### Title
CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing




### Journal
Genes and Development, 28(21)

### ISSN
0890-9369



### Authors
Chan, SL
Huppertz, I
Yao, C
et al.



### Publication Date
2014

### DOI
10.1101/gad.250993.114.

# CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing

**Serena L. Chan,[1] Ina Huppertz,[2,3] Chengguo Yao,[1] Lingjie Weng,[1,4,5] James J. Moresco,[6] John R. Yates III,[6] Jernej Ule,[2,3] James L. Manley,[7] and Yongsheng Shi[1]**

[1]Department of Microbiology and Molecular Genetics, School of Medicine, University of California at Irvine, Irvine, California 92697, USA; [2]Department of Molecular Neuroscience, University College London Institute of Neurology, London WC1N 3BG, United Kingdom; [3]Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom; [4]Institute for Genomics and Bioinformatics, [5]Department of Computer Science, University of California at Irvine Irvine, California 92697, USA; [6]Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA; [7]Department of Biological Sciences, Columbia University, New York, New York 10027, USA

**AAUAAA is the most highly conserved motif in eukaryotic mRNA polyadenylation sites and, in mammals, is specifically recognized by the multisubunit CPSF (cleavage and polyadenylation specificity factor) complex. Despite its critical functions in mRNA 3′ end formation, the molecular basis for CPSF–AAUAAA interaction remains poorly defined. The CPSF subunit CPSF160 has been implicated in AAUAAA recognition, but direct evidence has been lacking. Using in vitro and in vivo assays, we unexpectedly found that CPSF subunits CPSF30 and Wdr33 directly contact AAUAAA. Importantly, the CPSF30–RNA interaction is essential for mRNA 3′ processing and is primarily mediated by its zinc fingers 2 and 3, which are specifically targeted by the influenza protein NS1A to suppress host mRNA 3′ processing. Our data suggest that AAUAAA recognition in mammalian mRNA 3′ processing is more complex than previously thought and involves multiple protein–RNA interactions.**

[*Keywords*: mRNA 3′ processing; cleavage and polyadenylation; polyadenylation signal]

Supplemental material is available for this article.

Received August 13, 2014; revised version accepted September 23, 2014.

mRNA 3′ processing, typically involving an endonucleolytic cleavage and subsequent polyadenylation, is not only an essential step in eukaryotic gene expression but also a critical mechanism for gene regulation (Colgan and Manley 1997; Zhao et al. 1999; Millevoi and Vagner 2009; Chan et al. 2011). The majority of eukaryotic genes produce multiple mRNA isoforms by using alternative polyadenylation sites (PASs), and PAS selection is dynamically regulated during development (Shi 2012; Tian and Manley 2013). Aberrant mRNA alternative polyadenylation has been associated with a number of diseases, including cancer (Di Giammartino et al. 2011). As such, a central question in the mRNA 3′ processing field has been to understand how PASs are recognized and how PAS selection can be regulated. In metazoans, most PASs are defined by several key *cis*-elements, including the AAUAAA hexamer located 10–30 nucleotides (nt) upstream of the cleavage site, a U/GU-rich downstream element located ~30 nt downstream from the cleavage site, and other auxiliary sequences (Colgan and Manley 1997; Zhao et al. 1999; Chan et al. 2011). Among these sequences, the AAUAAA hexamer (or its close variant, AUUAAA) is the most highly conserved, and almost all single-nucleotide substitutions in this hexamer reduce or abolish mRNA 3′ processing in vitro (Sheets et al. 1990; Proudfoot 2011). Genetic mutations in this hexamer can significantly impact the efficiency of mRNA 3′ processing in vivo and cause human diseases, including thalassemia and thrombophilia (Danckwardt et al. 2008).

The AAUAAA hexamer is specifically recognized by CPSF (cleavage and polyadenylation specificity factor) (Takagaki et al. 1988; Keller et al. 1991), a protein complex that contains six subunits: CPSF160, Wdr33, CPSF100, CPSF73, Fip1, and CPSF30 (Mandel et al. 2008; Shi et al. 2009). It has been suggested that CPSF160 is responsible for recognizing AAUAAA based on several lines of evidence. First, two proteins of ~160 kDa and ~30 kDa in the purified CPSF complex that were thought to correspond to CPSF160 and

**Corresponding author: yongshes@uci.edu**
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.250993.114.

CPSF30, respectively, can be specifically cross-linked to AAUAAA-containing RNA substrates by UV irradiation (Keller et al. 1991). However, the identities of these proteins were not experimentally determined, and, more importantly, it was unclear whether either of these two proteins directly binds to AAUAAA. Second, the yeast homolog of CPSF160, Yhh1/Cft1p, has been shown to bind RNA (Dichtl et al. 2002). Third, recombinant human CPSF160 binds to RNAs in a pull-down assay and shows a preference for AAUAAA-containing RNAs (Murthy and Manley 1995). However, CPSF160 has been shown to bind to an enhancer element located 76 nt upstream of the AAUAAA in HIV mRNA PASs (Gilmartin et al. 1995). Additionally, a recent study that aimed to map protein–RNA interactions in vivo for mRNA 3′ processing factors failed to detect specific CPSF160 binding at the AAUAAA hexamer (Martin et al. 2012). Therefore, it remains unclear how CPSF recognizes AAUAAA in mammalian mRNA 3′ processing.

The smallest subunit of the CPSF complex, CPSF30, contains five zinc finger (ZF) domains and, in metazoans, a zinc knuckle (ZK) domain (Barabino et al. 1997). Recombinant CPSF30 has RNA-binding activity, and its ZK domain enhances this activity (Barabino et al. 1997). Interestingly, the influenza virus protein NS1A specifically interacts with CPSF30 via its ZF2 and ZF3 domains, and this interaction is responsible for suppressing host mRNA 3′ processing (Nemeroff et al. 1998; Twu et al. 2006). However, the role of CPSF30–RNA interaction in mRNA 3′ processing remains poorly characterized, and it is unclear why CPSF30 is specifically targeted by the viral protein.

In an effort to define the molecular basis for CPSF–AAUAAA interaction, we unexpectedly found that CPSF30 and Wdr33 directly bind to the AAUAAA hexamer. We demonstrated that CPSF30–RNA interaction is essential for mRNA 3′ processing and is primarily mediated by its ZF2 and ZF3 domains. By combining these new results with previous findings, we propose a new model for AAUAAA recognition by CPSF that requires multiple protein–RNA interactions.

## Results

### *Immunopurified CPSF specifically recognizes AAUAAA*

Several groups have previously purified CPSF from cultured mammalian cells or animal tissues using chromatography methods (Bienroth et al. 1991; Murthy and Manley 1992; Gilmartin et al. 1995). However, the yield of CPSF was generally low, and the purity and activities of the isolated complexes tended to vary, making it difficult to carry out detailed functional studies. To overcome these technical limitations, we established HEK293 cell lines that stably express Flag-tagged CPSF73 near the endogenous levels (Shi et al. 2009). To purify the CPSF complex, we prepared nuclear extract from these cells, carried out immunoprecipitation under stringent conditions using anti-Flag antibodies, and eluted immunoprecipitated proteins under native conditions using Flag peptides. Analyses of the eluted proteins by mass spectrometry and Western blotting suggested that the purified CPSF complex was intact, containing all known CPSF subunits plus an additional associated protein, symplekin (Table 1; Shi et al. 2009). It was noted that the purified CPSF complex contained two CPSF30 isoforms that are encoded by alternatively spliced mRNAs (Fig. 1A). The gel staining profile (Fig. 1A) and mass spectrometry analyses results (Table 1) suggest that all subunits are present in the purified CPSF complex at comparable levels. Furthermore, mass spectrometry analyses of biological replicates revealed that the molecular composition and apparent stoichiometry of the purified CPSF complex were highly consistent (Table 1). Therefore, these cell lines provide a reliable and efficient system for purifying CPSF.

Next we tested whether the immunopurified CPSF specifically recognizes the AAUAAA hexamer and functions in mRNA 3′ processing using multiple in vitro assays. As shown in Figure 1B, gel mobility shift assays demonstrated that a specific RNP complex was formed when CPSF was incubated with an AAUAAA-containing RNA substrate (WT: wild type) derived from the commonly used SV40 late (SVL) transcript PAS. A single-nucleotide substitution in the hexamer (AAGAAA) completely abolished this interaction (Fig. 1B, MT: mutant). In addition, we also carried out in vitro polyadenylation assays with purified CPSF and recombinant poly(A) polymerase (PAP), which have been shown to be sufficient for reconstituting AAUAAA-dependent polyadenylation (Christofori and Keller 1988; Takagaki et al. 1988). Consistent with the gel shift results (Fig. 1B), our in vitro polyadenylation assays showed that the AAUAAA-containing, but not the AAGAAA-containing, RNA substrate was specifically polyadenylated when incubated with CPSF and PAP (Fig. 1C). Similar results were obtained with the adenovirus L3 PAS RNA, another commonly used RNA substrate for mRNA 3′ processing assays (Supplemental Fig. S1). These results demonstrated that the immunopurified CPSF specifically recognizes the AAUAAA hexamer and is functional in mRNA polyadenylation and thus suitable for detailed characterizations of the CPSF–AAUAAA interaction.

**Table 1.** *Mass spectrometry analyses of the immunopurified CPSF complex*

| Protein name | Accession number | Molecular weight | Coverage (replicate 1) | Coverage (replicate 2) |
|---|---|---|---|---|
| CPSF160 | NP_037423.2 | 160.9 kDa | 56.2% | 61.3% |
| Wdr33 | NP_060853.3 | 145.9 kDa | 38.4% | 38.5% |
| CPSF100 | NP_059133.1 | 88.5 kDa | 68.8% | 67.9% |
| CPSF73 | NP_057291.1 | 77.5 kDa | 73.5% | 74.7% |
| Fip1 | NP_00128410.1 | 58.4 kDa | 52.5% | 49.0% |
| CPSF30 | NP_006684.1 | 30.3 kDa | 53.9% | 53.9% |
| Symplekin | NP_004810.2 | 141.1 kDa | 36.8% | 44.3% |

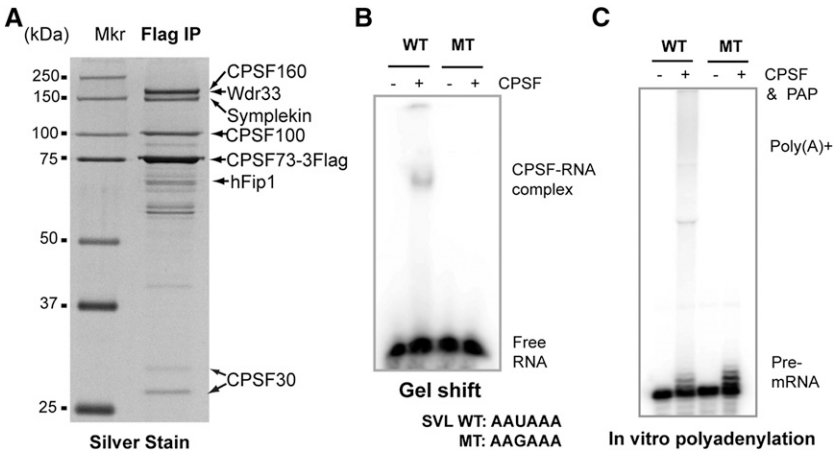


**Figure 1.** Immunopurified CPSF specifically recognizes AAUAAA. (*A*) Immunopurified CPSF complex was resolved by SDS-PAGE and visualized by silver staining. Key components are marked. (Mkr) Protein marker (the size of each band is labeled at the *left*). (*B*) Immunopurified CPSF was incubated with 5′ radiolabeled SVL RNA substrate and then resolved on a native gel and visualized by phosphorimaging. The free RNA and CPSF–RNA complex are marked. (WT) RNA substrate containing the wild-type hexamer AAUAAA; (MT) RNA substrate containing a mutant hexamer, AAGAAA. (*C*) The same RNA substrate used in *B* was incubated with CPSF and recombinant PAP in the presence of ATP. The RNAs were resolved on an 8% urea gel and visualized by phosphorimaging. The pre-mRNA substrate and the polyadenylated [poly(A)$^+$] RNA are marked.

*CPSF30 and Wdr33 directly bind to AAUAAA*

To define the molecular basis of CPSF–AAUAAA interaction, a key first step was to determine which CPSF subunits directly bind to the AAUAAA hexamer. To this end, we synthesized an L3 RNA substrate containing the wild-type AAUAAA hexamer by splinted ligation. This RNA was radiolabeled at a single position between the third and fourth nucleotide within the AAUAAA hexamer (Fig. 2A, top panel). Another RNA (mutant) was synthesized that contained a mutant hexamer (AAUAAC) but was otherwise identical in sequence and the position of the radiolabel (Fig. 2A, top panel). AAUAAC was chosen such that the nucleotides immediately adjacent to the radiolabel phosphate were identical between the wild-type and mutant RNAs, thereby minimizing the possible bias in UV cross-linking efficiency caused by nucleotide context.

To identify AAUAAA-interacting proteins, purified CPSF was incubated with either the wild-type or mutant RNA substrates, UV-irradiated, and resolved by SDS-PAGE. Two major radiolabeled bands of ~200 kDa and ~40 kDa and a minor band of ~80 kDa, were specifically detected in AAUAAA-containing samples (Fig. 2A, bottom panel, lane 1), but no radiolabeled bands were observed when the mutant RNA was used (Fig. 2A, bottom panel, lane 4). Consistent with our gel mobility shift and in vitro polyadenylation assay results (Fig. 1B,C), the UV cross-linking pattern again demonstrated that the immunopurified CPSF can specifically bind to AAUAAA-containing RNAs and that a single-nucleotide substitution in the hexamer abolished this interaction (Fig. 2A). Additionally, the UV cross-linking results suggested that CPSF–RNA interactions are mediated primarily by two proteins. To determine which, if any, of these proteins directly binds to AAUAAA, we treated the UV cross-linked CPSF–RNA complexes with RNase T1 or RNase I. RNase T1 cleaves 3′ of unpaired G residues, while RNase I cleaves at all nucleotides in ssRNAs. RNA cleavage by these RNases was experimentally confirmed, and the presence of CPSF did not interfere with RNA digestion (Supplemental Fig. S2). After RNase treatment, only the proteins that are directly cross-linked to the hexamer region should remain radiolabeled. Under these conditions, the two major bands remained labeled, but their apparent molecular weights decreased to ~160 kDa and ~30 kDa, respectively (Fig. 2A, bottom panel, cf. lanes 1 and 2,3). Such shifts were due to the trimming of the CPSF-associated RNA by the RNases (Supplemental Fig. S2). These data demonstrated that two CPSF subunits of ~160 kDa and ~30 kDa specifically and directly bind to the AAUAAA hexamer. It should be pointed out that the decreases in signal intensities of the two major bands following RNase treatments (Fig. 2A, cf. lanes 1 and 2,3) indicate that these two proteins may also cross-link to sequences outside AAUAAA.

Based on the prevalent model for mammalian mRNA 3′ processing, we suspected that the 160-kDa species corresponds to CPSF160, which has been implicated in AAUAAA recognition (Murthy and Manley 1995; Dichtl et al. 2002). Wdr33, a recently identified CPSF subunit with a molecular weight similar to that of CPSF160 (Shi et al. 2009), represents another candidate protein for the ~160-kDa species. The ~30-kDa species most likely corresponds to CPSF30 for several reasons. First, CPSF30 is the only CPSF subunit that has a molecular weight in the ~30-kDa range (Fig. 1A). Second, the ~30-kDa species appeared to contain a doublet (Fig. 2A, lanes 2,3) that is reminiscent of the two CPSF30 isoforms present in the purified CPSF complex (Fig. 1A). Finally, CPSF30 is a known RNA-binding protein (Barabino et al. 1997). To determine the identities of the two proteins that directly bind to AAUAAA, we first treated the UV cross-linked CPSF–wild-type RNA complex with RNase I and then denatured the CPSF–RNA complex by boiling it for 3 min in the presence of 0.5% SDS. We then carried out immunoprecipitation using antibodies against candidate proteins with the denatured CPSF–RNA complexes (see the Materials and Methods for details). As shown in Figure 2B (top panel), CPSF30, CPSF160, and Wdr33 were all specifically immunoprecipitated by their respective antibodies, while other CPSF subunits were not coprecipitated, confirming the specificity of this procedure.

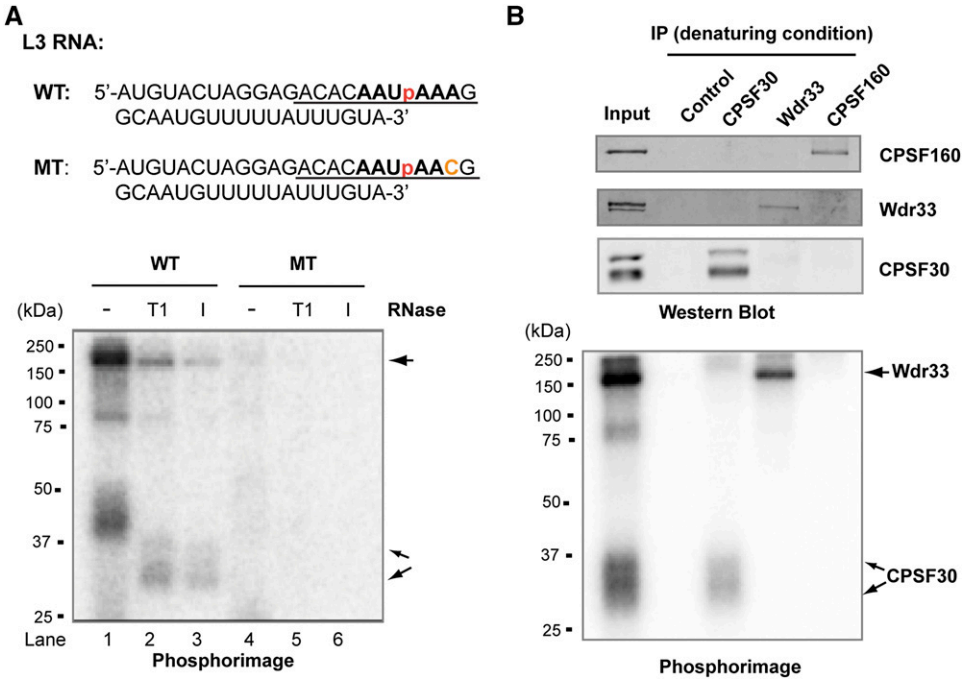


**Figure 2.** CPSF30 and Wdr33 directly bind to AAUAAA in vitro. (*A*, *top* panel) The L3 RNA substrate containing AAUAAA (wild type [WT]) or AAUAAC (mutant [MT]) hexamer. The red "p" denotes the position of the single $^{32}$P label. The RNase T1-resistant fragment is underlined. (*Bottom* panel) Immunopurified CPSF and site-specifically labeled RNA substrates were incubated and irradiated with UV (wavelength 254 nm). The cross-linked samples were left untreated (−) or incubated with RNase T1 or RNase I and then resolved on SDS-PAGE and visualized by phosphorimaging. The major radiolabeled bands are marked with arrows. (*B*) CPSF was incubated with RNA substrates, irradiated by UV, and digested with RNase I as described in *A*. The sample was then denatured before being subjected to immunoprecipitation (IP) using the specified antibodies. The immunoprecipitation samples were analyzed by Western blotting (*top* panel) or phosphorimaging (*bottom* panel).

Using this procedure, we found that the ~30-kDa radiolabeled species in the CPSF–RNA cross-linking product was specifically immunoprecipitated by anti-CPSF30 antibodies (Fig. 2B, bottom panel). Surprisingly, however, the ~160-kDa radiolabeled band was specifically and efficiently immunoprecipitated by anti-Wdr33 antibodies but not by anti-CPSF160 antibodies (Fig. 2B, bottom panel). These results provided direct evidence that CPSF30 and Wdr33 directly bind to the AAUAAA hexamer. Under these conditions, we did not detect cross-linking between CPSF160 and the AAUAAA. The implications of these findings are discussed below.

### *Mapping the CPSF30 and Wdr33 footprints in the AAUAAA region*

As both CPSF30 and Wdr33 bind directly to the AAUAAA hexamer, we next mapped their footprints in the AAUAAA hexamer region at a high resolution. To this end, we prepared a series of AAUAAA-containing RNAs that have sequences identical to those of the L3 wild-type RNA used in Figure 2A but were site-specifically radiolabeled at different sites at or near the AAUAAA hexamer, including 0, −13, −7, and +9 nt relative to the U in the AAUAAA hexamer (Fig. 3A, top panel). We then incubated these RNAs with CPSF and subjected them to UV cross-linking followed by RNase I digestion. Our results showed that both CPSF30 and Wdr33 cross-linked to 0 nt and −7 nt but weakly to −13 nt (Fig. 3A, bottom panel, lanes 6,4,2). In contrast, a prominent protein of ~70 kDa specifically cross-linked to −13 nt (Fig. 3A, lane 2). Using the same immunoprecipitation procedure as described above (Fig. 2B), we provided evidence that this ~70-kDa species most likely corresponds to Fip1 (Supplemental Fig. S3), a known CPSF subunit that has RNA-binding activity (Kaufmann et al. 2004; Lackford et al. 2014). At +9 nt, however, Wdr33 was the only major protein that cross-linked to RNA, and its cross-linking signal at this site was weaker compared with those at 0 and −7 nt (Fig. 3A, bottom panel, cf. lanes 8 and 4,6), indicating weaker Wdr33–RNA interaction at +9 nt.

We also mapped the footprints of CPSF30 and Wdr33 on SVL RNA (Fig. 3B). To this end, four different SVL-derived RNAs were synthesized and 3′ radiolabeled. These RNAs contain a unique G residue that is positioned at different positions relative to the AAUAAA hexamer such that RNase T1 digestion of these RNAs will generate a series of different radiolabeled RNA fragments (Fig. 3B, top panel; Supplemental Fig. S4). These RNAs were incubated with CPSF, UV-irradiated, and then treated with RNase T1. Under this condition, only the proteins that directly cross-link to the 3′ fragments (underlined in Fig. 3B, top panel) will be radiolabeled. Our results showed that Wdr33 cross-linked to SVL1–3 but not to SVL4 (Fig. 3B, bottom panel), suggesting that Wdr33 binds to sequences upstream of +12 nt, including the AAUAAA region. In contrast, CPSF30

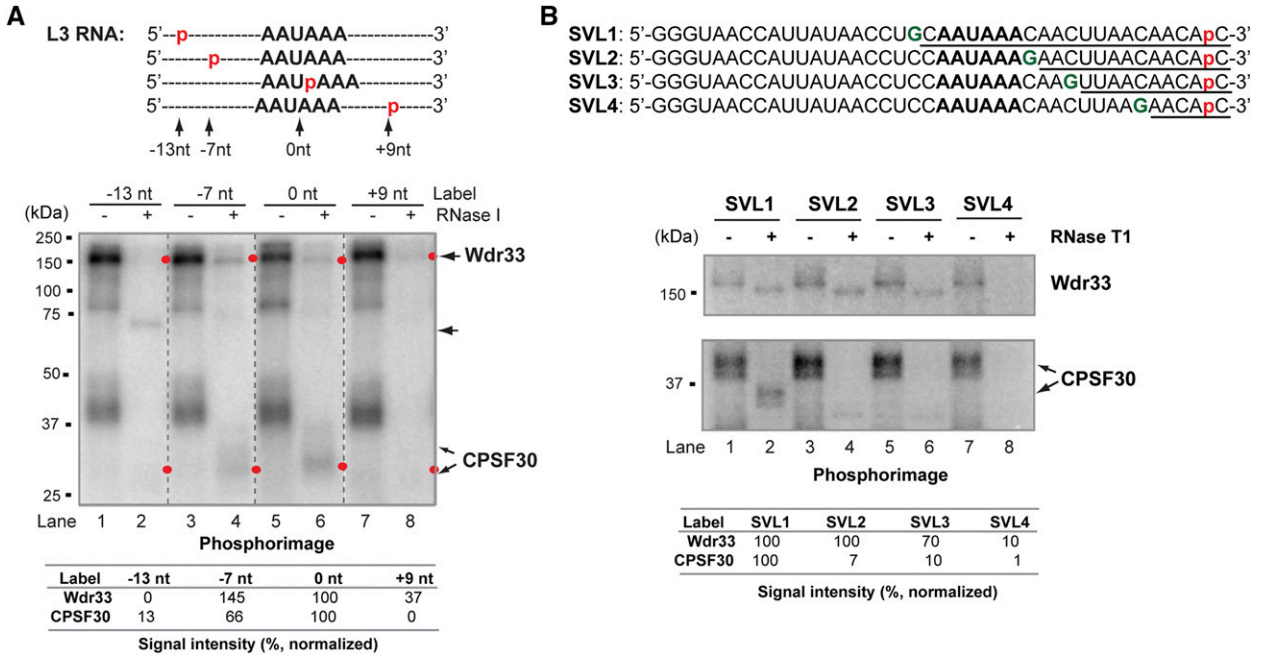


**Figure 3.** Mapping the footprints of CPSF30 and Wdr33 in the AAUAAA region in vitro. (*A*, *top* panel) The RNA substrates used. All RNAs have the same sequence as wild type (WT) in Figure 2 A, and the red "p" denotes the position of the single $^{32}$P label. Each RNA was incubated with CPSF, UV cross-linked, and treated with RNase I and visualized as described. The red dots mark the positions of Wdr33 and CPSF30. The signal intensities of these two species were quantified and normalized by those in the 0-nt sample (shown in lane *6*) and are listed in the *bottom* panel. (*B*) SVL-derived RNA substrates generated by in vitro transcription with T7 polymerase and 3′ radiolabeled with [$^{32}$P]pCp. Mapping was carried out the same way as described in *A* except that RNase T1 was used. A 6% gel was used to resolve the Wdr33 band (*top* blot), and a 10% gel was used for CPSF30 (*bottom* blot). Only proteins cross-linked to the underlined sequences remain radiolabeled. The signal intensities (normalized by the predigestion intensity) of CPSF30 and Wdr33 were normalized to those of SVL1 and are listed in the *bottom* panel.

bound to SVL1 but showed significantly lower interactions with SVL2 and SVL3, and no interaction was detected with SVL4 (Fig. 3B, bottom panel). These results provided evidence that CPSF30 binds directly to AAUAAA in SVL RNA. Together, our mapping results with two distinct RNA substrates suggest that both CPSF30 and Wdr33 bind specifically to a narrow region spanning the AAUAAA hexamer.

### *CPSF30 is essential for CPSF–RNA interaction*

Mammalian CPSF30 contains five ZF domains and one ZK domain (Fig. 4A). We next set out to map the protein domains of CPSF30 that mediate its interaction with RNA. To this end, we expressed Flag-tagged full-length (CPSF30 isoform 2: NP_001075028.1) or truncated CPSF30 in which one of the ZF or ZK domains was deleted in HEK293T cells by transient transfection and then purified them by immunoprecipitation with anti-Flag antibodies. As shown in Figure 4B (top panel), the full-length and truncated CPSF30 were specifically purified. We next incubated the recombinant CPSF30 with a radiolabeled AAUAAA-containing SVL RNA substrate followed by UV irradiation. Our results showed that the full-length CPSF30 was efficiently cross-linked to RNA (Fig. 4B, bottom panel). Although small amounts of intact CPSF complex might be copurified in this procedure, it does not seem to contribute significantly to the observed RNA interaction, as no Wdr33 cross-linking signal was observed (Fig. 4B, bottom panel). In addition, recombinant Flag-CPSF30, which was expressed using the baculovirus–insect cell system, also cross-linked to RNA (Supplemental Fig. S5). These results suggest that CPSF30 directly binds to RNA. However, CPSF30 UV cross-linking signals were nearly absent when ZF2 or ZF3 was deleted (Fig. 4B, bottom panel). Additionally, the UV cross-linking efficiency decreased by >60% when the ZK domain was removed (Fig. 4B, bottom panel). These data suggest that CPSF30–RNA interactions are primarily mediated by its ZF2 and ZF3 domains, while the ZK domain also contributes to this interaction.

To determine the role of CPSF30 in the overall CPSF–RNA interaction, we wanted to generate intact CPSF complexes containing a mutant CPSF30 that is deficient of RNA binding. To this end, we established HEK293 cell lines that stably express both Flag-tagged CPSF73 and HA-tagged full-length or truncated CPSF30 in which ZF2 or ZF3 was deleted (ΔZF2/3). Interestingly, the expression of exogenous CPSF30 led to a significant decrease in the level of the endogenous CPSF30 proteins (Supplemental Fig. S6), indicating a possible autoregulatory mechanism. A similar phenomenon has been described for a number of RNA-binding proteins, including splicing factors of the SR protein family (e.g., Wang et al. 1996). The low levels of the endogenous CPSF30 proteins in these cell lines

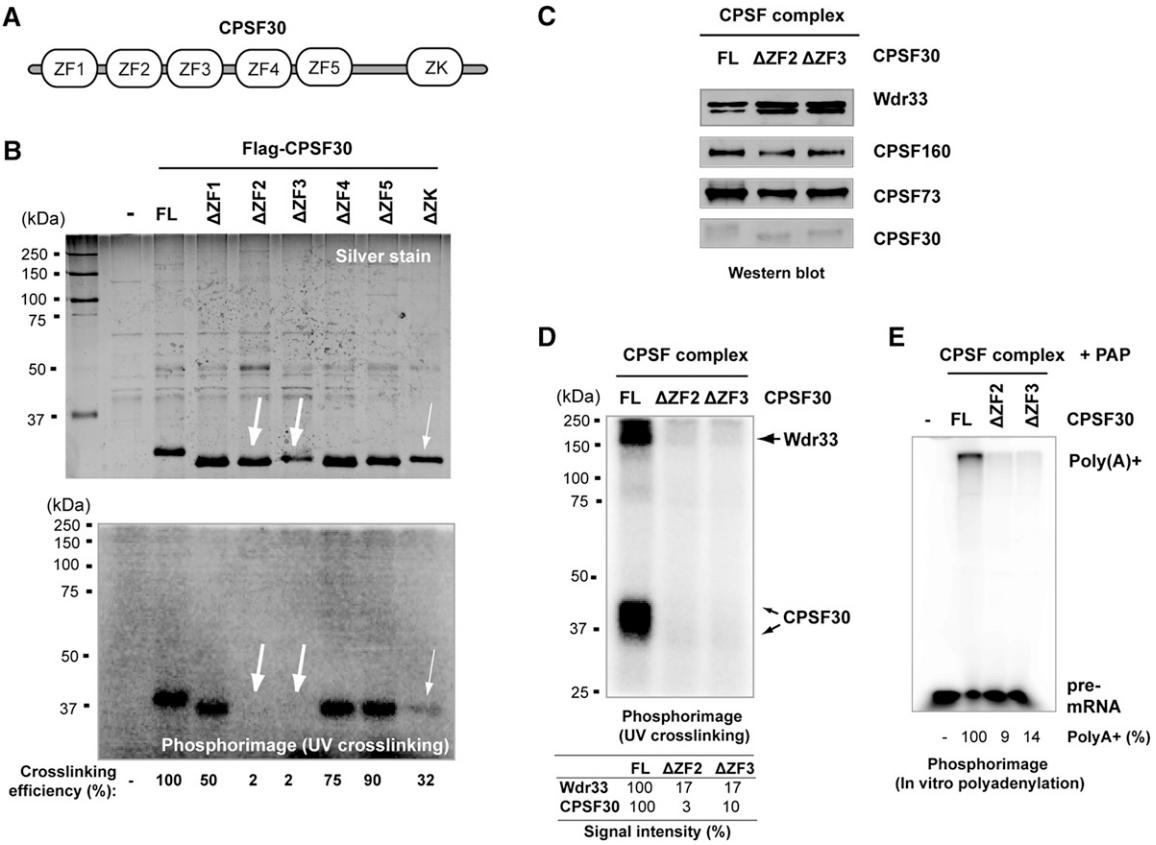


**Figure 4.** Mapping the CPSF30 domains necessary for RNA binding. (*A*) A schematic of the CPSF30 domain structure. Each circle represents a putative domain. The five ZF domains are marked as ZF1–ZF5. (*B*) Full-length (FL) or truncated CPSF30 in which one of the ZF or ZK domains are deleted (ΔZF1–5/ZK) were expressed in 293T cells by transient transfection and purified by immunoprecipitation (IP) using anti-Flag antibodies. (*Top* panel) The purified proteins were resolved by SDS-PAGE and visualized by silver staining. These proteins were incubated with SVL RNA substrate (wild type [WT] in Fig. 1A), UV-irradiated, resolved by SDS-PAGE, and visualized by phosphorimaging. The mutant proteins that showed the most significant differences in UV cross-linking compared with the full-length CPSF30 are marked by white arrows. Cross-linking efficiency was calculated as UV cross-linking signal intensity/protein staining intensity and normalized by that of the full-length sample and is listed. (*C*) CPSF complexes were purified by immunoprecipitation with anti-Flag antibodies and lysates from HEK293 cell lines that express CPSF73-3xFlag and HA-tagged full-length CPSF30 or the ZF2 or ZF2 deletion mutant (ΔZF2 or ΔZF2). The purified complexes were analyzed by Western blotting with antibodies against specified proteins. (*D*) CPSF complexes containing the full length or ΔZF2 or ΔZF3 were incubated with SVL RNA substrate, UV cross-linked, resolved by SDS-PAGE, and visualized by phosphorimaging. Wdr33 and CPSF30 signal intensities were normalized by those of the full-length sample and are listed at the *bottom*. (*E*) CPSF complexes containing the full length, ΔZF2, or ΔZF3 were incubated with SVL RNA substrate and recombinant PAP in the presence of ATP. The RNAs were resolved on an 8% urea gel and visualized by phosphorimaging. The percentage of poly(A)$^+$ RNA in each sample was calculated and normalized by that of the full-length sample and is listed.

allowed the purification of CPSF complexes that contain the full-length CPSF30, ΔZF2-CPSF30, or ΔZF3-CPSF30 by immunoprecipitation using anti-Flag antibodies. As shown in Figure 4C, Western blotting analyses detected similar levels of all tested CPSF subunits in these complexes, including the HA-tagged CPSF30. These results suggest that the deletion of either ZF2 or ZF3 did not interfere with the association between CPSF30 and the rest of the CPSF complex. We then incubated equivalent amounts of these complexes with radiolabeled AAUAAA-containing SVL RNAs followed by UV cross-linking. In line with our earlier results (Fig. 2A), Wdr33 and CPSF30 specifically cross-linked to the RNAs when CPSF complexes contained the full-length CPSF30 (Fig. 4D). However, when CPSF complex containing CPSF30-ΔZF2 or CPSF30-ΔZF3 was used, the cross-linking signals of both CPSF30 and Wdr33 dramatically decreased (Fig. 4D). Consistent with these results, when CPSF complexes containing the full-length CPSF30 or ΔZF2/3 mutant were used for in vitro polyadenylation assays, significantly lower levels of RNA polyadenylation were observed for CPSF30-ΔZF2/3-containing CPSF (Fig. 4E). In contrast, when CPSF containing CPSF30-ΔZK was used in the same assay, we did not observe a significant difference in polyadenylation efficiency compared with the CPSF complex containing the full-length CPSF30 (Supplemental Fig. S7). These results suggest that the CPSF30–RNA interaction is primarily mediated by its ZF2 and ZF3 domains and that CPSF30 RNA-binding activity is required for the entire CPSF complex to interact with RNA.

*Transcriptome-wide analyses of CPSF–RNA interaction in vivo*

In addition to the in vitro studies with selected RNA substrates as described above, we also characterized CPSF–RNA interactions in vivo through iCLIP (individual nucleotide resolution UV cross-linking and immunoprecipitation) analyses (Huppertz et al. 2014). Since CPSF160 was believed to be the main RNA-binding subunit of CPSF, we initially wanted to perform iCLIP analysis of CPSF160. To this end, we used a stable HeLa cell line that expresses GFP–CPSF160 in a doxycycline (Dox)-dependent manner (Supplemental Fig. S8A). After GFP-CPSF160 was induced, we irradiated the cells with UV-C (wavelength 254 nm) to induce protein–RNA cross-links. After immunoprecipitation with anti-GFP antibodies, however, very low amounts of RNA were recovered (Supplemental Fig. S8B), most likely due to low cross-linking efficiency. To overcome this limitation, we increased cross-linking efficiency through incorporation of the photoreactive ribonucleoside analog 4-thiouridine (4-SU) into nascent RNAs and then UV-A (wavelength 365 nm) irradiation (Hafner et al. 2010). Significantly higher amounts of RNA were recovered from GFP-CPSF160 immunoprecipitation under this condition (Supplemental Fig. S8B). To visualize the proteins that are cross-linked to RNAs, we next treated the cell lysates from UV cross-linked cells with RNase I, carried out immunoprecipitation using anti-GFP antibodies, radiolabeled the protein-associated RNAs, and resolved the immunoprecipitation sample by polyacrylamide gel electrophoresis. Strikingly, the most prominent radiolabeled species in the immunoprecipitation sample was not GFP-CPSF160 (~180 kDa) but an ~30-kDa band (Fig. 5A, lane 3). Based on our in vitro results, we suspected that this ~30-kDa species may correspond to CPSF30. To test this possibility, we subjected the anti-GFP immunoprecipitation sample to a second immunoprecipitation using anti-CPSF30 antibodies. Indeed, the ~30-kDa band was specifically immunoprecipitated by anti-CPSF30 antibodies (Fig. 5A, lane 6). These results point to CPSF30 as perhaps the major

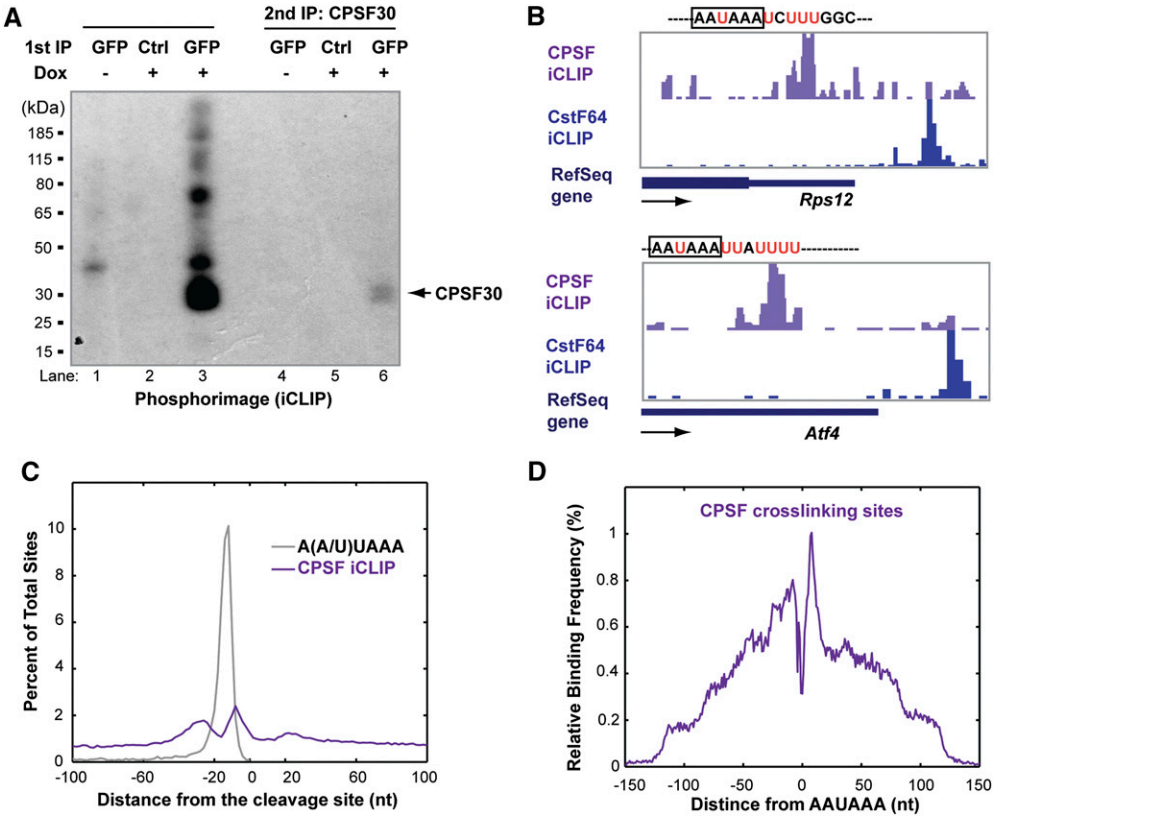


**Figure 5.** Transcriptome-wide analyses of CPSF–RNA interaction by iCLIP. (*A*) A tet-inducible GFP-CPSF160 expression HeLa cell line was untreated (−) or treated (+) with Dox, incubated with 4-SU, and UV-irradiated. (Lanes *1*–*3*) Cell lysates were treated with RNase I before being subjected to immunoprecipitation (IP) with control (Ctrl) or anti-GFP antibodies. For samples in lanes *4*–*6*, the immunoprecipitation samples were subjected to a second immunoprecipitation with anti-CPSF30 antibodies. All immunoprecipitation samples were 5′ labeled with $^{32}$P, resolved on a polyacrylamide gel, and visualized by phosphorimaging. (*B*) CPSF iCLIP tracks on the *Rps12* and *Atf4* genes. CstF64 iCLIP tracks based on data from Yao et al. (2012) were included for comparison. Each peak represents a unique cross-linking site. Nucleotide sequences around the iCLIP peaks are shown *above* the tracks. (*C*) Transcriptome-wide distribution of CPSF cross-linking sites relative to the cleavage sites (ends of transcripts). The gray line shows the distribution of the A(A/U)UAAA hexamer. The *Y*-axis is the percentage of CPSF cross-linking events or AAUAAA occurrence at that site over those in the entire region. (*D*) Transcriptome-wide distribution of CPSF cross-linking sites relative to AAUAAA (the underlined A corresponds to position 0 on the *X*-axis).

RNA-binding subunit of the CPSF complex in vivo. Although weak protein–RNA cross-linking signals were observed at ~160 kDa, they were significantly lower than the signals from CPSF30–RNAs (Fig. 5A, lane 3), indicating that Wdr33–RNA cross-linking products were not efficiently recovered under these conditions.

To analyze the in vivo CPSF–RNA interactions at the transcriptome level, we constructed iCLIP libraries from RNAs recovered in the anti-GFP immunoprecipitation sample and subjected them to high-throughput sequencing using the Illumina HiSeq platform. Using these data, we generated a CPSF–RNA interaction map at single-nucleotide resolution. As shown earlier (Fig. 5A), CPSF30 seems to be the major mediator of these interactions. Consistent with our in vitro data, CPSF–RNA in vivo cross-linking sites are highly enriched near the AAUAAA hexamer (Fig. 5B,C). As UV cross-linking only occurs at U residues in our procedure, CPSF cross-linking was often observed at the U runs nearest to the AAUAAA hexamer, as shown in the *Rps12* and *Atf4* transcripts (Fig. 5B). CPSF iCLIP signals are distinct from those of CstF64 (Fig. 5B; Yao et al. 2012), demonstrating the specificity of our iCLIP analyses. At the transcriptome level, CPSF cross-linking sites showed two peaks at −26 and −7 nt, respectively, relative to the cleavage site (Fig. 5C). The valley between the two peaks coincides with the position of the AAUAAA hexamer (Fig. 5C). To map CPSF-binding sites near AAUAAA at a higher resolution, we plotted the distribution of CPSF iCLIP sites relative to the AAUAAA itself. Strikingly, CPSF cross-linking sites displayed a bell-shaped peak that is only disrupted by a sharp valley near the center where AAUAAA resides (Fig. 5D). As mentioned earlier, such a decrease in CPSF cross-linking was most likely due to the low U content in the AAUAAA hexamer. As CPSF30 is the main mediator of CPSF–RNA interactions under the conditions used in this analysis (Fig. 5A), our iCLIP analyses provided in vivo evidence that CPSF30 binding encompasses the AAUAAA hexamer on a transcriptome-wide level.

## Discussion

A central question in the mRNA 3′ processing field has been how PASs are recognized and how PAS selection can be regulated. We provided critical and novel insights into the mechanism by which the AAUAAA hexamer is recognized by CPSF. Based on data presented here and previous studies, we propose a new model for CPSF–RNA interaction that involves multiple protein–RNA interactions (Fig. 6). First, we provided direct evidence that CPSF30 and Wdr33 directly bind to AAUAAA. These interactions anchor CPSF to RNA and provide specificity to mRNA 3′ processing. The individual contributions of CPSF30 and Wdr33 to CPSF RNA-binding specificity, however, remain to be determined. Second, previous in vitro and in vivo studies suggest that Fip1 binds to U-rich upstream sequences (Kaufmann et al. 2004; Martin et al. 2012; Lackford et al. 2014). Third, CPSF160 interacts with RNA over a broad region upstream of the cleavage sites, including the AAUAAA area (Gilmartin et al. 1995; Murthy and Manley 1995; Martin et al. 2012). Among these factors, we studied CPSF30 in detail and demonstrated that CPSF30 is essential for the entire CPSF complex to interact with RNAs. Interestingly, we found that CPSF30–RNA interactions are primarily mediated by its ZF2 and ZF3, the exact same domains targeted by the influenza protein NS1A to suppress host mRNA 3′ processing (Nemeroff et al. 1998; Twu et al. 2006). Together with these earlier studies, our results indicate that the influenza virus has evolved to precisely target this essential protein–RNA interaction to suppress host mRNA 3′ processing (Fig. 6). Additionally, it has been shown in plants that CPSF30 plays an important role in PAS selection, and its activity is regulated by calmodulin (Delaney et al. 2006; Thomas et al. 2012). Therefore, CPSF30–RNA interaction may serve as a hub for mRNA 3′ processing regulation.

We previously identified Wdr33 as a component of the mammalian mRNA 3′ processing complex and a bona fide subunit of the CPSF complex (Shi et al. 2009). The yeast Wdr33 homolog Pfs2p is a subunit of the CPF complex and an essential 3′ processing factor (Ohnacker et al. 2000). Pfs2 inactivation in fission yeast also leads to defects in chromosome replication and segregation (Wang et al. 2005). Wdr33/Pfs2 has been shown to regulate synapse and axon development in *Caenorhabditis elegans* (Van Epps et al. 2010). Mutations in the plant *Wdr33* homolog *FY* lead to aberrant mRNA alternative polyadenylation and defective flowering time control (Simpson et al. 2003). These studies strongly suggest that Wdr33 plays an essential role in mRNA 3′ processing and in mediating cross-talk between

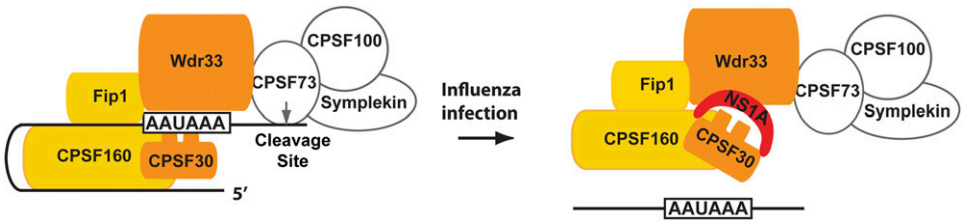


**Figure 6.** A new model for CPSF–RNA interaction. Multiple proteins mediate CPSF–RNA interactions: CPSF30 and Wdr33 directly interact with AAUAAA, and CPSF30 binds to AAUAAA via its ZF2 and ZF3 (shown as two fingers). Fip1 binds to auxiliary upstream sequences. CPSF160 binds to RNA over a broad region upstream of the cleavage sites. CPSF73, CPSF100, and symplekin do not participate in PAS recognition. During influenza viral infection, the viral protein NS1A specifically binds to the ZF2 and ZF3 of CPSF30 within the CPSF complex. As CPSF30–RNA interactions are primarily mediated by its ZF2 and ZF3 and are essential for CPSF binding to RNA, NS1A–CPSF30 interaction blocks CPSF–AAUAAA interaction and host mRNA 3′ processing.

mRNA 3′ processing and other cellular processes. However, the exact functions of Wdr33 in mRNA 3′ processing remained poorly understood. In this study, we provide in vitro evidence that Wdr33 directly binds to AAUAAA (Fig. 2). In addition, recent photoactivatable ribonucleoside-enhanced CLIP (PAR-CLIP) analyses suggest that Wdr33 binds to AAUAAA in vivo at the transcriptome level (see the accompanying study by Schönemann et al. 2014). Wdr33 contains a highly conserved WD40 repeat domain. This domain constitutes the majority of the yeast Pfs2p, while the metazoan Wdr33 proteins are significantly larger and contain additional domains (Chan et al. 2011; Xiang et al. 2014). Consisting of repeats of a 31- 60-residue motif that forms β-propeller structures, WD40 domains are best known for mediating protein–protein interactions (Stirnimann et al. 2010). However, several studies suggest that this domain may also participate in RNA interaction. For example, the WD40 domain of Gemin5 in the SMN complex binds to snRNAs in a sequence-specific manner (Lau et al. 2009). Two recent proteomic analyses of poly(A)$^+$ RNA-binding proteins identified 16 and 23 WD40-containig proteins in the mouse and human proteomes, respectively (Castello et al. 2012; Kwon et al. 2013). In future studies, it will be of great interest to characterize the Wdr33–RNA interactions through structural and biochemical analyses, which may provide important insights into the mechanisms of mRNA 3′ processing as well as WD40–RNA interactions in general.

CPSF160 has also long been implicated in RNA binding and AAUAAA recognition. Purified recombinant CPSF160 was found to bind AAUAAA with weak specificity in vitro (Murthy and Manley 1995), and the yeast CPSF160 homolog Yhh1/Cft1p avidly binds RNA around the PAS in vitro (Dichtl et al. 2002). While it is unclear why CPSF160 RNA binding was not detected in the present study or the accompanying one by Schönemann et al. (2014), our results suggest that the mechanism for AAUAAA recognition is more complex than previously thought. We propose several possible models to integrate our results with those from previous studies and help explain how these interactions could be accommodated and coordinated in the region around the AAUAAA. First, given the dynamic nature of mRNA 3′ processing (Chan et al. 2011), it is possible that CPSF160 and CPSF30/Wdr33 may bind to AAUAAA at different stages of the mRNA 3′ processing reaction. This may be similar to pre-mRNA splicing, in which the 5′ splice site is initially bound by the U1A protein and subsequently by U1 snRNA through base-pairing (Du and Rosbash 2002). Second, earlier biochemical studies suggested that the molecular weight of the CPSF complex ranged from ~350 kDa to ~500 kDa (Bienroth et al. 1991; Murthy and Manley 1992). More recent functional studies suggest that some CPSF subunits form stable subcomplexes (Sullivan et al. 2009; Schönemann et al. 2014). Therefore, there could be multiple alternative CPSF complexes or subcomplexes that rely on different protein–RNA interactions for AAUAAA recognition. Third, RNA sequences encompassing the AAUAAA hexamer could form secondary structures (Gilmartin et al. 1992; Hans and Alwine 2000), thereby allowing CPSF subunits to contact RNA in multiple locations. This may explain the previous observations that CPSF contacts both the AAUAAA hexamer and an enhancer element >70 nt upstream (Gilmartin et al. 1995) and that CPSF160 interacts with RNA in a broad region upstream of the cleavage sites (Martin et al. 2012). Finally, given the high degree of diversity in mammalian PAS sequences (Proudfoot 2011), CPSF160 and CPSF30/Wdr33 may be required for AAUAAA recognition in different PAS sequence contexts. This is reminiscent of the transcription factor TFIID, which relies on multiple protein–DNA interactions for the precise recognition of diverse promoter sequences (Cler et al. 2009). Interestingly, although CPSF160 shares little sequence homology with Wdr33, it was predicted to form a number of WD40-like β-propeller structures, including the putative conserved RNA-binding domain (Dichtl et al. 2002; Xiang et al. 2014). Therefore, both CPSF160 and Wdr33 may rely on β-propeller structures for their RNA-binding activities. In any event, our studies have provided evidence for a previously unappreciated complexity in AAUAAA recognition by CPSF.

## Materials and methods

### *RNA substrates*

SVL RNA and sequences used were as follows: SVL wild type, CUGCAAUAAACAAGUUAA (Figs. 1B,C, 4; Supplemental Figs. S5, S7); SVL mutant, CUGCAAGAAACAAGUUAA (Figs. 1B,C, 4); L3 wild type, AUGUACUAGGAGACACAAUAAAGGCAAU GUUUUUAUUUGUA (Figs. 2, 3A; Supplemental Figs. S1–S3); and L3 mutant, AUGUACUAGGAGACACAAUAACGGCAAU GUUUUUAUUUGUA (Fig. 2; Supplemental Fig. S1). The SVL RNA substrates were chemically synthesized and 5′ radiolabeled using Optikinase and γ-$^{32}$PATP. The L3 RNA substrates were site-specifically radiolabeled by splinted ligation of two chemically synthesized RNAs. For mapping experiments in Figure 3B, the RNAs were synthesized by in vitro transcription with DNA oligo templates and T7 polymerase.

### *Protein purification*

CPSF purification and mass spectrometry analyses were carried out as described before (Shi et al. 2009). To purify the CPSF complex shown in Figure 4, C–E, the HEK293-CPSF73-3xFlag cell line was retrovirally transduced to express HA-CPSF30 (full-length or truncated), and cell lysates were used for immunoprecipitation as described above. For purifying recombinant CPSF30 and its mutants, Flag-CPSF30 (full-length or truncated)-pCDNA3 plasmids were transiently transfected into HEK293T cells by using the calcium phosphate method. At 48 h post-transfection, whole-cell extracts were prepared from mock or transfected cells and used for immunoprecipitation.

### *In vitro assays*

For UV cross-linking assays, ~1 fmol of RNA substrates and ~60 fmol of CPSF were incubated and UV-irradiated according to a previously published procedure (Keller et al. 1991). Afterward, the samples were immediately mixed with SDS sample buffer or incubated with 20 U of RNase T1 or 60 U of RNase I for 5 min at 37°C. The cross-linking samples were resolved by SDS-PAGE and visualized by phosphorimaging. Immunoprecipitation of UV cross-linked protein was carried out as previously described (Luo et al. 1999) using the following antibodies: CPSF30 (Bethyl

Laboratories, A301-585A), Wdr33 (a gift from Dr. Rosenblatt), CPSF160 (Bethyl Laboratories, A301-580A), and Fip1 (A301-462A). In vitro polyadenylation assay reactions were assembled similarly to the UV cross-linking assay except that it contained 0.8 mM ATP and 20 ng of recombinant PAP. The samples were then incubated for 1.5 h at 30°C, and the extracted RNAs were resolved on an 8% urea gel and visualized by phosphorimaging.

*Cell culture*

HEK293, 293T, and HeLa FlpIn cell lines containing inducible GFP-CPSF160 were maintained in Dulbecco's modified Eagle medium (DMEM) with 10% fetal bovine serum (FBS) supplemented with the appropriate antibiotics (CPSF73-3xFlag cell line: 200 μg/mL G418; HeLa FlpIn GFP-CPSF160 cell line: 100 μg/mL hygromycin, 4 μg/mL blasticidine, 10 U/mL penicillin, 10 μg/mL streptomycin).

*iCLIP*

For iCLIP experiments, cells were preincubated with 4-SU for 1 h and cross-linked with $2 \times 400$ mJ/cm$^2$ and 365-nm UV light. Protein G Dynabeads (Life Technologies) were used for immunoprecipitation. iCLIP library construction and data analyses were performed as described (Huppertz et al. 2014). Sequencing data have been deposited to EMBL-EBI ArrayExpress Archive (accession no. E-MTAB-2939).

## Acknowledgments

We thank Dr. L. Tong and Dr. K.J. Hertel for critical reading of the manuscript, E. Wahle for communicating unpublished results, K. Neugebauer and O. Rosenblatt for providing antibodies, and S.-C. Ling for providing the GFP-CPSF160 cell line. This study was supported by the following grants: National Institutes of Health (NIH) GM090056 and American Cancer Society RSG-12-186 to Y.S., the University of California at Irvine Eugene Cota-Robles Fellowship to S.L.C., the Marie Curie Initial Training Network RNPnet 289007 to I.H., the National Center for Research Resources 5P41RR011823 and National Institute of General Medical Sciences 8 P41 GM103533 to J.R.Y., and NIH GM28983 to J.L.M.

## References

Barabino SM, Hubner W, Jenny A, Minvielle-Sebastia L, Keller W. 1997. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev* **11:** 1703–1716.

Bienroth S, Wahle E, Suter-Crazzolara C, Keller W. 1991. Purification of the cleavage and polyadenylation factor involved in the 3′-processing of messenger RNA precursors. *J Biol Chem* **266:** 19768–19776.

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149:** 1393–1406.

Chan S, Choi EA, Shi Y. 2011. Pre-mRNA 3′-end processing complex assembly and function. *Wiley Interdiscip Rev RNA* **2:** 321–335.

Christofori G, Keller W. 1988. 3′ cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP. *Cell* **54:** 875–889.

Cler E, Papai G, Schultz P, Davidson I. 2009. Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell Mol Life Sci* **66:** 2123–2134.

Colgan DF, Manley JL. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* **11:** 2755–2766.

Danckwardt S, Hentze MW, Kulozik AE. 2008. 3′ end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J* **27:** 482–498.

Delaney KJ, Xu R, Zhang J, Li QQ, Yun KY, Falcone DL, Hunt AG. 2006. Calmodulin interacts with and regulates the RNA-binding activity of an *Arabidopsis* polyadenylation factor subunit. *Plant Physiol* **140:** 1507–1521.

Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and consequences of alternative polyadenylation. *Mol Cell* **43:** 853–866.

Dichtl B, Blank D, Sadowski M, Hubner W, Weiser S, Keller W. 2002. Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO J* **21:** 4125–4135.

Du H, Rosbash M. 2002. The U1 snRNP protein U1C recognizes the 5′ splice site in the absence of base pairing. *Nature* **419:** 86–90.

Gilmartin GM, Fleming ES, Oetjen J. 1992. Activation of HIV-1 pre-mRNA 3′ processing in vitro requires both an upstream element and TAR. *EMBO J* **11:** 4419–4428.

Gilmartin GM, Fleming ES, Oetjen J, Graveley BR. 1995. CPSF recognition of an HIV-1 mRNA 3′-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev* **9:** 72–83.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141:** 129–141.

Hans H, Alwine JC. 2000. Functionally significant secondary structure of the simian virus 40 late polyadenylation signal. *Mol Cell Biol* **20:** 2926–2932.

Huppertz I, Attig J, D'Ambrogio A, Easton LE, Sibley CR, Sugimoto Y, Tajnik M, Konig J, Ule J. 2014. iCLIP: protein–RNA interactions at nucleotide resolution. *Methods* **65:** 274–287.

Kaufmann I, Martin G, Friedlein A, Langen H, Keller W. 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J* **23:** 616–626.

Keller W, Bienroth S, Lang KM, Christofori G. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. *EMBO J* **10:** 4241–4249.

Kwon SC, Yi H, Eichelbaum K, Fohr S, Fischer B, You KT, Castello A, Krijgsveld J, Hentze MW, Kim VN. 2013. The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol* **20:** 1122–1130.

Lackford B, Yao C, Charles GM, Weng L, Zheng X, Choi EA, Xie X, Wan J, Xing Y, Freudenberg JM, et al. 2014. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J* **33:** 878–889.

Lau CK, Bachorik JL, Dreyfuss G. 2009. Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains. *Nat Struct Mol Biol* **16:** 486–491.

Luo HR, Moreau GA, Levin N, Moore MJ. 1999. The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. *RNA* **5:** 893–908.

Mandel CR, Bai Y, Tong L. 2008. Protein factors in pre-mRNA 3′-end processing. *Cell Mol Life Sci* **65:** 1099–1122.

Martin G, Gruber AR, Keller W, Zavolan M. 2012. Genome-wide analysis of Pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3′ UTR length. *Cell Reports* **1:** 753–763.

Millevoi S, Vagner S. 2009. Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. *Nucleic Acids Res* **38:** 2757–2774.

Murthy KG, Manley JL. 1992. Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *J Biol Chem* **267:** 14804–14811.

Murthy KG, Manley JL. 1995. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3′-end formation. *Genes Dev* **9:** 2672–2683.

Nemeroff ME, Barabino SM, Li Y, Keller W, Krug RM. 1998. Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3′end formation of cellular pre-mRNAs. *Mol Cell* **1:** 991–1000.

Ohnacker M, Barabino SM, Preker PJ, Keller W. 2000. The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3′-end-processing complex. *EMBO J* **19:** 37–47.

Proudfoot NJ. 2011. Ending the message: poly(A) signals then and now. *Genes Dev* **25:** 1770–1782.

Schönemann L, Kühn U, Martin G, Schäfer P, Gruber AR, Keller W, Zavolan M, Wahle E. 2014. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev* (this issue). doi: 10.1101/gad.250985.114.

Sheets MD, Ogg SC, Wickens MP. 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res* **18:** 5799–5805.

Shi Y. 2012. Alternative polyadenylation: new insights from global analyses. *RNA* **18:** 2105–2117.

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL. 2009. Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol Cell* **33:** 365–376.

Simpson GG, Dijkwel PP, Quesada V, Henderson I, Dean C. 2003. FY is an RNA 3′ end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. *Cell* **113:** 777–787.

Stirnimann CU, Petsalaki E, Russell RB, Muller CW. 2010. WD40 proteins propel cellular networks. *Trends Biochem Sci* **35:** 565–574.

Sullivan KD, Steiniger M, Marzluff WF. 2009. A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. *Mol Cell* **34:** 322–332.

Takagaki Y, Ryner LC, Manley JL. 1988. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* **52:** 731–742.

Thomas PE, Wu X, Liu M, Gaffney B, Ji G, Li QQ, Hunt AG. 2012. Genome-wide control of polyadenylation site choice by CPSF30 in *Arabidopsis*. *Plant Cell* **24:** 4376–4388.

Tian B, Manley JL. 2013. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* **38:** 312–320.

Twu KY, Noah DL, Rao P, Kuo RL, Krug RM. 2006. The CPSF30 binding site on the NS1A protein of influenza A virus is a potential antiviral target. *J Virol* **80:** 3957–3965.

Van Epps H, Dai Y, Qi Y, Goncharov A, Jin Y. 2010. Nuclear pre-mRNA 3′-end processing regulates synapse and axon development in *C. elegans*. *Development* **137:** 2237–2250.

Wang J, Takagaki Y, Manley JL. 1996. Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes Dev* **10:** 2588–2599.

Wang SW, Asakawa K, Win TZ, Toda T, Norbury CJ. 2005. Inactivation of the pre-mRNA cleavage and polyadenylation factor Pfs2 in fission yeast causes lethal cell cycle defects. *Mol Cell Biol* **25:** 2288–2296.

Xiang K, Tong L, Manley JL. 2014. Delineating the structural blueprint of the pre-mRNA 3′-end processing machinery. *Mol Cell Biol* **34:** 1894–1910.

Yao C, Biesinger J, Wan J, Weng L, Xing Y, Xie X, Shi Y. 2012. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc Natl Acad Sci* **109:** 18773–18778.

Zhao J, Hyman L, Moore C. 1999. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* **63:** 405–445.