Alternative cleavage and polyadenylation: extent, regulation and function

Ran Elkon, Alejandro P. Ugalde and Reuven Agami

Abstract | The 3′ end of most protein-coding genes and long non-coding RNAs is cleaved and polyadenylated. Recent discoveries have revealed that a large proportion of these genes contains more than one polyadenylation site. Therefore, alternative polyadenylation (APA) is a widespread phenomenon, generating mRNAs with alternative 3′ ends. APA contributes to the complexity of the transcriptome by generating isoforms that differ either in their coding sequence or in their 3′ untranslated regions (UTRs), thereby potentially regulating the function, stability, localization and translation efficiency of target RNAs. Here, we review our current understanding of the polyadenylation process and the latest progress in the identification of APA events, mechanisms that regulate poly(A) site selection, and biological processes and diseases resulting from APA.

Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP). A method for profiling RNA that is bound to a specific protein. Cells are grown in a medium containing 4-thiouridine or 6-thioguanosine, which, when incorporated into RNA, allows for efficient ultraviolet crosslinking to RNA-binding proteins. The immunoprecipitated protein–RNA complexes are then used to generate libraries for deep sequencing.

Division of Gene Regulation, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. Correspondence to R.A.
e-mail: r.agami@nki.nl
doi:10.1038/nrg3482

The 3′ end of the vast majority of eukaryotic mRNAs contains a long stretch of untemplated adenosines termed the poly(A) tail. Polyadenylation activity was first observed more than 50 years ago[1]. It was recognized as a post-transcriptional modification added to the 3′ ends of mRNAs a decade later, and it took yet another decade to unravel the requirement of a dedicated apparatus for processing of 3′ ends of mRNAs (recently reviewed in REF. 2).

3′ end processing is a two-step nuclear process that involves an endonucleolytic cleavage of the transcribed transcript followed by the addition of the poly(A) tail (FIG. 1); this process is required for nuclear export and stability of mature transcripts and for efficient translation of mRNAs[3]. Intriguingly, poly(A) tails have a restricted length that can greatly differ between species (for example, in humans, some 250–300 adenines are added on average compared with 70–80 in yeast). The length of poly(A) tails is important, as mRNAs with tails that are too short are, in general, subjected to enzymatic degradation[4] or stored in a translationally dormant state[5].

Earlier work identified the canonical *cis*-acting RNA elements and several dozen core and auxiliary polypeptides involved in the process of cleavage and polyadenylation (FIG. 1). Recent work using photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) has globally mapped the binding of key cleavage and polyadenylation factors to 3′ ends of transcripts and has confirmed to a large extent the previously characterized target motifs[6].

Recent discoveries using transcriptome-wide techniques have revealed that most human genes contain more than one poly(A) site, thereby indicating that alternative polyadenylation (APA) is a widespread phenomenon that generates transcript isoforms with alternative 3′ ends. APA events can be classified into four general classes (FIG. 2): tandem 3′UTR APA and alternative terminal exon APA, which involve cleavage in 3′UTRs and are the most frequent APA forms; intronic APA, which is less frequent; and internal exon APA, which is the least frequent type. Through these four general classes, APA contributes to the complexity of the transcriptome by generating isoforms that differ either in their coding sequence or in their 3′UTRs. Through alteration of coding sequences, APA potentially regulates the function of the affected genes. Through alteration of 3′UTRs, APA potentially regulates the stability, cellular localization and translation efficiency of target RNAs, as 3′UTRs serve as major docking platforms for factors that control these regulatory layers (for example, microRNAs (miRNAs) and RNA-binding proteins)[7,8].

Recent years have witnessed tremendous progress in technologies for mapping poly(A) sites and identifying APA events, thereby substantially broadening our recognition of APA as an important layer of gene regulation. In this Review, we discuss these recent technological

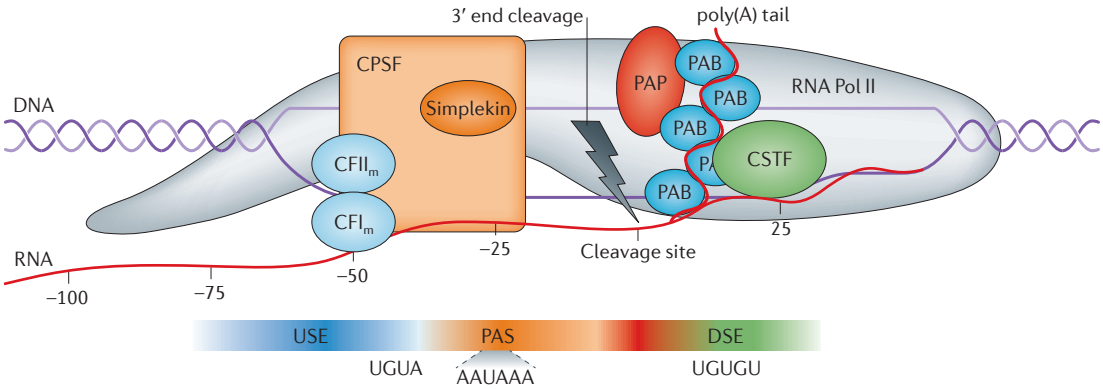

Figure 1 | **Core players involved in cleavage and polyadenylation.** Cleavage and polyadenylation of transcripts requires several *cis*-acting RNA elements and several dozen core and auxiliary polypeptides (reviewed in REFS 51,81,82). *Trans*-acting factors and their corresponding *cis*-elements are coloured similarly. The key *cis*-element that dictates cleavage is a 6 nt motif called the poly(A) signal (PAS), the canonical form of which is AAUAAA, but it can adopt more than ten variants thereof and is mostly located 15–30 nt upstream of the cleavage site. In addition to the PAS, nearby U- or GU-rich downstream sequence elements (DSEs) and less well-defined upstream sequence elements (USEs) enhance cleavage efficiency. Cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulating factor (CSTF) are two multi-polypeptide complexes that, respectively, recognize the PAS and DSEs to promote cleavage between these two elements[2,82]. CPSF consists of six polypeptides, named CPSF4, CPSF2, CPSF1, CPSF3, FIP1L1 and WDR33, where CPSF1 is the subunit that recognizes the PAS[82]. CSTF consists of three subunits (namely, CSTF1, CSTF2 and CSTF3) that are implicated in PAS selection. Additional factors that were found to be required for the cleavage and polyadenylation process are the highly evolutionarily conserved poly(A) polymerase (PAP), the scaffold protein simplekin and cleavage factor Im (CFIm) and CFIIm. CFIm is a tetramer composed of two subunits (namely, CPSF6 and CFIm25) that bind UGUA motifs upstream of the cleavage site[12,83] and two larger polypeptides. PAB, poly(A)-binding protein.

advances, the biological processes associated with extensive APA modulation in mammalian systems, our current understanding of regulatory mechanisms that control poly(A) site selection and finally the involvement of APA misregulation in human disease and the potential use of APA manipulation in therapy.

Global mapping of APA

EST databases. The wealth of sequence information in expressed sequence tag (EST) databases was the first resource that allowed APA analysis beyond the single-gene level[9,10]. Early studies in this direction used ESTs sequenced from mRNA 3′ ends and identified strong evidence for APA in humans and mice, with 54% and 32% of human and mouse transcripts, respectively, being alternatively cleaved[11]. This resource of putative poly(A) sites was then used to computationally derive poly(A) signals (PASs), which are generally enriched in the 50 nt region upstream of the poly(A) sites. In addition to the canonical PAS AAUAAA and its main variant AUUAAA, which were detected in ~60% and 15% of the poly(A) sites, respectively, nine additional PAS variants were identified and collectively appeared in ~14% of the sites. All of these variant signals show the characteristic position distribution peaking at ~15–20 nt upstream of the cleavage sites. Auxiliary upstream and downstream signals were also identified, including the UGUA motif that is typically located 40–100 nt upstream of the cleavage site and is bound by cleavage factor Im (CFIm)[12] (FIG. 1). Furthermore, these studies indicated that the actual cleavage position typically fluctuates by ~10 nt with respect to a major cleavage position.

These analyses also pointed out that in mRNAs containing two or more poly(A) sites, the proximal sites tend to use variant signals and are therefore generally weaker, whereas the most distal ones tend to use the canonical PAS and therefore are usually the strongest sites.

Microarray technology. Although they are useful in mapping poly(A) sites, EST databases allow only limited APA analyses that are restricted in coverage to biological conditions used for preparation of the EST libraries. Gene expression microarrays offer a more agile alternative for APA analysis but still rely on EST-derived annotations of poly(A) sites. Studies that use array data for APA analysis mostly focus on events of tandem 3′UTR APA. For this analysis, array probes are divided into 'common' and 'extended' 3′UTR regions (common probes detect both isoforms generated by proximal and distal sites usage, whereas extended probes detect only the isoform generated by cleavage at the distal poly(A) site). The ratio between the common and extended probes of a transcript is indicative of the relative usage of the corresponding proximal and distal poly(A) sites. Importantly, the first study that applied this approach detected global increase in the signal of the common probes relative to the extended ones in conditions of higher cellular proliferation, reflecting widespread 3′UTR shortening in these conditions (discussed below)[13].

Although supporting global analysis, microarray-based analysis of APA has several serious limitations: it is restricted by the design of the array probes, and the portion of genes covered by probes in a manner that allows APA analysis is limited; it relies on previously

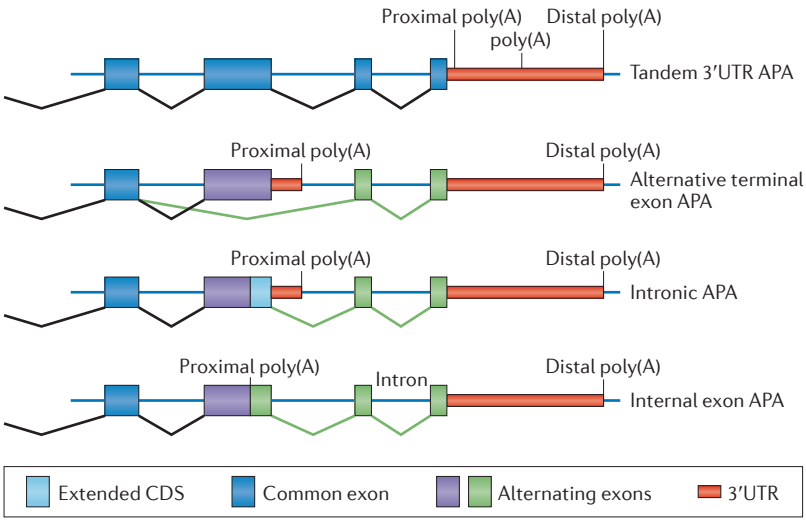

Figure 2 | **The four different APA types.** The simplest alternative polyadenylation (APA) type, which is termed tandem 3′ untranslated region (UTR) APA, involves the occurrence of alternative poly(A) sites within the same terminal exon and hence generates multiple isoforms that differ in their 3′UTR length without affecting the protein encoded by the gene. The other three types involve APA events, which potentially affect the coding sequences in addition to the 3′UTRs. These types are: alternative terminal exon APA, in which alternative splicing generates isoforms that differ in their last exon; intronic APA, which involves cleaving at the cryptic intronic poly(A) signal (PAS), extending an internal exon and making it the terminal one; and internal exon APA, which involves premature polyadenylation within the coding region.

annotated poly(A) sites and cannot identify new sites; and when transcripts have more than two poly(A) sites, quantification is difficult.

RNA-seq. The advent of RNA sequencing (RNA-seq) has boosted not only our ability to measure gene expression levels and to identify novel transcribed regions in the genome but has also greatly enhanced our ability to study alternative transcript processing under different conditions: most prominently alternative splicing, but also APA.

The first systematic application of RNA-seq to examine APA in humans profiled the transcriptomes of 15 diverse human tissues and cell lines[14]. Using EST-based poly(A) annotations, transcripts with alternative 3′UTR were identified, and read densities (reflecting isoform expression levels) that correspond to the common and extended 3′UTR regions were calculated for each such transcript in each tissue or cell line. Remarkably, ~75% (that is, 3,801 out of 5,136) of the transcripts with tandem 3′UTRs showed substantial tissue-regulated APA.

In principle, RNA-seq data can be used to map poly(A) sites on the basis of identification of sequenced reads that contain a stretch of untemplated adenosines at their 3′ end. Such reads potentially span the actual cleavage position and allow precise mapping of poly(A) sites with nucleotide resolution. A study that collected ultra-deep RNA-seq data (~1.2×10^9 sequenced reads) was able to define ~8,000 putative novel poly(A) sites[15]. This analysis demonstrated the potential of sequencing techniques for comprehensive identification of poly(A) sites but also indicated the low yield of standard RNA-seq in APA analysis, as only a very small fraction of reads actually spans cleavage sites and can be used for mapping poly(A) sites.

3′-end-enriched RNA-seq. The full potential of deep sequencing for the study of APA has started to materialize only in the past 2 years after the implementation of deep-sequencing techniques that are specifically adapted for poly(A) analysis. A recent review discussed the main deep-sequencing protocols for 3′ end transcript analysis[16], and we briefly describe them in BOX 1. These novel techniques allow studying APA on a truly comprehensive scale for the first time. Studies applying these techniques demonstrate that APA is pervasive in all eukaryotes examined so far, including yeast, plants, insects, vertebrates and mammals, bringing current estimates for the proportion of human genes subjected to APA to ~70–75%[17]. These recent discoveries highlight APA as a widespread regulatory mechanism that controls gene expression and expands protein diversity. In particular, they have brought about an exponential growth in the number of annotated poly(A) sites and have revealed that APA is a common phenomenon that also affects long non-coding RNAs (lncRNAs)[18]. In addition, as discussed below, they are highly instrumental in detecting biological conditions in which APA is modulated and in studying underlying mechanisms that regulate APA. The above techniques infer APA modulation on the basis of changes in relative levels of the proximal and distal poly(A) sites. However, to establish that APA modulation is indeed the mechanism that underlies the observed changes, these techniques should be accompanied by the experimental techniques discussed in BOX 2.

Biological processes and APA modulation

Our knowledge of biological processes that display extensive APA alterations has substantially increased in recent years. In this section, we review these connections between biological conditions and APA modulation (FIG. 3).

Development and cellular differentiation. EST data collected from divergent tissue sources led to the identification of many tissue-specific APA events as well as to global trends of poly(A) site selection in specific human tissues[19]. For example, transcripts in the nervous system and brain are characterized by preferential usage of distal poly(A) sites (generating isoforms with longer 3′UTRs), whereas in the placenta, ovaries and blood, they tend to use more proximal sites. A subsequent study used serial analysis of gene expression (SAGE) and expression array data sets to examine APA modulation during mouse embryonic development and unravelled that embryonic development is accompanied by progressive lengthening of 3′UTRs. This phenomenon was recapitulated *in vitro* using myogenesis of C2C12 myoblasts as a model[20]. Similar trends of extreme long and short 3′UTR lengths in the brain and gonads,

Serial analysis of gene expression
(SAGE). A method for quantitative and simultaneous analysis of a large number of transcripts; short sequence tags are isolated, concentrated and cloned; their sequencing reveals a gene expression pattern that is characteristic of the tissue or cell type from which the tags were isolated.

Induced pluripotent stem cells
(iPSCs). Pluripotent stem cells that are artificially derived from non-pluripotent cells, typically by genetic manipulation.

Striatal neurons
Neurons that lie in the striatum, which is an area of the brain involved in fine movements, emotion and cognition.

Hippocampal neurons
Neurons that lie in the hippocampus, which is a neurogenic region of the forebrain that has an important functions in learning and memory.

respectively, and of progressive 3′UTR lengthening during development were confirmed by additional studies and were also observed in *Drosophila melanogaster* and zebrafish (*Danio rerio*)[21–24]. Furthermore, it was observed that the generation of induced pluripotent stem cells (iPSCs) from differentiated cells is accompanied by global 3′UTR shortening[25]. The only exceptional case was the reprogramming of spermatogonial cells into iPSCs that involved 3′UTR lengthening, which is consistent with the observation that gonads are characterized by extreme short 3′UTRs[25].

Proliferation. Perhaps the most profound association of extensive APA modulation is the one observed with cellular proliferation. First, a widespread shift towards usage of proximal poly(A) sites was observed during T cell activation[13]. As T cell activation is accompanied with a dramatic increase in cell proliferation, and as a similar effect was also detected when human B cells and monocytes were stimulated, it was hypothesized that enhanced APA is linked to the proliferative state of cells. In agreement with this, analysis of a large panel of diverse human tissues and cell lines demonstrated a substantial anti-correlation between proliferation and 3′UTR length indices. This link between APA and proliferation can potentially explain the observed changes in APA during embryonic development and during the generation of pluripotent iPSCs, as development and differentiation are associated with decreased proliferation capacity.

Further support for the strong coupling between cell proliferation and APA was recently provided by a study that used 3′-seq to explore APA events associated with cellular proliferation in two human cellular models, each examined under proliferative and arrested conditions[26]. In both cell systems, broad induction of proximal poly(A) sites was observed in proliferation. Furthermore, this induction resulted not only in a broad 3′UTR shortening but also in a substantial enhancement of cleavage at intronic poly(A) sites.

Neuron activation. One of the first examples of APA modulation was identified in the gene brain-derived neurotrophic factor (*BDNF*). *BDNF* is induced by cortical neurons and is required for survival of striatal neurons in the brain. Two alternative poly(A) sites within the *BDNF* 3′ UTR generate short and long 3′UTR isoforms[27]. The distinct *BDNF* 3′UTRs differentially regulate BDNF translation in brain compartments and during neuronal activation, providing a unique means to control precisely the temporal and spatial production of BDNF within neurons[28]. Another study used expression arrays to identify activity-regulated genes that might control synapse development[29]. Rat hippocampal neurons were stimulated by extracellular potassium chloride, which leads to membrane depolarization and calcium influx, and were subjected to transcriptome profiling. One striking and unexpected feature of the activity-regulated genetic programme revealed by this study was that a group of genes that was induced by the stimulatory treatment exhibited a switch in poly(A) site usage. This switch favoured the production of shorter mRNAs that use internal sites, leading to the production of truncated mRNAs that may have different functions from their full-length counterparts.

Future directions for the role of APA. It is important to note that, despite the remarkable progress in our recognition of biological processes that involve extensive APA modulation, we still do not have a good understanding of the biological roles of these APA events. Future examination should clarify, for example, what the functional importance is of the progressive 3′UTR lengthening that accompanies differentiation and development and of the global 3′UTR shortening that accompanies cellular proliferation.

Mechanistic regulation of APA

The link between various biological processes and extensive APA modulation indicates that poly(A) site selection is under active control. Indeed, in recent years, we witnessed a substantial advance not only in recognizing the extent of APA but also in discovering underlying regulatory mechanisms. Below, we discuss key emergent principles (FIG. 4).

Box 1 | **Identifying APA events using sequencing based approaches**

Readers are referred to a recent review[16] for a more detailed discussion of sequencing-based methods.

3′ end RNA-seq. In this modified RNA sequencing (RNA-seq) protocol, polyadenylated transcripts are randomly fragmented, their 3′ ends are specifically selected by reverse transcription using an anchored oligo(dT), and sequencing adaptors are attached to generate a library for next-generation sequencing[75].

Poly(A) site sequencing (PAS-seq). This approach uses a custom sequencing primer that extends through the poly(A) site and therefore allows sequencing in the reverse orientation; this improves reads quality and cleavage site identification[22].

Poly(A)-seq. This approach is similar to PAS-seq but differs in its use of random priming for second-strand synthesis, allowing rapid library generation[76].

Sequencing APA sites (SAPAS). This method uses an amplification step that generates mutations in the poly(A) stretch and therefore allows sequencing from this region[67].

3′-seq. This was developed as a dedicated 3′ end RNA-seq protocol to quantify gene expression in poorly conserved archival tumour samples[77], and an adaptation of this method (3′-seq) was used to study APA[56].

A-seq. This method uses a stem–loop-anchored oligo(dT) primer to reduce the poly(A) tail in the final library[6].

The above sequencing protocols suffer from false-positive calls of poly(A) sites derived from internal priming of oligo(dT) primers. The following protocols have been introduced to circumvent this limitation without using computational filters.

Poly(A) position profiling by sequencing (3P-seq). This approach uses a split ligation to favour the capture of real 3′ ends[78].

3′ region extraction and deep sequencing (3′READS). This method was recently applied to reduce false positives significantly by capturing polyadenylated fragments with a special oligo(dT) containing 45 Ts[18].

Direct RNA sequencing (DRS). This method uses an independent sequencing platform that uses oligo(dT)-coated flow cells to sequence RNA directly from the 3′ end without any previous manipulation[79]. Despite the use of an oligo(dT), the authors reported no or very few poly(A) sites derived from internal priming.

APA, alternative polyadenylation.

Box 2 | **Experimental approaches to demonstrate APA regulation**

Besides alternative polyadenylation (APA), changes in mRNA stability driven by *trans*-acting factors, such as microRNAs and RNA-binding proteins (RBPs), can result in global changes in the expression of 3′ end isoforms that are identifiable by the methods in BOX 1. Therefore, to ascribe the observed changes in relative isoform levels to APA regulation, more direct evidence for this mode of gene regulation should be provided.

Experimental methods for indicating involvement of genuine APA

Total expression level of all isoforms of the target gene. One indication for APA modulation is given when the expression level of the total sum of all 3′ end isoforms of a gene — measured by isoform-specific northern blots, microarrays, specific reverse-transcription PCR (RT-PCR) analyses or dedicated 3′ end deep sequencing approaches — remains unaltered or, assuming mainly repressive role for 3′ untranslated regions (UTRs), changes in a way that is not consistent with higher stability of the longer isoform. Decreases in the overall level might indicate destabilization of the longer transcript as the underlying mechanisms.

Transcript stability. In cell-based systems, isoform stability can be directly measured by applying actinomycin D, which is an inhibitor of RNA polymerase II (RNA Pol II), or using 4-thiouridin[80] to rule out the possibility that changes in stability underlie the relative change in proximal or distal levels.

RNA Pol II occupancy. If APA is involved, consistent changes in RNA Pol II occupancy should occur, which can be measured by an RNA Pol II chromatin immune-precipitation (ChIP) protocol. For example, when the usage of a proximal poly(A) site is increased, less RNA Pol II occupancy should be detected downstream of it.

Cleavage assays. A more direct measurement of poly(A) site usage can be done *in vitro* using cleavage assay using nuclear extracts from the examined cells or tissues and RNA probes spanning the poly(A) site cleavage region[56].

Regulation by 3′-end-processing factors. One emerging APA regulatory theme is the negative correlation between 3′UTR length and expression level of the genes encoding proteins that are a part of, or proteins that associate with, the 3′-end-processing machinery. The progressive 3′UTR lengthening observed during mouse embryonic development was accompanied by a downregulation of 3′-end-processing genes[20]. Conversely, widespread 3′UTR shortening that was detected both in reprogramming of somatic cells into stem cells[25] and during the transition of arrested cells into proliferative condition[26] was accompanied by substantial upregulation of this set of genes. These observations indicate that global alterations in the activity of the poly(A) machinery provide one means for modulating APA. Recently, a role for E2F — a transcription factor involved in cell cycle regulation — was demonstrated in the transcriptional regulation of 3′-end-processing genes[26], providing one mechanistic link between APA and proliferation (FIG. 4a).

In addition to global co-transcriptional regulation of 3′-end-processing genes, more focused control of specific polyadenylation factors also contributes to APA regulation. A prominent example of this type of regulation is provided by immunoglobulin M (IgM). IgM is subjected to regulation by intronic APA, in which usage of the intronic poly(A) site generates an mRNA that encodes a secreted form of IgM, whereas cleavage at the distal site produces a membrane-bound IgM. It was demonstrated that this regulated poly(A) selection is controlled during B cell maturation in part by the concentration of the polyadenylation factor cleavage stimulation factor subunit 2 (CSTF2): low CSTF2 concentrations in early stages of B cell development favour cleavage at the canonical and stronger IgM poly(A) site, whereas high CSTF2 levels in activated B cells induce cleavage at the weaker proximal site[30,31]. In agreement with this, a recent transcriptome-wide study carried out in HeLa cells observed that co-depletion of CSTF2 and its paralogue CSTF64τ resulted in increased usage of distal poly(A) sites[32], indicating that when these factors are abundant, they increase the efficiency of cleavage at proximal sites.

Somewhat surprisingly, the opposite effect is observed for CFIm, although, like CSTF2, it is a component of the 3′ end cleavage and polyadenylation machinery. A depletion in the levels of the 25 kDa subunit of CFIm (namely, CFIm25, which is encoded by the gene *NUDT21*) results in preferential usage of 3′UTR proximal poly(A) sites, suggesting that this factor represses cleavage at those sites. A repressive role for CFIm25 in poly(A) site selection was first identified using systematic evolution of ligands by exponential enrichment (SELEX; an *in vitro* selection approach of RNA aptamers)[12]; this role was then demonstrated on several target transcripts[33], extended to the CFIm68 subunit[34] and was recently demonstrated on a transcriptomic scale, in which knockdown of CFIm68 resulted in a systematic enhancement of cleavage at proximal poly(A) sites[6]. Thus, CFIm might have not only a general role in 3′ end processing but may also have a regulatory role in poly(A) site selection. It still remains to be seen whether alterations in CFIm levels is used as a means to control APA *in vivo*.

Another layer of APA regulation is used by the *cis*-regulatory signals that are identified by the 3′-end-processing factors. First, as reviewed above, distal poly(A) sites preferentially use the canonical PAS, whereas the prevalence of PAS variants increases at proximal poly(A) sites. Therefore, distal sites are generally stronger than proximal ones. In addition, notably, ~20% of the human poly(A) sites do not contain a canonical PAS or any of its variants. It was demonstrated that 3′ end processing in such sites relies on strong binding either of the CFIm to the upstream UGUA element[35] or of CSTF to downstream element (DSE) signals together with presence of additional auxiliary elements[36].

Interplay between transcription and APA. The functional coupling between transcription and 3′ end processing is well documented and comprises many physical associations of RNA polymerase II (RNA Pol II) and other transcription factors with poly(A) regulatory proteins (reviewed in REFS 37,38 and also see REF. 39). Yet the impact of this link on APA regulation is just beginning to be elucidated. So far, two main principles have emerged.

First, it has been suggested that the interaction between the transcription and 3′-end-processing machineries increases poly(A) tail cleavage efficiency and thus also increases usage of proximal sites[40]. In support of this model, it was recently shown that the transcription elongation complex PAF1C is required for enhancement of 3′ end processing[41] and that knocking down the CDC73 (also known as parafibromin) subunit of PAF1C results in an increase in the relative levels of an isoform with a longer 3′UTR of one of the CDC73

Immunoglobulin
(Ig). An antigen receptor molecule produced by B cells that consists of two heavy chains and two light chains.

Systematic evolution of ligands by exponential enrichment
(SELEX). In the context of RNA, this is a method for identifying consensus protein-binding sequences on RNA substrates by *in vitro* selection of short RNAs that bind preferentially to RNA-binding proteins.

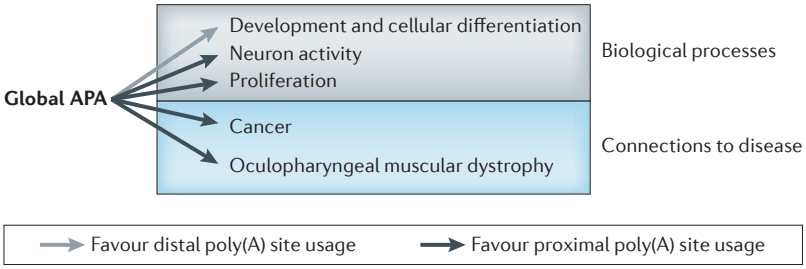

Figure 3 | **Biological processes that have been linked with broad APA modulation.** A schematic showing the biological processes and diseases that alternative polyadenylation (APA) has been linked with. In addition, the tendency towards distal or proximal poly(A) site usage is shown.

targets, indicating reduced cleavage at the proximal poly(A) site[42]. Similarly, ELL2 — another transcription elongation factor — was suggested to promote loading of the polyadenylation factor CSTF on the transcription machinery, thereby enhancing usage of the proximal poly(A) site of IgM[43], providing an additional mechanism for the switch of membrane-bound IgM to the secreted IgM form. Furthermore, a global positive correlation between gene expression level and relative usage of proximal poly(A) sites was observed in the human and mouse transcriptomes, and it was demonstrated using reporter assays that enhancement of transcriptional activity results in increased cleavage at proximal sites[44].

The second emerging principle for the interplay between transcription and APA involves kinetic coupling (FIG. 4b). As the proximal poly(A) sites are transcribed first and are therefore encountered first by the 3′-end-processing machinery, they have an advantage for being used over distal poly(A) sites[17]. That is, use of proximal poly(A) sites should positively correlate with the distance between consecutive poly(A) sites and should negatively correlate with transcription elongation rate. In accordance with this expectation, using a *D. melanogaster* strain with a lower transcriptional elongation rate, it was shown that reduced RNA Pol II elongation kinetics results in increased usage of proximal poly(A) sites in a number of transcripts[45]. Of note, this kinetic coupling resembles a mechanism for alternative splicing regulation, in which slow kinetics of RNA Pol II leads to preferential inclusion of otherwise skipped alternative exons[46]. Little is known about mechanisms that regulate transcription elongation rates, and it remains to be seen whether this kinetic coupling is used to regulate APA in physiological conditions.

APA and chromatin. Recent results have suggested that chromatin and epigenetic modifications affect APA. It was observed that poly(A) sites are strongly depleted of nucleosomes, whereas regions downstream of these sites are enriched for nucleosomes[47]. To some extent, nucleosome depletion at poly(A) sites is explained by base composition of sequences in these regions, which are A- and T-rich, as poly(dA:dT) DNA stretches have a low nucleosome affinity. Interestingly, examination of genes with multiple poly(A) sites showed that stronger poly(A) sites are associated with more pronounced nucleosome depletion at the site and more pronounced enrichment downstream from it, suggesting that nucleosome positioning might influence PAS use by, for example, affecting the rate of polymerase elongation. Yet at this stage, these observations are only correlative, and experimental studies are required in order to test this model and to establish a cause–effect relationship between nucleosome occupancy and poly(A) site selection.

Another way in which chromatin has been proposed to affect APA is through DNA methylation. This epigenetic effect on APA was first suggested using mouse tissues, in two cases of retrogenes (namely, *Mcts2* and *Napl15*), which are located within the introns of host genes (namely, *H13* and *Herc3*, respectively)[48,49]. In both cases, the promoters of the retrogenes are imprinted and are therefore silenced on the maternal allele, whereas they are unmethylated and active on the paternal allele. It was shown that when the retrogene is transcribed, an upstream intronic poly(A) is used by the host gene, whereas a downstream distal poly(A) site is used by the host gene in the allele on which the retrogene is silenced. These observations support a model in which transcriptional interference affected poly(A) site choice by the host gene but cannot exclude the involvement of a polyadenylation factor (or factors) that is sensitive to the methylation status of the DNA in the vicinity of the poly(A) site.

Interplay between splicing and APA. Numerous studies have reported multiple links between the splicing and 3′-end-processing machineries and have demonstrated that physical interactions between splicing and polyadenylation factors occurring at terminal introns of precursor mRNAs (pre-mRNAs) enhance cleavage efficiency at 3′UTR poly(A) sites[50,51]. Two types of APA events are affected by the interplay between splicing and 3′ end processing — alternative terminal exons and intronic APA (FIG. 2) — and recent studies are shedding light on underlying regulatory mechanisms. A first global analysis on this interplay used EST databases to identify events of intronic polyadenylation in hundreds of human genes[52]. Importantly, these events were associated with weak 5′ splicing sites (5′ss) and long introns, suggesting a dynamic competition between splicing and polyadenylation. In agreement with this model, increased cleavage at intronic poly(A) sites was observed in conditions that were associated with increased usage of 3′UTR proximal poly(A) sites, including proliferation[26], whereas decreased intronic cleavage was observed during development and differentiation[18]. Another indication for interplay between splicing and APA regulation was provided by an RNA-seq study that examined the transcriptomes of a diverse panel of human tissues and cell lines[14]. A strong correlation between patterns of alternative splicing and APA across the probed samples was observed, suggesting coordinated regulation of these processes. Furthermore, strong enrichment of well-known splicing-related regulatory motifs was also detected in 3′UTRs, suggesting that the factors binding these motifs function in the regulation of both splicing and APA.

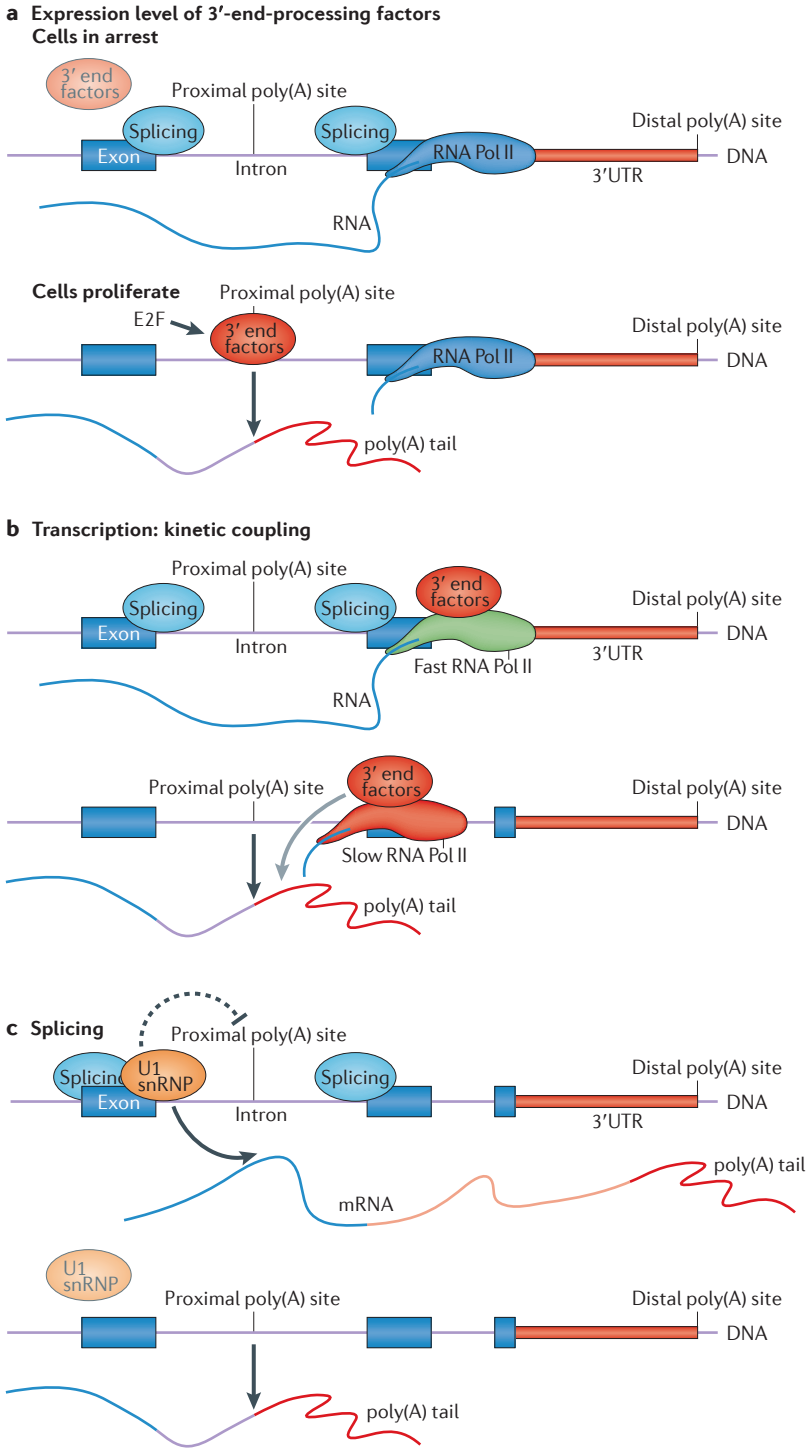

Figure 4 | **Mechanistic regulation of APA.** **a** | Low expression of genes encoding 3′-end-processing factors leads to the use of distal poly(A) sites, and intronic and proximal sites are spliced out as shown in the upper panel. Higher expression of 3′-end-processing factors may lead to enhanced cleavage and polyadenylation activity, thereby leading to elevated usage of proximal poly(A) sites (lower panel). E2F was indicated as one factor that transcriptionally co-regulates the 3′-end-processing genes and hence may lead to enhanced polyadenylation at proximal sites. **b** | Slower rate of transcription by RNA polymerase II (RNA Pol II) increases the chance of using proximal poly(A) sites. **c** | U1 small nuclear ribonucleoprotein (snRNP)-mediated interplay between precursor mRNA (pre-mRNA) splicing and alternative polyadenylation (APA). U1 induces splicing but also actively suppresses proximal poly(A) site use. Low U1 levels cause both splicing attenuation and induction in use of proximal, intronic poly(A) sites.

One such splicing factor is NOVA2, which is a neuron-specific RNA-binding protein. A genome-wide map of NOVA2–RNA interactions in the mouse brain not only confirmed previous roles of NOVA2 in the regulation of alternative splicing but also detected many NOVA2–RNA interactions in 3′UTRs, suggesting a role for NOVA2 in APA regulation[53]. Interestingly, it was found that in 9 out of 12 examined tandem 3′UTR APA events, *Nova2*-knockout mouse brains showed 3′UTR shortening compared with the wild type, suggesting that NOVA2 generally promotes the production of isoforms with longer 3′UTRs. Mechanistically, in these cases, NOVA2 binds close to the PAS of the proximal poly(A) sites, pointing to a possible interference with 3′ end cleavage by steric hindrance. Noting that brain tissue is characterized by a strong preference for usage of distal 3′UTR poly(A) sites, it was suggested that NOVA2 contributes to this tendency[53].

Another prominent example is the U1 small nuclear ribonucleoprotein (snRNP; here referred to simply as U1), which is a component of the major spliceosome (FIG. 4c). It was recently demonstrated that U1 knockdown resulted in extensive induction in usage of intronic poly(A) sites independently of the role of U1 in splicing[54]. This suggested a function for U1 in protecting pre-mRNAs from premature cleavage. A follow-up study by the same group showed that more moderate decreases in U1 levels alleviated the extent of premature intronic polyadenylation and instead resulted in a broad shift towards cleavage at 3′UTR proximal poly(A) sites[55]. Physiologically, although U1 is very abundant, and its level is not known to be regulated, it was suggested that transient U1 shortage relative to the target transcripts that it protects occurs in conditions in which overall transcription is rapidly upregulated. Such shortage should cause a shift towards cleavage at proximal poly(A) sites and thereby global 3′UTR shortening, similarly to the phenomenon observed in activated neuronal and immune cells and in proliferation and cancer. Whether such U1 shortage is involved in the enhanced APA associated with proliferation still remains to be tested.

Trans-*acting factors.* Recent studies have also discovered specific RNA-binding proteins (RBPs) that are involved in the control of APA by modulating cleavage at intronic and 3′UTR proximal poly(A) sites through their binding in close proximity to those sites. One such example is polyadenylation binding protein nuclear 1 (PABPN1). Compromised function of PABPN1, either by its knockdown or by expression of a variant mutated in patients with oculopharyngeal muscular dystrophy (OPMD), results in a broad enhancement of polyadenylation at 3′UTR proximal sites[56,57]. Mechanistically, it was shown that PABPN1 associates with proximal sites, thereby repressing their cleavage. Another RBP that was shown to regulate APA is the neuronal ELAV protein[58]. This factor was linked to the pronounced 3′UTR lengthening observed in neuronal tissues. First, ectopic expression of ELAV in non-neuronal *Drosophila melanogaster* cells induces 3′ extension, and further evidence supports

Table 1 | **Mutations in poly(A) *cis*-elements that cause or contribute to human diseases**

Disease	Disease description	Poly(A) site mutation	Affected gene	Comments	Refs
α-Thalassaemia	Thalassaemias are a group of common human genetic diseases that result from defects in haemoglobin production	AATAAA to AATAAG	*HBA2*		84
β-Thalassaemia		AATAAA to AACAAA	*HBB*	The mutation results in generation of an unstable transcript that is ~900 nt longer	85
Metachromatic leukodystrophy	A neurodegenerative disorder caused by null mutations in *ARSA*	AATAAC to AGTAAC	*ARSA*	Carriers of this mutation show reduced mRNA levels and enzyme activity but do not develop the disease symptoms; this condition is termed 'pseudodeficiency'	86,87
IPEX	A rare multifaceted and fatal disease	AATAAA to AATGAA	*FOXP3*		88
Fabry's disease	A rare and severe X-linked lysosomal storage disease caused by mutations in *GLA*	An AA dinucleotide deletion within the poly(A) site	*GLA*	*GLA* is one of the unusual mammalian genes that lack a 3′UTR and has its PAS within the CDS. The mutation in the poly(A) site results in aberrant 3′ end formation and multiple unstable transcripts	89

ARSA, arylsulfatase A; CDS, coding DNA sequence; *FOXP3*, forkhead box P3; *GLA*, alpha-galactosidase; *HBA2*, haemoglobin, alpha 2 (also known as alpha-2-globin); *HBB*, haemoglobin, beta (also known as beta-globin); IPEX, immune dysfunction, polyendocrinopathy, enteropathy, X linked; PAS, poly(A) signal; UTR, untranslated region. Only selected examples are discussed here. For a more thorough discussion, see REF. 64.

the binding of ELAV near proximal poly(A) sites and repressing their use[58]. Second, in humans, three ELAV homologues — HUB, HUC and HUD — are selectively expressed in neurons and, interestingly, they too were reported to repress cleavage at proximal poly(A) sites[59,60]. Another example is provided by polypyrimidine-tract-binding protein 1 (PTB1), which is mainly known for its role in splicing and can inhibit cleavage by competing with CSTF64 for binding to DSEs[61]. Interestingly, PTB can also augment 3′ end processing by increasing the binding of heterogeneous nuclear RNP H (HNRNPH) to auxiliary upsteam elements (USEs), and this in turn stimulates cleavage by recruiting CSTF and poly(A) polymerase-α (PAPα)[62].

A role for cytoplasmic polyadenylation element binding protein 1 (CPEB1) in APA regulation was recently reported[63]. CPEB1 is a well-known key regulator of cytoplasmic polyadenylation, which is a process in which specific cytoplasmic mRNAs, particularly transcripts encoding cell-cycle regulators, undergo poly(A) tail elongation to enhance their translational efficiency[5]. CPEB1 regulates cytoplasmic polyadenylation by binding to the *cis*-regulatory cytoplasmic polyadenylation element (CPE). CPEB1 binding to the CPE in the cytoplasm results in recruitment of CPSF to the PAS element, followed by recruitment of the cytoplasmic poly(A) polymerase GLD2. Interestingly, it was recently indicated that binding of nuclear CPEB1 to CPEs located close to proximal poly(A) sites recruits the 3′-end-processing machinery to these suboptimal sites and enhances their usage[63], thus generating isoforms with shorter 3′UTRs. This dual function of CPEB1 in regulating both nuclear and cytoplasmic polyadenylation coordinates the regulation of specific transcripts at the levels of mRNA processing and protein translation. It was also indicated that when the CPE is located near an alternative splice site, binding of the CPEB1–CPSF complex to the CPE can interfere with the recruitment of the splicing machinery to that site, thereby providing another example of the splicing–APA interplay.

Spliceosome
A ribonucleoprotein complex that is involved in splicing nuclear precursor mRNA (pre-mRNA). It is composed of five small nuclear ribonucleoproteins (snRNPs) and more than 50 non-snRNPs, which recognize and assemble on exon–intron boundaries to catalyse intron processing of the pre-mRNA.

Alternative polyadenylation and disease

APA misregulation in human disease. Mutations in PASs and other poly(A) *cis*-elements, and the consequent concomitant alteration in gene expression, have been shown to cause or to contribute to the development of several human genetic diseases. Several key examples are summarized in TABLE 1 (for a thorough review, see REF. 64). Moreover, in recent years, pathological conditions that are associated with extensive misregulation of APA are being uncovered. As discussed above, alanine expansion mutations in *PABPN1* that cause OPMD result in global induction of proximal poly(A) sites[56,57]. Global enhancement of cleavage at proximal poly(A) sites was also detected in conditions of cardiac hypertrophy, which is an enlargement of the heart that occurs in response to physiological or pathological stimuli[65].

Uncontrolled cellular proliferation is a hallmark of cancer development. As global APA modulation was reported during induced proliferation, it would be expected that APA be associated with cancer development too. Indeed, widespread increase in the use of proximal poly(A) sites was observed in cancer cells[66]. This induction was over and above the effect expected from their proliferative status, as cancer cell lines also showed enhanced APA when compared with non-transformed cell lines with similar proliferation rates. This study provided evidence that switching to shorter 3′UTRs allows proto-oncogenes to escape from inhibition by miRNAs, thereby enhancing their tumorigenic activity. Consistent with this, global induction of the use of proximal poly(A) sites was observed in a comparison between the breast cancer cell line MCF7 and the immortalized but non-transformed mammary epithelial cell line MCF10A[67]. Yet this study also indicated that the relationship between APA and cancer transformation is complicated, as a similar trend was not detected in the comparison between MCF10A and a different breast cancer cell line, namely, MDA-MB231. A more controlled analysis compared APA modulation

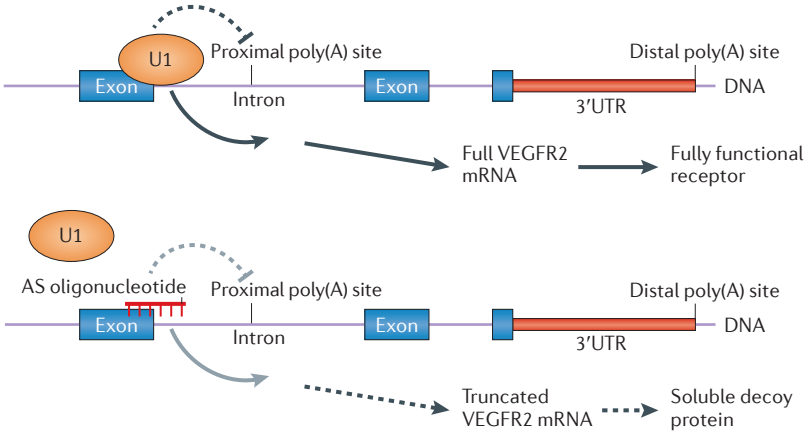

Figure 5 | **Therapeutic potential for intervention with APA.** A potential therapeutic implication for site-specific alternative polyadenylation (APA) modulation was recently demonstrated[72]. Vascular endothelial growth factor receptor 2 (*VEGFR2*) undergoes APA when either a fully functional receptor (that has been generated by use of the distal poly(A) site) or a soluble decoy protein (that has been generated from a proximal intronic site) are generated. As VEGFR2 mediates important angiogenic signalling in endothelial cells, inhibitory antisense oligonucleotides to the nearest 5′ splicing site upstream of the intronic poly(A) signal (PAS) results in increased expression of a soluble decoy VEGFR2 protein and blockade of angiogenic signals. U1, U1 small nuclear ribonucleoprotein (U1 snRNP); UTR, untranslated region.

in MCF10A cells examined under normal proliferation, arrested and transformed states[26]. In this model system, the effect of proliferation on APA was stronger than the effect imposed by oncogenic transformation.

Recently, comprehensive polyadenylation maps were generated in five tumour–normal pairs (including breast, colon, kidney, liver and lung tumours)[68]. Preferential cleavage at 3′UTR proximal poly(A) sites in the cancer samples was observed. Furthermore, APA modulation was also measured during colon cancer progression (from normal to adenoma to carcinoma), and several strong APA events were identified and validated[69], although in this study, no simple correlation between cancer progression and 3′UTR shortening was observed. One of the strongest effects observed in this study was a progressive increase in the usage of the proximal poly(A) site in the 3′UTR of pyridoxal kinase (PDXK) during progression of normal colon tissue to carcinomas. PDXK was recently shown to sensitize lung cancer cells to apoptosis induction by chemotherapeutics drugs[70]. It was therefore suggested that both the mRNA level and the proximal-to-distal poly(A) site usage ratio of PDXK can serve as a biomarker in lung and colon cancer. Another interesting link between cancer and APA was also recently reported[71]. This report showed that oestrogen treatment of oestrogen-receptor-positive breast cancer cells caused 3′UTR shortening of the CDC6 regulator of DNA replication, allowing this gene to evade 3′UTR-mediated repression and resulting in elevated CDC6 protein levels.

Given the high prevalence of APA, it can be expected that the involvement of its misregulation in the development of additional pathological conditions will be discovered in the coming years.

Angiogenesis
The formation of new blood vessels from pre-existing ones. It is often associated with cell division and the subsequent sprouting of the endothelial cells that contribute to the growing blood vessel.

Potential therapeutic modulation of APA. A promise for therapies that manipulate APA was recently provided[72]. It was first demonstrated that blocking U1 enhances intronic polyadenylation in numerous genes that encode receptor tyrosine kinases (RTKs), thereby producing soluble variant RTK isoforms lacking the *trans*-membrane domain. Importantly, these soluble variants can act in a dominant-negative way to block RTK signalling, which is frequently amplified in cancer. In addition, antisense oligonucleotides that bind to a specific 5′ss of the vascular endothelial growth factor receptor 2 (*VEGFR2*) pre-mRNA — a gene involved in angiogenesis — were developed to interfere with the binding of U1 to this location. The use of these oligonucleotides resulted in increased cleavage at nearby downstream intronic poly(A) sites and, consequently, in elevated production of soluble decoy VEGFR2 protein and attenuation of angiogenic signals (FIG. 5).

Conclusions and perspectives

Despite the major advance in our understanding of APA regulation, the crucial issues in this field are still open and require further research efforts. First, at the molecular level, modulating the 3′UTRs of genes is expected to have an impact on transcript stability, translation efficiency or cellular localization. Notably, so far, in most of the conditions in which global 3′UTR shortening or lengthening has been observed, a correlated change in the expression level of the affected genes has not been observed, indicating that for most of the affected genes, APA did not substantially change transcript stability[13,26]. One possible explanation for this observation is that only a minority of the affected genes contains regulatory sites for repressive factors (for example, miRNAs or RBPs) that are expressed or active under the examined conditions. The effect of APA on translational efficiency has so far not been systematically examined, but the advent of ribosome-profiling technique[73] now allows carrying out such examination on a global scale. Techniques for comprehensively monitoring the impact of APA on transcript localization are still not available.

Second, although novel deep-sequencing techniques tremendously extended the number of annotated poly(A) sites in eukaryote genomes and discovered extensive APA regulation in various conditions, our understanding of the biological importance of this phenomenon is still rudimentary. For example, causal APA events that affect proliferation, development and differentiation are still to be discovered and established in physiological settings. Furthermore, more than 70% of the human genes are potentially subjected to APA regulation as they contain multiple poly(A) sites. However, how many of these genes are actually regulated by APA in a way that has a substantial biological effect is still an open question that needs to be addressed in the future before we can appreciate not only the extent of APA but also its biological importance.

Last, given the high prevalence of APA, it can be expected that many more diseases associated with

APA misregulation will be detected in the coming years. Therefore, it is of great interest to explore further the therapeutic potential of external manipulation of APA. Antisense oligonucleotide-based methods that modulate alternative splicing are currently under clinical testing[74], and recent results hold promise that similar methods could be used to interfere with pathological conditions by manipulating APA[72].

1. Edmonds, M. & Abrams, R. Polynucleotide biosynthesis: formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei. *J. Biol. Chem.* **235**, 1142–1149 (1960).
2. Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes Dev.* **25**, 1770–1782 (2011).
3. Sachs, A. The role of poly(A) in the translation and stability of mRNA. *Curr. Opin. Cell Biol.* **2**, 1092–1098 (1990).
4. Guhaniyogi, J. & Brewer, G. Regulation of mRNA stability in mammalian cells. *Gene* **265**, 11–23 (2001).
5. D'Ambrogio, A., Nagaoka, K. & Richter, J. D. Translational control of cell growth and malignancy by the CPEBs. *Nature Rev. Cancer* **13**, 283–290 (2013).
6. 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).
7. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* **79**, 351–379 (2010).
8. Andreassi, C. & Riccio, A. To localize or not to localize: mRNA fate is in 3′UTR ends. *Trends Cell Biol.* **19**, 465–474 (2009).
9. Gautheret, D., Poirot, O., Lopez, F., Audic, S. & Claverie, J. M. Alternate polyadenylation in human mRNAs: a large-scale analysis by EST clustering. *Genome Res.* **8**, 524–530 (1998).
10. Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M. & Gautheret, D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* **10**, 1001–1010 (2000).
11. Tian, B., Hu, J., Zhang, H. & Lutz, C. S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **33**, 201–212 (2005).
12. Brown, K. M. & Gilmartin, G. M. A mechanism for the regulation of pre-mRNA 3′ processing by human cleavage factor Im. *Mol. Cell* **12**, 1467–1476 (2003).
13. 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).
This paper presents the first genomic demonstration of the association between APA and proliferation.
14. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
15. Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
16. Sun, Y., Fu, Y., Li, Y. & Xu, A. Genome-wide alternative polyadenylation in animals: insights from high-throughput technologies. *J. Mol. Cell. Biol.* **4**, 352–361 (2012).
17. Shi, Y. Alternative polyadenylation: new insights from global analyses. *RNA* **18**, 2105–2117 (2012).
18. Hoque, M. *et al.* Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nature Methods* **10**, 133–139 (2012).
19. Zhang, H., Lee, J. Y. & Tian, B. Biased alternative polyadenylation in human tissues. *Genome Biol.* **6**, R100 (2005).
20. 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).
21. Hilgers, V. *et al.* Neural-specific elongation of 3′ UTRs during *Drosophila* development. *Proc. Natl Acad. Sci. USA* **108**, 15864–15869 (2011).
22. Shepard, P. J. *et al.* Complex and dynamic landscape of RNA polyadenylation revealed by PAS-seq. *RNA* **17**, 761–772 (2011).
23. Smibert, P. *et al.* Global patterns of tissue-specific alternative polyadenylation in *Drosophila*. *Cell Rep.* **1**, 277–289 (2012).
24. Ulitsky, I. *et al.* Extensive alternative polyadenylation during zebrafish development. *Genome Res.* **22**, 2054–2066 (2012).
25. Ji, Z. & Tian, B. Reprogramming of 3′ untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS ONE* **4**, e8419 (2009).
26. Elkon, R. *et al.* E2F mediates enhanced alternative polyadenylation in proliferation. *Genome Biol.* **13**, R59 (2012).
27. Timmusk, T. *et al.* Multiple promoters direct tissue-specific expression of the rat *BDNF* gene. *Neuron* **10**, 475–489 (1993).
28. Lau, A. G. *et al.* Distinct 3′UTRs differentially regulate activity-dependent translation of brain-derived neurotrophic factor (BDNF). *Proc. Natl Acad. Sci. USA* **107**, 15945–15950 (2010).
29. Flavell, S. W. *et al.* Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* **60**, 1022–1038 (2008).
30. Takagaki, Y. & Manley, J. L. Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Mol. Cell* **2**, 761–771 (1998).
31. 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).
32. Yao, C. *et al.* Transcriptome-wide analyses of CstF64–RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc. Natl Acad. Sci. USA* **109**, 18773–18778 (2012).
33. Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A. & Handa, H. Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3′-UTRs. *Nucleic Acids Res.* **34**, 6264–6271 (2006).
34. Kim, S. *et al.* Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes Cells* **15**, 1003–1013 (2010).
35. Venkataraman, K., Brown, K. M. & Gilmartin, G. M. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev.* **19**, 1315–1327 (2005).
36. 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).
37. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416**, 499–506 (2002).
38. Perales, R. & Bentley, D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell* **36**, 178–191 (2009).
39. Shi, Y. *et al.* Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol. Cell* **33**, 365–376 (2009).
40. Di Giammartino, D. C., Nishida, K. & Manley, J. L. Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* **43**, 853–866 (2011).
41. Nagaike, T. *et al.* Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol. Cell* **41**, 409–418 (2011).
42. Rozenblatt-Rosen, O. *et al.* The tumor suppressor CDC73 functionally associates with CPSF and CstF 3′ mRNA processing factors. *Proc. Natl Acad. Sci. USA* **106**, 755–760 (2009).
43. Martincic, K., Alkan, S. A., Cheatle, A., Borghesi, L. & Milcarek, C. Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. *Nature Immunol.* **10**, 1102–1109 (2009).
44. Ji, Z. *et al.* Transcriptional activity regulates alternative cleavage and polyadenylation. *Mol. Syst. Biol.* **7**, 534 (2011).
This paper provides a nice demonstration of the interplay between APA and transcription.
45. Pinto, P. A. *et al.* RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO J.* **30**, 2431–2444 (2011).
46. de la Mata, M. *et al.* A slow RNA polymerase II affects alternative splicing *in vivo*. *Mol. Cell* **12**, 525–532 (2003).
47. 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).
48. Wood, A. J. *et al.* A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. *PLoS Genet.* **3**, e20 (2007).
49. Cowley, M., Wood, A. J., Bohm, S., Schulz, R. & Oakey, R. J. Epigenetic control of alternative mRNA processing at the imprinted Herc3/Nap1l5 locus. *Nucleic Acids Res.* **40**, 8917–8926 (2012).
50. Millevoi, S. *et al.* An interaction between U2AF 65 and CF I(m) links the splicing and 3′ end processing machineries. *EMBO J.* **25**, 4854–4864 (2006).
51. Millevoi, S. & Vagner, S. Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. *Nucleic Acids Res.* **38**, 2757–2774 (2010).
52. Tian, B., Pan, Z. & Lee, J. Y. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res.* **17**, 156–165 (2007).
53. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
54. Kaida, D. *et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664–668 (2010).
55. Berg, M. G. *et al.* U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**, 53–64 (2012).
This paper characterizes the role of U1 in the interplay between APA and splicing.
56. Jenal, M. *et al.* The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* **149**, 538–553 (2012).
This paper identifies PABPN1 as a regulator of APA and provides the first link between a human genetic disorder (specifically, OPMD) and broad APA misregulation.
57. de Klerk, E. *et al.* Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res.* **40**, 9089–9101 (2012).
58. Hilgers, V., Lemke, S. B. & Levine, M. ELAV mediates 3′ UTR extension in the *Drosophila* nervous system. *Genes Dev.* **26**, 2259–2264 (2012).
59. 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).
60. Zhu, H., Zhou, H. L., Hasman, R. A. & Lou, H. Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. *J. Biol. Chem.* **282**, 2203–2210 (2007).
61. Castelo-Branco, P. *et al.* Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol. Cell. Biol.* **24**, 4174–4183 (2004).
62. Danckwardt, S. *et al.* Splicing factors stimulate polyadenylation via USEs at non-canonical 3′ end formation signals. *EMBO J.* **26**, 2658–2669 (2007).
63. Bava, F. A. *et al.* CPEB1 coordinates alternative 3′-UTR formation with translational regulation. *Nature* **495**, 121–125 (2013).
This paper shows that CPEB1, which is the key regulator of cytoplasmic polyadenylation, also regulates nuclear APA.
64. Danckwardt, S., Hentze, M. W. & Kulozik, A. E. 3′ end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.* **27**, 482–498 (2008).
65. 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).

66. 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).
This paper is the first to demonstrate an association between enhanced APA and cancer.
67. Fu, Y. *et al.* Differential genome-wide profiling of tandem 3′ UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res.* **21**, 741–747 (2011).
68. Lin, Y. *et al.* An in-depth map of polyadenylation sites in cancer. *Nucleic Acids Res.* **40**, 8460–8471 (2012).
69. Morris, A. R. *et al.* Alternative cleavage and polyadenylation during colorectal cancer development. *Clin. Cancer Res.* **18**, 5256–5266 (2012).
70. Galluzzi, L. *et al.* Prognostic impact of vitamin B_6 metabolism in lung cancer. *Cell Rep.* **2**, 257–269 (2012).
71. Akman, B. H., Can, T. & Erson-Bensan, A. E. Estrogen-induced upregulation and 3′-UTR shortening of CDC6. *Nucleic Acids Res.* **40**, 10679–10688 (2012).
72. Vorlova, S. *et al.* Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Mol. Cell* **43**, 927–939 (2011).
This is the first demonstration, to our knowledge, of the therapeutic potential for external manipulation of APA.
73. Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. *Nature Protoc.* **7**, 1534–1550 (2012).
74. Kole, R., Krainer, A. R. & Altman, S. RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nature Rev. Drug Discov.* **11**, 125–140 (2012).
75. Yoon, O. K. & Brem, R. B. Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA* **16**, 1256–1267 (2010).
76. Derti, A. *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Res.* **22**, 1173–1183 (2012).
77. Beck, A. H. *et al.* 3′-end sequencing for expression quantification (3SEQ) from archival tumor samples. *PLoS ONE* **5**, e8768 (2010).
78. Jan, C. H., Friedman, R. C., Ruby, J. G. & Bartel, D. P. Formation, regulation and evolution of *Caenorhabditis elegans* 3′UTRs. *Nature* **469**, 97–101 (2011).
79. Ozsolak, F. *et al.* Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* **143**, 1018–1029 (2010).
80. Friedel, C. C., Dolken, L., Ruzsics, Z., Koszinowski, U. H. & Zimmer, R. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res.* **37**, e115 (2009).
81. Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11**, 2755–2766 (1997).
82. Mandel, C. R., Bai, Y. & Tong, L. Protein factors in pre-mRNA 3′-end processing. *Cell. Mol. Life Sci.* **65**, 1099–1122 (2008).
83. Yang, Q., Coseno, M., Gilmartin, G. M. & Doublie, S. Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. *Structure* **19**, 368–377 (2011).
84. Higgs, D. R. *et al.* Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature* **306**, 398–400 (1983).
85. Orkin, S. H., Cheng, T. C., Antonarakis, S. E. & Kazazian, H. H. Jr. Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human β-globin gene. *EMBO J.* **4**, 453–456 (1985).
86. Gieselmann, V., Polten, A., Kreysing, J. & von Figura, K. Arylsulfatase A pseudodeficiency: loss of a polyadenylylation signal and *N*-glycosylation site. *Proc. Natl Acad. Sci. USA* **86**, 9436–9440 (1989).
87. Barth, M. L., Fensom, A. & Harris, A. Prevalence of common mutations in the arylsulphatase A gene in metachromatic leukodystrophy patients diagnosed in Britain. *Hum. Genet.* **91**, 73–77 (1993).
88. Bennett, C. L. *et al.* A rare polyadenylation signal mutation of the *FOXP3* gene (AAUAAA– > AAUGAA) leads to the IPEX syndrome. *Immunogenetics* **53**, 435–439 (2001).
89. Yasuda, M., Shabbeer, J., Osawa, M. & Desnick, R. J. Fabry disease: novel α-galactosidase A 3′-terminal mutations result in multiple transcripts due to aberrant 3′-end formation. *Am. J. Hum. Genet.* **73**, 162–173 (2003).

Acknowledgements
This work was supported by funds from the European Research Council (ERC), the Dutch cancer foundation (KWF), Horizon and the Netherlands Organisation for Scientific Research (VICI-NWO).

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Reuven Agami's homepage: http://research.nki.nl/agamilab

ALL LINKS ARE ACTIVE IN THE ONLINE PDF