

Article

U1 snRNP proteins promote proximal alternative polyadenylation sites by directly interacting with 3' end processing core factors

Zhijie Hu^{1,†}, Mengxia Li^{1,†}, Zhanfeng Huo^{1,†}, Liutao Chen¹, Susu Liu¹, Ke Deng¹, Xin Lu¹, Shangwu Chen¹, Yonggui Fu^{1,*}, and Anlong Xu^{1,2,*} 

¹ State Key Laboratory of Biocontrol, Guangdong Province Key Laboratory of Pharmaceutical Functional Genes, Department of Biochemistry, School of Life Sciences, Sun Yat-sen University, Higher Education Mega Center, Guangzhou 510006, China

² School of Life Sciences, Beijing University of Chinese Medicine, Beijing 100029, China

[†] These authors contributed equally to this work.

* Correspondence to: Anlong Xu, E-mail: lssxl@mail.sysu.edu.cn; Yonggui Fu, E-mail: fuyg@mail.sysu.edu.cn

Edited by Zefeng Wang

In eukaryotic cells, both alternative splicing and alternative polyadenylation (APA) play essential roles in the gene regulation network. U1 small ribonucleoprotein particle (U1 snRNP) is a major component of spliceosome, and U1 snRNP complex can suppress proximal APA sites through crosstalk with 3' end processing factors. However, here we show that both knockdown and overexpression of SNRPA, SNRPC, SNRNP70, and SNRPD2, the U1 snRNP proteins, promote the usage of proximal APA sites at the transcriptome level. SNRNP70 can drive the phase transition of PABPN1 from droplet to aggregate, which may reduce the repressive effects of PABPN1 on the proximal APA sites. Additionally, SNRNP70 can also promote the proximal APA sites by recruiting CPSF6, suggesting that the function of CPSF6 on APA is related with other RNA-binding proteins and cell context-dependent. Consequently, these results reveal that, on the contrary to U1 snRNP complex, the free proteins of U1 snRNP complex can promote proximal APA sites through the interaction with 3' end processing machinery.

Keywords: U1 snRNP, 3' end processing factors, alternative polyadenylation, phase separation

Introduction

Maturation of precursor mRNA (pre-mRNA) in eukaryotic cells undergoes a series of co-transcriptional processing, including exon splicing and 3' end cleavage and polyadenylation (Licatalosi and Darnell, 2010). The 3' end processing machinery recognizes poly(A) signal (PAS) on pre-mRNA, then cleaves it and adds a poly(A) tail. Most genes contain multiple PASs, by which distinct mRNA isoforms with different lengths can be transcribed out, referred to as alternative polyadenylation (APA) (Shi, 2012; Tian and Manley, 2017). APA plays an important role in gene regulation network by increasing transcript and protein diversities (Fu et al., 2011; Mayr, 2016; Tian and Manley, 2017) and has been found to be related with many biology functions, such as development, tumorigenesis, morphogenesis, immune response, and neurons activation

(Li et al., 2012; Elkon et al., 2013; Carpenter et al., 2014; Jia et al., 2017; Gruber and Zavolan, 2019).

U1 snRNP complex, one of the important components of spliceosome, has much higher content in cells than other splicing components (U2, U4, U5, and U6 snRNPs) (Wahl et al., 2009; Wan et al., 2019). It consists of a non-coding small nuclear RNA (U1 snRNA), three specific associated proteins (SNRPA, SNRPC, and SNRNP70), and seven common Sm core proteins (SNRPB/B', D1, D2, D3, E, F, and G) (Guio and O'Reilly, 2015; Zhang et al., 2021). U1 snRNP is associated with tumorigenesis (Oh et al., 2020), embryonic stem cell (ESC) maintenance (Kainov and Makeyev, 2020), and the development of some neurological diseases, such as amyotrophic lateral sclerosis (Chi et al., 2018) and Alzheimer's disease (Bai, 2018).

Besides alternative splicing, U1 snRNP complex has also been found to be involved in regulation of APA (Kaida, 2016). U1 snRNP complex can suppress the usage of proximal poly(A) sites (Kaida et al., 2010; Oh et al., 2017; Venters et al., 2019). Several mechanisms have been proposed with protein–RNA, RNA–RNA, and protein–protein interactions. U1 snRNP complex binds to the pre-mRNA 5' splice site (5' ss) through RNA–RNA interaction of U1 snRNA and pre-mRNA, and then SNRPA in U1 snRNP

Received February 24, 2022. Revised July 14, 2022. Accepted September 6, 2022.

© The Author(s) (2022). Published by Oxford University Press on behalf of *Journal of Molecular Cell Biology*, CEMCS, CAS.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

complex can compete with CSTF2 to bind with poly(A) site, repressing the usage of adjacent cryptic poly(A) site (Shi et al., 2019; Deng et al., 2020). While binding to 5' ss, U1 snRNP complex can also directly interact with 3' end processing factors. It was found that the N-terminal region of SNRPA and SNRNP70 can crosstalk with poly(A) polymerase (PAP) and affect the catalytic polyadenylating activity of PAP (Gunderson et al., 1994, 1998). Moreover, the core proteins of U1 snRNP complex and the cleavage and polyadenylation factors (CPAFs) form the complex U1–CPAFs at intronic PASs and suppress the cryptic poly(A) site (So et al., 2019). Interestingly, U1 snRNP-free proteins that function independently of U1 snRNP complex can also participate in APA regulation (Ma et al., 2006; Workman et al., 2014). U1 snRNP-free SNRPA can interfere with 3' end processing machinery by steric hindrance effect. It was found that SNRPA could directly bind to 3' UTR adjacent of poly(A) sites of itself (Boelens et al., 1993) and SMN gene (Workman et al., 2014) and then inhibit polyadenylation of pre-mRNA. Additionally, non-snRNP SNRPA was down-regulated in differentiated B cells and reduced the repression of IgM secretory poly(A) site usage (Ma et al., 2006).

Nevertheless, it has also been found that some proteins of U1 snRNP complex may promote 3' end processing. SNRPA can enhance polyadenylation efficiency of SV40 late PAS by interacting with CPSF1 (Lutz et al., 1996), providing its enhancing ability of 3' end processing. APA at exon 4 of CT/CGRP gene was found to be activated by U1 snRNPs and ASF/SF2 (Lou et al., 1996). We also found that overexpression of SNRPA could facilitate switching to proximal PAS of STAT5B gene (Qiu et al., 2017). Homoplastically, the interaction between SNRNP70 and CFIm may mediate the crosstalk of U1 snRNP and CPAFs, which contributes to 3' end processing (Awasthi and Alwine, 2003).

As described above, U1 snRNP may have two sides of function in the regulation of polyadenylation, but the regulatory mechanism is largely unknown. Here, we found that both upregulation and downregulation of U1 snRNP proteins could promote the usage of proximal APA sites at the transcriptome level, but they showed totally different patterns of APA switching genes. We further investigated the molecular mechanisms and found that SNRNP70 and SNRPD2 can directly interact with 3' end processing factors. Furthermore, SNRNP70 can promote the proximal APA site through two ways, recruiting CPSF6 and suppressing PABPN1. All of these reveal that the U1 snRNP proteins in free state can promote the genes switching to proximal APA sites whereas U1 snRNP complex suppresses the proximal APA.

Results

Both overexpression and knockdown of U1 snRNP proteins can promote the usage of proximal APA sites

To investigate the effects of U1 snRNP proteins on APA regulation, we disturbed the expression of four U1 snRNP proteins (three specific proteins SNRPA, SNRPC, and SNRNP70 and one Sm protein SNRPD2) by overexpression and RNAi in HEK293T cell line (Figure 1A). Then, we performed 3' end sequencing with IVT-SAPAS (Fu et al., 2015) to measure APA site usage and 3' UTR length changes. On average, 45.1 million and 17.6 million raw

reads for each sample were acquired from the overexpressed and knockdown samples, respectively (Supplementary Table S1). Totally, an average of 5534 genes were found to have multiple PASs at last exon, referring as UTR-APA. The next analysis was based on these genes.

We firstly analyzed 3' UTR length changes gene-by-gene with the test of linear trend alternative to independence (Fu et al., 2011). Consistent with the previous finding of suppressive effects of U1 snRNP on proximal APA sites, we observed that more genes switched to shorter 3' UTRs with interfering the U1 snRNP genes except of SNRPC. Surprisingly, overexpression of these genes also resulted in the same trend (Figure 1B). To validate these results, we selected 20 genes from the intersect of genes regulated by the U1 snRNP proteins (Supplementary Figure S1A). We measured the expression levels of common and extended regions of these genes with real-time polymerase chain reaction (qRT-PCR) (Figure 1C). A higher common/extended (C/E) ratio means shorter 3' UTR. The results showed that 16 genes have the same trend with our sequencing data and 14 of them were statistically significant (Figure 1D and E).

Then, we investigated the difference of APA switching patterns between overexpression and knockdown of these U1 snRNP proteins. We took the Pearson's correlation r of each gene (calculated in the test of linear trend alternative to independence) as index of 3' UTR length changes. The index ranges from -1 to $+1$, and a negative value means shorter 3' UTR compared to the control sample, and vice versa. We calculated the correlations of these samples with the index. It was obvious that the samples could be clustered into two groups: overexpression and knockdown. There were higher correlations within groups but almost no correlations between the two groups (Figure 1F; Supplementary Figure S1B and C). This result uncovered totally different gene set with APA switching under overexpression and knockdown of U1 snRNP proteins. Intriguingly, we observed significantly higher correlation within the knockdown group than that within the overexpression group ($P = 0.031$ with Wilcoxon matched-pairs signed rank test) (Figure 1G). Additionally, the knockdown group also showed lower variance than the overexpression group, suggesting that knockdown of the four U1 snRNP proteins can affect similar gene sets, while overexpression of them may regulate different gene sets.

We also investigated the motifs enriched in the ± 100 bp sequence surrounding the proximal PASs of genes with 3' UTR shortening with the MEME software, by taking the proximal sites of genes without significant 3' UTR length change as the background. The motifs of G-rich, A-rich, C-rich, and CTKGG were significantly enriched near the proximal PASs of the target genes with overexpressed SNRPC, SNRNP70, and SNRPD2 (Supplementary Figure S1D), but only C-rich and CWG motifs were significantly enriched near the proximal PASs of the target gene with knockdown of SNRPA, SNRNP70, and SNRPD2 (Supplementary Figure S1E). APA switching genes with upregulated U1 snRNP proteins had more diverse motif enrichments than those in the knockdown group, suggesting

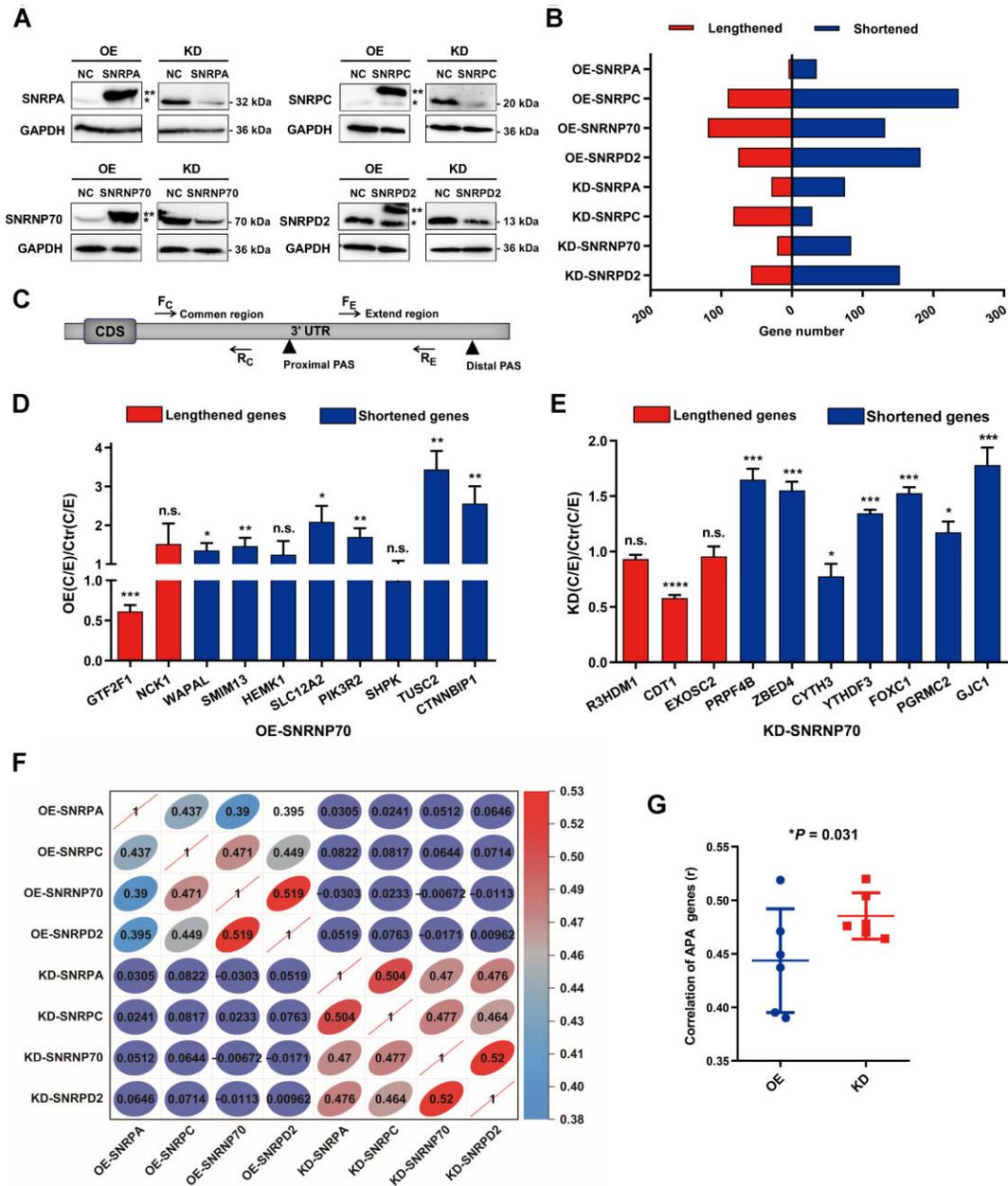


Figure 1 U1 snRNP proteins regulate APA in HEK293T cells. **(A)** Overexpression and knockdown efficiencies of U1 snRNP proteins revealed by western blotting. HEK293T cells were transfected with MYC-tagged plasmids and siRNAs to overexpress and knock down the four U1 snRNP proteins, respectively. * represents the band of endogenous proteins, and ** represents the band of MYC-tagged proteins. NC, negative control; OE, overexpression; KD, knockdown. GAPDH was used as an internal reference. **(B)** The number of APA genes significantly switched to lengthened (red) or shortened (blue) 3' UTRs in HEK293T cells with overexpression and knockdown of U1 snRNP proteins. **(C)** A diagram showing the design of qRT-PCR primers to validate APA switching genes. We measured the relative expression of common and extended regions of 3' UTR with qRT-PCR, and higher common/extended (C/E) ratio refers to shorter 3' UTR. **(D and E)** qRT-PCR validation of APA switching in HEK293T cells with overexpression **(D)** and knockdown **(E)** of SNRNP70. Red: lengthened genes; blue: shortened genes. Data were shown as mean \pm SD, $n = 3$. P-values were calculated with the student *t*-test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; n.s., no significance. **(F)** Pairwise Pearson's correlation of 3' UTR length changes in samples with overexpression and knockdown of U1 snRNP proteins. The correlation is color coded with a blue to red gradient: the blue corresponds to a weak correlation and the red to a strong positive correlation. **(G)** Correlations within the overexpression group and knockdown group. Data were shown as mean \pm SD. P-value was calculated with Wilcoxon matched-pairs signed rank test.

that knockdown of U1 snRNP proteins can regulate APA site switching of similar gene sets.

To further characterize the difference of gene sets under knockdown and overexpression of the four U1 snRNP proteins, respectively, we chose genes with $|r| > 0.1$ and got the intersection genes for knockdown and overexpression groups, respectively (Supplementary Figure S2A). We observed significantly higher ratios of overlapped genes in the knockdown group than in the overexpression group ($P = 0.031$ with Wilcoxon matched-pairs signed rank test) (Supplementary Figure S2B). For the knockdown group, the motifs of C-rich and T-rich were significantly enriched near the proximal PASs of the 94 overlapped genes with $r < -0.1$, while A-rich and G-rich were significantly enriched among the 91 genes with $r > 0.1$. However, no motifs were found with significant enrichment in the overlapped genes within the overexpression group (Supplementary Figure S2C). Consistently, the gene enrichment analysis result showed that the overlapped genes of the knockdown group but not the overexpression group could be enriched in some biological processes (Supplementary Figure S2D). These also prove that knockdown of U1 snRNP proteins regulates similar APA gene sets, while overexpression of them may regulate different gene sets.

U1 snRNP-free SNRPA has been observed in several cells (Boelens et al., 1993; Ma et al., 2006; Workman et al., 2014). Thus, the higher divergence of gene sets with APA switching caused by overexpression of U1 snRNP proteins indicates that the free proteins of U1 snRNP complex may affect APA directly.

SNRNP70 and SNRPD2 interact with 3' end processing factors

To understand the mechanism of APA regulation by U1 snRNP proteins, we searched the proteins interacting with four U1 snRNP proteins (SNRPA, SNRPC, SNRNP70, and SNRPD2) from the STRING database (Szklarczyk et al., 2019) and found intensive interactions of U1 snRNP proteins with the 3' end processing machinery factors (Supplementary Figure S3A). We overexpressed the proteins of SNRPA, SNRPC, SNRNP70, and SNRPD2 fused with a MYC-tag in HEK293T cells, respectively, and performed co-immunoprecipitation (co-IP) using MYC-tag antibody after RNase A treatment. The silver staining showed that SNRNP70 and SNRPD2 could pull down much more proteins than SNRPA and SNRPC (Supplementary Figure S3B). The mass spectrometry data reported previously (So et al., 2019) revealed a complex comprising U1 snRNP complex and 3' end processing factors by using crosslinking IP and mass spectrometry. To identify the direct interaction between U1 snRNP proteins and 3' end processing factors, we then performed co-IP experiments with 21 MYC-tagged 3' end processing core factors, including key components of CPSF (FIP1L1, CPSF1, CPSF2, CPSF3, CPSF4), CFIm (CPSF5, CPSF6, CPSF7), CFII (CLP1, PCF11), CSTF (CSTF1, CSTF2, CSTF2T, CSTF3), PAP (PAPOLA, PAPOLB, PAPOLG), PABP (PABPC1, PABPC4, PABPN1), and SYMPK, under the condition of RNase A treatment. Western blotting analysis showed that most of the 3' end processing core factors could efficiently pull down endogenous SNRNP70 and SNRPD2, but none of the factors could pull down endogenous SNRPA or

SNRPC (Supplementary Figure S4A), which was consistent with the silver staining result. Then, we applied more stringent co-IP conditions by increasing the washing times and the salt ion concentration of the wash buffer, and the result showed that CPSF6, CSTF1, PABPC1, and PABPN1 could still pull down endogenous SNRNP70 and SNRPD2 (Figure 2A). We also performed reverse co-IP to further validate these interactions. FLAG-SNRNP70 could pull down MYC-CPSF6 and MYC-PABPN1 (Figure 2B and C), but not MYC-CSTF1 or MYC-PABPC1 (Supplementary Figure S4B and C), and FLAG-SNRPD2 could only pull down MYC-PABPN1 (Figure 2D; Supplementary Figure S4D–F). These interactions were also validated for endogenous proteins with antibodies against SNRNP70 and SNRPD2 (Figure 2E and F). Consistently, we actually also found the higher correlation ($r = 0.519$) of APA site switching between SNRNP70 and SNRPD2 from APA sequencing data (Figure 1F; Supplementary Figure S1C), and both target gene sets of the two genes have similar motifs (Supplementary Figure S1D).

Given the interactions of SNRNP70 and SNRPD2 with the 3' end processing factors revealed by co-IP, we examined their localization in cultured HEK293T cells. Here, we also expressed mcherry-NLS, which contains a nuclear localization signal (NLS) as a negative control for nuclear protein. Immunofluorescence and intensity scan graph analysis showed the clear co-localization of SNRNP70 with CPSF6 and PABPN1 and of SNRPD2 with PABPN1 in the nucleus of HEK293T cells compared with the negative control, respectively (Figure 2G). We extracted the fluorescence intensity data of nucleus parts of 25 cells from different fields of view, and the Pearson's correlation coefficients showed higher correlations of SNRNP70 with CPSF6 and PABPN1 and of SNRPD2 with PABPN1, respectively, compared with the negative control (Figure 2H). The co-localization of these proteins further confirmed that SNRNP70 and SNRPD2 can interact with the 3' end processing factors.

Interaction of SNRNP70 with CPSF6 and PABPN1 is mediated through its LC1 domain

As a core component of U1 snRNP spliceosome complex, SNRNP70 contains four important functional domains: N-terminus, RNA recognition motif (RRM), and two low-complexity domains LC1 and LC2 (Bishof et al., 2018). To further determine which domain is involved in its interaction with CPSF6 or PABPN1, we truncated SNRNP70 into three fragments: N-terminus/RRM (N/R), LC1, and LC2 (Figure 3A). All of SNRNP70 fragments were fused with a MYC-tag at their N-terminus. We transfected MYC-tagged SNRNP70 fragments in HEK293T cells, by taking MYC-tagged empty vectors as the negative control, and then performed co-IP experiments with antibody against MYC-tag. The western blotting result showed that only LC1 domain of SNRNP70 could pull down endogenous CPSF6 and PABPN1 as SNRNP70-WT did (Figure 3B).

Next, we also analyzed the domains of CPSF6 and PABPN1. We truncated CPSF6 into three fragments, N/R, proline-rich (PRR) domain, and arginine-serine (RS) domain (Figure 3C), and fused them with a MYC-tag at their N-terminus. The co-IP result showed

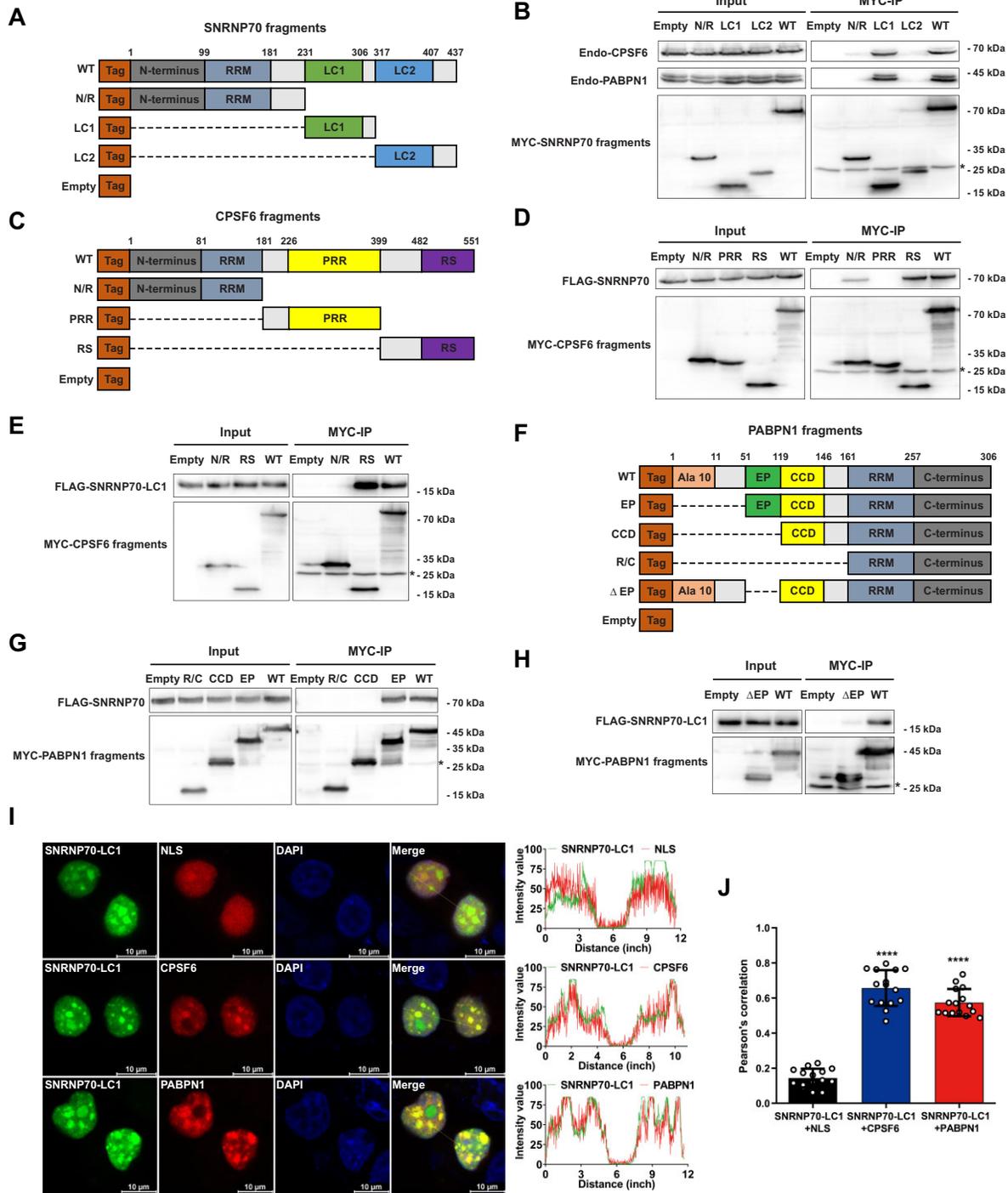


Figure 3 The LC1 domain of SNRNP70 interacts with CPSF6 and PABPN1. **(A)** Schematic representation of fragments of human SNRNP70 fused with MYC-tag: full-length (WT), N/R, LC1, and LC2. **(B)** HEK293T cells were transfected with MYC-tagged SNRNP70 fragments. Co-IP with RNase A treatment was carried out to pull down endogenous CPSF6 and PABPN1 by MYC-tag antibodies. * represents the light chain of antibody. MYC-tag (Empty) was used as a negative control. **(C)** Schematic representation of fragments of human CPSF6 fused with MYC-tag: WT, N/R, PRR, and RS. **(D)** HEK293T cells were co-transfected with FLAG-SNRNP70 and MYC-CPSF6 fragments. Co-IP analysis of FLAG-SNRNP70 with MYC-CPSF6 fragments using MYC-tag antibodies. * represents the light chain of antibody. MYC-tag (Empty) was used as a negative control. **(E)** Co-IP analysis of FLAG-SNRNP70-LC1 with MYC-CPSF6-N/R and MYC-CPSF6-RS using MYC-tag antibodies. * represents the light chain of antibody. MYC-tag (Empty) was used as a negative control. **(F)** Schematic representation of fragments of human PABPN1 fused with MYC-tag: WT (aa 1–306), EP (aa 51–306), CCD (aa 119–306), R/C (aa 161–306), and ΔEP (aa 1–50, 119–306). **(G)** Co-IP of FLAG-SNRNP70 with MYC-tagged fusion proteins of

that CPSF6-RS could pull down SNRNP70 as CPSF6-WT did, while CPSF6-N/R had a weak interaction with SNRNP70 (Figure 3D). To rule out indirect interaction, we further co-transfected SNRNP70-LC1 with CPSF6-N/R, CPSF6-RS, and CPSF6-WT, respectively, and found that only CPSF6-RS could pull down SNRNP70-LC1 as CPSF6-WT did (Figure 3E). These results suggest that the LC1 domain of SNRNP70 and RS domain of CPSF6 may mediate the interaction between them. Besides, PABPN1 contains five important functional domains: a stretch of 10 alanine residues at N-terminus, glutamic-proline (EP) region, coiled-coil domain (CCD), RRM, and C-terminus (Banerjee et al., 2013). We truncated PABPN1 into three fragments, RRM/C-terminus (R/C, aa 161–306), CCD (aa 119–306), and EP (aa 51–306), as shown in Figure 3F. MYC-tagged empty vector and PABPN1-WT were used as the negative control and positive control, respectively. The co-IP result showed that the EP (aa 51–306) of PABPN1, instead of CCD, RRM, or C-terminus, could pull down SNRNP70 as PABPN1-WT did (Figure 3G). To further examine the EP region for the interaction with SNRNP70, we constructed MYC-tagged PABPN1 with EP region deletion (PABPN1- Δ EP) (Figure 3F). Consistently, the PABPN1- Δ EP lost its ability to pull down SNRNP70-LC1 (Figure 3H), suggesting that the EP region of PABPN1 was responsible for the interactions with SNRNP70.

We analyzed the intracellular co-localization of SNRNP70-LC1 with CPSF6 and PABPN1 using laser scanning confocal microscopy. The results showed that SNRNP70-LC1 co-localized with CPSF6 and PABPN1 in the nucleus (Figure 3I). The fluorescence intensity correlation analysis showed higher correlations of SNRNP70-LC1 with CPSF6 and PABPN1 compared with the negative control (Figure 3J). All above results reveal that the LC1 domain of SNRNP70 contributes to the interaction of SNRNP70 with CPSF6 and PABPN1. Next, we would investigate the mechanism of APA regulation by SNRNP70 through interacting with PABPN1 and CPSF6.

SNRNP70 drives the phase transition of PABPN1 from droplet into aggregate

The LC1 domain of SNRNP70 has been characterized by liquid-liquid phase separation (LLPS) and shows the ability to self-assemble into droplets or aggregates due to the enrichment of basic (arginine) and acidic (aspartic/glutamic) amino acids (RD/RE motif) (Xue et al., 2019). Here, we indeed observed that SNRNP70 and SNRNP70-LC1 could form nuclear condensates in HEK293T nucleus, and fluorescence recovery after photobleaching (FRAP) results showed that these condensates

displayed dynamic droplet-like properties (Supplementary Figure S5A and B). Furthermore, we noticed that the EP domain of PABPN1, which interacts with SNRNP70, has an intrinsically disordered region. In order to explore whether the interaction between LC1 domain of SNRNP70 and PABPN1 is regulated by LLPS, we expressed and purified recombinant proteins EGFP-SNRNP70-LC1 and mCherry-PABPN1 *in vitro* (Supplementary Figure S5C). Both EGFP-SNRNP70-LC1 and mCherry-PABPN1 could undergo dynamic liquid phase at a certain concentration of protein and polyethylene glycol, and FRAP analyses showed that the bleached droplets recovered rapidly over the time course (Supplementary Figure S5D and E). Furthermore, EGFP-SNRNP70-LC1 could be fused with mCherry-PABPN1 into the incorporated droplets (Figure 4A), implying that the interaction of SNRNP70 with PABPN1 is characterized by LLPS. Actually, protein condensates could also be found from the co-localization of SNRNP70/PABPN1 and SNRNP70-LC1/PABPN1 in HEK293T nucleus (Figures 2G and 3I).

Since LLPS of SNRNP70 can produce protein assemblies of liquid-phase droplet or solid aggregate status in cells (Xue et al., 2019; Supplementary Figure S5A), we wondered whether SNRNP70 could affect the protein assembly of PABPN1. We transfected MYC-SNRNP70 or MYC-tagged empty vectors into HEK293T cells, and then detected the condensate status of MYC-SNRNP70 and endogenous PABPN1 by immunofluorescence assay with antibodies against MYC-tag and PABPN1. Consistently, overexpression of SNRNP70 resulted in coexistence of liquid-phase droplets and solid aggregates of SNRNP70 in cells and transformed endogenous PABPN1 from liquid-phase droplets into irregular condensates (Figure 4B). One obvious different feature between droplet and aggregate is the kinetics of their protein. In order to detect whether the irregular condensates of PABPN1 have undergone phase transition from droplets to aggregates, we further measured the mobility changes of EGFP-PABPN1 with FRAP assay while overexpressing MYC-SNRNP70 or MYC-tag. Obviously, EGFP-PABPN1 mainly existed in the form of liquid-phase droplets with co-transfection with MYC-tag, but almost all the droplets disappeared and assembled into irregular condensates when MYC-SNRNP70 was co-transfected (Figure 4C). The FRAP results showed that the aggregated PABPN1 recovered more slowly than the droplets (Figure 4C and D). It is known that PABPN1 can repress the proximal APA sites (de Klerk et al., 2012; Jenal et al., 2012). Thus, the transition of droplets into aggregates of PABPN1 induced by SNRNP70 may decrease its regulatory activity on the 3' end processing.

Figure 3 (continued) PABPN1 fragments using MYC-tag antibodies. * represents the light chain of antibody. MYC-tag (Empty) was used as a negative control. (H) Co-IP analysis of FLAG-SNRNP70-LC1 with MYC-PABPN1- Δ EP and MYC-PABPN1-WT using MYC-tag antibodies. * represents the light chain of antibody. MYC-tag was used as a negative control. (I) Co-localization of SNRNP70-LC1 with NLS, CPSF6, and PABPN1 in HEK293T cells. HEK293T cells were co-transfected with EGFP-SNRNP70-LC1 and mCherry-NLS/CPSF6/PABPN1 recombinant proteins. DAPI staining indicates the nucleus. mCherry-NLS was used as a negative control. Scan graph analysis shown in the right column represents the corresponding fluorescence intensity scanned by ImageJ software. (J) Pearson's correlation coefficients of two proteins in 15 cells from different fields of view were used to assess co-localization rate. Data were shown as mean \pm SD. *P*-values were calculated with the student *t*-test. *****P* < 0.0001.

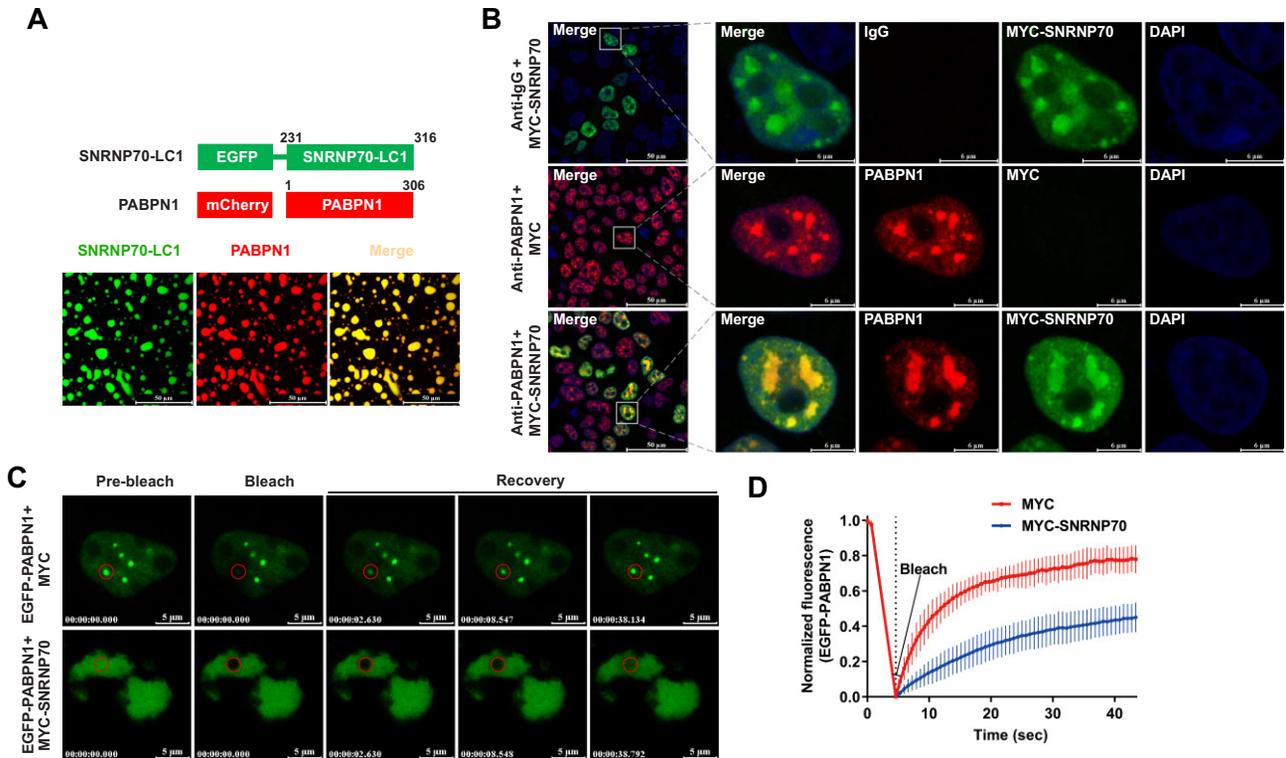


Figure 4 SNRNP70 promotes the phase transition of PABPN1. **(A)** Schematic picture of recombinant fluorescence-tagged proteins (EGFP-SNRNP70-LC1 and mCherry-PABPN1) and corresponding droplets. Droplet co-assembly of EGFP-SNRNP70-LC1 (30 μ M) with mCherry-PABPN1 (50 μ M) under 10% PEG 8000 *in vitro* was examined with fluorescence microscopy. **(B)** Confocal images of HEK293T cells transfected with MYC or MYC-SNRNP70. Cells were co-stained with MYC-tag antibody and endogenous PABPN1 antibody. Homologous IgG antibody was used as a control. DAPI staining indicates the nucleus. **(C)** HEK293T cells were co-transfected with MYC/MYC-SNRNP70 and EGFP-PABPN1. Fluorescence images show the FRAP recovery of the droplets or aggregates formed by EGFP-PABPN1. **(D)** Curves represent the recovery of normalized fluorescence intensity for EGFP-PABPN1. Colors represent different phase separation states: droplets (red) and gel-like aggregates (blue). The data were obtained from 12 cells.

SNRNP70 can promote the proximal APA sites by recruiting CPSF6

CPSF6 is an important component of the CFIm complex involved in mRNA 3' processing. It prefers to bind to UGUA motif upstream of distal APA sites and promotes the 3' end processing, resulting in the phenomenon of repression of proximal APA sites (Zhu et al., 2018). In order to further clarify the effects of SNRNP70 and CPSF6 interaction on APA regulation, we performed a tethering assay with the λ N-BoxB system. The tethering system consists of two parts: (i) a bicistronic dual luciferase (*Renilla* [Rluc] and *Firefly* [Fluc], Fluc is translated by the internal ribosomal entry site [IRES]) (Deng et al., 2018), with an insertion of 3 \times BoxB and \pm 50 bp sequence of candidate PASs between two luciferases; (ii) a λ N peptide-tagged SNRNP70-specifically-recognized BoxB motif (Figure 5A). Here, we used the proximal APA sites from the target genes of CTNNBIP1 and WAPAL, whose APA sites could be switched to proximal APA sites with upregulation of SNRNP70.

We firstly co-transfected the bicistronic reporter and SNRNP70 with or without a λ N-tag. Whereas both of SNRNP70 and λ N-SNRNP70 could significantly enhance the Rluc/Fluc ratios

compared to λ N control, λ N-SNRNP70 showed significantly greater effect than SNRNP70 ($P = 0.0026$ for CTNNBIP1 gene and $P = 0.0009$ for WAPAL gene) (Figure 5B). The Rluc/Fluc ratios at the mRNA level were consistent with those at the protein level (Supplementary Figure S6A). Next, we co-transfected CPSF6 with the bicistronic reporter and found that the Rluc/Fluc ratios were significantly reduced (Figure 5C), consistent with its effect on repressing the proximal APA sites (Zhu et al., 2018). Interestingly, when λ N-SNRNP70 was also transfected, the Rluc/Fluc ratios were fully rescued (Figure 5C). However, SNRNP70 without λ N-tag could only partially rescue the Rluc/Fluc ratios (Supplementary Figure S6B). These results suggest that the promotion of proximal APA sites by SNRNP70 needs its binding to the sites.

There are two ways to explain this rescuing effect of SNRNP70: (i) SNRNP70 and CPSF6 independently promote the proximal and distal APA sites, respectively (Figure 5D, top); (ii) SNRNP70 recruits CPSF6 and then promotes the proximal APA sites (Figure 5D, bottom). To investigate these two possibilities, we first transfected cells with CPSF6 siRNA and the scramble control, respectively, and then examined the effect of SNRNP70 with

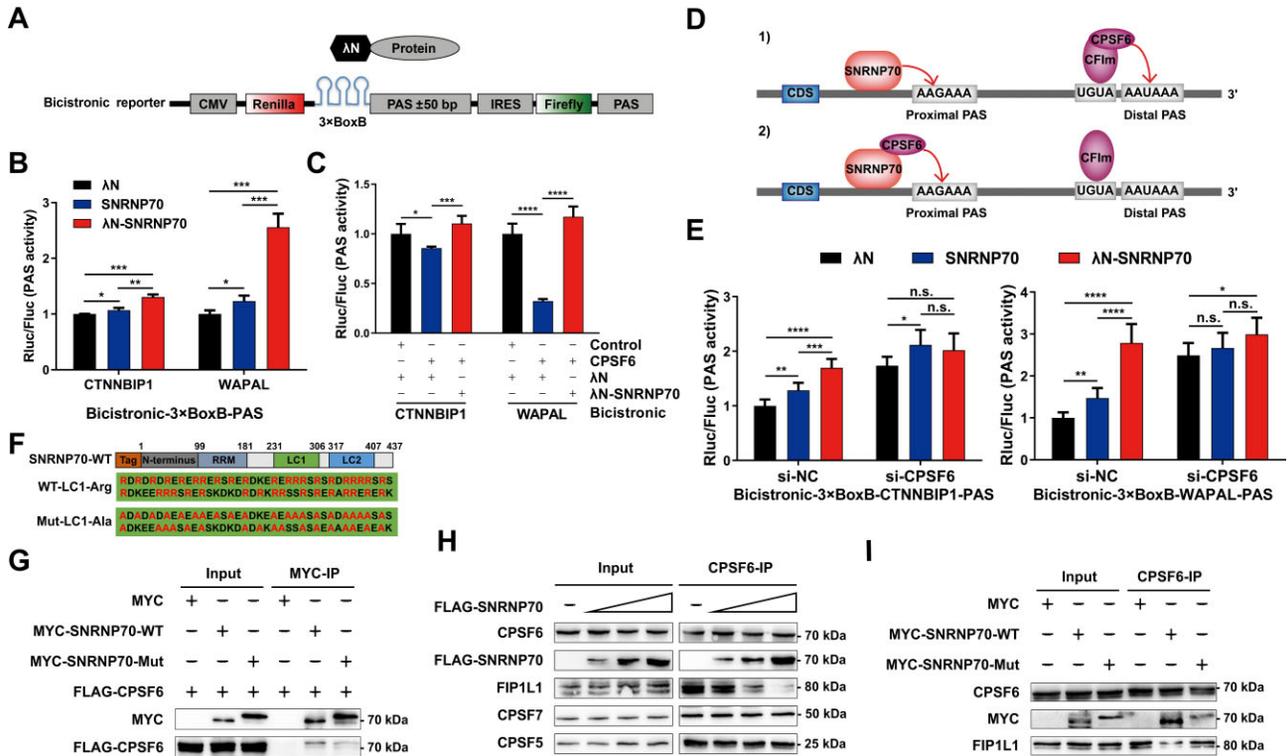


Figure 5 SNRNP70 promotes the proximal APA sites by binding with CPSF6. **(A)** Schematic diagram of the tethering assay. The 3×BoxB and ±50 bp sequence of PAS signal were inserted into a bicistronic reporting system containing two luciferases (*Renilla* and *Firefly*). λ-N-tag was fused to N-terminus of SNRNP70. **(B)** Tethering assay reveals that SNRNP70 can promote the proximal APA sites by binding to upstream of them. The sequences of the proximal APA sites of two genes (CTNNBIP1 and WAPAL) were cloned into the bicistronic reporter. The reporters were co-transfected into HEK293T cells with λN, SNRNP70, and λN-SNRNP70, respectively. Data were represented as mean ± SD, *n* = 3. *P*-values were calculated with the student *t*-test. **P* < 0.05; ***P* < 0.01; ****P* < 0.001. **(C)** SNRNP70 can rescue the repression of CPSF6 on the proximal APA sites. The bicistronic reporters were co-transfected with CPSF6, λN-SNRNP70, and empty vector control, respectively. Data were represented as mean ± SD, *n* = 4. *P*-values were calculated with the student *t*-test. **P* < 0.05; ****P* < 0.001; *****P* < 0.0001. **(D)** Proposed models for SNRNP70 to promote the proximal APA sites. Top: SNRNP70 and CPSF6 independently promote the proximal and distal APA sites, respectively; bottom: SNRNP70 recruits CPSF6 and then promotes the proximal APA sites. **(E)** Genetic interaction analysis shows that SNRNP70 promotes the proximal APA sites by recruiting CPSF6 to upstream of them. The bicistronic reporters and siRNA (si-NC/si-CPSF6) were co-transfected into HEK293T cells with λN, SNRNP70, and λN-SNRNP70, respectively. While tethering SNRNP70 to upstream of proximal poly(A) site could promote it, this ability was almost lost with knockdown of CPSF6. Data are mean ± SD, *n* = 6. *P*-values were calculated with the Student *t*-test. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001; n.s., no significance. Two-way ANOVA test also shows significant promotive effect by the interaction between SNRNP70 and CPSF6 (*P* = 0.014 for CTNNBIP1 gene and *P* = 0.0037 for WAPAL gene). **(F)** The arginines in LC1 domain (231–306) of SNRNP70 were mutated into alanines. **(G)** Co-IP results show that SNRNP70-Mut pulled down less CPSF6 than SNRNP70-WT did. **(H)** SNRNP70 can competitively bind to CPSF6 with FIP1L1 but not CPSF5 or CPSF7. Co-IP of FIP1L1, CPSF5, and CPSF7 with CPSF6 was performed using antibody against CPSF6 with gradient expression of FLAG-SNRNP70 in HEK293T cells. **(I)** HEK293T cells were transfected with MYC, MYC-SNRNP70-WT, and MYC-SNRNP70-Mut, respectively. Co-IP results show that the SNRNP70 mutant loses its ability to disturb the interaction of CPSF6 with FIP1L1.

tethering assay. Strikingly, the promotion of Rluc/Fluc ratios was not significant and the difference between λN-SNRNP70 and SNRNP70 disappeared when CPSF6 was knockdown. However, ANOVA test revealed that the effect of SNRNP70 and CPSF6 interaction on tethering was statistically significant (*P* = 0.014 for CTNNBIP1 and *P* = 0.0037 for WAPAL gene) (Figure 5E). These results imply that SNRNP70 can promote the proximal APA sites by binding to their vicinity region and recruiting CPSF6.

To further investigate the effect of SNRNP70 and CPSF6 interaction on APA, we tried to disrupt their interaction by mutations. The LC1 domain of SNRNP70, which mediates the interaction between SNRNP70 and CPSF6, is enriched with basic amino acid residues, especially arginines. We wondered whether the arginine-repeat region of SNRNP70 affects the interaction of SNRNP70 with CPSF6. Thus, we mutated all of the arginines in LC1 domain of SNRNP70 to alanines (Figure 5F) and found that less CPSF6 was pulled down by SNRNP70-Mut than

SNRNP70-WT in co-IP experiments (Figure 5G), suggesting that the mutant almost abrogated the interaction between SNRNP70 and CPSF6. We further examined the effect of mutated SNRNP70 on the usage of proximal PAS. The tethering assay results showed that SNRNP70-Mut and λ N-SNRNP70-Mut could not significantly raise Rluc/Fluc ratios as SNRNP70-WT and λ N-SNRNP70-WT did, respectively (Supplementary Figure S6C), indicating that the mutation lost the ability to promote the proximal APA sites. Consistently, we also found that SNRNP70-Mut was not able to rescue the CPSF6-repressed proximal APA sites (Supplementary Figure S6D). All above results reveal that the arginine-repeat region in LC1 domain of SNRNP70 does contribute to the regulation of SNRNP70 on APA, and SNRNP70 can promote the proximal APA sites by recruiting CPSF6.

SNRNP70 can compete with FIP1L1 to bind with CPSF6

The RS domains of CPSF6 can directly interact with the RE/D domain of FIP1L1, which is the subunit of the CPSF complex, and then promote distal PASs (Zhu et al., 2018). Our results also revealed that the RE/D-rich region in LC1 domain of SNRNP70 interacts with the RS domains of CPSF6. Thus, we wondered whether SNRNP70 affects the interaction between CPSF6 and other 3' end processing factors (CPSF5, CPSF7, and FIP1L1). We transfected FLAG-SNRNP70 with a gradient concentration, and then performed the co-IP experiment with antibody against CPSF6. We observed that less FIP1L1 was pulled down with the increased expression level of FLAG-SNRNP70 (Figure 5H). However, FLAG-SNRNP70 had no effect on the interaction of CPSF6 with CPSF5 or CPSF7. Moreover, SNRNP70-Mut could hardly inhibit the interaction between CPSF6 and FIP1L1 competitively (Figure 5I). This reveals that SNRNP70 can compete with FIP1L1 to bind with CPSF6, suggesting that the promotive effect of SNRNP70 on the proximal APA sites may be unrelated with FIP1L1.

Discussion

Alternative splicing and APA are indispensable steps in the regulation of mRNA transcription in eukaryotes, and they are closely related and mutually restricted (Elkon et al., 2013). It has been found that U1 snRNP complex involved in alternative splicing is closely associated with the regulation of APA. Here, we found that the protein components of U1 snRNP complex tend to promote the proximal APA sites whereas the whole complex suppresses them. We then further investigated the molecular mechanism of the protein components especially SNRNP70. We found that SNRNP70 could promote the proximal APA sites by interacting with the 3' end processing factors in two ways: (i) it drives the phase transition of PABPN1 from droplet to aggregate, which may interfere with the repressive effect of PABPN1 on the proximal APA sites; (ii) it binds to the proximal APA sites and recruits CPSF6, and then promotes the proximal sites. Consequently, the APA-regulating function of SNRNP70 in cells may depend on the expression levels of PABPN1 and CPSF6.

With knockdown of four genes encoding proteins of U1 snRNP complex, we found much more genes switched to the proximal

APA sites. This is consistent with the previous results from interfering with U1 snRNA. Perturbation of U1 snRNA with morpholino probes and overexpression revealed that U1 snRNP complex can inhibit cryptic APA sites in intron to protect the integrity of mRNA (Kaida et al., 2010; Berg et al., 2012; Oh et al., 2017) and suppress the proximal UTR-APA sites to regulate 3' UTR length (Oh et al., 2020) at the transcriptome-wide level. This suppressive effect typically requires the assembly of complete U1 snRNP complexes by U1 snRNA forming base-pairing with 5' ss (Shi et al., 2019; So et al., 2019; Venters et al., 2019). Moreover, we found a higher correlation of switching pattern among samples with knockdown of the four proteins. These reveal that knockdown of the genes may reduce U1 snRNP complex formation and weaken the inhibitory effect of U1 snRNP on the usage of PASs.

It has also been found that U1 snRNP-free SNRPA shows different regulatory effects on APA of several candidate genes in an U1-independent way. The free SNRPA could inhibit the 3' end processing of itself (Boelens et al., 1993) and SMN gene (Workman et al., 2014) and reduce their expression level. Elevated expression of free SNRPA was found in undifferentiated B cells but not differentiated B cells, which can repress the proximal APA sites of IgM and regulate the membrane and secretory isoform switch (Ma et al., 2006). However, it was also found that the free SNRPA could promote the 3' end processing *in vitro* (Liang and Lutz, 2006). Our group found the upregulation of SNRPA during the differentiation of T cells from naive to Th1 cell, which promoted the proximal APA sites of STAT5B gene (Qiu et al., 2017). Actually, the core proteins of U1 snRNP are expressed differently in different physiological states (Guio and O'Reilly, 2015). We analyzed the TCGA data on the GEPIA website and found that the expression levels of SNRPA, SNRPC, and SNRPD2 in almost all types of tumors were higher than those in normal tissues, while the expression levels of SNRNP70 were in the opposite direction (Supplementary Figure S7A). All of these suggest that the free proteins of U1 snRNP complex may play important roles in APA regulation when they are upregulated. Here, we found that overexpression of these proteins can also promote the proximal APA sites at the transcriptome level. Lower correlations of the APA switching pattern among samples with overexpression of these proteins indicate the different target genes of these proteins. In addition, we noted that the upregulation of any individual U1 snRNP protein did not affect the others (Supplementary Figure S7B and C), which indicates that their upregulation does not disrupt the assembly and function of the complex, thus protecting the pre-mRNA from abnormal premature termination. As a consequence, we can conclude here that U1 snRNP complex can suppress the proximal APA sites but the free proteins of U1 snRNP complex can directly promote them.

It has been found that both CPSF6 and PABPN1 can repress the proximal APA sites and promote the distal sites (Jenal et al., 2012; Zhu et al., 2018). We found the interaction of SNRNP70 with CPSF6 and PABPN1, mediated by the LC1 domain of SNRNP70, which is enriched with basic-acidic motifs. The LC1 domain of SNRNP70 modulates its phase separation and

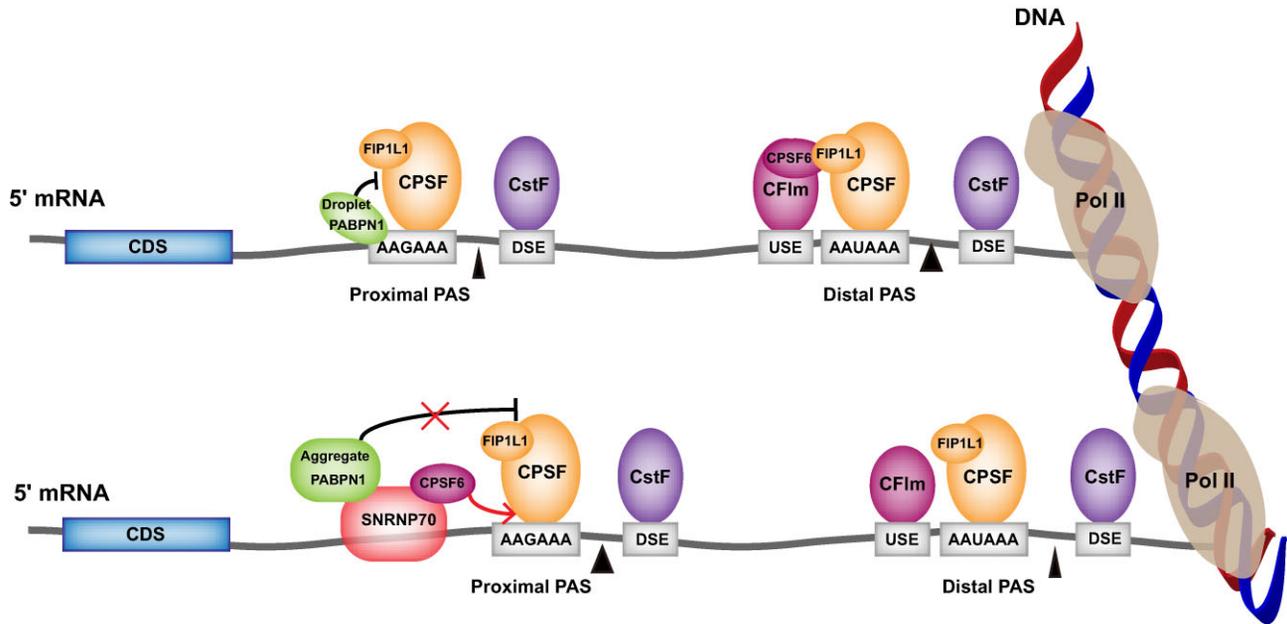


Figure 6 A proposed model of APA regulation by SNRNP70. PABPN1 binds to adjacent region of the proximal APA sites and represses them. CPSF6 promotes the distal APA sites by binding to UGUA motif and recruiting FIP1L1. The elevated SNRNP70 can decrease the repressive activity of PABPN1 on the proximal APA sites by phase transition and also recruit CPSF6 to promote the proximal APA sites.

aggregation (Xue et al., 2019; Greig et al., 2020), which was also confirmed here. We found that overexpression of SNRNP70 could promote the phase transition of PABPN1 from droplets into aggregates. It is increasingly clear that aggregates could be the next form of the droplets for some proteins, and this phase transition can affect the mobility and exchange efficiency of the proteins (Alberti et al., 2019; Xue et al., 2019). Aggregates of PABPN1 have less mobility than soluble droplets, which may decrease the repressive activity of PABPN1 on the proximal APA sites. Consistently, oculopharyngeal muscular dystrophy-associated PABPN1 mutant (trePABPN1) could induce protein aggregation (Raz et al., 2011) and lead to transcriptome-wide 3' UTR shortening (Jenal et al., 2012).

CPSF6 has been found to bind to UGUA motif upstream of poly(A) sites, recruit FIP1L1, and then activate the 3' end processing. Since UGUA motif is highly enriched in the region adjacent to the distal APA sites, CPSF6 seems to promote the distal APA sites and repress proximal APA sites, leading to longer 3' UTR (Zhu et al., 2018). With a λ -BoxB tethering assay, we validated that SNRNP70 can promote the usage of the proximal APA site by binding to upstream of it. When CPSF6 was knocked down, SNRNP70 lost its ability to promote the proximal sites though tethering on it. Furthermore, we found that the RD/RE motif in LC1 domain contributed to the interaction of SNRNP70 with CPSF6. Mutation of this motif cannot promote the proximal APA sites or rescue the repression of proximal APA site by CPSF6 as the wild-type. All of these illustrate that SNRNP70 can promote the proximal APA sites by recruiting CPSF6 (Figure 6). Our results suggest that the regulation of APA by CPSF6 may be to some extent cell context-dependent. In other word, there are

some other RNA-binding proteins with RD/RE motif, which may bind to the proximal APA sites, recruit CPSF6, and then activate these APA sites. Actually, CPSF6 has been found with elevated expression in liver cancer, and it may contribute to the shorter 3' UTR in cancer tissue than in normal tissue (Tan et al., 2021). In conclusion, we found that the free proteins of U1 snRNP complex can promote the proximal APA sites by directly interacting with the 3' end processing machinery factors through phase separation, and this regulatory effect is cell context-dependent, which is opposite to the effects of whole U1 snRNP complex.

Materials and methods

Genes cloning and plasmid construction

Total RNA extracted from HEK293T cells was used for cDNA synthesis by SuperScript® III Reverse Transcriptase (Invitrogen) with a random primer. The obtained cDNA was used as a template to amplify the U1 snRNP proteins and the 3' end processing factors with Phanta Max Super-Fidelity DNA Polymerase (Vazyme). Target fragments were cloned into pCMV-MYC, pCMV-FLAG, pmCherry-C1, and pEGFP-C1 expression vectors (Invitrogen) with ClonExpress® Ultra One Step Cloning Kit (Vazyme). The mutant vector MYC-SNRNP70-Mut was constructed by Tsingke. All constructed expression plasmids were confirmed by Sanger sequencing (Tsingke).

Cell culture and transfection

HEK293T cells were maintained in DMEM (Gibco) supplemented with 10% fetal bovine serum (ExCell Bio) at 37°C with 5% CO₂ incubation. HEK293T cells at ~40% confluence were transfected with siRNAs or plasmids using Lipofectamine RNAiMAX

Reagent (Invitrogen) and Transporter™ 5 Transfection Reagent (Polysciences) according to the manufacturer's instructions, respectively. The siRNA oligos of negative control (NC), U1 snRNP proteins, and CPSF6 were designed and synthesized (RiboBio). The siRNA sequences are listed in Supplementary Table S2.

IVT-SAPAS library and data analysis

HEK293T cells in 6-well plates were transfected with 50 μ M siRNAs of four U1 snRNP proteins and 2 μ g MYC-tagged plasmids for 48 h. The NC siRNA (RiboBio) and MYC-tagged empty vector were used as the control, respectively. Total RNA was extracted and APA sequencing was performed with IVT-SAPAS according to the previous report (Fu et al., 2015). Five overexpressed libraries and five knockdown libraries with different barcodes (Supplementary Table S3) were quantified and pooled together, respectively, and then sequenced with Hiseq 2500.

The sequencing reads were mapped to the human genome (hg19) after trimmed and filtered, and then the 3' UTR switching for each APA gene was analyzed as described previously (Fu et al., 2015). A test of linear trend alternative to independence was used to detect the genes with significant 3' UTR length changes between the control and overexpression/knockdown samples. We organized the read numbers of n APA sites for each gene in two samples into a $2 \times n$ table. Columns represent an APA site and are assigned with 3' UTR length, and rows represent the control and treatment samples and are assigned with 0 and 1, respectively. Then, Pearson's correlation r was calculated, and P -values were calculated with a statistic $M^2 = (n-1)r^2$, which is an approximated *chi*-square distribution with $df = 1$ for large samples. False discovery rate (FDR) was obtained with Benjamin–Hochberg method. A threshold of $FDR \leq 0.01$ & $|r| > 0.1$ was adopted to determine the significant genes with APA site switching.

Motif enrichment analysis

The 200-bp sequences of 3' UTR near the proximal poly(A) sites of the genes with significantly APA site switching were extracted by using the Bioconductor packages (BSgenome.Hsapiens.UCSC.hg19, PWMEnrich and seqLogo) of R software based on the recorded poly(A) site information. The genes with significantly changed 3' UTR were the target genes, and the other APA genes were the background genes. The target sequence and background sequence were respectively submitted to MEME software of The MEME Suite website (<http://meme-suite.org/>) for motif enrichment analysis.

qRT-PCR and APA sequencing validation

HEK293T cells in 12-well plates were transfected with 50 μ M SNRNP70 siRNA and 1 μ g MYC-tagged plasmid. Cells were harvested after 48 h of transfection and total RNA was extracted with TRIzol reagent (Invitrogen). cDNAs were synthesized by 500 ng total RNA with Evo M-MLV RT Kit (Accurate Biology), and qRT-PCR was run on a LightCycler480 apparatus (Roche) with SYBR Green pro Tag Mix (Accurate Biology). The relative mRNA expression of target genes was calculated by $2^{-\Delta\Delta CT}$ method. GAPDH was

used as internal control. The relative expression of common and extended regions of 3' UTR was measured by two pairs of primers (F_C/R_C and F_E/R_E) with qRT-PCR, and higher common/extended (C/E) ratio refers to shorter 3' UTR. Three biological repeats were performed and P -value was calculated with the student *t*-test. The qRT-PCR primers are listed in Supplementary Table S4.

Western blotting

HEK293T cells were harvested and lysed in RIPA buffer (Beyotime). After quantification with BCA reagent (Thermo), protein samples were separated by 6%–15% sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis and transferred to nitrocellulose filter membranes (Tanon). The nitrocellulose filter membrane was blocked in Tris-buffered saline with Tween 20 (TBST) buffer and 5% non-fat milk, incubated with diluted antibody overnight at 4°C. Then, the nitrocellulose filter membrane was washed with TBST buffer and incubated with correspond horseradish peroxidase (HRP)-labeled secondary antibodies for 1 h at room temperature. Protein signals were detected by Immobilon® Western Chemiluminescent HRP Substrate (Millipore).

Co-IP

HEK293T cells in 6-well plates were transfected with 2 μ g plasmids. Cells were washed with phosphate-buffered saline (PBS) and harvested after 48 h of transfection. Then, cells were lysed in co-IP lysis buffer (25 mM Tris-HCl, pH = 7.4, 150 mM NaCl, 1% NP-40, 1 mM EDTA, 1 mM PMSF (Sigma), 5% glycerol, 1 \times Protease inhibitor cocktail (Roche), and 1 mg/ml RNase A (Sigma)) at 4°C for 20 min, followed by centrifugation at 15000 *g* for 15 min at 4°C to remove cell debris. The supernatant was incubated with corresponding immunoprecipitation antibodies at room temperature for 2 h. The protein G magnetic beads (Invitrogen) were washed with washing buffer (50 mM Tris-HCl, pH = 7.4, and 300 mM NaCl) for pretreatment and incubated with supernatant samples for 1 h. Then, the beads were washed with washing buffer for three times, and proteins were eluted by boiling 15 min in 1 \times SDS loading buffer, followed by western blotting analysis.

Immunofluorescence

HEK293T cells were grown on a 15-mm glass-bottom cell culture dish (NEST). Cells were transfected with 1 μ g plasmid and rinsed briefly with ice-cold PBS after 24 h of transfection, followed by fixing cells with 4% paraformaldehyde in PBS for 20 min and washing three times with PBS. Then, the cells were permeabilized with 0.2% Triton X-100 in PBS for 15 min, blocked with 1% BSA in PBS for 30 min, incubated with corresponded antibody in PBST for 2 h, and washed three times with PBS, followed by incubating cells with the Alexa Fluor™ 488 or 568 secondary antibody in PBST for 1 h and washing three times with PBS. The last two steps are not required if the transfected protein carries a fluorescent label. Next, 1 μ g/ml DAPI (Sigma) was used to stain nuclei for 1 min followed by rinse with PBS. Then, the cells were imaged by Leica TCS SP8 STED 3X (Leica

Microsystems). The protein–protein co-localization coefficients were calculated by Leica microscope software. Cells from different fields of view were randomly extracted to assess co-localization coefficient.

Antibodies

The antibodies used in western blotting, co-IP, and immunofluorescence are as follows: MYC-tag (Sigma, M4439), FLAG-tag (Sigma, F1804), SNRPA (Abcepta, AW5557), SNRPC (Abcepta, AW5526), SNRNP70 (Santa Cruz Biotechnology, sc-390899), SNRPD2 (Abcam, ab198296), CPSF6 (Novus, NBP1-85676), PABPN1 (Abclonal, A1735), FIP1L1 (Novus, NBP1-85064), CPSF5 (Proteintech, 66335-1-Ig), CPSF7 (Proteintech, 55195-1-AP), GAPDH (Proteintech, 60004-1-Ig), IgG-rabbit (Cell Signaling Technology, 2729S), HRP-linked-mouse (Cell Signaling Technology, 7076S), HRP-linked-rabbit (Cell Signaling Technology, 7074S), Alexa Fluor™ 568 labeled-rabbit (Invitrogen, A-11011), Alexa Fluor™ 488 labeled-mouse (Invitrogen, A-11001).

Protein expression and purification

DNA fragments were cloned into the pET-32a-EGFP (*EcoRI* and *XhoI* digested) and pET-32a-pmCherry (*Sall* and *XhoI* digested) vectors. The proteins were expressed in BL21 (DE3) *Escherichia coli* cells (Tsingke) and induced by 0.6 mM IPTG. *E. coli* cells were cultured in standard LB medium at 37°C for 2–3 h until the OD600 reached 0.6–0.8. After that, the cells were cultured overnight at 20°C, 210 rpm for protein expression. Then, cells were harvested by centrifugation at 8000 rpm for 6 min. The 50 ml binding buffer (40 mM Na₂HPO₄, 120 mM NaH₂PO₄, 50 mM NaCl, and 20 mM imidazole, pH = 7.4) was used to re-suspend the bacteria. After breaking bacteria with high-pressure crusher, the fractured fluid was centrifuged at 10000 rpm for 20 min, filtered with 0.45 μm filter, and purified by Ni-NTA metal affinity column. Finally, the proteins were ultrafiltered into high-salt solution buffer (50 mM Tris–HCl, pH = 7.4, 300 mM KCl, 1 mM DDT, and 5% glycerol).

FRAP analyses

Confocal microscopy and FRAP analyses were conducted with the Leica TCS SP8 STED 3X (Leica microsystems) using a 100×, N.A. 1.4 oil-immersion apochromatic objective. Samples were applied to a 15-mm glass-bottom cell culture dish (NEST). A circular area of 1 μm radius or a speckle was bleached with the 65 mW argon laser and the PMT was used as detector. After completing the three steps of set up (setting parameters for pre- and post-bleach imaging), bleach (defining parameters for bleaching) and time course (defining numbers of pre-bleach, bleach, and post-bleach intervals), which depend on the sample and the purpose of the experiment, we obtained the displayed recovery showing all intensity values averaged over the regions of interest for all frames. The mean and standard deviation of fluorescence density were obtained from cells with different fields of view.

Tethering assay and dual-luciferase reporter assay

The bicistronic luciferase vectors were constructed as previously reported (Yao et al., 2012; Deng et al., 2018). A 3×BoxB sequence was synthesized (IGE Biotechnology) and two ±50 bp PAS sequences of target APA genes (CTNNBIP1 and WAPAL) were amplified from human cDNA. The 3×BoxB and PASs sequences were fused into the multiple cloning site of the bicistronic luciferase vectors. A λN-peptide was fused to the N-terminus of SNRNP70-WT and SNRNP70-Mut, and both were cloned into pCMV-MYC. HEK293T cells at 24-well plates were co-transfected with 300 ng of 3×BoxB-PAS-bicistronic luciferase vectors and 500 ng of λN-fused protein or control and collected and lysed after 48 h of transfection. The luciferase activity of Rluc and Fluc was measured with Dual-Luciferase® Reporter Assays reagent (Promega) by the GloMax Discover System (Promega), and the relative mRNA expression levels of Rluc and Fluc were examined by qRT-PCR. The qRT-PCR primers of tethering assay are listed in Supplementary Table S4.

Data availability

All IVT-SAPAS raw sequence data reported in this paper have been submitted to the NCBI Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>) under accession number SRR14842161.

Supplementary material

Supplementary material is available at *Journal of Molecular Cell Biology* online.

Funding

This work was supported by the National Natural Science Foundation of China (31971332 to Y.F., 91942301 and 81430099 to A.X., and 32000450 to L.C.).

Conflict of interest: none declared.

Author contributions: Z. Hu designed and performed the experiments, analyzed the data, and wrote the manuscript. M.L. performed the protein purification and FRAP experiments. Z. Huo performed IVT-SAPAS experiments. L.C. analyzed the APA sequencing data. S.L., K.D., and X.L. performed the co-IP and gene cloning experiments. S.C. analyzed the co-IP and phase separation data. Y.F. designed the study, analyzed the data, and wrote the manuscript. A.X. conceived and designed the study, edited the manuscript, and approved the final manuscript. All authors read and agreed on the final manuscript.

References

- Alberti, S., Gladfelder, A., and Mittag, T. (2019). Considerations and challenges in studying liquid–liquid phase separation and biomolecular condensates. *Cell* 176, 419–434.
- Awasthi, S., and Alwine, J.C. (2003). Association of polyadenylation cleavage factor I with U1 snRNP. *RNA* 9, 1400–1409.
- Bai, B. (2018). U1 snRNP alteration and neuronal cell cycle reentry in alzheimer disease. *Front. Aging Neurosci.* 10, 75.

- Banerjee, A., Apponi, L.H., Pavlath, G.K., et al. (2013). PABPN1: molecular function and muscle disease. *FEBS J.* 280, 4230–4250.
- Berg, M.G., Singh, L.N., Younis, I., et al. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53–64.
- Bishop, I., Dammer, E.B., Duong, D.M., et al. (2018). RNA-binding proteins with basic–acidic dipeptide (BAD) domains self-assemble and aggregate in Alzheimer’s disease. *J. Biol. Chem.* 293, 11047–11066.
- Boelens, W.C., Jansen, E.J., van Venrooij, W.J., et al. (1993). The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell* 72, 881–892.
- Carpenter, S., Ricci, E.P., Mercier, B.C., et al. (2014). Post-transcriptional regulation of gene expression in innate immunity. *Nat. Rev. Immunol.* 14, 361–376.
- Chi, B., O’Connell, J.D., Iocolano, A.D., et al. (2018). The neurodegenerative diseases ALS and SMA are linked at the molecular level via the ASC-1 complex. *Nucleic Acids Res.* 46, 11939–11951.
- de Klerk, E., Venema, A., Anvar, S.Y., et al. (2012). Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res.* 40, 9089–9101.
- Deng, Y., Shi, J., Ran, Y., et al. (2020). A potential mechanism underlying U1 snRNP inhibition of the cleavage step of mRNA 3’ processing. *Biochem. Biophys. Res. Commun.* 530, 196–202.
- Deng, Z., Zhang, S., Gu, S., et al. (2018). Useful bicistronic reporter system for studying poly(A) site-defining cis elements and regulation of alternative polyadenylation. *Int. J. Mol. Sci.* 19, 279.
- Elkon, R., Ugalde, A.P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation, and function. *Nat. Rev. Genet.* 14, 496–506.
- Fu, Y., Ge, Y., Sun, Y., et al. (2015). IVT-SAPAS: low-input and rapid method for sequencing alternative polyadenylation sites. *PLoS One* 10, e0145477.
- Fu, Y., Sun, Y., Li, Y., 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.
- Greig, J.A., Nguyen, T.A., Lee, M., et al. (2020). Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation. *Mol. Cell* 77, 1237–1250.
- Gruber, A.J., and Zavolan, M. (2019). Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* 20, 599–614.
- Guirou, J., and O’Reilly, D. (2015). Insights into the U1 small nuclear ribonucleoprotein complex superfamily. *Wiley Interdiscip. Rev. RNA* 6, 79–92.
- Gunderson, S.I., Beyer, K., Martin, G., et al. (1994). The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell* 76, 531–541.
- Gunderson, S.I., Polycarpou-Schwarz, M., and Mattaj, I.W. (1998). U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol. Cell* 1, 255–264.
- Jenal, M., Elkon, R., Loayza-Puch, F., et al. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* 149, 538–553.
- Jia, X., Yuan, S., Wang, Y., et al. (2017). The role of alternative polyadenylation in the antiviral innate immune response. *Nat. Commun.* 8, 14605.
- Kaida, D. (2016). The reciprocal regulation between splicing and 3’-end processing. *Wiley Interdiscip. Rev. RNA* 7, 499–511.
- Kaida, D., Berg, M.G., Younis, I., et al. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668.
- Kainov, Y.A., and Makeyev, E.V. (2020). A transcriptome-wide antitermination mechanism sustaining identity of embryonic stem cells. *Nat. Commun.* 11, 361.
- Li, Y., Sun, Y., Fu, Y., et al. (2012). Dynamic landscape of tandem 3’ UTRs during zebrafish development. *Genome Res.* 22, 1899–1906.
- Liang, S., and Lutz, C.S. (2006). p54nrb is a component of the snRNP-free U1A (SF-A) complex that promotes pre-mRNA cleavage during polyadenylation. *RNA* 12, 111–121.
- Licatalosi, D.D., and Darnell, R.B. (2010). RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.* 11, 75–87.
- Lou, H., Gagel, R.F., and Berget, S.M. (1996). An intron enhancer recognized by splicing factors activates polyadenylation. *Genes Dev.* 10, 208–219.
- Lutz, C.S., Murthy, K.G., Schek, N., 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.
- Ma, J., Gunderson, S.I., and Phillips, C. (2006). Non-snRNP U1A levels decrease during mammalian B-cell differentiation and release the IgM secretory poly(A) site from repression. *RNA* 12, 122–132.
- Mayr, C. (2016). Evolution and biological roles of alternative 3’UTRs. *Trends Cell Biol.* 26, 227–237.
- Oh, J.M., Di, C., Venters, C.C., et al. (2017). U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat. Struct. Mol. Biol.* 24, 993–999.
- Oh, J.M., Venters, C.C., Di, C., et al. (2020). U1 snRNP regulates cancer cell migration and invasion in vitro. *Nat. Commun.* 11, 1.
- Qiu, F., Fu, Y., Lu, C., et al. (2017). Small nuclear ribonucleoprotein polypeptide A-mediated alternative polyadenylation of STAT5B during Th1 cell differentiation. *J. Immunol.* 199, 3106–3115.
- Raz, V., Abraham, T., van Zwet, E.W., et al. (2011). Reversible aggregation of PABPN1 pre-inclusion structures. *Nucleus* 2, 208–218.
- Shi, J., Deng, Y., Huang, S., et al. (2019). Suboptimal RNA–RNA interaction limits U1 snRNP inhibition of canonical mRNA 3’ processing. *RNA Biol.* 16, 1448–1460.
- Shi, Y.S. (2012). Alternative polyadenylation: new insights from global analyses. *RNA* 18, 2105–2117.
- So, B.R., Di, C., Cai, Z., et al. (2019). A complex of U1 snRNP with cleavage and polyadenylation factors controls telescripting, regulating mRNA transcription in human cells. *Mol. Cell* 76, 590–599.e4.
- Szklarczyk, D., Gable, A.L., Lyon, D., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613.
- Tan, S., Zhang, M., Shi, X., et al. (2021). CPSF6 links alternative polyadenylation to metabolism adaption in hepatocellular carcinoma progression. *J. Exp. Clin. Cancer Res.* 40, 85.
- Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* 18, 18–30.
- Venters, C.C., Oh, J.M., Di, C., et al. (2019). U1 snRNP telescripting: suppression of premature transcription termination in introns as a new layer of gene regulation. *Cold Spring Harb. Perspect. Biol.* 11, a032235.
- Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718.
- Wan, R., Bai, R., and Shi, Y. (2019). Molecular choreography of pre-mRNA splicing by the spliceosome. *Curr. Opin. Struct. Biol.* 59, 124–133.
- Workman, E., Veith, A., and Battle, D.J. (2014). U1A regulates 3’ processing of the survival motor neuron mRNA. *J. Biol. Chem.* 289, 3703–3712.
- Xue, S., Gong, R., He, F.Q., et al. (2019). Low-complexity domain of U1-70K modulates phase separation and aggregation through distinctive basic–acidic motifs. *Sci. Adv.* 5, eaax5349.
- Yao, C., Biesinger, J., Wan, J., et al. (2012). Transcriptome-wide analyses of CstF64–RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc. Natl Acad. Sci. USA* 109, 18773–18778.
- Zhang, S., Aibara, S., Vos, S.M., et al. (2021). Structure of a transcribing RNA polymerase II–U1 snRNP complex. *Science* 371, 305–309.
- Zhu, Y., Wang, X.Y., Forouzmand, E., et al. (2018). Molecular mechanisms for CFIm-mediated regulation of mRNA alternative polyadenylation. *Mol. Cell* 69, 62–74.