Alterations in polyadenylation and its implications for endocrine disease

Anders Rehfeld[1], Mireya Plass[2], Anders Krogh[2] and Lennart Friis-Hansen[1]*

[1] Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[2] Department of Biology, The Bioinformatics Centre, University of Copenhagen, Copenhagen, Denmark

Introduction: Polyadenylation is the process in which the pre-mRNA is cleaved at the poly(A) site and a poly(A) tail is added – a process necessary for normal mRNA formation. Genes with multiple poly(A) sites can undergo alternative polyadenylation (APA), producing distinct mRNA isoforms with different 3′ untranslated regions (3′ UTRs) and in some cases different coding regions. Two thirds of all human genes undergo APA. The efficiency of the polyadenylation process regulates gene expression and APA plays an important part in post-transcriptional regulation, as the 3′ UTR contains various cis-elements associated with post-transcriptional regulation, such as target sites for micro-RNAs and RNA-binding proteins.

Implications of alterations in polyadenylation for endocrine disease: Alterations in polyadenylation have been found to be causative of neonatal diabetes and IPEX (immune dysfunction, polyendocrinopathy, enteropathy, X-linked) and to be associated with type I and II diabetes, pre-eclampsia, fragile X-associated premature ovarian insufficiency, ectopic Cushing syndrome, and many cancer diseases, including several types of endocrine tumor diseases.

Perspectives: Recent developments in high-throughput sequencing have made it possible to characterize polyadenylation genome-wide. Antisense elements inhibiting or enhancing specific poly(A) site usage can induce desired alterations in polyadenylation, and thus hold the promise of new therapeutic approaches.

Summary: This review gives a detailed description of alterations in polyadenylation in endocrine disease, an overview of the current literature on polyadenylation and summarizes the clinical implications of the current state of research in this field.

Keywords: polyadenylation, gene expression regulation, diabetes, IGF-1, pre-eclampsia, RET, antisense elements

INTRODUCTION

BACKGROUND

Almost all eukaryotic pre-messenger RNAs (pre-mRNAs) and several non-coding transcripts have poly(A) sites and are polyadenylated (Shepard et al., 2011; Derti et al., 2012; Lin et al., 2012). Several studies have mapped poly(A) sites genome-wide in humans (Tian et al., 2005; Yan and Marr, 2005; Ozsolak et al., 2010; Fu et al., 2011; Shepard et al., 2011; Derti et al., 2012; Lin et al., 2012). One of the latest found that more than two thirds of all genes have multiple poly(A) sites and can therefore undergo alternative polyadenylation (APA), producing distinct mRNA isoforms with different 3′ untranslated regions (3′ UTRs) (Derti et al., 2012). Polyadenylation and APA are emerging as important regulators of gene expression and alterations in polyadenylation have implications for many diseases (Danckwardt et al., 2008).

This review gives a detailed description of alterations in polyadenylation in endocrine disease, an overview of the current literature on polyadenylation and summarizes the clinical

Abbreviations: 3′ UTR, 3′ untranslated region; APA, alternative polyadenylation; ARE, AU-rich element; Aux-DSE, auxiliary downstream element; CBC, nuclear cap-binding complex; CF-I and CF-II, cleavage factor I and II; CGRP, calcitonin gene-related peptide; CIN, chromosome instability; CPSF, cleavage and polyadenylation specificity factor; CR-APA, coding region-APA; CstF, cleavage stimulation factor; CTD, C-terminal domain; DSE, downstream sequence element; FMR1, fragile X mental retardation 1; GRE, GU-rich element; GWAS, genome-wide association studies; HSF1, heat shock factor 1; IA-2, tyrosine phosphatase; IGF-1, insulin-like growth factor 1; IPEX, immune dysfunction, polyendocrinopathy, enteropathy, X-linked; LOH, loss of heterozygosity; MCL, mantle cell lymphoma; miRNA, micro-RNA; mRNA, messenger RNA; PABPN1, nuclear poly(A)-binding protein; PAP, poly(A) polymerase; PARP1, poly(ADP-ribose) polymerase 1; PAS, polyadenylation signal; PDGFRα, platelet-derived growth factor receptor alpha; PI4,5P2, phosphatidyli-nositol 4,5-bisphosphate; Pol-II, RNA-polymerase II; PTB, polypyrimidine tract binding protein; RBP, RNA-binding protein; RET, rearranged during transfection; Ser2, Serine 2; SERT, serotonin transporter; SF3a2, splicing factor 3a subunit 2; sFlt-1, soluble fms-like tyrosine kinase 1; siRNA, small interfering RNA; SLE, systemic lupus erythematosus; SNP, single nucleotide polymorphisms; SOCS3, suppressor of cytokine signaling 3; TB-RBP, testis brain RNA-binding protein; TCF/LEF, T-cell factor/lymphoid-enhancer factor; TCF7L2, transcription factor 7-like 2; U2AF1 and U2AF2, U2 auxiliary factor 1 and 2; USE, upstream sequence element; UTR-APA, untranslated region-APA.

implications of the current state of research in this field. We focus on the co-transcriptional, nuclear polyadenylation, as post-transcriptional and cytoplasmic polyadenylation [alterations to the already added poly(A) tail] is beyond the scope of this review.

mRNA TRANSCRIPTION AND POLYADENYLATION

Pre-mRNAs are transcribed from protein-coding genes by the RNA-polymerase II (Pol-II). The nascent pre-mRNA molecule emerging from the Pol-II goes through the co-transcriptional processes of 5′ capping, splicing, and polyadenylation before transcription is terminated. Polyadenylation consists of two steps: (1) Cleavage of the pre-mRNA at a poly(A) site and (2) Addition of an untemplated 3′ poly(A) tail to the upstream cleavage product. The specific cleavage position and the efficiency of the process depend on the interaction between trans-acting polyadenylation factors and cis-elements present in the pre-mRNAs. All eukaryotic pre-mRNAs, except some replication-dependent histone pre-mRNAs, as well as several non-coding transcripts, including micro-RNAs (miRNAs), have poly(A) sites and are polyadenylated (Shepard et al., 2011; Derti et al., 2012; Lin et al., 2012). Polyadenylation is vital for normal mRNA formation, transcription termination and mRNA export from the nucleus, and affects mRNA stability, subcellular localization, and translational efficiency (Sandberg et al., 2008; Ji and Tian, 2009). As polyadenylation is required for transcription termination, the efficiency of the polyadenylation process affects gene expression quantitatively (West and Proudfoot, 2009; Yang et al., 2009; Mapendano et al., 2010).

POST-TRANSCRIPTIONAL REGULATION OF GENE EXPRESSION

Tight regulation of gene expression is essential for cells to perform their normal functions and disturbances in gene expression underlie many diseases. Gene expression can be regulated at any step, from the pre-transcriptional epigenetic modifications of the chromatin to the post-translational modifications of the protein product. A quickly responding part of gene expression regulation is the post-transcriptional regulation exerted by RNA-binding proteins (RBPs) and miRNAs. RBPs and miRNAs mainly bind target sequences in the 3′ UTR of mRNAs, thereby controlling mRNA turnover and translation rate (Keene, 2007; Friedman et al., 2009). An example of post-transcriptional regulation is seen in the rapid regulation of insulin level and signaling in response to changes in glucose level. This regulation is exerted through altered stability and translation of insulin mRNA and insulin receptor mRNA, and helps to maintain glucose homeostasis (reviewed in Lee and Gorospe, 2010).

ROLE OF ALTERNATIVE POLYADENYLATION IN POST-TRANSCRIPTIONAL REGULATION

As RBPs and miRNAs mainly target the 3′ UTR, alterations to the 3′ UTR affect the post-transcriptional regulation. More than two thirds of all mammalian genes encode alternative mRNA isoforms with different 3′ UTRs through APA (Derti et al., 2012). APA therefore plays an important part in the post-transcriptional regulation, as it determines which isoform of the 3′ UTR is expressed (Thomsen et al., 2010). This is seen for the insulin receptor mRNA, from the example above (Levy et al., 1995). Characterization of polyadenylation is therefore essential for understanding both normal cell biology as well as disease. Recent advances in sequencing

technologies have made it possible to map poly(A) site usage genome-wide (Mangone et al., 2010; Ozsolak et al., 2010; Fox-Walsh et al., 2011; Fu et al., 2011; Jan et al., 2011; Shepard et al., 2011; Derti et al., 2012; Jenal et al., 2012; Lin et al., 2012; Martin et al., 2012; Pelechano et al., 2012; Yoon et al., 2012; Hoque et al., 2013; Wang et al., 2013; Wilkening et al., 2013). These techniques could be implemented in and improve diagnostics.

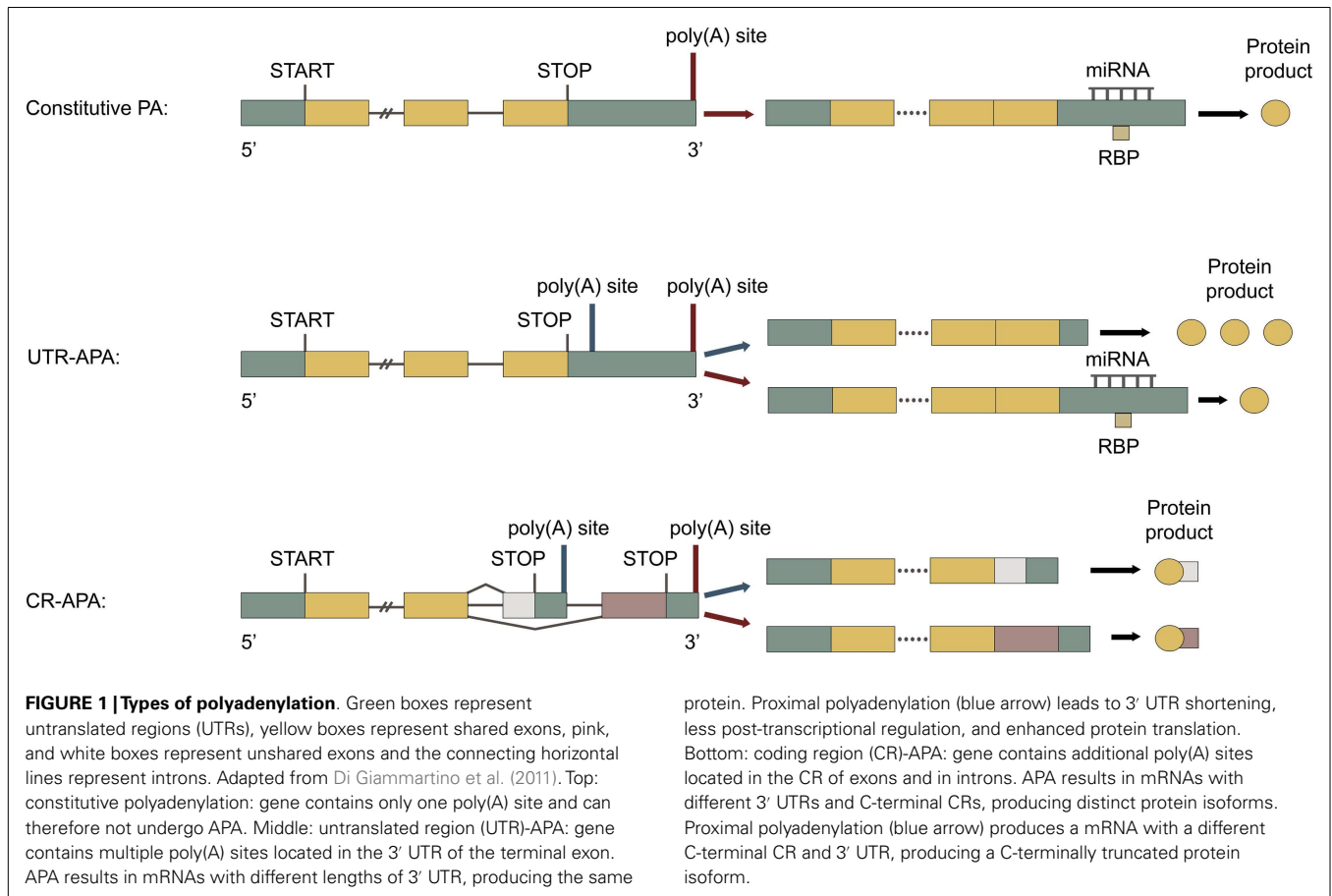## POLYADENYLATION

TYPES OF POLYADENYLATION

Constitutive polyadenylation

Genes with a single poly(A) site can only be constitutively polyadenylated (**Figure 1**, top), as seen for the *INS* gene encoding insulin (Garin et al., 2010). Approximately one third of all human genes only have one poly(A) site (Derti et al., 2012). For these genes the efficiency of polyadenylation regulates gene expression through changes in the transcription termination efficiency (West and Proudfoot, 2009; Yang et al., 2009; Mapendano et al., 2010). Weakened polyadenylation can lead to impaired gene expression and read-through transcription (Higgs et al., 1983), while enhanced polyadenylation can lead to upregulated gene expression (Danckwardt et al., 2004).

Alternative polyadenylation

For genes with multiple poly(A) sites, APA can take place (for review of APA see Neilson and Sandberg, 2010; Di Giammartino et al., 2011; Proudfoot, 2011). Only one of the possible poly(A) sites in a pre-mRNA is used for polyadenylation per mRNA transcription event. APA is typically divided into two categories, UTR-APA and coding region (CR)-APA.

Untranslated region alternative polyadenylation. Alternative polyadenylation occurring at alternative poly(A) sites located in the 3′ UTR of the last exon is called UTR-APA. It results in mRNAs with the same CR, but with different 3′ UTR length (**Figure 1**, middle), as seen for the *INSR* gene encoding insulin receptor (Levy et al., 1995). UTR-APA is the most abundant type of APA, accounting for more than half of the APA events (Yan and Marr, 2005; Shepard et al., 2011). About 70% human genes have multiple poly(A) sites in their 3′ UTR and undergo UTR-APA (Derti et al., 2012). The 3′ UTR contains various cis-elements associated with post-transcriptional gene regulation, such as target sites for miRNAs and RBPs, as well as AU-rich elements (AREs) and GU-rich elements (GREs). The 3′ UTR and the factors interacting with it largely determine mRNA stability, subcellular localization, and translational efficiency (Sandberg et al., 2008; Ji and Tian, 2009). It is therefore not surprising, that mutations in the 3′ UTR are associated with many diseases (reviewed in Chen et al., 2006a,b). The general length of the 3′ UTRs has been found to be inversely correlated with cellular proliferation (Sandberg et al., 2008; Elkon et al., 2012). Individual mRNAs with shorter 3′ UTRs are more stable (Mayr and Bartel, 2009; Hogg and Goff, 2010; Yepiskoposyan et al., 2011) and generally produce more protein (Sandberg et al., 2008; Mayr and Bartel, 2009; Singh et al., 2009). In one study 45% of the mRNAs with a general 3′ UTR shortening also had significantly changed expression levels, the majority being upregulated (de Klerk et al., 2012). This would correlate well with a heightened

FIGURE 1 | Types of polyadenylation. Green boxes represent untranslated regions (UTRs), yellow boxes represent shared exons, pink, and white boxes represent unshared exons and the connecting horizontal lines represent introns. Adapted from Di Giammartino et al. (2011). Top: constitutive polyadenylation: gene contains only one poly(A) site and can therefore not undergo APA. Middle: untranslated region (UTR)-APA: gene contains multiple poly(A) sites located in the 3′ UTR of the terminal exon. APA results in mRNAs with different lengths of 3′ UTR, producing the same protein. Proximal polyadenylation (blue arrow) leads to 3′ UTR shortening, less post-transcriptional regulation, and enhanced protein translation. Bottom: coding region (CR)-APA: gene contains additional poly(A) sites located in the CR of exons and in introns. APA results in mRNAs with different 3′ UTRs and C-terminal CRs, producing distinct protein isoforms. Proximal polyadenylation (blue arrow) produces a mRNA with a different C-terminal CR and 3′ UTR, producing a C-terminally truncated protein isoform.

mRNA stability for these transcripts. However, no such significant change in mRNA levels was found in several other studies, where general changes in 3′ UTR length were also seen (Sandberg et al., 2008; Fu et al., 2011; Elkon et al., 2012; Morris et al., 2012). The effects of general changes in 3′ UTR length thus remain unclear. During normal development, escape of miRNA-mediated regulation by APA is seen (Thomsen et al., 2010; Boutet et al., 2012). In addition escape of both miRNA- and RBP-mediated regulation by APA is seen in cancer cells (Boutaud et al., 2003; Sandberg et al., 2008; Mayr and Bartel, 2009). Genes with UTR-APA have common parts of the 3′ UTR expressed in all mRNA isoforms and alternative parts of the 3′ UTR, only expressed in the mRNAs with longer 3′ UTR isoforms. The alternative parts of the 3′ UTR are normally longer, more AU-rich and contain more cis-elements than the common parts (Sandberg et al., 2008; Ji and Tian, 2009). Poly(A) sites preferentially flank cis-elements in the 3′ UTR (Yoon et al., 2012), as seen for miRNA target sites (Ji and Tian, 2009). Interestingly, more than half of the miRNA targets are found downstream of the most proximal poly(A) site in the 3′ UTR (Legendre et al., 2006). The energy expenditure required for synthesizing the longer 3′ UTR isoforms, is used to provide a more refined control of gene expression, through post-transcriptional regulation, as seen for StAR mRNA (Zhao et al., 2005). Surprisingly, the 3′ UTRs of a large number of genes are also expressed alone, separately from their associated protein-coding sequences (Mercer et al., 2011). These expressed 3′ UTRs are suggested both to have trans-acting roles and to function as decoys, which can sequester trans-acting factors, such as miRNAs. UTR-APA in these transcripts would therefore affect these functions, such as generating multiple isoforms of a decoying 3′ UTR, each with different miRNA-binding profiles. For genes undergoing UTR-APA, the efficiency of polyadenylation contributes to gene expression through the transcription termination efficiency. In addition to this the choice of different poly(A) sites through APA plays an important part in post-transcriptional regulation, as it defines which isoform of the 3′ UTR is expressed (Thomsen et al., 2010). Dysfunctional polyadenylation can lead to changes in the APA pattern, altering the 3′ UTR, and to changes in polyadenylation efficiency, both affecting gene expression.

Coding region alternative polyadenylation. Alternative polyadenylation occurring at alternative poly(A) sites located in introns or in the CR of exons is called CR-APA. It results in mRNAs with both different C-terminal CRs and 3′ UTRs (**Figure 1**, bottom), as seen for the *CALCA* gene encoding calcitonin/CGRP (Amara et al., 1984; Zhou et al., 2007). CR-APA often occurs together with alternative splicing of the last exon and it is suggested that there is a dynamic competition between splicing and polyadenylation (Tian et al., 2007; Evsyukova et al., 2012). About 20% of human genes have at least one utilized intronic poly(A) site, usually located in large introns with weak 5′ splice sites (Tian et al., 2007). Thousands

of dormant intronic poly(A) sites have been characterized, which are normally suppressed and therefore not utilized (Yao et al., 2012a). Enhanced usage of intronic poly(A) sites is associated with proliferation (Elkon et al., 2012) and with a general transcriptional upregulation (Berg et al., 2012). Polyadenylation within the CR of an exon can convert a Tyr codon to a stop codon and thereby give rise to a functional mRNA (Yao et al., 2012b). Unlike mRNAs with acquired premature termination codons, which also generate C-terminal truncated proteins, the mRNAs generated by CR-APA are not targeted by nonsense-mediated decay (Lejeune and Maquat, 2005). For genes undergoing CR-APA, the efficiency of polyadenylation contributes to gene expression through the transcription termination efficiency. Additionally, the choice of different poly(A) sites through APA determines both the mRNA coding potential and 3′ UTR isoform. Dysfunctional polyadenylation can lead to changes in the APA pattern, altering the protein product and 3′ UTR, and to changes in polyadenylation efficiency, both affecting gene expression.

CIS-ELEMENTS AND THE TRANS-ACTING POLYADENYLATION FACTORS
Cis-elements guiding polyadenylation
The cis-elements guiding polyadenylation of the nascent pre-mRNA are positioned upstream or downstream relative to a poly(A) site (**Figure 2**) (reviewed in Tian and Graber, 2012). The poly(A) site is where the pre-mRNA molecule is cleaved, usually immediately 3′ of a CA dinucleotide, and where the poly(A) tail on the mature mRNA starts. Poly(A) sites normally contain two core cis-elements: (1) The polyadenylation signal (PAS) placed 10–30nt upstream of the poly(A) site, consisting of the canonical sequence element, AAUAAA, found in more than half of mammalian genes, or a weaker non-canonical variant (Beaudoing et al., 2000; Zhang et al., 2005; Ho and Gunderson, 2011; Wang et al., 2013). (2) The U/GU-rich downstream sequence element (DSE), located up to 30nt downstream of the poly(A) site. Additional elements such as U-rich upstream sequence elements (USE), located just upstream of the PAS (Moreira et al., 1998), G-rich auxiliary downstream elements (Aux-DSE) (Chen and Wilusz, 1998), located downstream of the DSE, and UGUA motifs (Brown and Gilmartin, 2003; Venkataraman et al., 2005; Yang et al., 2011) are found around some poly(A) sites and all act as enhancers of polyadenylation. The strength of a poly(A) site depends on the exact sequence of its surrounding cis-elements, as well as their spatial placement, which together determinates the affinity for the trans-acting factors. In genes with only one poly(A) site, the PAS is typically canonical (Beaudoing et al., 2000). In genes with multiple poly(A) sites, the most distal poly(A) site generally has a canonical PAS whereas the proximal poly(A) sites generally have a non-canonical PAS (Beaudoing et al., 2000; Tian et al., 2005, 2007; Shepard et al., 2011; de Klerk et al., 2012; Lin et al., 2012; Yoon et al., 2012; Wang et al., 2013). The proximal poly(A) sites of skipped terminal exons however tend to be associated with a canonical PAS as well (Tian et al., 2007). Poly(A) sites with non-canonical cis-elements are weaker and require additional factors for efficient utilization (Fu et al., 2011), as seen for the weak proximal poly(A) site in the *CALCA* gene encoding calcitonin/CGRP (Van Oers et al., 1994; Lou et al., 1998). The distal poly(A) sites with canonical cis-elements work as a secure last option for polyadenylation, ensuring

that transcription is terminated. This minimizes potentially damaging read-through transcription into downstream genes, which normally occurs for a small proportion of transcripts (Dresser et al., 1995). The arrangement of the cis-elements with one A-rich element associated with one or more U-rich elements around the poly(A) site is generally conserved across eukaryotes (Millevoi and Vagner, 2010). A comparison of orthologous human, rhesus, dog, mouse, and rat genes, showed that poly(A) site usage is strikingly similar in orthologous tissues between species (Tian et al., 2005; Derti et al., 2012). The canonical PASs are positionally conserved between species (Derti et al., 2012), opposed to the non-canonical PASs, which have been found to be less conserved (Ara et al., 2006). Intronic poly(A) sites are also less conserved than poly(A) sites in the 3′ UTR of terminal exons (Tian et al., 2007). The non-canonical proximal poly(A) sites tend to have more conserved flanking regions (Wang et al., 2008), in which conserved motifs have been found, possibly explaining the relative frequent usage of these poly(A) sites (Nunes et al., 2010; Ozsolak et al., 2010).

Trans-acting polyadenylation factors (the human 3′-end processing complex)
The human 3′-end processing complex consists of more than 80 proteins (Shi et al., 2009), which are highly conserved between species (Darmon and Lutz, 2012). An important part of this complex, the core polyadenylation machinery, consists of five multi-subunit protein factors: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I (CF-I), cleavage factor II (CF-II), and poly(A) polymerase (PAP) (**Table 1**; **Figure 2**) (reviewed in Chan et al., 2011). The endonucleolytic cleavage of the pre-mRNA is made by the CPSF-73 subunit at the poly(A) site (**Figure 2**) (Mandel et al., 2006). The pre-mRNA is typically cleaved immediately 3′ of a CA dinucleotide (Danckwardt et al., 2008), although the exact cleavage site is usually clustered within a few nucleotides (Beaudoing et al., 2000; Pauws et al., 2001; Tian et al., 2005; Lin et al., 2012). After cleavage an untemplated poly(A) tail of about 250 adenosine-nucleotides is added to the upstream cleavage product by PAP (**Figure 2**). This is stimulated by the CPSF through the nuclear poly(A)-binding protein (PABPN1), which binds the growing poly(A) tail and controls its length (Kühn et al., 2009). Many of the other proteins in the human 3′-end processing complex do not play a direct role in the 3′-end formation, but instead couple polyadenylation to other cellular processes.

REGULATION OF POLYADENYLATION
Polyadenylation is influenced by three factors: (1) The strength of the cis-elements. (2) The concentration and activity of polyadenylation factors. (3) The concentration and activity regulatory factors (Barabino and Keller, 1999). These determine the efficiency of polyadenylation and the pattern of APA. The regulation of constitutive and alternative polydenylation differs and even CR-APA and UTR-APA are regulated differently. This differentiated regulation of APA is seen during T-cell activation, where CR-APA occurs at both early and late stages of activation, utilizing both more proximal and more distal poly(A) sites. In contrast, UTR-APA is most evident in late stages of activation and dominantly utilize more
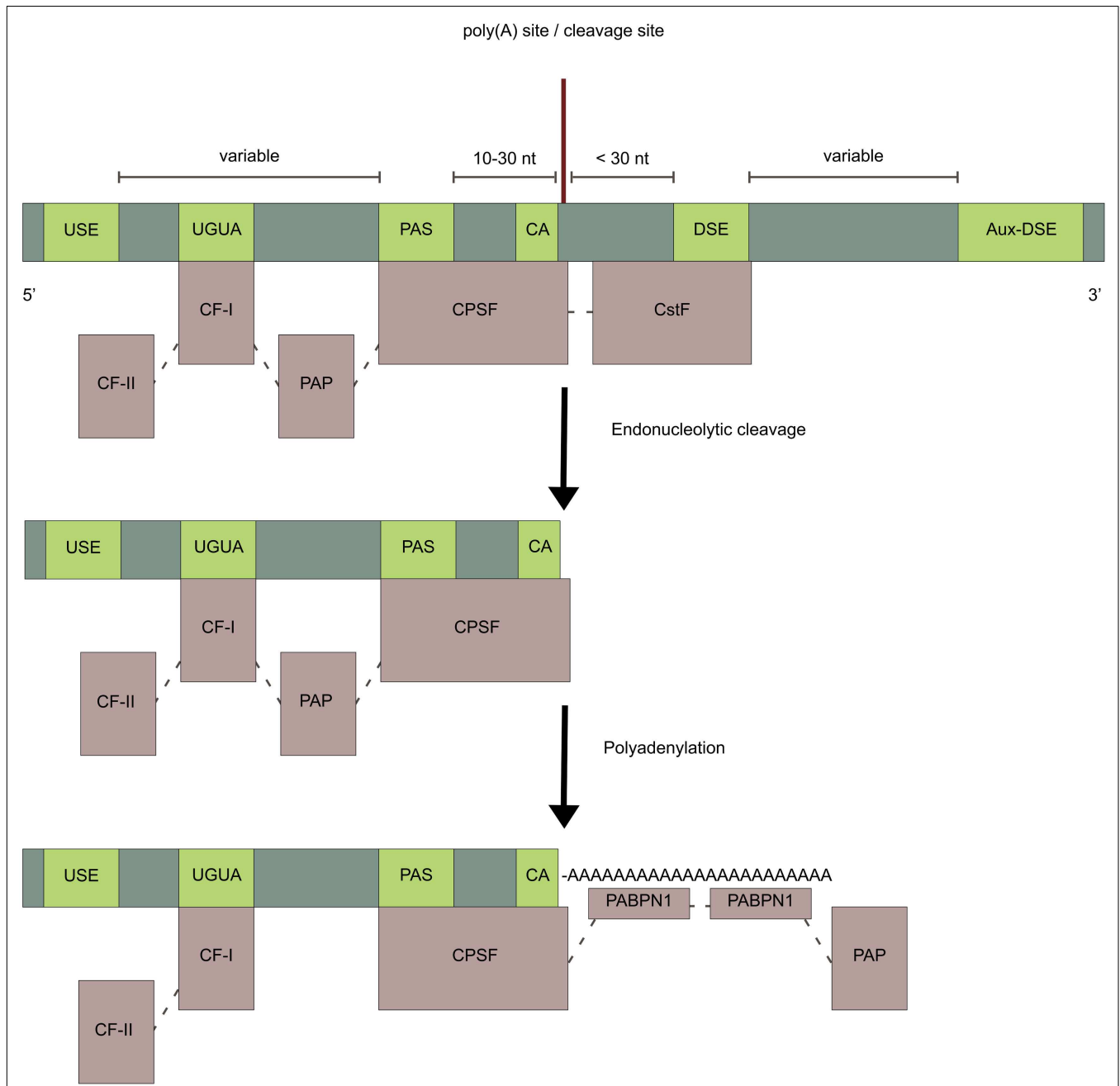
FIGURE 2 | Cis-elements and polyadenylation factors. Light-green boxes represent cis-elements in the pre-mRNA and light-purple boxes represent polyadenylation factors. The endonucleolytic cleavage of the pre-mRNA is made by the CPSF-subunit CPSF-73 at the poly(A) site, typically immediately 3' of a CA dinucleotide. After cleavage, an untemplated poly(A) tail of about 250 adenosine-nucleotides, is added to the upstream cleavage product by PAP. This is stimulated by the CPSF through the nuclear poly(A)-binding protein (PABPN1) which binds the growing poly(A) tail and controls its length.

proximal poly(A) sites (Sandberg et al., 2008). The differentiated regulation of APA is also seen during stem cell differentiation where UTR-APA generally switches toward utilizing more distal poly(A) sites, opposed to CR-APA which switches toward utilizing both more proximal and more distal poly(A) sites (Shepard et al., 2011). CR-APA often occurs together with alternative splicing of the last exon and it is suggested that there is a dynamic competition between splicing and polyadenylation (Tian et al., 2007; Evsyukova et al., 2012). CR-APA is thus both regulated by polyadenylation and splicing. The regulation of polyadenylation is complex and the involved factors interact with several other cellular processes. Part of the regulation is mediated through mechanistic interactions with polyadenylation factors, enhancing, or inhibiting their function, competing for the binding to their cis-elements or reducing

Table 1 | The core polyadenylation machinery.

Factor	Subunits	Necessary for	Interactions with other core factors	Target cis-element	Reference
CPSF	CPSF-30, CPSF-73, CPSF-100, CPSF-160, hFip1, Wdr33, and symplekin	Cleavage step and polyadenylation step	CstF, CF-II, and PAP	The PAS (AAUAAA), binding mediated by CPSF-160	Murthy and Manley (1995), Chan et al. (2011)
CstF	CstF-50, CstF-64, and CstF-77	Cleavage step	CPSF and CF-II	The DSE (U/GU-rich), binding mediated by CstF-64	MacDonald et al. (1994), Chan et al. (2011)
CF-I	CF-I-25 and either CF-I-68 or CF-I-59	Cleavage step	CF-II and PAP	The auxiliary motif UGUA	Venkataraman et al. (2005), Chan et al. (2011)
CF-II	hPcf11 and hClp1	Cleavage step	CPSF, CstF, and CF-I	None	Chan et al. (2011)
PAP	None	Cleavage step and polyadenylation step	CPSF and CF-I	None	Chan et al. (2011)

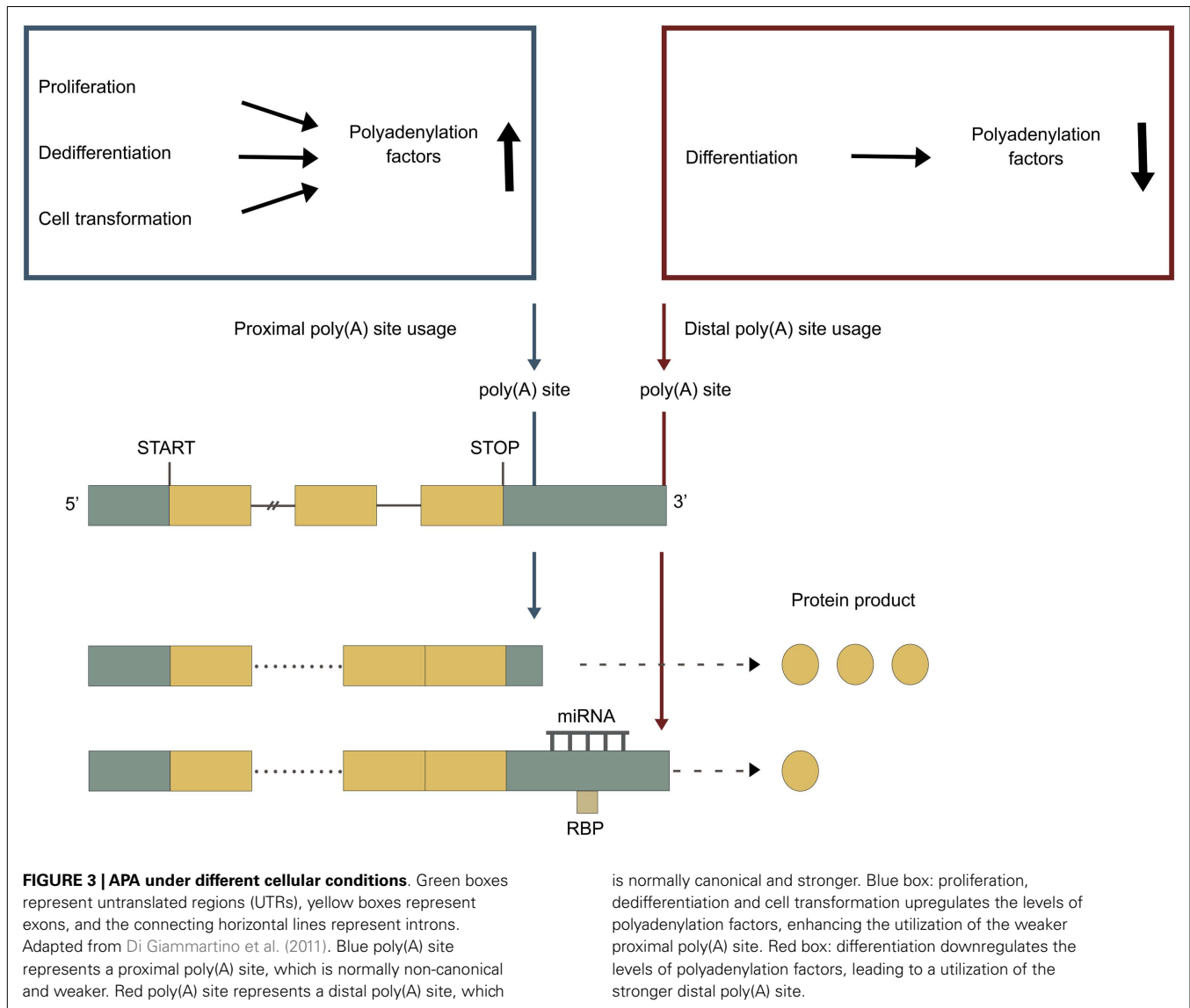their free levels in the nucleus (reviewed in Millevoi and Vagner, 2010).

Regulation by levels and variants of polyadenylation factors

It is hypothesized that higher levels of polyadenylation factors lead to improved utilization of weaker proximal poly(A) sites, encountered earlier during transcription (**Figure 3**). This is seen during proliferation, where upregulated levels of polyadenylation factors correlate with proximal poly(A) site usage (Elkon et al., 2012). It is also seen for many of the individual core polyadenylation factors (see sections below). Every component, but not every subunit, of the core polyadenylation machinery are regulated by post-translational modifications, such as methylation, sumoylation, acetylation, and phosphorylation (reviewed in Ryan and Bauer, 2008).

Cleavage and polyadenylation specificity factor.

During colorectal carcinoma progression and in breast cancer cell lines, upregulated levels of CPSF-subunits CPSF-73, CPSF-160, CPSF-30, and symplekin correlate with proximal poly(A) site usage (Mayr and Bartel, 2009; Morris et al., 2012). Interestingly, symplekin is also highly overexpressed in colon, lung, muscle, and prostate tumors where it promotes tumorigenesis (Buchert et al., 2010). CPSF-160 and symplekin are expressed in high levels in germ cells where the general 3′ UTR length is the shortest (Dass et al., 2001; Liu et al., 2007). CPSF-100 was conversely downregulated. During the induction of stem cells from somatic cells, all CPSF-subunits except CPSF-100 and hFip1, are upregulated. This correlates with a general shift toward proximal poly(A) site usage. During embryonic development all CPSF-subunits except hFip1 are downregulated. This correlates with a general shift toward distal poly(A) site usage (Ji and Tian, 2009; Ji et al., 2009). The examples indicate that upregulated levels of CPSF enhance polyadenylation, leading to proximal poly(A) site usage. Interestingly, knockdown of the CPSF-subunit hFip1 increases DNA damage and chromosome breakage (Stirling et al., 2012). This is properly caused by dysfunctional polyadenylation, leading to defective transcription termination, read-through transcription

and formation of R-loops (see section *Sequence changes in genes encoding polyadenylation factors*, page 15).

Cleavage stimulation factor.

Upregulated levels of the CstF subunit CstF-64 correlate with proximal poly(A) site usage, during colorectal carcinoma progression and in breast cancer cell lines (Mayr and Bartel, 2009; Morris et al., 2012). Upregulated levels of CstF-64 enhance utilization of proximal poly(A) sites in the mouse Testis Brain RNA-Binding Protein (*TB-RBP*) gene in male germ cells (Chennathukuzhi et al., 2001). CstF-64 is also upregulated in activated immune cells, enhancing the usage of proximal poly(A) sites (Shell et al., 2005; Sandberg et al., 2008). All CstF subunits are upregulated, during the induction of stem cells from somatic cells, correlating with a general shift toward proximal poly(A) site usage. Additionally, these subunits are downregulated during embryonic development, correlating with a general shift toward distal poly(A) site usage (Ji and Tian, 2009; Ji et al., 2009). When comparing the used poly(A) sites in genes that lengthen and genes that shorten during the induction of stem cells from somatic cells, there was major differences in the DSE. The DSE is where CstF binds, indicating that CstF plays a crucial role in the regulation of poly(A) site usage (Ji and Tian, 2009). The CstF-64 ortholog τCstF-64, expressed at highest levels in testis and brain, is necessary for spermatogenesis and fertilization (Dass et al., 2007; Hockert et al., 2011). Both CstF-64 and τCstF-64 are expressed in high levels in germ cells, which have the shortest general 3′ UTR length (Dass et al., 2001; Liu et al., 2007). Loss of τCstF-64 changes gene expression genome-wide in germ cells, leads to more frequent usage of distal poly(A) sites and causes more read-through transcription due to aberrant transcription termination (Li et al., 2012c). RNAi-mediated depletion of CstF-64 has a small effect on APA, but leads to upregulation of τCstF-64, suggesting that CstF-64 and τCstF-64 play redundant roles (Yao et al., 2012a). In support of this, during differentiation of C2C12 myoblasts τCstF-64 is also upregulated at the same time as the three subunits of CstF are downreulated (Ji et al., 2009). Codepletion of CstF-64 and τCstF-64 leads to significant changes in APA, with enhanced usage of distal poly(A) sites (Yao et al., 2012a). Together this indicates that upregulation

FIGURE 3 | APA under different cellular conditions. Green boxes represent untranslated regions (UTRs), yellow boxes represent exons, and the connecting horizontal lines represent introns. Adapted from Di Giammartino et al. (2011). Blue poly(A) site represents a proximal poly(A) site, which is normally non-canonical and weaker. Red poly(A) site represents a distal poly(A) site, which is normally canonical and stronger. Blue box: proliferation, dedifferentiation and cell transformation upregulates the levels of polyadenylation factors, enhancing the utilization of the weaker proximal poly(A) site. Red box: differentiation downregulates the levels of polyadenylation factors, leading to a utilization of the stronger distal poly(A) site.

of CstF enhances polyadenylation, leading to the usage of proximal poly(A) sites. A CstF-64 splice variant βCstF-64 is expressed specifically in nervous tissue and is thought to have regulatory functions here (Shankarling et al., 2009). Interestingly, CstF-64 increases fivefold during cell cycle, at G_0 to S phase transition (Martincic et al., 1998) and low levels of CstF-64 cause reversible cell cycle arrest while depletion causes apoptosis (Takagaki and Manley, 1998).

Cleavage factor I.   Cleavage factor I has been found to be important for non-canonical poly(A) site usage by enhancing the recruitment of CPSF (Brown and Gilmartin, 2003; Venkataraman et al., 2005; Sartini et al., 2008; Yang et al., 2011). In the germ cells, where the general 3′ UTR length is the shortest, the two CF-I subunits CF-I-25 and CF-I-68 are upregulated (Liu et al., 2007; Sartini et al., 2008). These two subunits are also upregulated during the induction of stem cells from somatic cells, correlating with a general shift toward proximal poly(A) site usage (Ji and Tian,

2009). These examples correlate well with the proposed function of CF-I in promoting non-canonical poly(A) site usage. However, reduced levels of CF-I has been found to enhance more proximal poly(A) site usage (Kubo et al., 2006; Kim et al., 2010). Loss of function of CF-I-25 and CF-I-68, but not of CF-I-59, have also been found to enhance genome-wide proximal poly(A) site usage (Gruber et al., 2012; Martin et al., 2012). This indicates that CF-I has other functions than enhancing the recruitment of CPSF, yet to be fully explained. Interestingly, downregulation of CF-I-68 has also been found to increase tumor invasiveness (Yu et al., 2008).

Cleavage factor II.   The CF-II subunit hClp1 is upregulated during the induction of stem cells from somatic cells, correlating with a general shift toward proximal poly(A) site usage (Ji and Tian, 2009). Likewise, hClp1 is downregulated during embryonic development, correlating with a general shift toward distal poly(A) site usage (Ji and Tian, 2009; Ji et al., 2009). In the germ cells, where there is a short general 3′ UTR length, hClp1 is also

upregulated (Liu et al., 2007). Low levels of the CF-II subunit hPcf11 reduces transcription termination efficiency, due to weakened polyadenylation (West and Proudfoot, 2008). Taken together it shows that upregulation of CF-II enhances polyadenylation, leading to proximal poly(A) site usage.

Poly(A) polymerase. In breast cancer cell lines, upregulated levels of PAP correlate with proximal poly(A) site usage (Mayr and Bartel, 2009). PAP is also upregulated during the induction of stem cells from somatic cells, correlating with a general shift toward proximal poly(A) site usage (Ji and Tian, 2009). These examples indicate that upregulation of PAP enhances polyadenylation, leading to usage of proximal poly(A) sites. PAP is post-translationally modified during cell cycle progression, with the modifications reflecting the proliferative state of the cell (Thomadaki et al., 2008). PAP modifications also correlate with apoptosis (Thomadaki et al., 2008). One of the PAP regulators, poly(ADP-ribose) polymerase 1 (PARP1), modifies PAP during heat shock, leading to an inhibition of polyadenylation (Di Giammartino et al., 2013). In addition to PAP, three nuclear non-canonical PAPs exists: Neo-PAP, TRAP, and Star-PAP. Neo-PAP has a function equal to that of PAP (Topalian et al., 2001). TRAP is only expressed in testis and has functions not fully understood (Lee et al., 2000). Star-PAP is necessary for the 3'-end formation of selected mRNAs, such as the cytoprotective enzyme heme oxygenase-1, which is upregulated in response to oxidative stress (Mellman et al., 2008) and the BIK protein, an initiator of mitochondrial apoptosis (Li et al., 2012b). Star-PAP activity is modulated directly by phosphatidylinositol 4,5-bisphosphate ($PI4,5P_2$) and Star-PAP associates with both $PIPKI\alpha$, which can generate new $PI4,5P_2$, and with the $PI4,5P_2$ sensitive protein kinase $CKI\alpha$, which can phosphorylate Star-PAP (Laishram et al., 2011). The phosphorylation of Star-PAP is critical for its activity and $PI4,5P_2$ sensitivity.

Other polyadenylation factors. The nuclear poly(A)-binding protein (PABPN1) is upregulated during the induction of stem cells from somatic cells, correlating with a general shift toward proximal poly(A) site usage (Ji and Tian, 2009). However, PABPN1 recently has been shown to suppress polyadenylation at weaker proximal poly(A) sites and loss of PABPN1 results in genome-wide 3' UTR shortening (de Klerk et al., 2012; Jenal et al., 2012). Loss of PABPN1 also increases apoptotic markers in affected cells (Bhattacharjee and Bag, 2012).

Regulation by non-polyadenylation factors
Splice factors. Many known splice factors have been shown to affect polyadenylation (**Table 2**).

Non-splice factors. Many non-splice factors have also been shown to affect polyadenylation (**Table 3**).

Regulation by transcription and co-transcriptional processes
Transcription. Polyadenylation is necessary for transcription termination (West and Proudfoot, 2009). It also affects transcription re-initiation at upstream promoters (Mapendano et al., 2010). The CTD of the Pol-II is necessary for the cleavage step in mammals (Licatalosi et al., 2002). It recruits CPSF, CstF, and CF-II and transfers them to their cis-elements in the pre-mRNA, as it emerges

from the Pol-II (McCracken et al., 1997; Lunde et al., 2010). The recruitment is mediated by phosphorylation of the Pol-II CTD serine 2 (Ser2) residues, which are progressively phosphorylated during Pol-II elongation (Ahn et al., 2004). Poly(A) sites located near the 5'-end of genes are normally not used, preventing premature 3'-end processing. This is probably because of the low Ser2 phosphorylation in the Pol-II CTD right after transcription initiation, resulting in inefficient recruitment of polyadenylation factors (Guo et al., 2011). The general transcription factor TFIIB is required for recruiting CPSF and CstF to the promoter (Wang et al., 2010). The transcription elongation rate affects the choice of poly(A) site, with a lower rate enhancing usage of proximal poly(A) sites (Pinto et al., 2011). Transcription elongation factor ELL2 also enhances polyadenylation (Martincic et al., 2009). Pol-II occupancy is found to be higher at distal poly(A) sites than at proximal poly(A) sites (Lin et al., 2012). Transcriptional activity regulates APA, with highly expressed genes expressing short 3' UTR isoforms and lowly expressed genes expressing long 3' UTR isoforms. This as the Pol-II is more likely to pause at poly(A) sites in the highly expressed genes (Ji et al., 2011). As the mRNAs with short 3' UTRs generated from the highly expressed genes are more stable and produce more protein, this effect potentiates the gene expression difference. Transcriptional activators have also been shown to enhance polyadenylation (Nagaike et al., 2011). The choice of alternative 5' UTRs and promoters affect the choice of poly(A) site and thus APA (Winter et al., 2007; Ji et al., 2011).

Capping. The presence of a 5'-end m7GpppG cap has been shown to positively affect the efficiency of polyadenylation (Cooke and Alwine, 1996). The nuclear cap-binding complex (CBC), bound to the 5' cap, is necessary for assembling a stable 3'-end processing complex. Depletion of CBC thus strongly reduces the efficiency of the cleavage step (Flaherty et al., 1997).

Splicing. The patterns of APA and alternative splicing correlate across tissues, indicating coordinated regulation of these processes (Wang et al., 2008). Because the used intronic poly(A) sites are usually located in large introns with weak 5' splice sites, which require more time to be spliced out, it has been suggested that there is a dynamic competition between splicing and polyadenylation (Tian et al., 2007; Evsyukova et al., 2012). Many of the trans-factors involved in polyadenylation are also known to take part in splicing (Shi et al., 2009; Evsyukova et al., 2012). Similarly, a number of known splice factors have also been shown to affect polyadenylation (**Table 2**).

Regulation by other processes
RNA structure. The secondary structure of RNA can affect polyadenylation. This is seen in HIV-1-strains where the proximal poly(A) site dominantly fold into a hairpin structure, thereby inhibiting polyadenylation (Gee et al., 2006). Highly used poly(A) sites are positively associated with an energetically favorable mRNA structure near poly(A) sites that exposes the PAS (Khaladkar et al., 2011).

Nucleosome positioning. Decreased nucleosome density is found around poly(A) sites, most significantly around the actively used

Table 2 | Splice factors affecting polyadenylation.

Splice factor	Motif	Mechanism	Effect on polyadenylation	Reference
FOX-1/FOX-2 (neuronal-specific)	UGCAUG	Motifs are found in 3′ UTRs between poly(A) sites, indicating a role in APA regulation	Unknown	Wang et al. (2008), Zhou et al. (2007)
Nova (neuronal-specific)	YCAY	Enhances or inhibits polyadenylation depending on the position of its motif relative to the PAS	Enhance/inhibit	Licatalosi et al. (2008)
U1	5′ splice site	Represses adjacent downstream intronic poly(A) site usage	Inhibit	Kaida et al. (2010), Berg et al. (2012), Vorlová et al. (2011), Vickers et al. (2011)
		Inhibits the PAP	Inhibit	Gunderson et al. (1998)
	Intronic enhancer element (with both 5′ and 3′ splice sites)	Enhances upstream polyadenylation	Enhance	Lou et al. (1998)
U1A (component of U1)	AUGCN(1-3)C	Binds motifs downstream of poly(A) site and inhibits CstF binding to the DSE	Inhibit	Phillips et al. (2004)
	USE	Enhances polyadenylation	Enhance	Hall-Pogar et al. (2007)
	AUUGCAC	Inhibits the PAP	Inhibit	Gunderson et al. (1997)
U2	3′ splice site	Enhances adjacent downstream poly(A) site usage	Enhance	Kyburz et al. (2006)
U2 auxiliary factor 1 (U2AF1)	USE	Enhances polyadenylation	Enhance	Danckwardt et al. (2007)
U2 auxiliary factor 2 (U2AF2)	3′ splice site	Enhances adjacent downstream poly(A) site usage	Enhance	Millevoi et al. (2006)
		Inhibits the PAP	Inhibit	Ko and Gunderson (2002)
hnRNP F	DSE	Competes with CstF for DSE binding	Inhibit	Veraldi et al. (2001)
hnRNP L	CA-rich elements	Inhibits the usage of intronic poly(A) sites	Inhibit	Hung et al. (2008)
hnRNP H	Aux-DSE	Enhances polyadenylaiton	Enhance	Dalziel et al. (2007)
hnRNP I (PTB)	DSE	Competes with CstF for DSE binding	Inhibit	Castelo-Branco et al. (2004)
	USE	Enhances polyadenylation	Enhance	Millevoi et al. (2009), Hall-Pogar et al. (2007), Blechingberg et al. (2007)
PTB associated splicing factor	USE	Enhances polyadenylation	Enhance	Hall-Pogar et al. (2007)
SR proteins	Unknown	Enhances polyadenylation	Enhance	Blechingberg et al. (2007)
SRp20	Intronic enhancer element (with both 5′ and 3′ splice sites)	Enhances upstream polyadenylation	Enhance	Lou et al. (1998)
SRp75	Unknown	Inhibits the PAP	Inhibit	Ko and Gunderson (2002)
SRm160	Unknown	Enhances polyadenylation	Enhance	McCracken et al. (2002, 2003)

Table 3 | Non-splice factors affecting polyadenylation.

Factor	Motif	Mechanism	Effect on polyadenylation	Reference
E2F	TTGGCGG	Motifs are enriched in promoter-regions of many polyadenylation factors and E2F enhances their expression	Upregulation of polyadenylation factors leading to 3′ UTR shortening	Elkon et al. (2012), Ji and Tian (2009)
Hu proteins	U-rich sequences	Inhibits CstF binding to the DSE at poly(A) sites containing U-rich sequences	Inhibit	Zhu et al. (2007)
NS1 (Influenza A virus protein)	None	Interacts with CPSF-30 and inhibits polyadenylation of host pre-mRNAs in infected cells	Inhibit	Nemeroff et al. (1998)
14-3-3ε	None	Inhibits the PAP and redistributes it to the cytoplasm	Inhibit	Kim et al. (2003)
CSR1	None	Redistributes CPSF-73 to the cytoplasm	Inhibit	Zhu et al. (2009)
IRBIT	None	In response to oxidative stress, IRBIT is phosphorylated, whereby it both inhibits PAP and redistributes the CPSF-subunit hFip1 to the cytoplasm	Inhibit	Kiefer et al. (2009)
p53	None	In response to DNA damage, p53 complexes with BARD1 and CstF-50, lowering the free levels of CstF-50	Inhibit	Nazeer et al. (2011)
BRCA1	None	In response to DNA damage, BRCA1 complexes with BARD1 and CstF-50, lowering the free levels of CstF-50	Inhibit	Kleiman and Manley (2001)
Cdc73	None	Associates with CPSF-CstF complexes and stimulates polyadenylation of specific target genes	Enhance	Rozenblatt-Rosen et al. (2009)
HSF1	None	In response to cellular stress, HSF1 complexes with symplekin and CstF-64 and enhances polyadenylation of Hsp70	Enhance	Xing et al. (2004)
P54	USE	Enhances polyadenylation	Enhance	Hall-Pogar et al. (2007)
p38 MAPK	None	In response to cellular stess, p38 MAPK phosphorylates RBPs blocking the USE, leading to their dissociation, and upregulates several polyadenylation factors, both events leading to enhanced polyadenylation	Enhance	Danckwardt et al. (2011)
T3	None	T3 stimulates the nuclear polyadenylaiton step, mediated by PAP, increasing poly(A) tail length of P450R	Enhance	Liu and Waxman (2002)
FSH	None	Stimulates proximal poly(A) site usage in *CREM* gene	Enhance	Foulkes et al. (1993)
Estrogen	None	Induces 3′ UTR shortening of the DNA replication regulator CDC6	Enhance	Akman et al. (2012)

ones (Spies et al., 2009). Increased nucleosome density is found downstream of poly(A) sites, most significantly downstream of the actively used ones (Spies et al., 2009; Khaladkar et al., 2011). Highly expressed genes have a lower nucleosome density around poly(A) sites than genes expressed at low levels (Ji et al., 2011).

Epigenetic modifications. Allele-specific poly(A) site usage is seen in an imprinted mouse gene (*H13*), as a result of DNA-methylation. A CpG island separates polyA sites utilized in *H13*. Alleles without methylation of the CpG island utilize a proximal poly(A) site, generating a truncated H13, whereas alleles with methylation of the CpG island, utilize downstream poly(A) sites (Wood et al., 2008). Histone methylation (H3K36me3) decreases downstream of used poly(A) sites (Lian et al., 2008; Wang et al., 2013). Highly expressed genes have increased levels of histone methylation (H3K4me3 and H3K36me3) around proximal poly(A) sites compared to around distal poly(A) sites (Ji et al., 2011; Lin et al., 2012). The pattern of histone modifications can thus characterize poly(A) sites and even discriminate between poly(A) sites with high and low usage, suggesting a link

between chromatin structure and mRNA structure (Khaladkar et al., 2011).

Circadian rhythms. The suppressor of cytokine signaling 3 (SOCS3) mRNA, encodes a protein important for the leptin signaling pathway. SOCS3 mRNA is expressed in two isoforms due to APA. The two isoforms are expressed in an oscillating pattern and in opposite phases (Ptitsyn and Gimble, 2007).

METHODS FOR CHARACTERIZING POLY(A) SITE USAGE
COMPLEMENTARY DNAs, EXPRESSED SEQUENCE TAGS, AND MICROARRAYS
Early studies to identify poly(A) sites in human were done by aligning complementary DNAs (cDNAs) and expressed sequence tags (ESTs) to the genome, thereby identifying the 3′ end of genes. These analyses revealed that more than half of human genes have multiple poly(A) sites and undergo APA (Beaudoing et al., 2000; Tian et al., 2005; Yan and Marr, 2005). Later studies using microarray platforms were designed to characterize APA in different cell types, cancers, and at different development stages. These analyses showed that a general 3′ UTR shortening associates with increased proliferation, dedifferentiation, and cell transformation (Sandberg et al., 2008; Ji and Tian, 2009; Singh et al., 2009) whereas a general 3′ UTR lengthening associates with cell differentiation (Ji and Tian, 2009; Ji et al., 2009). Poly(A) site usage were also found to differ according to tissue type, developmental stage, genotype, and cancer subtype (Kwan et al., 2008; Sandberg et al., 2008; Ji and Tian, 2009; Ji et al., 2009; Singh et al., 2009). The cDNA and EST methods have limited statistical power and microarrays are limited to comparing the results only with already known poly(A) sites and have difficulties with quantifying the different 3′-UTR isoforms produced by APA. Thus new methods for characterizing poly(A) site usage were needed.

HIGH-THROUGHPUT SEQUENCING TECHNIQUES
With the development of new high-throughput sequencing techniques, several methods have been developed to specifically characterize poly(A) site usage genome-wide (Mangone et al., 2010; Ozsolak et al., 2010; Fox-Walsh et al., 2011; Fu et al., 2011; Jan et al., 2011; Shepard et al., 2011; Derti et al., 2012; Jenal et al., 2012; Lin et al., 2012; Martin et al., 2012; Pelechano et al., 2012; Yoon et al., 2012; Hoque et al., 2013; Wang et al., 2013; Wilkening et al., 2013). These new techniques provide higher coverage than any of the previous methods, allow better detection of the poly(A) sites and give a more precise quantification of their relative usage. High-throughput sequencing studies have confirmed that more than half of human genes have multiple poly(A) sites and undergo APA (Derti et al., 2012). They have also confirmed that 3′ UTR shortening associates with proliferation (Elkon et al., 2012) and that 3′ UTR lengthening associates with differentiation (Shepard et al., 2011). However, they have not unanimously confirmed that 3′ UTR shortening associates with cell transformation, but instead found that APA patterns can shift for both more proximal and more distal poly(A) site usage during cell transformation (Fu et al., 2011; Elkon et al., 2012; Lin et al., 2012; Morris et al., 2012). Several computational studies have been performed to characterize patterns in the cis-elements around used poly(A)

sites obtained from experiments (Beaudoing et al., 2000; Legendre and Gautheret, 2003; Tian et al., 2005; Ara et al., 2006; Nunes et al., 2010; Ozsolak et al., 2010). Computational methods have also been developed to identify poly(A) sites (Tabaska and Zhang, 1999; Graber et al., 2002; Hajarnavis et al., 2004; Bajic et al., 2005; Cheng et al., 2006; Retelska et al., 2006; Akhtar et al., 2010).
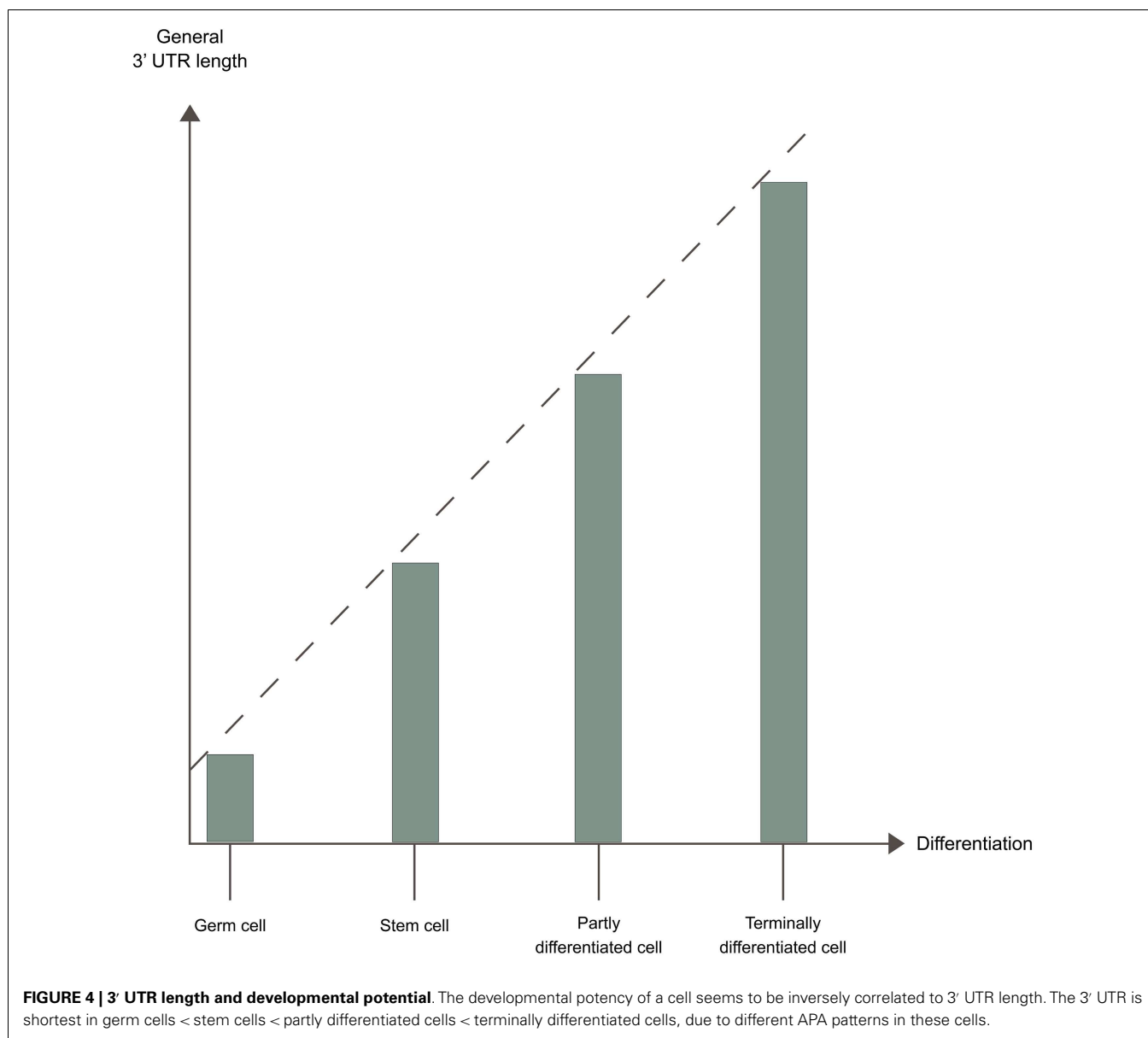
THE POLYADENYLATION PATTERN
CHANGED PATTERNS OF POLYADENYLATION UNDER DIFFERENT CELLULAR CONDITIONS
Changes in cellular conditions can cause a widespread dynamic shift in the use of poly(A) sites by APA, altering the 3′ UTRs, and in some cases the coding potential of the mRNAs. A shift toward more proximal poly(A) site usage associates with increased proliferation (Sandberg et al., 2008; Elkon et al., 2012) and with dedifferentiation (Ji and Tian, 2009). In contrast, shift toward more distal poly(A) site usage associates with cell differentiation (**Figure 3**) (Ji and Tian, 2009; Ji et al., 2009; Mangone et al., 2010; Hilgers et al., 2011; Shepard et al., 2011). The developmental potency of a cell seems to be inversely correlated to 3′ UTR length, with the 3′ UTR being shortest in germ cells < stem cells < partly differentiated cells < terminally differentiated cells (**Figure 4**) (Ji and Tian, 2009; Shepard et al., 2011). 3′ UTR shortening has also been reported to be associated with cell transformation (Mayr and Bartel, 2009; Singh et al., 2009; Lin et al., 2012; Morris et al., 2012). Recent studies have however found that this association is inconsistent (Fu et al., 2011; Elkon et al., 2012). Interestingly, there is an enrichment of binding sites for several transcription factors involved in proliferation/differentiation, including E2F, c-myc, and p53 in the promoter-regions of many of the genes coding for polyadenylation factors (Ji and Tian, 2009; Elkon et al., 2012). High levels of E2F have been shown to enhance the expression of polyadenylation factors. E2F transcription factors are upregulated during proliferation (Elkon et al., 2012). This might explain the shift in the APA during proliferation, while changes in levels of other transcription factors might explain the shift in APA during differentiation. Activation of neurons also induce 3′ UTR shortening for multiple genes (Flavell et al., 2008). This is due to the relative shortage of splice factor U1 generated by the transcriptional upregulation found in activated neurons, as shortage of U1 leads to enhanced usage of proximal poly(A) sites (Berg et al., 2012).

During cell differentiation/dedifferentiation
Generally the 3′ UTRs shorten when inducing pluripotent stem cells from somatic cell types, the opposite of what happens during embryonic development. When inducing from germ cells 3′ UTRs generally lengthen, the opposite of what happens during postnatal testis development (Ji and Tian, 2009). The genes with 3′ UTR shortening during the induction were more likely to have had 3′ UTR lengthening during embryonic development. Vice versa, the genes with 3′ UTR lengthening during the induction were more likely to have had 3′ UTR shortening during embryonic development (Ji and Tian, 2009). The proximal poly(A) sites highly responsive to cell state change during the induction of somatic cells, were also highly regulated during embryonic development, but in the opposite direction. When comparing these

FIGURE 4 | 3′ UTR length and developmental potential. The developmental potency of a cell seems to be inversely correlated to 3′ UTR length. The 3′ UTR is shortest in germ cells < stem cells < partly differentiated cells < terminally differentiated cells, due to different APA patterns in these cells.

poly(A) sites with less responsive proximal poly(A) sites in other genes, they were more conserved, had stronger DSE and lead to longer alternative 3′ UTRs (Ji and Tian, 2009). A recent high-throughput sequencing study also found a correlation between 3′ UTR lengthening and differentiation (Shepard et al., 2011).

During proliferation and cell transformation

A general shift toward more proximal poly(A) site usage is seen during increased proliferation, leading to 3′ UTR shortening. Enhanced usage of proximal intronic poly(A) sites is also seen, leading to changes in mRNA CRs (Sandberg et al., 2008; Elkon et al., 2012). 3′ UTR shortening has also been reported be associated to cell transformation (Mayr and Bartel, 2009; Singh et al., 2009; Morris et al., 2012), with the transformed cells expressing mRNAs with shorter 3′ UTRs compared with non-transformed cells with similar proliferation rate (Mayr and Bartel, 2009). 3′

UTR shortening has also been associated with poor cancer prognosis (Lembo et al., 2012). 3′ UTR shortening is, e.g., seen for fibroblast growth factor 2 (FGF-2), which is involved in tumor neovascularization. The distal poly(A) site is primarily utilized in fibroblasts, in contrast to in transformed cell lines, where more proximal poly(A) sites are utilized (Touriol et al., 1999). The 3′ UTR shortening is stronger during the transition from arrested to proliferative state than during the transition from proliferative to transformed state (Elkon et al., 2012). Recent studies have found the association between 3′ UTR shortening and cell transformation to be inconsistent. One study found an even distribution between genes with 3′ UTR shortening and lengthening in one transformed cell line (Elkon et al., 2012). Another study found a shift toward usage of more distal poly(A) sites for the majority of genes in another transformed cell line (Fu et al., 2011). Both these studies did however find a general 3′ UTR shortening in

other transformed cell lines (Fu et al., 2011; Elkon et al., 2012). Yet other studies found a general 3′ UTR shortening compared to matched normal tissue, both in tumor samples from five different tissues (Lin et al., 2012) and in several colorectal carcinomas (Morris et al., 2012). These examples illustrate that the shifts in poly(A) site usage in transformed cells is more complex than just a shift toward usage of more proximal poly(A) sites.

THE POLYADENYLATION PATTERN AS A BIOMARKER

Poly(A) site usage differs according to tissue type, developmental stage, genotype, and cancer subtype (Breton et al., 2001; Zhang et al., 2005; Kubo et al., 2006; Kwan et al., 2008; Sandberg et al., 2008; Wang et al., 2008; Ji and Tian, 2009; Ji et al., 2009; Singh et al., 2009; Thomsen et al., 2010; Derti et al., 2012; MacIsaac et al., 2012). This makes the polyadenylation pattern a candidate biomarker, which could be used for, e.g., cancer classification. Tissue-specific APA is common and often associated with the proximal non-canonical poly(A) sites (Beaudoing et al., 2000). It has been estimated that 52% of all CR-APA events and 80% of all UTR-APA events are regulated differentially between tissues (Wang et al., 2008), making tissue-specific UTR-APA events even more differentially regulated than tissue-specific alternative splicing.

ALTERATIONS IN POLYADENYLATION IN ENDOCRINE DISEASE

SEQUENCE CHANGES IN AND AROUND CIS-ELEMENTS

Changes in the sequence in and around cis-elements can disrupt normal polyadenylation and cause disease (reviewed in Danckwardt et al., 2008). Single nucleotide polymorphisms (SNPs) have been identified that can both create or disturb PASs (Thomas and Sætrom, 2012). SNPs in PASs can affect mRNA length and expression, and are significant predictors of gene expression levels (Thomas and Sætrom, 2012; Yoon et al., 2012). SNPs can also affect APA. A significant high fraction of the SNPs which affect APA, are linked to SNPs found by genome-wide association studies (GWAS). A high proportion of the APA-alleles created by these SNPs, are positively correlated with disease-risk alleles (Thomas and Sætrom, 2012). GWAS-indentified trait-associated SNPs are generally overrepresented in the 3′ UTRs of genes, possibly creating, or affecting the cis-elements involved in polyadenylation (Arnold et al., 2012).

Loss of function changes

A functional poly(A) site is required for polyadenylation and thus transcription termination (Connelly and Manley, 1988; West and Proudfoot, 2009). Loss of function changes in the cis-elements of poly(A) sites, can therefore lead to reduced gene expression caused by weakened polyadenylation efficiency. Such loss of function changes are associated with various diseases. They were first described in the PAS of the only poly(A) site in the *HBA2* gene causing α-thalassemias (Higgs et al., 1983; Harteveld et al., 1994) and in a PAS in the *HBB* gene causing β-thalassemias (Orkin et al., 1985; Jankovic et al., 1990; Rund et al., 1992; van Solinge et al., 1996). Loss of function changes causing disease, frequently seem to affect the PAS, which in one of the two core cis-elements found at poly(A) sites. Any possible base change in the canonical PAS (AAUAAA) significantly reduce polyadenylation efficiency

(Sheets et al., 1990). The PAS is responsible for binding the core polyadenylation factor CPSF, which is required for polyadenylation (Chan et al., 2011). The reduced efficiency of polyadenylation associated with changes to the canonical PAS, is thus caused by weakened recruitment of CPSF to poly(A) sites. The examples of endocrine diseases affected by loss of function changes in and around cis-elements of poly(A) sites are described below.

Neonatal diabetes. A loss of function mutation in the *INS* gene PAS has been shown to cause neonatal diabetes. This disruptive mutation changes the PAS of the only poly(A) site (AAUAAA to AAUAAG), reducing insulin mRNA expression to less than 3×10^{-4} percent (Garin et al., 2010). The authors hypothesize that this is due to impaired mRNA stability. However, such a low expression level is more likely a result of weakened polyadenylation efficiency of the only poly(A) site in the *INS* gene, leading to impaired transcription termination and thus reduced gene expression, as seen similarly for other genes (Higgs et al., 1983; Bennett et al., 2001; Stacey et al., 2011). The change of the PAS sequence (AAUAAA to AAUAAG) reduces polyadenylation efficiency about 98% (Sheets et al., 1990). As the PAS is crucial for binding CPSF, the reduced polyadenylation efficiency in the mutated *INS* gene is due to weakened CPSF recruitment (Chan et al., 2011).

Immune dysfunction, polyendocrinopathy, enteropathy, X-linked. A loss of function mutation in the PAS of the only poly(A) site in the *FOXP3* gene causes immune dysfunction, polyendocrinopathy, enteropathy, X-linked (IPEX). The mutation changes the PAS (AAUAAA to AAUGAA), leading to diminished FOXP3 mRNA expression (Bennett et al., 2001). This is probably due to impaired polyadenylation of the only poly(A) site in the *FOXP3* gene, leading to non-termination of transcription, as seen similarly for other genes (Higgs et al., 1983; Stacey et al., 2011). The change of the PAS sequence (AAUAAA to AAUGAA) reduces polyadenylation efficiency about 95% (Sheets et al., 1990). The PAS is vital for binding CPSF, hence the impaired polyadenylation of the mutated *FOXP3* gene is due to weakened binding of CPSF (Chan et al., 2011).

Mayer–Rokitanski–Küster–Hauser syndrome. The AMH gene encoding Anti-Müllerian Hormone is located only 739 bp downstream of the housekeeping gene encoding splicing factor 3a subunit 2 (SF3a2) (Dresser et al., 1995), a component of the splice factor U2 (Bennett and Reed, 1993). Read-through transcription of the *SF3a2* gene into the *AMH* gene, due to failed polyadenylation of the *SF3a2* gene, is normally seen for a small proportion of transcripts (Dresser et al., 1995). This leads to constitutive expression of AMH along with SF3a2 and results in higher than normal AMH levels. It is hypothesized, that polymorphisms around the *SF3a2* gene poly(A) site leading to reduced polyadenylation with more read-through transcription, could cause Mayer–Rokitanski–Küster–Hauser (MRKH) syndrome. MRKH syndrome is characterized by the Müllerian ducts failing to develop into the uterus, cervix, and upper vagina. Polymorphisms around the *SF3a2* gene poly(A) site were however not found in 30 MRKH patients examined in one study (Oppelt et al., 2005).

Hypothalamic-pituitary-adrenal axis dysregulation. In the serotonin transporter (*SERT*) gene a PAS U/G polymorphism exists in the distal, more canonical poly(A) site. The G-allele changes the PAS (AUUAAC to AGUAAC) leading to a lesser usage of the distal poly(A) site. This reduces total SERT expression to about half in the brain (Gyawali et al., 2010). This is sensible as the distal, canonical poly(A) site is most efficient for polyadenylation and most frequently utilized, therefore yielding higher mRNA expression levels (Wang et al., 2013). The G-allele is associated with an increased risk for panic disorder (Gyawali et al., 2010), as well as heightened anxiety and depressive symptoms (Hartley et al., 2012). Interestingly, the SSRI fluoxetine increases the usage of the distal poly(A) site, suggesting that the therapeutic effect of this compound depends on alterations in polyadenylation (Hartley et al., 2012). Serotonin has many functions in the endocrine system, e.g., in the regulation of the hypothalamic-pituitary-adrenal axis (Heisler et al., 2007). This suggests that the *SERT* gene polymorphism could have endocrine implications as well.

Cancer. In the PAS of the only poly(A) site in the *TP53* gene a A/C polymorphism exists. The C-allele changes the PAS (AAUAAA to AAUACA), leading to impaired polyadenylation and transcription termination. This results in decreased p53 mRNA expression and read-through trancription. The change of the PAS sequence (AAUAAA to AAUACA) reduces polyadenylation efficiency about 89% (Sheets et al., 1990). CPSF binds the PAS, so the reduced polyadenylation efficiency is due to weakened binding of CPSF (Chan et al., 2011). The C-allele is associated with various cancer diseases (Stacey et al., 2011; Zhou et al., 2012). This polymorphism could be important for endocrine tumor diseases as well, as loss of p53 is an important event in the pathogenesis of many tumors, e.g., thyroid tumors (Yoshimoto et al., 1992).

Gain of function changes

Gain of function changes in the cis-elements of poly(A) sites, can lead to enhanced gene expression caused by improved polyadenylation efficiency and to changes in APA. Such gain of function changes are associated with various diseases. They were first described in the CA dinucleotide immediately 5′ of the cleavage site in the *F2* gene encoding thrombin (Gehring et al., 2001; Ceelie et al., 2004; Danckwardt et al., 2004, 2006, 2007). Later they were also described in the DSE of the *F2* gene (Danckwardt et al., 2004), as well as in the DSE of the *FGG* gene encoding fibrinogen gamma (Lisman et al., 2005; Uitte de Willige et al., 2007). All of these mutations cause thrombophilia through enhanced polyadenylation efficiency, leading to upregulated gene expression. The examples of endocrine diseases affected by gain of function changes in and around cis-elements of poly(A) sites are described below.

Type I diabetes. The *GIMAP5* gene has two poly(A) sites in its 3′ UTR. In the PAS of the proximal poly(A) site a G/A polymorphism exists. The minor A-allele changes the PAS to a canonical one (AAUAGA to AAUAAA) leading to enhanced proximal poly(a) site usage (Shin et al., 2007). The change of the PAS sequence (AAUAGA to AAUAAA) increases polyadenylation efficiency about 25-fold (Sheets et al., 1990). The A-allele is associated

with high levels of tyrosine phosphatase (IA-2) autoantibodies, which are reported to be associated with clinical onset of type I diabetes (Shin et al., 2007). Interestingly, the G-allele is associated with an increased risk of systemic lupus erythematosus (SLE) (Hellquist et al., 2007). The A-allele induce a higher proportion of mRNAs with the short 3′ UTR isoform, but was not found to have an effect on total GIMAP5 mRNA levels (Hellquist et al., 2007). Differences in translational efficiency between the mRNAs with the long and short 3′ UTR isoforms are thus believed to be affecting the levels of IA-2 autoantibodies.

Type II diabetes. A proximal intronic poly(A) site in the transcription factor 7-like 2 (*TCF7L2*) gene has been found. Usage of this poly(A) site leads to production of a C-terminally truncated protein isoform that repress T-cell factor/lymphoid-enhancer factor (TCF/LEF)-dependent target genes (Locke et al., 2011). The authors hypothesize that this poly(A) site might explain why intronic SNPs in the *TCF7L2* gene are associated with a risk of type 2 diabetes, as they could enhance the utilization of this poly(A) site.

IGF-1 deficiency. A U/A polymorphism in a proximal poly(A) site PAS of the insulin-like growth factor 1 (*IGF-1*) gene 3′ UTR has been described. It changes the PAS (AAUAUA to AAAAUA), leading to expression of a shorter than normal IGF-1 mRNA. The U/A polymorphism has been reported in a child born small for gestation age with IGF-1 deficiency (Bonapace et al., 2003). A later study however reported that this polymorphism along with other polymorphisms in this PAS of the *IGF-1* gene do not cause IGF-1 deficiency nor growth impairment (Coutinho et al., 2007).

Endocrine tumor diseases. A C/T polymorphism in the *RET* proto-oncogene, six nucleotides upstream of an intronic proximal poly(A) site PAS of intron 19 exists (Gartner et al., 2005). The adjacent poly(A) site is commonly used for polyadenylation by CR-APA. The C/T polymorphism lies within a binding site of the nucleic acid binding protein Pbx1 (AAATTAG(C/T)T), which might affect polyadenylation at the adjacent poly (A) site. The T-allele carriers have the more canonical binding motif for Pbx1. Heterozygosity for the T-allele was found in a significantly higher number of patients with various endocrine tumors than in healthy controls. Homozygosity for the T-allele was exclusively found in DNA from endocrine tumors with a high malignant potential, as a result of loss of heterozygosity (LOH) in these tumors (Gartner et al., 2005). This suggests that Pbx1 enhances polyadenylation at the adjacent poly(A) site, leading to higher expression levels of RET.

Cancer. Gain of function mutations have been found in the *CCND1* gene encoding Cyclin D1 in mantle cell lymphoma (MCL) tumors. These mutations create new proximal poly(A) sites, giving rise to expression of truncated mRNAs lacking most of the 3′ UTR. These truncated mRNAs are associated with strongly proliferative MCL tumors and inferior survival (Wiestner et al., 2007). Such mutations could be important for endocrine tumor diseases as well, as high expression of Cyclin D1 is associated with many cancers, e.g., pancreatic endocrine tumors (Guo et al., 2003).

SEQUENCE CHANGES IN GENES ENCODING POLYADENYLATION FACTORS

Cancer

Mutations in yeast polydadenylation factors have been shown to disrupt normal transcription termination (Birse et al., 1998) and to cause inefficient transcription elongation (Luna et al., 2005). They have also been shown to cause genome instability, through formation of R-loops (Stirling et al., 2012). R-loops are persistent, transcription-associated RNA:DNA hybrids that expose damage-prone single stranded DNA (ssDNA) on the nonsense strand and may block the replication fork progression (Stirling et al., 2012). Non-terminated pre-mRNA transcripts, which form due to defective cleavage at the poly(A) site (Birse et al., 1998), could form the basis of these RNA:DNA hybrids found in R-loops. Genome instability is found in many human cancers, in part due to increased chromosome instability (CIN), which is linked to R-loops (Wahba et al., 2011). CIN is normally found as an early event in the oncogenic process and can lead to loss or gain of whole chromosomes or chromosomal fragments. Researchers working with yeast strains, each with mutations in one gene out of 305 genes involved with CIN, found that mutations in 44 of the genes resulted in an indicated increased CIN. Seven of these 44 genes encoded subunits of polyadenylation factors. They were homologs of CPSF-subunits CPSF-100, hFip1, and Wdr33, CstF subunit CstF-64, both subunits of CF-II and yeast factor HRP1, lacking a human homolog. The mutant stains had indicated increased CIN in all cell cycle stages (Stirling et al., 2012). Genome-wide analysis of fragile chromatin sites in these mutant stains supported a transcription-dependent mechanism of DNA damage, characteristic of R-loop formation. Increased RNA:DNA hybrid formation in polyadenylation factor mutant stains were found and the resulting CIN could be suppressed by RNaseH, which degrades the RNA and thereby removes the R-loops (Stirling et al., 2012). The researchers found that small interfering RNA (siRNA) knockdown of the CPSF-subunit hFip1 in human cells increased DNA damage and chromosome breakage (Stirling et al., 2012). Interestingly, hFip1 also comprises the N-terminal part of a fusion protein with platelet-derived growth factor receptor alpha (PDGFRα), which causes 14–60% the incidents of hypereosinophilic syndrome/chronic eosinophilic leukemia (Cools et al., 2003; Gotlib and Cools, 2008; Li et al., 2012a). Truncation fusions of yeast FIP1, analogous to those found for hFip1 in chronic eosinophilic leukemia, was found to cause genome instability in yeast (Stirling et al., 2012). All these findings indicate that the normal 3′-end processing complex maintains genome integrity by suppressing R-loop formation, thereby suppressing CIN, a function that may be impaired in many human cancers (Stirling et al., 2012). Mutations in genes encoding polyadenylation factors could be important for endocrine tumor diseases, as CIN is already known as a powerful prognostic indicator for endocrine pancreatic tumor patients (Jonkers et al., 2007).

SEQUENCE CHANGES IN GENES ENCODING REGULATORY FACTORS

Hereditary and sporadic parathyroid tumors

The tumor suppressor protein Cdc73 is mutationally inactivated in hereditary and sporadic parathyroid tumors (Rozenblatt-Rosen et al., 2009). It normally associates with CPSF-CstF complexes and enhances polyadenylation of specific target genes. The inactivation of Cdc73 leads to decreased association of CPSF-CstF complexes with these target genes, reducing their expression levels (Rozenblatt-Rosen et al., 2009). This suggests that impaired polyadenylation of these target genes plays a part in the parathyroid tumorigenesis.

OTHER MECHANISMS LEADING TO ALTERATIONS IN POLYADENYLATION

Fragile X-associated premature ovarian insufficiency

A shift in poly(A) site usage due to UTR-APA is seen in the fragile X mental retardation 1 (FMR1) gene premutation alleles compared to normal alleles (Tassone et al., 2011). The FMR1 gene 5′ UTR contains a CGG repeat element that is expanded to more than 200 CGG repeats and methylated in fragile X syndrome. The premutation alleles of the FMR1 gene, with 55–200 CGG repeats, are associated with various phenotypes, including fragile X-associated premature ovarian insufficiency (Toniolo and Rizzolio, 2007). These phenotypic manifestations could be a result of the UTR-APA of the FMR1 pre-mRNA. The choice of alternative 5′ UTRs affects the choice of poly(A) site (Winter et al., 2007; Ji et al., 2011). This could explain how the CGG expansion in the 5′ UTR of the FMR1 gene affects the choice of poly(A) site by altering the transcription start site (Tassone et al., 2011).

Diabetic nephropathy

A high-glucose-regulated gene, HGRG-14, found in human mesangial cells, is expressing a longer isoform of HGRG-14 mRNA due to UTR-APA under high-glucose conditions (Abdel Wahab et al., 1998). This results in lower HGRG-14 protein levels. The switch from the shorter to the longer isoform is detected within 2 h of exposure to high-glucose levels and the authors hypothesize that this may be involved in the pathogenesis of diabetic nephropathy (Abdel Wahab et al., 1998).

Medullary thyroid carcinoma

Human Thyroid C cells express an alternative isoform of calcitonin mRNA, containing all exons of the CALCA gene, producing calcitonin, and calcitonin carboxyl terminal peptide II (CCP II). Expression of this isoform is seen in low levels in normal subjects. In medullary thyroid carcinomas expression is more abundant, leading to higher plasma levels of CCP II (Minvielle et al., 1991). CCP II could thus be used as a diagnostic marker for medullary thyroid carcinomas. As the alternative processing of the CALCA gene occurs through APA, polyadenylation regulation may be aberrant in the medullary thyroid carcinomas.

Pre-eclampsia

Soluble fms-like tyrosine kinase 1 (sFlt-1) arises from CR-APA utilizing an intronic poly(A) site in the FLT1 gene encoding VEGF receptor 1 (Thomas et al., 2007). sFlt-1 binds VEGF and thereby reduces its free circulating level. Expression of sFlt-1 is upregulated in cytotrophoblasts during hypoxia and is suggested to induce endothelial damage (Nagamatsu et al., 2004; Zhao et al., 2012). sFlt-1 expression is increased in pre-eclampsia and evidence suggests that it causes the development of the disease (Shibata et al., 2005; Zhao et al., 2012). Selective expression of sFlt-1

can be induced by targeting the 5′ splice site with an antisense oligonucleotide, thereby inhibiting U1 binding (Vorlová et al., 2011). This hints that U1 plays a role in the upregulation of sFlt-1. Several techniques using antisense elements can inhibit poly(A) site usage (see section *Treatment by altering polyadenylation*) (Beckley et al., 2001; Fortes et al., 2003; Goraczniak et al., 2009; Vickers et al., 2011; Vorlová et al., 2011; Blazquez et al., 2012; Vickers and Crooke, 2012). These techniques could be used to lower the expression of sFlt-1, possibly alleviating the symptoms or even treating pre-eclampsia in the future. Interestingly, sFlt-1 is also increased in type II diabetes (Nandy et al., 2010).

Ectopic Cushing syndrome

Ectopic Cushing syndrome is cause by adrenocorticotropic hormone (ACTH) production in non-pituitary tumors, such as small cell lung cancers, which is normally non-suppressible (Parks et al., 1998). An ACTH-producing small cell lung cancer was shown to express a C-terminally truncated isoform of glucocorticoid receptor due to CR-APA using an intronic poly(A) site. This caused the ACTH production to be non-suppressible by exogenous glucocorticoid administration (Parks et al., 1998). Several techniques using antisense elements can inhibit poly(A) site usage (see Treatment by altering polyadenylation) (Beckley et al., 2001; Fortes et al., 2003; Goraczniak et al., 2009; Vickers et al., 2011; Vorlová et al., 2011; Blazquez et al., 2012; Vickers and Crooke, 2012). If these techniques were used to target the intronic poly(A) site utilized by CR-APA, the normal full length isoform of glucocorticoid receptor would be expressed. This would render the ACTH production suppressible by glucocorticoids, possibly treating the Cushing syndrome.

Adrenal and gonadal dysfunction

The gene encoding the steroidogenic acute regulatory (StAR) protein express two mRNA isoforms (1.6 and 3.5 kb) differing only in the length of their 3′ UTR due to UTR-APA (Duan et al., 2009). StAR regulates the rate-limiting step in steroid biosynthesis. ACTH stimulates the StAR function by increasing cAMP and cAMP-protein kinase A (PKA) activity, leading to preferential expression of the longer, less stable 3.5 kb StAR mRNA isoform (Zhao et al., 2005). PKA also stimulate transcription of the RBP TIS11b, which selectively binds the 3.5 kb mRNA isoform and destabilizes it, while at the same time enhancing StAR protein translation (Duan et al., 2009). Aberrant expression of StAR due to alterations in polydenylation could cause abnormal steroid production and adrenal/gonadal dysfunction, as seen for other disruptive mutations leading to enhanced or diminished StAR expression (Okuhara et al., 2008; Sahakitrungruang et al., 2010).

TREATMENT BY ALTERING POLYADENYLATION
ANTISENSE ELEMENTS TARGETING SPECIFIC POLY(A) SITES
Enhancing poly(A) site usage

Specific usage of selected intronic poly(A) sites can be induced by targeting the upstream 5′ splice site with an antisense oligonucleotide, thereby inhibiting U1 binding (Vorlová et al., 2011). As U1 associating with 5′ splice sites represses adjacent downstream intronic poly(A) site, this enhances the usage of these intronic poly(A) sites and leads to expression of C-terminal truncated proteins (Kaida et al., 2010; Vorlová et al., 2011; Berg et al., 2012).

Unlike mRNAs with acquired premature termination codons, which also generate C-terminal truncated proteins, the mRNAs generated by usage of intronic poly(A) sites are not targeted by nonsense-mediated decay (Lejeune and Maquat, 2005). Thousands of dormant intronic poly(A) sites have been characterized (Yao et al., 2012a). Induced usage of any of these sites can lead to the expression of specific C-terminal truncated protein isoforms.

Inhibiting poly(A) site usage

Different techniques using antisense elements have been used to effectively inhibit poly(A) site usage. They include antisense oligonucleotides (Vorlová et al., 2011), siRNAs (Vickers and Crooke, 2012), and U1 modifications (Beckley et al., 2001; Fortes et al., 2003; Goraczniak et al., 2009; Vickers et al., 2011; Blazquez et al., 2012). The techniques can both induce APA and lead to significantly reduced gene expression, depending on the poly(A) site targeted.

DRUGS TARGETING POLYADENYLATION FACTORS

The specific PAP inhibitor cordycepin is an adenosine-nucleotide analog. It reduces polyadenylation efficiency and causes defects in transcription termination (Kondrashov et al., 2012). Cordycepin has been shown to inhibit proliferation and induce apoptosis in various cancer cell lines (Thomadaki et al., 2008; Imesch et al., 2011). Cordycepin also inhibit the induction of inflammatory genes mediated by cytokines (Kondrashov et al., 2012).

PERSPECTIVES

Recently, the field of polyadenylation has seen major progress. Advances in high-throughput sequencing have made it possible to characterize polyadenylation genome-wide and lowered the price to ~100$ per sample (Wang et al., 2013). This has stimulated the field and more research on polyadenylation is now being published. This will help to find more endocrine pathways regulated by polyadenylation and thus to discover novel implications of alterations in polyadenylation for endocrine disease. The lowered cost of genome-wide characterization of polyadenylation will also push forward the implementation of this technique into the clinic to improve diagnostics. For instance, cancers show characteristic changes in APA patterns (Mayr and Bartel, 2009; Singh et al., 2009; Fu et al., 2011; Elkon et al., 2012; Lin et al., 2012; Morris et al., 2012). The inclusion of APA screenings in diagnostics could therefore help to produce more accurate diagnoses by, e.g., differentiating between cancer subtypes. It is therefore crucial to increase the awareness about the importance of polyadenylation among physicians and scientists working with endocrinology.

Alternative polyadenylation is increasingly being recognized as an important regulator of the human transcriptome, like alternative splicing. It has been estimated that 52% of all CR-APA events and 80% of all UTR-APA events are regulated differentially between tissues, making tissue-specific UTR-APA events even more differentially regulated than tissue-specific alternative splicing (Wang et al., 2008). Interestingly, changes in the APA pattern is seen during changes in cellular conditions, such as proliferation, differentiation, cellular transformation, and dedifferentiation (Sandberg et al., 2008; Ji and Tian, 2009; Mayr and Bartel, 2009; Singh et al., 2009; Fu et al., 2011; Elkon et al., 2012; Lin et al., 2012;

Morris et al., 2012). Unlike mRNAs created by alternative splicing, mRNAs generated by CR-APA are not targeted by nonsense-mediated decay, thus lead to the production of C-terminally truncated protein isoforms (Lejeune and Maquat, 2005). Such isoforms can have significantly altered functions (Vorlová et al., 2011) and have been suggested to cause diseases such as pre-ecclamsia (Zhao et al., 2012). C-terminally truncated protein isoforms generated by CR-APA could be causative of many other diseases. UTR-APA alters the 3′ UTR length and induces changes in stability and translational efficiency for individual mRNAs (Sandberg et al., 2008; Mayr and Bartel, 2009; Singh et al., 2009; Hogg and Goff, 2010; Yepiskoposyan et al., 2011). However, the effect of global changes in 3′ UTR length remains mysterious as no significant change in mRNA levels were seen in several cases with general changes in 3′ UTR length (Sandberg et al., 2008; Fu et al., 2011; Elkon et al., 2012; Morris et al., 2012). The latter findings could be explained by the mRNAs competing for trans-acting factors. For instance, in activated neurons, a general upregulation of pre-mRNAs, leads to a relative shortage of splice factor U1, which in turn, leads to changes in APA (Berg et al., 2012). Similarly, a general change in mRNA 3′ UTR length could lead to a relative shortage/abundance of trans-acting factors, giving rise to a relative weaker effect on gene expression levels, than seen for a single gene expressing an alternative 3′ UTR isoform. It would however still be interesting to investigate the changes in protein translation efficiency in cells with general changes in 3′ UTR length.

Both loss of function and gain of function changes in and around poly(A) sites have been shown to cause disease (reviewed in Danckwardt et al., 2008). It is probable that loss of function changes most frequently cause disease, when affecting the stronger, more canonical poly(A) site of a gene. Firstly, these sites are most efficient for polyadenylation and the most commonly utilized. Usage of these sites therefore give rise to higher mRNA expression levels (Wang et al., 2013). Secondly, these sites are often found as the most distal poly(A) site, which works as the last option for polyadenylation and ensures that transcription is terminated (Beaudoing et al., 2000; Tian et al., 2005, 2007; Shepard et al., 2011; de Klerk et al., 2012; Lin et al., 2012; Yoon et al., 2012; Wang et al., 2013). When loss of function changes affect these more canonical poly(A) sites, the effect on gene expression is therefore strong. This is seen for the *SERT* gene encoding a SERT. Here a loss of function change in the PAS of the distal, canonical poly(A) site, changes the PAS (AUUAAC to AGUAAC). This leads to lesser usage of the poly(A) site, reducing total SERT expression to about half (Gyawali et al., 2010). When loss of function changes affect the poly(A) site in genes where only one poly(A) site is found, the effect on gene expression is even stronger, as no alternative poly (A) sites may be used. This is seen for the *INS*, *FOXP3*, and *TP53* genes, where loss of function changes in the PAS of the only poly(A) site in these genes, lead to an almost diminished gene expression (Bennett et al., 2001; Garin et al., 2010; Stacey et al., 2011). When loss of function changes affect the weaker poly(A) sites in genes with multiple poly(A) sites, the effect is changes in the APA pattern, altering the 3′ UTR and in some cases the protein product. Gene expression is also affected, but often in a more subtle way. Interestingly, loss of function changes causing disease also commonly seem to affect the PAS (Higgs et al., 1983; Bennett

et al., 2001; Garin et al., 2010; Gyawali et al., 2010; Stacey et al., 2011). This is predictable, as the PAS is one of the two core cis-elements found at poly(A) sites. It is essential for binding CPSF, which is required for both the cleavage and the polyadenylation step (Chan et al., 2011). It is important to note that loss of function changes can also lead to subclinical manifestations, as seen in the metachromatic leukodystrophy pseudodeficiency phenotype. Here a change in the PAS (AAUAAC to AGUAAC) of a majorly used proximal poly(A) site in the *ARSA* gene reduces gene expression by 90%(Gieselmann et al., 1989). Conversely, it seems that gain of function changes associated with disease often affect proximal poly(A) sites. This is expected, as the proximal poly(A) sites are generally weaker and therefore have the biggest potential to be enhanced (Beaudoing et al., 2000; Tian et al., 2005, 2007; Shepard et al., 2011; de Klerk et al., 2012; Lin et al., 2012; Yoon et al., 2012; Wang et al., 2013). Interestingly, gain of function changes causing disease also commonly seem to affect other cis-elements than the PAS, as seen for the CA dinucleotide immediately 5′ of the cleavage site and for the DSE (Gehring et al., 2001; Ceelie et al., 2004; Danckwardt et al., 2004, 2006, 2007; Uitte de Willige et al., 2007). This is also anticipated, as improvement of auxiliary cis-elements can both enhance polyadenylation at poly(A) sites with a canonical PAS and at poly(A) sites with a non-canonical PAS.

Various techniques using specific antisense elements that enhance or inhibit the usage of specific poly(A) sites exists (Beckley et al., 2001; Fortes et al., 2003; Goraczniak et al., 2009; Vickers et al., 2011; Vorlová et al., 2011; Blazquez et al., 2012; Vickers and Crooke, 2012). These techniques hold great promises of novel therapeutic approaches, as they can both induce APA and change gene expression significantly. One obvious use would be to reverse the effect of disease-causing mutations affecting polyadenylation. However, they could also be used to treat diseases, where alterations in polyadenylation do not take part in the pathogenesis. One way of treating could be by inducing the expression of C-terminally truncated protein isoforms. This primarily leads to loss of function of the full length product, but the expressed C-terminally truncated protein isoforms can have beneficial functions on their own. This is seen for the VEGF receptor, where a C-terminally truncated isoform functions as a soluble decoy. Induction of this isoform resulted in an antiangiogenic effect both in targeted cells and in untreated cells exposed to the conditioned media (Vorlová et al., 2011). Another way to treat using these techniques, is by significantly reducing gene expression by inhibiting the more canonical poly(A) site (Goraczniak et al., 2009; Blazquez et al., 2012). A potential side-effect free usage of these techniques, would be to inhibit premature intronic poly(A) sites, which are used in diseased cells, but not in healthy cells. As polyadenylation is inhibited at the intronic poly(A) site, the targeted intron will be spliced out and degraded, leaving the healthy cells unaffected by the treatment.

In conclusion, the clinical implications of the current state of research in this field are: (1) Poly(A) site usage differs according to tissue type, developmental stage, genotype, and cancer subtype (Breton et al., 2001; Zhang et al., 2005; Kubo et al., 2006; Kwan et al., 2008; Sandberg et al., 2008; Wang et al., 2008; Ji and Tian, 2009; Ji et al., 2009; Singh et al., 2009; Thomsen et al., 2010; Derti et al., 2012; MacIsaac et al., 2012). Characterization of polyadenylation can therefore improve diagnostics and poly(A) site usage

would be a candidate diagnostic biomarker for, e.g., cancer classification. (2) Many diseases and symptoms are caused by defective polyadenylation (reviewed in Danckwardt et al., 2008). Several endocrine diseases are either caused by or associated with alterations in polyadenylation. These are neonatal diabetes (Garin et al., 2010), IPEX (Bennett et al., 2001), type I and II diabetes (Shin et al., 2007; Locke et al., 2011), pre-eclampsia (Zhao et al., 2012), fragile X-associated premature ovarian insufficiency (Tassone et al., 2011), ectopic Cushing syndrome (Parks et al., 1998), and several types of endocrine tumor diseases (Minvielle et al., 1991; Gartner et al., 2005). Consequently, the advances in this field therefore lead to a better understanding of these diseases and expose new possible drug targets. (3) Novel techniques using antisense elements that can both enhance and inhibit the usage of specific poly(A) sites have been developed (Beckley et al., 2001; Fortes et al., 2003; Goraczniak et al., 2009; Vickers et al., 2011; Vorlová et al., 2011;

Blazquez et al., 2012; Vickers and Crooke, 2012). These techniques hold the promise of novel therapeutic approaches, through desired induced changes in polyadenylation.

REFERENCES

Abdel Wahab, N., Gibbs, J., and Mason, R. M. (1998). Regulation of gene expression by alternative polyadenylation and mRNA instability in hyperglycaemic mesangial cells. *Biochem. J.* 336(Pt 2), 405–411.

Ahn, S. H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol. Cell* 13, 67–76.

Akhtar, M. N., Bukhari, S. A., Fazal, Z., Qamar, R., and Shahmuradov, I. A. (2010). POLYAR, a new computer program for prediction of poly(A) sites in human sequences. *BMC Genomics* 11:646. doi:10.1186/1471-2164-11-646

Akman, B. H., Can, T., and Erson-Bensan, A. E. (2012). Estrogen-induced upregulation and 3′-UTR shortening of CDC6. *Nucleic Acids Res.* 40, 10679–10688.

Amara, S. G., Evans, R. M., and Rosenfeld, M. G. (1984). Calcitonin/calcitonin gene-related peptide transcription unit: tissue-specific expression involves selective use of alternative polyadenylation sites. *Mol. Cell. Biol.* 4, 2151–2160.

Ara, T., Lopez, F., Ritchie, W., Benech, P., and Gautheret, D. (2006). Conservation of alternative polyadenylation patterns in mammalian genes. *BMC Genomics* 7:189. doi:10.1186/1471-2164-7-189

Arnold, M., Ellwanger, D. C., Hartsperger, M. L., Pfeufer, A., and Stümpflen, V. (2012). Cis-acting polymorphisms affect complex traits through modifications of microRNA regulation pathways. *PLoS ONE* 7:e36694. doi:10.1371/journal.pone.0036694

Bajic, V. B., Charn, T. H., Xu, J. X., Panda, S. K., and Krishnan, T. S. P.

(2005). Prediction models for DNA transcription termination based on SOM networks. *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, Shanghai, 5, 4791–4.

Barabino, S. M., and Keller, W. (1999). Last but not least: regulated poly(A) tail formation. *Cell* 99, 9–11.

Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* 10, 1001–1010.

Beckley, S. A., Liu, P., Stover, M. L., Gunderson, S. I., Lichtler, A. C., and Rowe, D. W. (2001). Reduction of target gene expression by a modified U1 snRNA. *Mol. Cell. Biol.* 21, 2815–2825.

Bennett, C. L., Brunkow, M. E., Ramsdell, F., O'Briant, K. C., Zhu, Q., Fuleihan, R. L., et al. (2001). A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA → AAUGAA) leads to the IPEX syndrome. *Immunogenetics* 53, 435–439.

Bennett, M., and Reed, R. (1993). Correspondence between a mammalian spliceosome component and an essential yeast splicing factor. *Science* 262, 105–108.

Berg, M. G., Singh, L. N., Younis, I., Liu, Q., Pinto, A. M., Kaida, D., et al. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53–64.

Bhattacharjee, R. B., and Bag, J. (2012). Depletion of nuclear poly(A) binding protein PABPN1 produces a compensatory response by cytoplasmic PABP4 and PABP5 in cultured human cells. *PLoS ONE* 7:e53036. doi:10.1371/journal.pone.0053036

Birse, C. E., Minvielle-Sebastia, L., Lee, B. A., Keller, W., and Proudfoot, N. J. (1998). Coupling termination of transcription to messenger RNA maturation in yeast. *Science* 280, 298–301.

Blazquez, L., Gonzalez-Rojas, S. J., Abad, A., Razquin, N., Abad, X., and Fortes, P. (2012). Increased in vivo inhibition of gene expression by combining RNA interference and U1 inhibition. *Nucleic Acids Res.* 40, e8.

Blechingberg, J., Lykke-Andersen, S., Jensen, T. H., Jørgensen, A. L., and Nielsen, A. L. (2007). Regulatory mechanisms for 3′-end alternative splicing and polyadenylation of the Glial Fibrillary Acidic Protein, GFAP, transcript. *Nucleic Acids Res.* 35, 7636–7650.

Bonapace, G., Concolino, D., Formicola, S., and Strisciuglio, P. (2003). A novel mutation in a patient with insulin-like growth factor 1 (IGF1) deficiency. *J. Med. Genet.* 40, 913–917.

Boutaud, O., Dixon, D. A., Oates, J. A., and Sawaoka, H. (2003). Tristetraprolin binds to the COX-2 mRNA 3′ untranslated region in cancer cells. *Adv. Exp. Med. Biol.* 525, 157–160.

Boutet, S. C., Cheung, T. H., Quach, N. L., Liu, L., Prescott, S. L., Edalati, A., et al. (2012). Alternative polyadenylation mediates microRNA regulation of muscle stem cell function. *Cell Stem Cell* 10, 327–336.

Breton, C., Di Scala-Guenot, D., and Zingg, H. H. (2001). Oxytocin receptor gene expression in rat mammary gland: structural characterization and regulation. *J. Mol. Endocrinol.* 27, 175–189.

Brown, K. M., and Gilmartin, G. M. (2003). A mechanism for the regulation of pre-mRNA 3′ processing by human cleavage factor Im. *Mol. Cell* 12, 1467–1476.

Buchert, M., Papin, M., Bonnans, C., Darido, C., Raye, W. S., Garambois, V., et al. (2010). Symplekin promotes tumorigenicity by up-regulating claudin-2 expression. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2628–2633.

Castelo-Branco, P., Furger, A., Wollerton, M., Smith, C., Moreira, A., and Proudfoot, N. (2004). Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol. Cell. Biol.* 24, 4174–4183.

Ceelie, H., Spaargaren-van Riel, C. C., Bertina, R. M., and Vos, H. L. (2004). G20210A is a functional mutation in the prothrombin gene; effect on protein levels and 3′-end formation. *J. Thromb. Haemost.* 2, 119–127.

Chan, S., Choi, E.-A., and Shi, Y. (2011). Pre-mRNA 3′-end processing complex assembly and function. *Wiley Interdisc. Rev. RNA* 2, 321–335.

Chen, F., and Wilusz, J. (1998). Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs. *Nucleic Acids Res.* 26, 2891–2898.

Chen, J.-M., Férec, C., and Cooper, D. N. (2006a). A systematic analysis of disease-associated variants in the 3′ regulatory regions of human protein-coding genes I: general principles and overview. *Hum. Genet.* 120, 1–21.

Chen, J.-M., Férec, C., and Cooper, D. N. (2006b). A systematic analysis of disease-associated variants in the 3′ regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3′ UTR variants. *Hum. Genet.* 120, 301–333.

Cheng, Y., Miura, R. M., and Tian, B. (2006). Prediction of mRNA polyadenylation sites by support vector machine. *Bioinformatics* 22, 2320–2325.

Chennathukuzhi, V. M., Lefrancois, S., Morales, C. R., Syed, V., and Hecht, N. B. (2001). Elevated levels of the polyadenylation factor CstF 64 enhance formation of the 1kB Testis brain RNA-binding protein (TB-RBP) mRNA in male germ cells. *Mol. Reprod. Dev.* 58, 460–469.

Connelly, S., and Manley, J. L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* 2, 440–452.

Cooke, C., and Alwine, J. C. (1996). The cap and the 3′ splice site similarly affect polyadenylation efficiency. *Mol. Cell. Biol.* 16, 2579–2584.

Cools, J., DeAngelo, D. J., Gotlib, J., Stover, E. H., Legare, R. D., Cortes, J., et al. (2003). A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. *N. Engl. J. Med.* 348, 1201–1214.

Coutinho, D. C., Coletta, R. R. D., Costa, E. M. F., Pachi, P. R., Boguszewski, M. C. S., Damiani, D., et al. (2007). Polymorphisms identified in the upstream core polyadenylation signal of IGF1 gene exon 6 do not cause pre- and postnatal growth impairment. *J. Clin. Endocrinol. Metab.* 92, 4889–4892.

Dalziel, M., Nunes, N. M., and Furger, A. (2007). Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3′ end processing. *Mol. Cell. Biol.* 27, 1568–1580.

Danckwardt, S., Gantzert, A.-S., Macher-Goeppinger, S., Probst, H. C., Gentzel, M., Wilm, M., et al. (2011). p38 MAPK controls prothrombin expression by regulated RNA 3′ end processing. *Mol. Cell* 41, 298–310.

Danckwardt, S., Gehring, N. H., Neu-Yilik, G., Hundsdoerfer, P., Pforsich, M., Frede, U., et al. (2004). The prothrombin 3′end formation signal reveals a unique architecture that is sensitive to thrombophilic gain-of-function mutations. *Blood* 104, 428–435.

Danckwardt, S., Hartmann, K., Katz, B., Hentze, M. W., Levy, Y., Eichele, R., et al. (2006). The prothrombin 20209 C → T mutation in Jewish-Moroccan Caucasians: molecular analysis of gain-of-function of 3′

end processing. *J. Thromb. Haemost.* 4, 1078–1085.

Danckwardt, S., Hentze, M. W., and Kulozik, A. E. (2008). 3′ end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.* 27, 482–498.

Danckwardt, S., Kaufmann, I., Gentzel, M., Foerstner, K. U., Gantzert, A.-S., Gehring, N. H., et al. (2007). Splicing factors stimulate polyadenylation via USEs at non-canonical 3′ end formation signals. *EMBO J.* 26, 2658–2669.

Darmon, S. K., and Lutz, C. S. (2012). mRNA 3′ end processing factors: a phylogenetic comparison. *Comp. Funct. Genomics* 2012, 876893.

Dass, B., Attaya, E. N., Michelle Wallace, A., and MacDonald, C. C. (2001). Overexpression of the CstF-64 and CPSF-160 polyadenylation protein messenger RNAs in mouse male germ cells. *Biol. Reprod.* 64, 1722–1729.

Dass, B., Tardif, S., Park, J. Y., Tian, B., Weitlauf, H. M., Hess, R. A., et al. (2007). Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20374–20379.

de Klerk, E., Venema, A., Anvar, S. Y., Goeman, J. J., Hu, O., Trollet, C., et al. (2012). Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res.* 40, 9089–9101.

Derti, A., Garrett-Engele, P., Macisaac, K. D., Stevens, R. C., Sriram, S., Chen, R., et al. (2012). A quantitative atlas of polyadenylation in five mammals. *Genome Res.* 22, 1173–1183.

Di Giammartino, D. C., Nishida, K., and Manley, J. L. (2011). Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* 43, 853–866.

Di Giammartino, D. C., Shi, Y., and Manley, J. L. (2013). PARP1 represses PAP and inhibits polyadenylation during heat shock. *Mol. Cell* 49, 7–17.

Dresser, D. W., Hacker, A., Lovell-Badge, R., and Guerrier, D. (1995). The genes for a spliceosome protein (SAP62) and the anti-Müllerian hormone (AMH) are contiguous. *Hum. Mol. Genet.* 4, 1613–1618.

Duan, H., Cherradi, N., Feige, J.-J., and Jefcoate, C. (2009). cAMP-dependent posttranscriptional regulation of steroidogenic acute regulatory (STAR) protein by the zinc finger protein ZFP36L1/TIS11b. *Mol. Endocrinol.* 23, 497–509.

Elkon, R., Drost, J., Van Haaften, G., Jenal, M., Schrier, M., Vrielink, J.

A. O., et al. (2012). E2F mediates enhanced alternative polyadenylation in proliferation. *Genome Biol.* 13, R59.

Evsyukova, I., Bradrick, S. S., Gregory, S. G., and Garcia-Blanco, M. A. (2012). Cleavage and polyadenylation specificity factor 1 (CPSF1) regulates alternative splicing of interleukin 7 receptor (IL7R) exon 6. *RNA* 19, 103–115.

Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I. W., and Gilmartin, G. M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3′ processing. *Proc. Natl. Acad. Sci. U.S.A.* 94, 11893–11898.

Flavell, S. W., Kim, T.-K., Gray, J. M., Harmin, D. A., Hemberg, M., Hong, E. J., et al. (2008). Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* 60, 1022–1038.

Fortes, P., Cuevas, Y., Guan, F., Liu, P., Pentlicky, S., Jung, S. P., et al. (2003). Inhibiting expression of specific genes in mammalian cells with 5′ end-mutated U1 small nuclear RNAs targeted to terminal exons of pre-mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8264–8269.

Foulkes, N. S., Schlotter, F., Pévet, P., and Sassone-Corsi, P. (1993). Pituitary hormone FSH directs the CREM functional switch during spermatogenesis. *Nature* 362, 264–267.

Fox-Walsh, K., Davis-Turak, J., Zhou, Y., Li, H., and Fu, X.-D. (2011). A multiplex RNA-seq strategy to profile poly(A+) RNA: application to analysis of transcription response and 3′ end formation. *Genomics* 98, 266–271.

Friedman, R. C., Farh, K. K.-H., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105.

Fu, Y., Sun, Y., Li, Y., Li, J., Rao, X., Chen, C., et al. (2011). Differential genome-wide profiling of tandem 3′ UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res.* 21, 741–747.

Garin, I., Edghill, E. L., Akerman, I., Rubio-Cabezas, O., Rica, I., Locke, J. M., et al. (2010). Recessive mutations in the INS gene result in neonatal diabetes through reduced insulin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3105–3110.

Gartner, W., Mineva, I., Daneva, T., Baumgartner-Parzer, S., Niederle, B., Vierhapper, H., et al. (2005).

A newly identified RET proto-oncogene polymorphism is found in a high number of endocrine tumor patients. *Hum. Genet.* 117, 143–153.

Gee, A. H., Kasprzak, W., and Shapiro, B. A. (2006). Structural differentiation of the HIV-1 polyA signals. *J. Biomol. Struct. Dyn.* 23, 417–428.

Gehring, N. H., Frede, U., Neu-Yilik, G., Hundsdoerfer, P., Vetter, B., Hentze, M. W., et al. (2001). Increased efficiency of mRNA 3′ end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat. Genet.* 28, 389–392.

Gieselmann, V., Polten, A., Kreysing, J., and Von Figura, K. (1989). Arylsulfatase A pseudodeficiency: loss of a polyadenylylation signal and N-glycosylation site. *Proc. Natl. Acad. Sci. U.S.A.* 86, 9436–9440.

Goraczniak, R., Behlke, M. A., and Gunderson, S. I. (2009). Gene silencing by synthetic U1 adaptors. *Nat. Biotechnol.* 27, 257–263.

Gotlib, J., and Cools, J. (2008). Five years since the discovery of FIP1L1-PDGFRA: what we have learned about the fusion and other molecularly defined eosinophilias. *Leukemia* 22, 1999–2010.

Graber, J. H., McAllister, G. D., and Smith, T. F. (2002). Probabilistic prediction of *Saccharomyces cerevisiae* mRNA 3′-processing sites. *Nucleic Acids Res.* 30, 1851–1858.

Gruber, A. R., Martin, G., Keller, W., and Zavolan, M. (2012). Cleavage Factor Im is a key regulator of 3′ UTR length. *RNA Biol.* 9, 1405–1412.

Gunderson, S. I., Polycarpou-Schwarz, M., and Mattaj, I. W. (1998). U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol. Cell* 1, 255–264.

Gunderson, S. I., Vagner, S., Polycarpou-Schwarz, M., and Mattaj, I. W. (1997). Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. *Genes Dev.* 11, 761–773.

Guo, J., Garrett, M., Micklem, G., and Brogna, S. (2011). Poly(A) signals located near the 5′ end of genes are silenced by a general mechanism that prevents premature 3′-end processing. *Mol. Cell. Biol.* 31, 639–651.

Guo, S. S., Wu, X., Shimoide, A. T., Wong, J., Moatamed, F., and Sawicki, M. P. (2003). Frequent overexpression of cyclin D1 in sporadic pancreatic endocrine tumours. *J. Endocrinol.* 179, 73–79.

Gyawali, S., Subaran, R., Weissman, M. M., Hershkowitz, D., McKenna, M. C., Talati, A., et al. (2010). Association of a polyadenylation polymorphism in the serotonin transporter and panic disorder. *Biol. Psychiatry* 67, 331–338.

Hajarnavis, A., Korf, I., and Durbin, R. (2004). A probabilistic model of 3′ end formation in *Caenorhabditis elegans*. *Nucleic Acids Res.* 32, 3392–3399.

Hall-Pogar, T., Liang, S., Hague, L. K., and Lutz, C. S. (2007). Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3′-UTR. *RNA* 13, 1103–1115.

Harteveld, C. L., Losekoot, M., Haak, H., Heister, G. A., Giordano, P. C., and Bernini, L. F. (1994). A novel polyadenylation signal mutation in the alpha 2-globin gene causing alpha thalassaemia. *Br. J. Haematol.* 87, 139–143.

Hartley, C. A., McKenna, M. C., Salman, R., Holmes, A., Casey, B. J., Phelps, E. A., et al. (2012). Serotonin transporter polyadenylation polymorphism modulates the retention of fear extinction memory. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5493–5498.

Heisler, L. K., Pronchuk, N., Nonogaki, K., Zhou, L., Raber, J., Tung, L., et al. (2007). Serotonin activates the hypothalamic-pituitary-adrenal axis via serotonin 2C receptor stimulation. *J. Neurosci.* 27, 6956–6964.

Hellquist, A., Zucchelli, M., Kivinen, K., Saarialho-Kere, U., Koskenmies, S., Widen, E., et al. (2007). The human GIMAP5 gene has a common polyadenylation polymorphism increasing risk to systemic lupus erythematosus. *J. Med. Genet.* 44, 314–321.

Higgs, D. R., Goodbourn, S. E., Lamb, J., Clegg, J. B., Weatherall, D. J., and Proudfoot, N. J. (1983). Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature* 306, 398–400.

Hilgers, V., Perry, M. W., Hendrix, D., Stark, A., Levine, M., and Haley, B. (2011). Neural-specific elongation of 3′ UTRs during Drosophila development. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15864–15869.

Ho, E. S., and Gunderson, S. I. (2011). Long conserved fragments upstream of Mammalian polyadenylation sites. *Genome Biol. Evol.* 3, 654–666.

Hockert, K. J., Martincic, K., Mendis-Handagama, S. M. L. C., Borghesi, L. A., Milcarek, C., Dass, B., et al. (2011). Spermatogenetic but not

immunological defects in mice lacking the τCstF-64 polyadenylation protein. *J. Reprod. Immunol.* 89, 26–37.

Hogg, J. R., and Goff, S. P. (2010). Upf1 senses 3′UTR length to potentiate mRNA decay. *Cell* 143, 379–389.

Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., et al. (2013). Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nat. Methods* 10, 133–139.

Hung, L.-H., Heiner, M., Hui, J., Schreiner, S., Benes, V., and Bindereif, A. (2008). Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* 14, 284–296.

Imesch, P., Hornung, R., Fink, D., and Fedier, A. (2011). Cordycepin (3′-deoxyadenosine), an inhibitor of mRNA polyadenylation, suppresses proliferation and activates apoptosis in human epithelial endometriotic cells in vitro. *Gynecol. Obstet. Invest.* 72, 43–49.

Jan, C. H., Friedman, R. C., Ruby, J. G., and Bartel, D. P. (2011). Formation, regulation and evolution of *Caenorhabditis elegans* 3′UTRs. *Nature* 469, 97–101.

Jankovic, L., Efremov, G. D., Petkov, G., Kattamis, C., George, E., Yang, K. G., et al. (1990). Two novel polyadenylation mutations leading to beta(+)-thalassemia. *Br. J. Haematol.* 75, 122–126.

Jenal, M., Elkon, R., Loayza-Puch, F., Van Haaften, G., Kühn, U., Menzies, F. M., et al. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* 149, 538–553.

Ji, Z., Lee, J. Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7028–7033.

Ji, Z., Luo, W., Li, W., Hoque, M., Pan, Z., Zhao, Y., et al. (2011). Transcriptional activity regulates alternative cleavage and polyadenylation. *Mol. Syst. Biol.* 7, 534.

Ji, Z., and Tian, B. (2009). Reprogramming of 3′ untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS ONE* 4:e8419. doi:10.1371/journal.pone.0008419

Jonkers, Y. M. H., Claessen, S. M. H., Perren, A., Schmitt, A. M., Hofland, L. J., De Herder, W., et al.

(2007). DNA copy number status is a powerful predictor of poor survival in endocrine pancreatic tumor patients. *Endocr. Relat. Cancer* 14, 769–779.

Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., et al. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668.

Keene, J. D. (2007). RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.* 8, 533–543.

Khaladkar, M., Smyda, M., and Hannenhalli, S. (2011). Epigenomic and RNA structural correlates of polyadenylation. *RNA Biol.* 8, 529–537.

Kiefer, H., Mizutani, A., Iemura, S.-I., Natsume, T., Ando, H., Kuroda, Y., et al. (2009). Inositol 1,4,5-triphosphate receptor-binding protein released with inositol 1,4,5-triphosphate (IRBIT) associates with components of the mRNA 3′ processing machinery in a phosphorylation-dependent manner and inhibits polyadenylation. *J. Biol. Chem.* 284, 10694–10705.

Kim, H., Lee, J. H., and Lee, Y. (2003). Regulation of poly(A) polymerase by 14-3-3epsilon. *EMBO J.* 22, 5208–5219.

Kim, S., Yamamoto, J., Chen, Y., Aida, M., Wada, T., Handa, H., et al. (2010). Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes Cells* 15, 1003–1013.

Kleiman, F. E., and Manley, J. L. (2001). The BARD1-CstF-50 interaction links mRNA 3′ end formation to DNA damage and tumor suppression. *Cell* 104, 743–753.

Ko, B., and Gunderson, S. I. (2002). Identification of new poly(A) polymerase-inhibitory proteins capable of regulating pre-mRNA polyadenylation. *J. Mol. Biol.* 318, 1189–1206.

Kondrashov, A., Meijer, H. A., Barthet-Barateig, A., Parker, H. N., Khurshid, A., Tessier, S., et al. (2012). Inhibition of polyadenylation reduces inflammatory gene induction. *RNA* 18, 2236–2250.

Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A., and Handa, H. (2006). 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.

Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdel, S., and Wahle, E. (2009). Poly(A) tail length is

controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *J. Biol. Chem.* 284, 22803–22814.

Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., et al. (2008). Genome-wide analysis of transcript isoform variation in humans. *Nat. Genet.* 40, 225–231.

Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3′ end processing and splicing. *Mol. Cell* 23, 195–205.

Laishram, R. S., Barlow, C. A., and Anderson, R. A. (2011). CKI isoforms α and ε regulate Star-PAP target messages by controlling Star-PAP poly(A) polymerase activity and phosphoinositide stimulation. *Nucleic Acids Res.* 39, 7961–7973.

Lee, E. K., and Gorospe, M. (2010). Minireview: posttranscriptional regulation of the insulin and insulin-like growth factor systems. *Endocrinology* 151, 1403–1408.

Lee, Y. J., Lee, Y., and Chung, J. H. (2000). An intronless gene encoding a poly(A) polymerase is specifically expressed in testis. *FEBS Lett.* 487, 287–292.

Legendre, M., and Gautheret, D. (2003). Sequence determinants in human polyadenylation site selection. *BMC Genomics* 4:7. doi:10.1186/1471-2164-4-7

Legendre, M., Ritchie, W., Lopez, F., and Gautheret, D. (2006). Differential repression of alternative transcripts: a screen for miRNA targets. *PLoS Comput. Biol.* 2:e43. doi:10.1371/journal.pcbi.0020043

Lejeune, F., and Maquat, L. E. (2005). Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* 17, 309–315.

Lembo, A., Di Cunto, F., and Provero, P. (2012). Shortening of 3′UTRs correlates with poor prognosis in breast and lung cancer. *PLoS ONE* 7:e31129. doi:10.1371/journal.pone.0031129

Levy, J. R., Hannah, S., Mooney, R. L., Hug, V., and Stevens, W. (1995). Sequence and functional characterization of the terminal exon of the human insulin receptor gene. *Biochim. Biophys. Acta* 1263, 253–257.

Li, B., Zhang, G., Li, C., He, D., Li, X., Zhang, C., et al. (2012a).

Identification of JAK2 as a mediator of FIP1L1-PDGFRA-induced eosinophil growth and function in CEL. *PLoS ONE* 7:e34912. doi:10.1371/journal.pone.0034912

Li, W., Laishram, R. S., Ji, Z., Barlow, C. A., Tian, B., and Anderson, R. A. (2012b). Star-PAP control of BIK expression and apoptosis is regulated by nuclear PIPKIα and PKCδ signaling. *Mol. Cell* 45, 25–37.

Li, W., Yeh, H.-J., Shankarling, G. S., Ji, Z., Tian, B., and Macdonald, C. C. (2012c). The τCstF-64 polyadenylation protein controls genome expression in testis. *PLoS ONE* 7:e48373. doi:10.1371/journal.pone.0048373

Lian, Z., Karpikov, A., Lian, J., Mahajan, M. C., Hartman, S., Gerstein, M., et al. (2008). A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3′ end RNA polyadenylation. *Genome Res.* 18, 1224–1237.

Licatalosi, D. D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J. B., et al. (2002). Functional interaction of yeast pre-mRNA 3′ end processing factors with RNA polymerase II. *Mol. Cell* 9, 1101–1111.

Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464–469.

Lin, Y., Li, Z., Ozsolak, F., Kim, S. W., Arango-Argoty, G., Liu, T. T., et al. (2012). An in-depth map of polyadenylation sites in cancer. *Nucleic Acids Res.* 40, 8460–8471.

Lisman, T., De Groot, P. G., Meijers, J. C. M., and Rosendaal, F. R. (2005). Reduced plasma fibrinolytic potential is a risk factor for venous thrombosis. *Blood* 105, 1102–1105.

Liu, D., Brockman, J. M., Dass, B., Hutchins, L. N., Singh, P., McCarrey, J. R., et al. (2007). Systematic variation in mRNA 3′-processing signals during mouse spermatogenesis. *Nucleic Acids Res.* 35, 234–246.

Liu, D., and Waxman, D. J. (2002). Post-transcriptional regulation of hepatic NADPH-cytochrome P450 reductase by thyroid hormone: independent effects on poly(A) tail length and mRNA stability. *Mol. Pharmacol.* 61, 1089–1096.

Locke, J. M., Da Silva Xavier, G., Rutter, G. A., and Harries, L. W. (2011). An alternative polyadenylation signal in TCF7L2 generates isoforms that inhibit T cell factor/lymphoid-enhancer factor (TCF/LEF)-dependent target genes. *Diabetologia* 54, 3078–3082.

Lou, H., Neugebauer, K. M., Gagel, R. F., and Berget, S. M. (1998). Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol. Cell. Biol.* 18, 4977–4985.

Luna, R., Jimeno, S., Marín, M., Huertas, P., García-Rubio, M., and Aguilera, A. (2005). Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. *Mol. Cell* 18, 711–722.

Lunde, B. M., Reichow, S. L., Kim, M., Suh, H., Leeper, T. C., Yang, F., et al. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* 17, 1195–1201.

MacDonald, C. C., Wilusz, J., and Shenk, T. (1994). The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol. Cell. Biol.* 14, 6647–6654.

MacIsaac, J. L., Bogutz, A. B., Morrissy, A. S., and Lefebvre, L. (2012). Tissue-specific alternative polyadenylation at the imprinted gene Mest regulates allelic usage at Copg2. *Nucleic Acids Res.* 40, 1523–1535.

Mandel, C. R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J. L., et al. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature* 444, 953–956.

Mangone, M., Manoharan, A. P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S. D., et al. (2010). The landscape of C. elegans 3′UTRs. *Science* 329, 432–435.

Mapendano, C. K., Lykke-Andersen, S., Kjems, J., Bertrand, E., and Jensen, T. H. (2010). Crosstalk between mRNA 3′ end processing and transcription initiation. *Mol. Cell* 40, 410–422.

Martin, G., Gruber, A. R., Keller, W., and Zavolan, M. (2012). Genome-wide analysis of pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3′ UTR length. *Cell Rep.* 1, 753–763.

Martincic, K., Alkan, S. A., Cheatle, A., Borghesi, L., and Milcarek, C. (2009). Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. *Nat. Immunol.* 10, 1102–1109.

Martincic, K., Campbell, R., Edwalds-Gilbert, G., Souan, L., Lotze, M. T., and Milcarek, C. (1998). Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the G0 to S phase transition. *Proc. Natl. Acad. Sci. U.S.A.* 95, 11095–11100.

Mayr, C., and Bartel, D. P. (2009). Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138, 673–684.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., et al. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357–361.

McCracken, S., Lambermon, M., and Blencowe, B. J. (2002). SRm160 splicing coactivator promotes transcript 3′-end cleavage. *Mol. Cell. Biol.* 22, 148–160.

McCracken, S., Longman, D., Johnstone, I. L., Cáceres, J. F., and Blencowe, B. J. (2003). An evolutionarily conserved role for SRm160 in 3′-end processing that functions independently of exon junction complex formation. *J. Biol. Chem.* 278, 44153–44160.

Mellman, D. L., Gonzales, M. L., Song, C., Barlow, C. A., Wang, P., Kendziorski, C., et al. (2008). A PtdIns4,5P2-regulated nuclear poly(A) polymerase controls expression of select mRNAs. *Nature* 451, 1013–1017.

Mercer, T. R., Wilhelm, D., Dinger, M. E., Soldà, G., Korbie, D. J., Glazov, E. A., et al. (2011). Expression of distinct RNAs from 3′ untranslated regions. *Nucleic Acids Res.* 39, 2393–2403.

Millevoi, S., Decorsière, A., Loulergue, C., Iacovoni, J., Bernat, S., Antoniou, M., et al. (2009). A physical and functional link between splicing factors promotes pre-mRNA 3′ end processing. *Nucleic Acids Res.* 37, 4672–4683.

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S. Z., Keller, W., Antoniou, M., et al. (2006). An interaction between U2AF 65 and CF I(m) links the splicing and 3′ end processing machineries. *EMBO J.* 25, 4854–4864.

Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. *Nucleic Acids Res.* 38, 2757–2774.

Minvielle, S., Giscard-Dartevelle, S., Cohen, R., Taboulet, J., Labye, F., Jullienne, A., et al. (1991). A novel calcitonin carboxyl-terminal peptide produced in medullary thyroid carcinoma by alternative RNA processing of the calcitonin/calcitonin gene-related peptide gene. *J. Biol. Chem.* 266, 24627–24631.

Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J. L., and Proudfoot, N. J. (1998). The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3′ end formation by two distinct mechanisms. *Genes Dev.* 12, 2522–2534.

Morris, A. R., Bos, A., Diosdado, B., Rooijers, K., Elkon, R., Bolijn, A. S., et al. (2012). alternative cleavage and polyadenylation during colorectal cancer development. *Clin. Cancer Res.* 18, 5256–5266.

Murthy, K. G., and Manley, J. L. (1995). The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3′-end formation. *Genes Dev.* 9, 2672–2683.

Nagaike, T., Logan, C., Hotta, I., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J. L. (2011). Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol. Cell* 41, 409–418.

Nagamatsu, T., Fujii, T., Kusumi, M., Zou, L., Yamashita, T., Osuga, Y., et al. (2004). Cytotrophoblasts up-regulate soluble fms-like tyrosine kinase-1 expression under reduced oxygen: an implication for the placental vascular development and the pathophysiology of preeclampsia. *Endocrinology* 145, 4838–4845.

Nandy, D., Mukhopadhyay, D., and Basu, A. (2010). Both vascular endothelial growth factor and soluble Flt-1 are increased in type 2 diabetes but not in impaired fasting glucose. *J. Investig. Med.* 58, 804–806.

Nazeer, F. I., Devany, E., Mohammed, S., Fonseca, D., Akukwe, B., Taveras, C., et al. (2011). p53 inhibits mRNA 3′ processing through its interaction with the CstF/BARD1 complex. *Oncogene* 30, 3073–3083.

Neilson, J. R., and Sandberg, R. (2010). Heterogeneity in mammalian RNA 3′ end formation. *Exp. Cell Res.* 316, 1357–1364.

Nemeroff, M. E., Barabino, S. M., Li, Y., Keller, W., and Krug, R. M. (1998). Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3′end formation of cellular pre-mRNAs. *Mol. Cell* 1, 991–1000.

Nunes, N. M., Li, W., Tian, B., and Furger, A. (2010). A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. *EMBO J.* 29, 1523–1536.

Okuhara, K., Abe, S., Kondo, T., Fujita, K., Koda, N., Mochizuki, H., et al. (2008). Four Japanese patients with adrenal hypoplasia congenita and hypogonadotropic

hypogonadism caused by DAX-1 gene mutations: mutant DAX-1 failed to repress steroidogenic acute regulatory protein (StAR) and luteinizing hormone beta-subunit gene promoter acti. *Endocr. J.* 55, 97–103.

Oppelt, P., Strissel, P. L., Kellermann, A., Seeber, S., Humeny, A., Beckmann, M. W., et al. (2005). DNA sequence variations of the entire anti-Mullerian hormone (AMH) gene promoter and AMH protein expression in patients with the Mayer-Rokitanski-Kuster-Hauser syndrome. *Hum. Reprod.* 20, 149–157.

Orkin, S. H., Cheng, T. C., Antonarakis, S. E., and Kazazian, H. H. (1985). Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *EMBO J.* 4, 453–456.

Ozsolak, F., Kapranov, P., Foissac, S., Kim, S. W., Fishilevich, E., Monaghan, A. P., et al. (2010). Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* 143, 1018–1029.

Parks, L. L., Turney, M. K., Detera-Wadleigh, S., and Kovacs, W. J. (1998). An ACTH-producing small cell lung cancer expresses aberrant glucocorticoid receptor transcripts from a normal gene. *Mol. Cell. Endocrinol.* 142, 175–181.

Pauws, E., Van Kampen, A. H., Van de Graaf, S. A., De Vijlder, J. J., and Ris-Stalpers, C. (2001). Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. *Nucleic Acids Res.* 29, 1690–1694.

Pelechano, V., Wilkening, S., Järvelin, A. I., Tekkedil, M. M., and Steinmetz, L. M. (2012). Genome-wide polyadenylation site mapping. *Meth. Enzymol.* 513, 271–296.

Phillips, C., Pachikara, N., and Gunderson, S. I. (2004). U1A inhibits cleavage at the immunoglobulin M heavy-chain secretory poly(A) site by binding between the two downstream GU-rich regions. *Mol. Cell. Biol.* 24, 6162–6171.

Pinto, P. A. B., Henriques, T., Freitas, M. O., Martins, T., Domingues, R. G., Wyrzykowska, P. S., et al. (2011). RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO J.* 30, 2431–2444.

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

Ptitsyn, A. A., and Gimble, J. M. (2007). Analysis of circadian pattern reveals tissue-specific alternative transcription in leptin signaling pathway. *BMC Bioinformatics* 8(Suppl. 7):S15. doi:10.1186/1471-2105-8-S7-S15

Retelska, D., Iseli, C., Bucher, P., Jongeneel, C. V., and Naef, F. (2006). Similarities and differences of polyadenylation signals in human and fly. *BMC Genomics* 7:176. doi:10.1186/1471-2164-7-176

Rozenblatt-Rosen, O., Nagaike, T., Francis, J. M., Kaneko, S., Glatt, K. A., Hughes, C. M., et al. (2009). The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3′ mRNA processing factors. *Proc. Natl. Acad. Sci. U.S.A.* 106, 755–760.

Rund, D., Dowling, C., Najjar, K., Rachmilewitz, E. A., Kazazian, H. H., and Oppenheim, A. (1992). Two mutations in the beta-globin polyadenylylation signal reveal extended transcripts and new RNA polyadenylylation sites. *Proc. Natl. Acad. Sci. U.S.A.* 89, 4324–4328.

Ryan, K., and Bauer, D. L. V. (2008). Finishing touches: post-translational modification of protein factors involved in mammalian pre-mRNA 3′ end formation. *Int. J. Biochem. Cell Biol.* 40, 2384–2396.

Sahakitrungruang, T., Soccio, R. E., Lang-Muritano, M., Walker, J. M., Achermann, J. C., and Miller, W. L. (2010). Clinical, genetic, and functional characterization of four patients carrying partial loss-of-function mutations in the steroidogenic acute regulatory protein (StAR). *J. Clin. Endocrinol. Metab.* 95, 3352–3359.

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., and Burge, C. B. (2008). Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* 320, 1643–1647.

Sartini, B. L., Wang, H., Wang, W., Millette, C. F., and Kilpatrick, D. L. (2008). Pre-messenger RNA cleavage factor I (CFIm): potential role in alternative polyadenylation during spermatogenesis. *Biol. Reprod.* 78, 472–482.

Shankarling, G. S., Coates, P. W., Dass, B., and Macdonald, C. C. (2009). A family of splice variants of CstF-64 expressed in vertebrate nervous systems. *BMC Mol. Biol.* 10:22. doi:10.1186/1471-2199-10-22

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

Shell, S. A., Hesse, C., Morris, S. M., and Milcarek, C. (2005). Elevated levels of the 64-kDa cleavage stimulatory factor (CstF-64) in lipopolysaccharide-stimulated macrophages influence gene expression and induce alternative poly(A) site selection. *J. Biol. Chem.* 280, 39950–39961.

Shepard, P. J., Choi, E.-A., Lu, J., Flanagan, L. A., Hertel, K. J., and Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* 17, 761–772.

Shi, Y., Di Giammartino, D. C., Taylor, D., Sarkeshik, A., Rice, W. J., Yates, J. R., et al. (2009). Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol. Cell* 33, 365–376.

Shibata, E., Rajakumar, A., Powers, R. W., Larkin, R. W., Gilmour, C., Bodnar, L. M., et al. (2005). Soluble fms-like tyrosine kinase 1 is increased in preeclampsia but not in normotensive pregnancies with small-for-gestational-age neonates: relationship to circulating placental growth factor. *J. Clin. Endocrinol. Metab.* 90, 4895–4903.

Shin, J.-H., Janer, M., McNeney, B., Blay, S., Deutsch, K., Sanjeevi, C. B., et al. (2007). IA-2 autoantibodies in incident type I diabetes patients are associated with a polyadenylation signal polymorphism in GIMAP5. *Genes Immun.* 8, 503–512.

Singh, P., Alley, T. L., Wright, S. M., Kamdar, S., Schott, W., Wilpan, R. Y., et al. (2009). Global changes in processing of mRNA 3′ untranslated regions characterize clinically distinct cancer subtypes. *Cancer Res.* 69, 9422–9430.

Spies, N., Nielsen, C. B., Padgett, R. A., and Burge, C. B. (2009). Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* 36, 245–254.

Stacey, S. N., Sulem, P., Jonasdottir, A., Masson, G., Gudmundsson, J., Gudbjartsson, D. F., et al. (2011). A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat. Genet.* 43, 1098–1103.

Stirling, P. C., Chan, Y. A., Minaker, S. W., Aristizabal, M. J., Barrett, I., Sipahimalani, P., et al. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev.* 26, 163–175.

Tabaska, J. E., and Zhang, M. Q. (1999). Detection of polyadenylation signals in human DNA sequences. *Gene* 231, 77–86.

Takagaki, Y., and Manley, J. L. (1998). 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.

Tassone, F., De Rubeis, S., Carosi, C., La Fata, G., Serpa, G., Raske, C., et al. (2011). Differential usage of transcriptional start sites and polyadenylation sites in FMR1 premutation alleles. *Nucleic Acids Res.* 39, 6172–6185.

Thomadaki, H., Scorilas, A., Tsiapalis, C. M., and Havredaki, M. (2008). The role of cordycepin in cancer treatment via induction or inhibition of apoptosis: implication of polyadenylation in a cell type specific manner. *Cancer Chemother. Pharmacol.* 61, 251–265.

Thomas, C. P., Andrews, J. I., and Liu, K. Z. (2007). Intronic polyadenylation signal sequences and alternate splicing generate human soluble Flt1 variants and regulate the abundance of soluble Flt1 in the placenta. *FASEB J.* 21, 3885–3895.

Thomas, F., and Sætrom, P. (2012). Single nucleotide polymorphisms can create alternative polyadenylation signals and affect gene expression through loss of microRNA-regulation. *PLoS Comput. Biol.* 8:e1002621. doi:10.1371/journal.pcbi.1002621

Thomsen, S., Azzam, G., Kaschula, R., Williams, L. S., and Alonso, C. R. (2010). Developmental RNA processing of 3′UTRs in Hox mRNAs as a context-dependent mechanism modulating visibility to microRNAs. *Development* 137, 2951–2960.

Tian, B., and Graber, J. H. (2012). Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA* 3, 385–396.

Tian, B., Hu, J., Zhang, H., and Lutz, C. S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* 33, 201–212.

Tian, B., Pan, Z., and Lee, J. Y. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res.* 17, 156–165.

Toniolo, D., and Rizzolio, F. (2007). X chromosome and ovarian failure. *Semin. Reprod. Med.* 25, 264–271.

Topalian, S. L., Kaneko, S., Gonzales, M. I., Bond, G. L., Ward, Y., and Manley, J. L. (2001). Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in

human tumors. *Mol. Cell. Biol.* 21, 5614–5623.

Touriol, C., Morillon, A., Gensac, M. C., Prats, H., and Prats, A. C. (1999). Expression of human fibroblast growth factor 2 mRNA is post-transcriptionally controlled by a unique destabilizing element present in the 3′-untranslated region between alternative polyadenylation sites. *J. Biol. Chem.* 274, 21402–21408.

Uitte de Willige, S., Rietveld, I. M., De Visser, M. C. H., Vos, H. L., and Bertina, R. M. (2007). Polymorphism 10034C>T is located in a region regulating polyadenylation of FGG transcripts and influences the fibrinogen gamma′/gammaA mRNA ratio. *J. Thromb. Haemost.* 5, 1243–1249.

Van Oers, C. C., Bakker, L., and Baas, P. D. (1994). The exon 4 poly(A) site of the human calcitonin/CGRP-I pre-mRNA is a weak site in vitro. *Biochim. Biophys. Acta* 1218, 55–63.

van Solinge, W. W., Lind, B., Van Wijk, R., Hart, H. C., and Kraaijenhagen, R. J. (1996). Clinical expression of a rare beta-globin gene mutation co-inherited with haemoglobin E-disease. *Eur. J. Clin. Chem. Clin. Biochem.* 34, 949–954.

Venkataraman, K., Brown, K. M., and Gilmartin, G. M. (2005). Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev.* 19, 1315–1327.

Veraldi, K. L., Arhin, G. K., Martincic, K., Chung-Ganster, L. H., Wilusz, J., and Milcarek, C. (2001). hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. *Mol. Cell. Biol.* 21, 1228–1238.

Vickers, T. A., and Crooke, S. T. (2012). siRNAs targeted to certain polyadenylation sites promote specific, RISC-independent degradation of messenger RNAs. *Nucleic Acids Res.* 40, 6223–6234.

Vickers, T. A., Sabripour, M., and Crooke, S. T. (2011). U1 adaptors result in reduction of multiple pre-mRNA species principally by sequestering U1snRNP. *Nucleic Acids Res.* 39, e71.

Vorlová, S., Rocco, G., Lefave, C. V., Jodelka, F. M., Hess, K., Hastings, M. L., et al. (2011). Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Mol. Cell* 43, 927–939.

Wahba, L., Amon, J. D., Koshland, D., and Vuica-Ross, M. (2011).

RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell* 44, 978–988.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.

Wang, L., Dowell, R. D., and Yi, R. (2013). Genome-wide maps of polyadenylation reveal dynamic mRNA 3′-end formation in mammalian cell lineages. *RNA* 19, 413–425.

Wang, Y., Fairley, J. A., and Roberts, S. G. E. (2010). Phosphorylation of TFIIB links transcription initiation and termination. *Curr. Biol.* 20, 548–553.

West, S., and Proudfoot, N. J. (2008). Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. *Nucleic Acids Res.* 36, 905–914.

West, S., and Proudfoot, N. J. (2009). Transcriptional termination enhances protein expression in human cells. *Mol. Cell* 33, 354–364.

Wiestner, A., Tehrani, M., Chiorazzi, M., Wright, G., Gibellini, F., Nakayama, K., et al. (2007). 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.

Wilkening, S., Pelechano, V., Järvelin, A. I., Tekkedil, M. M., Anders, S., Benes, V., et al. (2013). An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res.* 41, e65.

Winter, J., Kunath, M., Roepcke, S., Krause, S., Schneider, R., and Schweiger, S. (2007). Alternative polyadenylation signals and promoters act in concert to control tissue-specific expression of the Opitz syndrome gene MID1. *BMC Mol. Biol.* 8:105. doi:10.1186/1471-2199-8-105

Wood, A. J., Schulz, R., Woodfine, K., Koltowska, K., Beechey, C. V., Peters, J., et al. (2008). Regulation of alternative polyadenylation by genomic imprinting. *Genes Dev.* 22, 1141–1146.

Xing, H., Mayhew, C. N., Cullen, K. E., Park-Sarge, O.-K., and Sarge, K. D. (2004). HSF1 modulation of Hsp70 mRNA polyadenylation via interaction with symplekin. *J. Biol. Chem.* 279, 10551–10555.

Yan, J., and Marr, T. G. (2005). Computational analysis of 3′-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. *Genome Res.* 15, 369–375.

Yang, Q., Gilmartin, G. M., and Doublié, S. (2011). The structure of human cleavage factor I(m) hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5′ capping and splicing. *RNA Biol.* 8, 748–753.

Yang, Y., Ho, S. C. L., and Yap, M. G. S. (2009). Mutated polyadenylation signals for controlling expression levels of multiple genes in mammalian cells. *Biotechnol. Bioeng.* 102, 1152–1160.

Yao, C., Biesinger, J., Wan, J., Weng, L., Xing, Y., Xie, X., et al. (2012a). Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18773–18778.

Yao, P., Potdar, A. A., Arif, A., Ray, P. S., Mukhopadhyay, R., Willard, B., et al. (2012b). Coding region polyadenylation generates a truncated tRNA synthetase that counters translation repression. *Cell* 149, 88–100.

Yepiskoposyan, H., Aeschimann, F., Nilsson, D., Okoniewski, M., and Mühlemann, O. (2011). Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* 17, 2108–2118.

Yoon, O. K., Hsu, T. Y., Im, J. H., and Brem, R. B. (2012). Genetics and regulatory impact of alternative polyadenylation in human B-lymphoblastoid cells. *PLoS Genet.* 8:e1002882. doi:10.1371/journal.pgen.1002882

Yoshimoto, K., Iwahana, H., Fukuda, A., Sano, T., Saito, S., and Itakura, M. (1992). Role of p53 mutations in endocrine tumorigenesis: mutation detection by polymerase chain reaction-single strand conformation polymorphism. *Cancer Res.* 52, 5061–5064.

Yu, K., Ganesan, K., Tan, L. K., Laban, M., Wu, J., Zhao, X. D., et al. (2008). A precisely regulated gene expression cassette potently modulates metastasis and survival in multiple solid cancers. *PLoS Genet.* 4:e1000129. doi:10.1371/journal.pgen.1000129

Zhang, H., Lee, J. Y., and Tian, B. (2005). Biased alternative polyadenylation in human tissues. *Genome Biol.* 6, R100.

Zhao, D., Duan, H., Kim, Y.-C., and Jefcoate, C. R. (2005). Rodent StAR mRNA is substantially regulated by control of mRNA stability through sites in the 3′-untranslated region and through coupling to ongoing transcription. *J. Steroid Biochem. Mol. Biol.* 96, 155–173.

Zhao, Y., Koga, K., Osuga, Y., Nagai, M., Izumi, G., Takamura, M., et al. (2012). Thrombin enhances soluble Fms-like tyrosine kinase 1 expression in trophoblasts; possible involvement in the pathogenesis of preeclampsia. *Fertil. Steril.* 98, 917–921.

Zhou, H.-L., Baraniak, A. P., and Lou, H. (2007). Role for Fox-1/Fox-2 in mediating the neuronal pathway of calcitonin/calcitonin gene-related peptide alternative RNA processing. *Mol. Cell. Biol.* 27, 830–841.

Zhou, L., Yuan, Q., and Yang, M. (2012). A functional germline variant in the P53 polyadenylation signal and risk of esophageal squamous cell carcinoma. *Gene* 506, 295–297.

Zhu, H., Zhou, H.-L., Hasman, R. A., and Lou, H. (2007). Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. *J. Biol. Chem.* 282, 2203–2210.

Zhu, Z.-H., Yu, Y. P., Shi, Y.-K., Nelson, J. B., and Luo, J.-H. (2009). CSR1 induces cell death through inactivation of CPSF3. *Oncogene* 28, 41–51.