Published in final edited form as:

*Nat Rev Mol Cell Biol*. 2017 January ; 18(1): 18–30. doi:10.1038/nrm.2016.116.

# Alternative polyadenylation of mRNA precursors

**Bin Tian**[1] and **James L. Manley**[2]

[1]Department of Microbiology, Biochemistry and Molecular Genetics, New Jersey Medical School, Rutgers University, Newark, New Jersey 07103, USA

[2]Department of Biological Sciences, Columbia University, New York, New York 10027, USA

## Abstract

Alternative polyadenylation (APA) is an RNA-processing mechanism that generates distinct 3′ termini on mRNAs and other RNA polymerase II transcripts. It is widespread across all eukaryotic species and is recognized as a major mechanism of gene regulation. APA exhibits tissue specificity and is important for cell proliferation and differentiation. In this Review, we discuss the roles of APA in diverse cellular processes, including mRNA metabolism, protein diversification and protein localization, and more generally in gene regulation. We also discuss the molecular mechanisms underlying APA, such as variation in the concentration of core processing factors and RNA-binding proteins, as well as transcription-based regulation.

The transcriptome of eukaryotic cells is produced by three RNA polymerases, each with its own mechanisms for the maturation of the 3′ ends of nascent transcripts (reviewed in REF. 1). Protein-coding transcripts, or mRNAs, are transcribed by RNA polymerase II (Pol II). With the exception of the canonical, replication-dependent transcripts encoding histones in metazoans[2], the maturation of mRNA 3′ ends involves endonucleolytic cleavage of the nascent RNA followed by synthesis of a poly(A) tail on the 3′ terminus of the cleaved product by a poly(A) polymerase (PAP). These two coupled reactions, collectively referred to as cleavage and polyadenylation or, simply, polyadenylation, are intimately linked to transcription termination[1]. Polyadenylation also occurs for some other Pol II products, especially long non-coding RNAs (lncRNAs; non-coding transcripts of ~200 nt or longer).

The sequences in the mRNA precursor and the proteins required for polyadenylation are now well understood. The polyadenylation site, also known as the poly(A) site (PAS), is defined by surrounding RNA sequence elements (BOX 1), which are generally conserved across metazoans with some minor variations (BOX 1 and Supplementary information S1 (box)). However, major distinctions can be found in yeast and plant PASs[3] (Supplementary information S1 (box)). Notably, the key protein factors responsible for polyadenylation are conserved throughout eukaryotes, although the machinery in mammals, which comprises more than 20 core proteins (BOX 1), has differences in protein composition and subcomplex organization compared with the machinery in yeast[4–7].

**Competing interests statement**: The authors declare no competing interests.

It was first reported more than three decades ago that a gene can give rise to transcripts with multiple PASs and that differential usage of these sites can lead to the formation of distinct mRNA isoforms, a phenomenon termed alternative polyadenylation (APA; early studies were reviewed in REFS 8,9). From early studies using expressed sequence tags[10,11] and more recent analyses using high-throughput sequencing, we know that APA is very common and occurs most frequently in the 3′ untranslated region (3′ UTR) of mRNAs, and that it is used frequently in essentially all eukaryotes, from yeast to humans. For example, at least 70% of m ammalian mRNA-encoding genes express APA isoforms[12,13]. Substantial, albeit slightly lower, APA frequencies have been reported in simpler species (Supplementary information S1 (box)). In this Review, we discuss our current understanding of APA from genomic as well as molecular and cellular perspectives, focusing mostly on the mechanisms and consequences of APA in metazoans. Readers are referred to other reviews for discussions of some early studies and of work in other species[6,14–19].

## APA in 3′ UTRs

Most APA sites are located in 3′ UTRs. In line with the nomenclature used for alternative splicing, here we refer to the 3′ UTR portion upstream of the first, or proximal, PAS as the constitutive UTR (cUTR) and the portion downstream as the alternative UTR (aUTR) (FIG. 1a). APA occurring in the 3′ UTR, referred to hereafter as 3′ UTR-APA, gives rise to mRNA isoforms with significantly different 3′ UTR lengths. For example, for mouse transcripts, the median 3′ UTR lengths of shortest and longest APA isoforms differ about sevenfold, at 249 nt and 1,773 nt, respectively[13]. As 3′ UTRs contain *cis* elements that are involved in various aspects of mRNA metabolism, 3′ UTR-APA can considerably affect post-transcriptional gene regulation in various ways, including through the modulation of mRNA stability, translation, nuclear export and cellular localization, and even through effects on the localization of the encoded protein (FIG. 1b–d). One remarkable feature of 3′ UTR-APA is that it can be regulated globally, simultaneously involving numerous transcripts in a cell. This was first shown for different human tissues that display a biased preference for certain APA isoform types (BOX 2) and was later demonstrated in studies of proliferation- and differentiation-based changes in APA profiles (BOX 3).

### mRNA stability and translation

Perhaps the best studied consequence of 3′ UTR-APA is its effect on microRNA (miRNA) functions. miRNAs are small RNAs (~22 nt) that modulate the stability and/or translation of their target complementary mRNAs[20]. miRNA target sites are generally located in 3′ UTRs. In mammals, more than half of the conserved miRNA target sites are located in aUTRs[21,22]. Differential targeting of 3′ UTR-APA isoforms was first demonstrated in activated T cells and cancer cells, both of which display global 3′ UTR shortening compared with non-activated T cells and non-transformed cells, respectively[21,23]. A recent study showed that APA isoform expression influences about 10% of targeting by miRNAs between any two cell types analysed and, importantly, that the accuracy of target prediction can be improved if the cellular APA profile is considered[24]. Targeting by miRNAs is often influenced by target site location in the mRNA and by the surrounding sequences[20]. For example, target sites located near either end of a 3′ UTR tend to be more efficient than sites in the middle.

Consistent with this, target sites for certain pro-proliferation miRNAs are enriched in the region immediately upstream of the proximal PASs of pro-differentiation or anti-proliferation mRNAs; the shortening of 3′ UTRs during cell proliferation improves the targeting context for these miRNAs and can thus enhance their targeting efficiency and their promotion of cell proliferation[25].

3′ UTRs are also hotbeds for mRNA destabilization elements, which often function through RNA-binding proteins (RBPs). Well-characterized motifs include AU-rich elements (AREs), GU-rich elements (GREs) and PUF protein-binding elements[26]. As with miRNA target sites, inclusion or exclusion of these elements by 3′ UTR-APA can affect mRNA stability. For example, a genetic polymorphism leading to differential expression of two APA isoforms of human IFN-regulatory factor 5 (IRF5) is linked to the risk of developing systemic lupus erythematosus[27] (FIG. 1b). Because of the presence of an ARE in the aUTR, the two isoforms have different decay rates[27]. In addition, RNA–RNA interactions, such as base pairing between 3′ UTR-encoded Alu elements (which are the most abundant transposable elements in the human genome) and lncRNAs can lead to mRNA decay through STAU1-mediated mRNA decay[28]. Moreover, a long 3′ UTR is itself considered to be a feature that causes mRNA degradation through nonsense-mediated mRNA decay[29]. It is therefore generally believed that, owing to their tendency to harbour destabilizing elements and their sheer size, isoforms with long 3′ UTRs are less stable than short isoforms. However, this view has been challenged by a genome-wide study of the role of APA in mRNA decay in mouse cells. Using the transcription inhibitor actinomycin D (ActD) to measure mRNA stability, long isoforms were found to be only slightly less stable than short isoforms[30]. Possible ActD-related artefacts notwithstanding, this suggests that the fate of 3′ UTR-APA isoforms is more complex than was previously thought. For example, additional sequences such as stabilizing elements in aUTRs can also substantially affect mRNA decay[30–34]. Although our understanding is therefore far from complete, it is nonetheless now clear that many genes produce multiple mRNA isoforms with different decay rates, highlighting the importance of 3′ UTR-APA in modulating mRNA stability.

A related question is whether 3′ UTR-APA affects mRNA translation. Indeed, the above-mentioned study analysing the effects of APA in mouse cells reported that long isoforms were associated with slightly more ribosomes than were short isoforms[30]. As with the destabilization effects of longer 3′ UTRs, this may be attributable to both translation-enhancing and translation-suppressing elements in aUTRs. However, another study using human cells reported a role for 3′ UTR length in suppressing translation and also detailed variable effects of different 3′ UTR sequences on translation[35]. Hence, further work is required to delineate how various *cis* elements and 3′ UTR size per se affect the stability and translation of APA isoforms in different cell types and under different conditions, such as cell stress and differentiation.

## mRNA nuclear export and localization

Isoforms with a long 3′ UTR tend to be more abundant in the nucleus than in the cytoplasm[36,37]. This was observed initially in a global analysis of all transcribed sequences in human cells[37], and a more recent study found that ~10% of all detected 3′ UTR-APA

isoforms differed significantly in abundance between nuclear and cytoplasmic fractions[36]. Although nuclear retention was reported for long isoforms containing certain *cis* elements in the aUTR, such as inverted Alu repeats[38], it is still uncertain how much of the differential localization of the long isoforms is due to differences in mRNA stability rather than differences in nuclear export. In addition, if regulation of nuclear export is involved, exactly how *cis* elements in aUTRs and 3′ UTR size per se have an impact on export, and what the functional significance of APA might be, remains unclear.

A better understood role of aUTRs in mRNA localization is the control of subcellular localization in the cytoplasm. Such regulated mRNA localization can in turn facilitate localized translation, which is an efficient way to enrich proteins at a specific cellular location[39]. The relevance of APA for mRNA localization has been demonstrated for several transcripts in neuronal cells, in which localized translation in dendrites and axons is common. For example, a short isoform of the mRNA encoding brain-derived neurotrophic factor (BDNF) is restricted to the cell body, whereas the long isoform localizes to the dendrites, where it is translated[40] (FIG. 1c). Similarly, long and short isoforms of mRNAs encoding inositol monophosphatase 1 (REF. 41) and RAN[42] are localized to the axon and cell body, respectively. These reports suggest that long isoforms are more likely to be located in dendrites or axons than are short isoforms. Conversely, a recent study compared mRNA localization in neurites (dendrites and axons) versus the cell body for neuronal cell lines and for primary cortical neurons, and this study found that short and long isoforms are similarly enriched in neurites and in the cell body[43]. Future investigations are required to delineate the underlying mechanisms involved and to address whether, as in mRNA stability, *cis* elements can function in both enhancing and suppressing subcellular localization of mRNAs.

## Protein localization

Sequences in 3′ UTRs have been implicated in mRNA localization to the ER to facilitate the expression of membrane proteins[44,45]. A surprising recent study showed that the 3′ UTR can also regulate protein localization independently of mRNA localization[46] (FIG. 1d). Specifically, the aUTR of the mRNA encoding the transmembrane protein CD47 was found to act as a scaffold for a protein complex containing the RBP Hu antigen R (HUR; also known as ELAVL1) and the phosphatase 2A inhibitor SET; this complex is therefore recruited to the site of translation, resulting in the interaction of SET with the newly translated cytoplasmic domains of CD47 and the subsequent translocation of CD47 to the plasma membrane. The short mRNA isoform, which lacks the sequences necessary for assembly of the HUR–SET complex, gives rise to CD47 that is primarily localized at the ER. Thus, CD47 has a different localization, and hence a function, depending on whether it is translated from the short or long mRNA isoform. This mechanism has also been observed for transcripts encoding several other proteins, including CD44, α1 integrin (ITGA1) and TNF receptor superfamily member 13C (TNFRSF13C)[46].

## APA upstream of the last exon

A sizable fraction of APA sites are located upstream of the last exon, mostly in introns. For simplicity, we refer to this as upstream regions APA (UR-APA). In the mouse genome, for example, more than 40% of genes have PASs of this type[13]. UR-APA leads to the expression of alternative terminal exons and can result in changes to both the coding sequence and 3′ UTR of an mRNA. Depending on the configuration of splicing relative to the PAS, the resulting alternative terminal exons can be divided into two subtypes (FIG. 2a): skipped terminal exons, which are alternative upstream exons selected through splicing to be the terminal exons, and composite terminal exons, which are formed by the extension of an internal exon into the adjacent intron through inhibition of the 5′ splice site. In addition, a small fraction of PASs can be identified in internal exons, leading to transcripts without an in-frame stop codon, which are likely to be degraded rapidly through the non-stop decay pathway[47]. However, in some rare cases, truncated proteins can be produced when adenosine residues from the poly(A) tail are used to form a stop codon[48]. UR-APA is generally upregulated in proliferating cells and suppressed during cell differentiation[13,43,49], mirroring the use of proximal PASs in 3′ UTRs, suggesting that UR-APA and 3′ UTR-APA are mechanistically related in these conditions. Similar to 3′ UTR-APA, UR-APA can also affect gene expression in various ways, and this is addressed below.

### Protein diversification

Two classic APA events reported in the early 1980s, involving transcripts from the calcitonin-related polypeptide-α gene (*CALCA*) and the gene encoding the immunoglobulin M (IgM) heavy chain, are well-known examples of UR-APA. In the case of *CALCA*, alternative splicing and the use of a proximal PAS generates a transcript containing a skipped ter minal exon, and this mRNA isoform encodes the protein calcitonin, whereas the use of a distal PAS in the 3′-most exon generates an mRNA encoding calcitonin gene-related peptide 1 (CGRP)[50]. The regulation of APA is tissue specific in this case: when comparing expression levels of the two isoforms, the calcitonin-encoding isoform is more highly expressed in the thyroid, whereas the CGRP-encoding isoform predominates in the hypothalamus. In the case of IgM heavy chain mRNA, during B cell activation there is a switch from using a distal PAS in the 3′-most exon to using a proximal PAS in a composite terminal exon, which results in a shift in protein production from a membrane-bound form of the antibody to a secreted form[51]. Notably, bioinformatic analysis has identified at least 376 mouse genes that potentially use such a mechanism for regulating membrane anchoring[52]. Manipulation of UR-APA-based protein isoform switching has also been shown to be a promising therapeutic approach. For example, the addition of an antisense RNA that attenuates splicing triggers the activation of an intronic PAS in the mRNA encoding vascular endothelial growth factor receptor 2 (VEGFR2) and thus enforces the expression of a soluble version of VEGFR2, which functions antagonistically to the membrane-bound form and inhibits angiogenesis[53].

In addition to the generation of proteins with distinct functions, UR-APA can lead to the expression of truncated proteins with dominant negative functions. For example, retinoblastoma-binding protein 6 (RBBP6) is a recently characterized polyadenylation factor

(BOX 1 and discussed below) that produces several isoforms through differential RNA processing. One of these isoforms arises from the use of an intronic PAS, which generates a severely truncated protein called Iso3 (REF. 54) (FIG. 2b). Iso3, which is downregulated in several human cancers[55], is able to compete with full-length RBBP6 for association with the remainder of the polyadenylation machinery, thereby inhibiting polyadenylation and regulating APA.

### Repression of gene expression

UR-APA can also generate transcripts without apparent functions by utilizing PASs in promoter-proximal introns. For example, the gene encoding the mammalian polyadenylation factor cleavage stimulation factor 77 kDa subunit (CSTF77; also known as CSTF3) (BOX 1) has a highly conserved intronic PAS, the use of which results in a transcript that would produce a severely truncated, probably non-functional, protein[56]. Production of this UR-APA transcript is induced by high cellular levels of full-length CSTF77 protein, thereby forming a negative feedback loop to control the activity of CSTF77, which is import for cell cycle control[57] (FIG. 2c). This mechanism was origin ally proposed for the *Drosophila melanogaster* homologue of CSTF77, Suppressor of Forked[58], and may exist in transcripts encoding other polyadenylation factors, such as PAP[59]. Notably, UR-APA of CSTF77, as well as that of a large fraction of genes containing promoter-proximal intronic PASs, is also regulated by U1 small nuclear ribonucleoprotein (U1 snRNP) (see below), suggesting that the generation of truncated transcripts by UR-APA is a widespread mechanism for the inhibition of gene expression.

## Regulation of APA

As the consequences of APA for gene expression and cell function are becoming increasingly clear, it is important to understand the mechanisms that regulate APA. A growing number of APA-regulatory factors have been identified and characterized in the past few years (FIG. 3a); some have global effects on APA, whereas others have an impact on APA of specific genes, as described below.

### Polyadenylation factors that promote proximal PAS usage

One important mechanism of APA regulation involves modulation of the expression levels of core polyadenylation machinery components. This was first demonstrated for CSTF64 (also known as CSTF2), the RNA-binding subunit of the trimeric CSTF complex (BOX 1); strong upregulation of CSTF64 during B cell maturation results in higher levels of the complete CSTF complex and increased usage of the weaker upstream intronic PAS in the IgM heavy chain transcript[60]. Consistent with this, siRNA-mediated knockdown of CSTF64 and its paralogue, CSTF64 τ-variant (τCSTF64; also known as CSTF2T), was found to cause global 3′ UTR lengthening in HeLa cells[61]. However, it is notable that knockdown of each factor alone was not sufficient to elicit such an effect[61,62], indicating that there is at least partial redundancy between the two highly similar proteins. The role of CSTF64 in APA was recently highlighted by a bioinformatics study which revealed that general 3′ UTR shortening occurs in seven tested cancer types, five of which also exhibited upregulation of *CSTF64* mRNA expression[63]. Furthermore, a significant overlap between

the APA events in this study and those of the knockdown experiment in HeLa cells was observed[61], suggesting that CSTF64 has a role in 3′ UTR shortening in cancer[63].

Changes in the expression of other core polyadenylation factors can also lead to global 3′ UTR-APA. Indeed, the expression level of polyadenylation factors as a whole has been shown to correlate inversely with the global relative expression levels of isoforms with long 3′ UTRs during cell differentiation and de-differentiation[64]. Consistent with this, genes encoding polyadenylation factors tend to have proliferation factor-binding elements in their promoters, such as binding sites for the E2F transcription factors[49,64], which promote cell cycle progression from G1 to S phase. In addition to these global effects, certain transcripts contain APA sites that are sensitive to the levels of specific poly adenylation factors, and such transcripts often display distinct features. For example, PASs regulated by CSTF64 or τCSTF64 show strong enrichment for U- and GU-rich elements, whereas those regulated by factor interacting with PAP (FIP1) — a cleavage and polyadenylation specificity factor (CPSF) complex subunit, knockdown of which leads to 3′ UTR lengthen ing[62,65] — are more likely to possess an upstream U-rich sequence[61,62,65], consistent in both cases with the respective RNA-binding activities of the proteins (BOX 1). Notably, FIP1 is particularly important for embryonic stem cell (ESC) self-renewal and somatic cell reprogramming, and a FIP1-dependent APA programme correlates with a change in FIP1 expression during ESC differentiation and somatic cell reprogramming[65]. By contrast, APA regulation by PCF11, a cleavage factor II complex subunit that lacks RNA-binding activity, does not appear to involve any specific sequences around the PAS[62]. Interestingly, although knockdown of RBBP6 leads to global 3′ UTR lengthening, RBBP6 also appears to be important specifically for the accumulation of its target mRNAs, including those with AU-rich elements in their 3′ UTRs, suggesting that there is a connection between mRNA stability and 3′ end processing[54].

## Polyadenylation factors that promote distal PAS usage

A model to explain how elevated levels of core polyadenylation factors can enhance the use of proximal PASs was initially proposed as a result of the early IgM studies[60] and posits that these sites, which are typically weaker than downstream sites[11], are used on a ‘first come, first served’ basis when the core polyadenylation machinery is not limiting. It was therefore unexpected that downregulation of subunits of the core polyadenylation complex cleavage factor I (CFI) would result in increased use of upstream PASs. CFI is a heterodimer consisting of CFI25 (also known as CPSF5) and either of two closely related subunits, CFI68 (also known as CPSF6) or CFI59 (also known as CPSF7). Specifically, knockdown of CFI25 or CFI68, but puzzlingly not CFI59, leads to significant shortening of 3′ UTRs[62,66,67]. PASs in the last intron were also found to be activated[62], suggesting a role for the CFI complex in defining the terminal exon. An analysis of *cis* elements indicated that UGUA elements, to which CFI25 binds[68], are highly enriched in the upstream parts of distal PASs affected by a CFI25 or CFI68 deficiency[62,66,67]. Because the CFI complex exists as a dimer, one possible mechanism for CFI-based APA is that CFI complexes bind two UGUA elements, one upstream and one downstream of a proximal PAS, thereby leading to skipping of this PAS[69] (FIG. 3b). Whether this in fact occurs and why CFI59 behaves differently require further study. Importantly, CFI25 expression is downregulated in glioblastoma cells,

leading to the usage of upstream PASs and to enhanced tumorigenicity and increased tumour size; conversely, CFI25 overexpression inhibits tumour growth[70]. In addition, copy number variations of *NUDT21* (the gene encoding CFI25) were found in individuals with certain neuropsychiatric syndromes[71]. In lymphoblastoid cells of these individuals, increased CFI25 levels led to higher expression of a long isoform of the mRNA encoding methyl CpG-binding protein 2 (MECP2), resulting in reduced production of MECP2, probably owing to the presence of numerous miRNA target sites in its aUTR. Because MECP2 levels need to be tightly regulated in the brain and small fluctuations in abundance can lead to neurological malfunctions, *NUDT21* was suggested to be a candidate gene for causing intellectual disability and neuropsychiatric diseases[71].

Nuclear poly(A)-binding protein 1 (PABPN1; also known as PABP2) controls poly(A) tail length[72]. Surprisingly, given that PABN1 was not thought to participate in PAS choice, knockdown of PABPN1 induced global 3′ UTR shortening[73,74]. Ectopic expression of trePABPN1(A17), which is a short trinucleotide-repeat expansion mutant found in patients with autosomal-dominant oculopharyngeal muscular dystrophy (OPMD), led to a similar trend of 3′ UTR shortening in cultured cells and mouse tissues, suggesting a connection between 3′ UTR regulation and the aetiology of OPMD[73]. Indeed, PABPN1 can inhibit 3′ cleavage of target transcripts *in vitro*[73], but whether this occurs *in vivo* is unknown. It is worth noting that, although a PABPN1 deficiency elicits 3′ UTR shortening, there is no apparent correlation, at least in mouse myoblasts, between the aUTR size and the degree of APA regulation, whereas such a correlation is a common feature for APA regulation by other polyadenylation factors[62]. Thus, how PABPN1 affects APA remains to be determined. Knockdown of PABPN1 also led to increased expression of RNA species using PASs near gene promoters, in both sense and antisense directions, with the latter being more prominent[62]. This function is likely to be related to the role of PABPN1 in hyperpolyadenylation, which leads to RNA degradation by the nuclear exosome[75,76]. Indeed, a similar phenotype was observed following knockdown of the exosome factors RRP44 and RRP6 (also known as EXOSC10)[62]. Interestingly, hyperpolyadenylation was found to involve the canonical enzymes PAPα and PAPγ, rather than the non-canonical PAPs that are usually associated with exosome activity[77]. It was suggested that hyperpolyadenylation is an important nuclear RNA decay pathway that is responsible for the removal of transcripts that have been poorly spliced or retained in the nucleus. An outstanding question is how PABPN1 has different roles at the 5′ and 3′ ends of genes, and whether the apparent effects on APA may instead reflect differential nuclear RNA decay.

Another poly(A)-binding protein that functions in APA is polyadenylate-binding protein 1 (PABP1), which shuttles between the nucleus and cytoplasm and is possibly the major poly(A) tail-binding protein in the cytoplasm. Knockdown of PABP1 was also found to modulate 3′ UTR length[62]. Exactly how poly(A) tail-binding proteins function in APA, and whether their involvement is direct or indirect, needs to be further studied.

It is clear that core polyadenylation factors have substantial roles in APA regulation. Notably, although some polyadenylation factors seem to fall into either proximal-PAS-promoting or distal-PAS-promoting activity groups, this bifurcation may be oversimplifying the situation for other polyadenylation factors. For example, similar numbers of proximal

and distal PASs were found to be regulated by CSTF77 knockdown[57], and CPSF30 of *Arabidopsis thaliana* regulates a large number of APA events without a clear preference for proximal or distal PAS usage during the response to oxidative stress[78].

## The splicing connection and the role of U1 snRNP in APA

Splicing and polyadenylation are frequently interconnected, and this connection was initially suggested as a mechanism to facilitate the definition of 3′-terminal exons[79]. In the case of intronic PASs, splicing and polyadenylation are likely to be in competition with each other, as large introns with weak 5′ splice sites undergo polyadenylation at their internal PASs to a greater extent than other introns[80], and inhibition of splicing — for example, by ablation of the U2 snRNP component splicing factor 3B subunit 1 (SF3B1) — gener ally activates intronic PASs[62]. Multiple protein–protein interactions exist between core splicing factors and core polyadenylation factors, such as between U1 snRNP and CPSF 160 kDa subunit (CPSF160; also known as CPSF1)[81], U2 snRNP and the CPSF complex[82], and U2 auxiliary factor 65 kDa subunit (U2AF65; also known as U2AF2) and the CFI complex[83]. However, U1 snRNP, which recognizes 5′ splice sites, seems to have an active and, in many cases, probably distinct role in APA regulation. In early studies, U1 snRNP was shown to suppress PAS usage through inhibition of PAPα[84]. More recently, inhibition of U1 snRNP was found to result in the activation of cryptic PASs near transcription start sites[85], implying that U1 snRNP normally represses the use of such PASs, and mild attenuation of this inhibitory function caused increased usage of proximal PASs in 3′ UTRs[86]. This process, dubbed telescripting, involves the inhibition of polyadenylation by U1 snRNP binding to canonical 5′ splice sites, or similar sites, throughout the nascent RNA[87]. The existence of telescripting provides an answer to the long-standing question of why U1 snRNP is present at a much higher abundance than other snRNPs[88]. Telescripting has also been implicated in the global transcript shortening that occurs during transient transcription upregulation upon activation of neurons[86]. A similar phenomenon was recently observed in the human colon carcinoma cell line RKO after exposure to UV damage[89], such that 5′ intronic PASs were substantially activated, probably owing to reduced levels of U1 snRNA. Notably, 3′ UTR shortening was not obvious in these RKO cells. Whether the difference is due to cell specificity or some other factors remains to be seen.

## Regulation of APA by other RBPs

A growing number of RBPs have been found to interact with regions near PASs and to regulate PAS usage. RBPs already known to regulate splicing are very often also found to be regulators of APA, and they typically regulate PAS usage in a context-dependent manner, as was first shown for the neuronal RBP NOVA2 (REF. 90): binding of NOVA near the PAS is inhibitory, whereas binding distantly from the PAS enhances PAS usage. Below, we highlight some recent findings that exemplify both general rules and novel mechanisms. Readers are referred to other reviews[91] for more exhaustive information.

The ELAV (embryonic-lethal abnormal visual) proteins constitute an extensively studied family of RBPs that function in several aspects of mRNA metabolism, including APA. For example, *D. melanogaster* Elav was shown to mediate neuron-specific 3′ UTR lengthening by suppressing the use of proximal PASs[92] (FIG. 3c). Interestingly, this involves the

recruitment of Elav to paused Pol II near the promoter of the Elav-responsive gene, indicative of a link between APA and transcription (see below)[93]. In mammals, the Hu proteins, which are Elav homologues, inhibit the use of PASs with U-rich elements[94]. The mRNA encoding HUR, a ubiquitously expressed Hu protein, is also subjected to APA, either by HUR itself[95] or by neuron-specific Hu proteins, HUB (also known as ELAVL2), HUC (also known as ELAVL3) and HUD (also known as ELAVL4)[96]. This process balances the pro-differentiation activity of the neuron-specific Hu proteins with the pro-proliferation activity of HUR.

SR proteins are a family of conserved RBPs that contain RNA-binding domains and sequences rich in Ser-Arg dipeptide repeats (RS domains). First discovered as splicing factors[97], these proteins are now known to have various roles in mRNA biogenesis and metabolism[98]. Two of the twelve SR proteins, SRSF3 and SRSF7, were also found to regulate 3′ UTR length in mouse P19 cells, in which SRSF3 lengthens 3′ UTRs and SRSF7 has the opposite effect[99]. SRSF3 binding to the last exon also promotes mRNA nuclear export through an interaction with nuclear RNA export factor 1 (NXF1). Although it is unclear how SRSF3 and SRSF7 alter PAS choice, the observation that SR proteins can regulate APA and mRNA export suggests that there is a connection between these two processes. Indeed, multiple interactions have been reported between nuclear export factors and polyadenylation factors[100,101], and all of these interactions have been shown to affect PAS choice.

The usually fatal neurodegenerative disease amyotrophic lateral sclerosis (ALS) has long been linked to defects in RNA processing, in part because it can be caused by mutations in genes encoding RBPs, including TAR DNA-binding protein 43 (TDP43) and FUS[102]. Although these proteins have documented roles in splicing, recent studies have pointed to functions in APA as well. FUS binds nascent RNAs and interacts with the CPSF and CSTF complexes, and FUS knockdown leads to changes in APA[103]. FUS also binds the carboxy-terminal domain of the largest subunit of Pol II and prevents inappropriate hyperphosphorylation of this domain[104]. Loss of FUS leads to Pol II accumulation at transcription start sites and activation of promoter-proximal PASs (see below). Future studies are needed to elucidate how the function of FUS in Pol II phosphorylation is related to its RNA binding and its interaction with polyadenylation factors. Interestingly, whereas wild-type FUS binds mostly intronic sequences, an ALS-causing FUS mutant binds predominantly to sites in 3′ UTRs[105]. However, how this is related to FUS-mediated pathology in ALS remains to be determined.

The most frequent known cause of ALS (as well as of a dementia called frontotemporal dementia (FTD)) is an expansion of the hexanucleotide GGGGCC in the gene *C9ORF72*. It is notable that a global shortening of 3′ UTRs was observed in the cerebellum of patients with ALS who also had the gene expansion[106]. A number of mechanisms have been suggested for how the expansion can lead to ALS and FTD, including the sequestering of multiple RBPs, such as heterogeneous nuclear RNP H (HNRNPH), into nuclear foci[107], but additional studies are needed to investigate how APA is dysregulated in ALS and to discern the possible significance of 3′ UTR shortening to the aetiology of the disease. On a related note, sequestering of muscleblind-like RBPs by expanded CUG repeats is known to lead to

the disease myotonic dystrophy and was found to change the adult APA profile to match that seen in neonates in the skeletal muscles of human patients and mouse models[108], highlighting the importance of RBP-mediated APA regulation in development and disease.

## Regulation of APA by transcription

Polyadenylation is frequently, if not always, co-transcriptional. It is not surprising, therefore, that various aspects of the transcription process influence PAS choice. Similar to the effects of transcription on alternative splicing (reviewed in REF. 109), transcriptional influences on PAS usage can fall into two categories. First, alterations in the rate of Pol II elongation at the PAS region can control APA. Consistent with this idea, *cis* elements that cause pausing of Pol II, such as G-rich sequences, facilitate PAS usage[110]. Mutations in the genes encoding the transcription elongation factors TFIIS (also known as Dst1) and Spt5, as well as in the gene encoding Pol II subunit Rpb2, can enhance the usage of upstream PASs in yeast, probably by increasing pausing[111]; furthermore, in mouse plasma cells, the elongation factor ELL2 can modulate usage of the proximal PAS of immunoglobulin heavy chain pre-mRNA[112]. More direct evidence comes from the study of a mutant *D. melanogaster* strain expressing Pol II with a slower elongation rate[113], in which expression of the short isoform of the *polo* gene was increased. This is physiologically relevant because transgenic flies lacking the long *polo* isoform die at the pupa stage owing to perturbed proliferation of abdominal precursor cells.

A second link between transcription and polyadenylation involves transcription machinery-mediated recruitment of specific factors that can influence PAS choice. Certain transcription activators can promote efficient 3′ end processing both *in vivo*[114] and in a transcription-coupled 3′-end-processing assay *in vitro*[115], and this leads to more frequent use of proximal PASs. Indeed, global analyses indicate that when genes are expressed at high levels, mRNA isoforms with shorter 3′ UTRs tend to be more abundant[116,117] and that Pol II tends to pause at proximal PASs to a greater extent[116]. The recruitment of polyadenylation factors by transcription-activating factors may be responsible for the presence of several polyadenylation proteins at promoter regions[118–121], and can be carried out directly[122] or through the Pol II-associated factor (PAF) elongation complex[115]. The CDC73 (also known a parafibromin) subunit of PAF, encoded by a tumour suppressor gene that is mutated in hereditary and sporadic parathyroid tumours, directly interacts with the CPSF and CSTF complexes[120], and knockdown of CDC73 or another PAF subunit, PAF1, in mouse myoblasts leads to significant shortening of transcripts through activation of intronic PASs and proximal PASs in 3′ UTRs[123]. These findings, as well as the observation that upstream antisense transcripts which use PASs near promoters were also upregulated following PAF knockdown, are consistent with there being a role for PAF in releasing paused Pol II[124]. Also notable is the finding that knockdown of the PAF subunit SKI8, which associates with the exosome, leads only to 3′ UTR shortening, with little effect on PASs in upstream regions or around the promoter. How different components of PAF exert different effects at different regions of transcripts is an interesting question that remains to be answered.

Chromatin structure is intimately connected with transcription. Studies from yeast to humans have shown that the region around the PAS is generally depleted of nucleosomes[125],

presumably at least partially owing to the presence of AT-rich sequences, which are less favourable for nucleosome association, around the PAS[126]. However, the higher affinity of nucleosomes for sequences downstream of highly used alternative PASs[127] and the correlation between nucleosome levels and the accumulation of Pol II downstream of the PAS[128] suggest a possible active role of chromatin organization in APA. Consistent with this, variations in nucleosome density and levels of histone H3 Lys36 trimethyl ation — a histone modification that is enriched at gene 3′ ends — between genes expressed at different levels were found to be much greater at proximal PASs than at distal PASs[116]. Furthermore, a recent study found that a heterochromatin formation that causes Pol II pausing also promotes the usage of proximal PASs[36]. Conversely, the more open chromatin conformation during spermato genesis, as indicated by the histone H3 Lys4 trimethylation, was found to correlate with an increased usage of proximal PASs in 3′ UTRs[129], suggesting that open chromatin allows more efficient recruitment of the polyadenylation complex and facilitates PAS usage. Future studies will be needed to delineate and reconcile the data about how chromatin organization affects APA.

## Concluding remarks

Numerous advances in the past few years have substantially enriched our knowledge of APA, which is now acknowledged as an important and widespread mechanism for modulating gene expression. We expect that our understanding of APA will continue to grow rapidly in the next few years, as some of the outstanding questions outlined below are answered.

### Regulation of APA

Novel findings about the regulation of APA continue to emerge. For example, $N^6$-methyladenosine ($m^6A$) was found to be highly enriched in 3′-most exons, and a reduction of $m^6A$ levels has been shown to affect APA, mostly causing 3′ UTR shortening[130]. But what is the mechanistic basis for this form of APA regulation, and what is its physiological role? More generally, we now know that variations in the expression levels of different polyadenylation factors can alter PAS choice in different ways, but what is the mechanism underlying this? For example, do changes in levels of a complex subunit affect the levels of the intact functional complex, as shown in the initial studies of CSTF64-mediated APA regulation in B cell differentiation[60]? If so, do different cell types or conditions have different rate-limiting polyadenylation factors? Or might these variable effects reflect the existence of heterogeneity in the make-up of the polyadenylation machinery? It is also important to understand the interplay between the polyadenylation complex and the cellular RBPs in control of APA, and the contributions of promoter sequences and U1 snRNP are to APA regulation in different cells. Finally, we still know very little about how the expression of polyadenylation factors is itself regulated through transcriptional and post-transcriptional mechanisms. When all these layers of regulation are better understood, we will be in a position to ‘crack the APA code’.

## Consequences of APA

The recent findings that 3′ UTR-APA can regulate protein localization independently of mRNA localization[46] and that UR-APA can change mRNA localization through the inclusion of different terminal exons[43] are particularly intriguing. How widespread are the roles of 3′ UTR-APA and UR-APA in protein localization and mRNA localization, respectively? In conditions under which 3′ UTR length is globally regulated, such as in cancer cells, is protein localization also globally remodelled as a consequence? In addition, the impact of APA on mRNA decay and translation needs to be further analysed in more cellular conditions, such as during cell stress and differentiation. How different cell types that have different average 3′ UTR lengths, such as neuronal cells, regulate mRNA decay and translation remains to be addressed.

The connection between poly(A) tail length — an important feature for mRNA stability and translation[131] — and PAS choice is still poorly understood. Poly(A) tail synthesis is intimately related to the nuclear cleavage reaction but can be dynamically remodelled in the cytoplasm. For example, cytoplasmic polyadenylation elements (CPEs) near the PAS can modulate poly(A) tail length through CPE-binding proteins and the activity of non-canonical PAPs, as demonstrated in neurons and during early development[132,133]. Although several methods for measuring poly(A) tail length have been established[134,135], they have yet to be exploited to investigate the long-standing issue of how or whether poly(A) tail length is regulated by changing PAS choice. With the recent rapid advances in sequencing technologies, we expect this issue will be elucidated in a genome-wide manner.

Another outstanding question is how APA affects the functions of lncRNAs, which constitute a still-expanding class of transcripts with many structural and regulatory functions (reviewed in REF. 136). These lncRNAs could be extensively regulated by APA[13]. For example, APA isoforms of the lncRNA nuclear-enriched abundant transcript 1 were shown to have different functions in paraspeckle formation[137].

## Clinical implications

The first example showing that human disease can be caused by a malfunction of 3′ end processing comes from the identification of a mutation in the AAUAAA sequence of the PAS of the gene *HBA2*, which encodes haemoglobin subunit α2, in patients with α-thalassaemia[138]. The relevance of APA, as opposed to 3′ end processing per se, to human health was first demonstrated by the causative correlation between systemic lupus erythematosus and a single nucleotide polymorphism in the *IRF5* gene that affected APA in the transcript[27]. Similarly, a polymorphic PAS downstream element in the *ATP1B1* gene, encoding the $(Na^+ + K^+)$ATPase β1 subunit, was found to be associated with high blood pressure[139]. As the feasibility of performing population-wide, whole-genome sequencing improves, we will gain a systemic view of how APA is affected by disease-causing mutations and genetic variations, and more importantly, what the role of APA is in disease aetiology and in shaping human traits. Given the cell type specificity of APA profiles, it is conceivable that information about APA could assist in disease diagnosis, as has been demonstrated in certain cancers[63,140] and cardiac diseases[141–143]. Whether and in what way

such approaches could be used in clinical settings and whether APA, or 3′ end processing in general, could serve as a therapeutic target remain to be explored.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors thank members of their laboratories for helpful discussions, and I. Boluck for assistance with manuscript preparation. Work in the authors' laboratories was funded by grants GM84089 (B.T.), and GM28983 and GM118136 (J.L.M.) from the US National Institutes of Health.

## References

1. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. Genes Dev. 2009; 23:1247–1269. [PubMed: 19487567]
2. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet. 2008; 9:843–854. [PubMed: 18927579]
3. Tian B, Graber JH. Signals for pre-mRNA cleavage and polyadenylation. Wiley Interdiscip Rev RNA. 2012; 3:385–396. [PubMed: 22012871]
4. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3′-end processing. Cell Mol Life Sci. 2008; 65:1099–1122. [PubMed: 18158581]
5. Zhao J, Hyman L, Moore C. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev. 1999; 63:405–445. [PubMed: 10357856]
6. Shi Y, Manley JL. The end of the message: multiple protein–RNA interactions define the mRNA polyadenylation site. Genes Dev. 2015; 29:889–897. [PubMed: 25934501]
7. Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. Genes Dev. 1997; 11:2755–2766. [PubMed: 9353246]
8. Edwalds-Gilbert G, Veraldi KL, Milcarek C. Alternative poly(A) site selection in complex transcription units: means to an end? Nucleic Acids Res. 1997; 25:2547–2561. [PubMed: 9185563]
9. Barabino SM, Keller W. Last but not least: regulated poly(A) tail formation. Cell. 1999; 99:9–11. [PubMed: 10520989]
10. Gautheret D, Poirot O, Lopez F, Audic S, Claverie JM. Alternate polyadenylation in human mRNAs: a large-scale analysis by EST clustering. Genome Res. 1998; 8:524–530. This paper reports the first use of expressed sequence tags to identify APA sites genome-wide. [PubMed: 9582195]
11. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. 2005; 33:201–212. [PubMed: 15647503]
12. Derti A, et al. A quantitative atlas of polyadenylation in five mammals. Genome Res. 2012; 22:1173–1183. [PubMed: 22454233]
13. Hoque M, et al. Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. Nat Methods. 2013; 10:133–139. [PubMed: 23241633]
14. Mayr C. Evolution and biological roles of alternative 3′ UTRs. Trends Cell Biol. 2016; 26:227–237. [PubMed: 26597575]
15. Proudfoot NJ. Ending the message: poly(A) signals then and now. Genes Dev. 2011; 25:1770–1782. [PubMed: 21896654]
16. Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet. 2013; 14:496–506. [PubMed: 23774734]
17. Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. Mol Cell. 2011; 43:853–866. [PubMed: 21925375]

18. Tian B, Manley JL. Alternative cleavage and polyadenylation: the long and short of it. Trends Biochem Sci. 2013; 38:312–320. [PubMed: 23632313]
19. Hunt AG. Messenger RNA 3′ end formation in plants. Curr Top Microbiol Immunol. 2008; 326:151–177. [PubMed: 18630752]
20. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009; 136:215–233. [PubMed: 19167326]
21. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. Science. 2008; 320:1643–1647. This is the first report that global changes in APA occur as a consequence of changes in cell proliferation, specifically demonstrating the use of proximal 3′ UTR PASs during the activation of T cells. [PubMed: 18566288]
22. Ji Z, Lee JY, Pan Z, Jiang B, Tian B. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc Natl Acad Sci USA. 2009; 106:7028–7033. This article describes global APA regulation in embryonic development, connecting polyadenylation activity with APA during cell differentiation. [PubMed: 19372383]
23. Mayr C, Bartel DP. Widespread shortening of 3′ UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell. 2009; 138:673–684. This publication reports a connection between 3′ UTR shortening, especially in the transcripts of several proto-oncogenes, and cell transformation. [PubMed: 19703394]
24. Nam JW, et al. Global analyses of the effect of different cellular contexts on microRNA targeting. Mol Cell. 2014; 53:1031–1043. [PubMed: 24631284]
25. Hoffman Y, et al. 3′ UTR shortening potentiates microRNA-based repression of pro-differentiation genes in proliferating human cells. PLoS Genet. 2016; 12:e1005879. [PubMed: 26908102]
26. Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. Nat Rev Mol Cell Biol. 2007; 8:113–126. [PubMed: 17245413]
27. Graham RR, et al. Three functional variants of IFN regulatory factor 5 (*IRF5*) define risk and protective haplotypes for human lupus. Proc Natl Acad Sci USA. 2007; 104:6758–6763. [PubMed: 17412832]
28. Gong C, Maquat L. E lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. Nature. 2011; 470:284–288. [PubMed: 21307942]
29. Hogg JR, Goff SP. Upf1 senses 3′ UTR length to potentiate mRNA decay. Cell. 2010; 143:379–389. [PubMed: 21029861]
30. Spies N, Burge CB, Bartel D. 3′ UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. Genome Res. 2013; 23:2078–2090. This report presents a global analysis of the different effects of short and long 3′ UTRs on mRNA decay and translation. [PubMed: 24072873]
31. Ulitsky I, et al. Extensive alternative polyadenylationduring zebrafish development. Genome Res. 2012; 22:2054–2066. [PubMed: 22722342]
32. Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell. 2014; 156:812–824. [PubMed: 24529382]
33. Tycowski KT, Shu MD, Steitz JA. Myriad triple-helix-forming structures in the transposable element RNAs of plants and fungi. Cell Rep. 2016; 15:1266–1276. [PubMed: 27134163]
34. Lee JE, Lee JY, Wilusz J, Tian B, Wilusz CJ. Systematic analysis of *cis*-elements in unstable mRNAs demonstrates that CUGBP1 is a key regulator of mRNA decay in muscle cells. PLoS ONE. 2010; 5:e11201. [PubMed: 20574513]
35. Floor SN, Doudna JA. Tunable protein synthesis by transcript isoforms in human cells. eLife. 2016; 5:e10921. [PubMed: 26735365]
36. Neve J, et al. Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. Genome Res. 2016; 26:24–35. [PubMed: 26546131]
37. Djebali S, et al. Landscape of transcription in human cells. Nature. 2012; 489:101–108. [PubMed: 22955620]

38. Chen LL, Carmichael GG. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol Cell. 2009; 35:467–478. [PubMed: 19716791]
39. Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension Cell. 2009; 136:719–730. [PubMed: 19239891]
40. An JJ, et al. Distinct role of long 3′ UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. Cell. 2008; 134:175–187. [PubMed: 18614020]
41. Andreassi C, Riccio A. To localize or not to localize: mRNA fate is in 3′ UTR ends. Trends Cell Biol. 2009; 19:465–474. [PubMed: 19716303]
42. Yudin D, et al. Localized regulation of axonal RanGTPase controls retrograde injury signaling in peripheral nerve. Neuron. 2008; 59:241–252. [PubMed: 18667152]
43. Taliaferro JM, et al. Distal alternative last exons localize mRNAs to neural projections. Mol Cell. 2016; 61:821–833. [PubMed: 26907613]
44. Loya A, et al. The 3′-UTR mediates the cellular localization of an mRNA encoding a short plasma membrane protein. RNA. 2008; 14:1352–1365. [PubMed: 18492794]
45. Reid DW, Nicchitta CV. Diversity and selectivity in mRNA translation on the endoplasmic reticulum. Nat Rev Mol Cell Biol. 2015; 16:221–231. [PubMed: 25735911]
46. Berkovits BD, Mayr C. Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature. 2015; 522:363–367. This work discovers a novel mechanism by which the aUTR of a transcript functions as a scaffold for the assembly of specific protein complexes, which then modulate the subcellular localization of the encoded protein. [PubMed: 25896326]
47. Vasudevan S, Peltz SW, Wilusz CJ. Non-stop decay — a new mRNA surveillance pathway. Bioessays. 2002; 24:785–788. [PubMed: 12210514]
48. Yao P, et al. Coding region polyadenylation generates a truncated tRNA synthetase that counters translation repression. Cell. 2012; 149:88–100. [PubMed: 22386318]
49. Elkon R, et al. E2F mediates enhanced alternative polyadenylation in proliferation. Genome Biol. 2012; 13:R59. [PubMed: 22747694]
50. Amara SG, Jonas V, Rosenfeld MG, Ong ES, Evans RM. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. Nature. 1982; 298:240–244. [PubMed: 6283379]
51. Alt FW, et al. Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3′ ends. Cell. 1980; 20:293–301. [PubMed: 6771018]
52. Davis MJ, et al. Differential use of signal peptides and membrane domains is a common occurrence in the protein output of transcriptional units. PLoS Genet. 2006; 2:e46. [PubMed: 16683029]
53. Vorlova S, et al. Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. Mol Cell. 2011; 43:927–939. [PubMed: 21925381]
54. Di Giammartino DC, et al. RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3′ UTRs. Gene Dev. 2014; 28:2248–2260. [PubMed: 25319826]
55. Mbita Z, et al. De-regulation of the RBBP6 isoform 3/DWNN in human cancers. Mol Cell Biochem. 2012; 362:249–262. [PubMed: 22139301]
56. Pan Z, et al. An intronic polyadenylation site in human and mouse CstF-77 genes suggests an evolutionarily conserved regulatory mechanism. Gene. 2006; 366:325–334. [PubMed: 16316725]
57. Luo W, et al. The conserved intronic cleavage and polyadenylation site of CstF-77 gene imparts control of 3′ end processing activity through feedback autoregulation and by U1 snRNP. PLoS Genet. 2013; 9:e1003613. [PubMed: 23874216]
58. Audibert A, Simonelig M. Autoregulation at the level of mRNA 3′ end formation of the *suppressor of forked* gene of *Drosophila melanogaster* is conserved in *Drosophila virilis*. Proc Natl Acad Sci USA. 1998; 95:14302–14307. [PubMed: 9826695]
59. Zhao W, Manley JL. Complex alternative RNA processing generates an unexpected diversity of poly(A) polymerase isoforms. Mol Cell Biol. 1996; 16:2378–2386. [PubMed: 8628305]

60. Takagaki Y, Seipelt RL, Peterson ML, Manley JL. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell. 1996; 87:941–952. This study uncovers a mechanism of APA regulation in which increased expression of a core polyadenylation factor, CSTF64, during B cell differentiation shifts PAS usage to an upstream site in the IgM heavy chain pre-mRNA. [PubMed: 8945520]
61. Yao C, et al. Overlapping and distinct functions of CstF64 and CstF64τ in mammalian mRNA 3′ processing. RNA. 2013; 19:1781–1790. [PubMed: 24149845]
62. Li W, et al. Systematic profiling of poly(A)$^+$ transcripts modulated by core 3′ end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. PLoS Genet. 2015; 11:e1005166. [PubMed: 25906188]
63. Xia Z, et al. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. Nat Commun. 2014; 5:5274. [PubMed: 25409906]
64. Ji Z, Tian B. Reprogramming of 3′ untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. PLoS ONE. 2009; 4:e8419. [PubMed: 20037631]
65. Lackford B, et al. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. EMBO J. 2014; 33:878–889. [PubMed: 24596251]
66. 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; 1:753–763. [PubMed: 22813749]
67. Gruber AR, Martin G, Keller W, Zavolan M. Cleavage factor Im is a key regulator of 3′ UTR length. RNA Biol. 2012; 9:1405–1412. [PubMed: 23187700]
68. Brown KM, Gilmartin GM. A mechanism for the regulation of pre-mRNA 3′ processing by human cleavage factor $I_m$. Mol Cell. 2003; 12:1467–1476. [PubMed: 14690600]
69. Yang Q, Gilmartin GM, Doublié S. The structure of human Cleavage Factor $I_m$ hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5′ capping and splicing. RNA Biol. 2011; 8:748–753. [PubMed: 21881408]
70. Masamha CP, et al. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature. 2014; 510:412–416. [PubMed: 24814343]
71. Gennarino VA, et al. *NUDT21*-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. eLife. 2015; 4:e10782.
72. Kuhn U, et al. Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. J Biol Chem. 2009; 284:22803–22814. [PubMed: 19509282]
73. Jenal M, et al. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell. 2012; 149:538–553. [PubMed: 22502866]
74. de Klerk E, et al. Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. Nucleic Acids Res. 2012; 40:9089–9101. [PubMed: 22772983]
75. Bresson SM, Conrad NK. The human nuclear poly(A)-binding protein promotes RNA hyperadenylation and decay. PLoS Genet. 2013; 9:e1003893. [PubMed: 24146636]
76. Beaulieu YB, Kleinman CL, Landry-Voyer AM, Majewski J, Bachand F. Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. PLoS Genet. 2012; 8:e1003078. [PubMed: 23166521]
77. Bresson SM, Hunter OV, Hunter AC, Conrad NK. Canonical poly(A) polymerase activity promotes the decay of a wide variety of mammalian nuclear RNAs. PLoS Genet. 2015; 11:e1005610. [PubMed: 26484760]
78. Thomas PE, et al. Genome-wide control of polyadenylation site choice by CPSF30 in *Arabidopsis*. Plant Cell. 2012; 24:4376–4388. [PubMed: 23136375]
79. Niwa M, Rose SD, Berget SM. *In vitro* polyadenylation is stimulated by the presence of an upstream intron. Genes Dev. 1990; 4:1552–1559. [PubMed: 1701407]
80. Tian B, Pan Z, Lee JY. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res. 2007; 17:156–165. [PubMed: 17210931]

81. Lutz CS, et al. Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency *in vitro*. Genes Dev. 1996; 10:325–337. [PubMed: 8595883]
82. 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. [PubMed: 16857586]
83. Millevoi S, et al. An interaction between U2AF 65 and CF $I_m$ links the splicing and 3′ end processing machineries. EMBO J. 2006; 25:4854–4864. [PubMed: 17024186]
84. Gunderson SI, Polycarpou-Schwarz M, Mattaj IW. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. Mol Cell. 1998; 1:255–264. [PubMed: 9659922]
85. Kaida D, et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature. 2010; 468:664–668. This article describes a global activity of U1 snRNP in suppressing promoter-proximal PASs. [PubMed: 20881964]
86. Berg MG, et al. U1 snRNP determines mRNA length and regulates isoform expression. Cell. 2012; 150:53–64. [PubMed: 22770214]
87. Engreitz JM, et al. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. Cell. 2014; 159:188–199. [PubMed: 25259926]
88. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009; 136:701–718. [PubMed: 19239890]
89. Devany E, et al. Intronic cleavage and polyadenylation regulates gene expression during DNA damage response through U1 snRNA. Cell Discov. 2016; 2:16013. [PubMed: 27462460]
90. Licatalosi DD, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature. 2008; 456:464–469. This is the first demonstration that a splicing-regulatory RBP, NOVA, can also regulate APA. [PubMed: 18978773]
91. Zheng D, Tian B. RNA-binding proteins in regulation of alternative cleavage and polyadenylation. Adv Exp Med Biol. 2014; 825:97–127. [PubMed: 25201104]
92. Hilgers V, Lemke SB, Levine M. ELAV mediates 3′ UTR extension in the *Drosophila* nervous system. Genes Dev. 2012; 26:2259–2264. [PubMed: 23019123]
93. Oktaba K, et al. ELAV links paused Pol II to alternative polyadenylation in the *Drosophila* nervous system. Mol Cell. 2015; 57:341–348. [PubMed: 25544561]
94. Zhu H, Zhou HL, Hasman RA, Lou H. Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. J Biol Chem. 2007; 282:2203–2210. [PubMed: 17127772]
95. Dai W, Zhang G, Makeyev EV. RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage. Nucleic Acids Res. 2012; 40:787–800. [PubMed: 21948791]
96. Mansfield KD, Keene JD. Neuron-specific ELAV/Hu proteins suppress HuR mRNA during neuronal differentiation by alternative polyadenylation. Nucleic Acids Res. 2012; 40:2734–2746. [PubMed: 22139917]
97. Manley JL, Tacke R. SR proteins and splicing control. Genes Dev. 1996; 10:1569–1579. [PubMed: 8682289]
98. Howard JM, Sanford JR. The RNAissance family: SR proteins as multifaceted regulators of gene expression. Wiley Interdiscip Rev RNA. 2015; 6:93–110. [PubMed: 25155147]
99. Muller-McNicoll M, et al. SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. Genes Dev. 2016; 30:553–566. [PubMed: 26944680]
100. Tran DD, et al. THOC5 controls 3′ end-processing of immediate early genes via interaction with polyadenylation specific factor 100 (CPSF100). Nucleic Acids Res. 2014; 42:12249–12260. [PubMed: 25274738]
101. Johnson SA, Kim H, Erickson B, Bentley DL. The export factor Yra1 modulates mRNA 3′ end processing. Nat Struct Mol Biol. 2011; 18:1164–1171. [PubMed: 21947206]
102. Ling SC, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. Neuron. 2013; 79:416–438. [PubMed: 23931993]

103. Masuda A, et al. Position-specific binding of FUS to nascent RNA regulates mRNA length. Genes Dev. 2015; 29:1045–1057. [PubMed: 25995189]
104. Schwartz JC, et al. FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. Genes Dev. 2012; 26:2690–2695. [PubMed: 23249733]
105. Hoell JI, et al. RNA targets of wild-type and mutant FET family proteins. Nat Struct Mol Biol. 2011; 18:1428–1431. [PubMed: 22081015]
106. Prudencio M, et al. Distinct brain transcriptome profiles in *C9orf72*-associated and sporadic ALS. Nat Neurosci. 2015; 18:1175–1182. [PubMed: 26192745]
107. Lee YB, et al. Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep. 2013; 5:1178–1186. [PubMed: 24290757]
108. Batra R, et al. Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. Mol Cell. 2014; 56:311–322. [PubMed: 25263597]
109. Naftelberg S, Schor IE, Ast G, Kornblihtt AR. Regulation of alternative splicing through coupling with transcription and chromatin structure. Annu Rev Biochem. 2015; 84:165–198. [PubMed: 26034889]
110. Yonaha M, Proudfoot NJ. Specific transcriptional pausing activates polyadenylation in a coupled *in vitro* system. Mol Cell. 1999; 3:593–600. [PubMed: 10360175]
111. Cui Y, Denis CL. *In vivo* evidence that defects in the transcriptional elongation factors RPB2, TFIIS, and SPT5 enhance upstream poly(A) site utilization. Mol Cell Biol. 2003; 23:7887–7901. [PubMed: 14560031]
112. Martincic K, Alkan SA, Cheatle A, Borghesi L, Milcarek C. Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. Nat Immunol. 2009; 10:1102–1109. [PubMed: 19749764]
113. Pinto PA, et al. RNA polymerase II kinetics in *polo* polyadenylation signal selection. EMBO J. 2011; 30:2431–2444. [PubMed: 21602789]
114. Rosonina E, Bakowski MA, McCracken S, Blencowe BJ. Transcriptional activators control splicing and 3′ -end cleavage levels. J Biol Chem. 2003; 278:43034–43040. [PubMed: 12939267]
115. Nagaike T, et al. Transcriptional activators enhance polyadenylation of mRNA precursors. Mol Cell. 2011; 41:409–418. [PubMed: 21329879]
116. Ji Z, et al. Transcriptional activity regulates alternative cleavage and polyadenylation. Mol Syst Biol. 2011; 7:534. [PubMed: 21952137]
117. Ni T, et al. Distinct polyadenylation landscapes of diverse human tissues revealed by a modified PA-seq strategy. BMC Genomics. 2013; 14:615. [PubMed: 24025092]
118. Glover-Cutter K, Kim S, Espinosa J, Bentley DL. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat Struct Mol Biol. 2008; 15:71–78. [PubMed: 18157150]
119. Venkataraman K, Brown KM, Gilmartin GM. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. 2005; 19:1315–1327. [PubMed: 15937220]
120. Rozenblatt-Rosen O, et al. The tumor suppressorCdc73 functionally associates with CPSF and CstF 3′ mRNA processing factors. Proc Natl Acad Sci USA. 2009; 106:755–760. [PubMed: 19136632]
121. Calvo O, Manley JL. Strange bedfellows: polyadenylation factors at the promoter. Genes Dev. 2003; 17:1321–1327. [PubMed: 12782649]
122. Uhlmann T, Boeing S, Lehmbacher M, Meisterernst M. The VP16 activation domain establishes an active mediator lacking CDK8 *in vivo*. J Biol Chem. 2007; 282:2163–2173. [PubMed: 17135252]
123. Yang Y, et al. PAF complex plays novel subunit-specific roles in alternative cleavage and polyadenylation. PLoS Genet. 2016; 12:e1005794. [PubMed: 26765774]
124. Yu M, et al. RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science. 2015; 350:1383–1386. [PubMed: 26659056]

125. Jiang C, Pugh BF. Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet. 2009; 10:161–172. [PubMed: 19204718]

126. Kaplan N, et al. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature. 2009; 458:362–366. [PubMed: 19092803]

127. Spies N, Nielsen CB, Padgett RA, Burge CB. Biased chromatin signatures around polyadenylation sites and exons. Mol Cell. 2009; 36:245–254. [PubMed: 19854133]

128. Grosso AR, de Almeida SF, Braga J, Carmo-Fonseca M. Dynamic transitions in RNA polymerase II density profiles during transcription termination. Genome Res. 2012; 22:1447–1456. [PubMed: 22684278]

129. Li W, et al. Alternative cleavage and polyadenylation in spermatogenesis connects chromatin regulation with post-transcriptional control. BMC Biol. 2016; 14:6. [PubMed: 26801249]

130. Ke S, et al. A majority of $m^6A$ residues are in the last exons, allowing the potential for 3′ UTR regulation. Genes Dev. 2015; 29:2037–2053. [PubMed: 26404942]

131. Eckmann CR, Rammelt C, Wahle E. Control of poly(A) tail length. Wiley Interdiscip Rev RNA. 2011; 2:348–361. [PubMed: 21957022]

132. Schmidt MJ, Norbury CJ. Polyadenylation and beyond: emerging roles for noncanonical poly(A) polymerases. Wiley Interdiscip Rev RNA. 2010; 1:142–151. [PubMed: 21956911]

133. Mendez R, Richter JD. Translational control by CPEB: a means to the end. Nat Rev Mol Cell Biol. 2001; 2:521–529. [PubMed: 11433366]

134. Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature. 2014; 508:66–71. [PubMed: 24476825]

135. Chang H, Lim J, Ha M, Kim VN. TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. Mol Cell. 2014; 53:1044–1052. [PubMed: 24582499]

136. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011; 43:904–914. [PubMed: 21925379]

137. Naganuma T, et al. Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. EMBO J. 2012; 31:4020–4034. [PubMed: 22960638]

138. Higgs DR, et al. α-Thalassaemia caused by a polyadenylation signal mutation. Nature. 1983; 306:398–400. [PubMed: 6646217]

139. Prasad MK, et al. A polymorphic 3′ UTR element in ATP1B1 regulates alternative polyadenylation and is associated with blood pressure. PLoS ONE. 2013; 8:e76290. [PubMed: 24098465]

140. Singh P, et al. Global changes in processing of mRNA 3′ untranslated regions characterize clinically distinct cancer subtypes. Cancer Res. 2009; 69:9422–9430. [PubMed: 19934316]

141. Creemers EE, et al. Genome-wide polyadenylation maps reveal dynamic mRNA 3′-end formation in the failing human heart. Circ Res. 2016; 118:433–438. [PubMed: 26671978]

142. Soetanto R, et al. Role of miRNAs and alternative mRNA 3′-end cleavage and polyadenylation of their mRNA targets in cardiomyocyte hypertrophy. Biochim Biophys Acta. 2016; 1859:744–756. [PubMed: 27032571]

143. Park JY, et al. Comparative analysis of mRNA isoform expression in cardiac hypertrophy and development reveals multiple post-transcriptional regulatory modules. PLoS ONE. 2011; 6:e22391. [PubMed: 21799842]

144. Hu J, Lutz CS, Wilusz J, Tian B. Bioinformatic identification of candidate *cis*-regulatory elements involved in human mRNA polyadenylation. RNA. 2005; 11:1485–1493. [PubMed: 16131587]

145. Cheng Y, Miura RM, Tian B. Prediction of mRNA polyadenylation sites by support vector machine. Bioinformatics. 2006; 22:2320–2325. [PubMed: 16870936]

146. Nunes NM, Li W, Tian B, Furger A. A functional human poly(A) site requires only a potent DSE and an A-rich upstream sequence. EMBO J. 2010; 29:1523–1536. [PubMed: 20339349]

147. Sheets MD, Ogg SC, Wickens MP. 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. 1990; 18:5799–5805. [PubMed: 2170946]

148. Shi Y, et al. Molecular architecture of the human pre-mRNA 3′ processing complex. Mol Cell. 2009; 33:365–376. This report details the purification of an active polyadenylation complex on

substrate RNA and the identification of more than 80 core and associated proteins. [PubMed: 19217410]

149. Chan SL, et al. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. Genes Dev. 2014; 28:2370–2380. [PubMed: 25301780]
150. Schonemann L, et al. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. Genes Dev. 2014; 28:2381–2393. [PubMed: 25301781]
151. 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; 23:616–626. [PubMed: 14749727]
152. Takagaki Y, Manley JL. RNA recognition by the human polyadenylation factor CstF. Mol Cell Biol. 1997; 17:3907–3914. [PubMed: 9199325]
153. Chen F, Wilusz J. Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. Nucleic Acids Res. 1998; 26:2891–2898. [PubMed: 9611233]
154. Mandel CR, et al. Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. Nature. 2006; 444:953–956. [PubMed: 17128255]
155. Bai Y, et al. Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. Mol Cell. 2007; 25:863–875. [PubMed: 17386263]
156. Yang Q, Gilmartin GM, Doublié S. Structural basis of UGUA recognition by the Nudix protein $CFI_m25$ and implications for a regulatory role in mRNA 3′ processing. Proc Natl Acad Sci USA. 2010; 107:10062–10067. [PubMed: 20479262]
157. Hunt AG, Xing D, Li QQ. Plant polyadenylation factors: conservation and variety in the polyadenylation complex in plants. BMC Genomics. 2012; 13:641. [PubMed: 23167306]
158. Zhang H, Lee JY, Tian B. Biased alternative polyadenylation in human tissues. Genome Biol. 2005; 6:R100. This is the first demonstration that isoforms using proximal and distal PASs are expressed with bias in certain tissues, for example in the brain and blood. [PubMed: 16356263]
159. Beaudoing E, Gautheret D. Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST data. Genome Res. 2001; 11:1520–1526. [PubMed: 11544195]
160. Lianoglou S, Garg V, Yang JL, Leslie CS, Mayr C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev. 2013; 27:2380–2396. [PubMed: 24145798]
161. Liu D, et al. Systematic variation in mRNA 3′-processing signals during mouse spermatogenesis. Nucleic Acids Res. 2007; 35:234–246. [PubMed: 17158511]
162. Smibert P, et al. Global patterns of tissue-specific alternative polyadenylation in *Drosophila*. Cell Rep. 2012; 1:277–289. [PubMed: 22685694]
163. Lee JY, Ji Z, Tian B. Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3′-end of genes. Nucleic Acids Res. 2008; 36:5581–5590. [PubMed: 18757892]
164. Shepard PJ, et al. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA. 2011; 17:761–772. [PubMed: 21343387]
165. Dai W, et al. A post-transcriptional mechanism pacing expression of neural genes with precursor cell differentiation status. Nat Commun. 2015; 6:7576. [PubMed: 26144867]
166. Dass B, et al. Loss of polyadenylation protein τCstF-64 causes spermatogenic defects and male infertility. Proc Natl Acad Sci USA. 2007; 104:20374–20379. [PubMed: 18077340]
167. Sartini BL, Wang H, Wang W, Millette CF, Kilpatrick DL. Pre-messenger RNA cleavage factor I (CFIm): potential role in alternative polyadenylation during spermatogenesis. Biol Reprod. 2008; 78:472–482. [PubMed: 18032416]
168. Soumillon M, et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep. 2013; 3:2179–2190. [PubMed: 23791531]
169. Zhang P, et al. MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. Cell Res. 2015; 25:193–207. [PubMed: 25582079]
170. Goh WS, et al. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. Genes Dev. 2015; 29:1032–1044. [PubMed: 25995188]

171. Watanabe T, Cheng EC, Zhong M, Lin H. Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. Genome Res. 2015; 25:368–380. [PubMed: 25480952]
172. Bao J, et al. UPF2-dependent nonsense-mediated mRNA decay pathway is essential for spermatogenesis by selectively eliminating longer 3′ UTR transcripts. PLoS Genet. 2016; 12:e1005863. [PubMed: 27149259]
173. Fanourgakis G, Lesche M, Akpinar M, Dahl A, Jessberger R. Chromatoid body protein TDRD6 supports long 3′ UTR triggered nonsense mediated mRNA decay. PLoS Genet. 2016; 12:e1005857. [PubMed: 27149095]
174. Gruber AR, et al. Global 3′ UTR shortening has a limited effect on protein abundance in proliferating T cells. Nat Commun. 2014; 5:5465. [PubMed: 25413384]
175. Fu Y, et al. Differential genome-wide profiling of tandem 3′ UTRs among human breast cancer and normal cells by high-throughput sequencing. Genome Res. 2011; 21:741–747. [PubMed: 21474764]
176. Morris AR, et al. Alternative cleavage and polyadenylation during colorectal cancer development. Clin Cancer Res. 2012; 18:5256–5266. [PubMed: 22874640]
177. Flavell SW, et al. Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. Neuron. 2008; 60:1022–1038. [PubMed: 19109909]
178. Chang JW, et al. mRNA 3′-UTR shortening is a molecular signature of mTORC1 activation. Nat Commun. 2015; 6:7218. [PubMed: 26074333]

## Abbreviations

**PUF protein**
(Pumilio and FBF homology family protein). A member of a family of RNA-binding proteins that regulate aspects of mRNA metabolism by binding to specific sequences in 3′ untranslated regions

**STAU1-mediated mRNA decay**
An mRNA decay mechanism in which RNA structures in the 3′ untranslated region interact with double-stranded RNA-binding protein Staufen homologue 1 (STAU1) to mediate mRNA decay

**AU-rich element-mediated decay**
mRNA decay elicited by the presence of AU-rich elements (AREs) in the 3′ untranslated region

**PIWI-interacting RNAs**
Small non-coding RNAs that form RNA–protein complexes with PIWI proteins to silence transposable elements in germline cells of metazoans

**Non-stop decay**
An mRNA decay mechanism that specifically degrades mRNAs without a stop codon

**Exosome**
A nuclear or cytoplasmic multiprotein complex that degrades mRNAs through the activity of 3′-to-5′ exoribonucleases

**Non-canonical PAPs**

(Non-canonical poly(A) polymerases). Enzymes that have distinct structural features and are capable of synthesizing poly(A) tails but are not typically associated with the polyadenylation machinery

**Paused Pol II**

(Paused RNA polymerase II). Pol II that has paused in the promoter-proximal region of the mRNA and is poised for productive elongation

**Paraspeckle**

A dynamic nuclear compartment composed of RNA-binding proteins and RNAs. The functions of paraspeckles are not entirely clear

**Box 1**

## Core sequence elements and factors involved in cleavage and polyadenylation

Cleavage and polyadenylation (hereafter referred to as polyadenylation) is controlled by cis elements located upstream and downstream of the polyadenylation site (PAS) (see the figure). In vertebrates, upstream elements include the hexamers A[A/U]UAAA or other close variants, U-rich elements and UGUA elements. Downstream elements include U-rich and GU-rich (typically in the form of GUGU) elements. A CA sequence is often found immediately 5′ to the cleavage site (see the figure; indicated by a lightning bolt). In addition, upstream UAUA elements and downstream G-rich sequences can frequently be found near PASs, typically more than 40 nt away[144]. The 'strength' of a PAS seems to be defined by these sequence elements in a combinatorial manner[145]. A functional PAS in human cells can consist of only an A-rich upstream sequence and strong U-rich downstream elements[146]. It is notable that this make-up of the core PAS is analogous to the organization of core RNA polymerase II (Pol II) promoters (discussed in REF. 6).

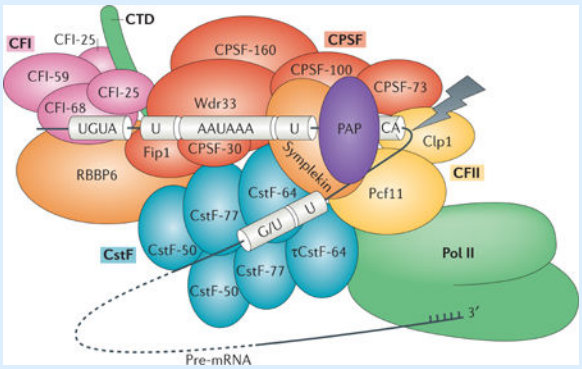
Not surprisingly, variants of the A[A/U]UAAA sequence are weaker at directing polyadenylation in vitro than the consensus hexamer[147]. However, variants are fairly common, especially for upstream APA sites[11]. For example, AAGAAA is commonly found in upstream (not 3′-most) APA sites[11], even though this sequence is essentially inactive in vitro. Other elements, such as upstream UGUA and downstream GU-rich sequences, also are less frequent in upstream APA sites, supporting the idea that such sites have suboptimal strength. This is probably important for the regulation of polyadenylation and could explain global APA regulation during cell proliferation and differentiation, when the concentration of polyadenylation factors changes[22]. By contrast, 3′-most PASs are typically strong, presumably to ensure proper transcription termination.

The polyadenylation machinery in metazoans is composed of ~20 core proteins, including four protein complexes and several single proteins[148]. The complexes are cleavage and polyadenylation specificity factor (CPSF), which contains CPSF 160 kDa subunit (CPSF160; also known as CPSF1), CPSF100 (also known as CPSF2), CPSF73 (also known as CPSF3), CPSF30 (also known as CPSF4), FIP1 (factor interacting with PAP) and WDR33; cleavage stimulation factor (CSTF), which contains CSTF 77 kDa subunit (CSTF77), CSTF50 (also known as CSTF1) and either CSTF64 or its paralogue, τCSTF64; cleavage factor I (CFI), which contains CFI 25 kDa subunit (CFI25) and either CFI68 or CFI59; and cleavage factor II (CFII), which contains PCF11 and CLP1. Single proteins include symplekin, poly(A) polymerase (PAP), retinoblastoma-binding protein 6 (RBBP6) and RNA polymerase II (Pol II), specifically the Pol II regulatory carboxy-terminal domain (CTD). Nuclear poly(A)-binding protein 1 (PABPN1) (not shown) is important for synthesis of an appropriately sized poly(A) tail and thus might also be considered to be a core factor.

As expected, many polyadenylation factors are RNA-binding proteins and display sequence-specific RNA binding. CPSF makes multiple RNA contacts, including with the AAUAAA element through CPSF30 and WDR33 (REFS 149,150), and with U-rich

sequences through FIP1 (REF 151). CFI25 binds the UGUA element[119], CFI68 and CFI59 also contact RNA, and CSTF64 and τCSTF64 interact with U- and GU-rich downstream elements[152,153]. CPSF73 is the endonuclease and has a preference for a CA dinucleotide at the cleavage site[154]. Some proteins that do not bind RNA have scaffolding functions, such as symplekin and the Pol II CTD. Both CFI and CSTF function as dimers in the polyadenylation machinery[155,156].

Despite considerable divergence between yeast and mammals in the core RNA sequences that constitute the PAS (see Supplementary information S1 (box)), nearly all mammalian polyadenylation factors have homologues in yeast, with the exception of the CFI proteins and CSTF50. However, there seems to be some variation in the make-up of subcomplexes in yeast[5], and the yeast polyadenylation factor Hrp1p, which interacts with UA-rich elements, is missing from metazoans. The polyadenylation machinery in plants is similar to that in metazoans, but with substantial gene (and thus protein) duplications[157].



**Box 2**

## Tissue-specific APA patterns

Alternative polyadenylation (APA) patterns are, to a great extent, tissue specific. Corroborating early analyses[158,159], deep-sequencing analyses have detailed the existence of tissue-specific APA profiles[12,160]. Some human tissues have a global tendency to favour certain APA isoform types[158]; for example, neuronal tissues favour isoforms that use distal polyadenylation sites (PASs) in 3′ untranslated regions (3′ UTRs), whereas the use of proximal PASs is favoured in blood cells and testis tissue[158,161]. The phenomenon in which mRNAs with extremely short or long 3′ UTRs are expressed in testis and brain, respectively, also occurs in flies[162]. Comparative analysis of the tissues of five different mammals indicated that APA profiles in different tissues are well conserved across species[12]. Importantly, ubiquitously expressed genes are more likely to express APA isoforms in different tissues than genes with a restricted tissue expression[160], raising the possibility that APA in the 3′ UTR (referred to as 3′ UTR-APA) has an important role in tissue-specific regulation of these genes. Consistent with this theory, genes that are evolutionarily old, which tend to be more widely expressed, are more likely to undergo APA than new genes[163].

Recent studies have shed some light on the mechanisms of expression of long 3′ UTR-APA isoforms in the brain. First, in *Drosophila melanogaster*, the RNA-binding protein (RBP) Embryonic-lethal abnormal visual (Elav) inhibits proximal PAS usage[92,93] (see main text). Second, during neuronal differentiation, when 3′ UTRs generally lengthen[164], AU-rich element-mediated decay (which targets isoforms with longer 3′ UTRs, as they are more likely to contain AU-rich elements (AREs)) seems to be suppressed[165]. Consistent with this, tristetraprolin, an RBP with a role in ARE-medicated decay, is downregulated by the microRNA (miRNA) miR-9 during neurogenesis[165]. Thus, the combined activity of RBPs that favour preferential selection of distal PASs and of those that stabilize mRNAs boost the abundance of long isoforms in neuronal cells, and this is important for neuronal cells presumably because of 3′ UTR-mediated mRNA localization in dendrites and axons (see main text).

In a recent study of 3′ UTR-APA isoform expression during spermatogenesis, 3′ UTRs were found to drastically shorten as spermatocytes differentiate into spermatids[129]. This too can be attributed to both regulation of PAS choice and mRNA stability. First, genes producing transcripts with shorter 3′ UTRs are more likely to undergo transcriptional upregulation and to reside in open chromatin, suggesting that APA is regulated by these features[129]. This is consistent with previous studies showing that genes transcribed at high rates tend to undergo more efficient polyadenylation and, presumably as a result, use more promoter-proximal PASs (see main text)[115,116]. This mechanism may function in conjunction with the unique regulation of polyadenylation factor expression during spermatogenesis, such as the regulation suggested for cleavage stimulation factor 64 kDa subunit τ-variant (τCSTF64τ)[166] and cleavage factor I[167]. Second, mRNA decay mechanisms are highly potent and responsible for the global elimination of RNAs during the transition from spermatocytes to spermatids, which contain ~12 pg and ~2.5 pg total RNA per cell, respectively[168]. Thus, long isoforms that contain destabilizing elements are

rapidly degraded, probably by multiple decay mechanisms, including the degradation of mRNAs containing transposable elements, mediated by PIWI-interacting RNAs[129,169–171], and the nonsense-mediated decay pathway, which degrades transcripts with abnormally long 3′ UTRs (among others)[172,173]. Importantly, transcripts with short 3′ UTRs, which lack these destabilizing elements or features and thus escape mRNA degradation, are thought to be stored for translation at a later developmental stage, when transcription in globally inhibited.

**Box 3**

## Global regulation of APA

Alternative polyadenylation (APA) can be globally regulated in response to changes in cell proliferation and differentiation. General shortening of 3′ untranslated regions (3′ UTRs) during T cell activation (which is accompanied by cell proliferation) was the first example of such regulation in response to changes in cell proliferation status[21]. It was proposed that transcripts with shorter 3′ UTRs evade targeting by microRNAs (miRNAs) and thus increase the protein output from a transcript. However, a more recent study using 3′ end sequencing and mass spectrometry showed that although 3′ UTR shortening during T cell activation is conserved between humans and mice, orthologous genes do not exhibit similar APA profiles[174]. In addition, 3′ UTR shortening was not accompanied by changes in mRNA and protein levels, suggesting that APA can have limited effects on overall protein output.

Global 3′ UTR regulation according to cellular proliferation and differentiation status has been observed in a number of biological systems and processes, including during embryonic development, the differentiation of myoblasts and of embryonic stem cells[22,164], and the de-differentiation of many cell types into induced pluripotent stem cells[64]. As most factors in the polyadenylation complex are highly expressed in proliferative cells compared to their levels in differentiated cells[22,64], it was hypothesized that global regulation of polyadenylation activity might underlie the global APA profile changes that are observed when cells alter their proliferation state. This view is consistent with reporter assays showing decreased proximal PAS usage in differentiated mouse myoblasts compared with proliferating cells[22].

The correlation between cellular proliferation status and 3′ UTR size extends to cancer cells. A meta-analysis of microarray data indicated that cancer cells have shorter 3′ UTRs than non-transformed cell lines[23], and a recent bioinformatics study of RNA-seq data from 358 tumour-versus-normal tissue pairs identified 1,346 genes for which the transcripts underwent significant and recurrent APA in seven tumour types[63]. The majority of transcripts (61–98%) displayed 3′ UTR shortening in the tumours. Another study compared cells (BJ primary fibroblast and the mammary epithelial cell line MCF10A) in the proliferating, growth-arrested and transformed states, and found that proliferation is a more important determinant of 3′ UTR length than transformation[49]. Notably, although global shortening of 3′ UTRs occurs in cancer cells, a substantial fraction of transcripts appear to undergo 3′ UTR lengthening, for example in breast cancer cells[175], colorectal cancer cells[176] and lymphoma cells[140]. Interestingly, genes producing transcripts that have lengthened 3′ UTRs in cancer cells seemed to be enriched for certain functional groups, such as cell–cell adhesion[140,175,176]. Overall, the APA profile in cancer cells may be more complicated than was initially suggested, and additional studies are needed to dissect the different groups of genes with respect to their APA regulation.

Finally, APA can be globally regulated by specific extracellular cues. For example, isoforms using the proximal PAS, which encode truncated proteins or have short 3′

UTRs, are generally upregulated when neuronal cells are activated by membrane depolarization agents[177], and activation of the mTOR pathway leads to global 3′ UTR shortening[178]. Although the mechanism underlying this APA is unknown, isoforms with shortened 3′ UTRs have greater translational potential as analysed by polysome profiling. This supports the idea that 3′ UTR shortening can increase protein output, a view that has been challenged in other settings, as described above. Interestingly, transcripts encoding proteins related to ubiquitin-mediated proteolysis, which is important for cell cycle progression, were most significantly affected by the 3′ UTR shortening elicited by mTOR activation.

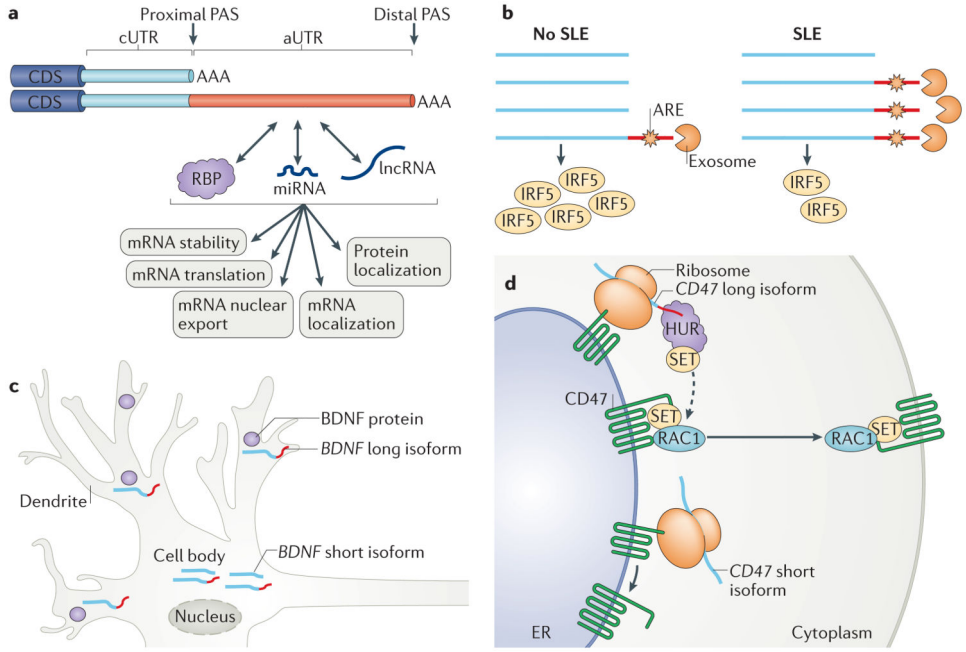



**Figure 1. 3′ UTR-APA**

**a** | Alternative polyadenylation (APA) leading to the production of two mRNA isoforms with different 3′ untranslated regions (3′ UTRs) — termed 3′ UTR-APA here — is shown. The 3′ UTR region upstream of the proximal polyadenylation site (PAS) is found in both short (top) and long (bottom) isoforms and is denoted the constitutive UTR (cUTR), whereas the downstream region is present in the long isoform only and is termed the alternative UTR (aUTR). Interactions between the aUTR and RNA-binding proteins (RBPs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) can have various functional consequences. The poly(A) tail is represented as AAA. **b** | In the case of the gene encoding human IFN-regulatory factor 5 (IRF5), APA of the transcript produces a long 3′ UTR isoform that is more rapidly degraded owing to the presence of a destabilizing AU-rich element (ARE) in the aUTR. The ARE and cytoplasmic exosome mediate mRNA decay. In patients with systemic lupus erythematosus (SLE), a single nucleotide polymorphism reducing the use of the proximal PAS leads to the production of long isoforms at the expense of short isoforms, which results in reduced IRF5 levels. **c** | Differential mRNA localization of brain-derived neurotrophic factor (BDNF) 3′ UTR-APA isoforms in neurons. The long isoform localizes to dendrites more than the short isoform, and this supports dendrite-localized protein synthesis. **d** | Differential localization of the transmembrane CD47 proteins encoded by long or short APA isoforms. Both isoforms are translated on the ER membrane. The aUTR of the long isoform is bound by the RBP Hu antigen R (HUR), which leads to the localization of CD47 protein to the cell membrane through a cascade of interactions (dashed arrow) involving the phosphatase 2A inhibitor SET and RAC1. The protein generated from the short isoform remains in the ER. CDS, coding sequence.

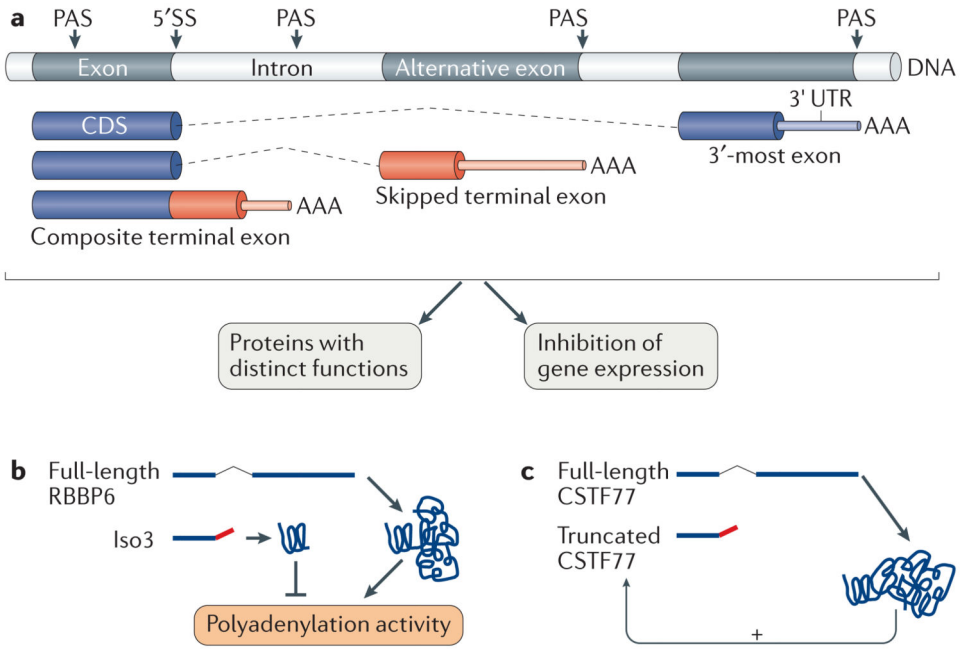



**Figure 2. UR-APA**

**a** | Alternative polyadenylation (APA) in upstream regions (URs) of mRNAs — termed UR-APA here — can lead to the production of isoforms with different 3′-terminal exons and, hence, different coding sequences and 3′ untranslated regions (3′ UTRs). Three isoforms are shown, with their respective terminal exon types indicated. Splicing is indicated by a dashed line. The 'canonical' isoform (top) is formed by the use of the polyadenylation site (PAS) in the 3′-most exon. The use of a PAS in an alternative exon that is excluded from the canonical isoform generates a transcript containing a skipped terminal exon (middle). Inhibition of splicing at the indicated 5′ splicing site (5′SS) results in the inclusion of part of the downstream intron and use of a PAS within that intron; such a transcript is described as containing a composite terminal exon (bottom). Regions not present in the canonical isoform are shown in red. The functional consequences of UR-APA are indicated. **b** | UR-APA of the transcript encoding polyadenylation factor retinoblastoma-binding protein 6 (RBBP6) produces an isoform encoding a dominant negative protein, Iso3. **c** | UR-APA of the mRNA encoding polyadenylation factor cleavage stimulation factor 77 kDa subunit (CSTF77) produces a short isoform that encodes a truncated protein with no apparent functions (not shown). The full-length protein activates the usage of the upstream PAS, thereby increasing the levels of the short mRNA and forming a negative feedback loop. CDS, coding sequence.

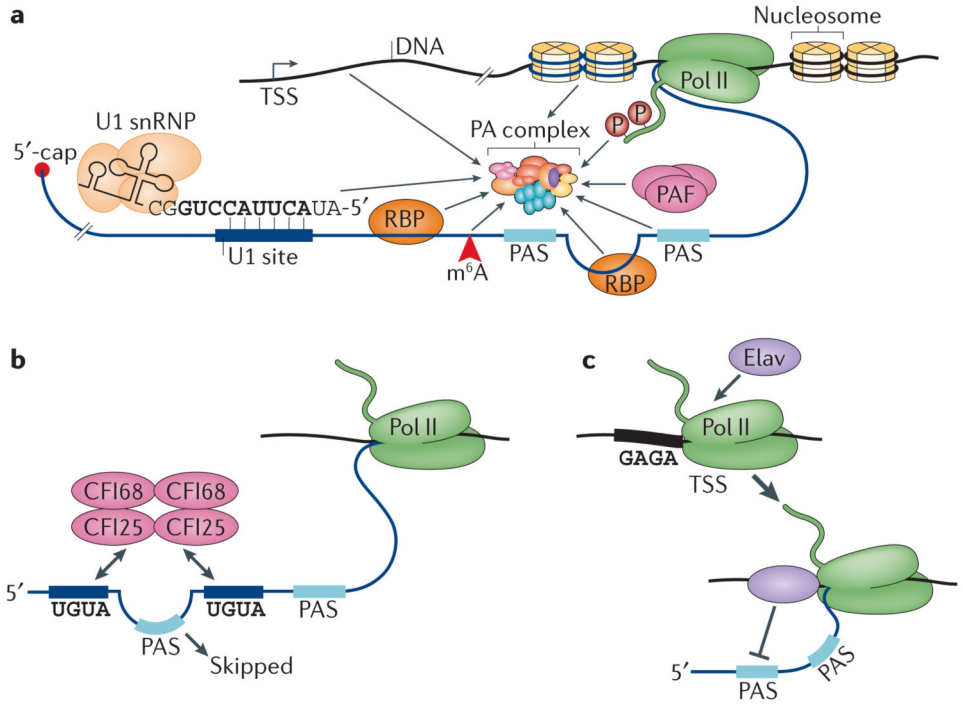



**Figure 3. Regulation of APA**

**a** | The choice of polyadenylation site (PAS) during alternative polyadenylation (APA) can be influenced by various factors, including the gene promoter at the transcription start site (TSS); recruitment of polyadenylation factors directly or of proteins that influence PAS choice; nucleosome density in the region around the PAS; RNA polymerase II (Pol II)-mediated transcription elongation by the Pol II-associated factor (PAF) complex; the function of various RNA-binding proteins (RBPs) associated with the nascent transcript; the presence of $N^6$-methyladenosine ($m^6A$); and inhibition of polyadenylation by the splicing factor U1 small nuclear ribonucleoprotein (U1 snRNP). See the main text for more details. **b** | A proposed model for the regulation of APA by the cleavage factor I (CFI) complex. Two UGUA elements upstream and downstream of a proximal PAS are recognized by the heterodimeric CFI complex, which consists of CFI 68 kDa subunit (CFI68) and CFI25, leading to skipping of the PAS. **c** | Regulation of neuronal APA in *Drosophila melanogaster* by the RBP Embryonic-lethal abnormal visual (Elav). Elav is recruited to Pol II at promoter regions that contain a GAGA sequence, which can cause Pol II pausing. Elav inhibits proximal PAS usage, leading to the expression of long APA isoforms during neurogenesis. PA complex, polyadenylation complex.