



# Cleavage and Polyadenylation Specific Factor 1 Promotes Tumor Progression *via* Alternative Polyadenylation and Splicing in Hepatocellular Carcinoma

## OPEN ACCESS

### Edited by:

Tyson Valentine Sharp,  
Queen Mary University of London,  
United Kingdom

### Reviewed by:

John Gordon Foster,  
Queen Mary University of London,  
United Kingdom  
Gang Wei,  
Fudan University, China  
Ihab Younis,  
Carnegie Mellon University in Qatar,  
Qatar

### \*Correspondence:

Xin Wang  
xin.wang@cityu.edu.hk  
Jing-ping Yun  
yunjp@sysucc.org.cn

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Molecular and Cellular Oncology,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

Received: 13 October 2020

Accepted: 04 February 2021

Published: 04 March 2021

### Citation:

Chen S, Zhu Z, Yang X, Liu L,  
He Y, Yang M, Guan X, Wang X and  
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.  
doi: 10.3389/fcell.2021.616835

Shi-lu Chen<sup>1,2†</sup>, Zhong-xu Zhu<sup>3†</sup>, Xia Yang<sup>1,2†</sup>, Li-li Liu<sup>1,2</sup>, Yang-fan He<sup>1,2</sup>,  
Ming-ming Yang<sup>1,2</sup>, Xin-yuan Guan<sup>1</sup>, Xin Wang<sup>3,4\*</sup> and Jing-ping Yun<sup>1,2\*</sup>

<sup>1</sup> State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China, <sup>2</sup> Department of Pathology, Sun Yat-sen University Cancer Center, Guangzhou, China, <sup>3</sup> Department of Biomedical Sciences, City University of Hong Kong, Hong Kong, China, <sup>4</sup> Key Laboratory of Bioclip Technology, Biotech and Health Centre, Shenzhen Research Institute, City University of Hong Kong, Shenzhen, China

Alternative polyadenylation (APA) is an important post-transcriptional regulatory mechanism required for cleavage and polyadenylation (CPA) of the 3' untranslated region (3' UTR) of mRNAs. Several aberrant APA events have been reported in hepatocellular carcinoma (HCC). However, the regulatory mechanisms underlying APA remain unclear. In this study, we found that the expression of cleavage and polyadenylation specific factor 1 (CPSF1), a major component of the CPA complex, was significantly increased in HCC tissues and correlated with unfavorable survival outcomes. Knockdown of CPSF1 inhibited HCC cell proliferation and migration, whereas overexpression of CPSF1 caused the opposite effect. Based on integrative analysis of Iso-Seq and RNA-seq data from HepG2.2.15 cells, we identified a series of transcripts with differential 3' UTR lengths following the knockdown of CPSF1. These transcripts were related to the biological functions of gene transcription, cytoskeleton maintenance, and endomembrane system transportation. Moreover, knockdown of CPSF1 induced an increase in alternative splicing (AS) events in addition to APA. Taken together, this study provides new insights into our understanding of the post-transcriptional regulatory mechanisms in HCC and implies that CPSF1 may be a potential prognostic biomarker and therapeutic target for HCC.

**Keywords:** hepatocellular carcinoma, alternative polyadenylation, cleavage and polyadenylation specific factor 1, cleavage and polyadenylation, alternative splicing

## INTRODUCTION

Liver cancer is one of the most aggressive malignancies worldwide, with 841,080 new cases and 781,631 cancer-related deaths reported in 2018 (Bray et al., 2018). According to GLOBOCAN statistics, the incidence and mortality rate of liver cancer were 905,677 and 830,180, respectively, in 2020. Hepatocellular carcinoma (HCC) constitutes approximately 80% of primary liver cancer cases and has extremely poor prognosis. Recent studies have identified many liver-specific oncogenes and

tumor suppressor genes that control HCC growth and metastasis, that have been driving research in targeted therapies (Chen et al., 2020). However, the improvement in HCC survival rate is still not acceptable; HCC ranked fourth as a leading cause of cancer-related deaths in 2018 because of poor response to systemic therapies. Therefore, it is necessary to continue searching for potential prognostic and therapeutic targets for HCC through enhancing our understanding of its cellular biological processes.

Post-transcriptional regulation is a common and vital regulatory mechanism that contributes to the diversity of gene expression. Owing to the rapid development of next-generation RNA sequencing, various post-transcriptional alterations in RNA processing have been discovered. In eukaryotic cells, the maturation of primary messenger RNAs (pre-mRNAs) involves 5' untranslated region (5' UTR) modification, removal of introns by splicing, and 3' untranslated region (3' UTR) cleavage and polyadenylation (CPA). Alternative polyadenylation (APA) is a general and evolutionarily conserved mechanism of transcriptional modification that involves endonucleolytic cleavage and addition of a polyadenosine tail of 200–250 nucleotides that generates different lengths of the 3'UTR from a single gene (Wang et al., 2008; Di Giammartino et al., 2011). APA at the RNA 3'UTR end of RNA results in various transcript isoforms that contribute to transcriptome diversity. Instead of affecting the protein-coding region, the 3' UTR region contains many recognition elements for microRNAs and RNA-binding proteins (RBPs) that affect the subcellular localization, intracellular trafficking, degradation, and translation rate of the transcripts in different cellular contexts (Mayr, 2016). Recent studies have revealed that APA is linked to changes in cellular states, including cancer and other diseases (Gruber and Zavolan, 2019). In HCC, a large number of APA events had been identified (Xiang et al., 2018). However, only a few studies have investigated the underlying mechanisms of APA and tumor progression. Thus, it is essential to identify the factors that control APA in HCC.

Alternative polyadenylation is mediated by a group of constant proteins that constitute the CPA multicomponent complexes, including four core components: cleavage and polyadenylation specific factors (CPSFs), cleavage stimulation factors (CSTFs), cleavage factor I (CFI), and CFII, along with numerous dynamic partners (Tian and Manley, 2017). CPSFs are site-specific cleavage subunits composed of CPSF1, CPSF2, CPSF3, CPSF4, and Fip1. As the largest component of the CPSFs complex, CPSF1 (also known as CPSF160) is primarily located in the nucleus and functions in pre-mRNA 3'UTR processing by recognizing a conserved AAUAAA polyadenylation signal near the polyadenylation site (pA site) (Murthy and Manley, 1995). Differential expression of CPSF1 has been reported to be closely associated with the progression of multiple human diseases. For example, CPSF1 binds to influenza virus NS1, resulting in the inhibition of wild-type influenza A virus replication (Du et al., 2020). CPSF1 expression decreases in diabetic retinopathy and mediates retinal vascular dysfunction *via* the MAPK/ERK pathway (Zhang et al., 2020). CPSF1 mutations are associated with early onset high myopia and involved in retinal ganglion cell axon projection (Ouyang et al., 2019).

In tumors, CPSF1 plays an oncogenic role in head and neck squamous cell carcinoma (Sakai et al., 2020), ovarian cancer (Zhang et al., 2017), and prostate cancer growth (Van Etten et al., 2017). However, its potential role in HCC has not been fully investigated.

In the present study, we investigated the clinical significance and biological function of CPSF1 in HCC. We found that CPSF1 was significantly upregulated in HCC. As a critical regulator of APA in HCC, CPSF1 accelerated cell proliferation and was correlated with unfavorable outcomes in patients with HCC. Our findings describe the oncogenic role of CPSF1 in HCC and suggest its potential role as a novel prognostic biomarker for HCC.

## MATERIALS AND METHODS

### Patients, Tissue Specimens, and Follow-Up

A total of 796 patients with primary HCC who underwent surgery between January 2000 and December 2010 at the Sun Yat-sen University Cancer Center (SYSUCC, Guangzhou, China) were included in this study. Paraffin-embedded tissue samples from these patients were re-embedded in new paraffin blocks for tissue microarrays (TMAs). Normal tissues were obtained from no less than 1 cm adjacent to tumor margin. This study was approved by the Institute Research Medical Ethics Committee of SYSUCC. All patients provided written informed consent for the use their tissues and data for research purposes. Pathological specimens and data were collected, and all samples were rendered anonymous. The overall survival (OS) follow-up period was defined as the interval from the date of surgery to the date of death or last follow-up. None of the patients received radiotherapy or chemotherapy prior to surgery.

### Hematoxylin–Eosin and Immunohistochemistry Staining

The HCC TMA blocks were cut into 4- $\mu$ m slices and mounted onto glass slides. These slides were then dewaxed and treated with 3% hydrogen peroxide in methanol and blocked with a Biotin-Blocking Kit (DAKO, Glostrup, Germany). HE and IHC staining of CPSF1 (1:200, Bethyl, A301-580) were then performed, and the slides were assessed by two independent pathologists who calculated the respective IHC scores. The IHC scores were determined by multiplying the staining intensity by the proportion of stained nuclei. The staining intensity was scored as four grades (0, 1, 2, and 3), which indicated a staining intensity from blank to strong staining. The proportion was scored as five grades (0, 1, 2, 3, and 4), which indicated a staining area of 0, 1–25, 25–50, 50–75, and 75–100%, respectively. The median IHC score was used as the cutoff value to separate patients into high and low CPSF1 expression groups.

### Cell Culture

Liver cell line QSG7701, and HCC cell lines HepG2.2.15 and HCCLM3 were obtained from the Cell Bank of Type

Culture Collection of Chinese Academy of Sciences Committee (Shanghai, China), and cultured in Dulbecco's modified Eagle's medium (Gibco, Grand Island, NY, United States) supplemented with 10% fetal bovine serum (Gibco, South America). All cells were grown in a humidified atmosphere at 37°C with 5% CO<sub>2</sub>.

## Generation of Cells With CPSF1 Overexpression and Knockdown

Full-length CPSF1 cDNA was cloned into the mammalian vector pENTER for overexpression and confirmed by sequencing. The plasmid was transfected into HCC cell lines using Lipofectamine<sup>TM</sup> 3000 reagent (Invitrogen, Carlsbad, CA, United States). For knockdown assays, small interfering RNAs (siRNAs) targeting CPSF1 were purchased from Shanghai GenePharma Co., Ltd. (Shanghai, China). Transfection was performed using Lipofectamine<sup>TM</sup> RNAiMAX (Invitrogen). The siRNA sequences used are listed in **Supplementary Table 1**.

## Protein Extraction and Western Blot Analysis

Total cellular proteins were extracted and resuspended in lysis buffer (Beyotime Biotechnology, Shanghai, China) supplemented with a protease inhibitor. Western blotting was performed using a standard protocol described previously (Chen et al., 2019). Briefly, total proteins (30 µg) were loaded in individual lanes and separated on an 8% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel. The following primary antibodies against the indicated proteins were used: CPSF1 (1:1,000, Bethyl, A301-580) and β-actin (1:2,000, Santa Cruz Biotechnology, Dallas, TX, United States). The secondary antibodies used were anti-rabbit (Cell Signaling Technology, 7074S, 1:3,000) and anti-mouse (Cell Signaling Technology, 7076S, 1:3,000), and both were purchased from Cell Signaling Technology. The protein band intensities were quantified using the Bio-Rad Molecular Imager ChemiDoc<sup>TM</sup> XRS + system.

## RNA Extraction and Quantitative Real-Time RT-PCR

Total RNA was isolated from fresh tissues and cell lines, and purified using TRIzol reagent (BIOO Scientific Co., Austin, TX, United States) according to the manufacturer's instructions. The mRNA was reverse-transcribed into cDNA using standard procedures with a Reverse Transcriptase Kit (Vazyme Biotech, Nanjing, China). SYBR Green-based quantitative real-time PCR (qRT-PCR) (Vazyme Biotech) was performed in 96-well plates. The mRNA levels were normalized to the expression of 18S mRNA. For APA validation, we used percentage of distal polyadenylation site usage index (PDUI) value by analyzing PDUI = long 3'UTR transcript/total transcript. For AS validation, we used percent spliced-in (PSI) value by analyzing PSI = skipping exon (SE) transcript/total transcript. Sequences of the oligonucleotide primers used for each transcript are listed in **Supplementary Table 1**.

## PacBio Sequencing and Transcriptome Construction by PacBio Reads

According to the Iso-Seq-v3 analysis pipeline<sup>1</sup>, subreads were used to call circular consensus sequences. Polished full-length non-concatemer (FLNC) reads were generated after primer, concatemer removal, and clustering. The FLNC reads were aligned against the reference genome (hg19/GRCh37) using GMAP ALIGNER (v2019-03-15) (Wu and Watanabe, 2005). Based on genome alignment, isoforms were collapsed by cDNA\_Cupcake<sup>2</sup>, followed by comparison with reference annotation (RefSeq hg19) using Cuffcompare (Trapnell et al., 2010). Isoforms with a complete match of the intron chain and potentially novel isoforms (with at least one known splice junction and occurring in both libraries) were selected. We then constructed a full transcriptome. The constructed transcriptome was used to identify alternative splicing (AS) events by SUPPA2 (Trincado et al., 2018). Seven types of AS events were identified, including SE, mutually exclusive exons (MX), alternative 5' splicing site (A5), alternative 3' splicing site (A3), retained intron (RI), alternative first exon (AF), and alternative last exon (AL).

## RNA-seq Data and Enrichment Analyses

Total RNA was extracted from HepG2.2.15 cells transfected with CPSF1 or scrambled siRNAs. Libraries were prepared using TruSeq Stranded mRNA Kits (Illumina) and subjected to 150 cycle paired-end sequencing on the Illumina HiSeq platform. Each group was sequenced in three replicates. Quality control and preprocessing of the sequencing data were performed using fastp-0.20.1 (Chen S. et al., 2018). Based on the constructed transcriptome, the abundance of transcripts was quantified by Salmon-1.0.0 (Patro et al., 2017) and gene models were generated in BED format. Wiggle files indicative of read coverage were generated by GenomeCoverageBed (BEDTools, v2.26.0) after sequence alignment by STAR-2.7.0f (Quinlan and Hall, 2010). Then, APA event analysis was performed using DaPars (version 0.9.1) to predict proximal polyadenylation sites and distal polyadenylation sites (Masamha et al., 2014). Significant APA events were determined using three criteria. First, the *P*-value of the PDUI differences should be <0.05. Second, the absolute mean difference of the PDUI must be ≥0.2. Third, the mean PDUI fold change must be ≥1.5. For functional analysis of the transcripts, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) gene sets were analyzed using DAVID<sup>3</sup>. For motif analysis, sequences between proximal and distal polyadenylation sites were collected in FASTA format and submitted to MEME suits<sup>4</sup>. For network analysis of functional protein association, CPSF1-associated proteins were filtered using experimental or curated data from the STRING database<sup>5</sup>. For APA data analyzed by DaPars, we downloaded the PDUI values of transcripts from<sup>6</sup>.

<sup>1</sup><https://github.com/PacificBiosciences/Iso-Seq>

<sup>2</sup>[https://github.com/Magdoll/cDNA\\_Cupcake](https://github.com/Magdoll/cDNA_Cupcake)

<sup>3</sup><https://david.ncifcrf.gov>

<sup>4</sup><http://meme-suite.org/>

<sup>5</sup><https://string-db.org/>

<sup>6</sup><http://tc3a.org/>



## Migration and Invasion Assays

Following transfection,  $5 \times 10^4$  cells were harvested and replated in the upper compartment of Transwell chambers (8- $\mu$ m pore size, Millipore, Billerica, MA, United States) in serum-free medium. For invasion assay, the upper chambers were coated with 50  $\mu$ L of 0.5 mg/mL Matrigel<sup>®</sup>. Fresh medium containing 10% fetal bovine serum was placed in the lower chamber. The fetal bovine serum-containing medium in the lower chamber serves as a chemoattractant for cell migration. After incubation for 36 h, the cells in the lower membrane were fixed with methanol for 30 min and then stained with 0.1% crystal violet. Three 20 $\times$  magnification fields were randomly chosen to count the cell number under a microscope.

## CCK8 and Colony Formation Assays

For Cell Counting Kit-8 (CCK8) assays, HCC cells were plated in 96-well plates and pretreated by transfection with CPSF1 plasmids or siRNAs. The cells were quantified on five consecutive days by incubating with the CCK8 detection reagent for 2 h at 37°C followed by colorimetric reading of absorbance at 450 nm using a microplate reader. For the colony formation assay, single-cell suspensions were plated in six-well plates. The culture medium was replaced every 72 h. After 14 days, the cells were fixed in methanol for 10 min, stained with crystal violet, and washed twice in PBS. Images were captured using a digital scanner to count the colonies.

## Anchorage-Independent Growth

Cultured HCC cells were plated in six-well plates and transfected with CPSF1 plasmids or siRNAs. Equal number of cells (1,000 cells) from each group was embedded in 2 mL of 0.35% agar solution and layered on top of 1.5 mL of 0.6% basal agar layer pre-coated in a six-well plate. Fresh medium (200  $\mu$ L) was added every 3 days. The plates were incubated at 37°C with 5% CO<sub>2</sub> for 3 weeks. Colony formation was determined by counting colonies larger than 50  $\mu$ m in 10 fields under the 4 $\times$  magnification under a microscope.

## Statistical Analyses

All data are presented as the mean  $\pm$  standard deviation. Statistical analyses were performed using SPSS (Version 20.0, IBM, Armonk, NY, United States), R (Version 3.6), and GraphPad PRISM software (Version 7.0, GraphPad Software Inc., La Jolla, CA, United States). Student's *t*-test, Pearson's  $\chi^2$  test, Kruskal–Wallis analysis, ANOVA analysis, Fisher's exact test, Kaplan–Meier method, and multivariate Cox proportional hazard regression models were used to analyze independent prognostic factors in OS and disease-free survival (DFS). *P* < 0.05 (two-tailed) was considered statistically significant.

## RESULTS

### CPSF1 Is Highly Expressed in HCC Tissues

First, we examined the genome alterations and expression of 22 major CPA factors that drive APA processing in HCC. As

indicated by the HCC patient dataset from The Cancer Genome Atlas (TCGA), CPSF1 mRNA was significantly increased in HCC compared to that in normal liver tissues (**Figure 1A**). Corresponding genome copy number data indicated that CPSF1 amplification correlated with its mRNA upregulation (**Figures 1A,B**). In 50 paired HCC tissues from the TCGA database, CPSF1 mRNA was relatively elevated in tumor tissues compared to the paired non-tumor tissues (**Figure 1C**). To further confirm the elevated mRNA expression of CPSF1 in HCC, we searched the data in other public databases. We found that CPSF1 mRNA expression was significantly increased in tumor tissues compared to that in normal tissues in the GEO Roessler liver dataset (**Figures 1D,E**) that was validated in 24 paired HCC samples collected at the SYSUCC (**Figure 1F**).

### High CPSF1 Expression Correlates With Poor Patient Prognosis in HCC

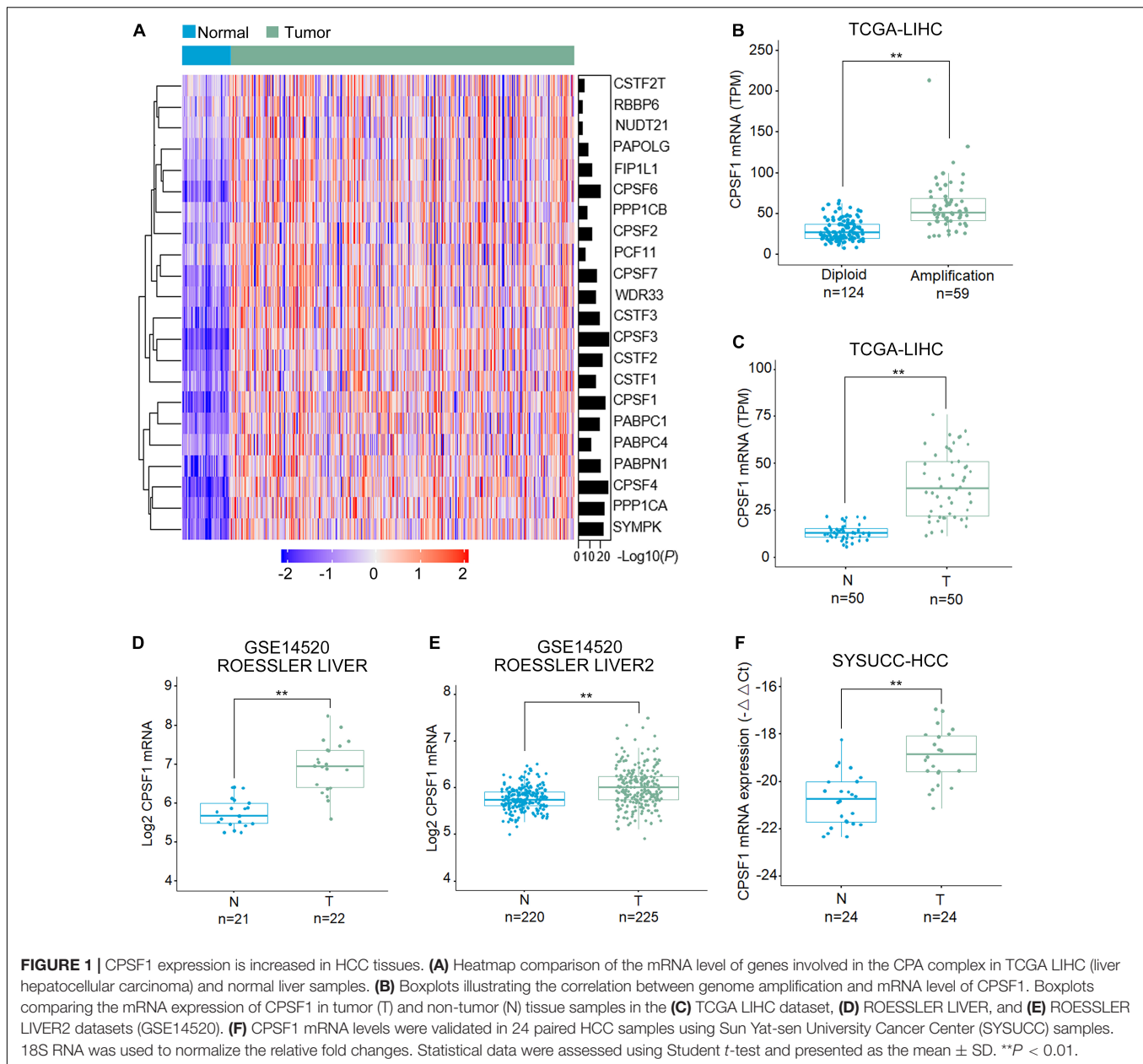
In terms of its expression and clinical significance, we found that CPSF1 mRNA was elevated from grade 1 to grade 3 (**Figure 2A**). Next, TMAs containing 786 HCC cases with complete clinicopathological data were evaluated by performing IHC staining for CPSF1 expression (**Supplementary Table 2**). CPSF1 was primarily localized in the nucleus. Statistical analysis revealed relatively higher expression of CPSF1 in tumor samples than in paired non-tumor samples (**Figure 2B**).

To identify the correlation between CPSF1 expression and patient outcomes, TCGA data were categorized into high and low groups based on the median CPSF1 mRNA level. Furthermore, the median IHC score of CPSF1 was selected as the cutoff value to group the HCC patients into two groups in the SYSUCC cohort. Survival analyses revealed that CPSF1 overexpression was significantly correlated with worse OS and DFS in TCGA patients (**Figure 2C**) and SYSUCC cohort (**Figure 2D**). Furthermore, CPSF1 was statistically associated with poor differentiation and relapse clinicopathological parameters (**Supplementary Table 2**). Stratified survival analysis revealed that the high CPSF1 group was associated with shorter OS in the larger tumor size ( $\geq 5$  cm), single or multiple tumors, high AFP levels, poor differentiation, TNM grades, and HBV-positive cases in **Supplementary Figure 1**. Univariate and multivariate Cox regression analyses indicated that CPSF1 overexpression was an independent predictor of OS (*P* < 0.001) (**Supplementary Table 3**). Taken together, these results indicate that an increase in CPSF1 expression is a prognostic hallmark of HCC.

### Silencing of CPSF1 Inhibits Cell Growth in HCC

To explore the biological function of CPSF1 in HCC cells, we depleted CPSF1 expression by transfecting two mixed siRNAs into HepG2.2.15 and HCCLM3 cell lines (**Figures 3A,B**). Compared to the control cells, silencing of CPSF1 decreased cell proliferation and colony formation as revealed by CCK8 and colony formation assays, respectively (**Figures 3C,E**). In contrast to CPSF1 silencing, transfection of a CPSF1 ectopic-expression vector into HepG2.2.15 and HCCLM3 cells increased cell growth (**Figures 3D,E**).





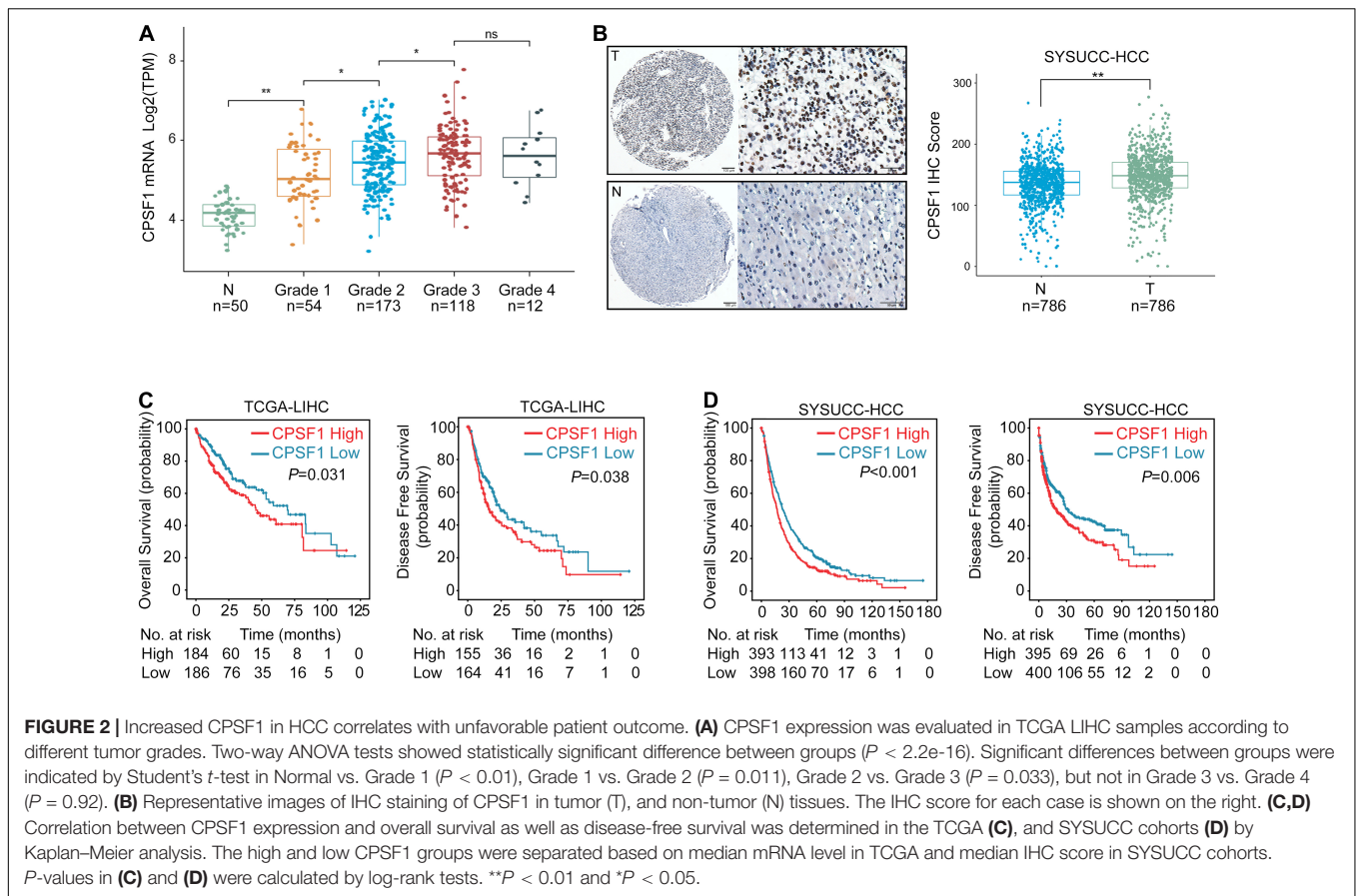
These results were further supported by anchorage-independent growth assays (**Figure 3F**). In cell cycle assays, silencing of CPSF1 increased G0/G1 and reduced G2/M cell proportion, whereas CPSF1 overexpression increased the proportion of G2/M cells in HepG2.2.15 cells (**Figure 3G**). To further demonstrate that CPSF1 can potentially trigger tumorigenesis in a non-transformed setting, we silenced or overexpressed CPSF1 in the normal liver epithelial cell line, QSG7701. The results indicated that overexpression of CPSF1 promoted cell proliferation and colony formation, whereas knockdown of CPSF1 reduced cell growth in QSG-7701 cells (**Supplementary Figure 2**).

Next, we analyzed the impact of CPSF1 on cell movement. Silencing of CPSF1 inhibited cell migration

and invasion, whereas CPSF1 overexpression increased cell migration and invasion, as detected by Transwell assays (**Figure 4**). These results demonstrated that CPSF1 serves key functions in HCC oncogenesis with a potential role in cell proliferation and metastasis.

### CPSF1 Modulates 3'UTR Alteration in HCC Cells

As CPSF1 is a key cleavage factor responsible for mRNA processing, we examined whether CPSF1 alters polyA site choices, at least in part, leading to changes in 3' UTR length and RNA transcription. We performed long-read sequencing and constructed a transcriptome.



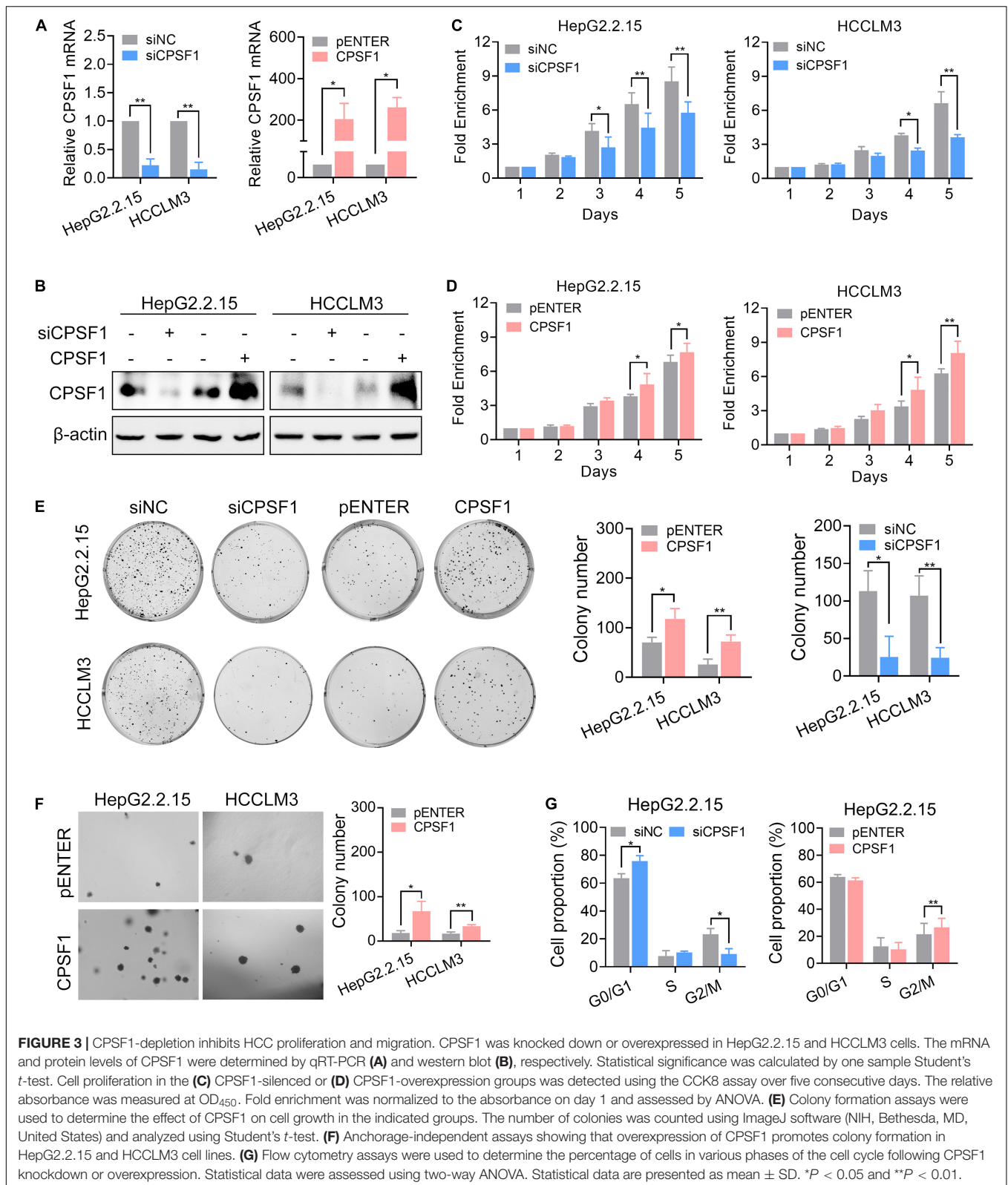
We performed ISO- and RNA-seq in control and CPSF1-siRNA transfected groups. The effect of CPSF1 siRNA transfection was confirmed by RNA-seq data (**Supplementary Figure 3**). A total of 9,595 transcripts were identified based on the Iso-Seq pipeline analysis, and these transcripts were used to construct the transcriptome. Short reads were mapped to the hg19 reference genome and used to quantify gene expression based on the constructed transcriptome. Knockdown of CPSF1 resulted in 707 upregulated and 262 downregulated transcripts, filtered by a fold change  $\geq 1.5$ , and a  $P < 0.05$  (**Figure 5A**). Gene set overrepresentation analysis identified significantly enriched KEGG pathways related to metabolism, cell cycle, protein processing, and nucleotide excision (**Figure 5B**) and GO biological processes such as cell division, protein folding, and mRNA splicing (**Figure 5C**). We performed GO functional enrichment analysis on the up- and downregulated transcripts separately. Results further indicated that the downregulated transcripts were mainly related to cell cycle, whereas the upregulated transcripts were related to RNA processing and protein folding (**Supplementary Figure 4**). The top 10 representative up- and downregulated transcripts are listed in **Figure 5D**.

Next, we analyzed the RNA-seq data in the control and CPSF1-siRNA transfected groups using DaPars algorithm to identify 3' UTR alteration profiles. The significantly differential APA genes were determined by three criteria:  $P < 0.05$ ,

$\Delta$ PDUI  $\geq 0.2$ , and absolute PDUI fold change  $\geq 1.5$ . Consequently, 36 shortened and seven lengthened transcripts were identified, of which 40 were protein-coding and three were non-coding transcripts (**Figure 6A**). Interestingly, previous reports indicated that CPSF1 binds to the canonical AAUAAA or similar motifs AUUAAA located 10–50 nt upstream of the 3' UTR. Analysis of the motif distribution of these 43 transcripts indicated a preferential high enrichment in the distal polyadenylation sites (**Figure 6B**), further suggesting that CPSF1 preferentially selects distal compared to proximal polyadenylation sites.

To test the correlation between transcript alteration and APA, we combined the analyses of altered mRNA levels and PDUI (**Figure 6C**). Upon depletion, we found that 15 transcripts were filtered with both altered mRNA expression and PDUI (**Figures 6C–E**), including four lengthened (REEP5, ZNF271P, TUBA1A, and MED29) and 11 shortened transcripts. Interestingly, in the 3' UTR shortened transcripts, three upregulated and eight downregulated transcripts were observed, while 3' UTR lengthening was upregulated. These included 13 protein-coding and two non-coding transcripts (ZNF271P and ARL6IP6) that were separated into three functions: gene transcription, endomembrane system, and cytoskeleton maintenance (**Figure 6F**).

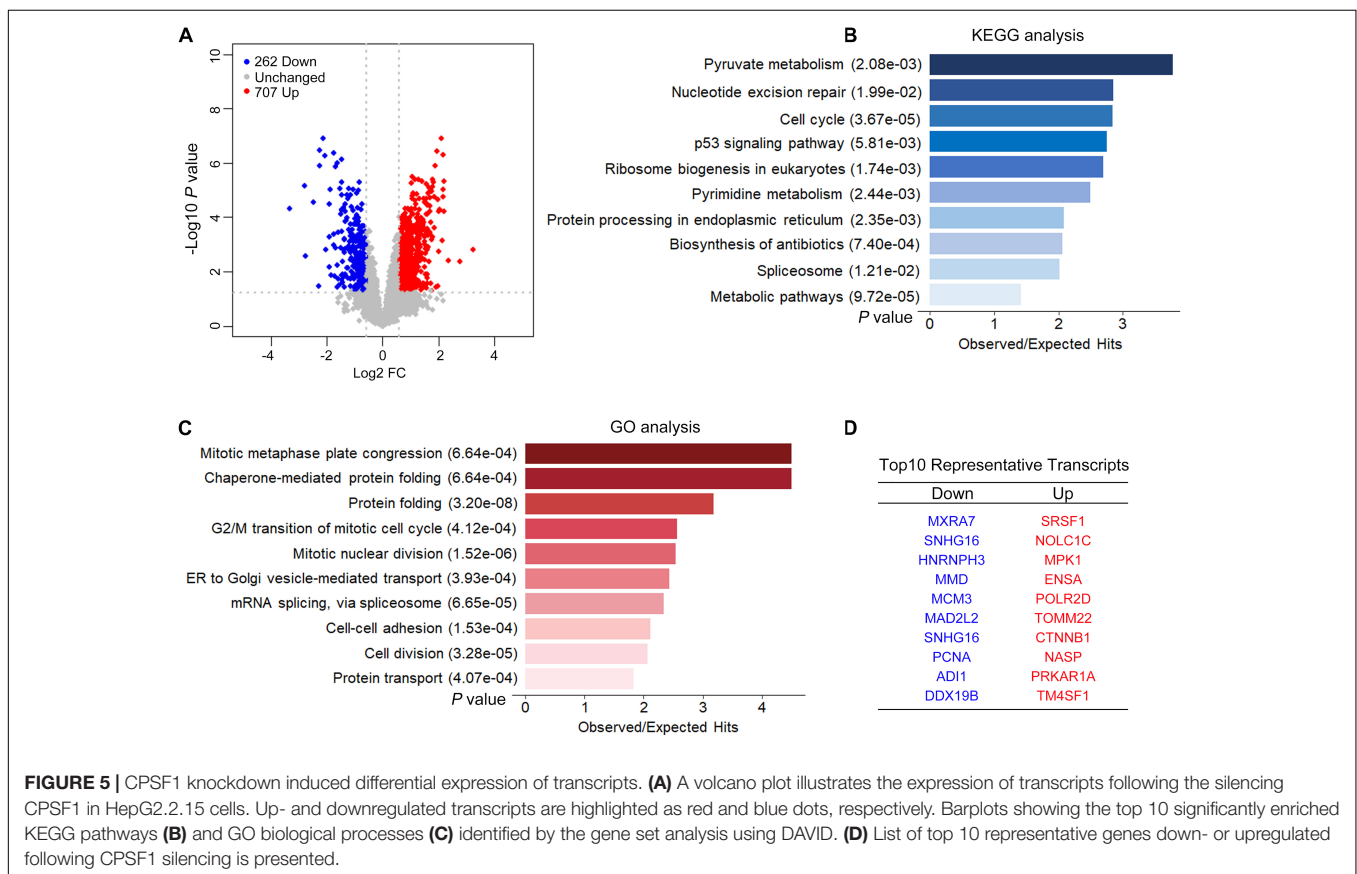
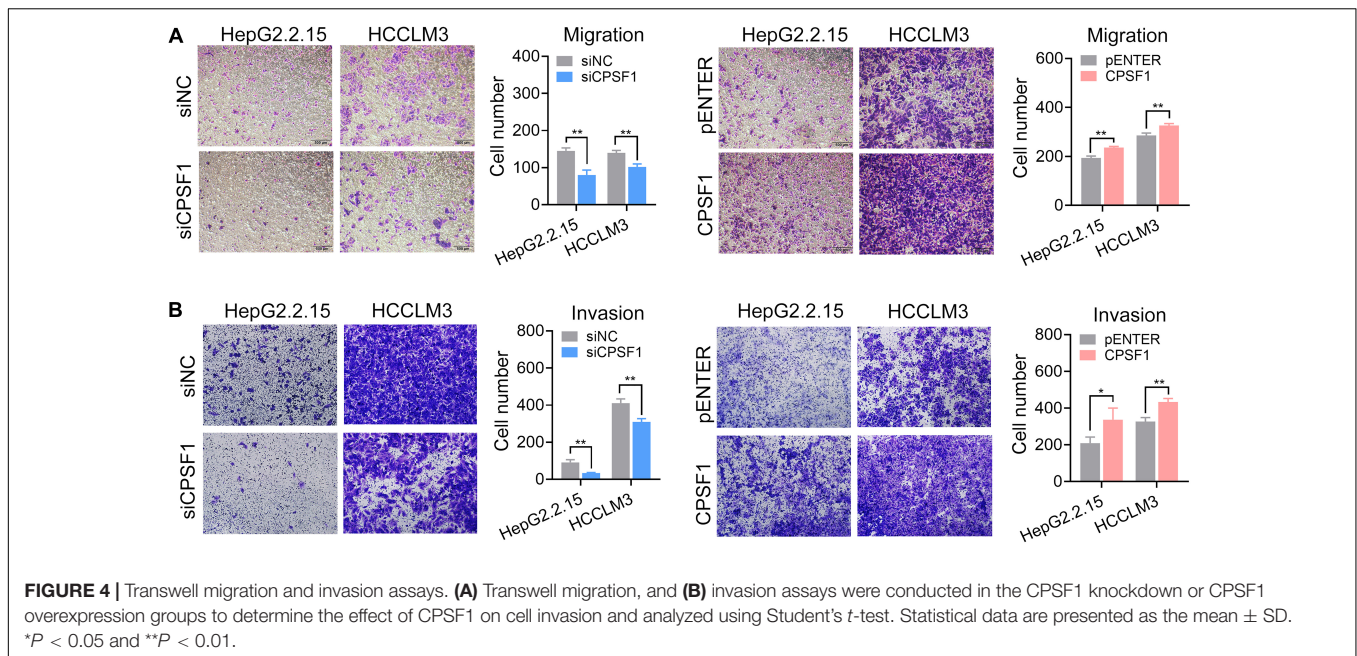
To validate these genes related to CPSF1 expression, we separated the TCGA APA events by CPSF1 mRNA median



expression score (high vs. low) to compare APAs between patient groups. Analysis of the 15 transcripts regulated by CPSF1 in our RNA-seq data (Figure 6E) revealed four overlapping genes

(RNF14, FAM177A1, MED29, and TUBA1A) (Supplementary Figure 5). These results indicate that complex regulatory mechanisms are involved in CPSF1-related gene alteration.

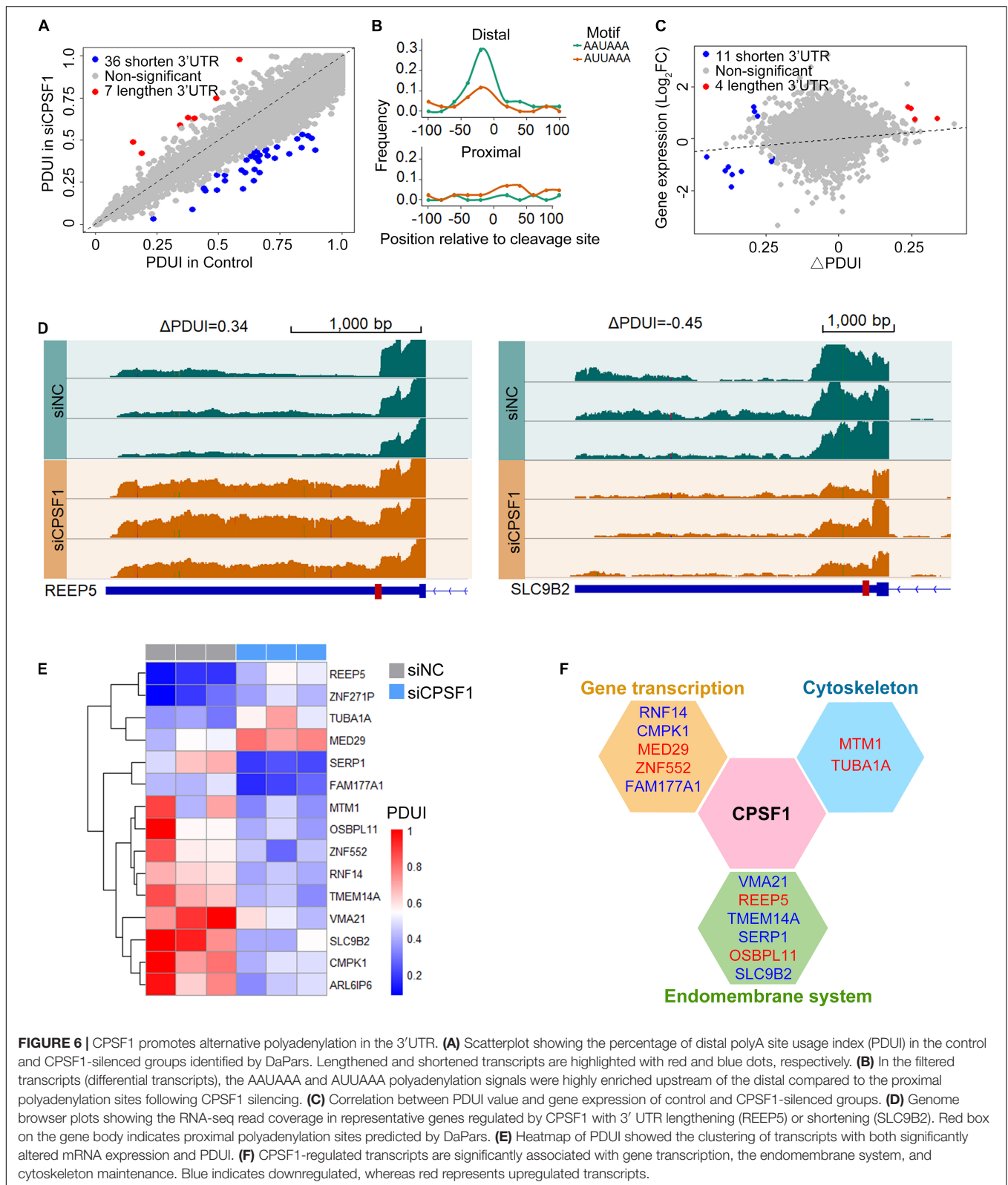




## CPSF1 Modulates Alternative Splicing in HCC Cells

In addition to APA, AS is a pre-mRNA processing event that leads to the diversity and maturation of RNA. Using GSEA of TCGA

LIHC data, we found that CPSF1 positively correlated with the function of spliceosome, suggesting a possible role of CPSF1 in AS, in addition to APA (Figure 7A). To test this hypothesis, we further examined the effect of CPSF1 on AS in our RNA-seq



**FIGURE 6 |** CPSF1 promotes alternative polyadenylation in the 3'UTR. **(A)** Scatterplot showing the percentage of distal polyA site usage index (PDI) in the control and CPSF1-silenced groups identified by DaPars. Lengthened and shortened transcripts are highlighted with red and blue dots, respectively. **(B)** In the filtered transcripts (differential transcripts), the AAUAAA and AUUAAA polyadenylation signals were highly enriched upstream of the distal compared to the proximal polyadenylation sites following CPSF1 silencing. **(C)** Correlation between PDI value and gene expression of control and CPSF1-silenced groups. **(D)** Genome browser plots showing the RNA-seq read coverage in representative genes regulated by CPSF1 with 3' UTR lengthening (REEP5) or shortening (SLC9B2). Red box on the gene body indicates proximal polyadenylation sites predicted by DaPars. **(E)** Heatmap of PDI showed the clustering of transcripts with both significantly altered mRNA expression and PDI. **(F)** CPSF1-regulated transcripts are significantly associated with gene transcription, the endomembrane system, and cytoskeleton maintenance. Blue indicates downregulated, whereas red represents upregulated transcripts.

data. Surprisingly, a total of 3,138 splicing events of seven types were identified, among which skipping exon (SE) was the most abundant type (Figure 7B and Supplementary Data 1). Next,

we filtered the differentially expressed AS events according to  $P < 0.05$  and identified 197 significant events that met these criteria (Figure 7C).





the expression of the transcripts due to various regulatory mechanisms involved at the mRNA level. For example, changes in the phosphorylation status of RNA polymerase II (RNAPII) affect its elongation rate within the transcription unit and its pausing at the end of 3' UTR (Tian and Manley, 2017). In addition, the *cis*-elements in the intron can induce longer 3' ends in histone-globin chimeric genes or enhance polyadenylation rates (Niwa et al., 1990; Pandey et al., 1990). Another reason is that post-transcriptional factors such as microRNAs and accessory RBPs bind to elements located near the regulated polyadenylation sites (Zhu et al., 2007; An et al., 2013). To confirm the hypothesis that microRNAs and RBPs are involved in the regulation of the 3' UTRs of the CPSF1-altered genes, we analyzed the sequences of the lengthened and shortened transcripts between distal and proximal polyadenylation sites. Using MEME suit analysis, we identified different motifs for lengthened and shortened transcripts. Tomtom analysis revealed that different RBPs and microRNAs bound to this region (Gupta et al., 2007; **Supplementary Figure 6A** and **Supplementary Data 2**), supporting motif analysis in the TargetScan database (**Supplementary Figure 6B**). Among the RBPs predicted by Tomtom analysis mentioned above, we found HNRNPL and HNRNPK that also interact with CPSF1 in the string database (**Supplementary Data 3**). These results suggested that besides the conserved motifs (AAUAAA or AUUAAA), other motifs within the 3' UTR transcripts may affect the binding of CPSF1.

Accumulating evidence suggests that AS and APA are not independent processes. The spliceosome and CPA complex are intimately interconnected and functionally intertwined (Martinson, 2011). For example, U1 snRNP, a major component of the spliceosome, protected pre-mRNAs from premature APA splicing. Distinct from its role in splicing, this resulted in differences in mRNA length and isoform expression (Kaida et al., 2010). CPA positions trending to usage of proximal polyadenylation sites with a decrease in U1 level yielded mRNAs with shorter 3' UTRs and alternatively spliced isoforms (Berg et al., 2012). Several splicing factors have been found to influence polyadenylation (Millevoi et al., 2006). CPSF1 plays a non-canonical role in AS regulation because the CPSF complex usually interacts with the spliceosome (Lutz et al., 1996; Kyburz et al., 2006). Although CPSF1 is the main component of the CPSF complex, it serves AS roles in addition to regulating the APA process. For example, CPSF1 promotes head and neck squamous cell carcinoma growth by regulating AS in cancer-associated genes, such as AKT2, HRAS, TGFBI, and UBE2C (Sakai et al., 2020). CPSF1 causes a splicing switch to inhibit the expression of AR variants and blocks the androgen-independent growth of CRPC cells (Van Etten et al., 2017). CPSF1 regulates the AS of exon 6 in interleukin 7 receptor (IL7R), which is essential for T-cell development and maintenance (Evsyukova et al., 2013). AS cleavage and APA are significant contributors to mRNA transcriptome diversity (Wang et al., 2008; Di Giammartino et al., 2011). These reactions are carried out by the spliceosome and CPA macromolecular complex composed of constant proteins along with the participation of numerous dynamic partners. We chose one representative gene, CMPK1, to validate the RNA-seq results as it was altered in both APA and

AS following knockdown of CPSF1 (**Supplementary Figure 7**). CCK8 results indicated that overexpression of CMPK1 rescued the effect of CPSF1 knockdown on the cellular phenotype (**Supplementary Figure 7**).

## CONCLUSION

Our results provide novel insights into the crucial role of the CPA core factor, CPSF1, and its related aberrant APA as well as AS events in HCC. Although further detailed examination is required to understand the exact mechanism underlying CPSF1-mediated cell growth and APA regulation in HCC, our results suggest that CPSF1 serves as an oncogene, and novel prognostic biomarker and potential therapeutic target for patients with HCC.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: SRA, PRJNA682912.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institute Research Medical Ethics Committee of Sun Yat-sen University Cancer Center. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SC, XW, and JY: conception and design of the study. SC, XY, YH, and MY: *in vitro* experiments. LL and XY: scoring and evaluation of IHC stained slides. ZZ and XW: RNA-seq analysis. SC, XW, and JY: drafting and revision of the manuscript. All authors approved the final version of the manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 81872012, 81802384, 82072611, and 82072853), the National Natural Science Foundation of Guangdong Province (No. 2018B030311005), a grant from Guangdong Basic and Applied Basic Research Foundation (2019B030302012), the Research Grants Council of the Hong Kong Special Administrative Region (Nos. CityU11102317, 11103718, 11103619, R4017-18, and C4041-17GF), the Certificate of China Postdoctoral Science Foundation Grant (No. 2019M663285), and China Postdoctoral Innovative Talent Support Program (No. BX20190392).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.616835/full#supplementary-material>

**Supplementary Figure 1** | Stratified analysis of CPSF1 expression and overall survival in HCC. Stratified analysis of the overall survival of patients with HCC and CPSF1 expression in different subgroups. *P*-values are derived from log-rank tests.

**Supplementary Figure 2** | CPSF1-depletion inhibits liver cell line proliferation. CPSF1 was silenced by transfection with CPSF1 siRNAs or overexpressed with plasmids in QSG7701 cells. **(A)** CPSF1 protein levels were determined by western blotting. Statistical significance was calculated using Student's *t*-test. **(B)** Cell proliferation in the CPSF1-silenced or CPSF1-overexpression groups was detected by CCK8 assay over five consecutive days. The relative absorbance was measured at OD<sub>450</sub>. Fold enrichment was normalized to the absorbance on day 1 and assessed using two-way ANOVA. **(C)** Colony formation assays determined the effect of CPSF1 on cell growth. The number of colonies was counted using ImageJ software (NIH, Bethesda, MD, United States) and analyzed using Student's *t*-test. Statistical data are presented as mean ± SD. \**P* < 0.05 and \*\**P* < 0.01.

**Supplementary Figure 3** | CPSF1-expression in RNA-seq data. **(A)** Genome browser plots showing the RNA-seq read coverage of CPSF1. **(B)** Gene expression of CPSF1 in the RNA-seq data was analyzed using Student's *t*-test. Statistical data are presented as mean ± SD. \**P* < 0.05.

## REFERENCES

- An, J., Zhu, X., Wang, H., and Jin, X. A. (2013). Dynamic interplay between alternative polyadenylation and microRNA regulation: implications for cancer (Review). *Int. J. Oncol.* 43, 995–1001. doi: 10.3892/ijo.2013.2047
- Berg, M. G., Singh, L. N., Younis, I., Liu, Q., Pinto, A. M., Kaida, D., et al. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53–64. doi: 10.1016/j.cell.2012.05.029
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., Jemal, A. et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492
- Chen, B., Garmire, L., Calvisi, D. F., Chua, M. S., Kelley, R. K., Chen, X. et al. (2020). Harnessing big 'omics' data and AI for drug discovery in hepatocellular carcinoma. *Nat. Rev. Gastroenterol. Hepatol.* 17, 238–251. doi: 10.1038/s41575-019-0240-9
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Chen, S. L., Zhang, C. Z., Liu, L. L., Lu, S. X., Pan, Y. H., Wang, C. H., et al. (2019). A GYS2/p53 negative feedback loop restricts tumor growth in HBV-related hepatocellular carcinoma. *Cancer Res.* 79, 534–545. doi: 10.1158/0008-5472.CAN-18-2357
- Chen, X., Zhang, J. X., Luo, J. H., Wu, S., Yuan, G. J., Ma, N. F., et al. (2018). CSTF2-Induced Shortening of the RAC1 3'UTR Promotes the Pathogenesis of Urothelial Carcinoma of the Bladder. *Cancer Res.* 78, 5848–5862. doi: 10.1158/0008-5472.CAN-18-0822
- Di Giammartino, D. C., Nishida, K., and Manley, J. L. (2011). Mechanisms and consequences of alternative polyadenylation. *Mol. Cell.* 43, 853–866. doi: 10.1016/j.molcel.2011.08.017
- Du, Y., Hultquist, J. F., Zhou, Q., Olson, A., Tseng, Y., Zhang, T. H., et al. (2020). mRNA display with library of even-distribution reveals cellular interactors of influenza virus NS1. *Nat. Commun.* 11:2449. doi: 10.1038/s41467-020-16140-9
- Evsyukova, I., Bradrick, S. S., Gregory, S. G., and Garcia-Blanco, M. A. (2013). Cleavage and polyadenylation specificity factor 1 (CPSF1) regulates alternative splicing of interleukin 7 receptor (IL7R) exon 6. *RNA* 19, 103–115. doi: 10.1261/rna.035410.112
- Fu, Y., Sun, Y., Li, Y., Li, J., Rao, X., Chen, C., et al. (2011). Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome. Res.* 21, 741–747. doi: 10.1101/gr.115295.110
- Gruber, A. J., and Zavolan, M. (2019). Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* 20, 599–614. doi: 10.1038/s41576-019-0145-z
- Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. S. (2007). Quantifying similarity between motifs. *Genome. Biol.* 8:R24. doi: 10.1186/gb-2007-8-2-r24
- Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., et al. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668. doi: 10.1038/nature09479
- Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Mol. Cell.* 23, 195–205. doi: 10.1016/j.molcel.2006.05.037
- Lam, S. H., and Gong, Z. (2006). Modeling liver cancer using zebrafish: a comparative oncogenomics approach. *Cell. Cycle* 5, 573–577. doi: 10.4161/cc.5.6.2550
- Lee, S. H., Singh, I., Tisdale, S., Abdel-Wahab, O., Leslie, C. S., Mayr, C. et al. (2018). Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* 561, 127–131. doi: 10.1038/s41586-018-0465-8
- Lutz, C. S., Murthy, K. G., Schek, N., O'Connor, J. P., Manley, J. L., Alwine, J. C. et al. (1996). Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency *in vitro*. *Genes. Dev.* 10, 325–337. doi: 10.1101/genad.10.3.325
- Martinson, H. G. (2011). An active role for splicing in 3'-end formation. *Wiley Interdiscip. Rev. RNA* 2, 459–470. doi: 10.1002/wrna.68
- Masamha, C. P., Xia, Z., Yang, J., Albrecht, T. R., Li, M., Shyu, A. B., et al. (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* 510, 412–416. doi: 10.1038/nature13261
- Mayr, C. (2016). Evolution and biological roles of alternative 3'UTRs. *Trends Cell Biol.* 26, 227–237. doi: 10.1016/j.tcb.2015.10.012
- Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S. Z., Keller, W., Antoniou, M., et al. (2006). An interaction between U2AF 65 and CF Im links the splicing and 3' end processing machineries. *EMBO J.* 25, 4854–4864. doi: 10.1038/sj.emboj.7601331
- Morris, A. R., Bos, A., Diosdado, B., Rooijers, K., Elkon, R., Bolijn, A. S., et al. (2012). Alternative cleavage and polyadenylation during colorectal cancer

- development. *Clin. Cancer Res.* 18, 5256–5266. doi: 10.1158/1078-0432.CCR-12-0543
- Murthy, K. G., and Manley, J. L. (1995). The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes. Dev.* 9, 2672–2683. doi: 10.1101/gad.9.21.2672
- Niwa, M., Rose, S. D., and Berget, S. M. (1990). *In vitro* polyadenylation is stimulated by the presence of an upstream intron. *Genes. Dev.* 4, 1552–1559. doi: 10.1101/gad.4.9.1552
- Ouyang, J., Sun, W., Xiao, X., Li, S., Jia, X., Zhou, L., et al. (2019). CPSF1 mutations are associated with early-onset high myopia and involved in retinal ganglion cell axon projection. *Hum. Mol. Genet.* 28, 1959–1970. doi: 10.1093/hmg/ddz029
- Pandey, N. B., Chodchoy, N., Liu, T. J., and Marzluff, W. F. (1990). Introns in histone genes alter the distribution of 3' ends. *Nucleic Acids Res.* 18, 3161–3170. doi: 10.1093/nar/18.11.3161
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Sakai, A., Ando, M., Fukusumi, T., Ren, S., Liu, C., Qualliotine, J., et al. (2020). Aberrant expression of CPSF1 promotes head and neck squamous cell carcinoma via regulating alternative splicing. *PLoS One* 15:e0233380. doi: 10.1371/journal.pone.0233380
- Sun, M., Ding, J., Li, D., Yang, G., Cheng, Z., Zhu, Q. et al. (2017). NUDT21 regulates 3'-UTR length and microRNA-mediated gene silencing in hepatocellular carcinoma. *Cancer Lett.* 410, 158–168. doi: 10.1016/j.canlet.2017.09.026
- Tan, S., Li, H., Zhang, W., Shao, Y., Liu, Y., Guan, H., et al. (2018). NUDT21 negatively regulates PSMB2 and CXXC5 by alternative polyadenylation and contributes to hepatocellular carcinoma suppression. *Oncogene* 37, 4887–4900. doi: 10.1038/s41388-018-0280-6
- Tian, B., and Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell. Biol.* 18, 18–30. doi: 10.1038/nrm.2016.116
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Trincado, J. L., Entizne, J. C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D. J., et al. (2018). SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome. Biol.* 19:40. doi: 10.1186/s13059-018-1417-1
- Van Etten, J. L., Nyquist, M., Li, Y., Yang, R., Ho, Y., Johnson, R., et al. (2017). Targeting a single alternative polyadenylation site coordinately blocks expression of androgen receptor mRNA splice variants in prostate cancer. *Cancer Res.* 77, 5228–5235. doi: 10.1158/0008-5472.CAN-17-0320
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi: 10.1038/nature07509
- Wu, T. D., and Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875. doi: 10.1093/bioinformatics/bti310
- Xiang, Y., Ye, Y., Lou, Y., Yang, Y., Cai, C., Zhang, Z., et al. (2018). Comprehensive characterization of alternative polyadenylation in human cancer. *J. Natl. Cancer. Inst.* 110, 379–389. doi: 10.1093/jnci/djx223
- Zhang, B., Liu, Y., Liu, D., and Yang, L. (2017). Targeting cleavage and polyadenylation specific factor 1 via shRNA inhibits cell proliferation in human ovarian cancer. *J. Biosci.* 42, 417–425. doi: 10.1007/s12038-017-9701-x
- Zhang, J., Zhang, X., Zou, Y., and Han, F. (2020). CPSF1 mediates retinal vascular dysfunction in diabetes mellitus via the MAPK/ERK pathway. *Arch. Physiol. Biochem.* 2020, 1–8. doi: 10.1080/13813455.2020.1722704
- Zhu, H., Zhou, H. L., Hasman, R. A., and Lou, H. (2007). Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. *J. Biol. Chem.* 282, 2203–2210. doi: 10.1074/jbc.M609349200

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Chen, Zhu, Yang, Liu, He, Yang, Guan, Wang and Yun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.