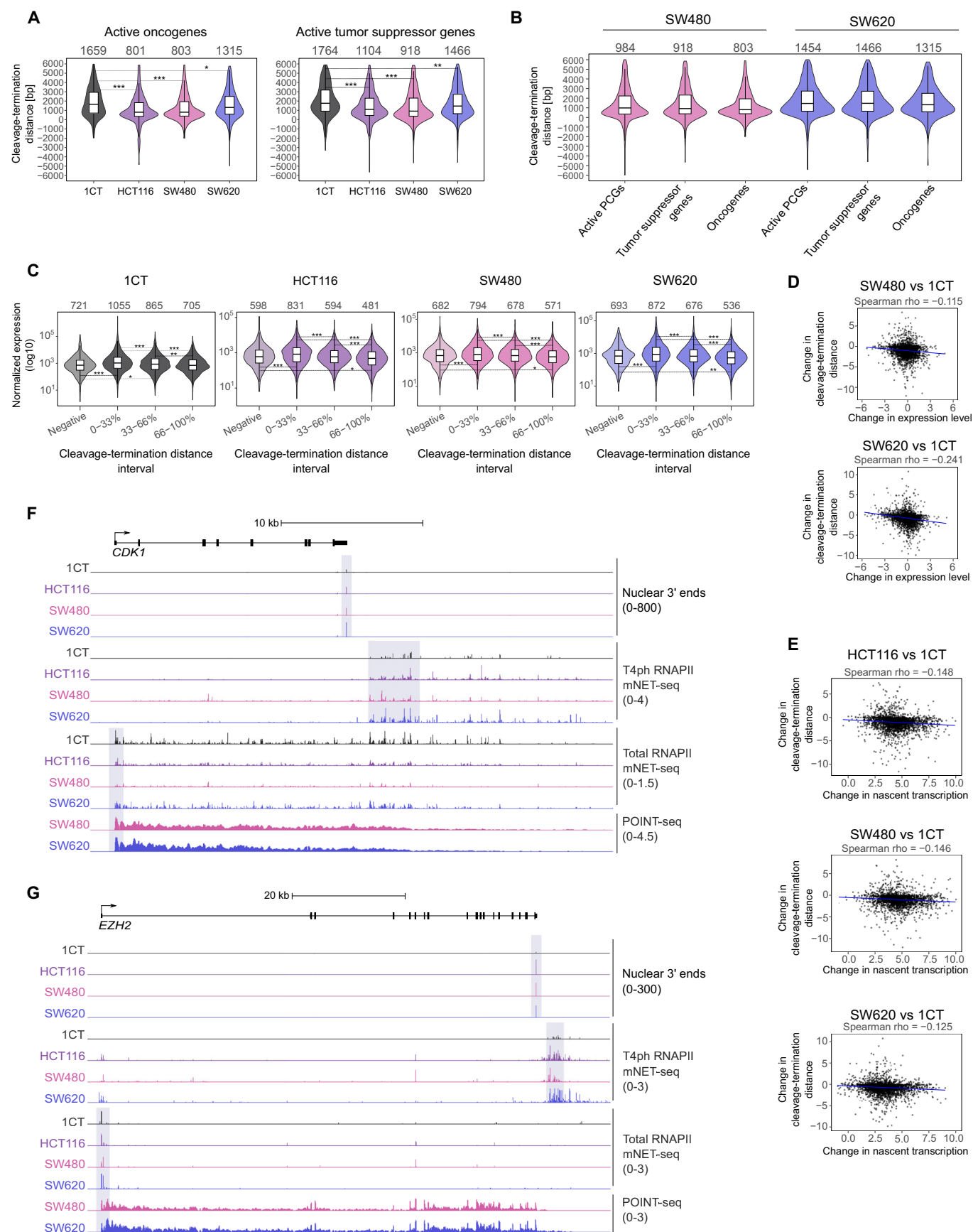
Figure EV2. Frequency of transcription-replication collisions upon transcription termination factor alterations.

Related to Fig. 2. (A) Bar graph showing the number of transcription-replication collisions (TRCs) (by T4ph RNAPII-PCNA PLA foci) in EdU-negative and EdU-positive cells. (A–G) Bar graphs represent average values, each dot represents one biological replicate ($N = 2-3$), calculated as the mean of 2–3 technical replicates (100–6400 analyzed cells per technical replicate). The exact numbers are provided in the Appendix Supplementary Methods. (B) Bar graph showing the rZscore for T4ph RNAPII-PCNA PLA signal in Edu-negative and Edu-positive cells. (C) Western blot showing the level of CPSF73 after acute CPSF73 depletion (dox IAA 3 h) in HCT116 CPSF73-AID cells. (D) Western blot showing the level of XRN2 after acute XRN2 depletion (IAA 3 h) and bar graph representing the number of transcription-replication collisions (TRCs) after XRN2 depletion (IAA 0–6 h) in HCT116 XRN2-AID cells. (E) Bar graph showing the % of EdU-positive cells in control (siLUC), PCF11 knockdown (siPCF11) or CPSF73 knockdown (siCPSF73) conditions in SW620 cells. (F) Bar graphs showing the number of TRCs in control (siLUC) or PCF11 (siPCF11) or CPSF73 (siCPSF73) knockdown conditions in SW620 cells (left) or after treatment of SW480 or SW620 cells with JTE-607 (right) obtained with the use of an antibody against Ser2-phosphorylated RNAPII. (G) Bar graph showing the number of TRCs after prolonged treatment of SW620 cells with siPCF11 or JTE-607 or combined treatment. (H) Bar graph representing the alternative polyadenylation (APA) analysis in the CRC progression model. (H–K) $N = 3$ biological replicates. The replicates were merged. (I) Genomic profile of *HTT*, showing nuclear 3' mRNA-seq signals. (J) Bar graph representing the APA analysis in the CRC progression model for separated oncogenes ($n = 533$) and tumor suppressor ($n = 764$). (K) Pairwise correlation between changes in gene expression and PAS usage (APA) in cancer cells compared to 1CT cells. Each bar represents a pairwise comparison, the bar width showing the Spearman correlation coefficient. Green bars are used for positive correlations and gray bars for non-significant correlations. Positive values indicate higher expression/distal APA, negative values indicate lower expression/proximal APA.



◀ **Figure EV3. Increase in premature transcription termination in CRC; uncoupling of changes in pre-mRNA 3' end cleavage and transcription termination during CRC progression.**

Related to Fig. 3. (A) Genomic profile of *AKT2*, showing nuclear 3' mRNA-seq and POINT-seq signals. Gray highlight indicates the region without nascent transcription in SW480 cells in control conditions. (B) Violin plot representation of T4ph RNAPII mNET-seq across the gene body of active PCGs ($n = 13,577$). RNAPII terminal pausing increases strongly in cells from primary CRC, then decreases at the transition to metastatic cell, yet remains upregulated compared to 1CT. The numbers above the plots represent the median. The boxplots display the median (center line), first and third quartiles (lower and upper bounds of the box), and whiskers extending to the most extreme data points within $1.5 \times \text{IQR}$ of the lower and upper quartiles. (B–G) T4ph mNET-seq, $N = 2$ (1CT, HCT116, SW480) or 3 (SW620) biological replicates; Nuclear 3' ends, three biological replicates. The replicates were merged. (C) Genomic profile of the tumor suppressor gene *TP53*, showing increased gene-body RNA terminal pausing in CRC cells. Gray highlights indicate regions of TSS-associated and full-length RNAPII pausing in 1CT cells (top 4 tracks, total RNAPII mNET-seq), and gene-body terminal pausing region in CRC cells (bottom tracks, T4ph RNAPII mNET-seq). (D) Genomic profile of *PCF11*, showing POINT-seq, T4ph mNET-seq, and nuclear 3' mRNA-seq signals. The gray highlight indicates the region of increased T4ph mNETseq and decreased POINT-seq signals. (E) Metagene profiles of T4ph-mNET-seq signal on genes having multiple PAS but not undergoing APA (no shift, upper panel) or genes with a single PAS in the CRC progression model (lower panel). The dotted line indicates the PAS position, and the gray highlight indicates the T4ph-mNET-seq signal maximum downstream of PAS in 1CT cells. (F, G) Genomic profiles of *ING5* (F), *AKIRIN2* (G), showing T4ph-mNET-seq and nuclear 3' mRNA-seq signals in the CRC progression model. The gray highlight indicates full-length RNAPII termination (F) or gene-body pausing (G) and PAS usage.



◀ **Figure EV4. Closer proximity of PAS and the transcription termination window predicts enhanced gene expression.**

Related to Fig. 4. (A) Violin plot representation of the cleavage-termination distance for active oncogenes (left) and tumor suppressor genes (right). (A–G) Based on merged biological replicates. $N = 3$ (nuclear 3' ends), 2–3 (T4ph mNET-seq), two (total-mNET-seq), 2 (POINT-seq) biological replicates. (A–C) *** $P_{\text{adj}} < 0.001$; ** < 0.01 ; * < 0.05 . Non-indicated differences are not significant. The pairwise Wilcoxon test exact p values are provided in Dataset EV1. The boxplots display the median (center line), first and third quartiles (lower and upper bounds of the box), and whiskers extending to the most extreme data points within $1.5 \times \text{IQR}$ of the lower and upper quartiles. (B) Violin plot representation of the cleavage-termination distance for active PCGs, tumor suppressor genes, and oncogenes in 1CT and SW620 cells. (C) Violin plot representation of the normalized expression of genes with negative or positive cleavage-termination distance. The subset of genes with negative cleavage-termination distance was ~7% of genes for 1CT and SW620, and ~11% for HCT116 and SW480. Genes with a positive cleavage-termination distance were divided into 0–33%, 33–66%, and 66–100% distance intervals (from shortest to longest distance). (D) Scatter plots showing pairwise comparison between change in gene expression levels and change in cleavage-termination distance in SW480 or SW620 cells relative to 1CT cells. (D, E) Positive values indicate an increase in expression, transcription and/or cleavage-termination distance; negative values correspond to a decrease. (E) Scatter plots showing the pairwise comparison between change in nascent transcription and change in cleavage-termination distance in CRC cells relative to 1CT cells. (F, G) Genomic profile of oncogenes *CDK1* (F) and *EZH2* (G) showing nuclear 3' mRNA-seq, T4ph-mNET-seq, total-mNET-seq signals, and POINT-seq signals. Gray highlights indicate PAS usage and RNAPII pausing in 1CT cells.

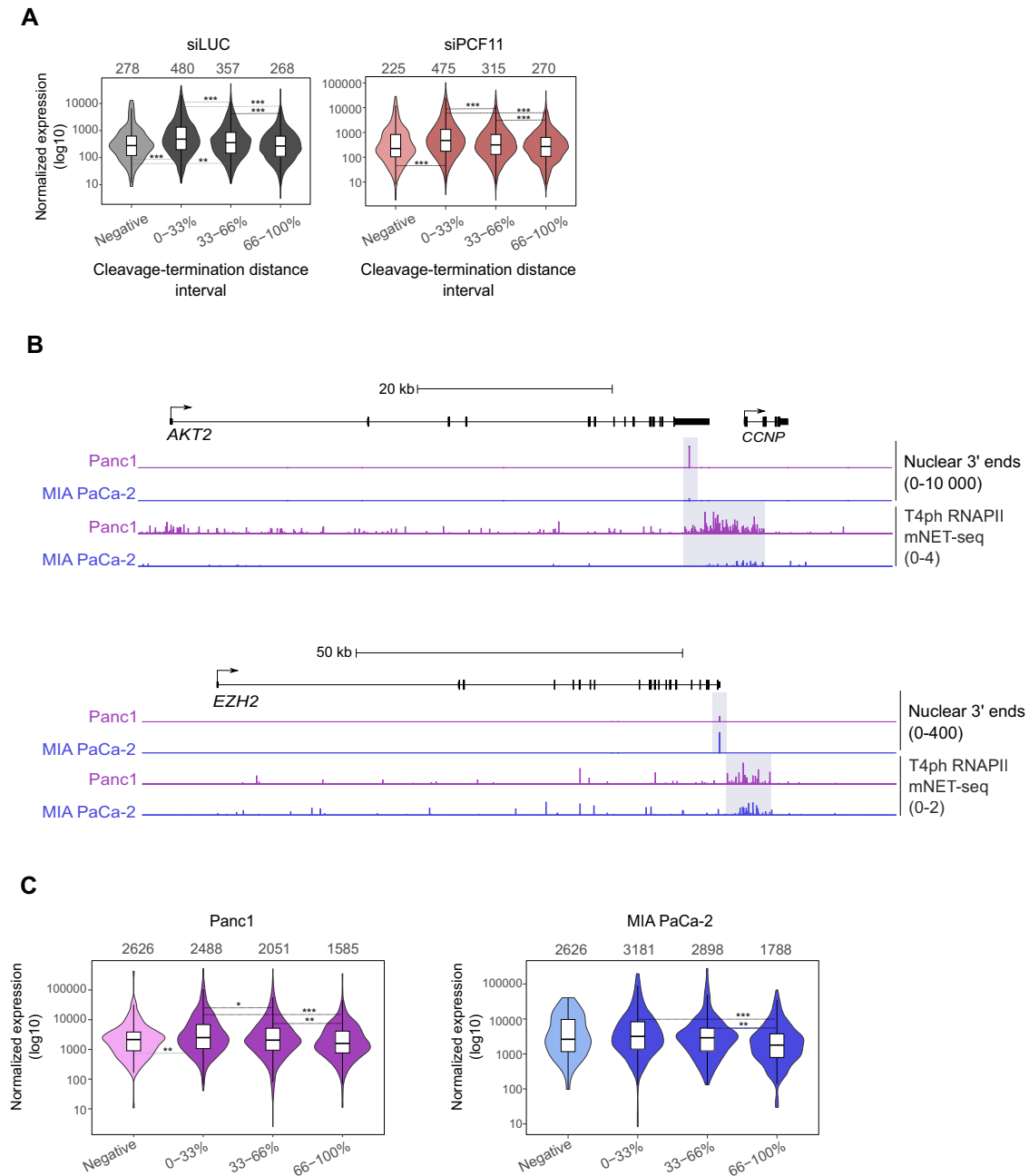


Figure EV5. Cleavage-termination distance in non-CRC cells.

Related to Fig. 5. (A) Violin plots representing normalized expression of genes with negative (~6% in siLUC control condition; ~2.5% in siPCF11 knockdown condition) or positive cleavage-termination distance in HeLa cells. The latter group was divided into 0-33%, 33-66%, and 66-100% distance intervals. The pairwise Wilcoxon test exact p values are provided in Dataset EV1. The boxplots display the median (center line), first and third quartiles (lower and upper bounds of the box), and whiskers extending to the most extreme data points within $1.5 \times \text{IQR}$ of the lower and upper quartiles. Based on three merged biological replicates. (B) Genomic profiles of *AKT2* and *EZH2* showing nuclear 3' mRNA-seq and T4ph-mNET-seq signals in pancreatic cancer cells. The gray highlight indicates PAS usage and RNAPII pausing in Panc1 cells. (B, C) Based on $N = 2$ (T4ph mNET-seq) or 3 (nuclear 3' ends) biological replicates. The replicates were merged. (C) Violin plots representing normalized expression of genes with negative (~8.5% in Panc1; ~7.2% in MIA PaCa-2) or positive cleavage-termination distance in pancreatic cancer cells. The latter group was divided into 0-33%, 33-66%, and 66-100% distance intervals. The pairwise Wilcoxon test exact p values are provided in Dataset EV1. The boxplots display the median (center line), first and third quartiles (lower and upper bounds of the box), and whiskers extending to the most extreme data points within $1.5 \times \text{IQR}$ of the lower and upper quartiles. Based on three merged biological replicates.