Review:

The polyadenylation code: a unified model for the regulation of mRNA alternative polyadenylation[*]

Ryan DAVIS, Yongsheng SHI[†‡]

(*Department of Microbiology and Molecular Genetics, School of Medicine, University of California, Irvine, Irvine, CA 92697, USA*)

[†]E-mail: yongshes@uci.edu

Abstract: The majority of eukaryotic genes produce multiple mRNA isoforms with distinct 3' ends through a process called mRNA alternative polyadenylation (APA). Recent studies have demonstrated that APA is dynamically regulated during development and in response to environmental stimuli. A number of mechanisms have been described for APA regulation. In this review, we attempt to integrate all the known mechanisms into a unified model. This model not only explains most of previous results, but also provides testable predictions that will improve our understanding of the mechanistic details of APA regulation. Finally, we briefly discuss the known and putative functions of APA regulation.

1 Introduction

For the majority of eukaryotic messenger RNAs (mRNAs), their 3' ends are formed co-transcriptionally through an endonucleolytic cleavage and polyadenylation (Colgan and Manley, 1997; Zhao *et al.*, 1999; Chan *et al.*, 2011). The polyadenylation site (PAS) is defined by a number of key *cis*-elements, including the AAUAAA element and the U/GU-rich downstream sequence element (DSE) (Fig. 1) (Tian and Graber, 2012). Several auxiliary elements have been identified that stimulate the PAS recognition, including the U-rich upstream sequence element (USE) and G-rich element (GRE) near DSE (Tian and Graber, 2012). These sequence elements are specifically recognized by a number of protein factors (Fig. 1). For example, the cleavage and polyadenylation specificity factor (CPSF) binds to the AAUAAA hexamer, cleavage stimulatory factor (CstF) binds to the DSE, and cleavage factor Im (CFIm) recognizes UGUA-containing USE (Colgan and Manley, 1997; Zhao *et al.*, 1999; Chan *et al.*, 2011). These factors bind to the PAS RNA sequence to form the core mRNA 3' processing complex and they in turn recruit cleavage factor IIm (CFIIm) and poly(A) polymerase (PAP) to initiate cleavage and polyadenylation (Shi *et al.*, 2009). The assembly of these factors on RNAs is facilitated by the C-terminal domain of the RNA polymerase II (RNAP II) (Hirose and Manley, 2000; Proudfoot *et al.*, 2002; Bentley, 2005).

The majority of eukaryotic genes have the capacity to generate multiple alternative 3' ends (Shi, 2012; Elkon *et al.*, 2013; Tian and Manley, 2013). This is due to the presence of multiple alternative PAS within the pre-mRNAs. Alternative polyadenylation (APA) isoforms of the same gene could encode different proteins and/or bear different 3' untranslated regions (UTRs). As 3' UTRs are rich in regulatory sequences, APA isoforms may have different stabilities, translation efficiencies, or subcellular localizations. Recent studies using high throughput tools have revealed that APA is highly regulated in development (Flavell *et al.*, 2008; Sandberg *et al.*, 2008; Ji *et al.*, 2009; Jan *et al.*, 2010;

Shepard *et al.*, 2011; Smibert *et al.*, 2012), and aberrant APA profiles have been detected in a number of human diseases (Mayr and Bartel, 2009; Park *et al.*, 2011; de Klerk *et al.*, 2012; Jenal *et al.*, 2012). Therefore, it is important to understand the "polyadenylation code", the rules that govern PAS selection and its regulation.
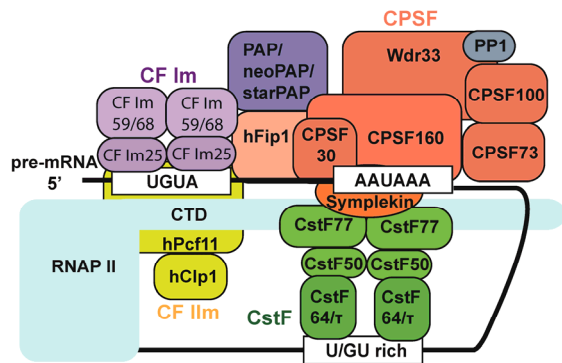

Fig. 1 A model for the mammalian mRNA 3' processing complex
The key *cis*-element in the RNA and *trans*-activing protein factors in the mammalian mRNA 3' processing complex are depicted in this schematic model

2 APA regulation: a unified mechanism

Studies of individual APA events and global analyses have revealed a number of mechanisms for APA regulation, which have been described in detail in a number of recent reviews (Shi, 2012; Elkon *et al.*, 2013; Mueller *et al.*, 2013; Tian and Manley, 2013). Instead of discussing these individual models, here we will attempt to integrate them into a unified mechanism for APA regulation and begin to put together a blueprint for the "polyadenylation code".

In the most simplistic view, the outcome of APA of a pre-mRNA is determined by the timing of cleavage/polyadenylation at competing PAS. As shown in Fig. 2, three key variables dictate the APA pattern in this tandem PAS-containing RNA substrate. $Tp1$ and $Tp2$ are the time that it takes for cleavage/polyadenylation (C/P) to occur at PAS1 and PAS2, respectively. Tt is the time it takes for the transcription machinery to traverse the region between the two PASs. Theoretically, if $Tp1 < (Tt+Tp2)$, then C/P will finish first at PAS1 to produce a shorter APA isoform. On the other hand, if $Tp1 > (Tt+Tp2)$, C/P will occur at PAS2, generating a longer APA isoform. These three

key parameters are subject to regulation by many factors, including the sequences of the alternative PASs, the distance between them, transcription elongation rate, and the concentrations of mRNA 3' processing factors and regulators. Below we will discuss the known mechanisms for APA regulation within the framework proposed in Fig. 2.
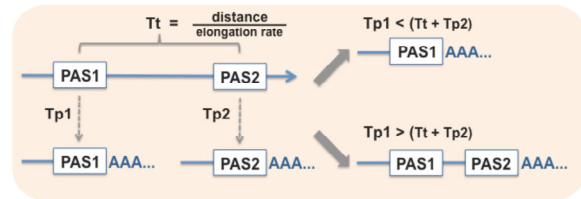

Fig. 2 A unified model for APA regulation
A pre-mRNA containing two alternative polyadenylation sites (PASs) is used as a template for APA regulation. Tt: the time it takes to transcrible the region between PAS1 and PAS2; $Tp1$: the time it takes for cleavage/polyadenylation at PAS1; $Tp2$: the time it takes for cleavage/polyadenylation at PAS2

2.1 "First come, first served" model

Based on our model, it is clear that the proximal PAS has an intrinsic advantage over the distal PAS, which has been described as "first come, first served" (Danckwardt *et al.*, 2008; Shi, 2012). This is because even if PAS1 and PAS2 have exactly the same sequence (i.e., $Tp1=Tp2$), it must be true that $Tp1 < (Tt+Tp2)$ and thus PAS1 will be preferentially used. The magnitude of the intrinsic advantage for the proximal PAS is determined by Tt. As Tt=distance/(elongation rate), our model predicts that the distance between the alternative PASs and the elongation rate of the transcription machinery in this region should play important roles in APA regulation. This conclusion is indeed supported by numerous previous studies as discussed below.

According to our model, a slower transcription elongation rate should result in higher Tt, thus giving greater advantage to the proximal PAS. This has indeed been shown to be the case by several studies. For example, a *Drosophila melanogaster* strain that harbors a mutated RNAP II with slower elongation rate displayed a significant shift in the PAS selection towards the proximal site in the *polo* gene (Pinto *et al.*, 2011). The mRNAs of the yeast *RPB2* gene show APA changes following UV-induced DNA damage

such that the mRNA isoforms using the distal PAS are preferentially produced during the recovery phase. However, addition of mycophenolic acid (MPA), a drug that reduces the transcription elongation rate by inhibiting GTP and UTP syntheses, led to a shift of polyadenylation to the proximal PAS (Yu and Volkert, 2013). UV-induced DNA damage has been suggested to inhibit global RNAP II elongation through hyperphosphorylation of its C-terminal domain (Muñoz *et al.*, 2009). However, the impact of UV-induced DNA damage on the global APA profile has not been characterized. In these two studies, the transcription elongation rates were modulated using artificial means. Under physiological conditions, the transcription elongation rate is influenced by the DNA and chromatin modifications, providing a potential mechanism for epigenetic regulation of APA (Brown *et al.*, 2012). For example, the mouse imprinted *H13* gene contains a CpG island between its alternative PASs and this CpG island is methylated in an allele-specific manner (Wood *et al.*, 2008). mRNAs produced by the unmethylated allele are mainly polyadenylated at the proximal PAS while those produced by the methylated allele preferentially use the distal PAS. It was proposed that the unmethylated CpG island functions as an internal promoter and binding of transcription factors may block or slow down the transcription from the upstream normal promoter. Such a slow-down of the transcription machinery leads to high usage of the proximal PAS. Similar regulation of APA by DNA methylation was later reported for the *Herc3* gene (Cowley *et al.*, 2012). For future studies, it will be important to measure the transcription elongation rate at the transcriptome level and to understand how it can be regulated under physiological and pathological conditions.

According to our model, a longer distance between the alternative PASs should lead to preferential usage of the proximal PAS. The first evidence for this was reported over twenty years ago. Using reporter assays, Denome and Cole (1988) demonstrated that insertion of additional sequences between two alternative PASs resulted in higher usage of the upstream PAS and a concomitant decrease in the usage of the downstream PAS. Lackford *et al.* (2014) provided evidence that this is also true for endogenous mRNAs. Fip1 is a subunit of the CPSF complex and has recently been identified as an APA regulator. Fip1 depletion in embryonic stem cells (ESCs) leads to a shift of polyadenylation to distal PAS (proximal-to-distal shift or PtoD) in the majority of affected mRNAs. A comparison of these mRNAs and those that displayed the opposite APA changes (distal-to-proximal or DtoP) revealed that the distance between the alternative PASs in PtoD genes is 4–5 times greater than that in DtoP genes (Lackford *et al.*, 2014), providing evidence that the distance between alterative PASs is an important determinant of APA regulation for endogenous mRNAs.

2.2 Survival of the fittest model

If the proximal PASs have an intrinsic advantage over the distal sites, how could distal PASs ever be selected? It turns out that the distal PASs in APA genes are often stronger than the proximal sites. Studies in fly, mouse, and human suggest that the distal PASs tend to have more canonical PAS features, including the AAUAAA hexamer and the U/GU-rich DSEs (Martin *et al.*, 2012; Smibert *et al.*, 2012; Tian and Graber, 2012; Lackford *et al.*, 2014). Reporter assays and *in vitro* assays provided evidence that the distal PASs are generally stronger than the proximal PASs (Yao *et al.*, 2012; Lackford *et al.*, 2014). Therefore, $Tp2 < Tp1$ in many APA genes. Importantly, similar to Tt, Tp is subject to regulation by a number of factors, including the concentrations of the core mRNA 3' processing factors and other regulatory factors. In this section, we will focus on the APA regulation by the concentrations of the core mRNA 3' processing factors, which was first reported by Takagaki *et al.* (1996) and Takagaki and Manley (1998). They showed that at high concentrations of CstF64, the lower-affinity proximal PAS in IgM pre-mRNA is preferentially used while a decrease in CstF64 levels leads to a shift to the stronger distal PAS. Several recent studies provided further support for this model. For example, depletion of CstF64/CstF64τ or the CPSF subunit Fip1 led to a shift of polyadenylation to the stronger distal PAS (Yao *et al.*, 2012; 2013; Lackford *et al.*, 2014). Interestingly, however, knockdown of the CFIm subunits, CFIm25 and CFIm68, resulted in the opposite APA changes, i.e., a shift to proximal PAS (Martin *et al.*, 2012). Although the underlying mechanism for CFIm-mediated APA regulation remains unclear, these observations suggest that the "survival of the fittest" model requires

further characterization. In addition, it should be pointed out that the concentrations of the mRNA 3' processing factors that are available for participating in the 3' processing reactions as discussed above may not be the same as the total concentration of these factors as they may be sequestered by other factors. For example, following UV-induced DNA damage, CstF seems to be bound and sequestered by the Brca1/Bard1 complexes (Kleiman and Manley, 2001).

The concentrations of mRNA 3' processing factors are regulated in development and in response to environmental stimuli. For example, activation of T cells or B cells induces the upregulation of CstF64, which in turn leads to APA changes (Takagaki *et al.*, 1996; Chuvpilo *et al.*, 1999). CstF64 levels are also regulated during the cell cycle (Martincic *et al.*, 1998; Sandberg *et al.*, 2008). In addition, the protein levels of Fip1 and other CPSF subunits decrease during ESC differentiation while the opposite was observed during somatic cell reprogramming to induced pluripotent stem cells (Ji and Tian, 2009; Lackford *et al.*, 2014). It remains poorly understood how the levels of mRNA 3' processing factors are regulated in development. Elkon *et al.* (2012) provided evidence that the transcription factor E2F plays a role in controlling the transcription of many mRNA 3' processing factors. In addition to transcription, the mRNA 3' processing factors can be regulated at the post-transcriptional and post-translational levels. For example, the *CstF77* gene contains an intronic PAS that is conserved from fly to human (Juge *et al.*, 2000; Pan *et al.*, 2006). The usage of this PAS gives rise to a short mRNA isoform, for which no protein product has been detected (Juge *et al.*, 2000; Luo *et al.*, 2013). Interestingly, CstF77 promotes the usage of this intronic PAS, thus turning off its own expression post-transcriptionally (Luo *et al.*, 2013). CstF64 and its paralog CstF64τ have also been shown to inhibit the expression of each other (Yao *et al.*, 2012; 2013). These mechanisms may function in maintaining the homeostasis of mRNA 3' processing activities.

2.3 Agonist/antagonist model

In addition to the concentrations of mRNA 3' processing factors as discussed above, the Tp values of alternative PASs can be individually modulated either positively or negatively by protein-RNA interactions of regulatory factors. Such regulatory factors can directly bind to the PAS sequence or to nearby *cis*-elements. This mode of APA regulation is referred to as the agonist/antagonist model (Shi, 2012).

Regulatory factors could directly bind to the sequences within or overlapping with the core PAS sequence, thereby competing with the mRNA 3' processing factors for binding to RNA. Several factors can compete with CstF for RNA binding. For example, hnRNP F and H2 bind to the G-rich element downstream of the cleavage site and these interactions may inhibit CstF64-RNA interaction and PAS recognition (Arhin *et al.*, 2002; Alkan *et al.*, 2006). In addition, the polypyrimidine tract binding (PTB) protein and Sexlethal, the master regulator of sex differentiation in flies, also compete with CstF64 for binding to the DSEs in many PASs (Castelo-Branco *et al.*, 2004; Gawande *et al.*, 2006). These studies suggest that CstF64-RNA interaction is a hot spot for regulation. Additionally CPSF-RNA interactions are also subject to regulation. Jenal *et al.* (2012) suggests that polyadenylate-binding nuclear protein 1 (PABPN1) may compete with CPSF for RNA binding. As proximal PASs are generally weaker and contain less canonical PAS features including the AAUAAA hexamer, PABPN1 binding to the A-rich sequences interferes with CPSF-AAUAAA interactions and thereby inhibits the recognition of proximal PAS.

Other regulatory factors bind to RNA sequences outside of the AAUAAA and DSE to modulate PAS selection. For example, the poly(C)-binding protein/α-complex protein (PCBP/αCP) proteins, previously shown to control the stability of *hα-globin* mRNAs, have recently been shown to regulate APA (Ji *et al.*, 2013). By directly binding to C-rich motifs located upstream of many PASs, PCBP/αCP stimulates the recognition of the nearby PAS by the core 3' processing machinery (Ji *et al.*, 2013). Similarly, the CPEB1, a protein known to stimulate cytoplasmic polyadenylation, binds to a specific *cis*-element called the cytoplasmic polyadenylation element (CPE) in the pre-mRNA and helps to recruit CPSF to suboptimal proximal PAS in many CPE-containing mRNAs (Bava *et al.*, 2013). Several mRNA 3' processing repressors have also been reported. For example, the RNA binding proteins floral binding protein 2 (FBP2) and FBP3 bind to the USE of the PAS in the prothrombin gene and inhibit cleavage (Danckwardt *et al.*, 2011). In response to stress, the p38 MAP kinase

pathway is activated and FBP2/3 becomes hyper-phosphorylated. The hyper-phosphorylation inhibits the FBP2/3-RNA interactions and activates the 3' processing of the prothrombin mRNAs. Additionally, the U1A protein binds to a *cis*-element located near the PAS of its own mRNAs and inhibits the polyadenylation step (Boelens *et al.*, 1993; Vagner *et al.*, 2000). This negative feed-back loop provides a mechanism for maintaining U1A protein levels (Boelens *et al.*, 1993). Finally, in addition to the aforementioned activators and suppressors, the RNA-binding proteins ESRP1/2 can both activate or inhibit PAS recognition in a position-dependent manner (Dittmar *et al.*, 2012).

3 APA regulation: biological functions

High throughput analyses using microarrays and deep sequencing have revealed widespread APA in the transcriptomes of yeast, nematode, fly, fish, mouse, and human (Jan *et al.*, 2010; Ozsolak *et al.*, 2010; Derti *et al.*, 2012; Ulitsky *et al.*, 2012). Additionally dynamic APA changes have been detected in development, in response to stress, and in a variety of human diseases (Shi, 2012; Elkon *et al.*, 2013; Tian and Manley, 2013). Despite our ever-increasing knowledge of APA regulation, the biological functions of APA remain poorly understood. Here we summarize the known and putative functions of APA regulation.

3.1 APA regulation in development

The Burge and Sharp groups carried out the first global analysis of APA regulation and their study revealed a correlation between cellular proliferation status and the global APA profile: proliferating cells tend to use proximal PAS and produce mRNAs with short 3' UTRs while quiescent/differentiated cells favor distal PAS and produce mRNAs with longer 3' UTRs (Sandberg *et al.*, 2008). The same trend was observed by the Tian group for the mouse development and by our group for the differentiation of ESCs into neurons (Ji *et al.*, 2009; Shepard *et al.*, 2011). The Tian group also detected 3' UTR shortening during somatic cell reprogramming (Ji and Tian, 2009). Finally, similar to somatic reprogramming, the Bartel group showed that polyadenylation of the mRNAs of a number of oncogenes shifts to proximal

PAS, leading to 3' UTR shortening (Mayr and Bartel, 2009). Although the correlation between the APA profile and the cellular proliferation/differentiation status is quite strong, the cause-effect relationship remains unclear. We have recently provided experimental evidence that APA regulation may play a direct role in determining cell fate (Lackford *et al.*, 2014). This conclusion is based on the following observations: (1) high levels of the mRNA 3' processing factor Fip1 are required for ESC self-renewal; (2) Fip1 depletion has relatively little effect on transcription, but leads to significant APA changes in many genes that result in 3' UTR lengthening; (3) 3' UTR lengthening in Fip1 APA target mRNAs suppresses their expression; (4) many Fip1 APA target mRNAs encode essential self-renewal factors; (5) Fip1 levels and Fip1-regulated APA events are developmentally regulated and are restored to an ESC-like state during somatic reprogramming. Together, these observations strongly suggest that Fip1-mediated APA regulation plays a direct role in cell fate specification. Further studies are needed to identify the key Fip1 APA target(s) that are required for ESC self-renewal and to understand how 3' UTR changes modulate the expression of these genes.

The impact of APA on the expression and function of individual genes is better understood. Based on reporter assays as well as studies on endogenous genes, the current consensus view is that the APA isoforms with shorter 3' UTRs tend to be more stable and produce more proteins (Mayr and Bartel, 2009; Sandberg *et al.*, 2008). This is because the extended UTR region found only in the longer APA isoforms may contain additional microRNA target sites and/or binding sites for protein factors that negatively regulate mRNA stability and translation. For example, Pax3 expression is important for the proliferation of muscle stem cells, but it is regulated tightly by miR206 (Boutet *et al.*, 2012). The expression of the shorter Pax3 APA isoforms that contain no miR206 target sites allows Pax3 to evade miRNA-mediated gene silencing (Boutet *et al.*, 2012). However, the advantage of shorter APA isoforms cannot be generalized. For example, only the longer APA isoform of the fly *polo* gene is efficiently translated (Pinto *et al.*, 2011). Transgenic flies, in which the distal PAS is deleted, cannot produce the longer isoform and die at the pupa stage due to a failure in the proliferation of

the precursor cells of the abdomen (Pinto *et al.*, 2011). The *brain-derived neurotrophic factor* (*BDNF*) gene provides one of the best studied examples of the functional difference between the APA isoforms. BDNF produces two APA isoforms that differ only in the 3' UTRs (An *et al.*, 2008). While the short isoform is restricted to the neuronal cell body, the long isoform is also found in the dentrites. Truncation of the extended 3' UTR leads to impaired dendritic targeting of BDNF mRNA. The compromised BDNF translation in dentrites causes defects in neurodevelopment and long-term potentiation (An *et al.*, 2008). Interestingly, insulin and leptin specifically stimulate the translation of the longer BDNF mRNA isoform and the truncation of the extended UTR prevents the hypothalamic neurons from responding to insulin and leptin signaling and in turn leads to energy imbalance and obesity (Liao *et al.*, 2012).

These studies suggest that the extended APA isoforms may have different stabilities, sub-cellular localizations, and translation efficiencies. Intriguingly however, a recent global study failed to detect significant differences in the stability and translation efficiency (measured by the association with the polysomes) of APA isoforms in mouse fibroblasts (Spies *et al.*, 2013). Further studies are needed to test this in other cell types under different conditions.

3.2 APA in human disease

Aberrant mRNA 3' processing and APA changes have been detected in several human diseases. Mayr and Bartel (2009) provided evidence for APA changes during cancer development that lead to 3' UTR shortening in the mRNAs of many genes, including proto-oncogenes. The shorter APA isoforms of proto-oncogenes seem more stable and produce higher levels of proteins, therefore more oncogenic. However, high throughput sequencing analyses of cancer and non-cancer cells derived from the same tissues failed to detect a consistent correlation between APA profile and the cancer state (Fu *et al.*, 2011; Elkon *et al.*, 2012; Lianoglou *et al.*, 2013). Therefore, further studies are necessary to determine whether and how APA regulation may contribute to cancer development.

Oculopharyngeal muscular dystrophy (OPMD) is an autosomal dominant disease caused by trinucleotide repeat expansion mutations in the *PABPN1* gene that leads to a mutant PABPN1 protein (trePABPN1) with a longer polyalanine tract (Brais, 2009). Jenal *et al.* (2012) provided evidence that trePABPN1 binds to the PABPN1 protein and sequesters it in many sub-cellular foci. The depletion of PABPN1 leads to APA changes and 3' UTR shortening in many genes through a mechanism described earlier (de Klerk *et al.*, 2012; Jenal *et al.*, 2012). These aberrant APA events have been suggested to contribute to the pathogenesis of OPMD. However, direct evidence is lacking and the key APA targets of PABPN1 in OPMD remain to be identified.

4 Conclusions and future perspectives

Despite the recent progress, the mechanisms and biological significance for APA regulation remain poorly understood. Our simple model provides a general framework that may be used to guide future studies. For example, our model predicts that the distance between alternative PASs and the transcription elongation rate play important roles in regulating APA as they determine Tt. Future studies are needed to measure the transcription elongation rate globally and gene-specifically and to determine its relationship with APA outcomes. In addition, it will be important to understand how Tp is regulated by the core mRNA 3' processing factors and regulatory factors. Ultimately, the integration of this information in a quantitative manner will be critical for deciphering the "polyadenylation code".

Compliance with ethics guidelines

Ryan DAVIS and Yongsheng SHI declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Alkan, S.A., Martincic, K., Milcarek, C., 2006. The hnRNPs F and H2 bind to similar sequences to influence gene expression. *Biochem. J.*, **393**(1):361-371. [doi:10.1042/ BJ20050538]

An, J.J., Gharami, K., Liao, G.Y., *et al.*, 2008. Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell*, **134**(1):

175-187. [doi:10.1016/j.cell.2008.05.045]

Arhin, G.K., Boots, M., Bagga, P.S., *et al.*, 2002. Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. *Nucl. Acids Res.*, **30**(8):1842-1850. [doi:10.1093/nar/30.8.1842]

Bava, F.A., Eliscovich, C., Ferreira, P.G., *et al.*, 2013. CPEB1 coordinates alternative 3'-UTR formation with translational regulation. *Nature*, **495**(7439):121-125. [doi:10.1038/nature11901]

Bentley, D.L., 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.*, **17**(3):251-256. [doi:10.1016/j.ceb.2005.04.006]

Boelens, W.C., Jansen, E.J., van Venrooij, W.J., *et al.*, 1993. The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell*, **72**(6): 881-892. [doi:10.1016/0092-8674(93)90577-D]

Boutet, S.C., Cheung, T.H., Quach, N.L., *et al.*, 2012. Alternative polyadenylation mediates microRNA regulation of muscle stem cell function. *Cell Stem Cell*, **10**(3):327-336. [doi:10.1016/j.stem.2012.01.017]

Brais, B., 2009. Oculopharyngeal muscular dystrophy: a polyalanine myopathy. *Curr. Neurol. Neurosci. Rep.*, **9**(1): 76-82. [doi:10.1007/s11910-009-0012-y]

Brown, S.J., Stoilov, P., Xing, Y., 2012. Chromatin and epigenetic regulation of pre-mRNA processing. *Hum. Mol. Genet.*, **21**(R1):R90. [doi:10.1093/hmg/dds353]

Castelo-Branco, P., Furger, A., Wollerton, M., *et al.*, 2004. Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol. Cell. Biol.*, **24**(10):4174-4183. [doi:10.1128/MCB.24.10.4174-4183.2004]

Chan, S., Choi, E.A., Shi, Y., 2011. Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip. Rev. RNA*, **2**(3):321-335. [doi:10.1002/wrna.54]

Chuvpilo, S., Zimmer, M., Kerstan, A., *et al.*, 1999. Alternative polyadenylation events contribute to the induction of NF-ATc in effector T cells. *Immunity*, **10**(2):261-269. [doi:10.1016/S1074-7613(00)80026-6]

Colgan, D.F., Manley, J.L., 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.*, **11**(21):2755-2766. [doi:10.1101/gad.11.21.2755]

Cowley, M., Wood, A.J., Bohm, S., *et al.*, 2012. Epigenetic control of alternative mRNA processing at the imprinted Herc3/Nap1l5 locus. *Nucl. Acids Res.*, **40**(18):8917-8926. [doi:10.1093/nar/gks654]

Danckwardt, S., Hentze, M.W., Kulozik, A.E., 2008. 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.*, **27**(3):482-498. [doi:10.1038/sj.emboj.7601932]

Danckwardt, S., Gantzert, A.S., Macher-Goeppinger, S., *et al.*, 2011. p38 MAPK controls prothrombin expression by regulated RNA 3' end processing. *Mol. Cell*, **41**(3): 298-310. [doi:10.1016/j.molcel.2010.12.032]

de Klerk, E., Venema, A., Anvar, S.Y., *et al.*, 2012. Poly(A) binding protein nuclear 1 levels affect alternative poly-adenylation. *Nucl. Acids Res.*, **40**(18):9089. [doi:10.1093/nar/gks655]

Denome, R.M., Cole, C.N., 1988. Patterns of polyadenylation site selection in gene constructs containing multiple polyadenylation signals. *Mol. Cell Biol.*, **8**:4829-4839.

Derti, A., Garrett-Engele, P., Macisaac, K.D., *et al.*, 2012. A quantitative atlas of polyadenylation in five mammals. *Genome Res.*, **22**(6):1173-1183. [doi:10.1101/gr.132563.111]

Dittmar, K.A., Jiang, P., Park, J.W., *et al.*, 2012. Genome-wide determination of a broad ESRP-regulated posttranscriptional network by high-throughput sequencing. *Mol. Cell. Biol.*, **32**(8):1468-1482. [doi:10.1128/MCB.06536-11]

Elkon, R., Drost, J., van Haaften, G., *et al.*, 2012. E2F mediates enhanced alternative polyadenylation in proliferation. *Genome Biol.*, **13**(7):R59. [doi:10.1186/gb-2012-13-7-r59]

Elkon, R., Ugalde, A.P., Agami, R., 2013. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.*, **14**(7):496-506. [doi:10.1038/nrg3482]

Flavell, S.W., Kim, T.K., Gray, J.M., *et al.*, 2008. Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron*, **60**(6):1022-1038. [doi:10.1016/j.neuron.2008.11.029]

Fu, Y., Sun, Y., Li, Y., *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**(5):741-747. [doi:10.1101/gr.115295.110]

Gawande, B., Robida, M.D., Rahn, A., *et al.*, 2006. *Drosophila* Sex-lethal protein mediates polyadenylation switching in the female germline. *EMBO J.*, **25**(6):1263-1272. [doi:10.1038/sj.emboj.7601022]

Hirose, Y., Manley, J.L., 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev.*, **14**:1415-1429.

Jan, C.H., Friedman, R.C., Ruby, J.G., *et al.*, 2010. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature*, **469**(7328):97-101. [doi:10.1038/nature09616]

Jenal, M., Elkon, R., Loayza-Puch, F., *et al.*, 2012. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell*, **149**(3):538-553. [doi:10.1016/j.cell.2012.03.022]

Ji, X., Wan, J., Vishnu, M., *et al.*, 2013. αCP poly(C) binding proteins act as global regulators of alternative polyadenylation. *Mol. Cell. Biol.*, **33**(13):2560-2573. [doi:10.1128/MCB.01380-12]

Ji, Z., 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**(12):e8419. [doi:10.1371/journal.pone.0008419]

Ji, Z., Lee, J.Y., Pan, Z., *et al.*, 2009. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *PNAS*, **106**(17):7028-7033. [doi:10.1073/pnas.0900028106]

Juge, F., Audibert, A., Benoit, B., *et al*., 2000. Tissue-specific autoregulation of *Drosophila suppressor of forked* by alternative poly(A) site utilization leads to accumulation of the suppressor of forked protein in mitotically active cells. *RNA*, **6**(11):1529-1538. [doi:10.1017/S1355838200001266]

Kleiman, F.E., Manley, J.L., 2001. The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. *Cell*, **104**(5):743-753. [doi:10.1016/S0092-8674(01)00270-7]

Lackford, B., Yao, C., Charles, G.M., *et al*., 2014. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.*, **33**(8):878-889. [doi:10.1002/embj.201386537]

Lianoglou, S., Garg, V., Yang, J.L., *et al*., 2013. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.*, **27**(21):2380-2396. [doi:10.1101/gad.229328.113]

Liao, G.Y., An, J.J., Gharami, K., *et al*., 2012. Dendritically targeted *Bdnf* mRNA is essential for energy balance and response to leptin. *Nat. Med.*, **18**(4):564-571. [doi:10.1038/nm.2687]

Luo, W., Ji, Z., Pan, Z., *et al*., 2013. The conserved intronic cleavage and polyadenylation site of *CstF-77* gene imparts control of 3' end processing activity through feedback autoregulation and by U1 snRNP. *PLoS Genet.*, **9**(7):e1003613. [doi:10.1371/journal.pgen.1003613]

Martin, G., Gruber, A.R., Keller, W., *et al*., 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**(6):753-763. [doi:10.1016/j.celrep.2012.05.003]

Martincic, K., Campbell, R., Edwalds-Gilbert, G., *et al*., 1998. Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the G_0 to S phase transition. *PNAS*, **95**(19):11095-11100. [doi:10.1073/pnas.95.19.11095]

Mayr, C., Bartel, D.P., 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, **138**(4):673-684. [doi:10.1016/j.cell.2009.06.016]

Mueller, A.A., Cheung, T.H., Rando, T.A., 2013. All's well that ends well: alternative polyadenylation and its implications for stem cell biology. *Curr. Opin. Cell Biol.*, **25**(2):222-232. [doi:10.1016/j.ceb.2012.12.008]

Muñoz, M.J., Pérez Santangelo, M.S., Paronetto, M.P., *et al*., 2009. DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell*, **137**(4):708-720. [doi:10.1016/j.cell.2009.03.010]

Ozsolak, F., Kapranov, P., Foissac, S., *et al*., 2010. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell*, **143**(6):1018-1029. [doi:10.1016/j.cell.2010.11.020]

Pan, Z., Zhang, H., Hague, L.K., *et al*., 2006. An intronic polyadenylation site in human and mouse *CstF-77* genes suggests an evolutionarily conserved regulatory mechanism. *Gene*, **366**(2):325-334. [doi:10.1016/j.gene.2005.09.024]

Park, J.Y., Li, W., Zheng, D., *et al*., 2011. Comparative analysis of mRNA isoform expression in cardiac hypertrophy and development reveals multiple post-transcriptional regulatory modules. *PLoS ONE*, **6**(7):e22391. [doi:10.1371/journal.pone.0022391]

Pinto, P.A., Henriques, T., Freitas, M.O., *et al*., 2011. RNA polymerase II kinetics in *polo* polyadenylation signal selection. *EMBO J.*, **30**(12):2431-2444. [doi:10.1038/emboj.2011.156]

Proudfoot, N.J., Furger, A., Dye, M.J., 2002. Integrating mRNA processing with transcription. *Cell*, **108**(4):501-512. [doi:10.1016/S0092-8674(02)00617-7]

Sandberg, R., Neilson, J.R., Sarma, A., *et al*., 2008. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*, **320**(5883):1643-1647. [doi:10.1126/science.1155390]

Shepard, P.J., Choi, E.A., Lu, J., *et al*., 2011. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA*, **17**(4):761-772. [doi:10.1261/rna.2581711]

Shi, Y., 2012. Alternative polyadenylation: new insights from global analyses. *RNA*, **18**(12):2105-2117. [doi:10.1261/rna.035899.112]

Shi, Y., Di Giammartino, D.C., Taylor, D., *et al*., 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell*, **33**(3):365-376. [doi:10.1016/j.molcel.2008.12.028]

Smibert, P., Miura, P., Westholm, J.O., *et al*., 2012. Global patterns of tissue-specific alternative polyadenylation in *Drosophila*. *Cell Rep.*, **1**(3):277-289. [doi:10.1016/j.celrep.2012.01.001]

Spies, N., Burge, C.B., Bartel, D.P., 2013. 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res.*, **23**(12):2078-2090. [doi:10.1101/gr.156919.113]

Takagaki, Y., 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**(6):761-771. [doi:10.1016/S1097-2765(00)80291-9]

Takagaki, Y., Seipelt, R.L., Peterson, M.L., *et al*., 1996. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell*, **87**(5):941-952. [doi:10.1016/S0092-8674(00)82000-0]

Tian, B., Graber, J.H., 2012. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdisc. Rev. RNA*, **3**(3):385-396. [doi:10.1002/wrna.116]

Tian, B., Manley, J.L., 2013. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem. Sci.*, **38**(6):312. [doi:10.1016/j.tibs.2013.03.005]

Ulitsky, I., Shkumatava, A., Jan, C.H., *et al*., 2012. Extensive alternative polyadenylation during zebrafish development.

Genome Res., **22**(10):2054-2066. [doi:10.1101/gr.139733.112]

Vagner, S., Ruegsegger, U., Gunderson, S.I., *et al*., 2000. Position-dependent inhibition of the cleavage step of pre-mRNA 3'-end processing by U1 snRNP. *RNA*, **6**(2):178-188.

Wood, A.J., Schulz, R., Woodfine, K., *et al*., 2008. Regulation of alternative polyadenylation by genomic imprinting. *Genes Dev.*, **22**(9):1141-1146. [doi:10.1101/gad.473408]

Yao, C., Biesinger, J., Wan, J., *et al*., 2012. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *PNAS*, **109**(46):18773-18778. [doi:10.1073/pnas.1211101109]

Yao, C., Choi, E.A., Weng, L., *et al*., 2013. Overlapping and distinct functions of CstF64 and CstF64tau in mammalian mRNA 3' processing. *RNA*, **19**(12):1781-1790. [doi:10.1261/rna.042317.113]

Yu, L., Volkert, M.R., 2013. UV damage regulates alternative polyadenylation of the *RPB2* gene in yeast. *Nucl. Acids Res.*, **41**(5):3104-3114. [doi:10.1093/nar/gkt020]

Zhao, J., Hyman, L., Moore, C., 1999. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.*, **63**(2):405-445.

中文概要：

本文题目：多腺苷酸化之密码：调解 mRNA 可变聚腺苷酸化的统一模型

The polyadenylation code: a unified model for the regulation of mRNA alternative polyadenylation

本文概要： 真核生物的大部分基因都通过可变聚腺苷酸化（APA）而产生多种不同的 mRNA 3'端。近期的研究表明，可变聚腺苷酸化在组织发展中被动态调节，并且会受环境刺激而自动调节。现有文献中表述了多种调节机制。本文整合所有现有的调节机制模型，进而提出一个综合的统一模型。这个模型不仅概括了已知的研究结果，而且为未来的研究提供了一个预测可变聚腺苷酸化的方法。最后，我们讨论已知和假设的可变聚腺苷酸化带来的功能。

关键词组： 基因表达；可变聚腺苷酸化；预测模型；mRNA