

# Regulation of gene expression by alternative polyadenylation in health and disease

Bin Tian <sup>1,2</sup> , Shan Yu <sup>1,2</sup> & Qiang Zhang<sup>1</sup>

## Abstract

More than half of the human protein-coding genes display alternative polyadenylation (APA), whereby 3'-end processing of the nascent RNA takes place at different sites. APA leads to mRNA isoforms containing different 3' untranslated regions (3'UTRs), which generally modulate mRNA metabolism *in cis* but can also exert cellular functions *in trans*. In addition, intronic APA alters protein sequences at the carboxy-terminal region or inhibits gene expression through premature transcription termination. APA is increasingly recognized as a key layer of transcriptomic regulation that defines cell identity and proliferation and/or differentiation states, as well as controlling cellular responses to environmental cues. The relevance of APA for human health is highlighted by the many pathological conditions that are associated with APA dysregulation, including cancer, developmental disorders and neurodegeneration, as well as the disease risks associated with a growing number of genetic variations shown to affect APA. Here, we discuss physiological and pathological APA dynamics, the human mutations and genetic variants that are associated with changes in APA, and our current understanding of the functional effects and regulatory mechanisms of APA.

## Sections

Introduction

APA dynamics in physiological conditions


APA in pathological contexts

Functional effects of 3'UTR-APA

Mechanisms for the regulation of APA

Conclusions and future perspectives

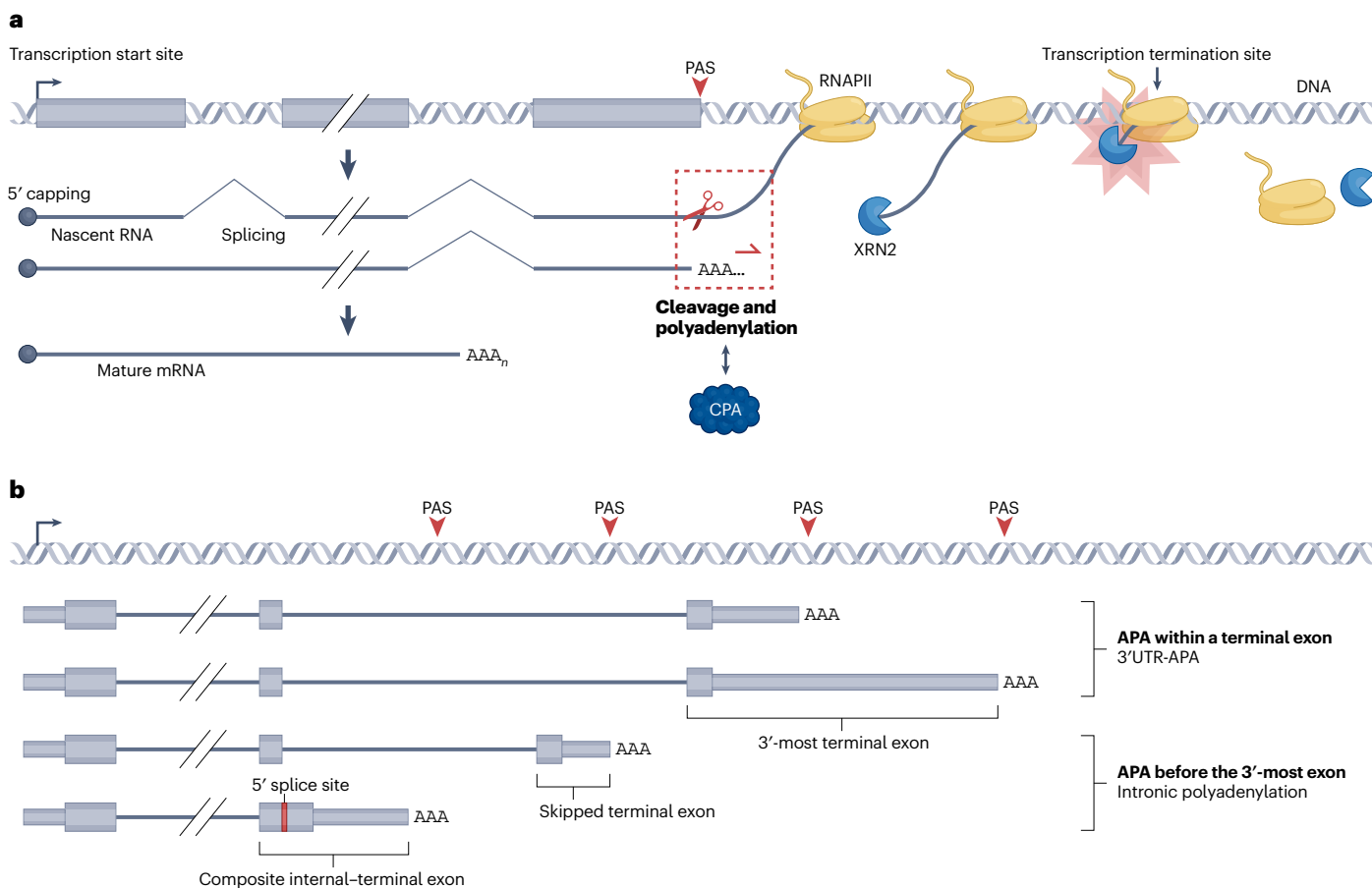
<sup>1</sup>Genome Regulation and Cell Signaling Program, Ellen and Ronald Caplan Cancer Center, The Wistar Institute, Philadelphia, PA, USA. <sup>2</sup>Center for Systems and Computational Biology, The Wistar Institute, Philadelphia, PA, USA.

 e-mail: [btian@wistar.org](mailto:btian@wistar.org)

## Introduction

Almost all eukaryotic mRNAs and long non-coding RNAs (lncRNAs; generally, longer than 200 nucleotides) that are transcribed by RNA polymerase II (RNAPII) use cleavage and polyadenylation (CPA) for 3'-end maturation (Fig. 1a). The CPA site (also known as the polyadenylation site (PAS)) is recognized by the CPA machinery through interactions between RNA motifs surrounding the PAS (known as PAS motifs) and the RNA-binding proteins (RBPs) in the CPA machinery (Box 1). CPA comprises two coupled enzymatic reactions (Fig. 1a) – cleavage of the nascent RNA and addition of a polyadenosine (poly(A)) tail (of up to 250 adenosines in human cells). Whereas the cleavage step affects transcription termination, which takes place within a few kilobases after the PAS<sup>1</sup> (Fig. 1a), the polyadenylation step has important roles in mRNA metabolism, such as nuclear export, stability and translation<sup>2</sup>. CPA is therefore a nexus for gene expression, connecting nuclear events with cytoplasmic fates for gene transcripts. In accord with this, mutations of CPA factor genes as well as those of PAS motifs in genes with crucial cellular functions have been implicated in many human diseases (Supplementary Tables 1 and 2).

More than half of the human mRNA genes harbour multiple PASs, leading to the expression of alternative polyadenylation (APA) isoforms (Fig. 1b). Those genes that display APA tend to have ubiquitous expression across cell types<sup>3</sup> and a long evolutionary history<sup>4</sup>. An APA event can be classified as one of two types depending on its functional consequences and relationship with splicing: APA in a terminal exon, typically the 3'-most one, alters the length of the 3' untranslated region (3'UTR) of mRNA and is hence named 3'UTR-APA (also known as tandem APA); whereas APA taking place upstream of the 3'-most exon, mainly in introns, leads to changes in both the 3'UTR and the coding sequence, and has been referred to variously as coding region APA (CR-APA), alternative last exon (ALE), premature CPA (PCPA) or intronic polyadenylation (IPA; the term we use here) (Fig. 1b). Depending on the splicing context, an IPA transcript could end with either a composite internal-terminal exon or a skipped terminal exon (Fig. 1b); this difference has important mechanistic and functional implications (see below). Across mammals, 3'UTR-APA events are generally more conserved than IPA events<sup>5</sup>; about 50% and 20% of human genes that display APA have conserved 3'UTR-APA and IPA events, respectively, in rodents<sup>5</sup>.



**Fig. 1 | The cleavage and polyadenylation mechanism and alternative polyadenylation.** **a**, Cleavage and polyadenylation (CPA) on a nascent RNA that is transcribed by RNA polymerase II (RNAPII). The site for CPA (also known as the polyadenylation site (PAS)) is indicated, which defines the transcript end site. CPA is shown as two separate reactions (boxed with dashed lines). Transcription termination, whereby RNAPII is evicted by the 5'-to-3' exonuclease XRN2, takes place downstream of the PAS, within up to several kilobases depending on the gene. **b**, Two main types of alternative polyadenylation (APA) are defined by the location of the alternative PAS in the context of splicing configuration:

APA within a terminal exon – typically the 3'-most exon – or between alternative terminal exons and the 3'-most exon. The former mainly regulates the size of the 3' untranslated region (3'UTR) and is therefore named 3'UTR-APA. The latter also alters the coding sequence and is often coupled with splicing; this is known as intronic polyadenylation (IPA). Two subtypes of IPA are based on the splicing configuration of the resulting terminal exon of the transcript – namely, a skipped terminal exon or a composite internal-terminal exon (containing a 5' splice site (5'SS)). Note that 3'UTR-APA can also take place in these alternative terminal exons, albeit less commonly than in the 3'-most exons.

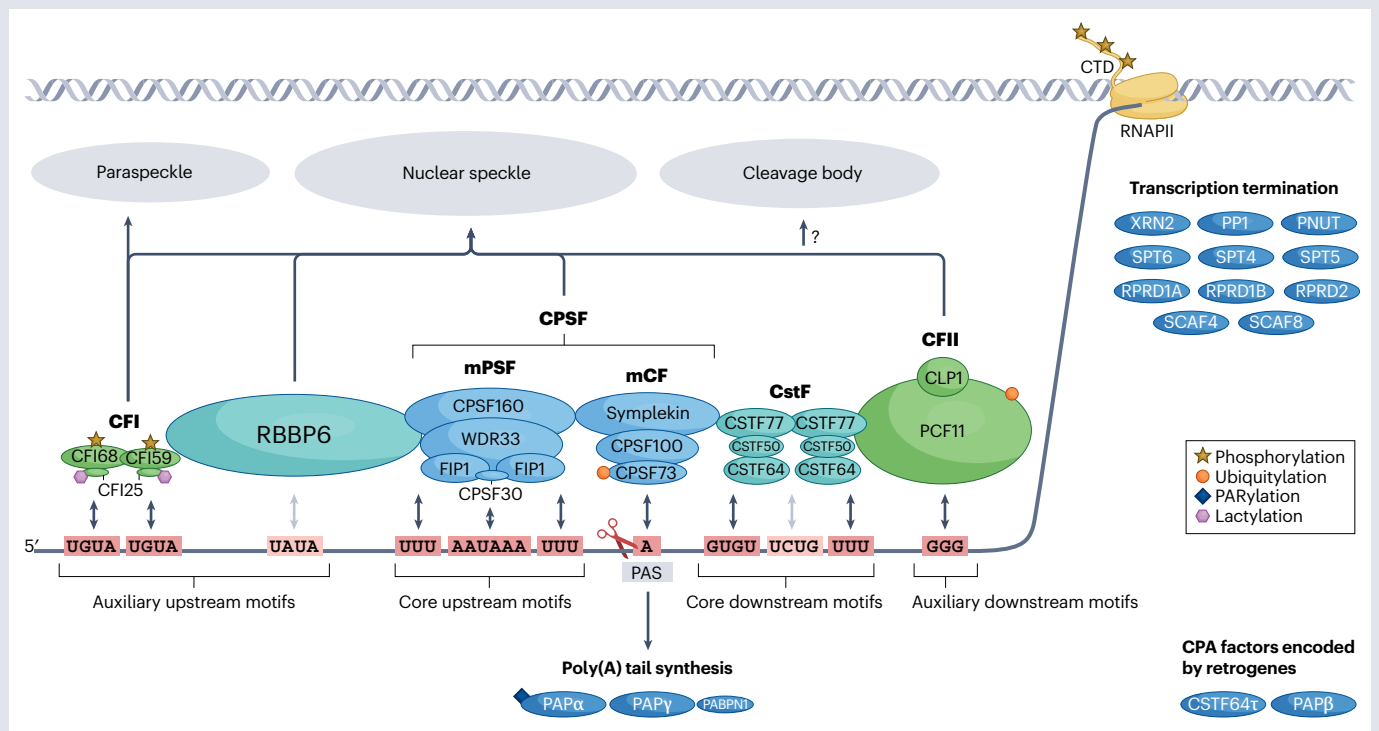
## Box 1 | Polyadenylation site motifs and the CPA machinery

The polyadenylation site (PAS) is defined by its surrounding motifs (referred to as PAS motifs) (see the figure). PAS motifs are largely conserved in vertebrates, vary slightly in lower metazoans, but differ greatly in *Saccharomyces cerevisiae* and plants<sup>219,220</sup>. The general usage level or 'strength' of a PAS is determined by the types of PAS motifs and their relative placements to one another<sup>151,221–224</sup>. The AAUAAA motif and its close variants (AUUAAA being the most prevalent) are known as the PAS hexamer which, together with the surrounding U-rich (UUU) motifs, are core motifs upstream of the PAS. Core motifs downstream of the PAS include GU-rich (GUGU) and U-rich sequences. Further upstream and downstream of the PAS are auxiliary motifs, including UGUA and G-rich (GGG) motifs, which are often absent from a proximal alternative PAS. Upstream UUAU and downstream UCUG motifs are also enriched for a strong PAS<sup>151,225</sup>, although their binding proteins are unclear (as indicated by their lighter colour in the figure).

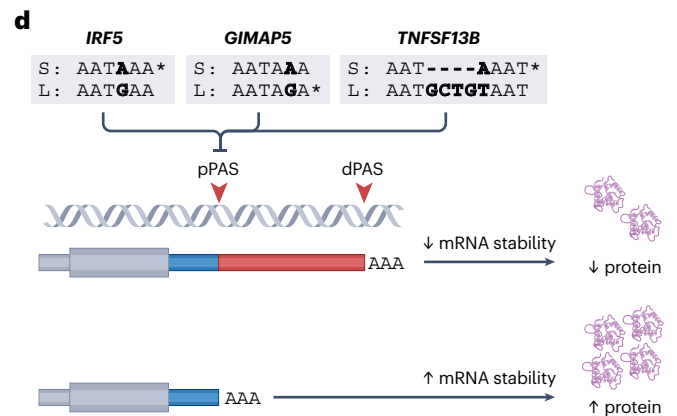
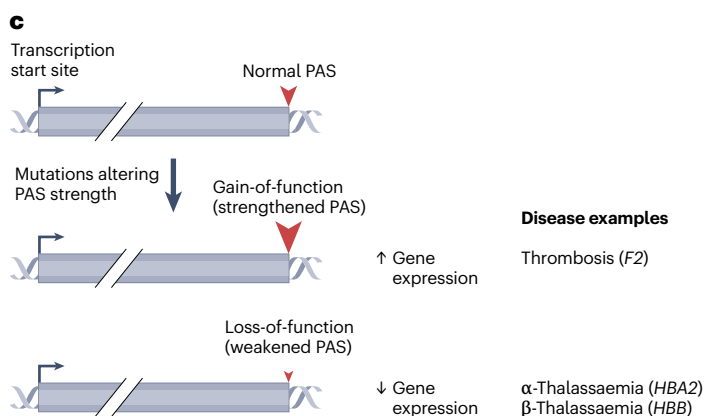
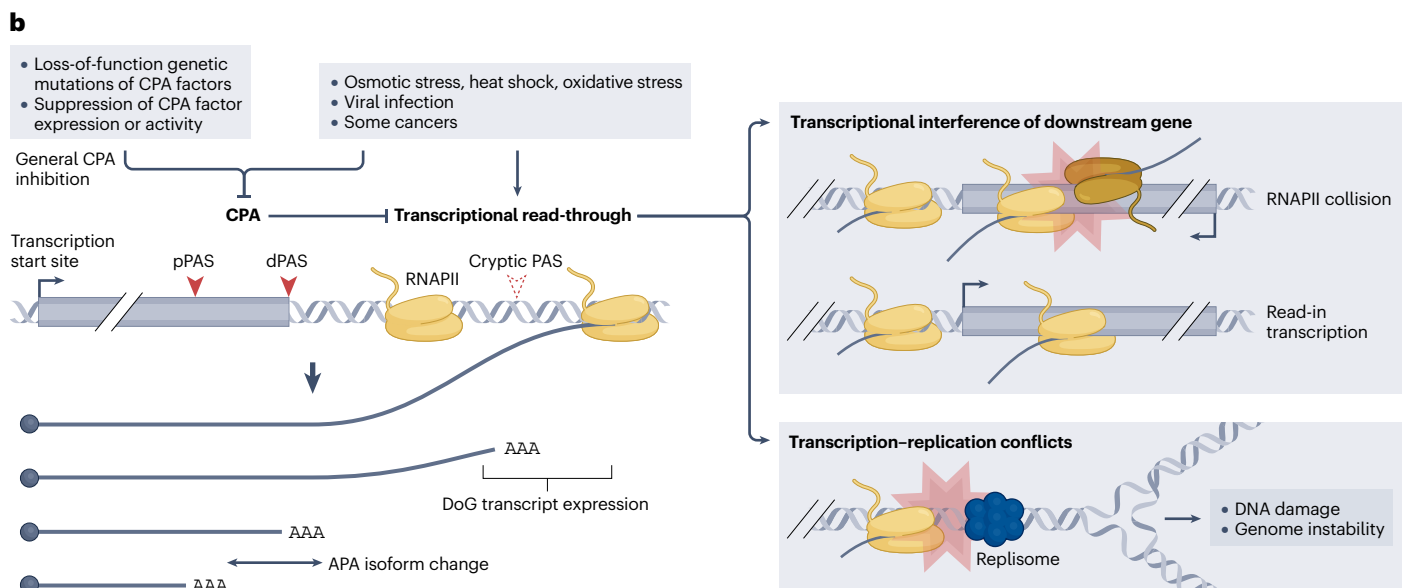
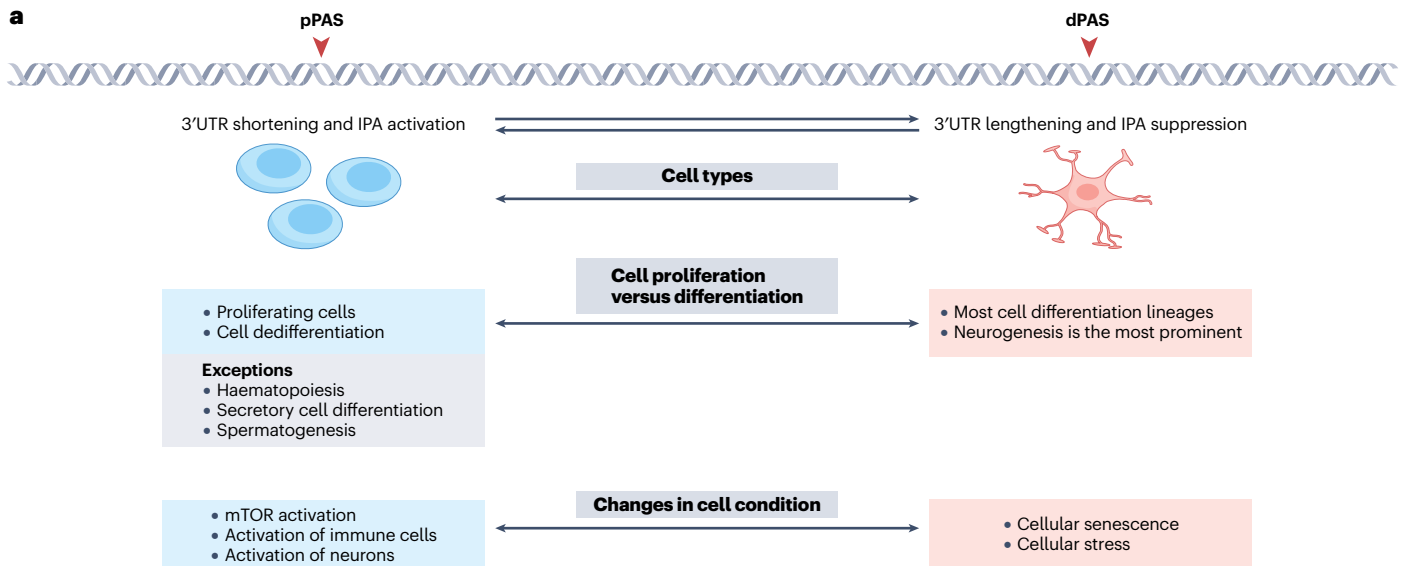
The cleavage and polyadenylation (CPA) machinery in human cells is a mega-Dalton complex composed of more than 20 core proteins, some of which form subcomplexes<sup>226</sup>. The cleavage and polyadenylation specificity factor (CPSF) subcomplex can be further divided into two distinct modules: the mammalian cleavage factor (mCF), which carries out the cleavage reaction and contains CPSF73, CPSF100 and SYMPLEKIN (encoded by *CPSF3*, *CPSF2* and *SYMPK*, respectively); and the mammalian polyadenylation specificity factor (mPSF), which recognizes core upstream PAS motifs and contains CPSF30, FIP1, WDR33 and CPSF160 (encoded by *CPSF4*, *FIP1L1*, *WDR33* and *CPSF1*, respectively) (two copies of FIP1 are shown in the figure, as reported previously<sup>227,228</sup>). Cleavage stimulation factor (CstF; containing CSTF50, CSTF64 and CSTF77, encoded by *CSTF1*,

*CSTF2* and *CSTF3*, respectively), cleavage factor I (CFI; containing CFI25, CFI59 and CFI68, encoded by *NUDT21*, *CPSF7* and *CPSF6*, respectively) and cleavage factor II (CFII; containing CLP1 and PCF11, encoded by *CLP1* and *PCF11*, respectively) subcomplexes enhance CPA through binding to their cognate motifs. Other single proteins in the CPA machinery include RBBP6, which activates CPSF73; RNA polymerase II (RNAPII) whose C-terminal domain (CTD) has a key role in CPA; poly(A) polymerases PAP $\alpha$  and PAP $\gamma$  (encoded by *PAPOLA* and *PAPOLG*, respectively), which are responsible for synthesis of the polyadenosine (poly(A)) tail; and PABPN1, which helps control the size of the poly(A) tail. In addition, CSTF64 $\tau$  and PAP $\beta$  (encoded by intron-less retrogenes *CSTF2T* and *PAPOLB*, respectively) are paralogues of CSTF64 and PAP $\alpha$ , respectively. Whereas PAP $\beta$  seems to be expressed exclusively in testis, CSTF64 $\tau$  expression is detected ubiquitously. Major transcription termination factors are also shown in the figure; some have global effects on alternative polyadenylation (APA), such as SCAF4 and SCAF8 (ref. 229).

Different CPA factors can have distinct localization in the nucleus, such as paraspeckles for CFI proteins and nuclear speckles for RBBP6, FIP1 and PCF11. Although it has not been fully elucidated, the CPA machinery is plausibly associated with the cleavage body<sup>230,231</sup> and its subnuclear localization might influence CPA differently on genes at different loci in the nucleus. An array of post-translational modifications have been identified on CPA factors, such as phosphorylation of the CTD of RNAPII and the arginine/serine-like domains of CFI68 and CFI59 (ref. 153), ubiquitylation of PCF11 (ref. 232) and CPSF73 (refs. 233,234), PARylation of PAP $\alpha$ <sup>235</sup> and lactylation of CFI25 (ref. 236). These modifications may regulate APA through specific signalling events.



# Review article



**Fig. 2 | Dynamics of alternative polyadenylation.** **a**, The global trend in expression of alternative polyadenylation (APA) isoforms – use of proximal polyadenylation site (pPAS) versus distal PAS (dPAS), which leads to shortening or lengthening of the 3' untranslated region (3'UTR) and activation or suppression of intronic polyadenylation (IPA), respectively – is associated with cell identity as well as changes in cell proliferation and differentiation state. Global APA profile changes are also responsive to some cell conditions, such as metabolic status (activation of mTOR), circadian rhythm, cellular senescence, various types of cellular stress as well as activation of cell signalling, such as in immune cells and neurons. **b**, General inhibition of cleavage and polyadenylation (CPA) – owing to loss-of-function genetic mutations or inhibition of CPA factors, or various forms of cell stress – can lead to APA isoform changes as well as transcriptional read-through, which results in the expression of downstream of gene (DoG) transcripts. Transcriptional read-through can cause transcriptional interference of downstream genes, especially in regions of the genome with high gene density, including head-on collision of RNA polymerase II (RNAPII) and production of chimeric transcripts from

multiple genes (read-in transcription). In addition, transcriptional read-through can lead to transcription–replication conflicts whereby RNAPII collides with the replisome, which leads to DNA damage and genome instability. As such, inadequate CPA induces replication stress in highly proliferative cells. **c**, Gain-of-function and loss-of-function mutations of PAS, that increase or decrease CPA activity, respectively, have been associated with human diseases through effects on gene expression level (prominent examples are shown; see Supplementary Table 2 for a more comprehensive list of these associations). The effect of loss-of-function mutations can be more apparent when the transcriptional activity of the affected gene is high. **d**, Three examples of genetic variants affecting the expression of APA isoforms (apaQTLs) for three genes (*IRF5*, *GIMAP5* and *TNFSF13B*) involved in autoimmunity. Genetic variants are shown in bold. Alleles that weaken or strengthen the pPAS give rise to a longer 3'UTR (designated 'L' alleles) or a shorter 3'UTR (designated 'S' alleles), respectively. Long 3'UTR isoforms have an increased rate of mRNA decay compared with short 3'UTR isoforms, which results in different protein levels. Asterisk indicates the autoimmune disease risk allele.

In this Review, we discuss the dynamics of APA in various physiological and pathological conditions, the human mutations and genetic variants that are associated with APA isoform changes, and our current understanding of the functional effects of APA and molecular mechanisms of its regulation. We also cover high-throughput methods and machine learning-based approaches to study APA. We focus on studies in metazoans, mostly humans. Readers are referred to other recent reviews for more in-depth coverage of the mechanism of CPA<sup>6,7</sup> or of gene regulation by APA in plants<sup>8</sup>.

## APA dynamics in physiological conditions

The widespread variability of APA isoforms was first revealed in the late 1990s by expressed sequence tag data<sup>9</sup>, which helped uncover distinct global trends of APA isoform expression in different human tissues, such as biased expression of distal PAS (dPAS; further from the 5' end of the gene) isoforms in neuronal tissues and of proximal PAS (pPAS; closer to the 5' end of the gene) isoforms in blood<sup>10</sup>. Subsequently, data from microarrays and serial analysis of gene expression provided evidence of APA isoform variations in cell proliferation, differentiation and development. Of note, it was reported that proliferating cells tend to have shortened mRNA 3'UTRs, as exemplified by activated T cells<sup>11</sup>, whereas progressive 3'UTR lengthening was shown to occur in mouse embryonic development and during differentiation of the myoblast C2C12 cell line<sup>12</sup>. Concurrently, general 3'UTR shortening was reported in cancer cells<sup>13</sup>. These early studies established APA as a prominent layer of transcriptomic regulation in different tissues and cell contexts (Fig. 2a). The advent of 3'-end RNA sequencing (RNA-seq) in the early 2010s enabled genome-wide profiling of APA isoforms with greater precision<sup>14–16</sup>. In addition, standard RNA-seq data have also been used to interrogate APA isoform expression when coupled with bioinformatics methods to demarcate PASs<sup>17</sup>. In the past few years, there has been a marked increase in the use of single-cell RNA sequencing (scRNA-seq) data to examine APA, owing to the fortuitous fact that most scRNA-seq methods give rise to reads biased to the 3' end of RNAs (Box 2).

## Global trends in cell growth and differentiation

In line with the global 3'UTR lengthening in cell differentiation and development that was reported in early studies using bulk samples<sup>12</sup>, analysis of scRNA-seq data for approximately two million nuclei from mouse embryonic days 9.5 to 13.5 identified a progressive 3'UTR

lengthening trend in all cell types across embryonic stages<sup>18</sup>. Nevertheless, the degree of APA regulation varies in different cell types, with neurogenesis lineages having more prominent APA than others<sup>18</sup>. This is in agreement with an early study in which APA events in myogenesis were compared with those in neurogenesis using bulk samples<sup>19</sup>. It was found that these two lineages share similar APA events but that the global shift in pPAS-to-dPAS use (resulting in 3'UTR lengthening) is executed to a much greater extent in neurogenesis than myogenesis<sup>19</sup>. One interpretation of these results is that there exists a common APA regulatory mechanism in cell differentiation that leads to transcript lengthening, and that neurogenesis has additional mechanism(s) that augment the common APA regulatory programme (such as Hu proteins, see 'RNA-binding proteins'). Notably, the neurogenesis-associated APA regulatory scheme is well conserved across species, as shown in fly<sup>20,21</sup> and worm<sup>22</sup>.

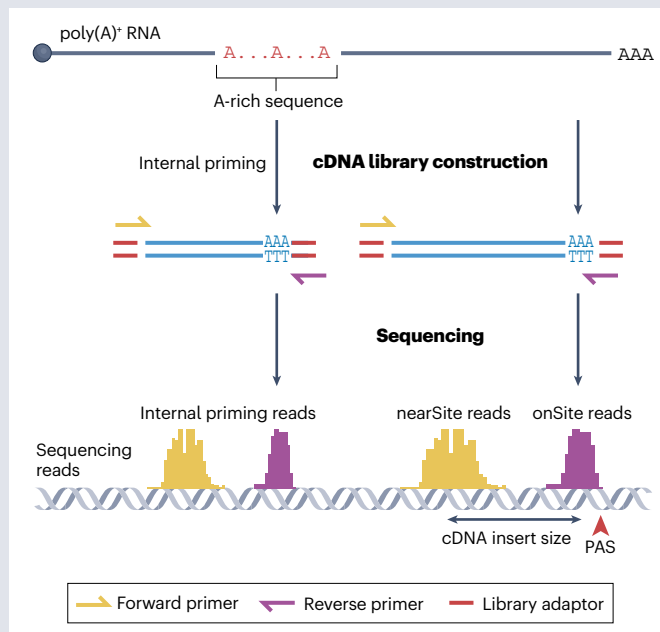
In another single cell-based study of pre-implantation development, it was shown that whereas there is continuous 3'UTR lengthening from zygote to morula in mice, human cells exhibit initial lengthening of 3'UTRs (from zygote to four-cell stage) followed by 3'UTR shortening<sup>23</sup>. However, similar pathways, such as RNA processing, are enriched for APA-regulated genes in both species, indicating that APA has conserved biological consequences. The global trend of 3'UTR lengthening during cell differentiation is likely to be related to the 3'UTR shortening that occurs during cell proliferation<sup>11</sup>, as differentiated cells are generally quiescent. In line with this, several cell differentiation lineages that involve increased proliferation have a global trend of 3'UTR shortening, including the activation of haematopoietic stem cells to differentiated multipotent progenitors<sup>24</sup> and the differentiation of lymphoid cells<sup>25</sup>.

A key question related to changes in 3'UTR length during proliferation and differentiation concerns the extent to which APA regulation can be attributed to certain phases of the cell cycle. Using scRNA-seq to examine APA events at different stages of the cell cycle – including G1/S, S, G2/M, M and M/G1, in mouse embryonic fibroblasts and two human cell lines<sup>26</sup> – it was found that, during the cell cycle, genes that are involved in cell cycle control have greater APA regulation than other genes and, importantly, that changes in APA isoforms are more prominent than changes in mRNA abundance<sup>26</sup>. These findings highlight a key role of APA in cell cycle regulation. However, it remains unclear to what extent cell proliferation-associated 3'UTR shortening is governed by cell cycle phases. It is also worth noting that although

## Box 2 | High-throughput approaches to study alternative polyadenylation

In the past few years, various high-throughput techniques, such as 3'-end RNA sequencing (RNA-seq), genome-wide CRISPR screens and massively parallel reporter assays, have enriched our understanding of aspects of alternative polyadenylation (APA). 3'-end RNA-seq methods generate reads either close to the polyadenylation site (PAS), termed 'nearSite' reads, or covering the PAS, termed 'onSite' reads (see the figure). One caveat, however, is that a sizable fraction of 3'-end reads can be derived from internal stretches of A-rich sequence, at which an oligo(dT) primer initiates reverse transcription for cDNA synthesis<sup>237</sup>. The template switch step in cDNA synthesis can also generate similar artefacts<sup>238</sup>. These issues can lead to false identification of a PAS, which need to be addressed by matching putative PAS to annotated PAS databases that either rigorously control for PAS quality by using motif information<sup>239</sup> or contain PAS not derived from oligo(dT) priming methods<sup>240</sup>. In addition, data from direct RNA-seq that contain the polyadenosine (poly(A)) tail information can be of use for data cleaning<sup>241</sup>.

Data from most single-cell RNA sequencing (scRNA-seq) methods, such as the 3' 10× Genomics protocol, are 3'-end reads and hence can be used to generate APA profiles at the single-cell level<sup>242</sup>. scRNA-seq coupled with gene perturbation by CRISPR interference (also known as Perturb-seq) has shed light on the distinct mechanisms by which cleavage and polyadenylation (CPA) factors regulate APA<sup>242,243</sup>. Machine-learning strategies have been used to examine PAS motifs by using massively parallel reporter assays together with 3'-end sequencing<sup>223,224</sup> and to uncover APA regulatory rules with respect to PAS motifs and the splicing context by using APA isoform expression data aggregated from different conditions<sup>151,220</sup>. Current deep learning



methods, such as deep neural networks, empower the construction of interpretable models from large volumes of high-throughput data. The resulting models have been used to assess the relevance of disease-associated genetic mutations and gene variants in APA regulation.

most studies have focused on 3'UTR-APA, IPA is also globally regulated and follows the same switch in pPAS-to-dPAS usage during embryonic development and cell differentiation<sup>16</sup>, indicating that both types of APA are regulated by similar mechanisms.

### Exceptions to general trends

The general APA trends associated with cell proliferation (3'UTR shortening) and differentiation (3'UTR lengthening) have several exceptions. Differentiation of spermatogonial stem cells to mature sperm cells involves stage-wise 3'UTR shortening and IPA activation<sup>27–29</sup>. This APA scheme, which culminates in spermatids after meiosis, seems to be coordinated with chromatin remodelling (the relaxation of histone proteins and their replacement with spermidine), as well as increased nonsense-mediated decay (NMD) activity that degrades mRNAs with long 3'UTRs<sup>30,31</sup>. As a result, mRNAs with shorter 3'UTRs can better escape from the large-scale culling of mRNAs that takes place during spermatogenesis, making them available for translation and protein production at later stages of sperm development when transcription ceases to operate. However, the differentiation of oogonia to oocytes (oogenesis), which also involves meiosis, does not have a similar global dPAS-to-pPAS shift<sup>28</sup>, which indicates that meiosis per se is not likely to be the cause of the unique APA programme in spermatogenesis.

Another exception to general APA trends is the differentiation of professional secretory cells. It was found that differentiation of human

embryonic stem cells to trophoblast-like cells in vitro elicits global 3'UTR shortening<sup>32</sup>. Further analysis of scRNA-seq data from placenta, which is composed of several trophoblast types, revealed that the differentiation of syncytiotrophoblast, a trophoblast subtype specialized in hormone production and secretion during pregnancy, involves a global dPAS-to-pPAS shift that is coupled with the expression of genes in the protein secretion pathway. This secretion-coupled mechanism of APA can also explain APA-mediated 3'UTR shortening during the differentiation of B cells to antibody-secreting plasma cells<sup>33</sup>, as well as during T cell activation<sup>11</sup>. However, the latter process has a concomitant increase in cell proliferation, which makes it difficult to separate APA mechanisms coupled with secretion from those connected with proliferation.

### Regulation by cell metabolism

As part of the transcriptomic programme, APA is highly dynamic upon changes to the metabolic state of cells. For example, activation of the mTOR pathway, which regulates cell metabolism in response to nutrient availability<sup>34</sup>, causes general 3'UTR shortening in multiple human and mouse cell types<sup>35</sup>. Interestingly, certain genes – such as those involved in protein processing in the endoplasmic reticulum and ubiquitin-mediated proteolysis – seem to be particularly affected by mTOR-mediated 3'UTR shortening<sup>35,36</sup>. Because mTOR pathway activation is an integral part of cell proliferation and growth, it is tempting to speculate that mTOR activation might have an important role

in mediating the global APA changes that occur in these processes. This could also explain the general 3'UTR shortening that occurs in cardiac hypertrophy<sup>37</sup>, where cells grow in size but not in number. The catabolic process of autophagy<sup>38</sup>, which involves inhibition of mTOR under conditions of nutrient starvation, was found to be augmented by perturbations of multiple CPA factors in a *Drosophila* model<sup>39</sup>, further highlighting the cross-talk between cell metabolism and APA regulation.

## Cellular response to stress

Stress conditions lead to global APA changes with variable consequences. Some types of stress cause 3'UTR lengthening, such as the ribotoxic stress elicited by anisomycin, which inhibits the peptidyl transferase activity of the ribosome<sup>40</sup>. However, some other stress conditions, especially those with genotoxic effects, have been reported to cause global 3'UTR shortening, including UV damage<sup>41,42</sup>, heat shock<sup>43,44</sup> and oxidative stress<sup>45</sup>. In particular, 3'UTRs markedly shorten during recovery from arsenic stress<sup>45</sup>, involving both the selection of pPAS and the decay of long 3'UTR isoforms<sup>45</sup>. Of note, some stress conditions such as heat shock, osmotic stress and oxidative stress can also induce transcriptional read-through (Fig. 2b) – when RNA polymerase continues transcribing past the normal termination sites – leading to the expression of downstream of gene (DoG) transcripts<sup>46,47</sup>. It is notable that when transcriptional read-through is caused by suppressed CPA – a failure to cleave the pre-mRNA at its usual PAS – there is also 3'UTR lengthening through APA, as shown in cells treated with JTE-607, a small-molecule inhibitor of the CPA endonuclease CPSF73 (also known as CPSF3)<sup>48,49</sup> (Box 1). By contrast, transcriptional read-through caused by defective termination per se does not seem to elicit substantial 3'UTR lengthening, as shown in cells with knockdown of the transcription termination factor *XRN2* (ref. 50). Notably, because some DoG RNAs are poly(A)<sup>+</sup> owing to the use of cryptic PAS in intergenic regions (Fig. 2b), poly(A)<sup>+</sup> DoG RNA expression can manifest as an extreme case of APA, where PAS selection is beyond the normal gene boundary.

## APA in pathological contexts

Dysregulated APA events are increasingly associated with human disease. Some pathological conditions involve global changes in APA profile related to cell identity or proliferation and/or differentiation state, as in cancer cells; others are associated with a large number of altered APA events owing to dysregulated core CPA factors or regulators, as in infection and some neurological conditions. In addition, mutations of CPA factors and APA regulators have been found to cause many human pathologies (Supplementary Table 1). Moreover, a growing number of mutations and genetic variants associated with PAS have been shown to alter APA with implications for pathological phenotypes and disease risks (Supplementary Tables 2 and 3).

## Inflammation and infection

APA has long been known for its role in adaptive immunity, including B cell differentiation and T cell activation<sup>11,33</sup>; its dynamics are also increasingly appreciated in innate immunity. Human macrophages display 3'UTR shortening after bacterial infection<sup>51</sup>, and scRNA-seq analysis of peripheral blood mononuclear cells from individuals with COVID-19 and healthy controls indicates that there is general 3'UTR shortening in all patient cell types, including monocytes and natural killer cells<sup>52</sup>. In accordance with these observations, ablations of CPA factor CFI68 (also known as CPSF6) (Box 1) in a human lung cell line and

a mouse fibroblast cell line, which led to pre-emptive 3'UTR shortening, enhanced the antiviral capacity of these cells against vesicular stomatitis virus. The authors attributed this phenomenon to the stabilization of key mRNAs encoding antiviral proteins through 3'UTR shortening<sup>53</sup>. Along similar lines, the CPSF73 inhibitor JTE-607, which generally suppresses pPAS use, has an anti-inflammatory effect<sup>54</sup>. This was shown to result from the suppressed expression of mRNAs encoding inflammatory cytokines<sup>55</sup>, which supports the notion that proper CPA activity is crucial for the rapid increase of cytokine gene expression during inflammation.

Several viruses inhibit CPA through interactions between the CPA machinery and viral proteins, such as the NS1 protein of influenza A virus<sup>56</sup> and the ICP27 protein of herpes simplex virus (HSV)<sup>57</sup>. CPA inhibition in cells infected by these viruses results in APA changes and increased expression of DoG transcripts<sup>58,59</sup>. Interestingly, it was found that HSV infection can activate both IPA (involving early transcription termination) and DoG RNA expression (involving transcriptional read-through) simultaneously in the same genes<sup>59,60</sup>, which suggests that the connection between PAS selection and transcription termination may depend on how much CPA activity is reduced. Therefore, although the virus-induced inhibition of CPA is thought to exert general suppression of host gene expression, it is tempting to speculate that mitigation of innate immunity or inflammation through CPA inhibition of specific antiviral genes may be a survival strategy used by viruses.

## Cancer

In general, cancer cells express transcripts with shortened 3'UTRs, which could be attributed plausibly to their high proliferation rate<sup>13,17</sup>. Importantly, the APA profile provides additional information for cancer prognosis beyond that obtained by measurement of gene expression levels<sup>17</sup>. As is the case for different cell differentiation lineages, the extent of APA varies between cancer types<sup>13,17</sup>. For example, kidney renal clear cell carcinoma typically has less 3'UTR shortening than some other cancer types, such as lung squamous cell carcinoma<sup>17</sup>. Notably, kidney renal clear cell carcinoma cells were also found to have increased expression of DoG transcripts<sup>61</sup>. Importantly, those individuals with kidney renal clear cell carcinoma who display shortened 3'UTRs have worse prognosis<sup>62</sup>. In some other cancer types, such as skin cutaneous melanoma, poor prognosis is associated with 3'UTR lengthening<sup>62</sup>, which indicates that the relationship between APA regulation and oncogenic development or severity is likely cancer type-specific. Notably, a recent study found that DoG transcript abundance is associated with poor patient survival for breast, colon and liver cancers<sup>63</sup>. Since DoG expression could happen together with 3'UTR lengthening owing to decreased CPA activity, it would be interesting to examine whether APA profiles could provide additional power to survival analyses for patients with these cancers. Along these lines, it is worth noting that inefficient CPA was found to induce replication stress<sup>64</sup>, such as malfunction of replication origins and transcription–replication conflicts (Fig. 2b), which results in DNA damage and genome instability that can further contribute to cancer development.

In keeping with the generally high expression of IPA isoforms in blood cells, IPA dysregulation has been reported in several types of leukaemia and lymphoma<sup>10</sup>. Widespread IPA activation in chronic lymphocytic leukaemia results in the expression of truncated proteins lacking C-terminal functional domains<sup>65</sup>. Importantly, owing to their relatively large size, some tumour-suppressor genes are particularly

## Box 3 | Genetic variants associated with alternative polyadenylation

The co-localization of apaQTLs (alternative polyadenylation (APA) isoform quantitative trait loci) with genetic variants that are identified by genome-wide association studies as being associated with clinical features — such as physiological traits, disease risk and pathological conditions — can provide insights into the aetiology of pathology and shed light on potential therapeutics (Supplementary Table 3). Furthermore, co-localization of apaQTLs with QTLs for other molecular phenotypes — such as mRNA and protein expression levels, alternative splicing and DNA methylation — can help identify molecular mechanisms that regulate APA or by which APA has functional effects. For example, by cross-referencing 3' untranslated region (3'UTR) apaQTLs with gene expression QTLs (eQTLs) and protein-level QTLs (pQTLs) from 46 tissue types, it was found that 78.5% of pQTLs match 3'UTR apaQTLs but not eQTLs, which suggests that many 3'UTR-APA events influence protein expression levels without affecting mRNA abundances<sup>79,80</sup>. In addition, a recent study focusing on rare genetic variants in non-coding regions identified 658 apaQTLs, of which ~18% are associated with gene expression changes, ~9% with splicing changes and ~3% with both<sup>244</sup>, indicating that there are complex relationships between APA, splicing and overall mRNA expression levels. Another study found that the co-localization of apaQTLs with eQTLs varies across cell conditions, suggesting that there are cell type-specific effects of APA on gene expression<sup>79</sup>. Moreover, a recent study of the co-localization of DNA methylation QTLs with apaQTLs reveals extensive involvement of DNA methylation in APA regulation across human cancers, particularly in low-grade glioma<sup>245</sup>. Therefore, co-localization of apaQTLs with other clinical or molecular features associated with the same genetic variants could shed much light on the biological relevance and regulatory mechanisms of APA. Although it has not been fully examined, this methodology could presumably be applied to genes that have a sole or one predominant polyadenylation site (PAS) to reveal the importance of CPA activity for their gene expression.

affected by this truncation mechanism in chronic lymphocytic leukaemia cells. By contrast, in multiple myeloma, a cancer that develops from antibody-secreting plasma cells, loss of IPA isoform expression predicts shorter progression-free survival<sup>66</sup>. Therefore, as with the solid tumours discussed above, the contribution of APA dysregulation to oncogenesis and cancer progression in haematopoietic cells appears highly dependent on the cancer type.

### Neurological disorders

The highly prominent preference for using dPAS over pPAS in neurons, resulting in both longer 3'UTRs and the selection of more-distal terminal exons, underscores the importance of expression of long transcripts for neuronal functions<sup>57</sup>. Consistently, dysregulated APA is emerging as a hallmark of neurodevelopmental disorders and neurodegeneration. Germline mutations in genes encoding several core CPA factors and transcription termination factors lead to neurodevelopmental defects (Supplementary Table 1). In addition, APA isoforms have been identified as a transcriptomic signature in individuals with amyotrophic

lateral sclerosis (ALS) and frontotemporal dementia (FTD)<sup>68,69</sup>. A case in point is the activation of an IPA isoform of *STMN2* (encoding the neuronal growth-associated factor stathmin 2) owing to nuclear loss of the RBP TDP-43 (encoded by *TARDBP*) in individuals with ALS or FTD who have *TARDBP* mutations<sup>70</sup>. Importantly, this IPA dysregulation event also takes place in individuals with familial or sporadic ALS or FTD who do not have *TARDBP* mutations, indicating that it is a convergent pathological event that drives neurodegeneration<sup>70</sup>.

### Disease-causing genetic mutations

PAS mutations have long been found to cause human genetic diseases<sup>71</sup> (Supplementary Table 2). For example, an A-to-G mutation that changes the strong PAS hexamer AATAAA in the *HBA2* gene (encoding haemoglobin A2) to the weaker hexamer AATAAG in individuals with  $\alpha$ -thalassaemia was the first example of a CPA-suppressing (loss-of-function) mutation to be reported<sup>72</sup>. Conversely, a G-to-A mutation at the cleavage site of the *F2* gene (encoding coagulation factor II, also known as prothrombin), which contributes to thrombophilia, was the first example of a CPA-enhancing (gain-of-function) mutation<sup>73</sup>. In these scenarios, CPA suppression or enhancement results in downregulation or upregulation of mRNA expression levels, respectively, which indicates that, in the cells that manifest these disease phenotypes, CPA might be a rate-limiting step for mRNA expression of affected genes<sup>74</sup> (Fig. 2c). In keeping with this, it is worth noting that somatic mutations of PAS hexamer are frequently associated with downregulated expression of tumour-suppressor genes in cancer<sup>75</sup>.

Whereas genes containing only one PAS may be more susceptible to PAS mutations because of the singular opportunity for CPA, PAS mutations in genes that undergo APA have also been found to alter gene expression (Supplementary Table 2). In accord with the notion that short 3'UTR mRNA isoforms are typically more stable than long 3'UTR isoforms<sup>5</sup>, mutations that create a PAS hexamer AATAAA in the 3'UTR of *CCND1* (encoding cyclin D1) result in 3'UTR shortening, and hence increased mRNA stability and protein production, leading to an increased proliferation rate of mantle-cell lymphoma cells and shorter survival time for patients carrying mutations<sup>76</sup>. Conversely, mutations that weaken a pPAS in the 3'UTR of *NAA10* lead to 3'UTR lengthening and downregulation of gene expression, causing Lenz microphthalmia syndrome<sup>77</sup>. In agreement with the general trend that IPA isoforms are unstable, a mutation changing AACAAA to AATAAA in intron 8 of *ERCC4* (encoding a subunit of the ERCC1-ERCC4 nucleotide excision repair endonuclease) activates IPA isoform expression, and hence reduces *ERCC4* expression, in a rare form of xeroderma pigmentosum<sup>78</sup>.

### Genetic variants associated with disease risk

In contrast to the small number of diseases caused by APA-altering mutations, a large number of genetic variants, mostly single nucleotide polymorphisms (SNPs), have been associated with APA isoform changes in several recent studies (Box 3 and summarized in Supplementary Table 3). For simplicity, here we use apaQTL (APA isoform quantitative trait locus) to refer to a genetic variant that is associated with a change in APA isoform expression. As expected, apaQTLs related to 3'UTR-APA are mostly enriched near the 3' end of the gene<sup>79,80</sup>. However, apaQTLs in other regions of the gene are also common<sup>79</sup>, in line with the notion that APA can be regulated by diverse mechanisms (see below).

It was estimated that 16–19% of the reported apaQTLs co-localize with genetic variants that are significantly associated with clinical phenotypes as identified by genome-wide association studies<sup>79,80</sup>. A few apaQTLs are of particular note because of their validated significance

for association with risk of the autoimmune disease systemic lupus erythematosus (SLE) (Fig. 2d). An apaQTL (rs10954213) in the 3'UTR of the transcription factor *IRF5* is associated with pPAS strength<sup>81</sup>, with the 'A' allele leading to a strong PAS hexamer (AATAAA) and the 'G' allele to a much weaker PAS hexamer (AATGAA). As such, the 'A' allele-containing *IRF5* gene preferentially expresses a shortened 3'UTR isoform that is more stable and produces a greater amount of IRF5 protein, which is a risk factor for SLE<sup>81</sup>. In a similar manner, an apaQTL (rs6598) in the 3'UTR of *GIMAP5* creates either an AATAAA ('A' allele) or AATAGA ('G' allele) PAS hexamer<sup>82</sup>, resulting in a stable, short 3'UTR isoform or an unstable, long 3'UTR isoform, respectively. Because the expression level of *GIMAP5* negatively correlates with SLE risk, the 'G' allele, which generates a less stable, long 3'UTR isoform, is associated with increased disease risk. Moreover, two genetic variants of *TNFSF13B* (which encodes B cell survival and activation factor (BAFF)) – one involving deletion and the other an SNP – together create an AATAAA hexamer, leading to 3'UTR shortening and increased mRNA expression owing to removal of a microRNA (miRNA) target site<sup>83</sup>. Increased expression of BAFF increases the risk of SLE, as well as multiple sclerosis.

## Functional effects of 3'UTR-APA

APA has diverse effects on gene expression; here, we discuss the general principles and highlight some recent findings relating to the functional roles of 3'UTR-APA. Readers can find additional information on this subject in earlier reviews<sup>84–86</sup>.

By definition, 3'UTR-APA changes the size and content of the 3'UTR of an mRNA (Fig. 3a). For simplicity, the 3'UTR sequence that is subject to regulation by APA (between the first and last PAS) is named the alternative 3'UTR (aUTR), and the sequence that is common to all isoforms is named the common 3'UTR (cUTR) (Fig. 3a). The median cUTR and aUTR sizes in human genes are ~280 nucleotides and ~900 nucleotides, respectively<sup>5</sup>. As such, 3'UTR-APA can generate mRNA isoforms that differ in length by more than fourfold.

## mRNA subcellular localization

3'UTRs have long been known to influence the subcellular localization of mRNAs in polarized cells, such as neurons. For example, the short and long 3'UTR isoforms of *BDNF* (encoding brain-derived neurotrophic factor) are enriched in somata and dendrites, respectively<sup>87</sup>. In line with this, using micro-dissected rat hippocampal slices and 3'-end RNA-seq, it was shown that, in general, the neuropil area (a synaptically dense region containing a relatively low number of cell bodies) has mRNAs with longer 3'UTRs than those in somata<sup>88</sup>. However, the opposite trend – of long 3'UTR isoforms being more localized in somata – has also been reported for many genes<sup>88</sup>. This supports the notion that binding to cognate RBPs, such as Pumilio (in *Drosophila*) or PUM2 (in humans)<sup>89</sup>, as opposed to 3'UTR size per se, is the driving force for differential localization of 3'UTR isoforms<sup>90</sup>. In addition, the neuropil-enriched transcripts tend to have a higher GC content and are more likely to fold into structured RNAs, which implicates the binding of RBPs to structured RNAs in determining mRNA localization. Importantly, neural activity drives changes in 3'UTR isoform abundance between the two compartments, indicating that there is localized regulation of mRNA metabolism through APA isoforms<sup>88</sup>.

The functions of the aUTR in mRNA localization have also been studied in cells that are not overtly polarized. Using APEX2-based proximity labelling, it was shown that transcripts with short 3'UTRs are more likely to be associated with nuclear pores than those with long 3'UTRs<sup>91</sup>. Although this study does not compare 3'UTR isoforms

directly, it implicates 3'UTR size as a parameter for nuclear export kinetics. At the steady state, long 3'UTR isoforms generally have a higher nuclear-to-cytoplasmic abundance ratio than do short 3'UTR isoforms<sup>92,93</sup>, which suggests that aUTRs may hinder nuclear export. However, owing to enriched UGUA motifs, aUTRs are also more likely to bind the cleavage factor I (CFI) complex component CFI68 (refs. 93–95), which also functions as a nuclear export adaptor. Therefore, it remains an open question as to how 3'UTR size versus CFI68 binding impacts the nuclear export kinetics of different 3'UTR isoforms.

The proximity labelling study also indicates a 3'UTR-mediated scheme for mRNA localization to certain organelles, such as mitochondria<sup>91</sup>. Along the same lines, using cell fractionation and 3'-end sequencing, another study found that aUTRs generally facilitate translation-independent endoplasmic reticulum association of mRNAs<sup>96</sup>. The authors also found that 3'UTR size, GC content and RNA structure are key features that can predict the potential of a given mRNA for translation-independent endoplasmic reticulum association. Although the underlying mechanism(s) are yet to be determined, some RBPs have previously been found to be endoplasmic reticulum bound<sup>97</sup>, which could potentially facilitate aUTR-mediated endoplasmic reticulum association. One ramification of translation-independent endoplasmic reticulum association is its influence on where in the cytoplasm an mRNA is translated, and hence the location of its newly made protein. Similarly, by analysing mRNAs that are enriched in the endoplasmic reticulum, cytosol and TIS granules (mesh-like condensate structures associated with the endoplasmic reticulum), it was found that most mRNAs have biased subcellular localization dependent on their sequence or functional features<sup>98</sup>. For example, mRNAs encoding transcription factors tend to be enriched in TIS granules, and endoplasmic reticulum-enriched mRNAs generally encode large proteins with high expression levels<sup>98</sup>.

## mRNA stability and translation

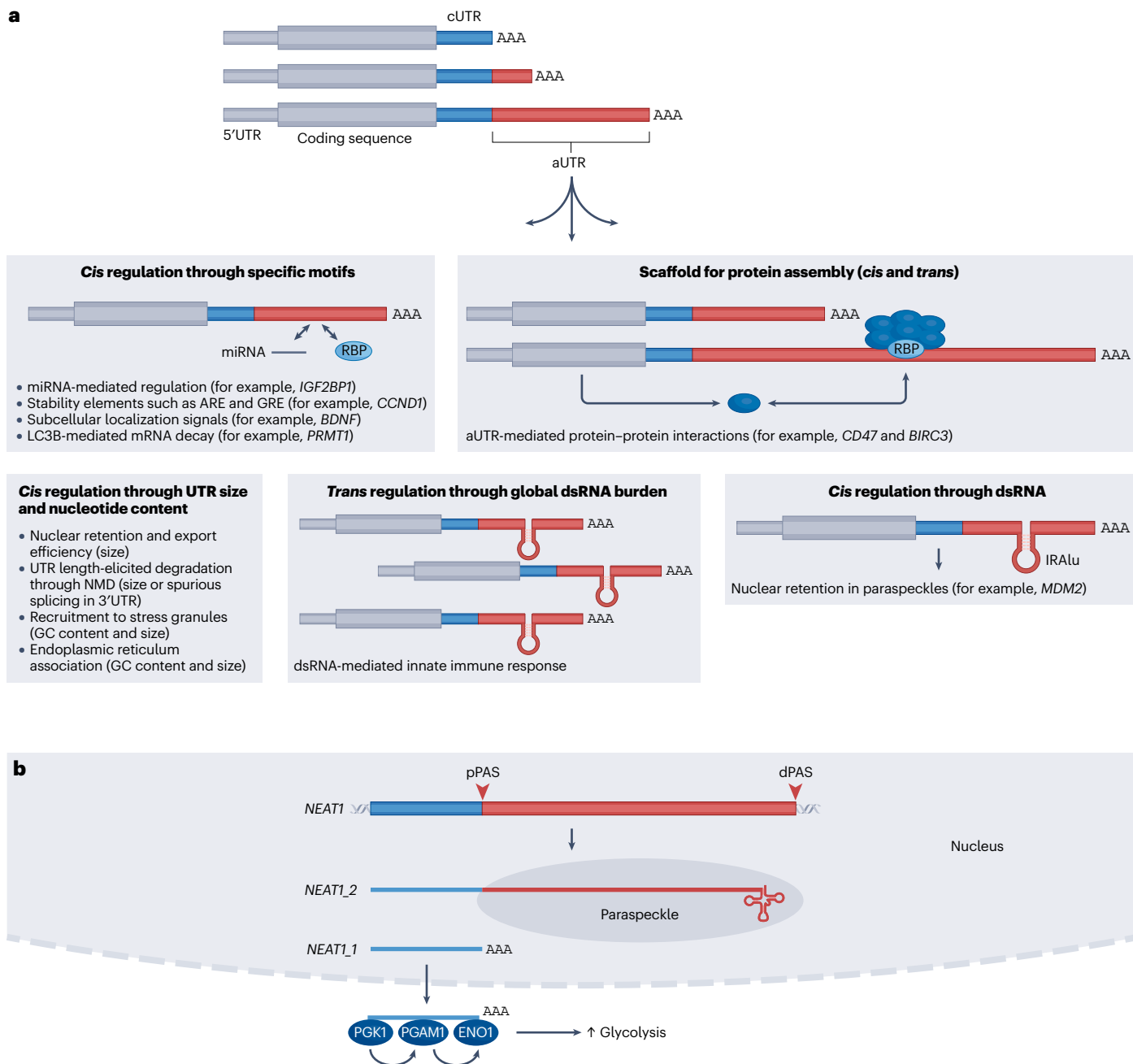
The 3'UTR contains many mRNA stability elements, such as AU-rich elements (AREs), GU-rich elements (GREs) and miRNA target sites<sup>99</sup>. Interestingly, the AAUAAA motif in the 3'UTR can also serve as a decay signal for LC3B-mediated mRNA decay, involving binding of mRNA by the lipidated autophagy regulator LC3B and subsequent mRNA deadenylation<sup>100</sup>. In addition, the size of the 3'UTR has a negative effect on mRNA stability through the NMD pathway, owing to either the length-dependent recruitment of UPF1 (ref. 101) or spurious splicing events in long 3'UTRs<sup>102</sup>. In several human cell lines<sup>103</sup>, it was reported that 3'UTRs in the newly made mRNA pool are globally longer than in the steady-state pool, which suggests that short 3'UTR isoforms are generally more stable than long 3'UTR isoforms. However, a less pronounced trend was reported in mouse NIH3T3 cells<sup>104</sup>. In addition, in activated T cells, which display global 3'UTR shortening, there are no discernible changes in mRNA levels that could be attributed to changes in 3'UTR lengths<sup>105</sup>. Therefore, the effects of aUTR on mRNA stability do not seem to be clear-cut, presumably owing to the variable milieu of mRNA-stabilizing and mRNA-destabilizing RBPs that determine the final outcomes of a specific 3'UTR isoform in different cells. A similarly complex relationship exists between 3'UTR isoforms and translation efficiency, as divergent results have been obtained in different cell types<sup>104–106</sup>. Of note, a recent analysis using a large number of ribosome profiling data reported a limited role of the 3'UTR in translation efficiency<sup>107</sup>.

In line with the cell type-specific influence on mRNA stability, it was found that long 3'UTR isoforms are relatively more stable in the

# Review article

neuronal cell line model SH-SY5Y than in the non-neuronal cell lines HEK293T and HepG2 (ref. 103). In keeping with this, mRNAs with greater stability have been shown to facilitate their localization to neurites<sup>108</sup>

and, in the neuropil compartment of the hippocampus, long 3'UTR isoforms are generally more stable than short 3'UTR isoforms<sup>88</sup>. In the latter study, the authors found that long 3'UTR isoforms for some



**Fig. 3 | Functional effects of 3' untranslated region alternative polyadenylation.**

**a**, 3' untranslated region (3'UTR) alternative polyadenylation (APA) isoforms have alternative 3'UTRs (aUTRs; shown in red), and hence different 3'UTR sizes. The region of 3'UTR common to all isoforms is known as the common 3'UTR (cUTR; shown in blue). aUTRs can contain double-stranded RNA (dsRNA) structures and various RNA regulatory motifs, which can affect mRNA metabolism in *cis* or *trans*. aUTR-mediated regulation of mRNA metabolism can be specific for particular RNA motifs or can be size based, as shown. Some example genes that are affected by these mechanisms are indicated. **b**, APA of the long non-coding RNA (lncRNA) *NEAT1* leads to two isoforms with different

subcellular localizations and functions. The long isoform *NEAT1\_2* (22.7 kb) is further processed at the 3' end to be a nuclear retained poly(A)<sup>-</sup> RNA that has an essential role in paraspeckle assembly. The poly(A)<sup>+</sup> short isoform *NEAT1\_1* (3.7 kb) can be exported into the cytoplasm and has a scaffolding function to hold together several glycolysis enzymes (PGK1, PGAM1 and ENO1), which increases the efficiency of glycolysis by substrate channelling. AAA, poly(A) tail; ARE, AU-rich element; dPAS, distal polyadenylation site; GRE, GU-rich element; IRAlu, inverted Alu; miRNA, microRNA; NMD, nonsense-mediated decay; pPAS, proximal polyadenylation site; RBP, RNA-binding protein.

genes can be shortened in a compartment-specific manner following neural activity, presumably through endonuclease cleavage or 3'-to-5' degradation. Notably, this post-transcriptional remodelling of APA isoforms was also reported in ageing brain, where translation-coupled mRNA decay was implicated as the mechanism<sup>109</sup>.

## Formation of double-stranded RNA

The Alu sequence is the major type of short interspersed nuclear element in the human genome; it accounts for ~11% of the entire genomic sequence and is often found in 3'UTRs. Two Alu sequences oriented in opposite directions, known as inverted Alu (IRAlu), can form an extended double-stranded RNA (dsRNA) structure. In most cells, IRAlu-containing mRNAs are retained in paraspeckles in the nucleus<sup>110</sup> – membraneless structures that are dynamically regulated under stress conditions – thereby preventing them from activating dsRNA-sensing mechanisms of the innate immunity system in the cytoplasm. It was recently shown that removal of IRAlu from the 3'UTR of *MDM2* mRNA through APA is crucial for its protein expression and for MDM2-mediated inhibition of tumour-suppressor protein p53 during tumorigenesis<sup>111</sup>. The same study also reported that longer 3'UTRs in neural progenitor cells cause downregulation of genes containing IRAlus in their aUTRs<sup>111</sup>. Notably, owing to globally increased expression of long 3'UTR isoforms, neuronal cells were found to have a higher dsRNA content overall, which predisposes them to dsRNA-elicited inflammation<sup>112</sup>.

## Formation of condensates

RNA–RBP condensates are membraneless structures formed by liquid–liquid phase separation<sup>113</sup>. Because of increased valency in RNA–RBP interactions, long 3'UTRs have greater potentials to form condensates<sup>114</sup>. For example, 3'UTR size is a key determinant of mRNA recruitment to stress granules, a type of condensate formed in the cytoplasm when mRNA translation is inhibited by stress<sup>115</sup>. On average, the shortest and longest 3'UTRs of a gene are comparable with the sizes of its 5'UTR (~280 nucleotides) and coding sequence (~1.2 kb), respectively<sup>5</sup>. Therefore, with respect to size, aUTRs could have a substantial impact on the formation of RNA–RBP condensates. On this note, it would be interesting to examine whether neurons, which have longer 3'UTRs than other cell types, are more prone to stress granule formation, which is a contributing factor in neurodegeneration<sup>116</sup>.

## RNA-mediated protein scaffolding

Some 3'UTRs can have a structural and scaffolding role in facilitating protein–protein interactions<sup>99</sup>. For example, it was shown that the aUTR of *CD47* facilitates the interaction between newly made CD47 protein and the adaptor protein SET, leading to increased plasma membrane localization of the 'don't eat me' signal through CD47 and better protection of cells from phagocytosis by macrophages<sup>117</sup>. Similarly, it was found that, despite similar mRNA localization of long and short 3'UTR isoforms of *BIRC3*, the resulting proteins bind distinct sets of partners, leading to divergent functions in the regulation of cell death (aUTR-independent) versus B cell migration (aUTR-dependent)<sup>118</sup>.

3'UTR-mediated protein–protein interactions are reminiscent of the interactions mediated by some lncRNAs. A case in point is the lncRNA gene *NEATI*, which generates two major APA isoforms: *NEATI\_1*, a 3.7 kb, short isoform; and *NEATI\_2*, a 22.7 kb, long isoform (Fig. 3b). *NEATI\_2* has a tRNA-like structure at its 3' end, resulting from cleavage by RNase P, and is essential for scaffolding paraspeckles in the

nucleus<sup>119,120</sup>. By contrast, *NEATI\_1* has a poly(A) tail, can be exported into the cytoplasm and is dispensable for paraspeckle formation. However, *NEATI\_1* can bind and hold together several glycolytic enzymes, including PGK1, PGAM1 and ENO1, thereby increasing the efficiency of glycolysis through substrate channelling across these enzymes<sup>121</sup>. As such, *NEATI\_1* has been implicated in promoting aerobic glycolysis in cancer cells<sup>121</sup>. It is yet to be seen whether the aUTRs of protein-coding genes can have similar *trans*-acting functions, independent of the protein encoded by the same mRNA.

## Functional effects of intronic APA

IPA isoforms have traditionally been less well studied than 3'UTR-APA isoforms, owing largely to their low expression levels and the prevalence of A-rich sequences in introns that can complicate the identification of intronic PAS (Box 2). However, with the advent of 3'-end sequencing methods and better annotation of IPA sites in the genome, understanding of the mechanisms of gene regulation by IPA, particularly its role in suppressing gene expression, has grown rapidly in recent years.

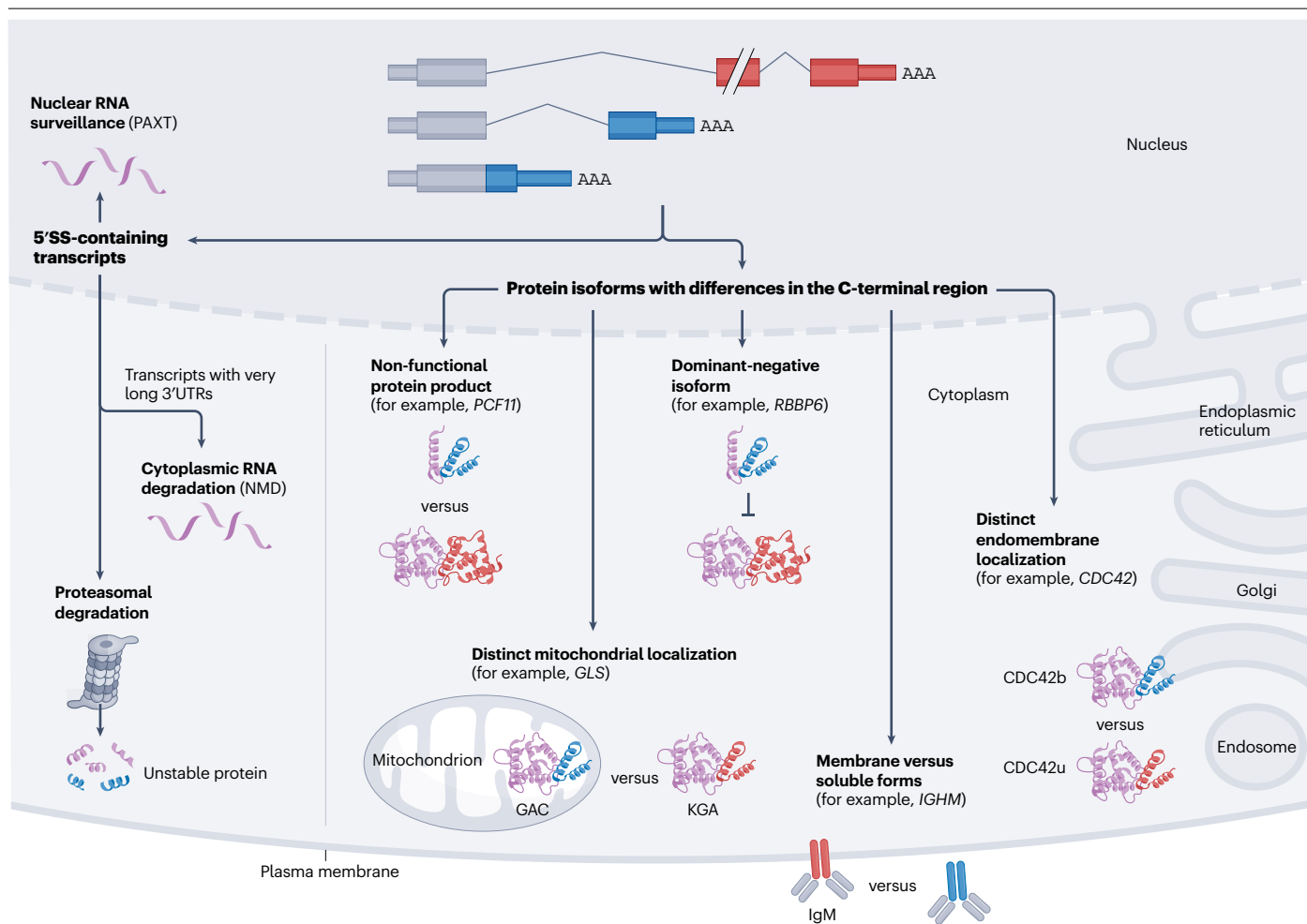
## Protein isoform regulation

By default, IPA changes the C-terminal region of protein products (Fig. 4). As such, one functional consequence is variation of protein localization that is based on the C-terminal region. For example, the *IGHM* gene (encoding the IgM antibody heavy chain)<sup>122,123</sup>, which was the first reported case of APA, produces a membrane-anchored protein isoform (CPA occurring in the last exon) in B cells and a secreted, soluble isoform (IPA isoform) in plasma cells. It has been estimated that more than 400 human genes could have similar IPA-mediated regulation of membrane-bound versus soluble isoforms<sup>124</sup> (Fig. 4). Notably, splicing-suppressing antisense oligonucleotides were found to activate IPA, thereby switching transmembrane receptor tyrosine kinases (RTKs), such as vascular endothelial growth factor receptor 2 (encoded by *KDR*), to their secreted, soluble isoforms with antagonistic functions<sup>125</sup>. This approach holds promise as a therapeutic modality for RTK-related diseases, such as cancer.

IPA-induced changes in the C-terminal region have also been shown to alter the location and/or activities of intracellular proteins (Fig. 4). For example, the *CDC42* gene expresses a ubiquitous protein isoform CDC42u (using a PAS in the last exon) and a brain-specific isoform CDC42b (using an intronic PAS). Although these isoforms differ only in the last few amino acids at their C termini, it was found that in astrocytes and neural precursors CDC42u is associated with the plasma membrane at the leading edge of migrating cells, whereas CDC42b is localized mainly to intracellular membrane compartments where it has a function in endocytosis<sup>126</sup> (Fig. 4). Another prominent example is the gene *GLS* (encoding glutaminase); protein products of the IPA isoform of *GLS* (termed GAC) are localized to mitochondria whereas those encoded by isoforms using the last exon PAS (termed KGA) are mostly cytosolic<sup>127</sup> (Fig. 4). Importantly, GAC is enzymatically more active than KGA, which is crucial for cancer cells that preferentially express the IPA isoform and use glutamine as a fuel for growth<sup>128</sup>.

## Premature transcription termination

An emerging regulatory scheme involving IPA is the suppression of gene expression by premature transcription termination<sup>129</sup>. For example, IPA is globally activated under conditions of genotoxic stress, resulting in widespread premature transcription termination. This may function as an adaptive mechanism to avoid the transcription of damaged DNA in



**Fig. 4 | Functional effects of intronic polyadenylation.** The consequences of intronic polyadenylation (IPA) include generation of unstable transcripts owing to the presence of a 5' splice site (5'SS) and/or having a very long 3' untranslated region (3'UTR), production of non-functional or dominant-negative protein products, or production of protein isoforms with distinct localizations and/or functions owing to differences in the C-terminal region. Several examples

are shown. Transcripts containing the composite internal-terminal exon are particularly unstable owing to 5'SS-mediated nuclear RNA surveillance executed by the poly(A) tail exosome targeting (PAXT) complex and long 3'UTR-associated cytoplasmic RNA degradation caused by nonsense-mediated decay (NMD). AAA, poly(A) tail.

stressed cells, which could otherwise result in aberrant transcripts as well as genome instability. Notably, IPA isoforms containing a 5' splice site (5'SS) (composite internal-terminal exon (Fig. 1b)) are generally unstable<sup>103,130</sup> owing to U1 small nuclear ribonucleoprotein (snRNP)-mediated nuclear retention of RNA<sup>131</sup> and RNA degradation through nuclear surveillance by the poly(A) tail exosome targeting (PAXT) complex<sup>130</sup> (Fig. 4). In addition, IPA isoforms containing a composite internal-terminal exon often have a long 3'UTR that could make them susceptible to cytoplasmic degradation by NMD<sup>101,102</sup> and/or encode suboptimal C-terminal protein sequences that are subject to rapid proteasomal degradation<sup>132</sup> (Fig. 4).

Notably, genes encoding several core CPA factors harbour conserved intronic PAS, including *RBBP6* (ref. 133), *CSTF3* (refs. 134,135), *PCF11* (refs. 136,137), *PAPOLA*<sup>138</sup> and *WDR33* (ref. 139). Whereas some protein products of IPA isoforms are non-functional (for example, *PCF11*), some others have dominant-negative functions (for example, *RBBP6*) (Fig. 4). Presumably, these conserved IPA events equip CPA

factor genes with negative-feedback regulatory mechanisms to modulate their protein activities (often the full-length protein) and maintain the homeostasis of CPA activity in the cell.

## Mechanisms for the regulation of APA

A growing number of APA regulatory mechanisms have been discovered in recent years, thanks to various sequencing techniques that interrogate nascent and mature RNAs as well as high-throughput and machine-learning approaches to uncover underlying rules for different regulators (Box 2).

### Core CPA factors

Regulation of core CPA factors (Box 1) has been implicated in global APA changes in many physiological and pathological processes, such as cell proliferation (regulated by *CSTF64* (ref. 140)), spermatogenesis (*CSTF64* (ref. 141), *CFI25* and *CFI68* (ref. 142), and *PCF11* (ref. 143)), pre-implantation embryonic development (*CFI25* and *CFI68* (ref. 23)),

haematopoiesis (CFI25 (ref. 24)), neurogenesis (PCF11 (ref. 144)), somatic cell reprogramming (CFI25 (ref. 145)), renewal of embryonic stem cells (FIP1 (ref. 146)), inflammation (CFI25 and CFI68 (ref. 147)), oncogenesis (FIP1 (ref. 148)) and tumour suppression (CFI25 (ref. 149)). Conversely, germline mutations of several CPA factor genes have been associated with human diseases (Supplementary Table 1).

Most core CPA factors promote pPAS usage in the last exon<sup>150</sup>, such as FIP1, RBBP6, CSTF64 (encoded by *CSTF2*) and its paralogue CSTF64 $\tau$  (encoded by *CSTF2T*), and PCF11. Conversely, CPA inhibition by genetic knockdown of CPA factors or by the small-molecule CPSF73 inhibitor JTE-607 leads to a global pPAS-to-dPAS usage shift<sup>48,49</sup>. This regulatory mode is in line with the first-come-first-served model for PAS usage<sup>151</sup> and is in accord with the fact that a pPAS is typically weaker than a dPAS as defined by surrounding RNA motifs<sup>5</sup> (Fig. 5a), a configuration that would confer regulatability to the suboptimal pPAS as well as ensure proper transcription termination after a strong dPAS.

By contrast, ablations of CFI25 (also known as CPSF5; encoded by *NUDT21*) and CFI68, two components of the CFI complex – as well as the poly(A) tail-binding proteins PABPN1 and PABPC1 – lead to global dPAS-to-pPAS shifts<sup>150,152</sup> (Fig. 5a). The mechanisms by which PABPN1 and PABPC1 regulate APA are not fully resolved, even though PABPN1 was shown to inhibit pPAS usage<sup>152</sup>. Much mechanistic insight, however, has been learned about APA regulation mediated by CFI25 and CFI68 (refs. 150,153,154), which involves binding of CFI25 to the UGUA motifs that are enriched in the upstream region of dPAS<sup>150,153,154</sup> and recruitment of the CPA machinery by CFI68 through interaction with FIP1 (ref. 153) and/or the formation of nuclear condensates<sup>155</sup>. On this note, nuclear condensate formation has recently been shown for RBBP6, which may function as the effector of PAS usage through its function in activation of the CPA endonuclease CPSF73 (ref. 156).

Notably, the distance between dPAS and pPAS is an important determinant of APA regulation for both the first-come-first-served scheme and the CFI complex-mediated mechanism<sup>150,151</sup>. Given that more than half of CPA events are estimated to occur near nuclear speckles<sup>156</sup>, it would be interesting to examine how these two modes of APA regulation are executed for genes in different loci of the nucleus. On this note, it was found that increased dwell time in the nucleus could subject APA isoforms with long 3'UTRs to additional CPA, giving rise to short 3'UTR isoforms<sup>157,158</sup>. This mechanism, known as 'sequential CPA', is in line with the global trend of 3'UTR shortening in cells with suppressed expression of nuclear export factors, such as NXF1 (ref. 93), or export adaptor proteins SRSF3 and SRSF7 (ref. 94). Sequential CPA may function as an adaptive mechanism to facilitate localization of mRNAs to cytoplasm through 3'UTR shortening when nuclear export is attenuated, such as during cellular stress<sup>159</sup>.

RNAPII is generally considered to be an essential component of the CPA machinery (Box 1). Its C-terminal domain (CTD), which contains 52 repeats of the heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser in human cells, is subject to several types of post-translational modification<sup>160</sup>. In particular, phosphorylation at different positions of the heptapeptide is coordinated with different stages of the transcription cycle as well as with chromatin configuration. For example, the phosphorylation statuses of Ser5, Ser2 and Thr4 are closely connected, respectively, with transcription initiation, elongation and termination. As such, kinases and phosphatases of the RNAPII CTD, such as CDK9 (ref. 161), CDK12 and CDK13 (refs. 162–164), PP1 (refs. 161,165) and PP2A (ref. 166), have been reported to have substantial influences on PAS selection.

## Splicing

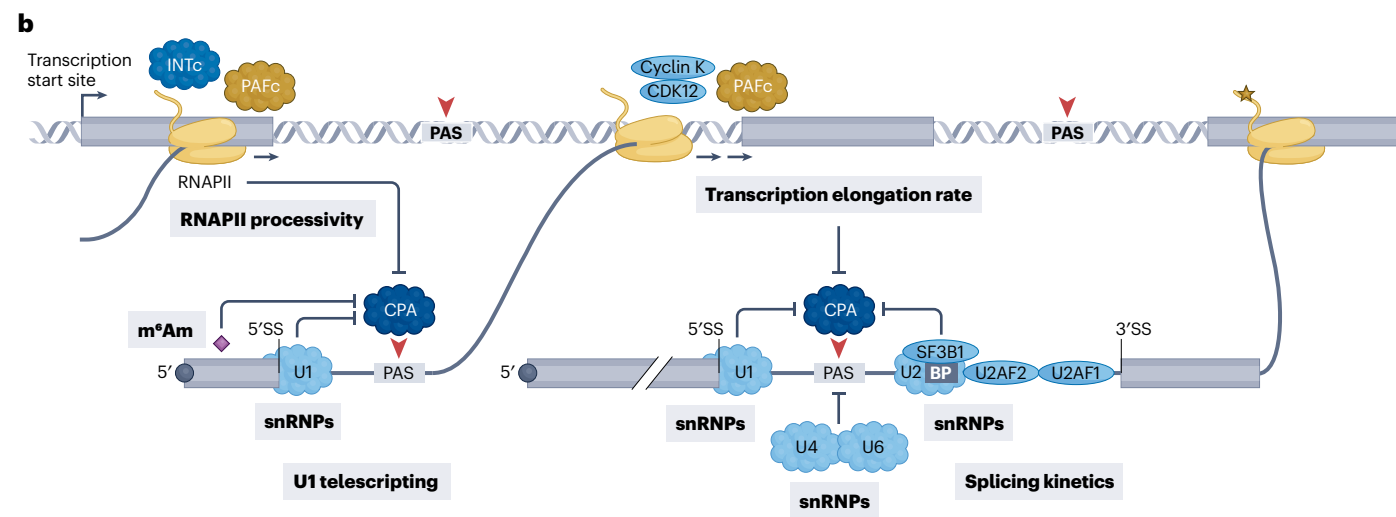
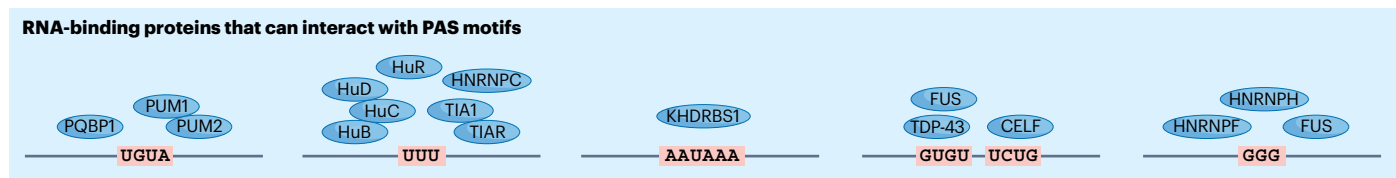
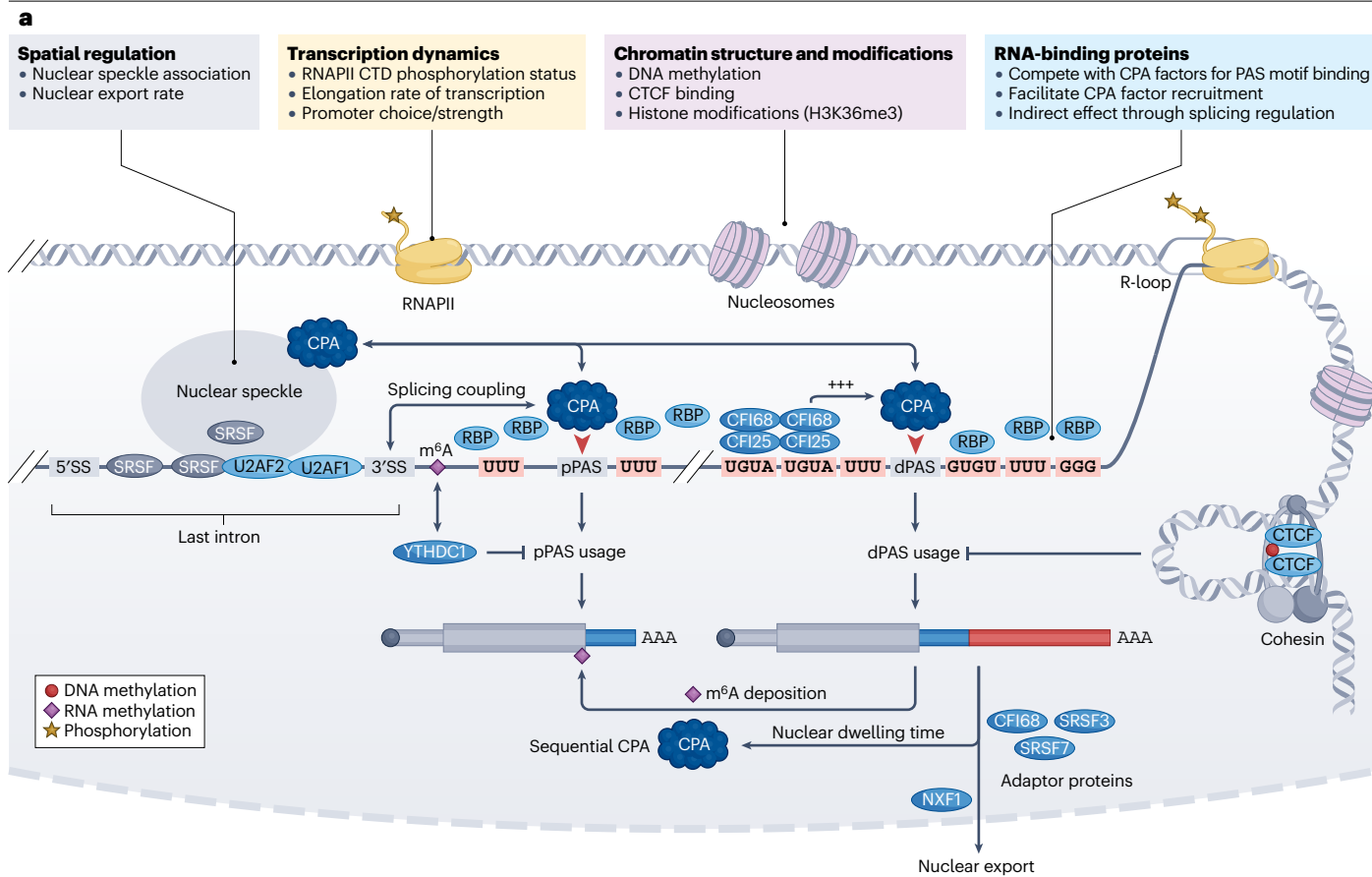
Owing to the cooperation between splicing of the last intron and CPA in the last exon<sup>151,167</sup>, removal of the penultimate intron tends to promote the usage of pPAS in the last exon (Fig. 5a). However, for intronic PAS, splicing is in a kinetic competition with CPA (Fig. 5b). As such, intron features that could reduce the splicing speed, such as weak splice sites and large intron size, facilitate IPA<sup>151</sup>. Consistently, knockdown of core splicing factors, such as SF3B1 and U2AF2 (ref. 150), or functional inhibition of splicing by antisense morpholinos, such as those targeting U1 (ref. 168), U4 (ref. 169) and U6 (ref. 170) snRNPs, all lead to marked IPA activation. The U1 snRNP-based suppression of CPA<sup>168</sup>, also known as U1 telescripting<sup>171</sup>, is of particular note. First, U1 is much more abundant in the cell than other splicing snRNPs, which supports the notion that it has additional roles beyond splicing<sup>168</sup>. Second, U1 telescripting shows a very strong IPA suppression effect at the 5' end of a gene, which gradually weakens towards the 3' end<sup>150,171</sup>. This polarity of IPA suppression is distinct from the IPA suppression by U2 or U6 snRNPs<sup>170</sup>. Third, in accordance with its bias to the 5' end of a gene, U1 snRNP has also been shown to control both promoter directionality<sup>172</sup> and alternative promoter selection<sup>173</sup>. Mechanistically, in addition to the inhibition of IPA through splicing, U1 snRNP increases transcriptional elongation in the intron, which accentuates kinetic suppression of IPA<sup>174</sup>, and exerts CPA inhibition through physical interactions with the CPA machinery<sup>175</sup>.

## Transcription dynamics

A growing body of evidence indicates that transcription initiation affects CPA activity and, hence, PAS choice. For example, in general, a strong promoter imparts a higher CPA activity on the pre-mRNA, resulting in preferential usage of pPAS<sup>176–178</sup>. Using long-read RNA-seq, it was found that promoter choice can be directly coupled with PAS selection in a tissue-specific manner and, mechanistically, involves histone acetylation<sup>179</sup>. More recently, it was shown that selection of a distal promoter (far from the gene body) tends to favour IPA events that result in the use of proximal terminal exons and, conversely, selection of a proximal promoter is more likely to be coupled with CPA in the 3'-most exon<sup>180</sup>. Mechanistically, the transcriptional elongation rate was found to be a contributing factor for this promoter-PAS coupling. Another recent study found that the Mediator protein complex, which has a key role in transcription initiation, interacts with the CPA factor FIP1 and, hence, regulates APA<sup>181</sup>. This finding is in line with the long-standing observation that CPA factors are present at both ends of the gene<sup>182</sup>. Therefore, the transcriptional elongation rate and CPA factor recruitment are two mechanisms that can connect the PAS choice to promoter strength and identity. In this context, it is also worth noting that transcription by immature RNAPII owing to suppression of the integrator complex (INTc) was found to activate IPA and hence premature transcription termination at the 5' end of genes<sup>183</sup>. Interestingly, the RNAs generated by this process contain dsRNA species that trigger the integrated stress response<sup>183</sup>.

In keeping with the kinetic nature of IPA regulation, the inhibition of transcription elongation factors, such as the polymerase-associated factor 1 (PAF1) complex and the cyclin-dependent kinase CDK12, activates IPA globally<sup>164,184</sup>. Of note, CDK12 mutations cause a 'BRCAness' phenotype that is typically associated with mutations of genes in the homologous recombination repair (HRR) pathway, such as *BRCA1* and *BRCA2*. Suppression of CDK12 activity preferentially activates IPA events in HRR genes and leads to their downregulation, thereby phenocopying their mutations<sup>164</sup>. Although it was shown that HRR genes tend to be long and hence more likely to harbour IPA events, it

# Review article



**Fig. 5 | Regulation of alternative polyadenylation.** **a**, Regulation of alternative polyadenylation (APA) in a terminal exon. Two alternative polyadenylation sites (PASs) are shown, namely the proximal PAS (pPAS) and the distal PAS (dPAS). The dPAS is typically stronger than the pPAS owing to differences in the surrounding PAS motifs, in particular the UGUA motifs upstream of the dPAS that recruit the CFI25–CFI68 complex. The major themes that influence the selection of an alternative PAS are shown, including spatial regulation, transcription dynamics, chromatin structure and modifications, and RNA-binding proteins (RBPs). RBPs that interact with core PAS motifs are shown underneath. The recruitment of cleavage and polyadenylation (CPA) factors to nuclear speckles is likely to regulate APA given that several CPA factors with APA functions localize to nuclear speckles, such as RBBP6, FIP1 and PCF11. Sequential APA is shown, which involves a mature, long APA isoform using a dPAS being subject to additional CPA post-transcriptionally in the nucleus, giving rise to a shorter isoform that is

normally derived directly from a pPAS. DNA methylation, RNA methylation and protein phosphorylation are shown. **b**, Regulation of APA in an intron (intronic polyadenylation (IPA)). Two intronic PAS are shown, with one being more proximal and the other more distal to the promoter. Five regulatory schemes are indicated, all of which function to inhibit IPA: maturation of RNA polymerase II (RNAPII) at the promoter by the integrator complex (INTc), which ensures transcriptional processivity; *N*<sup>6</sup>-methyladenosine (m<sup>6</sup>A) with 2'-O-methylation (m<sup>6</sup>Am) at the 5' end of pre-mRNA, which inhibits CPA by sequestering PCF11; U1 telescripting, which specifically inhibits IPA to prevent early transcription termination; the PAF complex (PAFc) and CDK12–cyclin K that control the elongation rate; and splicing kinetics modulated through U1, U2, U4 and U6 small nuclear ribonucleoproteins (snRNPs). BP, branch point; CTD, C-terminal domain; H3K36me3, histone H3 trimethylated at Lys36; 3'SS, 3' splice site; 5'SS, 5' splice site.

is not entirely clear why mutations of other elongation factors, which also control IPA in long genes, do not display the same BRCAness phenotype. One possibility is that the expression patterns of *CDK12* and *HRR* genes are generally correlated, leading to functional cooperation. In a similar vein, it is noteworthy that in *Drosophila* a mutant RNAPII with a slower elongation speed causes substantial IPA activation in the body but not in the brain<sup>185</sup>, indicating the importance of cell context in determining the outcome of transcription elongation-mediated IPA regulation.

## Chromatin structure and modifications

Chromatin configuration and post-translational modifications are integral components of pre-RNA processing (Fig. 5a). For example, genomic regions containing frequently used PAS are depleted of chromatin<sup>186</sup> and the pattern of histone H3 trimethylation at Lys36 (H3K36me3) correlates with the selection of alternative PAS<sup>178</sup>. More recent studies have indicated the relevance of the 3D genome structure to transcriptional activity<sup>187</sup>. Genes located in different compartments of the nucleus, such as near the nuclear lamina versus in nuclear speckles, have distinct features, including the gene size, gene density and GC content<sup>188</sup>. Although it has not yet been systematically examined, it is conceivable that CPA activity and APA regulatory mechanisms may differ in different gene compartments. In line with this, short and long genes seem to have distinct sensitivities to knockdown of the CPA factor *PCF11* (refs. 136,137) and to inhibition of CPSF73 by JTE-607 (ref. 48). Direct evidence of the impact of 3D genome structure on APA comes from the finding that CTCF and cohesin-mediated chromatin looping in the downstream region of a gene inhibits dPAS usage<sup>189</sup>. Consistently, DNA methylation of CTCF binding sites, which suppresses CTCF binding, promotes pPAS use<sup>189</sup>. The CTCF-mediated APA regulation is mechanistically similar to CPA modulation by blocking RNAPII elongation. Akin to this, it was reported that CRISPR–dCas9 targeting the non-template strand of DNA, which blocks RNAPII elongation, could promote the use of PAS situated before the blocking site<sup>190,191</sup>.

## RNA modifications

The *N*<sup>6</sup>-methyladenosine (m<sup>6</sup>A) modification is ubiquitous in mRNAs<sup>192</sup>. Interestingly, m<sup>6</sup>A levels are biased towards short 3'UTR isoforms<sup>192</sup>. In a specific case, the m<sup>6</sup>A level of an IPA isoform of the *HTT* gene (encoding the pathogenic Huntingtin protein) was found to correlate with its expression level in individuals with Huntington disease, as well as in cell lines and mouse models of the disease<sup>193</sup>. These global and gene-specific observations are in line with the proposition that m<sup>6</sup>A deposition in transcripts could be attributed to the RNA dwell time

in the nucleus<sup>158</sup>. Nevertheless, direct suppression of pPAS use by the m<sup>6</sup>A reader protein YTHDC1 was reported in MCF-7 and HEK293T cells, which involves sequestration of FIP1 through liquid–liquid phase separation and hence inhibition of FIP1–CPSF30 interaction<sup>194</sup>. In addition, m<sup>6</sup>A with 2'-O-methylation (m<sup>6</sup>Am), which is enriched towards the 5' end of mRNA<sup>195</sup>, was found to sequester and hence functionally suppress the core CPA factor PCF11, leading to pPAS inhibition<sup>196</sup>. Similar to *PCF11* knockdown<sup>197</sup>, this mechanism promotes the all-*trans* retinoic acid-based therapy of neuroblastoma, which involves induction of cancer cells to a less malignant state through cell differentiation.

## RNA-binding proteins

The large repertoire of RBPs in the cell have been increasingly recognized for their functions in APA<sup>86,198</sup>. By binding to their cognate motifs near the PAS, RBPs mostly hinder, but can also facilitate, the recruitment of the CPA machinery<sup>199</sup>. In addition, although it has not yet been systematically examined, RBPs that regulate splicing could, in principle, alter IPA activity. Here, we discuss a few examples of biological conditions in which APA regulation by RBPs is widespread, to illustrate the principles of RBP-mediated APA control.

Hu proteins – including the ubiquitously expressed HuR (encoded by *ELAVL1*) and three neuron-specific paralogues HuB, HuC and HuD (encoded by *ELAVL2*, *ELAVL3* and *ELAVL4*, respectively) – are RBPs that bind to U-rich regions and have extensive roles in mRNA metabolism. HuB, HuC and HuD downregulate HuR expression through APA, by promoting the expression of a translationally suppressed 3'UTR-APA isoform of HuR<sup>200</sup>. pPAS suppression by Hu proteins was found to have a widespread role in 3'UTR lengthening in neurogenesis in *Drosophila*<sup>201,202</sup>, in particular at the onset of neuronal differentiation<sup>203</sup>. Similar in mechanism but opposite in the direction of regulation, by interacting with UGUA motifs, PQBP1 was found to suppress the use of UGUA-enriched dPAS in neural projector cells. This regulatory scheme is weakened during neurogenesis, when PQBP1 levels decrease<sup>204</sup>.

Mutations of the genes encoding TDP-43 and FUS are associated with familial forms of ALS and FTD. TDP-43 binds GU-rich sequences that resemble the binding sites of CPA factors CSTF64 and CSTF64t (Box 1). Nuclear depletion of TDP-43 owing to its cytoplasmic aggregation is a hallmark of ALS and FTD<sup>205,206</sup>. Several recent studies have uncovered widespread APA changes in individuals with ALS or FTD who have cytoplasmic aggregation of TDP-43 (refs. 68,69,207,208). FUS preferentially binds GU-rich and G-rich sequences<sup>209,210</sup>, both of which are enriched in the downstream region of dPAS. Interestingly, APA regulation by FUS was found to involve both its RNA binding activity and its interaction with H3K36me3 (ref. 211).

APA regulation by RBPs is also common in cancer cells. For example, the U-rich motif-binding RBP HNRNPC is responsible for many APA events in metastases of colon cancer<sup>212</sup> and breast cancer<sup>213</sup>. KHDRBS1 (also known as SAM68), which has specificity for binding U(A/U)AA, interacts with the transcription termination factor XRN2 and leads to altered APA in prostate cancer<sup>50</sup>. Loss-of-function mutations of APC, a GC-rich motif-binding RBP, lead to 3'UTR lengthening in colorectal adenocarcinoma, in which APC mutations are frequent<sup>214</sup>.

## Conclusions and future perspectives

In the past few years, APA has emerged as a widespread gene regulatory mechanism with diverse biological and clinical consequences. Based on more than three decades of biochemical and molecular biology studies of CPA, and propelled more recently by advances in RNA-seq technologies and machine-learning methods, APA research has now come of age. Here, we outline some areas of APA research where outstanding questions remain to be addressed.

The APA isoform expression profile is closely connected with cell identity and is therefore informative as a diagnostic or prognostic tool. However, the accurate identification and quantification of APA isoforms is still challenging, especially when using standard RNA-seq data. Analysing the use of alternative PAS in the context of transcription start site and splicing events is important to understand the coupling between transcriptional and co-transcriptional events. Long-read sequencing, especially on RNA molecules directly, could help address these issues<sup>215</sup>. In addition, metabolic labelling with 4-thiouridine and APEX2-based proximal labelling could further shed light on the life cycle of APA isoforms in the cell, including the interplays between isoform-specific mRNA decay and translation and interactions with RBPs and subcellular locations.

APA regulation often displays a global trend, such as a dPAS-to-pPAS shift or vice versa, involving many genes in a concerted manner. Identification of the most consequential APA event(s) in a given condition is still challenging. APA perturbation tools, such as genomic removal of alternative PAS by CRISPR-Cas9 (ref. 62) or altering their use by dCas9-based or dCas13-based methods<sup>190,191,216</sup>, could be used to address this challenge. Similarly, little is known about the significance of APA of lncRNAs. At least ~15% of lncRNA genes undergo APA<sup>5</sup>. Although some lncRNA APA events, such as that of *NEAT1*, have been found to have important functional consequences, our understanding of lncRNA APA is, for the most part, still in its infancy. PAS-based screening tools are needed to determine the biological relevance of APA events of lncRNA genes.

Although much strides have been made in understanding the regulatory mechanisms of APA, some important aspects are still unclear. For example, largely underexplored are how APA is engaged in different signalling pathways and how chromatin structures and modifications, as well as DNA methylation patterns, might influence APA 'memory' and vice versa. How APA regulation is executed in a spatial manner in different compartments of the nucleus, such as nuclear speckles and nuclear lamina, is just beginning to be unveiled. Moreover, the ever-expanding repertoire of APA-regulating RBPs needs to be systematically and holistically examined in different physiological and pathological contexts.

The genetic information on APA regulation is a rich source of data for APA research, especially in relation to the clinically relevant apaQTLs. However, their validation is a daunting task. High-throughput methods, including single cell-based 3'-end sequencing and massively parallel reporter assays, coupled with machine-learning tools (Box 2) would be instrumental. The co-localization of apaQTLs to other

molecular features, such as gene expression, splicing, protein expression and DNA methylation, can shed light on the mechanisms and consequences of APA. In addition, large-scale validations using data mining tools in combination with high-throughput experimental analysis could offer a fruitful approach to delineate the relevance of genetically variable APA events in different cell contexts.

On the therapeutic front, antisense oligonucleotides have been successfully used to perturb APA<sup>217</sup>. IPA activation has been implicated in neoantigen production, which is potentially relevant to cancer immunotherapies<sup>218</sup>. Targeting APA events, either specifically on individual genes or globally across genes, could be an attractive therapeutic modality for various clinical conditions (such as those listed in Supplementary Tables 1–3) as we gain more knowledge about the mechanisms, consequences and genetics of APA-mediated gene regulation.

Published online: 15 January 2026

## References

- Rodríguez-Molina, J. B., West, S. & Passmore, L. A. Knowing when to stop: transcription termination on protein-coding genes by eukaryotic RNAPII. *Mol. Cell* **83**, 404–415 (2023).
- Passmore, L. A. & Collier, J. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat. Rev. Mol. Cell Biol.* **23**, 93–106 (2022).
- Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S. & Mayr, C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes. Dev.* **27**, 2380–2396 (2013).
- Lee, J. Y., Yeh, I., Park, J. Y. & Tian, B. PolyA\_DB 2: mRNA polyadenylation sites in vertebrate genes. *Nucleic Acids Res.* **35**, D165–D168 (2007).
- Wang, R., Zheng, D., Yehia, G. & Tian, B. A compendium of conserved cleavage and polyadenylation events in mammalian genes. *Genome Res.* **28**, 1427–1441 (2018).
- Boreikaite, V. & Passmore, L. A. 3'-End processing of eukaryotic mRNA: machinery, regulation, and impact on gene expression. *Annu. Rev. Biochem.* **92**, 199–225 (2023).
- Liu, L. & Manley, J. L. Modulation of diverse biological processes by CPSF, the master regulator of mRNA 3' ends. *RNA* **30**, 1122–1140 (2024).
- Hunt, A. G. Review: Mechanisms underlying alternative polyadenylation in plants—looking in the right places. *Plant. Sci.* **324**, 111430 (2022).
- Beaudoing, E. & Gautheret, D. Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST data. *Genome Res.* **11**, 1520–1526 (2001).
- Zhang, H., Lee, J. Y. & Tian, B. Biased alternative polyadenylation in human tissues. *Genome Biol.* **6**, R100 (2005).
- Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008).
- Ji, Z., Lee, J. Y., 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* **106**, 7028–7033 (2009).
- Mayr, C. & Bartel, D. P. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673–684 (2009).
- Shepard, P. J. et al. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-seq. *RNA* **17**, 761–772 (2011).
- Ozsolak, F. et al. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* **143**, 1018–1029 (2010).
- Hoque, M. et al. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat. Methods* **10**, 133–139 (2013).
- Xia, Z. et al. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat. Commun.* **5**, 5274 (2014).
- Agarwal, V., Lopez-Darwin, S., Kelley, D. R. & Shendure, J. The landscape of alternative polyadenylation in single cells of the developing mouse embryo. *Nat. Commun.* **12**, 5101 (2021).
- Guvenc, A. & Tian, B. Analysis of alternative cleavage and polyadenylation in mature and differentiating neurons using RNA-seq data. *Quant. Biol.* **6**, 253–266 (2018).
- Hilgers, V. et al. Neural-specific elongation of 3' UTRs during *Drosophila* development. *Proc. Natl Acad. Sci. USA* **108**, 15864–15869 (2011).
- Smbirt, P. et al. Global patterns of tissue-specific alternative polyadenylation in *Drosophila*. *Cell Rep.* **1**, 277–289 (2012).
- Ulitsky, I. et al. Extensive alternative polyadenylation during zebrafish development. *Genome Res.* **22**, 2054–2066 (2012).
- Li, N. et al. CFIm-mediated alternative polyadenylation safeguards the development of mammalian pre-implantation embryos. *Stem Cell Rep.* **18**, 81–96 (2023).
- Sommerkamp, P. et al. Differential alternative polyadenylation landscapes mediate hematopoietic stem cell activation and regulate glutamine metabolism. *Cell Stem Cell* **26**, 722–738.e7 (2020).
- Qiang, J. et al. Single-cell landscape of alternative polyadenylation in human lymphoid hematopoiesis. *J. Mol. Cell Biol.* **16**, mjae027 (2024).

26. Wang, J. et al. Comprehensive mapping of alternative polyadenylation site usage and its dynamics at single-cell resolution. *Proc. Natl Acad. Sci. USA* **119**, e2113504119 (2022).
27. Li, W. et al. Alternative cleavage and polyadenylation in spermatogenesis connects chromatin regulation with post-transcriptional control. *BMC Biol.* **14**, 6 (2016).
28. West, S. M. et al. Developmental dynamics of gene expression and alternative polyadenylation in the *Caenorhabditis elegans* germline. *Genome Biol.* **19**, 8 (2018).
29. Lee, S. et al. Diverse cell-specific patterns of alternative polyadenylation in *Drosophila*. *Nat. Commun.* **13**, 5372 (2022).
30. Bao, J. et al. UPF2-dependent nonsense-mediated mRNA decay pathway is essential for spermatogenesis by selectively eliminating longer 3'UTR transcripts. *PLoS Genet.* **12**, e1005863 (2016).
31. 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.* **12**, e1005857 (2016).
32. Cheng, L. C. et al. Widespread transcript shortening through alternative polyadenylation in secretory cell differentiation. *Nat. Commun.* **11**, 3182 (2020).
33. Takagaki, Y., Seipelt, R. L., Peterson, M. L. & Manley, J. L. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87**, 941–952 (1996).
34. Laplante, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012).
35. Chang, J. W. et al. mRNA 3'-UTR shortening is a molecular signature of mTORC1 activation. *Nat. Commun.* **6**, 7218 (2015).
36. Chang, J. W. et al. An integrative model for alternative polyadenylation, IntMAP, delineates mTOR-modulated endoplasmic reticulum stress response. *Nucleic Acids Res.* **46**, 5996–6008 (2018).
37. Park, J. Y. et al. Comparative analysis of mRNA isoform expression in cardiac hypertrophy and development reveals multiple post-transcriptional regulatory modules. *PLoS One* **6**, e22391 (2011).
38. Zhao, J., Zhai, B., Gygi, S. P. & Goldberg, A. L. mTOR inhibition activates overall protein degradation by the ubiquitin proteasome system as well as by autophagy. *Proc. Natl Acad. Sci. USA* **112**, 15790–15797 (2015).
39. Tang, H. W. et al. The TORC1-regulated CPA complex rewires an RNA processing network to drive autophagy and metabolic reprogramming. *Cell Metab.* **27**, 1040–1054.e8 (2018).
40. Holler, I. et al. The differential expression of alternatively polyadenylated transcripts is a common stress-induced response mechanism that modulates mammalian mRNA expression in a quantitative and qualitative fashion. *RNA* **22**, 1441–1453 (2016).
41. Devany, E. et al. Intronic cleavage and polyadenylation regulates gene expression during DNA damage response through U1 snRNA. *Cell Discovery* **2**, 16013 (2016).
42. Williamson, L. et al. UV irradiation induces a non-coding RNA that functionally opposes the protein encoded by the same gene. *Cell* **168**, 843–855.e13 (2017).
43. Cugusi, S. et al. Heat shock induces premature transcript termination and reconfigures the human transcriptome. *Mol. Cell* **82**, 1573–1588.e10 (2022).
44. Fink, E. E. et al. Heat shock induces alternative polyadenylation through dynamic DNA methylation and chromatin looping. *Cell Stress Chaperones* **30**, 100084 (2025).
45. Zheng, D. et al. Cellular stress alters 3'UTR landscape through alternative polyadenylation and isoform-specific degradation. *Nat. Commun.* **9**, 2268 (2018).
46. Vilborg, A., Passarelli, M. C., Yario, T. A., Tycowski, K. T. & Steitz, J. A. Widespread inducible transcription downstream of human genes. *Mol. Cell* **59**, 449–461 (2015).
47. Vilborg, A. et al. Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proc. Natl Acad. Sci. USA* **114**, E8362–E8371 (2017).
48. Cui, Y. et al. Elevated pre-mRNA 3' end processing activity in cancer cells renders vulnerability to inhibition of cleavage and polyadenylation. *Nat. Commun.* **14**, 4480 (2023).
49. Liu, L. et al. The anticancer compound JTE-607 reveals hidden sequence specificity of the mRNA 3' processing machinery. *Nat. Struct. Mol. Biol.* **30**, 1947–1957 (2023).
50. Pieraccioni, M. et al. The transcriptional terminator XRN2 and the RNA-binding protein Sam68 link alternative polyadenylation to cell cycle progression in prostate cancer. *Nat. Struct. Mol. Biol.* **29**, 1101–1112 (2022).
51. Pai, A. A. et al. Widespread Shortening of 3' untranslated regions and increased exon inclusion are evolutionarily conserved features of innate immune responses to infection. *PLoS Genet.* **12**, e1006338 (2016).
52. Li, G. W. et al. SCAPTURE: a deep learning-embedded pipeline that captures polyadenylation information from 3' tag-based RNA-seq of single cells. *Genome Biol.* **22**, 221 (2021).
53. Ge, Y. et al. Downregulation of CPSF6 leads to global mRNA 3' UTR shortening and enhanced antiviral immune responses. *PLoS Pathog.* **20**, e1012061 (2024).
54. Kakutani, M., Takeuchi, K., Waga, I., Iwamura, H. & Wakitani, K. JTE-607, a novel inflammatory cytokine synthesis inhibitor without immunosuppression, protects from endotoxin shock in mice. *Inflamm. Res.* **48**, 461–468 (1999).
55. Kakegawa, J., Sakane, N., Suzuki, K. & Yoshida, T. JTE-607, a multiple cytokine production inhibitor, targets CPSF3 and inhibits pre-mRNA processing. *Biochem. Biophys. Res. Commun.* **518**, 32–37 (2019).
56. Twu, K. Y., Noah, D. L., Rao, P., Kuo, R. L. & Krug, R. M. The CPSF30 binding site on the NS1A protein of influenza A virus is a potential antiviral target. *J. Virol.* **80**, 3957–3965 (2006).
57. Wang, X. et al. Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. *Nat. Commun.* **11**, 293 (2020).
58. Bauer, D. L. V. et al. Influenza virus mounts a two-pronged attack on host RNA polymerase II transcription. *Cell Rep.* **23**, 2119–2129.e3 (2018).
59. Hennig, T. et al. HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. *PLoS Pathog.* **14**, e1006954 (2018).
60. Wang, X. et al. Mechanism and consequences of herpes simplex virus 1-mediated regulation of host mRNA alternative polyadenylation. *PLoS Genet.* **17**, e1009263 (2021).
61. Grosso, A. R. et al. Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *eLife* **4**, e09214 (2015).
62. Gabel, A. M. et al. Multiplexed screening reveals how cancer-specific alternative polyadenylation shapes tumor growth in vivo. *Nat. Commun.* **15**, 959 (2024).
63. Abe, K. et al. Downstream-of-gene (DoG) transcripts contribute to an imbalance in the cancer cell transcriptome. *Sci. Adv.* **10**, eadh9613 (2024).
64. Teloni, F. et al. Efficient pre-mRNA cleavage prevents replication-stress-associated genome instability. *Mol. Cell* **73**, 670–683.e12 (2019).
65. Lee, S. H. et al. Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. *Nature* **561**, 127–131 (2018).
66. Singh, I. et al. Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat. Commun.* **9**, 1716 (2018).
67. Zylka, M. J., Simon, J. M. & Philpot, B. D. Gene length matters in neurons. *Neuron* **86**, 353–355 (2015).
68. Arnold, F. J. et al. TDP-43 dysregulation of polyadenylation site selection is a defining feature of RNA misprocessing in amyotrophic lateral sclerosis and frontotemporal dementia. *J. Clin. Invest.* **135**, e182088 (2025).
69. Zeng, Y. et al. TDP-43 nuclear loss in FTD/ALS causes widespread alternative polyadenylation changes. *Nat. Neurosci.* **28**, 2180–2189 (2025).
70. Melamed, Z. et al. Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat. Neurosci.* **22**, 180–190 (2019).
71. Grzechnik, P. & Mischo, H. E. Fateful decisions of where to cut the line: pathology associated with aberrant 3' end processing and transcription termination. *J. Mol. Biol.* **437**, 168802 (2025).
72. Higgs, D. R. et al.  $\alpha$ -Thalassaemia caused by a polyadenylation signal mutation. *Nature* **306**, 398–400 (1983).
73. Gehring, N. H. et al. Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat. Genet.* **28**, 389–392 (2001).
74. Nourse, J., Spada, S. & Danckwardt, S. Emerging roles of RNA 3'-end cleavage and polyadenylation in pathogenesis, diagnosis and therapy of human disorders. *Biomolecules* **10**, 915 (2020).
75. Kainov, Y., Hamid, F. & Makeyev, E. V. Recurrent disruption of tumour suppressor genes in cancer by somatic mutations in cleavage and polyadenylation signals. *eLife* **13**, RP99040 (2024).
76. Wiestner, A. et al. Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood* **109**, 4599–4606 (2007).
77. Johnston, J. J. et al. NAA10 polyadenylation signal variants cause syndromic microphthalmia. *J. Med. Genet.* **56**, 444–452 (2019).
78. Senju, C. et al. Deep intronic founder mutations identified in the ERCC4/XPF gene are potential therapeutic targets for a high-frequency form of xeroderma pigmentosum. *Proc. Natl Acad. Sci. USA* **120**, e2217423120 (2023).
79. Li, L. et al. An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat. Genet.* **53**, 994–1005 (2021).
80. Mittleman, B. E. et al. Alternative polyadenylation mediates genetic regulation of gene expression. *eLife* **9**, e57492 (2020).
81. Graham, R. R. et al. Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc. Natl Acad. Sci. USA* **104**, 6758–6763 (2007).
82. Hellquist, A. et al. The human GIMAP5 gene has a common polyadenylation polymorphism increasing risk to systemic lupus erythematosus. *J. Med. Genet.* **44**, 314–321 (2007).
83. Idda, M. L. et al. Cooperative translational control of polymorphic BAFF by NF90 and miR-15a. *Nucleic Acids Res.* **46**, 12040–12051 (2018).
84. Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* **18**, 18–30 (2017).
85. Gruber, A. J. & Zavolan, M. Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* **20**, 599–614 (2019).
86. Mitschka, S. & Mayr, C. Context-specific regulation and function of mRNA alternative polyadenylation. *Nat. Rev. Mol. Cell Biol.* **23**, 779–796 (2022).
87. An, J. J. et al. Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell* **134**, 175–187 (2008).
88. Tushue, G. et al. Alternative 3' UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments. *Neuron* **98**, 495–511.e6 (2018).
89. Martnez, J. C. et al. Pum2 shapes the transcriptome in developing axons through retention of target mRNAs in the cell body. *Neuron* **104**, 931–946.e5 (2019).
90. Chekulaeva, M. Mechanistic insights into the basis of widespread RNA localization. *Nat. Cell Biol.* **26**, 1037–1046 (2024).
91. Fazal, F. M. et al. Atlas of subcellular RNA localization revealed by APEX-seq. *Cell* **178**, 473–490.e26 (2019).
92. Neve, J. et al. Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Res.* **26**, 24–35 (2016).
93. Chen, S. et al. The mRNA export receptor NXF1 coordinates transcriptional dynamics, alternative polyadenylation, and mRNA export. *Mol. Cell* **74**, 118–131.e7 (2019).

94. Muller-McNicoll, M. et al. SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes. Dev.* **30**, 553–566 (2016).
95. Ruepp, M. D. et al. Mammalian pre-mRNA 3' end processing factor CFIm68 functions in mRNA export. *Mol. Biol. Cell* **20**, 5211–5223 (2009).
96. Cheng, L. C. et al. Alternative 3' UTRs play a widespread role in translation-independent mRNA association with the endoplasmic reticulum. *Cell Rep.* **36**, 109407 (2021).
97. Cui, X. A., Zhang, H. & Palazzo, A. F. p180 promotes the ribosome-independent localization of a subset of mRNA to the endoplasmic reticulum. *PLoS Biol.* **10**, e1001336 (2012).
98. Horste, E. L. et al. Subcytoplasmic location of translation controls protein output. *Mol. Cell* **83**, 4509–4523.e11 (2023).
99. Mayr, C. What are 3' UTRs doing? *Cold Spring Harb. Perspect. Biol.* **11**, a034728 (2019).
100. Hwang, H. J. et al. LC3B is an RNA-binding protein to trigger rapid mRNA degradation during autophagy. *Nat. Commun.* **13**, 1436 (2022).
101. Hogg, J. R. & Goff, S. P. Upf1 senses 3'UTR length to potentiate mRNA decay. *Cell* **143**, 379–389 (2010).
102. Karousis, E. D., Gypas, F., Zavolan, M. & Muhlemann, O. Nanopore sequencing reveals endogenous NMD-targeted isoforms in human cells. *Genome Biol.* **22**, 223 (2021).
103. Guvenek, A. et al. Neuronal cells display distinct stability controls of alternative polyadenylation mRNA isoforms, long non-coding RNAs, and mitochondrial RNAs. *Front. Genet.* **13**, 840369 (2022).
104. Spies, N., Burge, C. B. & Bartel, D. P. 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res.* **23**, 2078–2090 (2013).
105. Gruber, A. R. et al. Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat. Commun.* **5**, 5465 (2014).
106. Floor, S. N. & Doudna, J. A. Tunable protein synthesis by transcript isoforms in human cells. *eLife* **5**, e10921 (2016).
107. Zheng, D. et al. Predicting the translation efficiency of messenger RNA in mammalian cells. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-025-02712-x> (2025).
108. Loedige, I. et al. mRNA stability and m<sup>6</sup>A are major determinants of subcellular mRNA localization in neurons. *Mol. Cell* **83**, 2709–2725.e10 (2023).
109. Sudmant, P. H., Lee, H., Dominguez, D., Heiman, M. & Burge, C. B. Widespread accumulation of ribosome-associated isolated 3' UTRs in neuronal cell populations of the aging brain. *Cell Rep.* **25**, 2447–2456.e4 (2018).
110. Chen, L. L. & Carmichael, G. G. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol. Cell* **35**, 467–478 (2009).
111. Ku, J. et al. Alternative polyadenylation determines the functional landscape of inverted Alu repeats. *Mol. Cell* **84**, 1062–1077.e9 (2024).
112. Dorrity, T. J. et al. Long 3'UTRs predispose neurons to inflammation by promoting immunostimulatory double-stranded RNA formation. *Sci. Immunol.* **8**, eadg2979 (2023).
113. Matheny, T., Van Treeck, B., Huynh, T. N. & Parker, R. RNA partitioning into stress granules is based on the summation of multiple interactions. *RNA* **27**, 174–189 (2021).
114. Han, T. W. et al. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779 (2012).
115. Khong, A. et al. The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Mol. Cell* **68**, 808–820.e5 (2017).
116. Wolozin, B. & Ivanov, P. Stress granules and neurodegeneration. *Nat. Rev. Neurosci.* **20**, 649–666 (2019).
117. Berkovits, B. D. & Mayr, C. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**, 363–367 (2015).
118. Lee, S. H. & Mayr, C. Gain of additional BIRC3 protein functions through 3'-UTR-mediated protein complex formation. *Mol. Cell* **74**, 701–712.e9 (2019).
119. Hirose, T., Ninomiya, K., Nakagawa, S. & Yamazaki, T. A guide to membraneless organelles and their various roles in gene regulation. *Nat. Rev. Mol. Cell Biol.* **24**, 288–304 (2023).
120. Wilusz, J. E. et al. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes. Dev.* **26**, 2392–2407 (2012).
121. Park, M. K. et al. NEAT1 is essential for metabolic changes that promote breast cancer growth and metastasis. *Cell Metab.* **33**, 2380–2397.e9 (2021).
122. Kehry, M. et al. The immunoglobulin  $\mu$  chains of membrane-bound and secreted IgM molecules differ in their C-terminal segments. *Cell* **21**, 393–406 (1980).
123. Rogers, J. et al. Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin  $\mu$  chain. *Cell* **20**, 303–312 (1980).
124. Davis, M. J. et al. Differential use of signal peptides and membrane domains is a common occurrence in the protein output of transcriptional units. *PLoS Genet.* **2**, e46 (2006).
125. Vorlová, S. et al. Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Mol. Cell* **43**, 927–939 (2011).
126. Ravichandran, Y. et al. The distinct localization of CDC42 isoforms is responsible for their specific functions during migration. *J. Cell Biol.* **223**, e202004092 (2024).
127. Cassago, A. et al. Mitochondrial localization and structure-based phosphate activation mechanism of glutaminase C with implications for cancer metabolism. *Proc. Natl Acad. Sci. USA* **109**, 1092–1097 (2012).
128. Xu, L. et al. A glutaminase isoform switch drives therapeutic resistance and disease progression of prostate cancer. *Proc. Natl Acad. Sci. USA* **118**, e2012748118 (2021).
129. Kamieniarz-Gdula, K. & Proudfoot, N. J. Transcriptional control by premature termination: a forgotten mechanism. *Trends Genet.* **35**, 553–564 (2019).
130. Soles, L. V. et al. A nuclear RNA degradation code is recognized by PAXT for eukaryotic transcriptome surveillance. *Mol. Cell* **85**, 1575–1588.e9 (2025).
131. Lee, E. S. et al. ZFC3H1 and U1-70K promote the nuclear retention of mRNAs with 5' splice site motifs within nuclear speckles. *RNA* **28**, 878–894 (2022).
132. Kesner, J. S. et al. Noncoding translation mitigation. *Nature* **617**, 395–402 (2023).
133. Di Giandomartino, D. C. et al. RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs. *Genes. Dev.* **28**, 2248–2260 (2014).
134. 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.* **9**, e1003613 (2013).
135. 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* **95**, 14302–14307 (1998).
136. Kamieniarz-Gdula, K. et al. Selective roles of vertebrate PCF11 in premature and full-length transcript termination. *Mol. Cell* **74**, 158–172.e9 (2019).
137. Wang, R., Zheng, D., Wei, L., Ding, Q. & Tian, B. Regulation of intronic polyadenylation by PCF11 impacts mRNA expression of long genes. *Cell Rep.* **26**, 2766–2778.e6 (2019).
138. Zhao, W. & Manley, J. L. Complex alternative RNA processing generates an unexpected diversity of poly(A) polymerase isoforms. *Mol. Cell Biol.* **16**, 2378–2386 (1996).
139. Liu, L. & Manley, J. L. Non-canonical isoforms of the mRNA polyadenylation factor WDR33 regulate STING-mediated immune responses. *Cell Rep.* **43**, 113886 (2024).
140. Martincic, K. et al. Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the GO to S phase transition. *Proc. Natl Acad. Sci. USA* **95**, 11095–11100 (1998).
141. Dass, B. et al. Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility. *Proc. Natl Acad. Sci. USA* **104**, 20374–20379 (2007).
142. Sartini, B. L., Wang, H., Wang, W., Millette, C. F. & Kilpatrick, D. L. Pre-messenger RNA cleavage factor I (CFIm): potential role in alternative polyadenylation during spermatogenesis. *Biol. Reprod.* **78**, 472–482 (2008).
143. Gallicchio, L. et al. A developmental mechanism to regulate alternative polyadenylation in an adult stem cell lineage. *Genes. Dev.* **38**, 655–674 (2024).
144. LaForce, G. R. et al. Suppression of premature transcription termination leads to reduced mRNA isoform diversity and neurodegeneration. *Neuron* **110**, 1340–1357.e7 (2022).
145. Brumbaugh, J. et al. Nudt21 controls cell fate by connecting alternative polyadenylation to chromatin signaling. *Cell* **172**, 629–631 (2018).
146. Lackford, B. et al. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* **33**, 878–889 (2014).
147. Chen, Y. et al. Alternative mRNA polyadenylation regulates macrophage hyperactivation via the autophagy pathway. *Cell Mol. Immunol.* **21**, 1522–1534 (2024).
148. Davis, A. G. et al. Alternative polyadenylation dysregulation contributes to the differentiation block of acute myeloid leukemia. *Blood* **139**, 424–438 (2021).
149. Masamha, C. P. et al. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* **510**, 412–416 (2014).
150. Li, W. et al. Systematic profiling of poly(A)<sup>+</sup> transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet.* **11**, e1005166 (2015).
151. Stroup, E. K. & Ji, Z. Deep learning of human polyadenylation sites at nucleotide resolution reveals molecular determinants of site usage and relevance in disease. *Nat. Commun.* **14**, 7378 (2023).
152. Jenal, M. et al. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* **149**, 538–553 (2012).
153. Zhu, Y. et al. Molecular mechanisms for CFIm-mediated regulation of mRNA alternative polyadenylation. *Mol. Cell* **69**, 62–74.e4 (2018).
154. Martin, G., Gruber, A. R., 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.* **1**, 753–763 (2012).
155. Liu, S. et al. CPSF6 regulates alternative polyadenylation and proliferation of cancer cells through phase separation. *Cell Rep.* **42**, 113197 (2023).
156. Yoon, Y. et al. RBBP6 anchors pre-mRNA 3' end processing to nuclear speckles for efficient gene expression. *Mol. Cell* **85**, 555–570.e8 (2025).
157. Tang, P. et al. Alternative polyadenylation by sequential activation of distal and proximal polyA sites. *Nat. Struct. Mol. Biol.* **29**, 21–31 (2022).
158. Tang, P. et al. Nuclear retention coupled with sequential polyadenylation dictates post-transcriptional m<sup>6</sup>A modification in the nucleus. *Mol. Cell* **84**, 3758–3774.e10 (2024).
159. Seidler, J. F. & Strasser, K. Understanding nuclear mRNA export: survival under stress. *Mol. Cell* **84**, 3681–3691 (2024).
160. Harlen, K. M. & Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* **18**, 263–273 (2017).
161. Parua, P. K. et al. A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature* **558**, 460–464 (2018).
162. Fan, Z. et al. CDK13 cooperates with CDK12 to control global RNA polymerase II processivity. *Sci. Adv.* **6**, eaaz5041 (2020).
163. Krajewska, M. et al. CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nat. Commun.* **10**, 1757 (2019).
164. Dubburly, S. J., Boutz, P. L. & Sharp, P. A. CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature* **564**, 141–145 (2018).
165. Cortazar, M. A. et al. Control of RNA Pol II speed by PNUMS-PP1 and Spt5 dephosphorylation facilitates termination by a “sitting duck torpedo” mechanism. *Mol. Cell* **76**, 896–908.e4 (2019).

166. Tellier, M. et al. CDK9 and PP2A regulate RNA polymerase II transcription termination and coupled RNA maturation. *EMBO Rep.* **23**, e54520 (2022).
167. Reimer, K. A., Mimoso, C. A., Adelman, K. & Neugebauer, K. M. Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol. Cell* **81**, 998–1012.e7 (2021).
168. Kaida, D. et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664–668 (2010).
169. Feng, Q. et al. U4 snRNP inhibits premature cleavage and polyadenylation of pre-mRNAs. *Proc. Natl Acad. Sci. USA* **121**, e2406710121 (2024).
170. Feng, Q. et al. Functional inhibition of core spliceosomal machinery activates intronic premature cleavage and polyadenylation of pre-mRNAs. *Cell Rep.* **44**, 115376 (2025).
171. Berg, M. G. et al. U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**, 53–64 (2012).
172. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363 (2013).
173. Kim, G., Carroll, C. L., Wakefield, Z. P., Tuncay, M. & Fiszbein, A. U1 snRNP regulates alternative promoter activity by inhibiting premature polyadenylation. *Mol. Cell* **85**, 1968–1981.e7 (2025).
174. Mimoso, C. A. & Adelman, K. U1 snRNP increases RNA Pol II elongation rate to enable synthesis of long genes. *Mol. Cell* **83**, 1264–1279.e10 (2023).
175. So, B. R. et al. A complex of U1 snRNP with cleavage and polyadenylation factors controls telescripting, regulating mRNA transcription in human cells. *Mol. Cell* **76**, 590–599.e4 (2019).
176. Kwon, B. et al. Enhancers regulate 3' end processing activity to control expression of alternative 3'UTR isoforms. *Nat. Commun.* **13**, 2709 (2022).
177. Nagaike, T. et al. Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol. Cell* **41**, 409–418 (2011).
178. Ji, Z. et al. Transcriptional activity regulates alternative cleavage and polyadenylation. *Mol. Syst. Biol.* **7**, 534 (2011).
179. Alfonso-Gonzalez, C. et al. Sites of transcription initiation drive mRNA isoform selection. *Cell* **186**, 2438–2455.e22 (2023).
180. Calvo-Roitberg, E. et al. mRNA initiation and termination are spatially coordinated. *Science* **390**, eado8279 (2025).
181. Lu, C. et al. Mediator regulates transcriptional termination through crosstalk with pre-mRNA 3' end processing factors. *Mol. Cell* **85**, 2147–2164.e10 (2025).
182. Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D. L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat. Struct. Mol. Biol.* **15**, 71–78 (2008).
183. Baluapuri, A. et al. Integrator loss leads to dsRNA formation that triggers the integrated stress response. *Cell* **188**, 3184–3201.e21 (2025).
184. Yang, Y. et al. PAF complex plays novel subunit-specific roles in alternative cleavage and polyadenylation. *PLoS Genet.* **12**, e1005794 (2016).
185. Liu, X. et al. Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in *Drosophila melanogaster*. *RNA* **23**, 1807–1816 (2017).
186. Spies, N., Nielsen, C. B., Padgett, R. A. & Burge, C. B. Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* **36**, 245–254 (2009).
187. Rippe, K. & Papanonis, A. RNA polymerase II transcription compartments—from factories to condensates. *Nat. Rev. Genet.* **26**, 775–788 (2025).
188. Chaturvedi, P. & Belmont, A. S. Nuclear speckle biology: at the cross-roads of discovery and functional analysis. *Curr. Opin. Cell Biol.* **91**, 102438 (2024).
189. Nanavaty, V. et al. DNA methylation regulates alternative polyadenylation via CTCF and the cohesin complex. *Mol. Cell* **78**, 752–764.e6 (2020).
190. Shin, J. et al. CRISPRpas: programmable regulation of alternative polyadenylation by dCas9. *Nucleic Acids Res.* **50**, e25 (2022).
191. Zukher, I., Dujardin, G., Sousa-Luis, R. & Proudfoot, N. J. Elongation roadblocks mediated by dCas9 across human genes modulate transcription and nascent RNA processing. *Nat. Struct. Mol. Biol.* **30**, 1536–1548 (2023).
192. Molinie, B. et al. m<sup>5</sup>A-LAIC-seq reveals the census and complexity of the m<sup>5</sup>A epitranscriptome. *Nat. Methods* **13**, 692–698 (2016).
193. Pupak, A. et al. m<sup>5</sup>A modification of mutant huntingtin RNA promotes the biogenesis of pathogenic huntingtin transcripts. *EMBO Rep.* **25**, 5026–5052 (2024).
194. Chen, L. et al. Nuclear m<sup>5</sup>A reader YTHDC1 suppresses proximal alternative polyadenylation sites by interfering with the 3' processing machinery. *EMBO Rep.* **23**, e54686 (2022).
195. Jin, H., Shi, Z., Zhou, T. & Xie, S. Regulation of m<sup>5</sup>A RNA modification and its implications in human diseases. *J. Mol. Cell Biol.* **16**, mjae012 (2024).
196. An, H. et al. m<sup>5</sup>A sequesters PCF11 to suppress premature termination and drive neuroblastoma differentiation. *Mol. Cell* **84**, 4142–4157.e14 (2024).
197. Ogorodnikov, A. et al. Transcriptome 3' end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. *Nat. Commun.* **9**, 5331 (2018).
198. Panthi, A. & Lynch, K. W. RNA processing in innate immunity: regulation by RNA-binding proteins. *Trends Biochem. Sci.* **50**, 610–621 (2025).
199. Zheng, D. & Tian, B. RNA-binding proteins in regulation of alternative cleavage and polyadenylation. *Adv. Exp. Med. Biol.* **825**, 97–127 (2014).
200. Mansfield, K. D. & Keene, J. D. Neuron-specific ELAV/Hu proteins suppress HuR mRNA during neuronal differentiation by alternative polyadenylation. *Nucleic Acids Res.* **40**, 2734–2746 (2012).
201. Wei, L. et al. Overlapping activities of ELAV/Hu family RNA binding proteins specify the extended neuronal 3' UTR landscape in *Drosophila*. *Mol. Cell* **80**, 140–155.e6 (2020).
202. Carrasco, J. et al. ELAV and FNE determine neuronal transcript signatures through Exon-activated rescue. *Mol. Cell* **80**, 156–163.e6 (2020).
203. Carrasco, J., Mateos, F. & Hilgers, V. A critical developmental window for ELAV/Hu-dependent mRNA signatures at the onset of neuronal differentiation. *Cell Rep.* **41**, 111542 (2022).
204. Liu, X. et al. Dynamic regulation of alternative polyadenylation by PQBP1 during neurogenesis. *Cell Rep.* **43**, 114525 (2024).
205. Neumann, M. et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
206. Ling, S. C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79**, 416–438 (2013).
207. Bryce-Smith, S. et al. TDP-43 loss induces extensive cryptic polyadenylation in ALS/FTD. *Nat. Neurosci.* **28**, 2190–2200 (2025).
208. Vogler, T. O. et al. Form amyloid-like myo-granules in regenerating muscle. *Nature* **563**, 508–513 (2018).
209. Takeda, J. I., Masuda, A. & Ohno, K. Six GU-rich (6GU<sub>6</sub>) FUS-binding motifs detected by normalization of CLIP-seq by Nascent-seq. *Gene* **618**, 57–64 (2017).
210. Bhatt, U. et al. G-quadruplex structural motifs modulate protein–RNA interactions within the transcriptome. *Genome Biol.* **26**, 324 (2025).
211. Jia, J. et al. FUS reads histone H3K36me3 to regulate alternative polyadenylation. *Nucleic Acids Res.* **52**, 5549–5571 (2024).
212. Fischl, H. et al. hnRNPc regulates cancer-specific alternative cleavage and polyadenylation profiles. *Nucleic Acids Res.* **47**, 7580–7591 (2019).
213. Navickas, A. et al. An mRNA processing pathway suppresses metastasis by governing translational control from the nucleus. *Nat. Cell Biol.* **25**, 892–903 (2023).
214. Gabel, A. M., Belleville, A. E., Thomas, J. D., Pineda, J. M. B. & Bradley, R. K. APC mutations dysregulate alternative polyadenylation in cancer. *Genome Biol.* **25**, 255 (2024).
215. Alfonso-Gonzalez, C. & Hilgers, V. Elucidating the coordination of RNA processing using short-read and long-read RNA-sequencing methods. *Nat. Rev. Mol. Cell Biol.* <https://doi.org/10.1038/s41580-025-00895-4> (2025).
216. Tian, S. et al. CRISPR-iPAS: a novel dCAS13-based method for alternative polyadenylation interference. *Nucleic Acids Res.* **50**, e26 (2022).
217. Vickers, T. A., Wyatt, J. R., Burckin, T., Bennett, C. F. & Freier, S. M. Fully modified 2' MOE oligonucleotides redirect polyadenylation. *Nucleic Acids Res.* **29**, 1293–1299 (2001).
218. Liu, X. et al. DIPAN: detecting personalized intronic polyadenylation derived neoantigens from RNA sequencing data. *Comput. Struct. Biotechnol. J.* **23**, 2057–2066 (2024).
219. Tian, B. & Graber, J. H. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA* **3**, 385–396 (2012).
220. Stroup, E. & Ji, Z. Delineating yeast cleavage and polyadenylation signals using deep learning. *Genome Res.* **34**, 1066–1080 (2024).
221. Nunes, N. M., 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.* **29**, 1523–1536 (2010).
222. Cheng, Y., Miura, R. M. & Tian, B. Prediction of mRNA polyadenylation sites by support vector machine. *Bioinformatics* **22**, 2320–2325 (2006).
223. Bogard, N., Linder, J., Rosenberg, A. B. & Seelig, G. A Deep neural network for predicting and engineering alternative polyadenylation. *Cell* **178**, 91–106.e23 (2019).
224. Linder, J., Koplik, S. E., Kundaje, A. & Seelig, G. Deciphering the impact of genetic variation on human polyadenylation using APARENT2. *Genome Biol.* **23**, 232 (2022).
225. Hu, J., Lutz, C. S., Wilusz, J. & Tian, B. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA* **11**, 1485–1493 (2005).
226. Shi, Y. et al. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell* **33**, 365–376 (2009).
227. Muckenfuss, L. M., Migenda Herranz, A. C., Boneberg, F. M., Clerici, M. & Jinek, M. Fip1 is a multivalent interaction scaffold for processing factors in human mRNA 3' end biogenesis. *eLife* **11**, e80332 (2022).
228. Hamilton, K. & Tong, L. Molecular mechanism for the interaction between human CPSF30 and hFip1. *Genes. Dev.* **34**, 1753–1761 (2020).
229. Gregersen, L. H. et al. SCAF4 and SCAF8, mRNA anti-terminator proteins. *Cell* **177**, 1797–1813.e18 (2019).
230. Li, R. et al. Proximal proteomics reveals a landscape of human nuclear condensates. *Nat. Cell Biol.* **27**, 2198–2213 (2025).
231. Yoon, Y., Liu, L., Quan, C. & Shi, Y. Emerging roles of biomolecular condensates in pre-mRNA 3' end processing. *Wiley Interdiscip. Rev. RNA* **16**, e70024 (2025).
232. Yang, S. W. et al. A cancer-specific ubiquitin ligase drives mRNA alternative polyadenylation by ubiquitinating the mRNA 3' end processing complex. *Mol. Cell* **77**, 1206–1221.e7 (2020).
233. Heller-Trulli, D., Liu, H., Mukherjee, S. & Moore, C. L. UBE3D regulates mRNA 3' end processing and maintains adipogenic potential in 3T3-L1 cells. *Mol. Cell Biol.* **42**, e0017422 (2022).
234. Lin, P. et al. RBBP6 maintains glioblastoma stem cells through CPSF3-dependent alternative polyadenylation. *Cell Discov.* **10**, 32 (2024).
235. Di Giammartino, D. C., Shi, Y. & Manley, J. L. PARP1 represses PAP and inhibits polyadenylation during heat shock. *Mol. Cell* **49**, 7–17 (2013).
236. Lin, J. et al. NUDT21 lactylation reprograms alternative polyadenylation to promote cuproptosis resistance. *Cell Discov.* **11**, 52 (2025).

237. Li, W. V., Zheng, D., Wang, R. & Tian, B. MAAPER: model-based analysis of alternative polyadenylation using 3' end-linked reads. *Genome Biol.* **22**, 222 (2021).
238. Balazs, Z. et al. Template-switching artifacts resemble alternative polyadenylation. *BMC Genomics* **20**, 824 (2019).
239. Herrmann, C. J. et al. PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3' end sequencing. *Nucleic Acids Res.* **48**, D174–D179 (2020).
240. Wang, R., Nambiar, R., Zheng, D. & Tian, B. PolyA\_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res.* **46**, D315–D319 (2018).
241. Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **16**, 1297–1305 (2019).
242. Kowalski, M. H. et al. Multiplexed single-cell characterization of alternative polyadenylation regulators. *Cell* **187**, 4408–4425.e23 (2024).
243. Fansler, M. M., Mitschka, S. & Mayr, C. Quantifying 3'UTR length from scRNA-seq data reveals changes independent of gene expression. *Nat. Commun.* **15**, 4050 (2024).
244. Zou, X. et al. Impact of rare non-coding variants on human diseases through alternative polyadenylation outliers. *Nat. Commun.* **16**, 682 (2025).
245. Li, Y. et al. Alternative polyadenylation quantitative trait methylation mapping in human cancers provides clues into the molecular mechanisms of APA. *Am. J. Hum. Genet.* **111**, 562–583 (2024).

## Acknowledgements

The authors thank members of the Tian laboratory for helpful discussions. Research in the Tian laboratory was supported by National Institutes of Health (NIH) grant R35GM153277.

## Author contributions

All authors researched data for the article. B.T. and S.Y. contributed substantially to discussion of the content. B.T. wrote the article with input from S.Y. All authors reviewed and/or edited the manuscript before submission.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41576-025-00928-w>.

**Peer review information** *Nature Reviews Genetics* thanks Valérie Hilgers and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature Limited 2026