

POINT OF VIEW

From polyadenylation to splicing: Dual role for mRNA 3' end formation factors

Ashish Misra and Michael R. Green

Howard Hughes Medical Institute and Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA USA

ABSTRACT

Recent genome-wide protein–RNA interaction studies have significantly reshaped our understanding of the role of mRNA 3' end formation factors in RNA biology. Originally thought to function solely in mediating cleavage and polyadenylation of mRNAs during their maturation, 3' end formation factors have now been shown to play a role in alternative splicing, even at internal introns—an unanticipated role for factors thought only to act at the 3' end of the mRNA. Here, we discuss the recent advances in our understanding of the role of 3' end formation factors in promoting global changes in alternative splicing at internal exon-intron junctions and how they act as cofactors for well known splicing regulators. Additionally, we review the mechanism by which these factors affect the recruitment of early intron recognition components to the 5' and 3' splice site. Our understanding of the roles of 3' end formation factors is still evolving, and the final picture might be more complex than originally envisioned.

Abbreviations: CLIP, genome-wide crosslinking immunoprecipitation; CFIm/CFIIm, cleavage factor Im/IIm; CPSF, cleavage polyadenylation specificity factor; CSTF, cleavage stimulation factor; poly(A), polyadenosine; snRNP, small nuclear ribonucleoprotein; UTR, untranslated region.

ARTICLE HISTORY

Received 11 September 2015
Revised 15 October 2015
Accepted 19 October 2015

KEYWORDS

Alternative splicing; CPSF; mRNA 3' end processing; polyadenylation; SYMPK

Polyadenylation complexes and their Role in 3' end processing

The maturation of most pre-mRNAs to mRNAs occurs co-transcriptionally with the addition of a 7-methyl guanosine cap to the 5' end, removal of intronic sequences by splicing, and endonucleolytic cleavage at the 3' end followed by the synthesis of a polyadenosine [poly(A)] tail. The site of 3' end cleavage is determined by the interaction of specific sequence elements within the pre-mRNA with a multiprotein complex called the 3' end processing complex, a macromolecular assembly consisting of poly(A) polymerase and 4 multisubunit complexes: cleavage polyadenylation specificity factor (CPSF), cleavage factors Im and IIm (CFIm and CFIIm), and cleavage stimulation factor (CSTF) (Fig. 1).

The core component of the 3' end processing complex is CPSF. CPSF specifically recognizes the core sequence element AAUAAA in the polyadenylation site and catalyzes cleavage. The multisubunit CPSF complex consists of CPSF1 (also called CPSF160), CPSF2 (also called CPSF100), CPSF3 (the cleavage endonuclease; also called CPSF73), CPSF4 (also called CPSF30), FIP1L1 (also called FIP1), and WDR33. Against the long held

understanding that the CPSF1 subunit directly contacts RNA via the AAUAAA sequence, recent genome-wide crosslinking immunoprecipitation (CLIP) studies combined with in vitro observations have revealed that it is WDR33 and CPSF4 that directly contact the AAUAAA element.^{1,2} FIP1L1 has been shown to bind to U-rich sequences upstream of AAUAAA and, along with CPSF1, is responsible for recruiting poly(A) polymerase to the 3' end processing site.^{1,3,4}

The CFIm complex facilitates assembly of the 3' end processing complex and enhances the efficiency of polyadenylation site cleavage. CFIm consists of NUDT21 (also called CFIm25) and one of the 2 related proteins, CPSF7 (also called CFIm59) or CPSF6 (also called CFIm68). All three CFIm subunits are involved in binding to the RNA element UGUA, which is located upstream of the AAUAAA sequence.⁵

The CSTF complex consists of CSTF1 (CSTF50), CSTF2 (CSTF64) and CSTF3 (also called CSTF77). CSTF2 specifically recognizes the U/GU-rich rich elements located downstream of the AAUAAA sequence, and is required for cleavage but not polyadenylation.^{6,7}

CPSF, CSTF, and CFIm directly bind to RNAs to form a core complex that, in turn, recruits other factors,

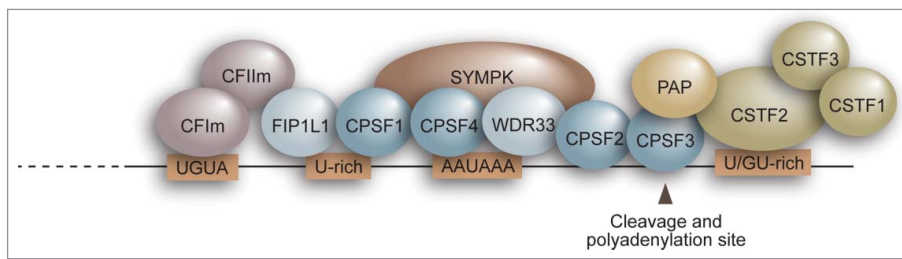


Figure 1. Schematic representation showing proteins involved in mRNA 3' end processing and the RNA sequence elements to which they bind. PAP, poly(A) polymerase.

including CFIIm, the scaffolding protein SYMPK, and poly(A) polymerase, to assemble the active 3' end processing complex.^{1,8}

Apart from the above-mentioned complex involved in 3' end processing of a large number of mRNAs, several lines of evidence point to the existence of heterogeneous CPSF complexes. For example, Schonemann et al.² showed that a relatively stable complex can be reconstituted with WDR33, CPSF4, FIP1L1, and CPSF1 and is active, with poly(A) polymerase, in AAUAAA-dependent poly(A) synthesis. Moreover, a distinct subcomplex consisting of CPSF2, CPSF3 and SYMPK functions in 3' processing of nonpolyadenylated histone mRNA precursors.⁹

Functional links between polyadenylation and splicing

Pre-mRNA splicing is catalyzed by a dynamic ribonucleoprotein complex called the spliceosome, which consists of 5 small nuclear ribonucleoproteins (snRNPs) and numerous associated proteins. Initial hints about the crosstalk between splicing and polyadenylation came from in vitro experiments in which single- or multi-intron pre-mRNA substrates were added to nuclear extracts that were competent for both splicing and 3' end processing. The results revealed that whereas splicing of a single-intron pre-mRNA required active 3' end processing, mutation of the AAUAAA sequence in a multi-intron pre-mRNA inhibited splicing of the 3' terminal intron but not internal introns.¹⁰⁻¹² Two factors, poly(A) polymerase and the essential splicing factor U2AF65, were implicated in the coupling of 3' end processing and the splicing of the 3' terminal intron. Based on these results, it was postulated that the polyadenylation site functions essentially as a splicing enhancer, with splicing enhancement mediated by specific interactions between poly(A) polymerase and U2AF65.^{12,13} These observations were later confirmed by in vivo experiments showing that CPSF4 was required for

splicing of a 3' terminal intron.¹⁴ Similar experiments by Kyburz et al.¹⁵ showed that direct interaction between the spliceosome component U2 snRNP and CPSF helps enhance 3' end processing.

Additional evidence of a link between splicing and polyadenylation was put forward by an RNA-seq study that examined the transcriptomes of a diverse panel of human tissues and cell lines.¹⁶ They observed a strong correlation between patterns of alternative splicing and alternative polyadenylation, suggesting these processes were coordinately regulated. Furthermore, they also detected a strong enrichment of well-known splicing-related regulatory motifs, such as those for the well characterized splicing regulators RBFOX2 and NOVA2, in 3' untranslated regions (UTRs), suggesting that these factors might have a role in regulating both splicing and alternative polyadenylation. These observations were later supported by genome-wide RNA-protein interaction maps of RBFOX2 and NOVA2, which revealed a number of RNA-protein interactions in 3' UTRs, suggesting a role for RBFOX2 and NOVA2 in regulating alternative polyadenylation.¹⁷⁻¹⁹ Furthermore, experiments performed in *Nova2*-knockout mice showed that they have shortened 3' UTRs compared to control animals, suggesting that NOVA2 promotes the production of isoforms with longer 3' UTRs.¹⁷

Several other studies have demonstrated a dual role for proteins in both splicing and polyadenylation. For example, knockdown of U1 snRNP, a spliceosome component that plays a critical role in 5' splice site recognition, results in increased usage of cryptic intronic polyadenylation sites.²⁰ These observations pointed toward a second major role of U1 snRNP beyond its critical function in splicing. Interestingly, a role for cytoplasmic polyadenylation element binding protein 1 (CPEB1) in alternative polyadenylation regulation and alternative splicing was also reported recently.²¹ It was shown that when the CPE-driven alternative polyadenylation site is present in the proximity of an alternative splice site, the CPEB1-CPSF complex interferes with the recruitment of the splicing machinery, thereby affecting splicing.

Role of the CPSF/SYMPK complex in promoting alternative splicing

Our recent study combining genome-wide RNA interference screening with genome-wide RNA association and transcriptome-wide alternative splicing revealed that CPSF and SYMPK, but not other mRNA 3' end formation factors, have a global role in promoting alternative splicing at internal exon-intron junctions.²² These factors are involved in promoting both exon inclusion and exon exclusion. Genome-wide individual-nucleotide resolution CLIP (iCLIP) analysis identified a large number of CPSF2 binding peaks in coding regions and introns within the mRNA and, contrary to what was expected, did not show a distinct binding peak near the cleavage and polyadenylation site. Our iCLIP results are consistent with genome-wide protein–RNA association studies showing that CPSF proteins do not show preferential binding near the cleavage and polyadenylation site.²³

Although previous studies had shown that CPSF1, CPSF3 and CPSF4 bind directly to RNA,^{4,24,25} our study was the first to demonstrate a direct RNA-binding activity for CPSF2. Our results are consistent with previous observations with the yeast CPSF2 homolog, Ydh1, which has been shown to bind directly to RNA in UV cross-linking experiments.^{26,27} Like CPSF3, CPSF2 belongs to a special class of nucleic acid-processing enzymes, called the β -CASP family of proteins, that contain a metallo- β -lactamase fold as well as a distinct globular (β -CASP) domain; however, CPSF2 lacks residues in the metallo- β -lactamase motif that are critical for zinc binding, making it incapable of catalysis. We employed RNA pull-down experiments using purified proteins containing domain truncations to show that the metallo- β -lactamase and β -CASP motifs of CPSF2, but not the C-terminal domain (amino acids 669–782), are required for its RNA-binding activity.²² Detailed biochemical analysis will be required to pinpoint the residues underlying the sequence-specific RNA-binding property of CPSF2. Interestingly, CPSF2 and CPSF3 are known to be tightly associated,²⁸ and one intriguing possibility is that the direct RNA-binding activity of CPSF2 may be critical for enhancing the catalytic (endonuclease) activity of CPSF3.

Combining biochemical experiments with bioinformatic analysis, we showed that the CPSF/SYMPK complex was recruited to RNA by well-known splicing factors (and RNA-binding proteins) such as RBFOX2, NOVA2 and HNRNPA1.²² The analysis also revealed that not all splicing events regulated by RBFOX2 require CPSF/SYMPK. At splicing events regulated by both RBFOX2 and CPSF/SYMPK, CPSF/SYMPK is required for alternative splicing and is

either recruited by pre-mRNA-bound RBFOX2 or associates with the pre-mRNA as an RBFOX2/CPSF/SYMPK complex. By contrast, at RBFOX2-regulated, CPSF/SYMPK-independent exons, CPSF/SYMPK is neither required for alternative splicing nor bound nearby. These results point toward the possibility that RBFOX2 exists in both free and CPSF/SYMPK-associated forms, or that there are other proteins substituting for the role of the CPSF/SYMPK complex at CPSF/SYMPK-independent exons. Further experiments are required to understand the mechanistic basis underlying this difference in CPSF/SYMPK binding. One possibility is that differences in the intrinsic strength of RBFOX2- and/or CPSF-binding sites contributes to differential CPSF/SYMPK binding. For example, strong RBFOX2-binding sites may be able to recruit free RBFOX2, whereas weak RBFOX2-binding sites may require the additional binding enhancement provided by CPSF/SYMPK in order to achieve stable RBFOX2 association. It is also possible that there are other sequence and/or structural features within the RNA that contribute to these differences in binding. For instance, the CPSF/SYMPK complex may only be required to facilitate splicing in cases where the exon-intron boundary is not clearly defined, or where the splice sites and/or binding sites for splicing factors (RBFOX2, NOVA2 or HNRNPA1) are masked by secondary structures present within the RNA. Another factor that might have a role in alternative splicing events being dependent or independent of CPSF/SYMPK could be splice site strength. Our analysis of splice site strength of exon exclusion events showed that CPSF/SYMPK-dependent exons have significantly weaker 3' splice sites compared to CPSF/SYMPK-independent exons.²²

Once bound to the RNA, the CPSF/SYMPK complex functions as a splicing cofactor by facilitating the ability of RBFOX2, NOVA2, or HNRNPA1 to regulate binding of the early intron recognition factors U2AF and U1 snRNP.²² Our findings are consistent with the general notion that frequently it is the earliest events in spliceosome assembly that are modulated to regulate alternative splicing. Through a series of RNA pull-down, mutational analysis and directed CLIP experiments, we showed that in the case of exon inclusion, binding of RBFOX2/NOVA2/HNRNPA1 and CPSF/SYMPK leads to recruitment of the U1 snRNP component U1 70K and U2AF65 to the 5' splice site and 3' splice site, respectively (Fig. 2, top). By contrast, in the case of exon exclusion, binding of RBFOX2/NOVA2/HNRNPA1 and CPSF/SYMPK interferes with recruitment of U2AF

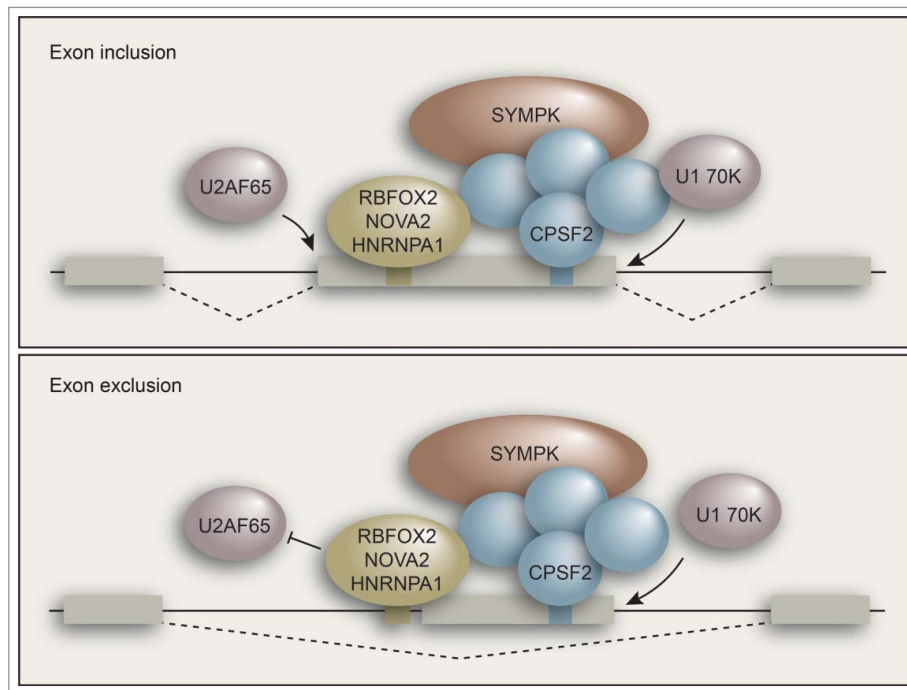


Figure 2. Model for CPSF/SYMPK-mediated regulation of alternative splicing of internal exons.

to the 3' splice site upstream of the excluded exon (Fig. 2, bottom).

Meta-analysis of our iCLIP and RNA-seq data showed that in cases of exon inclusion, the binding of RBFOX2 and CPSF2 is enriched inside the alternatively spliced exon and near the 5' splice site, which is consistent with their role in promoting 5' splice site recognition. In the case of exon exclusion, RBFOX2 and CPSF2 binding is significantly enriched immediately upstream of the 3' splice site, suggesting that they might sterically interfere with U2AF65 binding to the 3' splice site. However, we cannot rule out the possibility that they could interfere with communication between U1 snRNP and U2AF and thereby affect splicing. Further experiments will be required to pinpoint the exact mechanism by which the CPSF/SYMPK complex contributes to exon exclusion.

Collectively, our results show a strong correlation between CPSF/SYMPK affected splicing outcome and U2AF65 and U1 70K binding to RNA. However, genome-wide CLIP analysis of U2AF65 and U1 70K in CPSF and/or RBFOX2/NOVA2/HNRNPA1 knock-down cells are required to generalize our mechanistic observations at the transcriptome-wide level. Such studies are also likely to bring exceptions to the rule, which would be interesting cases for future studies. Furthermore, development of an efficient *in vitro* splicing assay system will be required to complement our cell-based studies and investigate the detailed

mechanism of CPSF/SYMPK-mediated splicing regulation.

Concluding remarks

The mechanistic basis by which the CPSF/SYMPK complex affects alternative splicing needs further exploration. It will be particularly important to determine if the CPSF/SYMPK complex plays a role in facilitating RBFOX- and NOVA-directed regulation of alternative polyadenylation. It will also be interesting to study the underlying reasons why CPSF/SYMPK affects splicing at only a subset of RBFOX2-regulated alternative splicing events. Are there other proteins that substitute for CPSF/SYMPK at those junctions? Comprehensive proteomic analysis will be required to address this issue. Proteomic analysis will also allow us to identify other proteins that might be part of this complex and have a role in promoting alternative splicing. Another question arising from our studies is whether the CPSF/SYMPK complex is capable of carrying out splicing on its own. Although knockdown and overexpression experiments of single components have already provided valuable clues into the role of CPSF/SYMPK factors, further analysis to study the synergy of different proteins in regulating splicing and polyadenylation will be needed in order to more precisely decode the role of these factors and identify the other sub-complexes that might exist. We

envisage that answering these questions will significantly advance our understanding of the role of CPSF/SYMPK in regulating alternative splicing and polyadenylation.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Sara Deibler for editorial assistance.

Funding

This work was supported by a grant from the National Institutes of Health (R01GM035490) to M.R.G. M.R.G. is an investigator of the Howard Hughes Medical Institute.

References

- Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR, 3rd, Ule J, Manley JL, Shi Y. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev* 2014; 28:2370-80; PMID:25301780; <http://dx.doi.org/10.1101/gad.250993.114>
- Schonemann L, Kuhn U, Martin G, Schafer P, Gruber AR, Keller W, Zavolan M, Wahle E. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev* 2014; 28:2381-93; PMID:25301781; <http://dx.doi.org/10.1101/gad.250985.114>
- Kaufmann I, Martin G, Friedlein A, Langen H, Keller W. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J* 2004; 23:616-26; PMID:14749727; <http://dx.doi.org/10.1038/sj.emboj.7600070>
- Murthy KG, Manley JL. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* 1995; 9:2672-83; PMID:7590244; <http://dx.doi.org/10.1101/gad.9.21.2672>
- Yang Q, Gilmartin GM, Doublet S. Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. *Proc Natl Acad Sci U S A* 2010; 107:10062-7; PMID:20479262; <http://dx.doi.org/10.1073/pnas.1000848107>
- MacDonald CC, Wilusz J, Shenk T. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol Cell Biol* 1994; 14:6647-54; PMID:7935383; <http://dx.doi.org/10.1128/MCB.14.10.6647>
- Takagaki Y, Manley JL. RNA recognition by the human polyadenylation factor CstF. *Mol Cell Biol* 1997; 17:3907-14; PMID:9199325; <http://dx.doi.org/10.1128/MCB.17.7.3907>
- Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, 3rd, Frank J, Manley JL. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 2009; 33:365-76; PMID:19217410; <http://dx.doi.org/10.1016/j.molcel.2008.12.028>
- Sullivan KD, Steiniger M, Marzluff WF. A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. *Mol Cell* 2009; 34:322-32; PMID:19450530; <http://dx.doi.org/10.1016/j.molcel.2009.04.024>
- Berget SM. Exon recognition in vertebrate splicing. *J Biol Chem* 1995; 270:2411-4; PMID:7852296; <http://dx.doi.org/10.1074/jbc.270.6.2411>
- Niwa M, Berget SM. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev* 1991; 5:2086-95; PMID:1657710; <http://dx.doi.org/10.1101/gad.5.11.2086>
- Vagner S, Vagner C, Mattaj IW. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev* 2000; 14:403-13; PMID:10691733; <http://dx.doi.org/10.1101/gad.14.4.403>
- Gunderson SI, Vagner S, Polycarpou-Schwarz M, Mattaj IW. Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. *Genes Dev* 1997; 11:761-73; PMID:9087430; <http://dx.doi.org/10.1101/gad.11.6.761>
- Li Y, Chen ZY, Wang W, Baker CC, Krug RM. The 3'-end-processing factor CPSF is required for the splicing of single-intron pre-mRNAs in vivo. *RNA* 2001; 7:920-31; PMID:11421366; <http://dx.doi.org/10.1017/S1355838201010226>
- Kyburz A, Friedlein A, Langen H, Keller W. Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Mol Cell* 2006; 23:195-205; PMID:16857586; <http://dx.doi.org/10.1016/j.molcel.2006.05.037>
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008; 456:470-6; PMID:18978772; <http://dx.doi.org/10.1038/nature07509>
- Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 2008; 456:464-9; PMID:18978773; <http://dx.doi.org/10.1038/nature07488>
- Weyn-Vanhenhenryck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z, Xue C, Herre M, Silver PA, Zhang MQ, et al. HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep* 2014; 6:1139-52; PMID:24613350; <http://dx.doi.org/10.1016/j.celrep.2014.02.005>
- Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem

- cells. *Nat Struct Mol Biol* 2009; 16:130-7; PMID:19136955; <http://dx.doi.org/10.1038/nsmb.1545>
20. Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 2010; 468:664-8; PMID:20881964; <http://dx.doi.org/10.1038/nature09479>
 21. Bava FA, Eliscovich C, Ferreira PG, Minana B, Ben-Dov C, Guigo R, Valcarcel J, Mendez R. CPEB1 coordinates alternative 3'-UTR formation with translational regulation. *Nature* 2013; 495:121-5; PMID:23434754; <http://dx.doi.org/10.1038/nature11901>
 22. Misra A, Ou J, Zhu LJ, Green MR. Global Promotion of Alternative Internal Exon Usage by mRNA 3' End Formation Factors. *Mol Cell* 2015; 58:819-31; PMID:25921069; <http://dx.doi.org/10.1016/j.molcel.2015.03.016>
 23. Martin G, Gruber AR, Keller W, Zavolan M. 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 Rep* 2012; 1:753-63; PMID:22813749; <http://dx.doi.org/10.1016/j.celrep.2012.05.003>
 24. Barabino SM, Hubner W, Jenny A, Minvielle-Sebastia L, Keller W. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev* 1997; 11:1703-16; PMID:9224719; <http://dx.doi.org/10.1101/gad.11.13.1703>
 25. Ryan K, Calvo O, Manley JL. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* 2004; 10:565-73; PMID:15037765; <http://dx.doi.org/10.1261/rna.5214404>
 26. Dichtl B, Keller W. Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. *EMBO J* 2001; 20:3197-209; PMID:11406596; <http://dx.doi.org/10.1093/emboj/20.12.3197>
 27. Zhao J, Kessler MM, Moore CL. Cleavage factor II of *Saccharomyces cerevisiae* contains homologues to subunits of the mammalian Cleavage/ polyadenylation specificity factor and exhibits sequence-specific, ATP-dependent interaction with precursor RNA. *J Biol Chem* 1997; 272:10831-8; PMID:9099738; <http://dx.doi.org/10.1074/jbc.272.16.10831>
 28. Dominski Z, Yang XC, Purdy M, Wagner EJ, Marzluff WF. A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol Cell Biol* 2005; 25:1489-500; PMID:15684398; <http://dx.doi.org/10.1128/MCB.25.4.1489-1500.2005>