Elucidating the Roles of PARP1 and RBBP6 in the Regulation of pre-mRNA 3'-end Processing

Dafne Campigli Di Giammartino

Submitted in the partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

ABSTRACT

Elucidating the Roles of PARP1 and RBBP6 in the Regulation of pre-mRNA 3'-end Processing

Dafne Campigli Di Giammartino

The mature 3' ends of most mRNAs are created by a two-step reaction that involves an endonucleolytic cleavage of the pre-mRNA followed by polyadenylation of the upstream product. The 3' processing machinery is composed of four multisubunit complexes, which, together with a few other proteins, constitute the core components required for cleavage and polyadenylation. A proteomic analysis led to the identification of approximately 80 proteins that associate with the human pre-mRNA 3' processing complex, including new core 3' factors and other proteins that might mediate crosstalk between 3' processing and other nuclear pathways. This thesis focuses on two of the newly identified proteins, which we found particularly intriguing: PARP1 and RBBP6.

PARP1 is an enzyme that, when activated, catalyzes the polymerization of ADP-ribose units from donor NAD molecules to acceptor proteins, a reaction known as PARylation. This post-translational modification has been shown to modulate critical events such as DNA damage response and transcription. We found that PARP1 binds PAP, the enzyme responsible for polyadenylating the 3' ends of mRNAs, and modifies it by PARylation. In vivo PAP is PARylated during heat shock, leading to inhibition of polyadenylation in a PARP1-dependent manner. Finally, we show that the observed inhibition reflects decreased PAP association with 3' end of

genes. These results identify PARP1 as a regulator of polyadenylation during thermal stress and show for the first time that PARylation can control gene expression by modulating processing of mRNA.

The second project involves RBBP6, a large multidomain protein that is known to interact with p53 and Rb. The N-terminal part of the human RBBP6 includes a DWNN domain, which is particularly interesting because it adopts a ubiquitin-like fold and, in addition to forming part of the full-length RBBP6 protein, is also expressed as a small protein (RBBP6 isoform3) which has been shown to be downregulated in several human cancers. We found that RBBP6 is essential for the cleavage activity of the 3' processing complex and that an N-terminal derivative of RBBP6 (RBBP6-N), containing only the DWNN, Zinc and Ring domains, is enough to rescue cleavage activity. The RBBP6 and RBBP6 isoform3 can compete with each other in binding to Cstf64 (an interaction mediated by the DWNN domain). In addition, overexpression of isoform3 inhibits cleavage raising intriguing possibilities of modulation of 3' processing by fine-tuning the levels of the two RBBP6 isoforms. To better characterize the function of RBBP6 globally, we also performed genome-wide analysis, both by microarray and deep sequencing. Following RBBP6 knockdown we observed a general lengthening of 3' UTRs accompanied by an overall downregulation in gene expression, especially of RNAs with AU-rich 3'UTRs. We show that this is the result of a defect in their 3' cleavage and subsequent degradation by the exosome. All together our results point to a role for RBBP6 as a new core 3' processing factor able to regulate the expression of AU-rich mRNAs.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

First of all I thank Professor James Manley for giving me the opportunity to do research in his laboratory and for supporting me during several challenging periods in my PhD with his optimistic incitements to persist and keep trying. I appreciate the freedom he gives his students to pursue their ideas. The most important thing I learned from him is how to think independently.

I must thank Yongsheng Shi, a former postdoc in our laboratory for introducing me to the RNA world and teaching me, when I was a research assistant, many of the techniques I used during my PhD years. Not to mention the fact that both of my research projects stem out of his work in the Manley's lab. I also thank my laboratory roommates Patricia Richard and Emanuel Rosonina for being such a pleasant company and being always available to share their experience and help in any scientific issue. I am thankful to all lab members, in particular JingPing Hsin for being always eager to discuss with interest any matter I brought to his attention.

I want to thank my parents who always supported me in every choice I made in my life and never stopped believing in me. In particular I am grateful to my mother and mother in law who helped me, during the final months of the PhD, with my newborn Lia.

Finally, and most importantly, I thank my husband, Raffaele. Without his love and constant support I would have never been able to accomplish my PhD.

PREFACE

This thesis is divided into four parts. The first part includes a chapter written for the *Encyclopedia of Biological Chemistry* (2013, second edition, Elsevier) entitled "mRNA Polyadenylation in Eukaryotes", and a section dedicated to new factors associated with the 3' processing complex (which will be published as part of a review on 3' processing). The second chapter is a review entitled "Mechanisms and Consequences of Alternative Polyadenylation" published in *Molecular Cell* (2011, 43:853-66). The third chapter of this thesis is a paper published in *Molecular Cell* entitled "PARP1 Represses PAP and Inhibits Polyadenylation During Heat Shock" (2013, 49:7-17). The fourth chapter is a manuscript in preparation named "RBBP6 is a Core Component of the Human 3' Processing Complex and a Regulator of Expression of mRNAs with AU-rich 3'UTRs". The appendix is a paper published in *Molecular Cell* entitled "Molecular Architecture of the Human pre-mRNA 3' Processing Complex" (2009, 33:365-76) to which I contributed before the beginning of the PhD and which represents the starting point for the works presented in chapters three and four of this dissertation.

CHAPTER1

Introduction

The first part of this introduction was written in 2010 and published in 2013 as a chapter in the *Encyclopedia of Biological Chemistry* (pages 188-193, second edition, Elsevier). The second part is dedicated to new factors associated with the 3' processing complex.

mRNA Polyadenylation in Eukaryotes

J L Manley and D C Di Giammartino, Columbia University, New York, NY, USA

Glossary

Capping The process by which a 7-methyl guanosine is added to the 5′ end of a precursor-messenger RNA (pre-mRNA) by a 5′ to 5′ triphosphate linkage. 5′ Capping is important for regulation of mRNA nuclear export, prevention of degradation by exonucleases, promotion of translation, and 5′ proximal intron excision.

Messenger ribonucleic acid (mRNA) A molecule transcribed from DNA by an RNA polymerase. Mature mRNAs contain 5′ and 3′ untranslated regions (UTRs), a modified base at the 5′ end called 'cap', a polyA tail at the 3′ end, and a protein-coding region in between. mRNAs are exported to the cytoplasm where they will be translated into a polymer of amino acids which constitutes a protein.

Polyadenylation The process by which poly(A) polymerase (PAP) adds an unencoded poly(A) tail at the 3′ end of a maturing mRNA. The site for poly(A) addition is created by a large protein complex that cleaves the pre-mRNA at the polyadenylation site in the 3′ UTR. More than one polyadenylation site might be present allowing a single gene to encode multiple transcripts with distinct 3′ ends, an event called 'alternative polyadenylation'.

Splicing It is a two-step reaction that leads to the removal of an intron and the joining of two adjacent exons. The spliceosome is the large complex that directs this step of pre-mRNA maturation and is made of approximately 200 proteins and five small RNAs.

Transcription The process by which an RNA polymerase synthesizes an RNA molecule on a DNA template. In eukaryotes, there are three RNA polymerases, but only RNA polymerase II is responsible for transcription of protein-coding genes into mRNA.

3′ UTR The untranslated region at the 3′ end of an mRNA, after the stop codon. It includes regulatory regions such as those that direct cleavage and polyadenylation as well as sequences that regulate mRNA translation and stability by serving as binding sites for microRNAs or regulatory proteins.

Introduction

Eukaryotic precursor-messenger RNAs (pre-mRNAs) must undergo several processing events before the mature mRNA can be transported out of the nucleus and translated into proteins. Transcription, capping at the 5′ end, splicing of introns, and polyadenylation of the 3′ end are all complex reactions that require numerous proteins (and in the case of splicing, RNAs). Interestingly, these reactions have all been found to be interrelated, such that proteins involved in one step often also participate in the other/s. It is the accurate sequence of events of this highly regulated process that leads to the formation of an mRNA.

All eukaryotic pre-mRNAs, with the exception of replication-dependent histone transcripts, acquire a polyadenylate tail at their 3′ end. The maturation of the 3′ end of the pre-mRNA is achieved by a two-step reaction that involves an endonucleolytic cleavage of a phosphodiester bond in the pre-mRNA transcript, followed by addition of a polyadenylate tail on the upstream cleavage product, which corresponds to the mature mRNA. In order for these vents to occur, four multi-subunit protein complexes, with additional accessory factors, have to bind specific RNA sequences that are found in the primary transcript, and this dictates the precise site where cleavage will occur. The importance of this step is emphasized by the absolute requirement of genes encoding 3′ processing factors for cell viability in yeast and higher eukaryotes, and by the association of some human diseases with aberrant polyadenylation events. Such diseases can be caused either by mutations in the *cis*-acting RNA sequences or by mutations/deregulation of the *trans*-acting protein factors involved.

Polyadenylation is a fundamental step of gene expression and its success will determine many aspects of mRNA metabolism: transcription termination by RNAP II (which occurs past the polyadenylation site and has been shown to be dependent on the 3′ processing), mRNA stability, mRNA export to the cytoplasm, and the efficiency of translation of the mRNAs into proteins are all dependent on 3′ processing. Therefore, the effectiveness by which the 3′ protein complex assembles onto defined sequence elements in the 3′ untranslated region (UTR) will determine the final level of protein expression.

Recent studies indicate that more than 50% of human genes encode transcripts that contain multiple polyadenylation sites. These alternative sites either can be found in internal introns and therefore alternative polyadenylation will be coupled to alternative splicing events, producing different protein isoforms, or can all be in the 3′ UTR, resulting in transcripts encoding the same protein but with 3′ UTRs of different length. The length of the 3′ UTR can, in turn, affect the stability, localization, transport, and translational properties of the mRNA. Interestingly, differential processing at multiple poly(A) sites can be influenced by physiological conditions such as differentiation and development, or by pathological events such as cancer and viral infection.

RNA Sequences That Direct Cleavage and Polyadenylation

In mammalian cells, three *cis*-elements define the cleavage site. The highly conserved hexanucleotide AAUAAA is found 10–30 nucleotides (nt) upstream of the cleavage site. Extensive

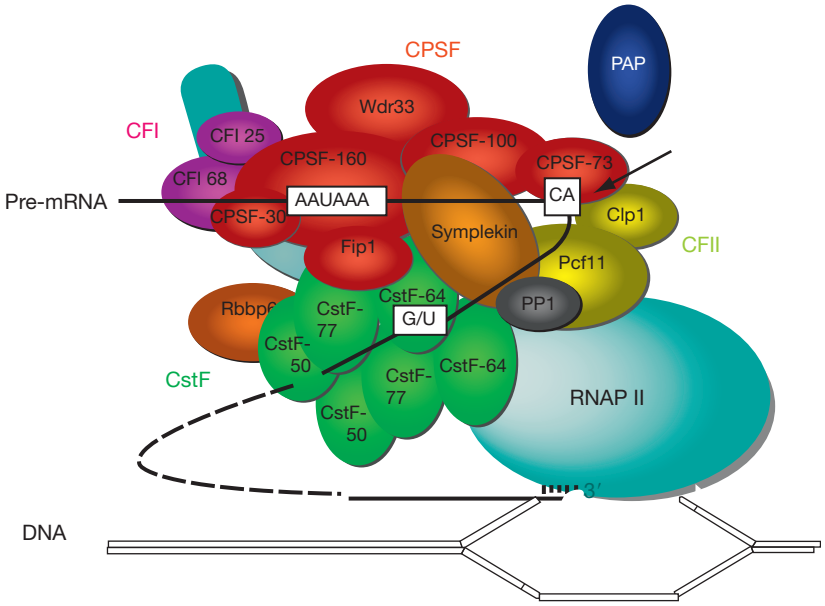


Figure 1 A schematic model of the mammalian pre-mRNA 3′-end processing machinery.

mutagenesis analysis and naturally occurring mutations have proved that this sequence is essential for both cleavage and specific polyadenylation. Variants of the AAUAAA motif are usually associated with alternatively used poly(A) sites.

The second motif of the core polyadenylation signal is the downstream element (DSE), which is located within 30-nt downstream of the cleavage site and consists of a U-rich or/and a GU-rich sequence in the form of YGUGUUYY (Y = pyrimidine). This sequence is poorly conserved and mutations or small deletions have only a weak effect on the efficiency of 3′ processing.

The cleavage site will be determined by the distance between the AAUAAA and the DSE. The sequence where cleavage will occur is not conserved, but, in 70% of the cases, is located immediately after an adenosine residue that is very often preceded by a cytosine residue. In addition to these three core polyadenylation sequences, an upstream (located upstream of the AAUAAA) and/or downstream (located after the DSE) auxiliary element are sometimes present: the first is a U-rich sequence while the second is a G-rich sequence, but these are very poorly conserved.

In yeast, the RNA signals that direct polyadenylation are different from those used in higher eukaryotes in both sequence and organization and appear to be much less conserved. The yeast poly(A) site is defined by four elements: the AU-rich efficiency element, the A-rich positioning element, the actual cleavage site, and the U-rich elements positioned upstream and downstream of the cleavage site.

Protein Complexes Involved in 3′ Processing

The development of appropriate techniques to reproduce the cleavage and polyadenylation reactions *in vitro*, by using cell extracts and *in vitro*-transcribed pre-mRNAs, allowed the biochemical characterization of the proteins involved in 3′-end maturation of pre-mRNAs.

Two processes as simple as cleavage and polyadenylation, which could be catalyzed by the action of only two enzymes (an endonuclease and a poly(A) polymerase (PAP)), are in fact supported by a vast number of protein factors (**Figure 1**) emphasizing the necessity to finely regulate the process and to coordinate it with other nuclear events.

The core molecular machinery responsible for 3′ end formation in mammals includes four multi-subunit protein complexes: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I (CFI), and cleavage factor II (CFII). In addition, PAP, symplekin, poly(A)-binding protein II (PABPII), and the C-terminal domain (CTD) of the RNA polymerase II largest subunit are also part of the core machinery. Other proteins have been discovered in a recent proteomic study to associate with the core factors and are very likely to be part of the multi-subunit complexes mentioned above; although they are annotated in **Table 1**, they will not be discussed here, due to the lack of enough published data on their function.

Mammalian CPSF contains five subunits (CPSF-160, CPSF-100, CPSF-73, CPSF-30, and Fip1), which are all essential for efficient cleavage and polyadenylation. The CPSF complex binds to the AAUAAA sequence, mainly through CPSF-160, but the interaction with the RNA is greatly enhanced by cooperative binding between CPSF and CstF. CPSF-160 is known to bind CstF-77 and PAP, and it also associates with proteins involved in transcription initiation (TFIID) and elongation (PolII CTD). CPSF-73 has recently been reported to be the actual endoribonuclease, responsible for the cleavage at the poly(A) site, but it is the entire 3′ processing complex that defines the exact cleavage site and confers sequence specificity. CPSF-73 contains a metallo-β-lactamase domain and is the founding member of the β-CASP protein superfamily.

Table 1 Characteristics of the protein factors of the mammalian pre-mRNA 3′-end processing complex

Protein factor (processing step)	*Subunits*	*Yeast homologue (sub-complex)*	*Sequence characteristics*	*Protein function*	*Interacting proteins*
CPSF (cleavage and polyadenylation)	CPSF-160	Cft 1p/Yhh1p (CPF)	Three possible β-propellers	Binds the AAUAAA sequence	CstF-77, Pol II CTD, PAP, Fip1, TFIID
	CPSF-100	Cft2p/Ydh1p (CPF)	Non-metal binding β-lactamase domain		CPSF-73 CstF-64 symplekin
	CPSF-73	Brr5p/Ysh1p (CPF)	Metallo β-lactamase domain	Endonuclease	CPSF-100, CstF64, symplekin
	CPSF-30	Yth1p (CPF)	Five zinc fingers and one zinc knuckle	Binds U-rich RNA sequences	Fip1
	Fip 1	Fip1p (CPF)	Pro-rich sequence. RD-rich sequence. Arg-rich sequence		PAP, CPSF-160, CPSF-70, CstF-77
	WDR33	Pfs2 (PFI)	WD repeats		
CstF (cleavage)	CstF-77	Rna14p (CFIA)	HAT domain, Proline rich sequence	Scaffolding protein, links CstF and CPSF	CPSF-160, CstF-64, CstF-50 Fip1
	CstF-64	Rna15p (CFIA)	RRM, pro/gly-rich sequence. MEARA/G pentapeptide motif	Binds to G/U rich sequences	CstF-77, Symplekin
	CstF-50		Seven WD40 repeats	Regulatory role during DNA damage	CstF-77, Pol II CTD, BARD1
CFIm (cleavage)	CFI-25		NUDIX domain, PAP interaction domain	Helps binding AAUAAA	PAP, CFI-68, PABPII
	CFI-68		RBD, SR protein homology in C-terminus	Helps binding AAUAAA	CFI-25
CFIIm (cleavage)	hClp1	Clp1p (CFIA)	Walker A and B motifs for ATP binding	Tethers CPSF with CF Im	CFIm, CPSF-100, CPSF-73, CstF-64, symplekin
	Pcf11	Pcf11p (CFIA)	polII CTD interacting motif, two zinc fingers		Pol II CTD
Symplekin (cleavage)		Pta1p (CPF)	HEAT fold	Mediates interaction between CPSF and CstF	CstF, CPSF. Ssu72, pol II CTD
PAP (cleavage and polyadenylation)		Pap1p	Catalytic core at N-terminus, C-terminus contains RBS, bipartite NLS, ST-rich region	Catalyzes the addition of the poly(A) tail to cleaved mRNA, non specific activity by itself	CPSF-160, Fip1, CFIm
PABPII (polyadenylation)		Pab1p	Two RRM domains	Responsible for processive elongation and control of poly(A) tail length, stabilizes the tail by binding	CPSF-30
Pol II CTD (cleavage)		Pol II CTD	YSPTSPS repeats (52 in humans)	Essential for co-transcriptional recruitment of CPSF and CstF and for cleavage	CPSF-160, CstF-77, CstF-50, Pcf11
PP1 (polyadenylation)		Glc7p		Type 1 protein phosphatase	
RBBP6		Mpe1p (CPF)	RS domain, RING finger, zinc knuckle, DWNN, pro-rich		p53, Rb

CstF consists of three subunits (CstF-77, CstF-64, and CstF-50), which are necessary for cleavage but not for polyadenylation. CstF-64 contains a typical RNP-type RNA-binding domain (RBD) that binds to the G/U-rich element in the mRNA precursor. In addition, it contains 12 repeats of a pentapeptide motif, MEAR(A/G), the function of which is unknown. In humans, a second isoform, known as tCstF-64, exists. CstF-77 contains a half A TPR (HAT) motif, which mediates its homodimerization. Both CstF-77 and CstF-50 bind specifically to the CTD of Pol II but the latter does so with higher efficiency. The WD-40 repeats in CstF-50 are involved in its interaction with CstF-77. CstF-50 was found to interact by yeast two-hybrid screen with breast cancer 1 (BRCA1) associated ring domain (BARD1), a protein associated with the tumor-suppressor BRCA1. Interestingly, the CstF/BARD1/BRCA1 complex is stabilized under conditions of DNA damage, leading to inhibition of 3′ processing (see below for more details). CFI and CFII are required only for the cleavage reaction. The former is a heterodimer made up of CF-25 and one of three related subunits of 59-, 68-, or 72-kDa subunits; its primary function is to provide additional recognition of the pre-mRNA and aid the definition of the proper polyadenylation site. In addition, CFI-68 has recently been reported to be involved in mRNA export to the cytoplasm. CFII also consists of two subunits, Pcf11 and Clp1. Interestingly, Pcf11 has a PolII CTD-interacting domain (CID), and mutations in this domain cause incorrect transcription termination.

PAP is the polymerase responsible for the addition of the poly(A) tail to the cleaved mRNA. Several isoforms of this protein have been reported to arise from the same gene, with PAPII (referred here as PAP) being the predominant nuclear species in most cells. Two additional nuclear poly(A) polymerases transcribed from different genes are of particular interest: neo-PAP, which is overexpressed in human tumors and star-PAP, which is required for the polyadenylation of a group of genes encoding proteins involved in detoxification and/or oxidative stress response. PAP interacts with CPSF160, Fip1, and CF-25. Its N-terminal domain coordinates two metal ions (Mg or Mn) that are required for catalysis, while the CTD is subject to extensive posttranslational modifications that modulate PAP's activity (see below). CPSF and PAP are sufficient for poly(A) addition to a pre-cleaved RNA substrate but rapid elongation and control of poly(A) length (up to 200–300 bases in mammals) requires PABPII. In mammalian cells, PABPII binds nascent tracts of adenylate residues and, along with CPSF, stimulates PAP to switch from distributive synthesis to processive synthesis.

Symplekin is thought to function as a scaffold in the 3′ processing complex. Recently, it has also been found to associate with a CTD phosphatase, Ssu72, and to stimulate its catalytic activity. Significantly, this interaction appears to be required for transcription-coupled polyadenylation, but not for polyadenylation of a presynthesized substrate.

The CTD is also required for efficient cleavage *in vitro*. In humans, the heptapeptide YSPTSPS is repeated 52 times and is subject to extensive phosphorylation events that are specific to transcriptional stages. One of these, phosphorylation of the ser 2 position, can enhance 3′ processing *in vivo*. As mentioned earlier, several 3′ processing factors can associate with Pol II CTD at the promoter and remain associated during transcription.

The yeast 3′ processing complex shares some similarities with the mammalian complex, reflected in a high number of homologous proteins, but the overall complex composition is different. The yeast sub-complexes include the cleavage factor IA (CFIA), cleavage factor IB (CFIB), and cleavage and polyadenylation factor (CPF). The latter, in turn, is composed of CFII and polyadenylation factor I (PFI). CFII and PFI contain subunits that are homologous to the mammalian CPSF, while the CFIA complex shares homologous subunits with mammalian CFII and CstF.

Regulation of mRNA 3′ End Formation

One way to regulate cleavage and polyadenylation of pre-mRNA is through posttranslational modification of the 3′ processing factors. Importantly, all types of modifications that will be mentioned are covalent but reversible, such that they provide an efficient but temporary way to control gene expression. Some of these modifications are cell-cycle dependent, such as PAP's phosphorylation, while others are triggered by environmental conditions such as stress and DNA damage. Phosphorylated serine, threonine, and, to a lesser extent, tyrosine residues have been detected within the core 3′ processing proteins, either by phosphoproteomic screens or by specific biochemical studies. The most explicative example is the cell-cycle-dependent phosphorylation of PAP. During mitosis, PAP is hyperphosphorylated by Cdc2/Cyclin B, which reduces its enzymatic activity, contributing to a general repression of mRNA and protein production during mitosis. Another important posttranslational modification is sumoylation of lysine residues. Proteins modified by small ubiquitin-like modifier (SUMO) are CPSF-73, symplekin, and PAP. This modification was shown to enhance formation of the 3′ processing complex as well as the nuclear localization of PAP. Other modifications include lysine acetylation in PAP and CFI25 and methylation of arginine residues in CFI68, although the functional significance of these is unclear.

Posttranslational modifications can regulate the formation of the 3′ end of pre-mRNAs by different mechanisms: modification of a protein can interfere with its binding to the rest of the complex, and the absence of that particular protein may, in turn, lead to destabilization of the whole complex, as is the case for sumoylation. Conversely, a posttranslational modification may also lead to recruitment of a binding partner that would not otherwise be associated with the complex. An additional possibility is that the modification leads to a conformational change or a change in the charge of the protein that inhibits the binding of the protein to the mRNA. Some types of modifications, such as sumoylation, are well known to regulate the localization of the modified proteins and affect the nucleo-cytoplasmic shuttling, while others can lead to degradation of the target, for example, in the presence of DNA damage, for example, after UV treatment, the CTD will be ubiquitinated by a pathway involving BRCA1/BARD1, which are recruited to the site of RNA processing by CstF-50. Ubiquitination of CTD causes its proteosomal

degradation, leading to inhibition of transcription. These events would also inhibit the cleavage of pre-mRNA, since the 3′ processing factors cannot be recruited properly without the CTD.

In the past 10 years, it has become increasingly evident that usage of alternative poly(A) sites is a common mechanism to regulate gene expression. Fifty percent of human genes encode multiple transcripts derived from alternative polyadenylation. In many mRNAs, the choice of the site changes in response to physiological conditions such as cell growth, development, and differentiation, while in other cases it is triggered by external conditions such as cancer and viral infection. Alternative polyadenylation may or may not be affected by splicing, depending on if the alternative site is present in an intron or in the 3′ UTR. A classical example of an alternative polyadenylation is the IgM heavy chain expression, in which increased accumulation of CstF in plasma cells is sufficient to switch the heavy chain from the membrane-bound form to the secreted form during differentiation of B lymphocytes.

Several recent reports have examined alternative polyadenylation using microarray-based techniques and found a general correlation between 3′ UTR shortening and cell proliferation and transformation, while there is a preferred usage of distal poly(A) sites during differentiation. 3′ UTR shortening can have striking functional consequences, for example, a shorter transcript produces more protein, an effect that can be explained by the absence of inhibitory sequences, such as miRNA-binding sites or other destabilizing elements that would otherwise be present in the longer transcript. The length of the 3′ UTR might affect not only the stability but also the localization, transport, and translational properties of the mRNA. The mechanisms that regulate such global events are mostly unknown and intense research is currently being carried out in order to better understand this phenomenon at the molecular level.

Coupling of 3′ End Processing with Other Steps of Gene Expression

The maturation of the 3′ end of pre-mRNA is closely connected to most steps of gene expression. Starting from transcription initiation, CPSF was first found to associate with TFIID and later, both CPSF and CstF were shown to be located at the transcription start site by chromatin immunoprecipitation experiments. These factors are believed to travel along the gene, during transcription elongation, through their interaction with the CTD. Once the polyadenylation signals are transcribed, the 3′ processing factors associate with the signal sequences in the newly synthesized pre-mRNA. The connection with transcription does not end here. Indeed, transcription termination is dependent on the molecular machinery responsible for 3′ end formation, and an intact polyadenylation signal has long been known to be necessary for transcription termination of protein-coding genes in human and yeast cells. In the last 10 years, it became increasingly evident that not only transcription and mRNA maturation, but virtually all steps of gene expression are coupled processes that can regulate one another.

A link between splicing and 3′ processing has been established by several studies. In one example, the splicing factor U2AF65, by binding to the polypyrimidine tract at the last intron 3′ splice site, stimulates both cleavage and polyadenylation by recruiting the CFI complex to the poly(A) site. The coupling is bidirectional, as evidenced by the fact that U2AF interaction with PAP stimulates splicing of the last intron. Another example of interconnection with the splicing machinery is through the splicing repressor polypyrimidine tract-binding protein (PTB). PTB can play a repressive role by competing with CstF binding to the DSE, but it can also have a stimulating function when associated to hnRNP H. In this case, PTB increases binding of hnRNP H to the G-rich auxiliary element, which in turn stimulates cleavage by recruiting CstF and PAP.

A poly(A) tail has also been proved to be important for an efficient translation of the mRNA into protein; in the cytoplasm the PABPs will facilitate the formation of a loop structure by interacting on one side with the poly(A) tail and on the other with the elongation initiation factor eIF4G, which in turn binds to the 5′ cap of the mRNA, thereby forming a closed circle which enhances translation. Other essential functions of the poly(A) tail are to promote transport of mRNA from the nucleus to the cytoplasm and to enhance mRNA stability, by preventing the action of the exosome complex of 3′→5′ exonucleases.

Studies over the past 20 years have identified numerous factors involved in pre-mRNA 3′ end processing, but more research is needed to elucidate the functional properties of many of the new protein factors found to be associated with this complex, especially those that might connect cleavage and polyadenylation with other nuclear events. An extensive characterization of the cross talk between 3′ processing and other cellular processes will be important for better understanding gene regulation on a global level.

See also: **Molecular Biology:** Messenger RNA Processing in Eukaryotes; RNA Polymerase II and Its General Transcription Factors; RNA Polymerase II Elongation Control in Eukaryotes; Transcription Termination.

Further Reading

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

Calvo O and Manley J (2003) Strange bedfellows: Polyadenylation factors at the promoter. *Genes and Development* 17: 1321–1327.

Colgan DF, Murthy KG, Prives C, and Manley JL (1996) Cell-cycle related regulation of poly(A) polymerase by phosphorylation. *Nature* 384: 282–285.

Danckwardt S, Hentze MW, and Kulozik AE (2008) 3′ end mRNA processing: Molecular mechanisms and implications for health and disease. *EMBO Journal* 27: 482–498.

Glover-Cutter K, Kim S, Espinosa J, and Bentley DL (2008) RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature Structural and Molecular Biology* 15: 71–78.

Hirose Y and Manley JL (1998) RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395: 93–96.

Mandel CR, Bai Y, and Tong L (2008) Protein factors in pre-mRNA 3′-end processing. *Cellular and Molecular Life Sciences* 65: 1099–1122.

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

Moore MJ and Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136: 688–700.
Proudfoot NJ, Furger A, and Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* 108: 501–512.
Ryan K and Bauer DLV (2008) Finishing touches: Post-translational modification of protein factors involved in mammalian pre-mRNA 3′ end formation. *International Journal of Biochemistry and Cell Biology* 40: 2384–2396.
Shatkin AJ and Manley JL (2000) The ends of the affair: Capping and polyadenylation. *Nature Structural Biology* 7: 838–842.
Shi Y, Campigli Di Giammartino D, Taylor D, et al. (2009) Molecular architecture of the human pre-mRNA 3′ processing complex. *Molecular Cell* 33: 365–376.
Zhao J, Hyman L, and Moore C (1999) Formation of mRNA 3′ ends in eukaryotes: Mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews* 63: 405–445.

New factors associated with the mammalian 3' processing complex

In the past decades, biochemical studies of individual 3' processing factors have greatly contributed to our understanding of the molecular mechanisms underlying the maturation of 3' ends of pre-mRNAs. More recently, some unique insights have been obtained from a purification and proteomic analysis of the entire 3' processing complex in its functional form (Shi et al., 2009, see appendix). The complex was purified at the "post-assembly" stage, when the complex had been assembled on a substrate RNA but little processing had occurred. Glycerol gradient sedimentation combined with RNA tag-based affinity purification led to identification of ~85 proteins that associated with substrates with intact processing signals, but not with RNAs containing AAUAAA mutations. These included nearly all previously identified 3' processing factors, with the exception of Clp1, a component of CFII. In fact, the other identified CFII subunit, Pcf11, was barely detectable, indicating that CFII may associate with the complex only transiently. CFII is the least understood of the 3' cleavage factors: Pcf11 interacts with the CTD of RNA polymerase II and is involved in transcription termination (Meinhart and Cramer, 2004), while Clp1 has been reported to have an RNA 5'-kinase activity that is important for tRNA splicing and activation of siRNAs (Weitzer and Martinez, 2007). Therefore, even if not tightly associated with the 3' processing complex, the function of CFII might be critical in connecting 3' processing to other nuclear pathways.

The canonical PAP (PAPα) was also absent from the proteins purified with the 3' complex. This suggests that it might be recruited at a later stage, which is in fact consistent with the earliest biochemical fractionation studies (Takagaki et al., 1988), or that another related protein might have taken its place. In this regard, it is interesting that PAPγ (also

known as neoPAP) was found to associate with the complex, although at low levels, indicating that PAPα and PAPγ may play redundant roles. A recent crystal structure of PAPγ shows it shares a conserved catalytic binding pocket while residues at the surface are more divergent (Yang et al., 2014). The diversity in the C-terminal domain of these two proteins could contribute to differential regulation, as this region is known to be critical for regulation of PAPα activity through post-translational modifications (e.g., Colgan et al., 1996; Vethantham et al., 2008), and distinct isoforms can be produced that result from alternative splicing that affects this region (Zhao and Manley, 1996). Interestingly, PAPα has been reported to be phosphorylated throughout the cell cycle and downregulated by hyperphosphorylation during M phase while PAPγ did not show evidence of phosphorylation or alternative isoforms (Topalian et al., 2001).

An additional protein that was identified in the proteomic analysis mentioned above is CstF64 tau, a conserved paralog of CstF64. CstF64 tau was shown initially to express specifically in the testis and brain (Wallace et al., 1999) and has been reported to mediate tissue-specific APA regulation (Li et al., 2012). However, its presence in the 3' complex purified from HeLa cells hinted to a more general role in pre-mRNA processing. Indeed, a recent study showed that CstF64 tau is widely expressed in mammalian tissues and has a similar RNA-binding pattern as CstF64 in vitro and in vivo (Yao et al., 2013). Also, the two proteins play redundant roles in alternative polyadenylation (APA) regulation such that depletion of either induces up-regulation of the other resulting in few changes in APA, but co-depletion leads to greater APA changes (Yao et al., 2012). Nonetheless, a significant difference between CstF64 and CstF64 tau is that the former binds symplekin with much higher affinity than the latter (Yao et al., 2013). Both proteins contain a "hinge" domain,

initially shown to mediate CstF64 binding to symplekin (Takagaki and Manley, 2000), and both paralogs bind symplekin in vitro. However, the interaction with CstF64 tau is inhibited by its C-terminal Pro-Gly rich domain, which is the most divergent region between the two proteins. It is therefore possible that their association with the 3' processing complex might be modulated by differential protein-protein interactions that depend on the Pro-Gly rich region, and these interactions, in turn, might reflect distinct functions in some aspects of mRNA 3' processing.

The proteomic purification of the 3' complex led to the identification of three proteins that were not previously implicated in mRNA 3' processing in mammals and are homologues of known yeast 3' processing factors: PP1, WDR33 and RBBP6; each of these proteins is discussed below.

PP1

PP1 is a serine/threonine phosphatase homologous to Glc7, which in yeast is known to play a role in poly(A) synthesis but not cleavage (He and Moore, 2005). Depletion of Glc7 in yeast has been shown to cause shortened poly(A) tails in vivo; similarly, Shi et al. (2009) showed that HeLa nuclear extract (NE) depleted of PP1 displayed inhibited poly(A) synthesis activity, which could be restored by adding back recombinant PP1. Glc7 dephosphorylates Pta1 (He and Moore, 2005) and therefore PP1 is likely to dephosphorylate symplekin, the mammalian homolog of Pta1. Since symplekin acts as a scaffolding protein in the 3' complex, it is possible that different states of phosphorylation of symplekin might affect its ability to interact with CPSF/CstF, ultimately modulating the efficiency of 3' end formation. However, the phosphorylation status of symplekin is currently unknown.

Given the presence of PP1, it is not surprising that the PP1 regulatory protein, PNUTS, was found in the 3' complex as well. PNUTS is known to form a stable complex with PP1 in mammalian cell lysates and has been shown to inhibit its catalytic activity (Kim et al., 2003). In addition, PNUTS has been shown to bind RNA in vitro (Kim et al., 2003), raising the possibility that PNUTS could have a direct function in recruiting PP1 to the 3' complex.

PP1 is a multifunctional protein that plays a role in regulating different aspects of mRNA maturation. For example, PP1 is known to be required for the second step of pre-mRNA splicing, targeting specific snRNP proteins (Shi et al., 2006), while in yeast, Glc7 functions in mRNA export, through dephosphorylation of Npl3 (Gilbert and Guthrie, 2004). Future studies will reveal if PP1 can function to bridge 3' end formation with such activities as splicing and mRNA export in mammalian cells.

WDR33

Another core subunit not identified in the mammalian 3' processing complex until recently is WDR33. The main characteristic of this protein is the fact that it contains a WD40 repeats region, a domain that is present in proteins involved in a wide range of cellular processes, as well as in the 3' factor CstF50. The underlying common function of most WD40-repeat proteins is that they coordinate multi-protein complex assemblies, where the repeating units serve as a scaffold for protein interactions (Xu and Min, 2011); in addition, WD40 domains have also been reported to bind both ubiquitin (Pashkova et al., 2010) and phosphorylated Ser/Thr (Reinhardt and Yaffe, 2013). The yeast homolog of WDR33 is Pfs2, which has been shown to be essential for 3' processing and might play a role in tethering the yeast CPF and CFIA complexes together (Ohnacker et al., 2000).

Similarly to Pfs2, WDR33 has recently been shown to interact with CPSF components and its depletion from HeLa NE abolishes both cleavage and polyadenylation (Shi et al., 2009). It remains to be determined if WDR33 coordinates the interaction between CstF and CPSF in mammalian cells the same way as Pfs2 does with the yeast homologues. For the future it will also be of interest to investigate whether the WD40 domain has a role in mediating the interaction of WDR33 with the other 3' factors and if this implicates Ser/Thr phosphorylated residues in the binding partners.

RBBP6

The third protein identified in the purification of the mammalian 3' complex that shares homology with a known yeast 3' processing factors is RBBP6. Its yeast counterpart is Mpe1, which is an essential gene. Mpe1 is an integral subunit of CPF and is required for both cleavage and polyadenylation (Vo et al., 2001). RBBP6 was first cloned in 1995 as a protein that interacts with the tumor suppressor Rb (Sakai et al., 1995) and later was shown to bind another tumor suppressor, p53 (Simons et al., 1997). The ~250 KD RBBP6 shares with Mpe1 three conserved domains in its N-terminus but has a unique long C-terminal extension that mediates the binding to p53 and Rb, raising the possibility that it may have a potential role in integrating 3' processing with these nuclear pathways. Chapter four of this thesis includes a more extensive introduction to RBBP6 and characterizes its function in polyadenylation, providing evidence that it is indeed a core component of the 3' processing machinery.

Other factors

In addition to the known core 3' processing factors and the other proteins discussed above, the proteomic analysis of the 3' complex allowed detection of about fifty proteins

that co-purified with the active complex (Shi et al., 2009). These include splicing factors such as U2AF65 and U1-70K, which were already found to mediate crosstalk between splicing and 3' processing (Awasthi and Alwine, 2003; Gunderson et al., 1998; Vagner et al., 2000) and additional splicing factors that have not been shown yet to take part in 3' processing. Among them there are several proteins that bind the pre-mRNA at the 3' splice sites and participate in the assembly of early spliceosomal complexes, for example SF1, which binds to the branch point sequence (Berglund et al., 1997) and several subunits of the multiprotein complexes SF3a and SF3b, which are part of the U2 snRNP that bind in close proximity of the branch point (Gozani et al., 1996). Finding these factors in the 3' complex confirms the physical interaction between the polyadenylation and splicing machineries and may indicate that these proteins may play a role in coupling splicing, for example of the terminal intron, to polyadenylation of pre-mRNAs.

The tumor suppressor Cdc73 is a component of the RNA pol II-associated PAF complex and was found to bind the 3' complex as well. Around the same time a biochemical study was published which indeed confirmed that Cdc73 functionally associates with CPSF and CstF. The authors suggest that Cdc73 might regulate mRNA processing by facilitating the recruitment of 3' factors to transcribed loci (Rozenblatt-Rosen et al., 2009). Other evidences of PAF's involvement in 3' processing come from yeast, where PAF has been shown to affect poly(A) tail length (Mueller et al., 2004) and poly(A) site selection (Penheiter et al., 2005). In addition, another study (Nagaike et al., 2011) found a role of PAF in mediating stimulation of mRNA 3′ processing by transcriptional activators, confirming its potential role in bridging 3' processing to transcription. It is, however, significant that none of the other five subunits of the PAF complex, except Cdc73, co-purified with the 3'

complex in the proteomic study (Shi et al., 2009) indicating the possibility that Cdc73 might actually have also a PAF-independent role in 3' processing of pre-mRNAs.

Another RNA pol II associated complex that was found to associate with the 3' processing machinery is the integrator. The integrator mediates 3' processing of U1 and U2 small nuclear RNAs (snRNAs) (Baillat et al., 2005), and several reports have emerged that suggest this complex might be multifunctional and play roles in various types of gene expression regulation beyond snRNA (Kapp et al., 2013; Takata et al., 2012; Zhang et al., 2013). Since almost all components of this complex were found in the proteomic analysis, it is very likely that the integrator might have also a yet undiscovered function in 3' end formation of mRNA as well.

Another interesting connection that stems out of the same proteomic purification and awaits to be confirmed biochemically involves the NEXT complex. NEXT is a multiprotein complex that is required for exosome-mediated degradation of noncoding RNAs such as promoter upstream transcripts (PROMPTs, also known as upstream antisense RNAs, uaRNAs), which are processed by the canonical 3' cleavage machinery (Almada et al., 2013; Ntini et al., 2013) but are rapidly degraded. Several of the known components of NEXT (Lubas et al., 2011), such as MTR4, ZCCHC8 and RBM7, were found to associate with the 3' complex (Shi et al., 2009), meaning that NEXT binds probably as a complex and might therefore be involved in the degradation of mRNAs by recruiting the exosome, perhaps as a quality control step. In fact some of the canonical exosome components were found as well in the proteomic purification of the 3' complex, including the catalytic subunit exosome10 (also known as Rrp6) and a number of non-catalytic

subunits; although judging by the number of peptides the exosome is only loosely associated with the 3' machinery.

Finally, links between DNA damage response factors and the 3' processing complex are especially intriguing. Previous studies have described similar connections, showing that 3' processing is inhibited following DNA damage, concomitantly with an increased interaction between CstF50 and the BARD/BRCA1 complex (Kleiman and Manley, 1999, 2001). This interaction was later shown also to stimulate the deadenylation activity of PARN during DNA damage leading to RNA degradation (Cevher et al., 2010). In addition, p53 also interacts with CstF50 and BARD1 and has an inhibitory effect on 3' processing of housekeeping genes following UV treatment (Nazeer et al., 2011).

The proteomic study mentioned above identified DNA-PK as associated with the pre-mRNA 3' processing complex. DNA-PK is a nuclear serine/threonine kinase that is comprised of a regulatory subunit, containing the Ku70/86 components, and a catalytic subunit, DNA-PK_{cs}; interestingly all of these subunits were found to associate with the 3' complex. DNA-PK is a molecular sensor for DNA damage: it is involved in DNA nonhomologous end joining and is required for double-strand break repair and VDJ recombination (reviewed in Collis et al., 2005). DNA-PK must be bound to DNA to express its catalytic properties but the fact that it is associated with the 3' complex raises the possibility that RNA might activate it as well. The main question would be to understand if it functions to somehow connect the cellular double strand break response to 3' processing or if it has a separate function in the maturation of 3' ends of mRNAs. One way to check this could be by using one of the several commercially available small molecule inhibitors of DNA-PK (Davidson et al., 2013), and check if it affects 3' cleavage activity, either in the

presence or not of DNA damage. It is very likely that one of the 3' processing factors might be a target for phosphorylation by DNA-PK; in fact several 3' factors have been reported to be phosphorylated (reviewed in Ryan and Bauer, 2008) and in most of the cases the kinase is unknown. Interestingly DNA-PK was shown also to phosphorylate and modulate PARP1 activity (Ariumi et al., 1999), another protein identified in association with the 3' complex.

PARP1 is an enzyme that catalyzes the post-translational modification known as Poly(ADP-Ribosyl)ation (PARylation) and which is known to take part in several cellular processes, including DNA damage detection and repair, chromatin modification and transcription (reviewed in Ji and Tulin, 2010; Krishnakumar and Kraus, 2010). Chapter 3 of this thesis shows that PARP1, although not a core component of the 3' processing complex, is able to prevent polyadenylation during heat shock by PARylating PAP and inhibiting its activity.

Altogether, it is becoming evident that the 3' processing machinery is much more complex than previously thought. The need to connect 3' cleavage and polyadenyaltion with many cellular pathways such as transcription, splicing, tumorigenesis and DNA damage, might explain why such big machinery is involved in cleavage and polyadenylation. Coordination between the different steps of genes expression is critical for proper cellular function and future studies need to focus on understanding in more detail the roles of the many new proteins discussed above.

In this thesis we tried to unravel the functions of two new factors associated with the 3' complex: RBBP6 and PARP1.

References

Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360-363.

Ariumi, Y., Masutani, M., Copeland, T.D., Mimori, T., Sugimura, T., Shimotohno, K., Ueda, K., Hatanaka, M., and Noda, M. (1999). Suppression of the poly(ADP-ribose) polymerase activity by DNA-dependent protein kinase in vitro. Oncogene *18*, 4616-4625.

Awasthi, S., and Alwine, J.C. (2003). Association of polyadenylation cleavage factor I with U1 snRNP. RNA *9*, 1400-1409.

Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell *123*, 265-276.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell *89*, 781-787.

Cevher, M.A., Zhang, X., Fernandez, S., Kim, S., Baquero, J., Nilsson, P., Lee, S., Virtanen, A., and Kleiman, F.E. (2010). Nuclear deadenylation/polyadenylation factors regulate 3' processing in response to DNA damage. The EMBO Journal *29*, 1674-1687.

Colgan, D.F., Murthy, K.G., Prives, C., and Manley, J.L. (1996). Cell-cycle related regulation of poly(A) polymerase by phosphorylation. Nature *384*, 282-285.

Collis, S.J., DeWeese, T.L., Jeggo, P.A., and Parker, A.R. (2005). The life and death of DNA-PK. Oncogene *24*, 949-961.

Davidson, D., Amrein, L., Panasci, L., and Aloyz, R. (2013). Small Molecules, Inhibitors of DNA-PK, Targeting DNA Repair, and Beyond. Frontiers in Pharmacology *4*, 5.

Gilbert, W., and Guthrie, C. (2004). The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA. Molecular Cell *13*, 201-212.

Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes & Development *10*, 233-243.

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. Molecular Cell *1*, 255-264.

He, X., and Moore, C. (2005). Regulation of yeast mRNA 3' end processing by phosphorylation. Molecular Cell *19*, 619-629.

Ji, Y., and Tulin, A.V. (2010). The roles of PARP1 in gene control and cell differentiation. Current Opinion in Genetics & Development *20*, 512-518.

Kapp, L.D., Abrams, E.W., Marlow, F.L., and Mullins, M.C. (2013). The integrator complex subunit 6 (Ints6) confines the dorsal organizer in vertebrate embryogenesis. PLoS Genetics *9*, e1003822.

Kim, Y.M., Watanabe, T., Allen, P.B., Kim, Y.M., Lee, S.J., Greengard, P., Nairn, A.C., and Kwon, Y.G. (2003). PNUTS, a protein phosphatase 1 (PP1) nuclear targeting subunit. Characterization of its PP1- and RNA-binding domains and regulation by phosphorylation. The Journal of Biological Chemistry *278*, 13819-13828.

Kleiman, F.E., and Manley, J.L. (1999). Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50. Science *285*, 1576-1579.

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.

Krishnakumar, R., and Kraus, W.L. (2010). The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. Molecular Cell *39*, 8-24.

Li, W., Yeh, H.J., Shankarling, G.S., Ji, Z., Tian, B., and MacDonald, C.C. (2012). The tauCstF-64 polyadenylation protein controls genome expression in testis. PloS One *7*, e48373.

Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A., and Jensen, T.H. (2011). Interaction profiling identifies the human nuclear exosome targeting complex. Molecular Cell *43*, 624-637.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature *430*, 223-226.

Mueller, C.L., Porter, S.E., Hoffman, M.G., and Jaehning, J.A. (2004). The Paf1 complex has functions independent of actively transcribing RNA polymerase II. Molecular Cell *14*, 447-456.

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

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

Ntini, E., Jarvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R.*, et al.* (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nature Structural & Molecular Biology *20*, 923-928.

Ohnacker, M., Barabino, S.M., Preker, P.J., and Keller, W. (2000). The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex. The EMBO Journal *19*, 37-47.

Pashkova, N., Gakhar, L., Winistorfer, S.C., Yu, L., Ramaswamy, S., and Piper, R.C. (2010). WD40 repeat propellers define a ubiquitin-binding domain that regulates turnover of F box proteins. Molecular Cell *40*, 433-443.

Penheiter, K.L., Washburn, T.M., Porter, S.E., Hoffman, M.G., and Jaehning, J.A. (2005). A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. Molecular Cell *20*, 213-223.

Reinhardt, H.C., and Yaffe, M.B. (2013). Phospho-Ser/Thr-binding domains: navigating the cell cycle and DNA damage response. Nature Reviews Molecular Cell Biology *14*, 563-580.

Rozenblatt-Rosen, O., Nagaike, T., Francis, J.M., Kaneko, S., Glatt, K.A., Hughes, C.M., LaFramboise, T., Manley, J.L., and Meyerson, M. (2009). The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. Proceedings of the National Academy of Sciences of the United States of America *106*, 755-760.

Ryan, K., and Bauer, D.L. (2008). Finishing touches: post-translational modification of protein factors involved in mammalian pre-mRNA 3' end formation. The International Journal of Biochemistry & Cell Biology *40*, 2384-2396.

Sakai, Y., Saijo, M., Coelho, K., Kishino, T., Niikawa, N., and Taya, Y. (1995). cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product. Genomics *30*, 98-101.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Molecular Cell *33*, 365-376.

Shi, Y., Reddy, B., and Manley, J.L. (2006). PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. Molecular Cell *23*, 819-829.

Simons, A., Melamed-Bessudo, C., Wolkowicz, R., Sperling, J., Sperling, R., Eisenbach, L., and Rotter, V. (1997). PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb. Oncogene *14*, 145-155.

Takagaki, Y., and Manley, J.L. (2000). Complex protein interactions within the human polyadenylation machinery identify a novel component. Molecular and Cellular Biology *20*, 1515-1525.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1988). Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell *52*, 731-742.

Takata, H., Nishijima, H., Maeshima, K., and Shibahara, K. (2012). The integrator complex is required for integrity of Cajal bodies. Journal of Cell Science *125*, 166-175.

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. Molecular and Cellular Biology *21*, 5614-5623.

Vagner, S., Vagner, C., and Mattaj, I.W. (2000). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. Genes & Development *14*, 403-413.

Vethantham, V., Rao, N., and Manley, J.L. (2008). Sumoylation regulates multiple aspects of mammalian poly(A) polymerase function. Genes & Development *22*, 499-511.

Vo, L.T., Minet, M., Schmitter, J.M., Lacroute, F., and Wyers, F. (2001). Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. Molecular and Cellular Biology *21*, 8346-8356.

Wallace, A.M., Dass, B., Ravnik, S.E., Tonk, V., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., and MacDonald, C.C. (1999). Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells. Proceedings of the National Academy of Sciences of the United States of America *96*, 6763-6768.

Weitzer, S., and Martinez, J. (2007). The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs. Nature *447*, 222-226.

Xu, C., and Min, J. (2011). Structure and function of WD40 domain proteins. Protein & Cell *2*, 202-214.

Yang, Q., Nausch, L.W., Martin, G., Keller, W., and Doublie, S. (2014). Crystal structure of human poly(a) polymerase gamma reveals a conserved catalytic core for canonical poly(a) polymerases. Journal of Molecular Biology *426*, 43-50.

Yao, C., Biesinger, J., Wan, J., Weng, L., Xing, Y., Xie, X., and Shi, Y. (2012). Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. Proceedings of the National Academy of Sciences of the United States of America *109*, 18773-18778.

Yao, C., Choi, E.A., Weng, L., Xie, X., Wan, J., Xing, Y., Moresco, J.J., Tu, P.G., Yates, J.R., 3rd, and Shi, Y. (2013). Overlapping and distinct functions of CstF64 and CstF64tau in mammalian mRNA 3' processing. RNA *19*, 1781-1790.

Zhang, F., Ma, T., and Yu, X. (2013). A core hSSB1-INTS complex participates in the DNA damage response. Journal of Cell Science *126*, 4850-4855.

Zhao, W., and Manley, J.L. (1996). Complex alternative RNA processing generates an unexpected diversity of poly(A) polymerase isoforms. Molecular and Cellular Biology *16*, 2378-2386.

CHAPTER 2

Mechanisms and consequences of alternative polyadenylation

This chapter was previously published as a review in *Molecuar Cell* (2011, 43:853-866)

Molecular Cell
Review

Mechanisms and Consequences of Alternative Polyadenylation

Dafne Campigli Di Giammartino,[1,2] Kensei Nishida,[1,2] and James L. Manley[1,*]
[1]Department of Biological Sciences, Columbia University, New York, NY 10027, USA
[2]These authors contributed equally to this work
*Correspondence: jlm2@columbia.edu
DOI 10.1016/j.molcel.2011.08.017

Alternative polyadenylation (APA) is emerging as a widespread mechanism used to control gene expression. Like alternative splicing, usage of alternative poly(A) sites allows a single gene to encode multiple mRNA transcripts. In some cases, this changes the mRNA coding potential; in other cases, the code remains unchanged but the 3′ UTR length is altered, influencing the fate of mRNAs in several ways, for example, by altering the availability of RNA binding protein sites and microRNA binding sites. The mechanisms governing both global and gene-specific APA are only starting to be deciphered. Here we review what is known about these mechanisms and the functional consequences of alternative polyadenylation.

Introduction

Regulation of mRNA processing is well known to play a fundamental role in determining the outcome of gene expression, but alternative polyadenylation (APA) has only recently gained attention as a major player influencing the dynamics of gene regulation. The maturation of 3′ ends of mRNA precursors (pre-mRNAs), although a relatively simple process, has been known for some time to require a complex set of protein factors. One explanation for this has been that the complexity reflects the importance of regulating 3′ end formation. It is well established that polyadenylation can contribute in several ways to gene control (Colgan and Manley, 1997; Barabino and Keller, 1999); however, in the past few years it has become clear that regulation of APA is considerably more widespread than previously thought and can affect gene expression in multiple ways. In this review, we discuss both the mechanisms and the consequences of APA and how regulated mRNA 3′ processing contributes to cell growth control and disease. We begin by providing some background and a brief overview of 3′ processing and its regulation.

The mature 3′ ends of nearly all eukaryotic mRNAs, with the exception of replication-dependent histone transcripts, are created by a two-step reaction that involves an endonucleolytic cleavage of the pre-mRNA, followed by synthesis of a polyadenylate tail onto the upstream cleavage product. This relatively simple reaction requires numerous protein factors that are directed to the correct cleavage site by sequence elements within the pre-mRNA (reviewed in Colgan and Manley, 1997; Mandel et al., 2008; Millevoi and Vagner, 2010; Zhao et al., 1999). The core molecular machinery responsible for 3′ end formation in mammals includes four multisubunit protein complexes, CPSF (cleavage and polyadenylation specificity factor), CstF (cleavage stimulation factor), CFI and CFII (cleavage factors I and II), as well as additional accessory factors and the single subunit poly(A) polymerase (PAP). RNA polymerase II (RNAP II), and specifically the C-terminal domain of its largest subunit, also plays an important role in processing. The assembly of the 3′ end processing complex on the pre-mRNA begins with the cooperative interaction of CPSF and CstF with specific sequences; the canonical poly(A) signal AAUAAA located upstream of the cleavage site, recognized by CPSF (specifically by the CPSF160 subunit); and a less defined downstream U/GU-rich region that constitutes the binding site for CstF (through the CstF64 subunit). Usage of one poly(A) site over another is often attributed to the relative ''strength'' of these core elements, but in fact auxiliary sequences and protein factors play a role in influencing poly(A) site choice in different contexts. Indeed, today we know that many more proteins than previously thought are involved in the fine-tuning of 3′ end formation (Shi et al., 2009), and a good number of these likely mediate crosstalk between pre-mRNA maturation and other nuclear events.

Polyadenylation influences many aspects of mRNA metabolism. Transcription termination by RNAP II, mRNA stability, mRNA export to the cytoplasm, and the efficiency of translation are all dependent on 3′ processing. These topics have all been reviewed recently and won't be discussed here (Ji et al., 2011; Richard and Manley, 2009; Vinciguerra and Stutz, 2004; Zhang et al., 2010).

In recent years it has become increasingly evident that APA is extensively used to regulate gene expression. For example, 50% or more of human genes encode multiple transcripts derived from APA (Tian et al., 2005). We will consider here two general classes of APA. In some cases the alternative poly(A) sites are located in internal introns/exons, and therefore APA events will produce different protein isoforms; we will refer to this type as CR-APA (coding region-APA). In other cases, APA sites are all located in the 3′ untranslated region (3′ UTR), resulting in transcripts with 3′ UTRs of different length but encoding the same protein; we refer to this type of APA as UTR-APA (Figure 1).

While CR-APA can affect gene expression qualitatively by producing distinct protein isoforms, UTR-APA has the potential to affect expression quantitatively. 3′ UTRs often harbor microRNA (miRNA) binding sites and/or other regulatory sequences, such as AU-rich elements (AREs) (Barreau et al., 2005; Fabian et al., 2010). Longer 3′ UTRs will more likely possess such

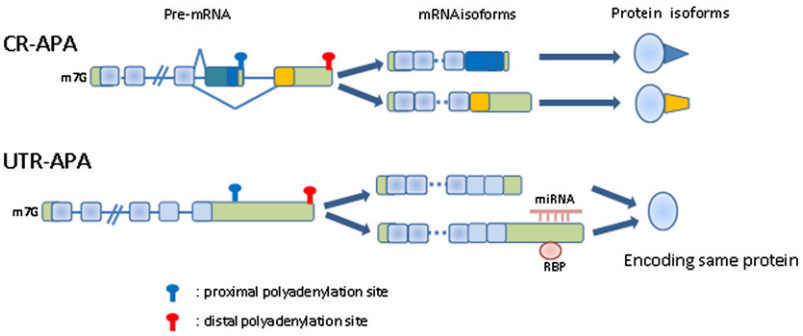


Figure 1. Schematic Representation of CR-APA and UTR-APA

CR-APA produces mRNA isoforms with distinct C-terminal coding regions, resulting in distinct protein isoforms. UTR-APA produces distinct mRNA isoforms with different-length 3′ UTRs but encodes the same protein. Longer 3′ UTRs usually contain *cis*-regulatory elements, such as miRNA and/or protein binding sites, which often bring about mRNA instability or translational repression. CR-APA, coding region-alternative polyadenylation; UTR-APA, 3′ UTR-alternative polyadenylation. Light green boxes, untranslated regions; light blue boxes, shared coding regions; dark blue and yellow boxes, unshared coding regions; lines, introns.

signals, or more of them, and the mRNA will therefore likely be more prone to negative regulation. Indeed, the amount of protein generated by an mRNA has been shown to depend on its 3′ UTR length, such that transcripts with shorter 3′ UTRs produce higher levels of protein (Mayr and Bartel, 2009; Sandberg et al., 2008). Furthermore, as discussed below, the length of the 3′ UTR can affect not only the stability but also the localization, transport, and translational properties of the mRNA.

Differential processing at multiple poly(A) sites can be influenced by physiological conditions such as cell growth, differentiation, and development or by pathological events such as cancer. The mechanisms that regulate such global events are mostly unknown, and intense research is currently being carried out in order to better understand this phenomenon at the molecular level. In this review, we will provide an overview of current studies on APA, both on genome-wide analyses and specific examples, focusing on the possible mechanisms of regulation and the functional consequences of differential poly(A) site usage.

Genome-wide Analyses of APA

Analyses of APA at the global level have been largely responsible for the appreciation that APA constitutes a significant contributor to gene regulation across species. Genome-wide studies carried out in humans, mice, worms, yeast, plants, and algae revealed that the number of genes encoding transcripts with multiple poly(A) sites ranges from 10% to 15% in *S. cerevisiae* (Nagalakshmi et al., 2008) to ~54% in humans (Tian et al., 2005). Significantly, orthologous human and mouse genes were found to have a high similarity in the numbers of 3′ ends mapped for each gene (Ara et al., 2006; Tian et al., 2005), indicating that APA sites have been actively selected during evolution. Interestingly, however, as shown by a genome-wide bioinformatic analysis, the majorities of tissue-specific and noncanonical poly(A) sites seem to be species specific and are not themselves conserved (Ara et al., 2006). This suggests that gain or loss of APA sites is a frequent event in mammalian genomes, implying that very often novel sites would be quickly lost if their presence is either neutral or deleterious.

Through genome-wide analysis of APA, it has been possible to define a pattern that relates the proliferation and differentiation status of cells with the length of 3′ UTRs (Figure 2). Specifically, states of increased proliferation, dedifferentiation, and disease (i.e., cancer) are associated with a general shortening in 3′ UTR length, while 3′ UTRs tend to be longer during late developmental stages and cell differentiation (Ji et al., 2009; Ji and Tian, 2009; Mayr and Bartel, 2009; Sandberg et al., 2008; Shepard et al., 2011; Wang et al., 2008a). In *C. elegans*, the length of the 3′ UTRs correlates inversely with animal age (Mangone et al., 2010). Interestingly, a considerable number of miRNAs diminish in expression over adult life in *C. elegans* (Ibáñez-Ventoso and Driscoll, 2009), suggesting that a relaxation in miRNA-3′ UTR control of mRNA stability/translation might be a general feature of advancing adult life.

Given the fact that different types of APA exist, it is interesting to note that CR-APA and UTR-APA can be differently regulated. During T cell activation, for example, CR-APA events occur at both early and late stages of activation (Sandberg et al., 2008). Moreover, proximal-to-distal and distal-to-proximal shifts in APA were similarly represented, whereas changes in UTR-APA were mostly evident during late stages of activation with a clear pattern of increased usage of the proximal site. This suggests that the regulation of different types of APA (CR-APA versus UTR-APA) may rely at least in part on different mechanisms. For example, CR-APA often occurs in conjunction with splicing of an overlapping intron, and it is thus possible that splicing regulation may also affect APA. Also, the observation that

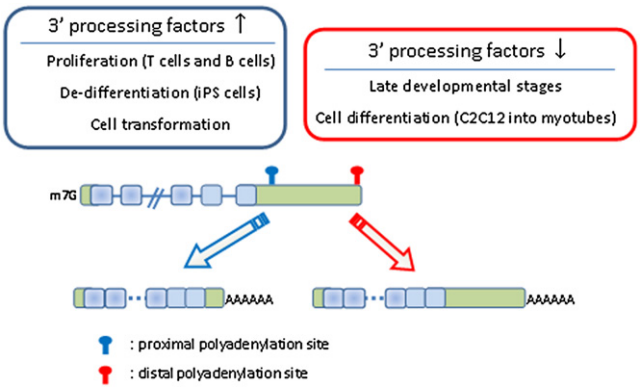


Figure 2. Connecting APA to Cellular Proliferative and Developmental States

Enhanced proliferation such as during dedifferentiation (e.g., in the generation of iPS cells), T cell activation, or cellular transformation are associated with upregulation in expression of certain 3′ processing factors and with increased usage of proximal poly(A) sites. Late developmental stages and cellular differentiation (e.g., differentiation of C2C12 into myotubes) are associated with downregulation of expression of 3′ processing factors and increased usage of distal poly(A) sites.

changes in UTR-APA is not an early event during T cell activation suggests that it is perhaps necessary to enhance the expression or activity of basal or auxiliary 3′ processing factors, which then function at later times. A consequence of the differential temporal behavior of CR-APA versus UTR-APA in T cell activation is that during early activation, there will be more APA events affecting the protein isoform produced, while during later stages, APA will lead to transcripts that differ in the length of their 3′ UTR, and therefore the main effect will be changes in the abundance of the proteins produced.

To understand how 3′ UTR lengthening is related to regulation of biological processes, the association of selected genes with gene ontology (GO) terms was examined during different developmental/differentiation states. 3′ UTR lengthening during mouse embryonic development coincides with upregulation of genes involved in morphogenesis and differentiation, such as cell morphogenesis and extracellular structure organization, and with downregulation of genes involved in proliferation, such as DNA replication and cell-cycle phase (Ji et al., 2009). The same pattern is detected during differentiation of proliferative C2C12 muscle cells into myotubes. In contrast, during generation of human and mouse induced pluripotent stem (iPS) cells, most of the same GO terms displayed regulation in the opposite direction. It is of particular interest that 3′ processing factors, such as CPSF and CstF components, were found to be strongly upregulated during generation of iPS cells (Ji and Tian, 2009). This may hint at a regulatory mechanism where the abundance of 3′ processing factors in undifferentiated cells (such as iPS) facilitates the usage of the proximal poly(A) site, which usually has a ''weaker'' consensus than the distal site (see below), thereby generating transcripts with shorter 3′ UTRs. Since both early embryonic and iPS cells are rapidly proliferating, a significant question is whether differentiation per se affects APA in a system where proliferation and differentiation could be uncoupled. For example, the leukemic cell line HL60 is capable of differentiating into neutrophils or monocytes (in response to different stimuli) even when the cell cycle is blocked in early G1 or S phase, indicating that differentiation and proliferation can be regulated independently (Brown et al., 2002). It would be of interest to compare changes in usage of APA sites before and after differentiation, independently from alterations in the proliferation rate.

Cancer cells provide an important subset of proliferating cells. In this regard, it is remarkable that in primary tumor samples from a mouse leukemia/lymphoma model (Singh et al., 2009), APA seems to define molecular signatures that can distinguish similar tumor subtypes with high accuracy. Mice lacking p53 and the core NHEJ factor DNA ligase IV develop pro-B cell lymphomas with frequent genomic amplification of *c-Myc* (designated LPC), while mice lacking p53 and the accessory NHEJ factor Artemis develop lymphomas with either *c-Myc* or *N-Myc* amplification (APC or APN, respectively). While LPC, APC, and APN lymphomas are histologically and immunophenotypically indistinguishable, using microarray analysis, specific sets of transcripts with differential 3′ UTR processing were identified between these lymphoma subtypes. The diagnostic capacity of these assignments was confirmed by analysis of unknown samples, which were correctly assigned at rates of 100% for LPC, 92% for APC, and 74% for APN. These results anticipate the possibility of future usage of APA as a molecular biomarker with prognostic potential. In accordance with previous findings (Mayr and Bartel, 2009), shortening of 3′ UTRs in the cancer cells compared to normal (pro-B) cells was the most common pattern observed, although some transcripts with elongated 3′ UTRs were also detected. In addition, levels of a number of mRNAs encoding 3′ processing factors, notably those encoding CstF subunits, were upregulated in the lymphomas. These data support a model in which changes in expression and/or stoichiometry of 3′ processing factors lead to changes in poly(A) site selection, for instance, enhanced expression of these factors might help increase utilization of suboptimal proximal poly(A) signals and thereby contribute to the tumor-specific shortening of 3′ UTRs (see below for further discussion of this hypothesis).

Bioinformatic approaches and genomic studies have also been used to shed light on the link between differential usage of APA sites in relation to tissue specificity (Wang et al., 2008a; Zhang et al., 2005). Using expressed sequence tag data, 42 distinct human tissue types were analyzed, revealing considerable tissue-specific APA. For example, retina, placenta, blood, and ovary were more likely to use proximal poly(A) sites, while tissues from bone marrow, uterus, brain, and nervous system showed increased usage of the distal poly(A) sites. These tissue-specific preferences are observed on a global rather than gene-specific level, indicating that the mechanism may lie in tissue-specific regulation or expression of polyadenylation factors. It will be of interest to compare the proliferation potential of these two groups of tissues and access whether it correlates with the usage of the proximal/distal poly(A) sites. For example, brain tissues are known to have low mitotic activity, suggesting that decreased proliferation is associated with tissues harboring transcripts with longer 3′ UTRs.

Another important role of large-scale studies is the contribution they have made to our understanding of the role that *cis* sequences play in APA. Computational analyses have indicated that variations of the canonical AAUAAA sequence are relatively frequent, occurring in more than 30% of 3′ ends (Tian et al., 2005). Interestingly, while the canonical sequence predominates in genes with a unique poly(A) site, the less-conserved, variant poly(A) sites occur frequently in genes with multiple poly(A) sites. In these cases, the variant sites are usually located promoter proximal, whereas canonical poly(A) signal often appears downstream of variant sites (Beaudoing et al., 2000). This suggests that the efficient utilization of the proximal alternative poly(A) signals is likely dependent on additional auxiliary factors, different abundance of core 3′ processing factors, and/or auxiliary surrounding RNA sequences. Indeed, genome-wide analyses recently identified conserved motifs, mostly around alternative poly(A) sites, that might help explain, at least in part, how the choice between the usage of the different sites is made (Nunes et al., 2010; Ozsolak et al., 2010). For example, a genome-wide analysis of over 10,000 human poly(A) sites shows that about one-third of noncanonical, proximal, poly(A) signals tend to have higher frequency of U and GU nucleotides downstream of the poly(A) site compared with canonical poly(A) signals, implying that a strong CstF binding site might compensate for the absence of a consensus hexanucleotide (Nunes et al.,

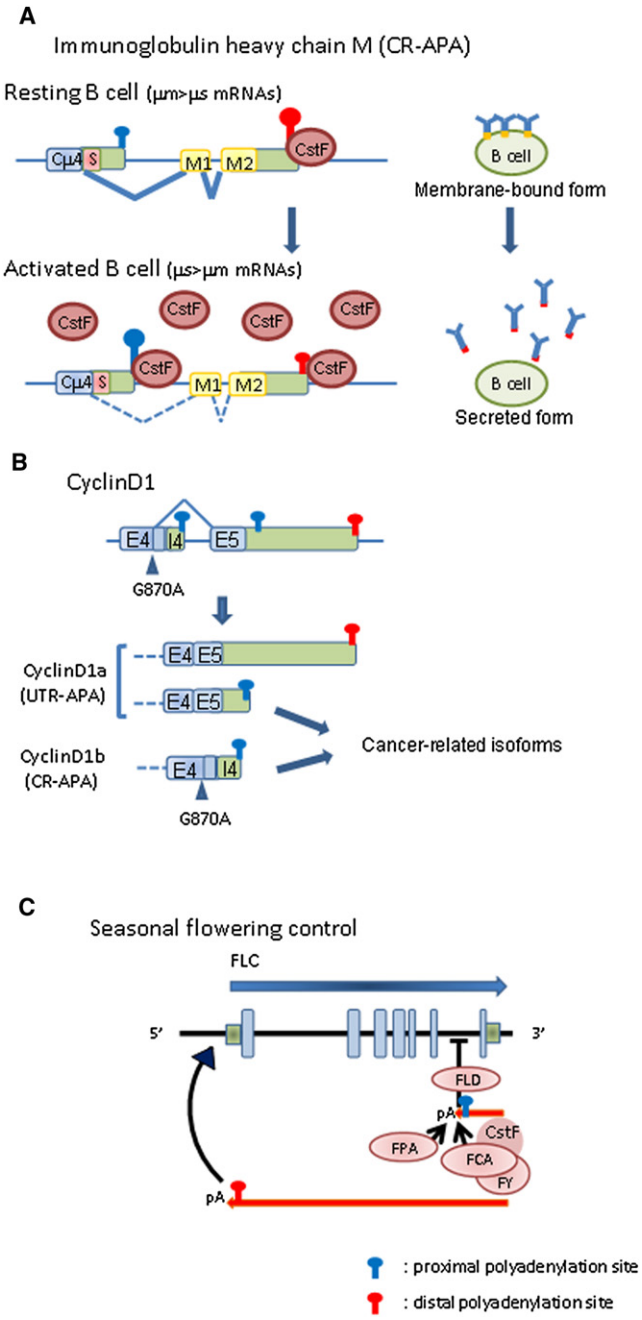


Figure 3. Examples of Gene Regulation by APA

(A) The immunoglobulin heavy-chain M gene is partly shown; a constant region (Cμ4) is shared by both μm and μs mRNAs, while exons M1 and 2 (yellow boxes) and S (red boxes) are specific to μm and μs mRNAs, respectively. In resting B cells, the amount of CstF is limiting, and the distal poly(A) site, which binds CstF more avidly, is preferentially used, resulting in production of the membrane-bound form of IgM (μm). In activated B cells, the concentration of CstF is elevated and no longer limiting, so the proximal, first transcribed poly(A) site is preferentially selected, leading to production of secreted-form IgM (μs). Additional factors, such as the transcription factor Ell2 (see text), may also contribute to the switch.

(B) Cyclin D1 is subject to both UTR-APA and CR-APA. Two major isoforms are created by CR-APA: cyclin D1a (full-length isoform) and cyclin D1b (truncated isoform). The truncated isoform is associated with a polymorphism at the end of exon 4 (E4) (G870A, arrowhead), resulting in increased usage of the poly(A) site located in intron 4 (I4). This isoform is retained in the nucleus and is associated with increasing transforming capability. UTR-APA of Cyclin D1 leads to increased usage of the weak proximal poly(A) site in cancer cells generally or in the usage of a newly mutational-derived proximal poly(A) site in mantle cell lymphoma. In both, mRNAs with shorter 3′ UTRs are generated. Light green boxes, untranslated regions; light blue boxes, shared coding regions; lines, introns.

(C) Seasonal flowering control by antisense RNA transcript. FPA and FCA promote selection of the proximal poly(A) site of an antisense transcript that initiates downstream of the *FLC* gene (red arrows) by stimulating 3′ end formation at that site. 3′ end processing at the proximal poly(A) site recruits the histone demethylase, FLD, which induces histone modifications on internal nucleosomes that result in silencing the sense FLC transcript (blue arrow). In the absence of FPA and FCA, the distal poly(A) site of the antisense transcripts is selected. This may facilitate the recruitment of positive transcription factors to the FLC promoter, resulting in activation of FLC transcription.

2010). How these variations in core sequences, as well as other signals, contribute to APA is discussed below.

Specific Examples of APA

There are now a growing number of examples of specific APA events for which the function and/or mechanism is at least reasonably well understood. In this section we discuss several of these, highlighting those that play important roles in cell growth and differentiation or disease. In the following sections, we discuss in more detail the known mechanisms and functions of APA.

Immunoglobulin M Heavy Chain

The immunoglobulin (Ig) M heavy-chain gene provided perhaps the first example of APA, specifically of CR-APA, as a regulatory mechanism (Alt et al., 1980; Early et al., 1980; Rogers et al., 1980; reviewed in Peterson, 2007). During the transition of a B cell to a plasma cell, the IgM protein switches from a membrane-bound form to a secreted form. This switch is caused in large part by the selection of one of two poly(A) sites. The secreted form is produced by using a proximal poly(A) site, while the membrane-bound form is produced from the spliced Cu4-M1 mRNA by using distal poly(A) site (Figure 3A). The switch from membrane-bound to secreted form IgM in LPS-induced mouse primary B cells was shown to be accompanied with a specific increase of CstF64 protein levels (Takagaki et al., 1996). Moreover, overexpression of CstF64 in a B cell line was enough to induce the switch from membrane-bound to secreted form by preferentially using the proximal poly(A) site. In the same context, conditional knockdown of CstF64 also showed a relative enhancement of distal poly(A) site usage (Takagaki and Manley, 1998).

Subsequent investigations of IgM switching mechanisms revealed that ELL2, a protein related to the transcription elongation factor ELL, may also contribute to selection of the proximal poly(A) site. Martincic et al. (2009) provided evidence that ELL2 and CstF64 track together with RNAP II across the IgM gene. Like CstF64, ELL2 levels were induced in LPS-activated B cells. This may provide an additional mechanism to enhance CstF levels at the proximal poly(A) site, increasing the efficiency with which it is utilized.

Germ Cell-Specific APA

Mammalian testes have unique APA processing characteristics. The canonical AAUAAA sequence is infrequent in testis-specific mRNAs, which often use proximal poly(A) sites that are not

efficiently polyadenylated in somatic cells (Liu et al., 2007a; MacDonald and Redondo, 2002; McMahon et al., 2006; Zhang et al., 2005). A CstF64 variant (τCstF64) is highly expressed in male germ cells compared to other tissues (Monarez et al., 2007), and it may contribute to the different cleavage specificity observed in germ cells. In agreement with this hypothesis, knockout of *Cstf2t*, the gene encoding τCstF64, in mice resulted in a spermatogenetic defect, but no significant influence on somatic cells (Dass et al., 2007; Hockert et al., 2011). In addition, microarray experiments showed that transcripts encoding a number of core polyadenylation factors were significantly more abundant in germ cells than somatic cells (Liu et al., 2007a). Furthermore, during spermatogenesis, *τCstF64* levels were found to increase, while those of *CstF64* decreased (Liu et al., 2007a). Germ cell-specific and stage-specific APA events may thus be induced by altered expression levels of 3′ processing factors, including τCstF64.

One interesting example of germ cell-specific APA is provided by transcripts encoding a transcription factor, BZW1, which exist as three mRNA isoforms created by UTR-APA. The two longer isoforms are expressed ubiquitously at low levels, while the shortest is expressed at high levels only in testis, especially spermatogonia. Expression of EGFP-BZW1 fusion genes with distinct BZW1 3′ UTRs showed that the shortest transcript had the lowest translation efficiency, suggesting that BZW1 expression is fine-tuned through 3′ UTR length in a cell type-specific manner (Yu et al., 2006). This result is contrary to the expectation that shorter 3′ UTRs produce more protein than those with longer 3′ UTRs. In this case, low expression of the short isoform of BZW1 may be due to its unusually short 3′ UTR, which is ~25 times shorter than the average (500 nt) 3′ UTR in testis germ cells (Sood et al., 2006) and may negatively affect translational efficiency (Tanguay and Gallie, 1996).

Disease-Related APA

Only a few studies have focused on the pathophysiology of diseases related to APA (Chen et al., 2006). However, it is well established that 3′ UTRs play an important role in various diseases and their progression (Conne et al., 2000). We describe two disease-related examples reflecting changes in APA caused by mutated poly(A) signals; one is the equivalent of a loss-of-function mutation and the other of a gain-of-function mutation.

IPEX (immune dysfunction, polyendocrinopathy, enteropathy, X-linked), a disease characterized by dysfunction of regulatory T cells and subsequent autoimmunity, is caused by mutations in the *FOXP3* gene, which encodes a transcription factor containing a forkhead DNA binding domain. Most of the reported mutations affect the forkhead domain, resulting in disruption of DNA binding. However, a rare mutation lies within the poly(A) signal (AAUAAA → AAUGAA). This mutation leads to skipping of the first poly(A) signal and usage of the next signal, located 5.1 kb downstream. This appears to result in an unstable mRNA, leading to a decrease of FOXP3 protein and in this way leading to IPEX (Bennett et al., 2001).

The loss of controlled cell-cycle progression is a critical event in tumorigenesis. Cyclin D1 regulates progression through G1-S phase by its association with cyclin-dependent kinase 4 or 6 (Knudsen et al., 2006). Two major isoforms, cyclin D1a and b, are created by alternative splicing/polyadenylation (CR-APA) (Figure 3B). Cyclin D1a mRNA is full length, whereas cyclin D1b mRNA is cleaved at an APA site within an intron. Cyclin D1b protein is constitutively nuclear, resulting in increased transforming capability (Lu et al., 2003; Solomon et al., 2003). High expression of cyclin D1b is observed in several human cancers, including breast and prostate cancer (Burd et al., 2006; Wang et al., 2008b). A G870A polymorphism at the end of exon 4 has been associated with production of the cyclin D1b isoform (Comstock et al., 2009; Knudsen et al., 2006). This polymorphism may cause impaired recognition by the splicing machinery, resulting in APA using the intron 4 poly(A) signal (Betticher et al., 1995).

Cyclin D1 levels can also be elevated by UTR-APA. Wiestner et al. (2007) investigated cyclin D1 expression in positive mantle cell lymphoma (MCL) patients. They found that patients who have isoforms of cyclin D1a mRNA with short 3′ UTRs had a median survival shorter than patients not expressing this isoform. Sequencing revealed that these short 3′ UTR-containing isoforms all contained mutated polyadenylation signals that would be predicted to strengthen the poly(A) signal (e.g., AAUAAUCAA → AAUAAA, 3 base pair deletion; AAUAAU → AAUAAAU, an A insertion). Since full-length cyclin D1a mRNA contains mRNA destabilizing elements, the truncated mRNAs will be more stable. In the same context, Mayr and Bartel (2009) showed that cancer cell lines preferentially expressed shorter 3′ UTR mRNAs of some oncogenes, including cyclin D1, whose shorter 3′ UTR mRNA isoform was produced by usage of a proximal poly(A) signal (AAGAAA) and not by mutations that create a new proximal poly(A) signal, as in the case of MCL. Although the mechanisms of generating cyclin D1 isoforms with shorter 3′ UTRs are different in the two cases described here, the final outcome, increased expression of cyclin D1, is the same.

Mechanisms Regulating APA

As discussed above, variations in the levels or activity of core polyadenylation factors can determine APA patterns. Another mechanism for regulating APA involves gene/tissue-specific RNA binding proteins. Interestingly, this is in many ways analogous to the control of alternative splicing (Chen and Manley, 2009), and as with splicing, it is probably the combined effect of multiple *trans*-acting factors that determines the probability of using each poly(A) site (Figure 4A). Additionally, and again analogous to splicing, regulation likely involves *cis*-acting elements not only on the nascent mRNA but also at the DNA/chromatin level (Figures 4B and 4C). Here we discuss what is known about the regulation of APA and speculate about additional possible mechanisms.

Regulation of APA by trans-Acting Factors

One way to regulate the choice of alternative poly(A) sites is by differential expression of general polyadenylation factors. This mechanism could, for example, promote the usage of an APA site that inefficiently recruits the 3′ processing machinery due to the presence of suboptimal *cis*-acting elements by increasing the concentration of one or more limiting processing factor. A well-known example of this model of action occurs during B cell differentiation. As discussed above, upregulation of CstF64, and indeed the CstF complex, results in a switch

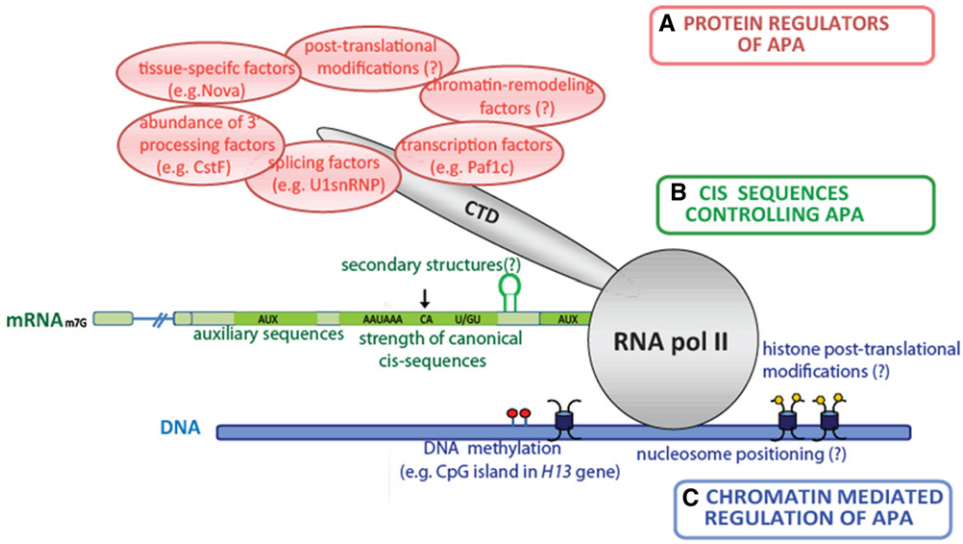


Figure 4. Mechanisms Regulating APA

(A–C) The choice of using one poly(A) site over another is dictated by a combination of several features, including variations in the abundance or activity of *trans*-acting factors such as core 3′ processing proteins and tissue-specific RNA-binding proteins, as well as through interaction with splicing and transcription factors (A), and combinations of *cis*-acting RNA elements, such as the strength of binding sites for core 3′ processing factors, auxiliary sequences, and/or new motifs directing the interaction of protein components with the mRNA and perhaps RNA secondary structures (B). APA is likely also influenced by chromatin, including nucleosome positioning around the poly(A) site, DNA methylation, and histone posttranslational modifications (C).

from distal to proximal poly(A) site selection, resulting in conversion of IgM heavy chain from membrane-bound to secreted form (Takagaki and Manley, 1998; Takagaki et al., 1996). This was shown to reflect a greater affinity of the purified CstF complex for the distal GU-rich downstream element relative to the corresponding promoter-proximal site, leading to a model in which the stronger, high-affinity site is utilized under conditions of limiting CstF, while at high concentrations of CstF, the first site encountered during transcription, i.e, the proximal site, is preferentially used. This model not only explains the switch in IgM pre-mRNA APA during B cell activation, but also provides a mechanistic explanation for the more recent global observations, also discussed above, that promoter proximal APA sites are frequently "weaker" than downstream sites. Thus, the switch to proximal sites that occurs generally in proliferating cells could be brought about by increased levels of CstF or other processing factors, which as we discussed is indeed frequently observed. It is also noteworthy that the global studies revealed that variations in the AAUAAA sequence frequently characterized the proximal sites, but the studies with IgM indicate that the nature of the GU-rich sequence can also influence APA.

Another example of regulation of APA by CstF is provided by control of the mRNA encoding the transcription factor NF-ATc during T cell activation. The transcription factor NF-ATc can be synthesized in three prominent isoforms, two long forms that are expressed in naive T cells and a shorter form arising from usage of a proximal poly(A) site during differentiation to effector T cells. Analogous to the situation in B cells, CstF64 levels are low in naive T cells when the distal poly(A) site is used, but increase during T cell activation when APA switches to the proximal site. Again, this switch appears to exploit the relatively low affinity of the proximal site for CstF64 (Chuvpilo et al., 1999).

The 3′ processing factor CFI has also been shown capable of influencing APA site choice, at least in human tissue culture cells. As opposed to the examples of CstF-mediated regulation of APA, where low protein levels promote the usage of the distal poly(A) site, reduced levels of CFI-25, achieved by siRNA knock down, resulted in an upstream shift in poly(A) site selection in transcripts of several genes tested (Kubo et al., 2006). These results suggest that CFI may be selectively recruited to the distal poly(A) site, perhaps by sequence-specific RNA binding. Indeed, previous studies have shown that CFI preferentially binds to RNA sequences containing UGUAN (Brown and Gilmartin, 2003; Venkataraman et al., 2005). Additional work is required to determine if alterations in CFI levels is a physiological mechanism of APA control.

An important question in considering the role of general poly(A) factors in APA regulation is whether the levels of any of these factors change in a systematic way in response to changes in proliferation and/or during differentiation. Indeed, genome-wide studies have found that expression of most 3′ processing factors does change in ways that correlate with APA changes. For example, expression of most of the core polyadenylation factors, including CstF and CPSF subunits, RBBP6 (Shi et al., 2009), and symplekin, was found to be upregulated during generation of iPS cells derived from different cell types, correlating with a general trend of 3′ UTR shortening, while the same factors were downregulated in differentiated embryonic tissues where longer 3′ UTRs are observed (Ji and Tian, 2009). In agreement with this, many of these same genes are downregulated during differentiation of C2C12 myoblasts into myotubes, when 3′ UTRs are lengthened (Ji et al., 2009). Moreover, Mayr and

Bartel (2009) found that genes encoding several 3′ processing factors were upregulated in cancer cells, correlating with shorter 3′ UTRs (Figure 2). The most striking difference was in levels of CstF64 and CPSF160, which as mentioned directly recognize the GU-rich region and AAUAAA hexanucleotide, respectively. These findings together support the view that changes in concentrations of core poly(A) factors in conjunction with relatively weak proximal poly(A) sites indeed plays an important, general role in controlling APA. However, as we discussed above with respect to the IgM gene, it is likely that other factors can contribute to APA control, perhaps providing redundancy, functioning together with the core factors, and/or allowing more gene-specific regulation. Consistent with this, proximal APA sites displaying higher variation of usage in different human tissues tend to be flanked by sequences with higher conservation rate (Wang et al., 2008a).

A number of RNA binding proteins have been implicated in APA control. An example of a tissue-specific factor, initially characterized as a splicing factor but that also controls APA, is Nova2 (Licatalosi et al., 2008). RNAs extracted from brains of WT versus *Nova2* knockout mice were hybridized to exon arrays, and the pattern of APA was found to be altered in ~300 transcripts. A Nova2 binding site, YCAY, was identified flanking the Nova2-regulated alternative poly(A) sites; moreover, the position of Nova2 binding was found to determine whether the protein acts to promote or inhibit poly(A) site use. In transcripts where Nova2 enhances poly(A) site use, it binds to more distal elements, where it possibly antagonizes the action of (unknown) auxiliary factors. In cases where Nova2 has an inhibitory effect, binding sites are located within 30 nt of the poly(A) signal sequences, and binding therefore likely interferes with the formation of the 3′ processing complex. Therefore, the position of Nova2 binding may determine the outcome of poly(A) site selection in a manner analogous to its action on splicing regulation (Ule et al., 2006). Another example of a ''splicing factor'' that can regulate polyadenylation is the polypyrimidine tract binding protein, PTB. PTB can compete with CstF binding to the downstream sequence element (Castelo-Branco et al., 2004) or can stimulate 3′ processing (Moreira et al., 1998) by increasing the binding of heterogeneous nuclear ribonucleoprotein H (hnRNP H, like Nova2, better known as a splicing factor) to the G-rich auxiliary element, which in turn stimulates cleavage by recruiting CstF and PAP (Danckwardt et al., 2007; Millevoi et al., 2009).

Additional genome-wide analysis also implicates hnRNP H in APA regulation. Specifically, Katz et al. (2010) used a statistical model to infer isoform regulation from RNA-seq data. The results showed that upon hnRNP H knockdown, preferential use of distal poly(A) sites was observed. This effect could be due to either hnRNP H-mediated inhibition of distal poly(A) sites or by direct activation of proximal sites. The authors found that genes with higher expression of shorter 3′ UTRs in the presence of hnRNP H displayed higher binding of hnRNP H near the proximal poly(A) site, implying that the second mechanism is the one used. Since this would imply a role of hnRNP H in recruitment of 3′ processing factors, this finding is in agreement with the fact that hnRNP H has been previously shown to exert a stimulatory role by interacting with PAP (Millevoi et al., 2009). High levels of hnRNP H have been observed in certain cancers (Honoré et al., 2004), suggesting that this protein contributes to the shortening of 3′ UTRs observed in cancer cells.

Several bona fide splicing factors are also known to influence 3′ processing (reviewed in Millevoi and Vagner, 2010). For example, the splicing factor U2AF65 binds to the polypyrimidine tract at the last intron 3′ splice site, stimulating both cleavage and polyadenylation by recruiting the CFI complex to the poly(A) site (Millevoi et al., 2006). Likewise, the SF3B component of U2 snRNP and the SR-related protein SRm160 have both been reported to influence 3′ processing by interacting with the CPSF complex (Kyburz et al., 2006; McCracken et al., 2002). It is an intriguing possibility that the interplay between factors involved in splicing of the 3′-terminal exon and polyadenylation factors in the 3′ UTR, as well as the physical distance between these two protein complexes, contributes to APA. U1 snRNP has also been shown to affect poly(A) site utilization, but independent of its role in splicing (Kaida et al., 2010). When binding of U1 snRNP to 5′ splice sites was blocked using an antisense morpholino oligonucleotide (AMO), premature polyadenylation in many pre-mRNAs at cryptic poly(A) sites, frequently in introns near the start of the transcript, was detected. This effect was proved to be specific to U1 snRNP and not dependent on splicing, since splicing inhibition by using an AMO to U2 snRNP did not have the same effect. Binding of U1 snRNP in the proximity of cryptic poly(A) sites likely blocks their use by inhibiting recruitment of core 3′ processing factors to these sites. Whether this provides a mechanism to regulate APA remains to be determined.

As with other gene regulatory mechanisms, APA is likely to be modulated by cell signaling pathways. Although little is so far known about this, a potentially interesting example is provided by the mechanism that upregulates the levels of the protease thrombin under conditions of stress, which is achieved through 3′ end processing regulation (Danckwardt et al., 2011). Stress conditions, such as inflammation, activate the kinase p38 MAPK, which on the one hand phosphorylates the RNA-binding proteins FBP2 and FBP3. Once phosphorylated, FBP2/3 no longer bind to a highly conserved upstream sequence element (USE) in the thrombin mRNA. On the other hand, activation of p38 MAPK signaling also upregulates the levels of 3′ processing factors as well as of proteins involved in splicing regulation. Interestingly, USE-RNP complexes were shown to include CPSF/CstF components and splicing regulators. The data suggest that p38 MAPK activation during stress leads to dissociation of FBP2/3 from the USE so that the USE is now able to counterbalance the relatively inefficient 3′ cleavage site by recruiting 3′ processing factors, leading to polyadenylation of the thrombin pre-mRNA. This finding has important implications, as deregulated thrombin expression, leading to the pathogenesis of thrombophilia, can result from point mutations in the 3′ UTR that improve the strength of the cleavage site (Gehring et al., 2001). Although not a direct example of APA, this emphasizes the existence of potential mechanisms by which 3′ processing of a weak poly(A) site, such as typical proximal poly(A) sites, can be selectively enhanced under specific physiological conditions.

Core 3′ processing factors are also regulated by posttranslational modification (reviewed in Ryan and Bauer, 2008). The best-studied example of this to date is provided by PAP. For

example, during mitosis PAP is hyperphosphorylated by Cdc2/Cyclin B, which reduces its activity and contributes to a general repression of mRNA and protein production during mitosis (Colgan et al., 1996). PAP was also shown to be sumoylated, a modification that is important both for its nuclear localization and its stability (Vethantham et al., 2008). Examples of posttranslational modifications of core processing factors that influence APA have not yet been reported, but are likely to exist.

The process of transcription and transcription-related proteins appear capable of affecting the choice of APA site. The coupling between transcription and 3′ processing is well established (reviewed in Hirose and Manley, 2000; Perales and Bentley, 2009; Proudfoot et al., 2002). The CTD of RNAP II is necessary for efficient 3′ processing in vivo and in vitro (Hirose and Manley, 1998; McCracken et al., 1997), the CTD interacts with 3′ processing factors such as CPSF and CstF (Glover-Cutter et al., 2008; Licatalosi et al., 2002), CPSF interacts with the transcription factor TFIID (Dantonel et al., 1997), 3′ processing factors have been detected by ChIP assays at both ends of genes (Glover-Cutter et al., 2008; Rozenblatt-Rosen et al., 2009; Venkataraman et al., 2005), and the 3′ processing factor symplekin binds to and stimulates the CTD phosphatase Ssu72, which is necessary for efficient transcription-coupled polyadenylation in vitro (Xiang et al., 2010). It has also recently been shown that a transcriptional activator can enhance the efficiency of transcription-coupled 3′ processing, in a manner that requires the transcription elongation complex PAF1c (Nagaike et al., 2011). PAF1c is a multifunctional complex implicated in various aspects of transcription (Rosonina and Manley, 2005) and is known to associate with 3′ processing factors (Rozenblatt-Rosen et al., 2009).

The above findings indicate multiple mechanisms by which 3′ end formation can be coupled to transcription. An explanation for this extensive coupling is that it serves to increase the efficiency by which nascent transcripts are cleaved, by facilitating recruitment of processing factors to the site of processing. But how might this influence APA? As discussed by Nagaike et al. (2011), an attractive model is that increasing the efficiency of 3′ processing along transcribed genes will tend to favor use of proximal poly(A) sites. Given that transcriptional activators can enhance processing efficiency, use of proximal poly(A) sites has the potential to further enhance expression of activated genes by removing repressive elements from the 3′ UTR. In support of this mechanism, knockdown of a PAF1c subunit led to increased accumulation of 3′ extended transcripts of a PAF1c target gene (Rozenblatt-Rosen et al., 2009). It will be of interest to determine whether this provides a general mechanism of APA control.

An additional recent study emphasizes the potential connection between transcription elongation rate and APA. Pinto et al. (2011) found that a mutant *Drosophila* strain with a reduced elongation rate (because of a mutation in the RNAP II largest subunit) displays increased usage of proximal poly(A) sites in a number of alternatively polyadenylated transcripts, suggesting that RNAP II elongation may have an important role in poly(A) site selection. A mechanistic explanation for these findings is simply that a slower RNAP II would enable the proximal poly(A) signal to be exposed to the 3′ processing complex for a longer time before the second poly(A) site is transcribed, increasing the efficiency with which it is used. This scenario is analogous to the effect that lower transcriptional rate has on alternative splicing: a human RNAP II carrying the equivalent of the abovementioned *Drosophila* mutation, when introduced into human cells, was shown to lead to the inclusion of otherwise skipped alternative exons in several transcripts (de la Mata et al., 2003).

RNA Signals that Modulate APA

As mentioned in the Introduction, specific RNA sequences in the pre-mRNA define the binding sites for different components of the 3′ processing complex, dictating the precise site where cleavage will occur. These are usually termed the ''core'' polyadenylation elements, while there are also less-defined auxiliary downstream and upstream elements. As discussed below, the ''strength'' of the core elements in combination with auxiliary elements is likely to play a critical role in selection of APA sites.

Large-scale computational analyses of 3′ UTRs have revealed interesting features of *cis*-acting elements in regulating usage of alternative poly(A) sites. As expected, poly(A) sites containing the consensus sequence AAUAAA are used more frequently than other variants. Nonetheless, usage of variant hexamers is not uncommon (Hu et al., 2005; Jan et al., 2011; Tian et al., 2005). Importantly, these variant sequences are usually found in a promoter-proximal position within the 3′ UTR, and the ones used more often are characterized by increased sequence conservation around the poly(A) site. This suggests that appropriate context can compensate for lack of a strong poly(A) site, probably by enhanced recruitment of 3′ processing factors, such as CstF, to these sites.

Analysis of APA in 15 human tissues using deep sequencing found a set of heptanucleotides showing high conservation located in the region between APA sites (Wang et al., 2008a). These include seed matches to a number of miRNAs, as expected, but also a consensus binding motif for FOX1/FOX2 (or other proteins with the same RNA-binding specificity). FOX1/FOX2 are well-characterized tissue-specific splicing factors (reviewed in Kuroyanagi, 2009), but such a strong conservation of their binding sequence in 3′ UTR regions suggests that they have additional roles. It will be interesting to determine if such roles are connected to regulation of APA and/or to determining mRNA localization and stability.

To examine the sequence patterns governing APA, Ozsolak et al. (2010) used direct RNA sequencing to analyze RNA samples extracted from human liver, human brain, and yeast. Three new motifs were identified near human poly(A) sites: a TTTTTTTTT motif positioned ~21 nt upstream of the poly(A) site, an AAWAAA motif (where W represents either A or T) positioned upstream of the poly(A) site, and a palindromic sequence, CCAGSCTGG (S = C/G) found downstream of the poly(A) site. The palindromic sequence strongly co-occurs with the TTTTTTTTT motif and with another sequence that was later found using a less stringent scan (RGYRYRGTGG, where R = A/G and Y = C/T). These sequences are present in intragenic and newly found intergenic poly(A) sites (likely to represent novel mRNAs), whereas they do not co-occur and actually anti-correlate with the canonical AATAAA signal localization. The anticorrelation hints at a possible role for these sequences in coordinating APA events. An interesting analogy is with TATA-less

promoters, which use the same set of core transcription factors but involving different interactions with promoter sequences (Juven-Gershon et al., 2008; Sikorski and Buratowski, 2009). Another possibility is that these new motifs function by directing the binding of yet unknown proteins, which in turn affect the recruitment and formation of the 3′ processing factors. A third possibility is that, under certain conditions, the affinity of CPSF and CstF complexes to RNA sequences might be modulated by mechanisms such that posttranslational modifications or association with other factors shift their binding from the canonical sequences to these new motifs, thereby affecting APA.

Finally, although not yet documented, it is possible that secondary structures and stem-loop motifs in 3′ UTRs may affect APA. Such structures in 3′ UTRs have been shown to regulate stability and other aspects of mRNA metabolism (Erlitzki et al., 2002; Liu et al., 2010b), and it is possible that they could also enhance or inhibit the binding of protein factors involved in 3′ processing and, as a result, modulate APA.

Chromatin and Epigenetic-Mediated Regulation of APA

An important recent discovery is that chromatin structure and epigenetic marks can act as regulators of alternative splicing (Fox-Walsh and Fu, 2010; Luco et al., 2010, 2011). Recent data suggest that 3′ end processing might be similarly modulated by chromatin and histone modifications. While much attention has focused on nucleosome organization around the promoter regions, little is known about their organization at the end of genes. The first evidence for a connection between polyadenylation and histone positioning was reported in *S. cerevisiae*, where antibodies against tagged histones H3 and H4 were used to perform a ChIP-seq analysis. Significantly, the 3′ region near the poly(A) site was shown to be depleted of nucleosomes (Mavrich et al., 2008; Shivaswamy et al., 2008; Spies et al., 2009). The depletion of nucelosomes in this area could be caused by the nucleotide sequence itself, which might have lower intrinsic affinity for nucleosomes (as shown for poly(dA:dT) streches), or by the possibility that a nucleosome-excluding DNA-binding protein associates near the poly(A) site.

Sites of mRNA polyadenylation and transcription termination by RNAP II are closely spaced in yeast genes (reviewed in Richard and Manley, 2009), so it could be that the nucleosome-free regions are related to transcription and not 3′ processing. More recently, however, a confirmation of strong nucleosome depletion around human poly(A) sites was obtained, suggesting that these regions are indeed connected to 3′processing. Spies et al. (2009) analyzed two previously published ChIP-Seq data sets from human T cells (Barski et al., 2007; Schones et al., 2008) and found that the dip in nucleosome density observed at the AATAAA sequence (and variants) was even more pronounced around actively used poly(A) sites (in genes with multiple poly(A) sites), suggesting either that additional sequences around the poly(A) signal, such as T-rich stretches, may play a role in nucleosome positioning or that a yet unknown nucleosome-excluding DNA binding protein maybe be commonly bound near the poly(A) sequence. Moreover, higher downstream nucleosome density, from approximately +75 to +375 downstream of the poly(A) signal, was observed to be associated with higher poly(A) site usage. Whether nucleosome positioning affects APA, for example, by influencing the rate of polymerase elongation, or if the opposite is true, via a 3′ complex-dependent recruitment of a chromatin remodeling factor, remains to be clarified.

Genomic imprinting has also been implicated in APA regulation. Alternative poly(A) sites on transcripts of the mouse imprinted gene *H13* (encoding for a signal peptide peptidase) have been found to be utilized in an allele-specific manner, such that two proximal poly(A) sites are used in the maternal allele, while a distal poly(A) site (one of three distal sites) is preferentially used in the paternal derived allele (Wood et al., 2008). The two clusters of poly(A) signals are separated by a CpG island, which is located 0.5–3 kb downstream of the first cluster and ~20 kb upstream of the second cluster of poly(A) sites. This CpG island has been shown to be specifically methylated only on the maternal allele. Since the maternal and paternal alleles are exposed to the same array of *trans*-regulatory factors, allelic differences in APA of this imprinted locus must be the result of epigenetic regulation. A possible explanation for this is that methylation of the CpG island on maternally derived alleles recruits an inhibitory factor that prevents binding of polyadenylation factors to the upstream poly(A) sites, and therefore the distal poly(A) site is used. CpG binding proteins have been shown to be able to indirectly change chromatin structure. For example the protein CFP1 binds specifically to nonmethylated CpGs and changes chromatin by recruiting a methyltransferase, which leads to increased H3K4me3 (Thomson et al., 2010). Similarly, changes in chromatin induced by the specific state of CpG methylation, if it occurs proximal to poly(A) sites, could affect their utilization.

While additional work is required, it seems that nucleosome positioning and epigenetic marks can affect the outcome of gene expression through regulation of APA. The precise mechanisms involved are not yet known but in theory could influence APA either indirectly, for example by influencing the transcription rate and therefore allowing more time for the assembly of the 3′ complex, or directly, by facilitating recruitment of components or modulators of the 3′ processing machinery. The latter case would be analogous to the mechanism by which recognition of H3K4me3 by CHD1 functions, at the 5′ ends of actively transcribed genes, to recruit core spliceosomal components, therefore facilitating the efficiency of pre-mRNA splicing (Sims et al., 2007). However, at this point it is difficult to establish a cause or effect relationship between APA and epigenetic marks (as well as nucleosome positioning). It is possible that poly(A) site selection may induce specific chromatin marks, perhaps through 3′ processing complex-dependent recruitment of chromatin modifiers, rather than chromatin marks acting to promote particular APA patterns.

Biological Functions of APA

UTR-APA produces mRNA isoforms that either contain or lack a full complement of *cis*-regulatory elements (e.g., AREs or miRNA binding sites), depending on the choice of proximal versus distal poly(A) sites. Thus, the landscape of such sequences throughout 3′ UTRs can determine the robustness of APA as a regulatory mechanism. In this regard, Legendre

et al.(2006) carried out a systematic examination of 3′ UTRs produced by APA and found that 52% of miRNA target sites are located downstream of the first poly(A) site. Sandberg et al. (2008) also found that in T cells mRNAs with longer 3′ UTRs have a 2.1-fold higher number of miRNA target sites than those with shorter 3′ UTRs. AREs have been estimated to be present in ~10%–15% of all transcripts (Halees et al., 2008) and were shown to interact with several proteins, some of which contribute to mRNA stability (reviewed in Barreau et al., 2005), and others control translation (reviewed in Espel, 2005). In addition, cooperation between miRNAs and ARE binding proteins has been documented in ARE-mediated mRNA degradation (Jing et al., 2005).

As previously discussed, states of increased cell proliferation are associated with generation of transcripts having shorter 3′ UTRs. This results in increased gene expression, consistent with the need of faster proliferating cells to produce more proteins. Indeed, Sandberg et al. (2008) showed that luciferase reporters with short 3′ UTRs from several genes produced about twice as much luciferase than those with longer 3′ UTRs. For example, one of the tested genes, *Hip2*, contains conserved binding sites for miR-21 and miR-155. Expression of the longer *Hip2* 3′ UTR isoform is decreased during T cell activation, while protein levels of Hip2 are increased. Mutation of these sites resulted in the same luciferase levels as the reporter with the shorter *Hip2* 3′ UTR produced. Likewise, Mayr and Bartel (2009) also showed that the longer 3′ UTRs of *IMP-1*, *Cyclin D2*, or *DICER1* genes negatively affected expression of similar luciferase constructs and that this could be partially reversed by specific deletions of miRNA sites (let-7 in *IMP-1*, miR103/107, and/or let7 in *DICER1* and miR15/16 in *Cyclin D2*). Significantly, some miRNAs, including let-7 and miR15/16, have been reported to act as tumor suppressors (Calin et al., 2002; Yu et al., 2007). Extending this notion, escape from miRNA-mediated regulation can induce increased oncogene protein synthesis, suggesting that loss of 3′ UTR regulatory elements by APA contributes to oncogenic transformation. Notably, deletions of miRNA binding sites within the full-length 3′ UTRs caused only a quarter to two-thirds increase in protein levels compared to levels observed with the shortened 3′ UTRs produced by APA (Mayr and Bartel, 2009). Thus, other regulatory factors such as RNA binding proteins likely influence this process.

Another mechanism by which UTR-APA can influence protein expression is via regulating mRNA localization. Since localization is mainly dictated by *cis*-elements found within the 3′ UTR (Kislauskis and Singer, 1992; Andreassi and Riccio, 2009), this process can be modulated by APA. Examples include ASH1 mRNA in budding yeast (Takizawa et al., 1997), bicoid mRNA in *Drosophila* embryos (Johnstone and Lasko, 2001), VegT1 mRNA in *Xenopus* oocytes (King et al., 2005), β-actin mRNA in human fibroblasts (Condeelis and Singer, 2005), and MBP mRNA in oligodendrocytes (Smith, 2004). Strikingly, high-resolution in situ hybridization techniques revealed that more than 70% of transcripts in *Drosophila* embryos are expressed in spatially distinct patterns (Lécuyer et al., 2007). Thus, mRNA localization is a global phenomenon, conserved from yeast to mammals. Asymmetric localization is observed in highly polarized cells like differentiated neurons where APA events are often observed and where mRNA localization is used to promote rapid local protein synthesis.

Several examples illustrate the role of APA in mRNA localization. One is the brain-derived neurotrophic factor BDNF. The brain produces two BDNF transcripts encoding the same protein, with either a short or a long 3′ UTR (Timmusk et al., 1993). The long BDNF mRNA was found to be preferentially targeted to dendrites in cultured rat neurons. In addition, a significant reduction of dendritic BDNF mRNA was observed in hippocampal and cortical neurons of mutant mice that lack the long 3′ UTR mRNA isoform due to the insertion of three strong poly(A) sites after the first BDNF poly(A) site (An et al., 2008). Furthermore, the long and the short 3′ UTRs are differently regulated in translation: while the short 3′ UTR BDNF mRNA is predominantly associated with polyribosomes, the long 3′ UTR BDNF mRNA is largely sequestered into translationally dormant ribonucleoprotein particles. After neuronal stimulation, polyribosome association with the long 3′ UTR mRNA was increased, accompanied by increased BDNF protein, although levels of BDNF mRNAs were not changed. These observations show that the long 3′ UTR mRNA specifically undergoes robust translational activation in the hippocampus before transcriptional upregulation of BDNF, while the short 3′ UTR mRNA mediates active translation to maintain basal levels (Lau et al., 2010). Another example of 3′ UTR-mediated localization is provided by the calmodulin-dependent protein kinase II, CaMKIIα. CaMKIIα mRNAs also have different-length 3′ UTRs (Bulleit et al., 1988), and again, the longer isoform specifically localizes in dendrites (Blichenberg et al., 2001), suggesting that a similar mechanism might exist as with BDNF mRNA. Indeed, a number of mRNAs localized in dendrites possess 3′ UTR sequences required for localization (Andreassi and Riccio, 2009), although in most of these, the role of APA has not been investigated. It is possible that 3′ UTR signals regulated by APA may provide a general mechanism for localizing mRNAs to soma and dendrites, as well as to other subcellular destinations.

APA also plays a role in control of gene expression in plants. For example, the control of seasonal flowering has a complex but unique gene-regulation mechanism that involves APA (Figure 3C) (Hornyik et al., 2010; Liu et al., 2010a). Flowering time is negatively regulated by expression of the *FLC* gene. Two RNA binding proteins, FPA and FCA, act independently to repress *FLC* expression and thereby allow flowering. Both FPA and FCA have been shown to repress *FLC* expression by mediating APA of a noncoding antisense transcript. A promoter situated downstream of the poly(A) site of *FLC* and on the opposite strand generates antisense transcripts that have alternative poly(A) sites: one cluster of poly(A) sites (proximal) is located opposite the terminal intron of *FLC*, and another cluster (distal) is located opposite the *FLC* promoter. Both FPA and FCA promote usage of the proximal poly(A) sites. Interestingly, mutants of CstF components (CstF64 and CstF77) showed elevation of sense *FLC* transcripts and reduction of antisense *FLC* transcripts, suggesting *FLC* antisense transcripts are sensitive to CstF activity (Liu et al., 2010a). However, FLD, a histone H3 lysine 4 (H3K4me2) demethylase, is also required for effective *FLC* silencing (Bäurle and Dean, 2008; Liu et al., 2007b, 2010a). In addition, another layer of complexity is added by the fact that

FCA interacts with FY (the homolog of the 3′ processing factor WDR33) (Shi et al., 2009) to promote proximal poly(A) site selection in its own pre-mRNA, resulting in production of a nonfunctional, truncated FCA-mRNA (Simpson et al., 2003). How does selection of the proximal poly(A) site in the antisense RNA transcript promote silencing of *FLC*? Perhaps, as suggested by Rosonina and Manley (2010), when the proximal poly(A) sites in the antisense transcript are used, the recruited FLD demethylase catalyzes removal of the transcriptionally active chromatin mark H3K4me2 in the body of the *FLC* gene, leading to *FLC* silencing, while utilization of the distal poly(A) site of the antisense transcript facilitates recruitment of positive factors to the *FLC* promoter, resulting in enhanced *FLC* mRNA expression.

Regulation by antisense APA has the potential to extend beyond plants. A genome-wide analysis of polyadenylated RNAs from both yeast and human liver cells (Ozsolak et al., 2010) demonstrated that antisense transcription is very common in eukaryotes, being present in at least ~60% of yeast-annotated open reading frames and as much as 30% in human liver. It is therefore possible that gene regulation through APA of antisense transcripts may occur also in mammalian genes in a way analogous to the *FLC* gene in yeast.

Concluding Remarks

While the existence of APA has been known for some time, whether or not it constitutes a general mechanism of gene control has until recently been unclear. In the last several years, however, genome-wide analyses have shown that APA is in fact widespread in mammalian cells, is regulated during development and differentiation, and can become deregulated in disease. One of the most interesting questions is how mechanistically alternative poly(A) sites are selected. As we have discussed, this will likely involve core polyadenylation factors, such as CstF64 in B cell activation (Takagaki et al., 1996). Indeed, since CPSF160 and CstF64 are upregulated in cancer cells (Mayr and Bartel, 2009) and several 3′ processing factors are downregulated during myoblast differentiation (Ji et al., 2009), regulation of APA by varying the levels of core factors may be a general mechanism. However, given how widespread we now know APA to be, it is likely that other factors are involved. Similar to control of alternative splicing (Chen and Manley, 2009), this is likely to reflect a combination of core processing factors and gene-specific RNA binding proteins. Indeed, APA is coupled with splicing events as well as with transcription, suggesting that numerous auxiliary factors involved in these processes may affect APA. And as we have seen, chromatin modifications also play a role in APA. In keeping with this, analysis of the poly(A) "proteome" revealed over 80 proteins (Shi et al., 2009), potentially linking polyadenylation efficiency, and hence APA regulation, with multiple cellular processes.

So far, only a few examples linking aberrant APA directly with known diseases have been documented. However, mutations in 3′ UTRs, including poly(A) signal sequences, have been associated with a number of medically relevant issues (reviewed in Danckwardt et al., 2008), including the early examples of α- and β-thalassemias (Higgs et al., 1983; Orkin et al., 1985) as well as the abovementioned examples of thrombophilia (Gehring et al., 2001), IPEX (Bennett et al., 2001), and Cyclin D1-related cancers (Burd et al., 2006; Wang et al., 2008b; Wiestner et al., 2007). It is therefore likely that, similar to diseases reflecting aberrant splicing (Cooper et al., 2009), more examples of diseases caused by changes in APA will emerge.

ACKNOWLEDGMENTS

We would like to thank Emanuel Rosonina and Tristan Coady for critical reading of the manuscript. Work from the authors' lab was supported by NIH grant R01 GM28983.

REFERENCES

Alt, F.W., Bothwell, A.L., Knapp, M., Siden, E., Mather, E., Koshland, M., and Baltimore, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3′ ends. Cell *20*, 293–301.

An, J.J., Gharami, K., Liao, G.Y., Woo, N.H., Lau, A.G., Vanevski, F., Torre, E.R., Jones, K.R., Feng, Y., Lu, B., and Xu, B. (2008). Distinct role of long 3′ UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. Cell *134*, 175–187.

Andreassi, C., and Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3′UTR ends. Trends Cell Biol. *19*, 465–474.

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

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

Barreau, C., Paillard, L., and Osborne, H.B. (2005). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res. *33*, 7138–7150.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bäurle, I., and Dean, C. (2008). Differential interactions of the autonomous pathway RRM proteins and chromatin regulators in the silencing of Arabidopsis targets. PLoS ONE *3*, e2733.

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.

Bennett, C.L., Brunkow, M.E., Ramsdell, F., O'Briant, K.C., Zhu, Q., Fuleihan, R.L., Shigeoka, A.O., Ochs, H.D., and Chance, P.F. (2001). A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA—>AAUGAA) leads to the IPEX syndrome. Immunogenetics *53*, 435–439.

Betticher, D.C., Thatcher, N., Altermatt, H.J., Hoban, P., Ryder, W.D., and Heighway, J. (1995). Alternate splicing produces a novel cyclin D1 transcript. Oncogene *11*, 1005–1011.

Blichenberg, A., Rehbein, M., Müller, R., Garner, C.C., Richter, D., and Kindler, S. (2001). Identification of a cis-acting dendritic targeting element in the mRNA encoding the alpha subunit of Ca2+/calmodulin-dependent protein kinase II. Eur. J. Neurosci. *13*, 1881–1888.

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.

Brown, G., Drayson, M.T., Durham, J., Toellner, K.M., Hughes, P.J., Choudhry, M.A., Taylor, D.R., Bird, R., and Michell, R.H. (2002). HL60 cells halted in G1 or S phase differentiate normally. Exp. Cell Res. *281*, 28–38.

Bulleit, R.F., Bennett, M.K., Molloy, S.S., Hurley, J.B., and Kennedy, M.B. (1988). Conserved and variable regions in the subunits of brain type II Ca2+/calmodulin-dependent protein kinase. Neuron *1*, 63–72.

Burd, C.J., Petre, C.E., Morey, L.M., Wang, Y., Revelo, M.P., Haiman, C.A., Lu, S., Fenoglio-Preiser, C.M., Li, J., Knudsen, E.S., et al. (2006). Cyclin D1b

variant influences prostate cancer growth through aberrant androgen receptor regulation. Proc. Natl. Acad. Sci. USA *103*, 2190–2195.

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., et al. (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc. Natl. Acad. Sci. USA *99*, 15524–15529.

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.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat. Rev. Mol. Cell Biol. *10*, 741–754.

Chen, J.M., Férec, C., and Cooper, D.N. (2006). 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.

Chuvpilo, S., Zimmer, M., Kerstan, A., Glöckner, J., Avots, A., Escher, C., Fischer, C., Inashkina, I., Jankevics, E., Berberich-Siebelt, F., et al. (1999). Alternative polyadenylation events contribute to the induction of NF-ATc in effector T cells. Immunity *10*, 261–269.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev. *11*, 2755–2766.

Colgan, D.F., Murthy, K.G., Prives, C., and Manley, J.L. (1996). Cell-cycle related regulation of poly(A) polymerase by phosphorylation. Nature *384*, 282–285.

Comstock, C.E., Augello, M.A., Benito, R.P., Karch, J., Tran, T.H., Utama, F.E., Tindall, E.A., Wang, Y., Burd, C.J., Groh, E.M., et al. (2009). Cyclin D1 splice variants: polymorphism, risk, and isoform-specific regulation in prostate cancer. Clin. Cancer Res. *15*, 5338–5349.

Condeelis, J., and Singer, R.H. (2005). How and why does beta-actin mRNA target? Biol. Cell *97*, 97–110.

Conne, B., Stutz, A., and Vassalli, J.D. (2000). The 3′ untranslated region of messenger RNA: A molecular 'hotspot' for pathology? Nat. Med. *6*, 637–641.

Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. Cell *136*, 777–793.

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

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., Gantzert, A.S., Macher-Goeppinger, S., Probst, H.C., Gentzel, M., Wilm, M., Gröne, H.J., Schirmacher, P., Hentze, M.W., and Kulozik, A.E. (2011). p38 MAPK controls prothrombin expression by regulated RNA 3′ end processing. Mol. Cell *41*, 298–310.

Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. Nature *389*, 399–402.

Dass, B., Tardif, S., Park, J.Y., Tian, B., Weitlauf, H.M., Hess, R.A., Carnes, K., Griswold, M.D., Small, C.L., and Macdonald, C.C. (2007). Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility. Proc. Natl. Acad. Sci. USA *104*, 20374–20379.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol. Cell *12*, 525–532.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. Cell *20*, 313–319.

Erlitzki, R., Long, J.C., and Theil, E.C. (2002). Multiple, conserved iron-responsive elements in the 3′-untranslated region of transferrin receptor mRNA enhance binding of iron regulatory protein 2. J. Biol. Chem. *277*, 42579–42587.

Espel, E. (2005). The role of the AU-rich elements of mRNAs in controlling translation. Semin. Cell Dev. Biol. *16*, 59–67.

Fabian, M.R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. Annu. Rev. Biochem. *79*, 351–379.

Fox-Walsh, K., and Fu, X.D. (2010). Chromatin: the final frontier in splicing regulation? Dev. Cell *18*, 336–338.

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

Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat. Struct. Mol. Biol. *15*, 71–78.

Halees, A.S., El-Badrawi, R., and Khabar, K.S. (2008). ARED Organism: expansion of ARED reveals AU-rich element cluster variations between human and mouse. Nucleic Acids Res. *36* (Database issue), D137–D140.

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.

Hirose, Y., and Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93–96.

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

Hockert, K.J., Martincic, K., Mendis-Handagama, S.M., Borghesi, L.A., Milcarek, C., Dass, B., and MacDonald, C.C. (2011). Spermatogenetic but not immunological defects in mice lacking the τCstF-64 polyadenylation protein. J. Reprod. Immunol. *89*, 26–37.

Honoré, B., Baandrup, U., and Vorum, H. (2004). Heterogeneous nuclear ribonucleoproteins F and H/H' show differential expression in normal and selected cancer tissues. Exp. Cell Res. *294*, 199–209.

Hornyik, C., Terzi, L.C., and Simpson, G.G. (2010). The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. Dev. Cell *18*, 203–213.

Hu, J., Lutz, C.S., Wilusz, J., and Tian, B. (2005). Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. RNA *11*, 1485–1493.

Ibáñez-Ventoso, C., and Driscoll, M. (2009). MicroRNAs in C. elegans Aging: Molecular Insurance for Robustness? Curr. Genomics *10*, 144–153.

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.

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.

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. USA *106*, 7028–7033.

Ji, X., Kong, J., and Liebhaber, S.A. (2011). An RNA-protein complex links enhanced nuclear 3′ processing with cytoplasmic mRNA stabilization. EMBO J. *30*, 2622–2633.

Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin, S.C., Gram, H., and Han, J. (2005). Involvement of microRNA in AU-rich element-mediated mRNA instability. Cell *120*, 623–634.

Johnstone, O., and Lasko, P. (2001). Translational regulation and RNA localization in Drosophila oocytes and embryos. Annu. Rev. Genet. *35*, 365–406.

Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., and Kadonaga, J.T. (2008). The RNA polymerase II core promoter - the gateway to transcription. Curr. Opin. Cell Biol. *20*, 253–259.

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

Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat. Methods *7*, 1009–1015.

King, M.L., Messitt, T.J., and Mowry, K.L. (2005). Putting RNAs in the right place at the right time: RNA localization in the frog oocyte. Biol. Cell *97*, 19–33.

Kislauskis, E.H., and Singer, R.H. (1992). Determinants of mRNA localization. Curr. Opin. Cell Biol. *4*, 975–978.

Knudsen, K.E., Diehl, J.A., Haiman, C.A., and Knudsen, E.S. (2006). Cyclin D1: polymorphism, aberrant splicing and cancer risk. Oncogene *25*, 1620–1628.

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.

Kuroyanagi, H. (2009). Fox-1 family of RNA-binding proteins. Cell. Mol. Life Sci. *66*, 3895–3907.

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.

Lau, A.G., Irier, H.A., Gu, J., Tian, D., Ku, L., Liu, G., Xia, M., Fritsch, B., Zheng, J.Q., Dingledine, R., et al. (2010). Distinct 3′UTRs differentially regulate activity-dependent translation of brain-derived neurotrophic factor (BDNF). Proc. Natl. Acad. Sci. USA *107*, 15945–15950.

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell *131*, 174–187.

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.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (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., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464–469.

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

Liu, F., Quesada, V., Crevillén, P., Bäurle, I., Swiezewski, S., and Dean, C. (2007b). The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. Mol. Cell *28*, 398–407.

Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2010a). Targeted 3′ processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. Science *327*, 94–97.

Liu, X., Jiang, Y., and Russell, J.E. (2010b). A potential regulatory role for mRNA secondary structures within the prothrombin 3′UTR. Thromb. Res. *126*, 130–136.

Lu, F., Gladden, A.B., and Diehl, J.A. (2003). An alternatively spliced cyclin D1 isoform, cyclin D1b, is a nuclear oncogene. Cancer Res. *63*, 7056–7061.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. Science *327*, 996–1000.

Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. Cell *144*, 16–26.

MacDonald, C.C., and Redondo, J.L. (2002). Reexamining the polyadenylation signal: were we wrong about AAUAAA? Mol. Cell. Endocrinol. *190*, 1–8.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3′-end processing. Cell. Mol. Life Sci. *65*, 1099–1122.

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

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.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. *18*, 1073–1083.

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., Patterson, S.D., Wickens, M., and Bentley, D.L. (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.

McMahon, K.W., Hirsch, B.A., and MacDonald, C.C. (2006). Differences in polyadenylation site choice between somatic and male germ cells. BMC Mol. Biol. *7*, 35.

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

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

Monarez, R.R., MacDonald, C.C., and Dass, B. (2007). Polyadenylation proteins CstF-64 and tauCstF-64 exhibit differential binding affinities for RNA polymers. Biochem. J. *401*, 651–658.

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.

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.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344–1349.

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.

Orkin, S.H., Cheng, T.C., Antonarakis, S.E., and Kazazian, H.H., Jr. (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., John, B., and Milos, P.M. (2010). Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell *143*, 1018–1029.

Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol. Cell *36*, 178–191.

Peterson, M.L. (2007). Mechanisms controlling production of membrane and secreted immunoglobulin during B cell development. Immunol. Res. *37*, 33–46.

Pinto, P.A., Henriques, T., Freitas, M.O., Martins, T., Domingues, R.G., Wyrzykowska, P.S., Coelho, P.A., Carmo, A.M., Sunkel, C.E., Proudfoot, N.J., and

Moreira, A. (2011). RNA polymerase II kinetics in polo polyadenylation signal selection. EMBO J. *30*, 2431–2444.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501–512.

Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. Genes Dev. *23*, 1247–1269.

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., and Wall, R. (1980). Two mRNAs with different 3′ ends encode membrane-bound and secreted forms of immunoglobulin mu chain. Cell *20*, 303–312.

Rosonina, E., and Manley, J.L. (2005). From transcription to mRNA: PAF provides a new path. Mol. Cell *20*, 167–168.

Rosonina, E., and Manley, J.L. (2010). Alternative polyadenylation blooms. Dev. Cell *18*, 172–174.

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

Ryan, K., and Bauer, D.L. (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.

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.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887–898.

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., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3′ processing complex. Mol. Cell *33*, 365–376.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol. *6*, e65.

Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. Curr. Opin. Cell Biol. *21*, 344–351.

Simpson, G.G., Dijkwel, P.P., Quesada, V., Henderson, I., and Dean, C. (2003). FY is an RNA 3′ end-processing factor that interacts with FCA to control the Arabidopsis floral transition. Cell *113*, 777–787.

Sims, R.J., 3rd, Millhouse, S., Chen, C.F., Lewis, B.A., Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol. Cell *28*, 665–676.

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

Smith, R. (2004). Moving molecules: mRNA trafficking in Mammalian oligodendrocytes and neurons. Neuroscientist *10*, 495–500.

Solomon, D.A., Wang, Y., Fox, S.R., Lambeck, T.C., Giesting, S., Lan, Z., Senderowicz, A.M., Conti, C.J., and Knudsen, E.S. (2003). Cyclin D1 splice variants. Differential effects on localization, RB phosphorylation, and cellular transformation. J. Biol. Chem. *278*, 30339–30347.

Sood, P., Krek, A., Zavolan, M., Macino, G., and Rajewsky, N. (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. Proc. Natl. Acad. Sci. USA *103*, 2746–2751.

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.

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.

Takagaki, Y., Seipelt, R.L., Peterson, M.L., and Manley, J.L. (1996). The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell *87*, 941–952.

Takizawa, P.A., Sil, A., Swedlow, J.R., Herskowitz, I., and Vale, R.D. (1997). Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast. Nature *389*, 90–93.

Tanguay, R.L., and Gallie, D.R. (1996). Translational efficiency is regulated by the length of the 3′ untranslated region. Mol. Cell. Biol. *16*, 146–156.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082–1086.

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.

Timmusk, T., Palm, K., Metsis, M., Reintam, T., Paalme, V., Saarma, M., and Persson, H. (1993). Multiple promoters direct tissue-specific expression of the rat BDNF gene. Neuron *10*, 475–489.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J., and Darnell, R.B. (2006). An RNA map predicting Nova-dependent splicing regulation. Nature *444*, 580–586.

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.

Vethantham, V., Rao, N., and Manley, J.L. (2008). Sumoylation regulates multiple aspects of mammalian poly(A) polymerase function. Genes Dev. *22*, 499–511.

Vinciguerra, P., and Stutz, F. (2004). mRNA export: an assembly line from genes to nuclear pores. Curr. Opin. Cell Biol. *16*, 285–292.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008a). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Wang, Y., Dean, J.L., Millar, E.K., Tran, T.H., McNeil, C.M., Burd, C.J., Henshall, S.M., Utama, F.E., Witkiewicz, A., Rui, H., et al. (2008b). Cyclin D1b is aberrantly regulated in response to therapeutic challenge and promotes resistance to estrogen antagonists. Cancer Res. *68*, 5628–5638.

Wiestner, A., Tehrani, M., Chiorazzi, M., Wright, G., Gibellini, F., Nakayama, K., Liu, H., Rosenwald, A., Muller-Hermelink, H.K., Ott, G., 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.

Wood, A.J., Schulz, R., Woodfine, K., Koltowska, K., Beechey, C.V., Peters, J., Bourc'his, D., and Oakey, R.J. (2008). Regulation of alternative polyadenylation by genomic imprinting. Genes Dev. *22*, 1141–1146.

Xiang, K., Nagaike, T., Xiang, S., Kilic, T., Beh, M.M., Manley, J.L., and Tong, L. (2010). Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. Nature *467*, 729–733.

Yu, M., Sha, H., Gao, Y., Zeng, H., Zhu, M., and Gao, X. (2006). Alternative 3′ UTR polyadenylation of Bzw1 transcripts display differential translation efficiency and tissue-specific expression. Biochem. Biophys. Res. Commun. *345*, 479–485.

Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J., and Song, E. (2007). let-7 regulates self renewal and tumorigenicity of breast cancer cells. Cell *131*, 1109–1123.

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

Zhang, X., Virtanen, A., and Kleiman, F.E. (2010). To polyadenylate or to deadenylate: that is the question. Cell Cycle *9*, 4437–4449.

Zhao, J., Hyman, L., and 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*, 405–445.

CHAPTER 3

PARP1 represses PAP and inhibits polyadenylation during heat shock

This chapter was previously published as a paper in *Molecular Cell* (2013, 49:7-17)

Molecular Cell
Article

PARP1 Represses PAP and Inhibits Polyadenylation during Heat Shock

Dafne Campigli Di Giammartino,[1] Yongsheng Shi,[1,2] and James L. Manley[1,*]
[1]Department of Biological Sciences, Columbia University, New York, NY 10027, USA
[2]Present address: Department of Microbiology and Molecular Genetics, University of California, Irvine, Irvine, CA 92697, USA
*Correspondence: jlm2@columba.edu
http://dx.doi.org/10.1016/j.molcel.2012.11.005

SUMMARY

The 3′ ends of most eukaryotic mRNAs are produced by an endonucleolytic cleavage followed by synthesis of a poly(A) tail. Poly(A) polymerase (PAP), the enzyme that catalyzes the formation of the tail, is subject to tight regulation involving several post-translational modifications. Here we show that the enzyme poly(ADP-ribose) polymerase 1 (PARP1) modifies PAP and regulates its activity both in vitro and in vivo. PARP1 binds to and modifies PAP by poly(ADP-ribosyl)ation (PARylation) in vitro, which inhibits PAP activity. In vivo we show that PAP is PARylated during heat shock, leading to inhibition of polyadenylation in a PARP1-dependent manner. The observed inhibition reflects reduced RNA binding affinity of PARylated PAP in vitro and decreased PAP association with non-heat shock protein-encoding genes in vivo. Our results provide direct evidence that PARylation can control processing of mRNA precursors, and also identify PARP1 as a regulator of polyadenylation during thermal stress.

INTRODUCTION

Eukaryotic mRNA precursors undergo several processing events before the mature mRNA is transported out of the nucleus and translated into protein. Transcription, capping, splicing, and polyadenylation are all complex reactions that require numerous protein factors, and which we now know are all interconnected (reviewed in Hirose and Manley [2000]; Moore and Proudfoot [2009]). Consistent with this, polyadenylation contributes to many aspects of mRNA metabolism, including transcription termination by RNA polymerase II (RNAP II), mRNA stability, mRNA export to the cytoplasm, and the efficiency of translation (reviewed in Millevoi and Vagner [2010]; Richard and Manley [2009]). The importance of polyadenylation is emphasized by the growing appreciation of its role in gene control and by the association of a number of human diseases with aberrant polyadenylation (reviewed in Danckwardt et al. [2008]; Di Giammartino et al. [2011]).

Polyadenylation consists of two reactions: an endonucleolytic cleavage followed by synthesis of the poly(A) tail onto the 5′ cleaved product. While these two reactions are catalyzed by the action of only two enzymes (CPSF73 and PAP, respectively), they are supported by a large number of protein factors, reflecting the necessity of finely regulating the process and coordinating it with other nuclear events. A proteomic analysis revealed the complexity of the molecular apparatus responsible for the generation of mature mRNA 3′ ends, identifying ~80 proteins that are associated with the 3′ processing complex (Shi et al., 2009). These comprise several new core 3′ processing factors as well as other proteins that may mediate crosstalk between pre-mRNA maturation and other cellular events.

Among the factors detected by Shi et al. (2009) that had not been previously identified in the mammalian 3′ processing complex was PARP1. PARP1 is an abundant nuclear enzyme that has been implicated in the DNA damage detection and repair pathway and in regulation of gene expression, especially through chromatin modification and transcription regulation (Ji and Tulin, 2010; Krishnakumar and Kraus, 2010; Rouleau et al., 2010). PARP1 is responsible for initiation, elongation, and branching of ADP-ribose units from donor NAD^+ molecules onto target proteins, leading to the posttranslational modification known as PARylation. Although PARP1 is the major target of its own activity, through an automodification reaction, a number of other covalently PARylated proteins have been described, including histones, chromatin remodeling proteins, and transcription factors. PARylation influences the activity of target proteins by modulating their protein-nucleic acid interactions, enzymatic activity, protein-protein interactions, and/or subcellular localization.

PARP1 is known to be activated by a variety of stresses (reviewed by Luo and Kraus [2012]). These include exposure to reactive oxygen, alkylating agent, ionizing radiation, and heat shock. In *Drosophila*, PARP1 has been shown to be potently activated upon heat shock and is crucial for the formation of heat shock-induced puffs on heat shock protein (*hsp*) genes (Tulin and Spradling, 2003). In humans, PARP1 is involved in regulation of the highly inducible *hsp70* gene (Ouararhni et al., 2006). Many of the proteins implicated in 3′ processing are subject to posttranslational modifications (reviewed in Ryan and Bauer [2008]). PAP, in particular, is subject to phosphorylation, which has been shown to inhibit PAP activity in vitro and during M phase (Colgan et al., 1996); acetylation, which disrupts its nuclear localization and association with the 3′ processing complex (Shimazu et al., 2007); and sumoylation, which is important for PAP's nuclear localization and stability and can also downregulate its activity (Vethantham et al., 2008).

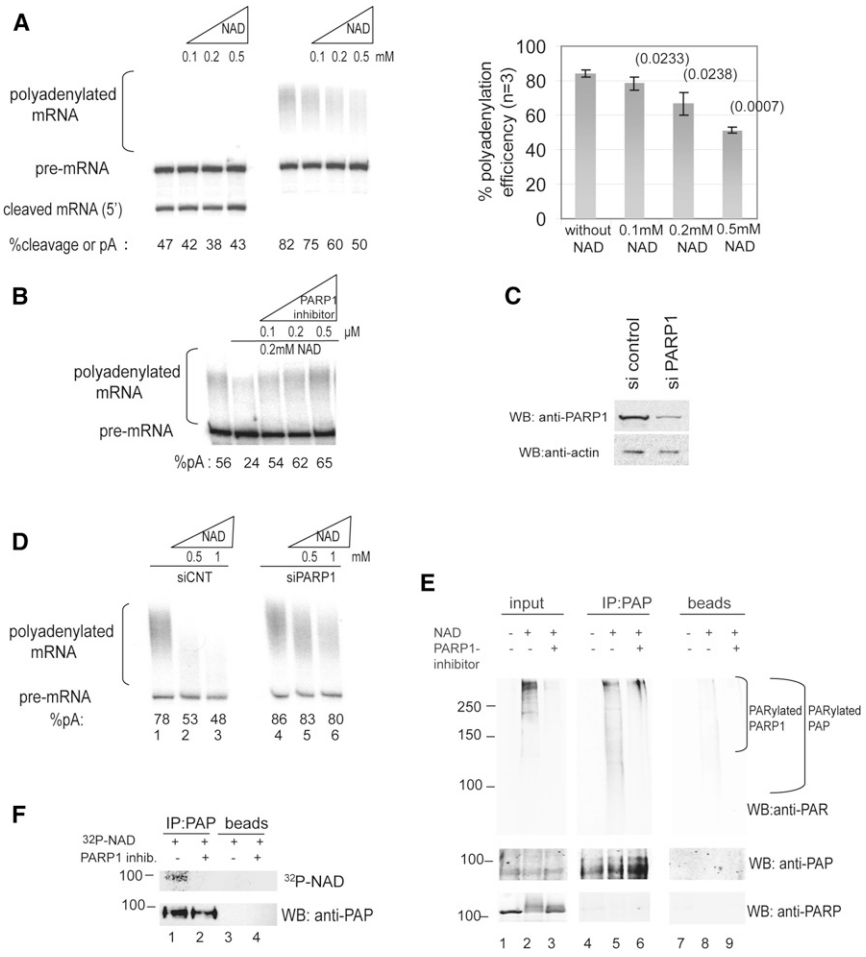


Figure 1. Activation of PARP1 by NAD$^+$ Inhibits Polyadenylation In Vitro and Induces PARylation of PAP in NE

(A) 3′ cleavage and polyadenylation assays were carried out using internally ^{32}P labeled RNA substrate and HeLa NE in the presence of the indicated concentrations of NAD$^+$. RNAs were purified, resolved by denaturing PAGE, and visualized by autoradiography. Positions of precursor and products are indicated. Percentage of cleavage/polyadenylation efficiency is indicated at the bottom and was calculated by dividing pA signal by total signal (which equals pA plus pre-mRNA). The graph (right) represents the mean percentage of three experiments, error bars represent standard deviation, and p values are indicated in parentheses.

(B) Polyadenylation assay in the presence of the indicated amounts of NAD$^+$ and PARP1inhibitor XI. Polyadenylation efficiency is indicated at the bottom.

(C) HeLa cells were transfected with a nontargeting siRNA or siRNA against PARP1. Western blots (WB) of NE were carried out with the indicated antibodies.

(D) NEs were made from the cells transfected with a nontargeting siRNA (lanes 1–3) or siRNA against PARP1 (lanes 4–6) and subsequently used in polyadenylation assays with increasing amounts of NAD$^+$. Polyadenylation efficiency is indicated at the bottom.

(E) PAP was immunoprecipitated from NE, either without incubation with NAD$^+$ (lane 4), with incubation with 0.5 mM NAD$^+$ (lane 5), or after incubation with both NAD and 0.5 μM of PARP inhibitor XI (lane 6). Western blots were carried out using the indicated antibodies. Lanes 1–3, input samples corresponding to lanes 4–6. Lanes 7–9, samples incubated with Sepharose A beads without antibody. MWs (in KDa) are indicated on the left.

(F) PAP was immunoprecipitated from NE with (lanes 2 and 4) or without (lanes 1 and 3) PARP inhibitor XI, in the presence of 0.4 uM ^{32}P-NAD$^+$. The membrane was exposed to a phosphorscreen (upper panel) followed by western blot with an anti-PAP antibody (lower panel).

Here we present evidence that PAP is a direct, physiologically significant PARP1 target. We show that PARP1 can bind and PARylate PAP in vitro, and that this inhibits PAP polyadenylation activity. Furthermore, we show that PAP is PARylated by PARP1 in vivo in response to heat shock, and that this is responsible for an observed inhibition of polyadenylation that occurs during heat shock. Finally, our data indicate that the mechanism of inhibition relies on decreased RNA binding by PARylated PAP. In vivo this is reflected by dissociation of PAP from transcribed genes upon heat shock, although, interestingly, *hsp* genes are resistant to this inhibition. Our data thus provide evidence that PARP1-mediated PARylation functions directly in control of pre-mRNA processing, and also define PARylation as a regulator of PAP and mRNA 3′ formation during heat shock.

RESULTS

Activation of PARP1 Inhibits Polyadenylation In Vitro

Our discovery that PARP1 can associate with the polyadenylation machinery raised the question of whether PARP1 might be able to regulate 3′ processing. We first wished to determine if activation of PARP1 affects 3′ processing in vitro. To this end, we added increasing amounts of NAD$^+$ to HeLa nuclear extracts (NEs) to activate endogenous PARP1 (Ogata et al., 1981). Extracts were then incubated with a ^{32}P-labeled polyadenylation substrate (SV40 late RNA, SVL) under conditions that allow either cleavage only or coupled cleavage and polyadenylation. Strikingly, NAD$^+$ potently inhibited polyadenylation, both the extent and length of the poly(A) tail, but not cleavage of the substrate (Figure 1A, left panel). Quantitation (right panel) indicates that addition of increasing concentrations of NAD$^+$ resulted in a dose-dependent decrease in polyadenylation such that 0.5 mM NAD$^+$ inhibited polyadenylation by ~40%. To assess whether the effect of NAD$^+$ was specific and indeed dependent on activation of PARP1, we repeated the coupled cleavage/polyadenylation assay as above, adding NAD$^+$ either alone, which again resulted in polyadenylation inhibition, or in the presence of increasing amounts (0.1–0.5 uM) of the PARP inhibitor XI, which led to restoration of polyadenylation activity (Figure 1B). Another PARP inhibitor (3-ABA) gave the same

results (see Figure S1 online). Moreover, to confirm that the NAD^+ effect was dependent on the presence of PARP1, NEs were prepared from HeLa cells treated for 72 hr with siRNA against PARP1 or a nontargeting siRNA (Figure 1C). Figure 1D shows that upon PARP1 knockdown, NAD^+ had almost no effect on polyadenylation, confirming that PARP1 was responsible for the observed inhibition. (The slight inhibitory effect of NAD^+ after knockdown of PARP1 can be attributed to the presence of low levels of PARP1 remaining following siRNA treatment; see Figure 1C). Polyadenylation in the absence of NAD^+ was not affected by PARP1 knockdown (compare lines 1 and 4 in Figure 1D), indicating that PARP1 does not have a constitutive role in 3′ processing.

Since the inhibitory effect of activated PARP1 was evident only on the second step of 3′ processing, we reasoned that PAP might be a PARP1 substrate. To test this, we immunoprecipitated PAP from NE or NE that had been incubated with 0.5 mM NAD^+ for 30 min at 30°C. Western blot of the NEs with an anti-PAR antibody detected a high molecular weight (MW) smear, reflecting PARylated proteins (mostly auto-PARylation) in NE that contained NAD^+ (Figure 1E, lanes 1–3 in upper panel), indicating that PARP1 was successfully activated by this treatment. Because of the heterogeneity in the length of the ADP-ribose chain, PARylation is typically detected as a smear starting extending upward from the modified protein (e.g., Hossain et al., 2009). Consistent with this, and suggesting that PAP was indeed PARylated in NE containing NAD^+, a smear extending upward from the position of PAP was detected in the PAP IP from NAD^+-containing NE blotted with the anti-PAR antibody (Figure 1E, lanes 4–6, upper panel). This smear did not come from PARylated PARP1 that might have coimmunoprecipitated with PAP, as PARP1 was not detected by western blotting with an anti-PARP1 antibody (lanes 4–6, lower panel). Moreover, addition of 0.5 uM PARP inhibitor XI abolished PAP modification (compare lanes 5 and 6 in Figure 1E, upper panel). These results indicate that under conditions in which polyadenylation was impaired by activation of PARP1 with NAD^+, PAP was indeed efficiently PARylated.

To provide additional evidence that PAP was PARylated in NE, we used ^{32}P-NAD^+ in the assay to enhance sensitivity and allow better visualization of modified PAP. Specifically, we used a low concentration of total NAD^+ (0.4 μM) to limit extension of PAR chains and therefore allow detection of the target protein as a discrete band rather than a smear (e.g., Lönn et al., 2010). Following IP with anti-PAP antibodies, SDS-PAGE, and transfer, the membrane was first exposed to a phosphor screen (Figure 1F, upper panel) and then subjected to western blot with anti-PAP (Figure 1F, lower panel). Significantly, a radioactive band was indeed detected at the position of PAP (lane 1, upper panel), which was not observed in the presence of inhibitor XI (lane 2) or when using only protein A Sepharose in the IP (lanes 3 and 4). These results provide strong confirmatory evidence that PAP is indeed PARylated in HeLa NE.

Purified PARP1 PARylates PAP In Vitro and Inhibits Its Intrinsic Activity

We next wished to examine more directly the effect of PARP1-catalyzed PARylation on PAP activity. We first asked whether PAP can be PARylated using recombinant proteins purified from *E. coli*. Figure S2A shows a Coomassie-stained gel of the two purified proteins, MBP-PARP1 and His-PAP. We employed an in vitro PARylation assay in which His-PAP was incubated with MBP-tagged PARP1 in the presence of $MgCl_2$, NAD^+, and sheared salmon sperm DNA. Under these conditions, PARP1 uses NAD^+ as a substrate to attach ADP-ribose units onto itself and target proteins, generating ADP-ribose chains that can be as long as 200 ADP-ribose units (de Murcia et al., 1983; Gagné et al., 2001). As expected (Ogata et al., 1981), PARP1 was a good acceptor of ADP-ribose units, resulting in automodification, as detected by anti-PAR western, when incubated under activating conditions (Figure 2A, compare lane 2 to lane 1). When PAP was included in the reaction, an additional smear starting from the MW of PAP and extending upward, representing PARylated PAP, was detected upon PARP1 activation (Figure 2A, compare lane 4 to lane 3).

To confirm and extend these findings, we performed two additional experiments. First, we repeated the above assay using increasing amounts of PAP. Importantly, this increased the signal detected by the anti-PAR antibody (Figure 2B, lanes 1–4), indicating that the PAR detected was from PARylated PAP (since the PARP concentration was kept constant). We then added increasing amounts of purified PARG, which hydrolyzes ADP-ribose units, to reaction mixtures, which caused the signal of PAR incorporation to collapse to the MW of PAP (Figure 2B, lanes 5–7), confirming that the PAR signal came from PARylated PAP. Second, we again used ^{32}P-NAD^+ as a source of NAD^+, which resulted in detection of a radioactive band at the MW of PAP (Figure 2C, upper panel, lane 4). Together, these results confirm that PAP is a PARP1 target in vitro.

We next tested whether PAP and PARP1 stably associate with each other in vitro. To this end, His-PAP was bound to nickel beads and its ability to bind MBP-PARP1 determined by incubating the two proteins at 4°C, followed by washes with high salt buffer (500 mM NaCl). Proteins were resuspended in denaturing loading buffer and resolved by SDS-PAGE. Western blot with anti-PARP1 antibodies revealed that MBP-PARP1 indeed bound to His-PAP on nickel beads (Figure 2D, lane 6), but not to nickel beads alone (Figure 2D, lane 5).

We next wished to determine the effect of PARylation on PAP intrinsic activity. To this end, we repurified PARylated PAP after in vitro PARylation and tested its activity in a nonspecific polyadenylation assay (e.g., Takagaki et al., 1988). In this assay, the presence of Mn^{2+} renders PAP independent from other 3′ processing factors, and purified PAP in the presence of ATP can therefore polyadenylate by itself essentially any RNA substrate. Instead of purifying total PAP after PARylation, which would include both modified and unmodified protein, we isolated specifically the PARylated fraction of PAP after in vitro PARylation (NAD^+ was not added in a control sample). We first used amylose beads to remove MBP-PARP1 (Figure S2B). The supernatant containing PAP was then incubated with an ADP-ribose affinity resin. The resin was washed and modified PAP eluted with free ADP-ribose (for the control, the free ADP-ribose was added directly to the supernatant). Following dialysis, the unmodified control and PARylated PAP proteins were quantified by dot blotting (Figure S2C; see the Experimental Procedures) and their

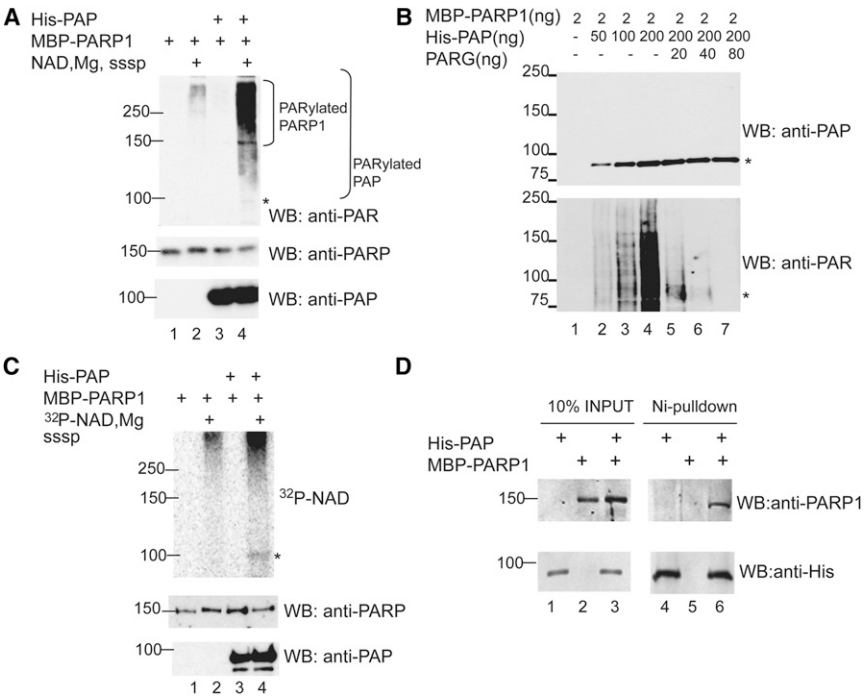


Figure 2. PARP1 Interacts with and PARylates PAP In Vitro

(A) In vitro PARylation reactions were carried out with the indicated purified recombinant proteins. Reaction mixtures in lanes 1 and 2 contained PARP1 alone; those in lanes 3 and 4 also contained PAP. In lanes 2 and 4, PARP1 was activated. Proteins were resolved by SDS-PAGE, and western blots were carried out with the indicated antibodies.

(B) In vitro PARylation as in (A) using increasing amounts of PAP (lanes 1–5) followed by incubation with increasing amounts of PARG (lanes 5–7).

(C) In vitro PARylation reactions as in (A) with the exception that 4 μM of ^{32}P-NAD$^+$ was used along with 0.05 mM NAD$^+$.

(D) Purified recombinant MBP-PARP1and His-PAP were used in "pull-down" assays using nickel beads. Following SDS-PAGE, proteins were detected by western blot with the antibodies indicated. MWs (in KDa) are indicated on the left. In (A)–(C), an asterisk indicates PAP.

concentration equalized. A dot blot also demonstrated that the recovered PAP was indeed PARylated (Figure 3A). Consistent with the results obtained with NAD$^+$-supplemented NE (Figure 1), the nonspecific polyadenylation assay revealed that PARylated PAP was essentially inactive (Figure 3B; compare lanes 2 and 3 to lanes 4 and 5). Together, our findings indicate that PARP1 directly PARylates PAP, thereby inhibiting its activity.

Heat Shock Inhibits Polyadenylation in a PAP- and PARP1-Dependent Way

We next investigated whether the link between PARP1 and polyadenylation we characterized in vitro also exists in vivo. Given that PARP1 is activated in vivo by a variety of stimuli, such as DNA damage, oxidative stress, and heat shock (Luo and Kraus, 2012), we exposed cells to several conditions known to activate endogenous PARP1, and as an initial approach tested whether any of these treatments affected the 3′ processing activity of NEs prepared from these cells. NEs were prepared from HeLa cells exposed to hydrogen peroxide, γ-IR, or heat shock and used in polyadenylation assays with a precleaved RNA substrate to examine polyadenylation uncoupled from cleavage. PARP1 was indeed activated by all of these conditions (Figure S3A), although it is possible that other PARPs may have contributed to the observed increase in PAR. While hydrogen peroxide or γ-IR treatments had no detectable effect on polyadenylation activity (Figures S3B and S3C), a drastic inhibition in activity was observed with NEs prepared from cells that had been exposed to 43°C for 1 hr (Figure 3C). Polyadenylation was strongly inhibited even with a milder heat shock, carried out at 41°C (Figure 3D). If this inhibition specifically reflected repression of PAP, then addition of purified PAP should restore activity. Figure 3E shows that addition of 5 ng of purified His-PAP restored polyadenylation, demonstrating that PAP activity was indeed impaired by heat shock.

We next investigated the effect of heat shock on polyadenylation in vivo. For this, we used a previously described HEK293 cell line that stably expresses a tetracycline-(tet) inducible β-globin transgene integrated in the genome through site-specific recombination (de Almeida et al., 2010). This system was particularly suitable for our purposes because it enabled us to measure the effect of heat shock on 3′ processing in a way that was largely independent of effects that heat shock might have on transcription. Moreover, since transcription is inducible, any effects that a short heat shock might have on polyadenylation would not be masked by polyadenylated transcripts that accumulated before heat treatment. After 3 hr of induction, β-globin mRNA expression was induced ~8-fold (Figure 4A). Cells were then incubated for 30 min at 37°C or 43°C followed by nuclear RNA extraction and reverse transcription either with random hexamer primers (to measure total mRNA) or with an oligo(dT) primer (to measure polyadenylated mRNAs). Figure 4B shows that the ratio of polyadenylated to total β-globin mRNA was indeed reduced after heat shock treatment (as measured by real-time PCR). To determine if the inhibition of polyadenylation was dependent on PARP activity, we treated cells with a cell-permeable PARP1 inhibitor (3-ABA) at the time of β-globin induction and during heat shock, and analyzed transcripts as above. Polyadenylation activity was fully restored by the PARP inhibitor (Figure 4B). Similar results were obtained with another PARP inhibitor, PJ34 (Figure S3D). Together, these results provide strong evidence that heat shock-induced inhibition of polyadenylation requires PARP activity.

To rule out the possibility that the observed decrease in polyadenylated mRNA during heat shock reflects degradation of β-globin mRNA that accumulated prior to heat shock, rather than inhibition of polyadenylation, we measured the β-globin

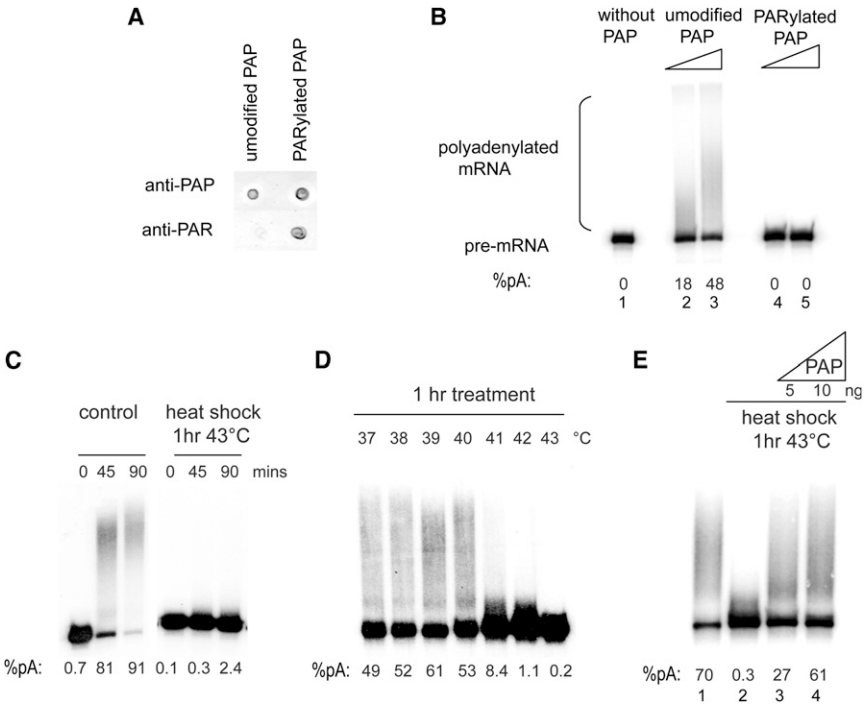


Figure 3. Heat Shock Inhibits Polyadenylation in a PAP-Dependent Manner

(A) Dot blot of 2 μl of PAP protein that was repurified after in vitro PARylation or mock PARylation as described in the Experimental Procedures. Anti-PAP or anti-PAR antibodies were used to visualize the extent of recovery and PARylation of the repurified proteins.

(B) Nonspecific polyadenylation assays with ^{32}P-labeled SVL RNA were performed in the absence of PAP (lane 1) or with increasing amounts (2 and 4 ng) of purified mock-PARylated (lanes 2 and 3) or PARylated PAP (lanes 4 and 5).

(C) In vitro polyadenylation was carried out for the indicated times using SVL RNA and HeLa NEs made from untreated cells or cells that were heat shocked for 1 hr at 43°C.

(D) In vitro polyadenylation as in (A) using NEs made from cells that were treated at the indicated temperatures for 1 hr.

(E) Polyadenylation assays as in (A) using NE from untreated cells (lane 1), or from cells that were heat shocked 1 hr at 43°C (lanes 2–4). NEs in lanes 3 and 4 were supplemented with the indicated amounts of recombinant His-PAP In (B)–(E), polyadenylation efficiencies are indicated at the bottom.

mRNA half-life following the heat treatment. For this purpose, the β-globin gene was induced for 3 hr and cells were either incubated at 37°C or 43°C for 30 min (as in Figure 4B), followed by extensive washes to remove tet from the medium and stop transcription. mRNA was extracted at the indicated time points following tet removal, and the percentage of remaining mRNA was plotted against time. As shown in Figure 4C, there was no detectable change in half-life after heat shock.

We next wished to provide additional evidence that polyadenylation is inhibited during heat shock, and that this inhibition is a general phenomenon, not specific to the β-globin gene reporter used above. To this end, we employed ^{3}H uridine labeling and oligo(dT) selection to measure newly synthesized polyadenylated mRNA during heat shock. HeLa cells were heat shocked for 30 min (^{3}H uridine added at the beginning of treatment), nuclear RNA was extracted, and polyadenylated RNA selected by oligo(dT) and quantitated by scintillation counting. The results (Figure 4D) reveal a 60% decrease in accumulation of nuclear polyadenylated RNA during heat shock. As with the β-globin gene, addition of 3-ABA restored, albeit partially, polyadenylation (Figure 4D). The 3-ABA-resistant fraction may reflect some inhibition of transcription (see the Discussion), but together our results provide strong evidence that polyadenylation is inhibited in a PARP-dependent manner during heat shock.

PARP1 PARylates PAP In Vivo during Heat Shock

We next asked whether PAP is in fact PARylated during heat shock. For this analysis, we used two different cell lines, HeLa (Figure S4A) and MCF-7 (Figure 5A). Cells were subject to heat shock and PAP immunoprecipitated from cell extracts. Using an anti-PAR antibody for western, we detected PARylated PAP in the samples from heat-shocked cells but not from control cells or from cells heat shocked in the presence of 3-ABA (compare lane 3 with lane 9, and lane 3 with lane 4 in Figure 5A). The signal did not derive from PARylated PARP1 because PARP1 did not immunoprecipitate with PAP under the conditions used (Figure 5A; see also Figure 1E). As the PAP precipitate revealed a faint signal at the PARP1 MW, which could be automodified PARP1, we treated the PAP precipitate with PARG (Figure S4B) but still did not detect PARP1 immunoprecipitating with PAP. Also, in agreement with the fact that polyadenylation was not inhibited by hydrogen peroxide-mediated activation of PARP1 (see above), PAP was not PARylated under this condition (Figure S4C).

PAP Dissociates from mRNA Transcripts during Heat Shock

We next wished to investigate the mechanism by which PARylation inhibits PAP activity. Given that PARylation can affect protein-protein interactions of modified targets (e.g., Huang et al., 2006; Kim et al., 2004), we first asked whether PARylation of PAP affects assembly of the 3′ processing complex. NEs were prepared from control cells or cells incubated 1 hr at 43°C, incubated with a ^{32}P-labeled RNA under conditions that allow formation of the 3′ processing complex, and loaded on a nondenaturing agarose gel. The results (Figure S5A) indicate that heat shock did not detectably affect the assembly or stability of the 3′ processing complex.

PARylation is also known to affect modified proteins by causing their dissociation from DNA (reviewed by Kraus, 2008). We therefore asked whether PARylation of PAP alters the enzyme's ability to bind its substrate RNA. To address this, we first used an in vitro assay. We incubated unmodified or

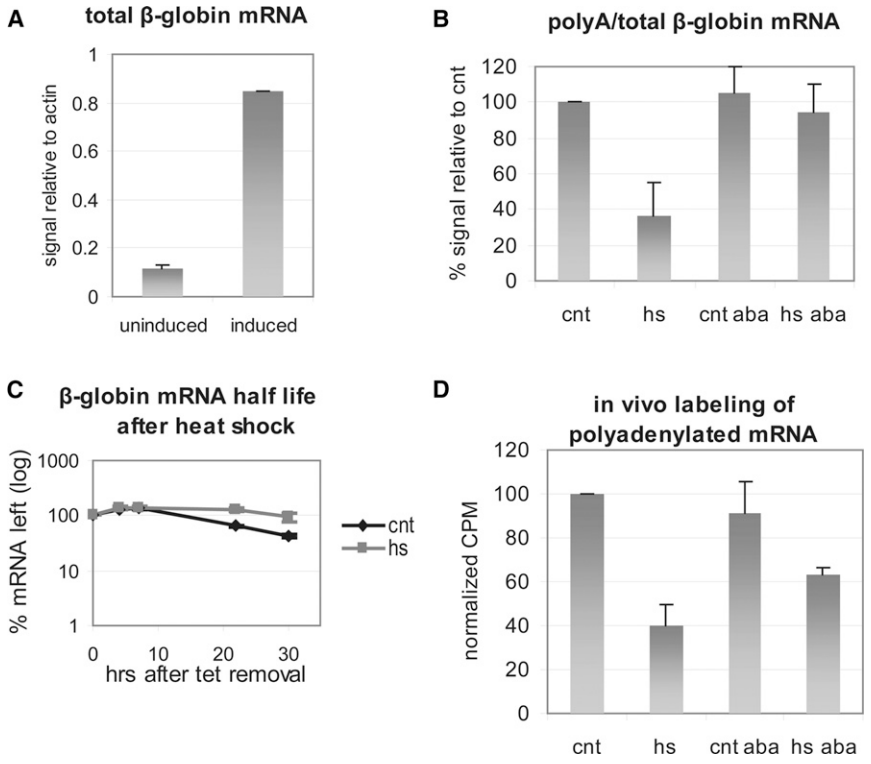


Figure 4. Heat Shock Inhibits Polyadenylation In Vivo in a PARP-Dependent Manner

(A) RNA was extracted from 293 cells stably transfected with an inducible β-globin transgene following 3 hr induction with tet or without induction. Following qPCR, the amount of induced β-globin mRNA relative to endogenous actin mRNA was calculated before and after induction.

(B) Nuclear RNA was extracted from cells induced for 3 hr as in (A) that were either kept at 37°C (cnt) or heat shocked at 43°C (hs) for 30 min. Where noted, the PARP inhibitor 3-ABA was added at the time of induction. qPCR was used to calculate the relative amount of β-globin polyadenylated mRNA compared total β-globin RNA as described in the Supplemental Experimental Procedures.

(C) The β-globin transgene was induced for 3 hr as in (A). Cells were either treated for 30 min at 43°C (hs) or left untreated (cnt). To measure β-globin half-life, RNA samples were extracted at the indicated times following tet removal.

(D) Following 30 min labeling with ^{3}H uridine, nuclear RNA was extracted from HeLa cells that were either kept at 37°C (cnt) or heat shocked at 43°C (hs) for 30 min. Where noted, the PARP inhibitor 3-ABA was added together with ^{3}H uridine. Poly (A)$^{+}$ fraction was isolated and quantitated by scintillation counting. Counts per minute (CPM) relative to cnt are shown. Results from three independent experiments are shown represented as mean and standard error.

PARylated PAP (after repurification as in Figure 3A) with SVL RNA under conditions used for polyadenylation but omitting ATP and $MgCl_2/Mn^{2+}$ to prevent poly(A) synthesis, and then loaded the samples onto a nondenaturing polyacrylamide gel (Figure 5B). Unmodified PAP bound the RNA in a concentration-dependent manner (lanes 2 and 3), while binding by PARylated PAP was greatly diminished (lanes 4 and 5), indicating that PARylation interferes with PAP binding to the RNA substrate.

We next asked whether PAP substrate binding is compromised by PARylation in vivo. For this analysis, we used the inducible β-globin cell line described above. We examined first whether PAP association with the 3′ end of the activated β-globin gene could be detected and, if so, whether it was reduced following heat shock. Chromatin immunoprecipitation (ChIP) was performed using anti-PAP antibodies, and PAP association with the 3′ end of the gene was quantified by real-time PCR (Figure 6A). In the absence of induction, a signal slightly above background was detected, likely reflecting a low level of transcription in the absence of tet (see Figure 4A and Figure 6D). When tet was added for 6 hr, PAP association with the 3′ end of the β-globin gene significantly increased, providing evidence that PAP was present at the 3′ end of the actively transcribed gene. Strikingly, PAP chromatin association was reduced rapidly, after only 5 min heat shock at 43°C. This effect was dependent on PARP1, as addition of 3-ABA during the last 4 hr of tet induction prevented the decrease in PAP crosslinking. ChIP with anti-PAP antibodies before and after heat shock was also performed on the 3′ ends of two endogenous genes, *C-MYC* and *GAPDH*, and a similar decrease in PAP crosslinking to these genes was observed (Figure S5B).

In order for cells to cope with stress, *hsp* genes must be transcribed and their transcripts processed by polyadenylation. Given the reduced association of PAP with non-*hsp* genes described above, we next asked how PAP association with the 3′ ends of three *hsp* genes—*hsp70*, *hsp90*, and *hsp27* (i.e., *HSPA1A, HSP90AA1*, and *HSPB1*)—is affected by heat shock. The same cells, conditions for ChIP and qPCR analysis, were used as in Figure 6A. Figure 6B shows that, in contrast with the results obtained with the non-*hsp* genes, association of PAP with each of the *hsp* genes did not decrease, and in fact slightly increased, after a heat shock of 5 min.

To gain a better understanding of the mechanism by which PARP1 inhibits PAP, we performed ChIP with an anti-PARP1 antibody to examine PARP1 association with the β-globin gene. The results (Figure 6C) indicate that PARP1 was also present at the 3′ end of the gene. Following heat shock, a 30% decrease in PARP1 chromatin association was observed, indicating that, similarly to PAP, although not as sharply, PARP1 was released from the 3′ end of β-globin upon activation by heat shock. To exclude the possibility that PAP and PARP1 chromatin association was reduced following heat shock because of a possible reduction in transcription, we performed a ChIP experiment using an antibody against RNAP II. Heat shock did not reduce RNAP II density at the 3′ end of the gene (Figure 6D), nor at its promoter (Figure S5C), and a slight increase (~20%) in RNAP II occupancy was indeed observed.

While we discuss below how *hsp* genes might evade PARP1-mediated inhibition of PAP recruitment and polyadenylation following heat shock, our findings together implicate PARylation of PAP as a significant aspect of the cellular response to stress.

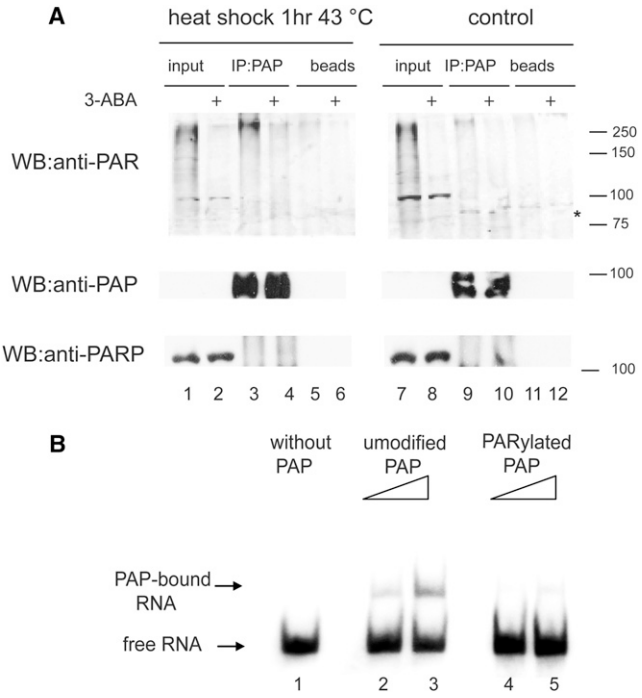


Figure 5. PARP1 PARylates PAP during Heat Shock In Vivo

(A) PAP was immunoprecipitated from untreated MCF-7 cells (lanes 7–12) or from cells that were heat shocked for 1 hr at 43°C (lanes 1–6). Where indicated, 3-ABA was added just before the heat shock (lanes 4, 6, 10, and 12). Following IP, samples were subjected to SDS-PAGE, and western blots (WB) were carried out with the indicated antibodies. The asterisk indicates the position of PAP.

(B) Gel shift assay without PAP (lane 1) or with increasing amounts of mock-PARylated (lanes 2 and 3) or PARylated (lanes 4 and 5) PAP samples that were repurified after in vitro PARylation as in Figure 2C and incubated with a ^{32}P-labeled RNA (SVL). Samples were resolved in a 5% nondenaturing polyacrylamide gel. The gel was dried and exposed to a PhosphorImager screen.

DISCUSSION

Our results have provided several insights into mechanisms of gene control. First, our findings have expanded the function of PARP1 and PARylation into the area of posttranscriptional regulation and specifically polyadenylation of mRNA precursors. Second, we have provided yet another mechanism by which PAP can be regulated, highlighting the importance of controlling PAP activity. Finally, we provide evidence that polyadenylation is inhibited following heat shock. Similar to splicing inhibition, inhibition of polyadenylation provides an additional layer of protection against the deleterious effects of heat shock, by ensuring that production of mature mRNAs is repressed so that new proteins will not be produced in a stressed environment prone to misfolding and other detrimental effects. Below we discuss the implications of our findings with respect to the role of both polyadenylation and PARylation in the regulation of gene expression.

As mentioned in the introduction, PAP is a well-known target for several posttranslational modifications. Consistent with the need to regulate PAP activity tightly under different conditions, PAP was previously reported to be phosphorylated, sumoylated, and acetylated, and now we have identified PARylation as an additional PAP modification. The reversible nature of these modifications is particularly suitable for regulating gene expression by modulating PAP, allowing the cell to respond efficiently to a changing cellular context such as during an environmental stress (e.g., PARylation during heat shock) or during the cell cycle (e.g., phosphorylation during M phase). All of the previously characterized posttranslational modifications occur in the C-terminal domain of PAP; they are in fact situated very closely and sometimes overlap (reviewed by Ryan and Bauer [2008]). PARylation, in contrast, does not seem to occur in this region, as a truncated version of PAP that lacks this domain can still be PARylated in vitro (our unpublished data). Since PARylated PAP loses its ability to bind RNA, we speculate that the most reasonable site for PARylation is the RNA-binding region itself.

Interesting questions are how and when PARP1 associates with PAP to block polyadenylation. Here we show that PARP1 colocalizes with PAP on the 3′ end of the β-globin gene. PARP1 has been shown previously to associate both with the 3′ processing machinery (Shi et al., 2009) and, similar to several components of the 3′ processing complex, with the promoters of numerous genes. Indeed, ChIP-chip experiments coupled to gene expression microarrays indicated that PARP1 binding is enriched at 90% of promoters of actively transcribed genes in human MCF-7 cells (Krishnakumar et al., 2008). In addition, and also analogous to components of the 3′ processing complex (reviewed in Hsin and Manley [2012]), PARP1 was found to interact with RNAP II (Carty and Greenleaf, 2002). This is consistent with the idea that PARP1 on promoters associates with RNAP II, which then facilitates its recruitment to the 3′ processing complex cotranscriptionally. However, it is also possible that PARP1 associates independently with promoters and then with PAP near the 3′ end of genes.

Several of the multiple factors required for mRNA 3′ end formation have previously been shown to associate with transcribed genes. These include in mammals CPSF, CstF, and CFI (Glover-Cutter et al., 2008; Rozenblatt-Rosen et al., 2009; Venkataraman et al., 2005). PAP has long been known to associate only loosely with the other core polyadenylation factors (Takagaki et al., 1988), but recently it was reported to crosslink to both 5′ and 3′ end of genes in yeast, where it is necessary for gene looping (Medler et al., 2011). These data are consistent with our results showing that PAP can be recruited to the 3′ ends of transcribed genes in human cells, implying that, even if polyadenylation occurs after release of the mRNA from RNAP II, PAP joins the 3′ processing complex cotranscriptionally. An attractive model is that PAP and PARP1 are recruited to transcribed genes together, and then upon PARP1 activation, for example by heat shock, PAP and PARP1 PARylation occurs rapidly, leading to dissociation of both from non-heat shock protein-encoding genes (see Figure 7).

Gene expression in mammalian cells is regulated at multiple levels during heat shock. From a posttranscriptional perspective, extensive studies have shown that pre-mRNA splicing (Shin et al. [2004] and references therein) and protein translation (Cuesta et al. [2000] and references therein) are inhibited following heat shock, and our data add polyadenylation to this list.

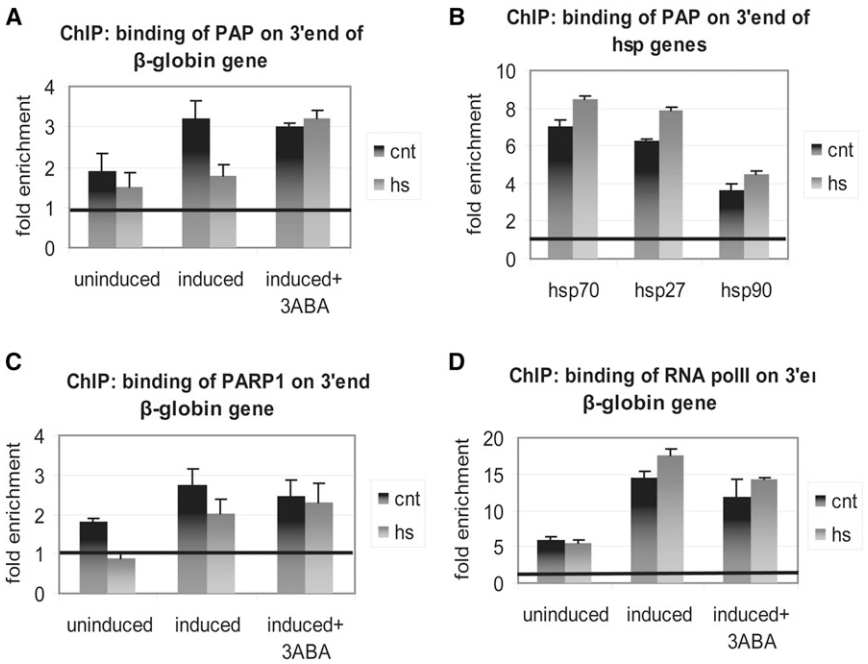


Figure 6. PARylated PAP Is Unable to Bind RNA In Vitro and Dissociates from the 3′ End of Non-*hsp* Genes during Heat Shock In Vivo

(A) ChIP was carried out using an anti-PAP antibody and amplifying a 3′ end region of the β-globin gene from the inducible 293 cell line. Cells were either uninduced or induced with tet for 6 hr, with or without 3-ABA for the last 4 hr of induction, as indicated. Cells were heat shocked for 5 min at 43°C. Results were analyzed as described in the Supplemental Experimental Procedures and quantified as fold change over background.

(B) ChIP in uninduced 293 cells using anti-PAP antibody as above. Cells were either untreated or heat shocked for 5 min at 43°C. Primers specific for the 3′ end of the indicated *hsp* genes were used for qPCR. The results were analyzed as in Figure 6A.

(C) ChIP using anti-PARP1 antibody was carried out and analyzed as in (A).

(D) ChIP using antibody against RNA polymerase II was carried out and analyzed as in (A). Results from three independent experiments are shown represented as mean and standard error.

Transcriptional regulation, however, appears to be more complex, and the effect that heat shock has on ongoing transcription in mammalian cells seems to be gene specific. Early studies showed that, following heat treatment of HeLa cells, rRNA but not mRNA transcription was inhibited (Sadis et al., 1988; Warocquier and Scherrer, 1969). Interestingly, mRNA export into the cytoplasm was found to be repressed (Sadis et al., 1988), which is consistent with a defect in 3′ processing. In addition, transcription of specific genes, e.g., *c-fos* (Andrews et al., 1987) was shown not to be affected by heat shock. Our own experiments showing a decrease in accumulation of newly synthesized polyadenylated nuclear RNA following heat shock may reflect inhibition of transcription as well as polyadenylation. The fact that 3-ABA only partially rescued the inhibition indicates that PARylation, likely of PAP, might not be the only determinant of repression, and we suggest that the 3-ABA-resistant fraction reflects inhibition of transcription of a subset of genes.

In contrast to most genes, heat shock protein-encoding genes must be expressed robustly during heat shock. A mechanism must therefore exist to ensure that PARP1 activation does not negatively affect polyadenylation of *hsp* transcripts. Consistent with this, our results showed that PAP association with *hsp* genes was not reduced, and actually increased, following heat shock. We propose two possible mechanisms that explain how PAP association with *hsp* and non-*hsp* genes is differentially regulated following heat shock. In the first, HSF1 (heat shock factor 1) plays the key role in determining specificity. HSF1 is constitutively expressed but becomes rapidly activated during heat shock by entering the nucleus and binding as a trimer to HSE (heat shock element) sequences present in heat-inducible promoters (reviewed in Shamovsky and Nudler [2008]). Notably, HSF1 has been shown to associate with 3′ processing components during heat shock (Xing et al., 2004). Therefore, in addition to its function in stimulating transcription of *hsp* genes, HSF1 may also act to enhance polyadenylation of the resulting *hsp* transcripts. Intriguingly, HSF1 has been shown to contain a PAR-binding motif (Fossati et al., 2006), and it is therefore tempting to speculate that this motif allows HSF1 to bind PARylated PAP. This might then explain our observation that PAP is not only retained on activated *hsp* genes but that its association enhanced after stress, ensuring that even if PAP is PARylated, *hsp* mRNAs will still be polyadenylated. Another possibility is that PAP PARylation does not occur on *hsp* genes. By this model, clearance of PARP1 from *hsp* promoters during heat shock (as has been shown for *hsp70.1*; Ouararhni et al., 2006) prevents the enzyme from associating with the 3′ processing complex on *hsp* transcripts, thereby preventing PAP PARylation on *hsp* genes and allowing polyadenylation of *hsp* transcripts to occur unabated.

PARylation has also been implicated in control of alternative splicing (Ji and Tulin, 2009). However, the effects on splicing appear to be indirect and not mediated by PARylation of the splicing factors involved. While these previous studies, along with ours, support a role for PARP1 in regulating gene expression through modulation of mRNA processing, our experiments have provided evidence that direct PARylation of a core processing factor, PAP, can modulate its activity and thereby influence processing of mRNA precursors.

Our results show that only heat shock, and not γ-IR or oxidative stress, was able to redirect PARP1 activity toward PAP and inhibit polyadenylation. Our current understanding of how PARP1 is activated by different stimuli is partial, and it is therefore difficult to explain this specificity. Three different modes of PARP1 activation have been described: DNA damage, post-translational modification, and binding to protein partners (reviewed in Luo and Kraus [2012]). The only mechanistic insight

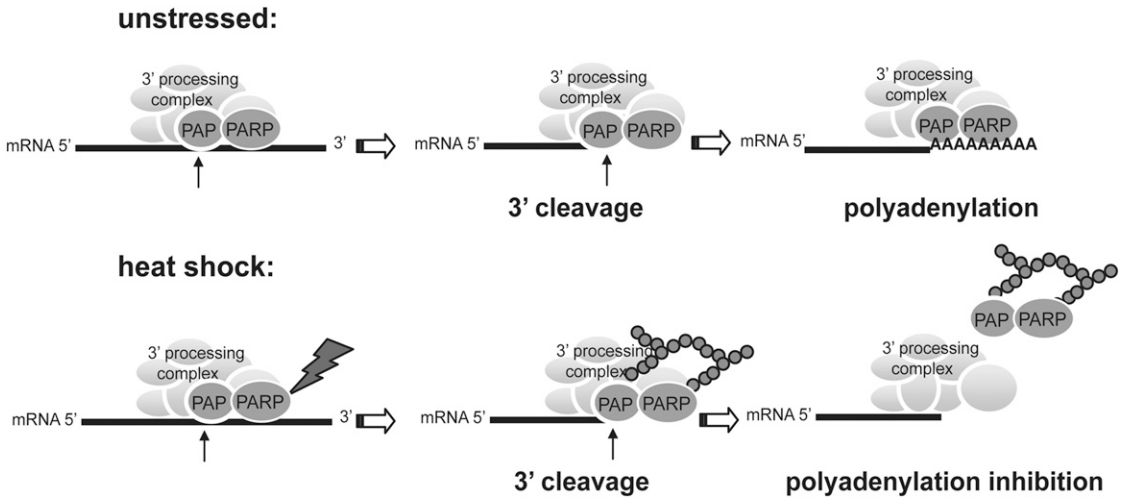


Figure 7. Model for PARP-Mediated Regulation of Polyadenylation during Heat Shock

In unstressed cells, 3′ cleavage and polyadenylation occur normally. During thermal stress, PARP1 becomes activated and PARylates PAP (and itself). The modified PAP and PARP1 dissociate from the 3′ end of genes, leading to polyadenylation inhibition and therefore arresting the production of new proteins under thermal stress.

available is with respect to DNA damage, where binding of PARP1 to damaged DNA induces a structural distortion that destabilizes its catalytic domain, leading to activation (Langelier et al., 2012). While any of the above three modes may apply, the molecular mechanism of PARP1 activation during heat shock is currently unknown.

In conclusion, our data support a model in which PARP1 activation following heat shock leads to PARylation of PAP, which in turn prevents PAP from associating with most mRNA transcripts, inhibiting their polyadenylation (Figure 7). Moreover, our ChIP results suggest that PAP is released from mRNAs concomitantly with PARP1. Ultimately, this mechanism, together with heat-induced repression of splicing, inhibits the maturation of newly synthesized transcripts, thereby preventing protein production in an environment otherwise prone to protein misfolding and aggregation. Our discovery that PAP is PARylated during heat shock provides an additional mechanism by which gene expression can be modulated at the posttranscriptional level, and also adds another layer of complexity to the functions of PARP1.

EXPERIMENTAL PROCEDURES

Cell Culture and Cell Treatments

HeLa and MCF-7 were cultured in DMEM with 10% fetal bovine serum. HEK293 cells stably expressing a tetracycline (tet)-inducible β-globin transgene were grown as previously reported. siRNA (200 pmol) against PARP1 (AAGAUAGAGCGUGAAGGCGAA) or nontargeting control (Dharmacon) was transfected into ~3 × 10^6 HeLa cells using Oligofectamine (Invitrogen). Cells were harvested 72 hr posttransfection directly into loading buffer or used to prepare NE. To analyze polyadenylation during heat shock in vivo, the HEK293 cells stably expressing the inducible β-globin transgene were treated for 3 hr with 1 ug/ml tet. Where noted, cells were incubated during the time of induction with 5 mM 3-ABA (Calbiochem). HEK293 cells were heat shocked for 30 min at 43°C in an incubator, and nuclear RNA was then extracted with 10 mM Tris (pH 7.4), 100 mM NaCl, 2.5 mM $MgCl_2$, and 50 μg digitonin followed by purification with TRIZOL reagent (Invitrogen) and DNaseI treatment (Fermentas). For the in vivo labeling experiment, 3 × 10^6 HeLa cells were incubated with 200 μCi of ^{3}H uridine for 30 min either at 37°C or 43°C. Where noted, 5 mM 3-ABA was added together with the tritiated uridine. HeLa nuclear RNA was extracted in NP-40 buffer (10 mM Tris [pH 7.4], 0.15% NP-40, 150 mM NaCl) followed by purification with TRIZOL. Poly(A) selection was performed using magnetic oligo(dt) beads (Novagen), and the eluted RNA was collected for scintillation counting.

To measure the half-life of β-globin transcripts following heat shock, 293 cells stably expressing the inducible β-globin transgene were treated for 3 hr with 1 ug/ml tet, and cells were then heat shocked for 30 min at 43°C in an incubator, after which they were washed several times and incubated with tet-free DMEM. RNA was extracted at the indicated time points using TRIZOL. Samples were subjected to DNaseI treatment, phenol/chloroform, ethanol precipitation, and reverse transcription.

Immunoprecipitation of PARylated PAP and Western Blotting

HeLa or MCF-7 cells (~8 × 10^6) were heat shocked for 1 hr at 43°C. Where noted, 10 mM 3-ABA was added just before heat shock. After two washes with cold PBS containing 0.1 mM tannic acid (a PARG inhibitor), cells were resuspended in two volumes of IP lysis buffer (see the Supplemental Experimental Procedures). Cells were lysed for 45 min on ice and then centrifuged at 13,000 g for 15 min at 4°C. After measuring protein concentrations using the Bradford assay and equalizing protein amounts, a combination of two anti-PAP antibodies was added to the supernatants, and tubes were rotated for 1 hr at 4°C. Protein A Sepharose beads (20 μl) (GE Healthcare) were then added and samples incubated overnight with rotation at 4°C. Samples were then washed four times in lysis buffer and resuspended in 2× loading buffer. Where noted, the PAP precipitate was then incubated with 40 ng of PARG for 10 min at 37°C in a buffer containing 10 mM NaH_2PO_4 (pH 7.5), 10 mM KCl, and 1 mM DTT. See the Supplemental Experimental Procedures for western blot and immunoprecipitation of PARylated PAP from NE protocols.

In Vitro 3′ Processing Assays

^{32}P-labeled simian virus 40 late (SVL) full-length or precleaved RNA substrates were prepared as described previously (Ryner et al., 1989). For 3′ cleavage assays, reaction mixtures consisted of 40% NE (prepared as described in Kleiman and Manley [2001]), 0.2–0.5 ng labeled RNA, 0.25 U RNasin (Promega), 2 mM EDTA, 250 ng tRNA, 2.5% polyvinyl alcohol (PVA), 20 mM creatine phosphate, 8 mM Tris (pH 7.9), 10% glycerol, 25 mM $NH_4(SO_4)_2$, 0.2 mM DTT, 0.2 mM PMSF. Polyadenylation assays contained the same reagents, with

the omission of EDTA and addition of 1 mM $MgCl_2$ and 1 mM ATP. See the Supplemental Experimental Procedures for nonspecific polyadenylation protocol.

In Vitro PARylation

The indicated amounts of His-PAP and MBP-PARP1 were incubated in Buffer D with 1 mM NAD^+ (or 0.05 mM NAD^+ and 0.4 μM ^{32}P-NAD^+), 400 ng sssp (sheared salmon sperm DNA), and 10 mM $MgCl_2$. PARylation reactions were carried out at 37°C for 10 min and stopped by adding 2× loading buffer. Where noted, the indicated amount of PARG was then added to reaction mixtures for an additional 10 min at 37°C. After 5 min boiling, proteins were resolved by SDS-PAGE and subjected to ECL with the relevant antibodies (anti-PAR from Biomol, anti-PAP and anti-PARP as above) or subjected to autoradiography.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and five figures and can be found with this article at http://dx.doi.org/10.1016/j.molcel.2012.11.005.

ACKNOWLEDGMENTS

We thank Dr. Maria Carmo-Fonseca for providing us with the β-globin inducible stable 293 cell line and Dr. W. Lee Kraus for providing the His-PARG and His-PARP1 constructs. We are also grateful to members of our lab for helpful discussion and to Dr. Emanuel Rosonina for critical reading of the manuscript. This work was supported by NIH grant R01 GM28983.

Received: June 21, 2012
Revised: September 21, 2012
Accepted: November 1, 2012
Published online: December 6, 2012

REFERENCES

Andrews, G.K., Harding, M.A., Calvet, J.P., and Adamson, E.D. (1987). The heat shock response in HeLa cells is accompanied by elevated expression of the c-fos proto-oncogene. Mol. Cell. Biol. *7*, 3452–3458.

Carty, S.M., and Greenleaf, A.L. (2002). Hyperphosphorylated C-terminal repeat domain-associating proteins in the nuclear proteome link transcription to DNA/chromatin modification and RNA processing. Mol. Cell. Proteomics *1*, 598–610.

Colgan, D.F., Murthy, K.G., Prives, C., and Manley, J.L. (1996). Cell-cycle related regulation of poly(A) polymerase by phosphorylation. Nature *384*, 282–285.

Cuesta, R., Laroia, G., and Schneider, R.J. (2000). Chaperone hsp27 inhibits translation during heat shock by binding eIF4G and facilitating dissociation of cap-initiation complexes. Genes Dev. *14*, 1460–1470.

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.

de Almeida, S.F., García-Sacristán, A., Custódio, N., and Carmo-Fonseca, M. (2010). A link between nuclear RNA surveillance, the human exosome and RNA polymerase II transcriptional termination. Nucleic Acids Res. *38*, 8015–8026.

de Murcia, G., Jongstra-Bilen, J., Ittel, M.E., Mandel, P., and Delain, E. (1983). Poly(ADP-ribose) polymerase auto-modification and interaction with DNA: electron microscopic visualization. EMBO J. *2*, 543–548.

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

Fossati, S., Formentini, L., Wang, Z.Q., Moroni, F., and Chiarugi, A. (2006). Poly(ADP-ribosyl)ation regulates heat shock factor-1 activity and the heat shock response in murine fibroblasts. Biochem. Cell Biol. *84*, 703–712.

Gagné, J.P., Shah, R.G., and Poirier, G.G. (2001). Analysis of ADP-ribose polymer sizes in intact cells. Mol. Cell. Biochem. *224*, 183–185.

Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat. Struct. Mol. Biol. *15*, 71–78.

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

Hossain, M.B., Ji, P., Anish, R., Jacobson, R.H., and Takada, S. (2009). Poly(ADP-ribose) polymerase 1 interacts with nuclear respiratory factor 1 (NRF-1) and plays a role in NRF-1 transcriptional regulation. J. Biol. Chem. *284*, 8621–8632.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. *26*, 2119–2137.

Huang, J.Y., Chen, W.H., Chang, Y.L., Wang, H.T., Chuang, W.T., and Lee, S.C. (2006). Modulation of nucleosome-binding activity of FACT by poly(ADP-ribosyl)ation. Nucleic Acids Res. *34*, 2398–2407.

Ji, Y., and Tulin, A.V. (2009). Poly(ADP-ribosyl)ation of heterogeneous nuclear ribonucleoproteins modulates splicing. Nucleic Acids Res. *37*, 3501–3513.

Ji, Y., and Tulin, A.V. (2010). The roles of PARP1 in gene control and cell differentiation. Curr. Opin. Genet. Dev. *20*, 512–518.

Kim, M.Y., Mauro, S., Gévry, N., Lis, J.T., and Kraus, W.L. (2004). NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. Cell *119*, 803–814.

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.

Kraus, W.L. (2008). Transcriptional control by PARP-1: chromatin modulation, enhancer-binding, coregulation, and insulation. Curr. Opin. Cell Biol. *20*, 294–302.

Krishnakumar, R., and Kraus, W.L. (2010). The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. Mol. Cell *39*, 8–24.

Krishnakumar, R., Gamble, M.J., Frizzell, K.M., Berrocal, J.G., Kininis, M., and Kraus, W.L. (2008). Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. Science *319*, 819–821.

Langelier, M.F., Planck, J.L., Roy, S., and Pascal, J.M. (2012). Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1. Science *336*, 728–732.

Lönn, P., van der Heide, L.P., Dahl, M., Hellman, U., Heldin, C.H., and Moustakas, A. (2010). PARP-1 attenuates Smad-mediated transcription. Mol. Cell *40*, 521–532.

Luo, X., and Kraus, W.L. (2012). On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1. Genes Dev. *26*, 417–432.

Medler, S., Al Husini, N., Raghunayakula, S., Mukundan, B., Aldea, A., and Ansari, A. (2011). Evidence for a complex of transcription factor IIB with poly(A) polymerase and cleavage factor 1 subunits required for gene looping. J. Biol. Chem. *286*, 33709–33718.

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

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688–700.

Ogata, N., Ueda, K., Kawaichi, M., and Hayaishi, O. (1981). Poly(ADP-ribose) synthetase, a main acceptor of poly(ADP-ribose) in isolated nuclei. J. Biol. Chem. *256*, 4135–4137.

Ouararhni, K., Hadj-Slimane, R., Ait-Si-Ali, S., Robin, P., Mietton, F., Harel-Bellan, A., Dimitrov, S., and Hamiche, A. (2006). The histone variant mH2A1.1 interferes with transcription by down-regulating PARP-1 enzymatic activity. Genes Dev. *20*, 3324–3336.

Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. Genes Dev. *23*, 1247–1269.

Rouleau, M., Patel, A., Hendzel, M.J., Kaufmann, S.H., and Poirier, G.G. (2010). PARP inhibition: PARP1 and beyond. Nat. Rev. Cancer *10*, 293–301.

Rozenblatt-Rosen, O., Nagaike, T., Francis, J.M., Kaneko, S., Glatt, K.A., Hughes, C.M., LaFramboise, T., Manley, J.L., and Meyerson, M. (2009). The

tumor suppressor Cdc73 functionally associates with CPSF and CstF 3′ mRNA processing factors. Proc. Natl. Acad. Sci. USA *106*, 755–760.

Ryan, K., and Bauer, D.L. (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.

Ryner, L.C., Takagaki, Y., and Manley, J.L. (1989). Multiple forms of poly(A) polymerases purified from HeLa cells function in specific mRNA 3′-end formation. Mol. Cell. Biol. *9*, 4229–4238.

Sadis, S., Hickey, E., and Weber, L.A. (1988). Effect of heat shock on RNA metabolism in HeLa cells. J. Cell. Physiol. *135*, 377–386.

Shamovsky, I., and Nudler, E. (2008). New insights into the mechanism of heat shock response activation. Cell. Mol. Life Sci. *65*, 855–861.

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

Shimazu, T., Horinouchi, S., and Yoshida, M. (2007). Multiple histone deacetylases and the CREB-binding protein regulate pre-mRNA 3′-end processing. J. Biol. Chem. *282*, 4470–4478.

Shin, C., Feng, Y., and Manley, J.L. (2004). Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. Nature *427*, 553–558.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1988). Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell *52*, 731–742.

Tulin, A., and Spradling, A. (2003). Chromatin loosening by poly(ADP)-ribose polymerase (PARP) at Drosophila puff loci. Science *299*, 560–562.

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.

Vethantham, V., Rao, N., and Manley, J.L. (2008). Sumoylation regulates multiple aspects of mammalian poly(A) polymerase function. Genes Dev. *22*, 499–511.

Warocquier, R., and Scherrer, K. (1969). RNA metabolism in mammalian cells at elevated temperature. Eur. J. Biochem. *10*, 362–370.

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.

Molecular Cell, Volume *49*

Supplemental Information

PARP1 Represses PAP and Inhibits Polyadenylation during Heat Shock

Dafne Campigli Di Giammartino, Yongsheng Shi, and James L. Manley

Supplemental Information Inventory

Figure S1

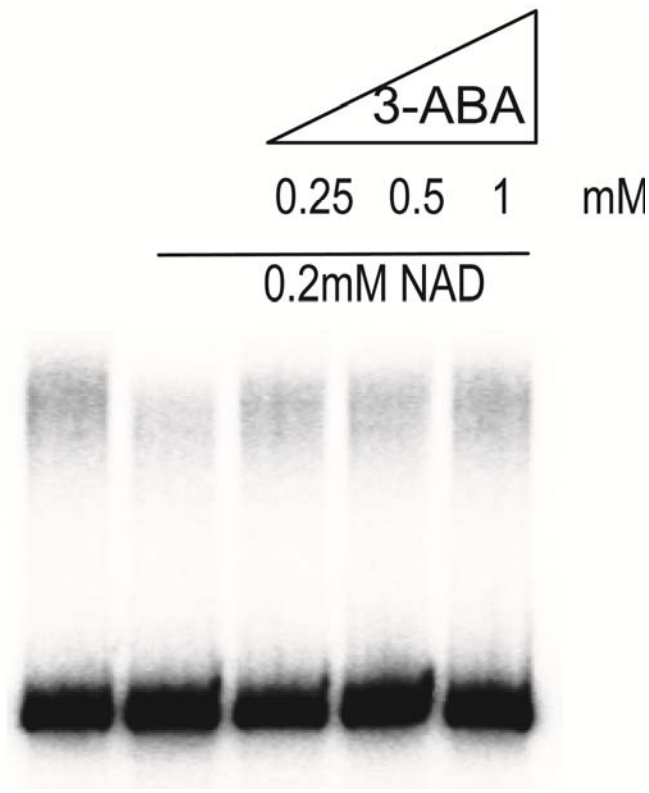
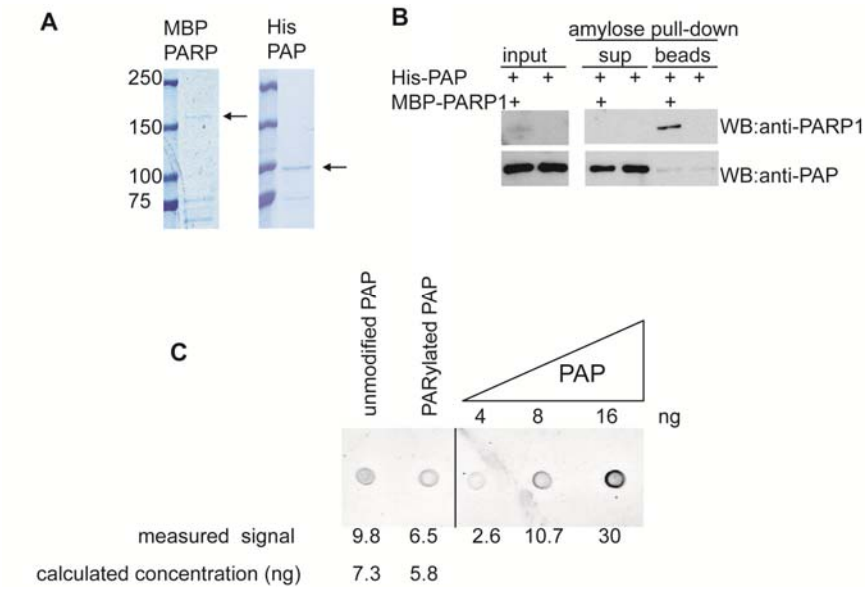


Figure S2


A
MBP PARP
His PAP
250
150
100
75
B
amylose pull-down
input
sup
beads
His-PAP + + + + + +
MBP-PARP1 + + +
WB:anti-PARP1
WB:anti-PAP
C
unmodified PAP
PARylated PAP
PAP
4
8
16
ng
measured signal 9.8 6.5 2.6 10.7 30
calculated concentration (ng) 7.3 5.8

FigureS3

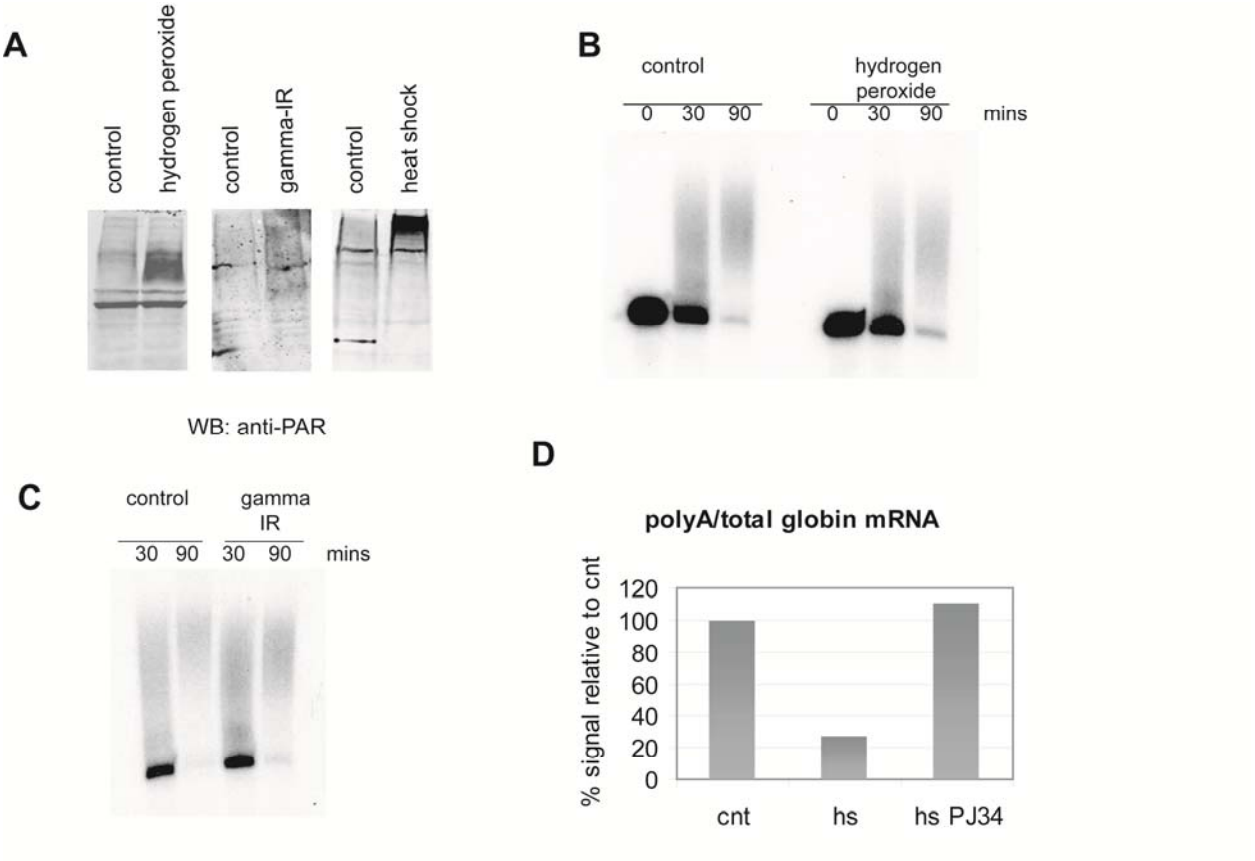

A
control
hydrogen peroxide
control
gamma-IR
control
heat shock
WB: anti-PAR
B
control
hydrogen peroxide
0 30 90
0 30 90
mins
C
control
gamma IR
30 90
30 90
mins
D
polyA/total globin mRNA
% signal relative to cnt
120
100
80
60
40
20
0
cnt
hs
hs PJ34

Figure S4

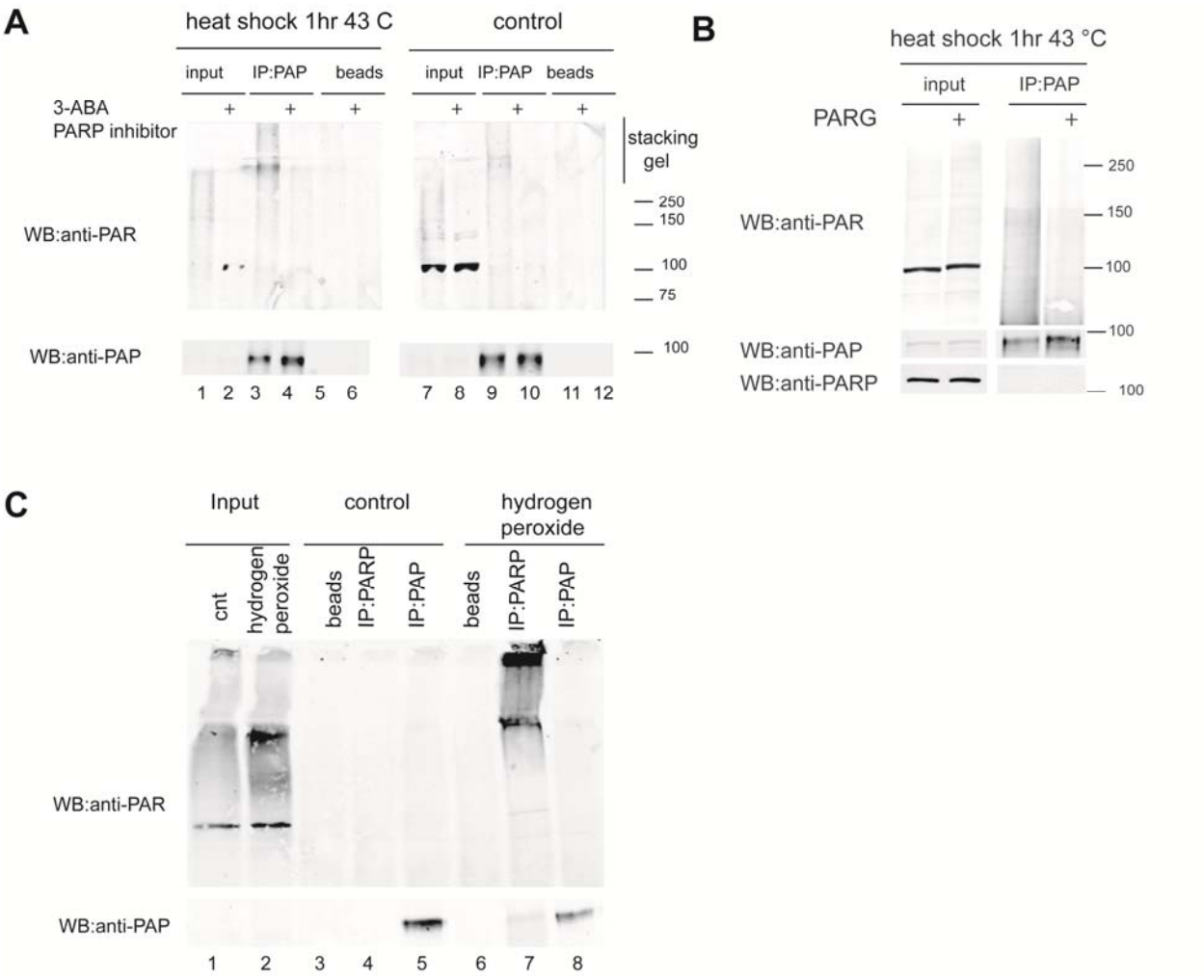
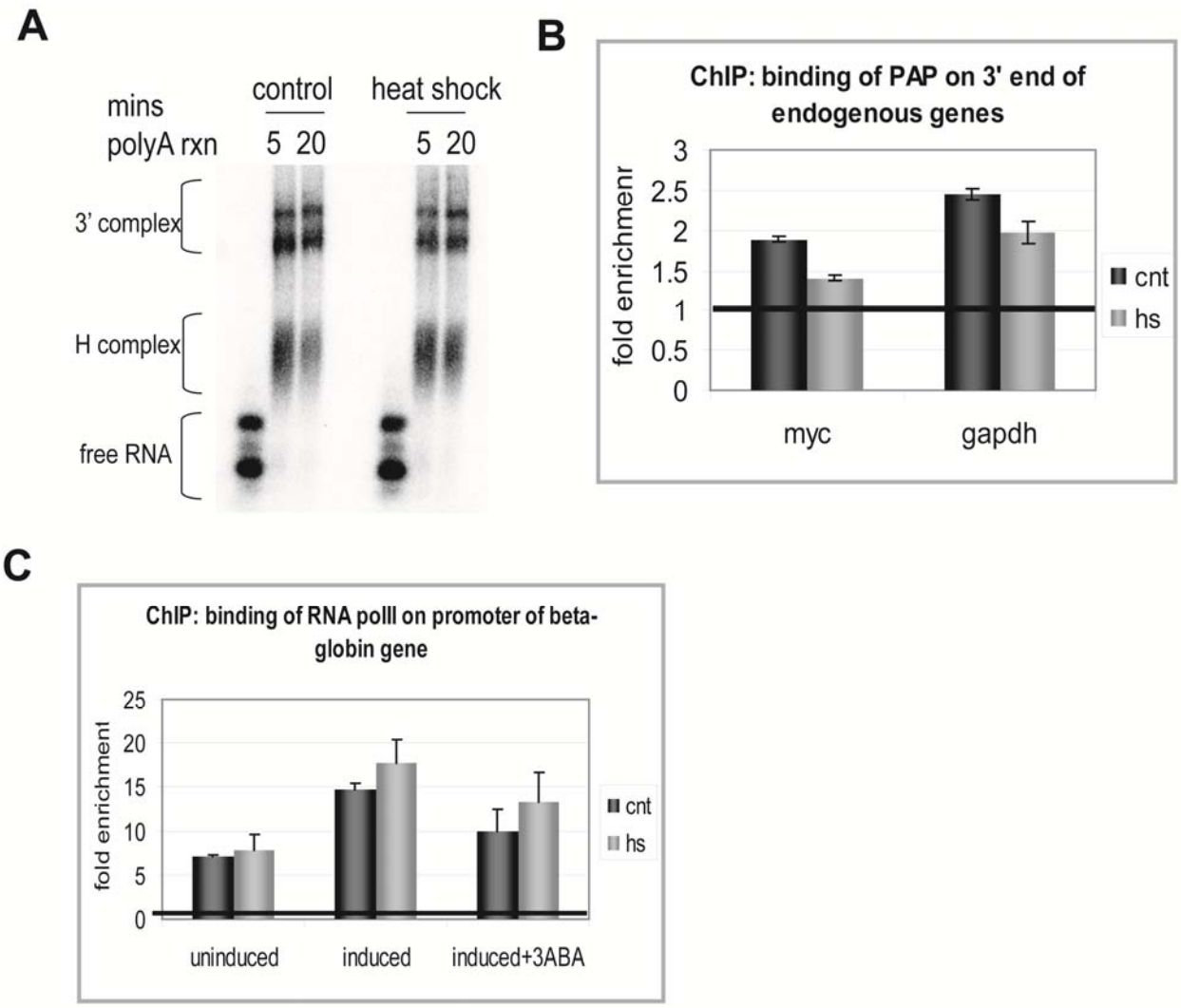

A
heat shock 1hr 43 C
control
input
IP:PAP
beads
input
IP:PAP
beads
3-ABA
PARP inhibitor
+
+
+
+
+
+
stacking gel
WB:anti-PAR
250
150
100
75
WB:anti-PAP
100
1 2 3 4 5 6 7 8 9 10 11 12
B
heat shock 1hr 43 °C
input
IP:PAP
PARG
+
+
WB:anti-PAR
250
150
100
100
WB:anti-PAP
WB:anti-PARP
100
C
Input
control
hydrogen peroxide
cnt
hydrogen peroxide
beads
IP:PARP
IP:PAP
beads
IP:PARP
IP:PAP
WB:anti-PAR
WB:anti-PAP
1 2 3 4 5 6 7 8

Figure S5


A
mins
polyA rxn
control
heat shock
5 20
5 20
3' complex
H complex
free RNA
B
ChIP: binding of PAP on 3' end of endogenous genes
fold enrichmenr
3
2.5
2
1.5
1
0.5
0
cnt
hs
myc
gapdh
C
ChIP: binding of RNA polII on promoter of beta-globin gene
fold enrichment
25
20
15
10
5
0
cnt
hs
uninduced
induced
induced+3ABA

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. The PARP1 inhibitor 3-ABA restores polyadenylation activity of NE treated with NAD.

3' cleavage and polyadenylation assays were carried out using internally ^{32}P labeled RNA substrate and HeLa NE in the presence of 0.2mM NAD^+ and the indicated amounts of the PARP inhibitor 3-ABA. RNAs were purified, resolved by denaturing PAGE and visualized by autoradiography.

Figure S2. Quantitation of PARylated PAP after re-purification following in vitro PARylation.

(A) comassie stain of MBP-PARP1 and His-PAP after purification from bacteria (bands representing the purified proteins are indicated by an arrow). **(B)** In vitro PARylation was carried out with MBP-PARP1 and His-PAP, followed by amylose pulldown. Samples from supernatant and amylose beads precipitate (after washes with 1M NaCl) were loaded on a SDS-PAGE followed by western blot with the indicated antibodies.**(C)**2 ul of unmodified (mock-PARylated) or PARylated PAP samples that were re-purifed after in vitro PARylation were blotted on a nitrocellulose membrane together with known increasing amounts of PAP protein. Western was performed using anti-PAP antibody. Anti-rabbit IR Dye was used as secondary antibody (LICOR) and the fluorescent signals were detected using the Odissey infrared imaging system. Image J software was used to quantify the blots relative to dilutions of the known concentrations of PAP. Unmodified and PARylated PAP samples were then adjusted to equal concentration

(Figure 3A) before using them in non-specific polyadenylation assays (Figure 3B) and gel shift assays (Figure 5B).

Figure S3. Activation of PARP1 with hydrogen peroxide or gamma-IR does not lead to polyadenylation inhibition.

(A) HeLa cells were treated with 500uM hydrogen proxide fro 5 minutes, 3.5 Gy of gamma irradiation or 1 hr heat shock at 43°C and extracts made. Samples were loaded on SDS-PAGE followed by western blot with anti-PAR antibody **(B)** In vitro polyadenylation reactions were carried out for the indicated times using in vitro transcribed SVL pre-mRNA and NE from untreated HeLa cells or cells treated with 500uM hydrogen peroxide for 5 minutes. Following incubation, RNAs were isolated, resolved by denaturing PAGE, and subjected to autoradiography. **(C)** In vitro polyadenylation reactions were carried out for the indicated times using in vitro transcribed SVL pre-mRNA and NE made from untreated HeLa cells or cells treated with 3.5 Gy of gamma-irradiation (cells were collected 10 minutes after treatment). Following incubation, RNAs were isolated, resolved by denaturing PAGE, and subjected to autoradiography. **(D)** Nuclear RNA was extracted from cells induced for 3 hrs as in Figure 4B that were either kept at 37°C (cnt) or heat shocked at 43°C (hs) for 30 mins. Where noted 120uM of the PARP inhibitor PJ34 was added at the time of induction. After reverse transcription, cDNAs were subject to qPCR to calculate the relative amount of β-globin polyadenylated mRNA (cDNA generated with oligo(dt) primers) compared total β-globin RNA (cDNA generated with random hexamers). The values obtained from RT-PCR were analyzed using the relative standard curve method and normalized to actin mRNA levels. Values are presented relative to the control sample which was set at 100%.

Figure S4. PAP is PARylated in vivo under heat shock but not following hydrogen peroxide treatment.

(A) PAP was IPed from untreated HeLa cells (lanes 7 to 12) or from cells that were heat shocked for 1hr at 43°C (lanes 1 to 6). Where indicated, 3-ABA was added just before the heat shock (lanes 4, 6, 10 and 12). Lane 1 displays the input for lanes 3 and 5, lane 2 the input for lanes 4 and 6, lane 7 the input for lanes 9 and 11 and lane 8 the input for lanes 10 and 12. Sepharose A beads without antibody were used for IPs shown in lanes 5, 6, 11 and 12. Following IP, samples were subjected to SDS-PAGE and western blots (WB) were carried out with the indicated antibodies. **(B)** PAP was IPed as in (A). Where indicated the precipitated PAP was then incubated with PARG to remove the modification. Samples were then loaded on a SDS-PAGE and western blots carried out with the indicated antibodies**. (C)** PAP or PARP1 were IPed from untreated HeLa cells (lanes 3 to 5) or cells that were treated for 5 minutes with 500uM hydrogen peroxide (lanes 6 to 8). Lane 1 displays the input for lanes 3 to 5, lane2 displays input for lanes 6 to 8. Sepharose A beads without antibody were used as control (lanes 3 and 6). Following IP samples were subjected to SDS-PAGE and western blot (WB) was carried out using the indicated antibodies.

(B)

Figure S5. PARylated PAP dissociates from the 3' end of *c-myc* and *gapdh* genes but does not affect 3' complex formation.

(A) Gel shift assay using ^{32}P labeled, in vitro-transcribed RNA and NE from HeLa cells that were either untreated or heat shocked 1 hr at 43°C. NEs and RNA were incubated under

conditions that allow pre-mRNA 3' processing for 5 or 20 minutes and then loaded on a 1.5% low-melting-point agarose gel. The gel was then dried and exposed to a PhosphorImager screen.

(B) ChIP was carried out in uninduced 293 cells using an anti-PAP antibody and amplifying the 3' end region of *C-MYC* (n=3) or *GAPDH* (n=3) genes. The results were analyzed as in Figure 6B.

(C) ChIP was carried out as in Figure 6D with the exception that a different set of primers were used in order to amplify the promoter region of β-globin

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Immunoprecipitation of PARylated PAP and western blotting (continued)

IP Lysis buffer (50mM Tris pH7.5, 150mM NaCl, 0.3% NP-40, 0.2% Triton, 0.1mM EDTA, 10% glycerol) with protease inhibitors (0.1mM tannic acid, 25uM NaF, 0.1mM NaVO4, 0.1mM PMSF, 5ug/ml Aprotonin, 0.5ug/ml Pepstatin, 0.5ug/ml Leupeptin). Proteins were resolved by 6% SDS-PAGE and westerns were carried out using as primary antibodies: anti-PAR in TBS, anti-PAP in TBS, anti-PARP in PBS and as secondary antibodies anti-mouse or anti-rabbit (LICOR). Fluorescent signals were detected by the Odyssey infrared imaging system. IP of PARylated PAP from NE was carried out as above except that the NE was treated with 0.5 mM NAD^+ for 30 min at 30°C (or 0.4uM ^{32}P-NAD, Perkin Elmer 800Ci/mmol) and then diluted with 4 volumes of IP buffer with inhibitors (see above). Anti-PAP antibody was added overnight and the next day pre-blocked beads were added for 1 hr. Samples were then washed 4 times in IP buffer (for ^{32}P-NAD experiments, samples were washed 6 times in IP buffer with 300mM NaCl and 0.6% NP-40) and subjected to SDS-PAGE and western blot as above.

In vitro 3' processing assays (continued)

Non-specific polyadenylation mixtures contained 2.5% PVA, 1mM $MnCl_2$, 100ng BSA, 1mM ATP, 0.5U RNasin, 10mM Hepes pH 7.9, 25mM $NH_4(SO_4)_2$, 0.2mM PMSF, 0.2mM DTT and 0.2ng his-PAP. Reaction mixtures were incubated 30 mins at 30°C. In all cases, reaction mixtures were incubated at 30°C for up to 90 mins, followed by proteinase K treatment, phenol/chloroform extraction, ethanol precipitation and separation on 6% urea-acrylamide gels.

Following autoradiography the signal was quantified with ImageJ and the percentage efficiency calculated as % of pA signal divided by total signal (which equals pA plus pre-mRNA).

Antibodies

For western: anti-PAR antibody was purchased from Tulip biolabs (1020N), anti-PARP from Santa Cruz (sc7150), anti-Actin from Sigma (A2066), anti-His from Santa Cruz (sc803), anti-MBP from NEB(E8038S), anti-PAP from Bethyl (A301-010). For IP and ChIPs: anti-PARP was purchased from Active Motif (39559), N20 for RNAP II was from Santa Cruz (sc899), a mixture of two different antibodies, both from Bethyl, were used for PAP (A301-09 and A301-08), the PAR resin was purchased from Tulip biolabs (2302).

Re-purification of PAP after in vitro PARylation

Following in vitro PARylation, reaction mixtures were diluted with 5 volumes binding buffer (20 mM Tris pH7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol, 1 mM PMSF, 100 uM 3-ABA). 20 ul amylose resin was added and incubated 2 hrs at 4°C rotation. The pellet containing MBP-PARP1 was discarded while the PAP-containing supernatant was further diluted 1:2 and incubated with 20 ul polyADP-ribose affinity resin (Tulip biolabs) overnight. The resin was washed 4 times in 0.5 ml binding buffer with 0.1% NP40 and then incubated 4 hrs at 4°C in 20 ul binding buffer supplemented with 40 mM free ADP-ribose . The eluted PARylated PAP was dialyzed 2 hrs in Buffer D in a mini-dialyzer device. 2 ul of the PARylated PAP or unmodified PAP was used for dot blot. Quantitation of LICOR-scanned blots was done using ImageJ.

Chromatin IP

The HEK-293 cells expressing the inducible β-globin gene were treated with 1 mg/ml tet for 6 hrs. (The same calls and protocol were used but without addition of tet for analysis of *C-MYC*, *GAPDH* and *hsp* genes.) Where mentioned, 5 mM 3-ABA was included during the last 4 hrs of induction. Cells were then exposed to 5 mins heat shock at 43°C in an incubator. 1% formaldehyde was then added to the cells for 15 mins at room temperature, followed by glycine to a final concentration of 125 mM for 5 mins. Cells were gently washed in PBS and collected in 2 ml LB1 buffer with inhibitors (50 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% triton). After 5 mins rotation, cells were pelleted and resuspended in 2 ml LB2 buffer (10 mM Tris pH 8, 200 mM NaCl, 1 mMEDTA, 0.5 mM EGTA). Finally pellets were resuspended in 0.5 ml LB3 (as LB2 but with 100 mM NaCl and 0.1% Na deoxycholate and 0.5% N-Laurylsarcosine). Sonication was carried out with a Branson Sonifier 250 (15 pulses,10 times). TritonX100 was added to a final concentration of 0.5% and samples were cleared by centrifuging at 13000g for 10 mins. 4 ug of the indicated antibodies were added to the supernatant with 20 ul protein A sepharose and incubated overnight at 4°C with rotation. Beads were then washed twice in Buffer 1 (20 mM Tris pH 8, 150 mM NaCl, 0.1% SDS, 1% triton, 2 mM EDTA), once in Buffer 2 (as Buffer 1 but with 500 mM NaCl), once quickly in Buffer 3 (10 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 1% NP40, 1% Na deoxycholate) and once with TE. 200 ul elution buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS) was added and samples incubated at 65°C with rotation for 6 hrs. After adjusting the pH to ~5.5 (by adding a final concentration of 50mM sodium acetate pH5.2), samples were subjected to DNA clean-up (using the Qiagen PCR clean-up kit) and eluted in 100 ul of 1 mM Tris pH 8.0.

RT-PCR

Real-time PCR was performed in 96 well plates with Maxima Sybergreen (Roche) using StepOnePlus (Applied Biosystems). For the ChIP experiments the results were analyzed using the ΔΔct method (Livak and Schmittgen, 2001), normalizing to input and an intergenic region (~2 Kb upstream of *C-MYC*).

For calculating the polyA/total mRNA signal and for measuring the β-globin mRNA half life, cDNAs obtained with either random hexamers, oligo dt or a mixture of both were used for RT-PCR. Values obtained after RT-PCR were analyzed with the relative standard curve method, before and after induction, normalized to actin mRNA levels and then the ratio relative to the control condition (considered as 100%) was calculated.

Primers

The primers used for amplification of β-globin target the 3' end of the gene (exon3): F-AAGCTCGCTTTCTTGCTGTC, R-GATGCTCAAGGCCCTTCATA. The primers for the intergenic region used as endogenous control are:

F-AAGACGCTTTGCAGCAAAATC

R-AGGCCTTTGCCGCAAAC

Primers for hsp genes are:

hsp70 F- GCCTTTCCAAGATTGCTGTT

hsp70R- TGCATGTAGAAACCGGAAAA

hsp90F-TCTCTCCACAGGGCTTGTTT

hsp90R-ACTCCCCTTTCCCCCTAAAT

hsp27F-TGCAAAATCCGATGAGACTG

hsp27R-TTTGACAGGTGGTTGCTTTG.

Primers for *C-MYC:*

F-ACACAATGTTTCTCTGTAAATATTGCCA,

R-ACTAGGATTGAAATTCTGTGTAACTGCT

Primers for GAPDH:

F-CCCTGTGCTCAACCAGT

R-CTCACCTTGACACAAGCC.

In vitro pull-down assay

200 ng his-PAP purified from *E. coli* was conjugated to nickel beads and incubated with 200 ng of MBP-PARP in 1 ml of binding buffer (20 mM Hepes pH 7.9, 20% glycerol, 200 mM NaCl, 0.05% NP-40, 0.5 mM PMSF and 1 mg/ml BSA) for 2 hrs at 4°C. Samples were washed three times with high salt binding buffer (500 mM NaCl) and once more with binding buffer. After resuspending the beads in loading buffer and boiling 5 mins, samples were resolved by 6% SDS-PAGE, followed by western blot.

Gel shift assay

Gel shift assays were carried out by incubating recombinant proteins with ^{32}P-labelled RNA substrate under the same conditions used for polyadenylation except that $MnCl_2$ and ATP were omitted and 100 ng BSA was added. After 10 mins incubation at 30 °C, sample were loaded on a 5% non-denaturing polyacrylamide gel. The gel was then dried and analyzed by Phosphorimager (Molecular Dynamics Storm 860).

3’ processing complex formation assay

Cleavage reactions with SVL substrate were set up as described above using NE from untreated cells or cells that were previously heat shocked for 1 hr at 43°C. Samples were incubated for 5 or 20 mins at 30°C followed by 10 minutes incubation on ice with 5 ug/ul heparin. The RNA-protein complexes were then resolved on 1.5% low-melting-point agarose gel. The gel was dried and exposed to a phosphorimager screen.

RBBP6 is a component of the human 3' processing complex and a regulator of mRNAs with AU-rich 3'UTRs

Dafne C. Di Giammartino[1], Wencheng Li[2], Jossie Yashinskie[1], Mainul Hoque[2], Bin Tian[2] and James L. Manley[1]

1. Columbia University, Department of Biological Sciences, New York, NY 10027.

2. Rutgers New Jersey Medical School, Department of Biochemistry and Molecular Biolology, Newark, NJ 07103.

Abstract

Maturation of the 3' ends of most mRNAs is mediated by a large protein complex, the 3' processing complex. Here we show that RBBP6, identified initially as an Rb and p53 binding protein, is a component of this complex, required for 3' cleavage in HeLa nuclear extract. RBBP6 associates with other core factors and this interaction, as well as cleavage activity, requires an unusual ubiquitin-like domain known as the DWNN. The DWNN is also expressed, via alternative RNA processing, as a single domain protein (iso3), and we show that iso3, which is downregulated in cancer, competes with RBBP6 for binding the core machinery, thereby inhibiting 3' processing. We next performed genome-wide RNA analyses, and observed following RBBP6 knockdown a downregulation in gene expression accompanied by increased usage of distal poly(A) sites. RNAs with AU-rich 3'UTRs, such as c-Fos and c-Jun, were especially enriched in the downregulated transcripts, which we show is the result of defective 3' processing and degradation by the exosome. Our results indicate that RBBP6 is a novel 3' processing factor that regulates expression of RNAs with AU-rich 3'UTRs.

Introduction

The 3' ends of nearly all polyadenylated RNAs are produced by a two-step reaction involving endonucleolytic cleavage of the transcript followed by synthesis of the poly(A) tail. This step in gene expression, which is necessary for mRNA stability, export, and translation (Moore and Proudfoot, 2009), is mediated by a massive protein machinery, the 3' processing complex (Mandel et al. 2008; Moore and Proudfoot 2009). The core complex includes four multi-subunit protein complexes: cleavage/polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF), cleavage factor I (CFI) and cleavage factor II (CFII). Additional proteins, such as poly(A) polymerase (PAP), symplekin, poly(A) binding protein II (PABPII) and RNA polymerase II, specifically the C-terminal domain of its largest subunit (CTD), also play important roles. The site where polyadenylation occurs, the poly(A) site, is defined by multiple cis-elements that contact several subunits of this machinery. Most transcripts contain more than one potential poly(A) site, and the selection of alternative sites is an important aspect of gene control (Di Giammartino et al. 2011; Elkon et al. 2013).

A large number of proteins, that associate with the core 3' processing factors, were discovered in a proteomic analysis of the human complex assembled on substrate RNA (Shi et al. 2009). Many of these, such as for example PARP-1 (Di Giammartino et al. 2013), are thought to connect 3' processing to other nuclear events, while others (e.g., WDR330) were previously undiscovered components of the human core 3' processing machinery. One protein that could conceivably fall in either category was RBBP6 (retinoblastoma binding protein 6). RBBP6 is a large (~250 KD) multidomain protein that is similar in its N-terminus to the yeast 3' processing factor Mpe1, which is an integral subunit of the yeast

CPF complex (cleavage and polyadenylation factor) and is strictly required for 3' end formation (Vo et al. 2001). Mpe1 is required for cell viability and absence of RBBP6 homologues leads to embryonic lethality in mouse (Li et al. 2007), flies (Mather et al. 2005), and worms (Huang et al. 2013).

RBBP6 has a number of features that suggest important roles in linking 3' end formation with other cellular processes. Homologues of RBBP6 are found in all eukaryotes and share three well conserved domains at their N-terminus. The first is called the "domain with no name" or DWNN, which adopts a ubiquitin-like fold (Pugh et al. 2006). In addition to forming part of full-length RBBP6, this domain is also expressed in vertebrates as a small protein containing the DWNN and a short C-terminal tail (isoform 3, iso3) (Pugh et al. 2006), which has been shown to be down regulated in several human cancers (Mbita et al. 2012). The second conserved domain is a CCHC zinc knuckle. This type of zinc finger is found also in a number of splicing factors and in the 3' processing factor CPSF30 where it functions in RNA binding (Barabino et al. 1997). The third domain is a RING finger, a domain found in E3-ubiquitin ligases. The RING domain of RBBP6 binds to YB-1, a multifunctional RNA-binding protein, and YB-1 was shown to be a substrate of RBBP6 for ubiquitination, leading to YB-1 degradation by the proteasome (Chibi et al. 2008). Mammalian RBBP6 also includes a long C-terminal extension containing several additional significant domains. One is an RS domain characteristic of SR proteins and other proteins involved in pre-mRNA splicing. Similar domains are also present in two other 3' processing factors, CFI-68 and Fip1 (Boucher et al. 2001). RBBP6 was first identified as an interactor with the tumor suppressor protein Rb (Saijo et al. 1995; Sakai et al. 1995) and was subsequently shown to interact with another tumor suppressor, p53 (Simons et al. 1997).

RBBP6 interferes with binding of p53 to DNA and also facilitates interaction between p53 and its negative regulator Mdm2, leading to enhanced p53 ubiquitination and degradation. Moreover, disruption of *RBBP6* in mice leads to early embryonic lethality, but a p53-null mutation partially rescues viability (Li et al. 2007). The RBBP6 C-terminal region contains domains responsible for interaction with both tumor suppressors.

Here we describe experiments that establish RBBP6 as a *bona fide* 3'processing factor in vitro that functions in polyadenylation control in vivo. We show that nuclear extracts (NE) prepared from HeLa cells following RBBP6 knockdown (KD) were defective in 3' cleavage, but not poly(A) synthesis, and that activity could be rescued by adding a recombinant RBBP6 N-terminal derivative (RBBP6-N) containing only the DWNN, Zinc knuckle and RING domains. When expressed in vivo, RBBP6-N co-IPed with the 3' processing factors CPSF and CstF; the binding being particularly strong to CstF64 and mediated by the DWNN domain. Consistent with this, RBBP6 iso3 outcompeted RBBP6-N for binding to CstF64 and inhibited cleavage when added to NE, or when overexpressed in cells. We also performed genome-wide analyses and observed following RBBP6 KD a general downregulation in transcript levels accompanied by increased usage of distal poly(A) sites resulting in a global lengthening of 3' UTRs. Interestingly, RNAs with AU-rich 3'UTRs, such as c-Fos and c-Jun, were especially enriched in the downregulated transcripts, which we show was the result of defective 3' processing coupled with degradation by the exosome.

Results

RBBP6 is an essential 3' processing factor

Our previous proteomic analysis showed that RBBP6 is physically associated with the active 3' processing complex (Shi et al. 2009). To determine if RBBP6 is also in fact necessary for 3' processing activity we prepared NEs from HeLa cells treated for 72hrs with siRNA against RBBP6 or a non-targeting siRNA (Figure 1A is a western blot showing KD efficiency) and used the NE for in vitro 3' processing assays. We took advantage of the fact that the two steps of pre-mRNA 3' processing, cleavage and poly(A) synthesis, which are tightly coupled in vivo, can be analyzed separately in vitro (see Methods). Figure 1B shows that upon RBBP6 KD the NE retained poly(A) synthesis activity when incubated with a pre-cleaved ^{32}P-labeled simian virus 40 late (SVL) RNA (compare lanes 1 and 2), while cleavage of a longer SVL RNA was largely inhibited (compare lanes 3 and 4).

We next wished to determine whether the reduced 3' processing activity of the NE was a direct consequence of the absence of RBBP6. Because of the homology of the N-terminal region of RBBP6 to the yeast cleavage and polyadenylation factor Mpe1 (Vo et al. 2001), and because all RBBP6 homologues in eukaryotes include the three domains shown schematically in Figure 1C (RBBP6-N), but not the long C-terminal extension present only in vertebrates (Pugh et al. 2006) (Figure 1C top), we suspected that the N-terminal part of the protein might be sufficient to support 3' processing. Therefore we expressed and purified the RBBP6-N derivative from *E. coli* (Coomassie stained gel shown in Figure S1A) and repeated the 3' cleavage assay as in Figure 1B but adding increasing amounts of purified RBBP6-N. Figure 1D shows that RBBP6-N fully restored the cleavage activity of NE made after RBBP6 KD.

RBBP6-N interacts with CPSF/CstF

We next asked whether RBBP6 interacts in vivo with other 3' processing factors. Since the N-terminal part of the protein was sufficient to restore 3' cleavage after KD of the endogenous protein, we first used only that part of the protein to analyze its interactions. We cloned RBBP6-N expressing sequence into a Flag-tagged vector and transfected it into 293T cells. Figure 2A shows that following immunoprecipitation (IP) with an anti-Flag antibody under mild conditions (150 mM NaCl), endogenous CPSF and CstF components co-IPed with RBBP6-N. These interactions were specific to CPSF/CstF complexes because CF25 and symplekin did not show a similar interaction with RBBP6. Moreover, when performing the same co-IP under more stringent conditions (500 mM NaCl) only binding to CstF64 was retained (Figure 2B). To provide evidence that the interaction between RBBP6 and CstF is physiologically relevant, we repeated the IP with endogenous proteins. Figure 2C shows that the anti-CstF64 antibody co-IPed RBBP6, confirming the interaction (the reciprocal co-IP is shown in Figure S1B). Also, RNA does not mediate binding since the interaction was resistant to RNAse treatment (Figure 2C).

We then asked if the presence of RBBP6 is important for assembly of the 3' processing complex. NE made after RBBP6 (or control) KD was briefly incubated with an in vitro transcribed SVL RNA under cleavage conditions, to allow formation of the complex, and then loaded on a non-denaturing gel. As shown in Figure S1C, the NE made after RBBP6 siRNA showed no defect in 3' complex formation, indicating that, similar to its yeast counterpart (Vo et al. 2001), RBBP6 is not necessary for 3' complex assembly.

Finally, as an additional way to characterize RBBP6 involvement in pre-mRNA processing, we examined whether RBBP6 could bind RNA. A gel shift assay with purified recombinant RBBP6-N and SVL RNA shows that this part of the protein could bind RNA in

vitro (Figure 3A). The binding was resistant to the addition of up to ten-fold excess of a non-specific competitor (tRNA), while it was reduced to 50% when supplementing the NE with equivalent amounts of cold and hot SVL, and eliminated completely when adding ten-fold excess of this specific unlabelled competitor. As indicated by the gel shift assays, the interaction displayed a Kd of ~60 nM (Figure 3C, first panel), although we have been unable to obtain any evidence for specificity (results not shown).

The DWNN is essential for 3' processing activity

As mentioned above RBBP6-N is comprised of three domains, of which the most N-terminal is the DWNN. This domain is particularly interesting because in addition to being present in full-length RBBP6, it is also expressed in vertebrates as a small protein, iso3 (see Introduction and Discussion). Since the function of DWNN is totally unknown we wanted to investigate if it contributes to the activity of RBBP6-N in 3' processing. To this end we first expressed a truncated RBBP6-N that lacks the entire DWNN (ΔDWNN), and purified the protein from *E. coli* (schematic diagram in Figure 1C and a Coomassie stained gel of ΔDWNN is shown in Figure S1A). As shown in Figure 3B, ΔDWNN, in contrast to RBBP6-N, was unable to reconstitute 3' cleavage activity of NE prepared after RBBP6 KD. Figure 3B (right panel) displays quantification of cleavage efficiency in three separate experiments as normalized to siCNT.

The ΔDWNN protein could be defective in 3' processing because it lost its ability to bind RNA or to interact with CstF. Surprisingly, when a gel shift assay was done with purified ΔDWNN (Figure 3C, second panel), RNA binding was actually enhanced relative to RBBP6-N (Figure 3C, first panel), indicating that the DWNN is in some way inhibitory to RNA binding. Consistent with this, incubation of the RNA with a recombinant protein that

includes the "DWNN only" showed no binding (Figure 3C, third panel) and the same was observed when using recombinant iso3 (Figure 3C, fourth panel). Addition of increasing amounts of the DWNN only (or iso3) was not sufficient to disrupt the interaction between RNA and the ΔDWNN protein (Figure S1D), indicating that the inhibitory effect cannot be exerted in trans.

DWNN/ iso3 binds to CstF and inhibits cleavage in vitro

Another possible reason why ΔDWNN was unable to reconstitute cleavage activity of NE after RBBP6 KD could be that it lost its ability to interact with CstF. To test this possibility we transiently transfected 293T cells with Flag-tagged RBPP6-N or Flag-ΔDWNN vectors. IP was carried out with an anti-Flag antibody similar to Figure 2A. Figure 4A shows that while, as expected, RBBP6-N bound to endogenous CstF, ΔDWNN did not. The Flag-DWNN only protein, on the other hand, was sufficient to bind CstF. These results show that the DWNN is necessary and sufficient for the interaction between RBBP6 and CstF.

In light of the above, an intriguing possibility is that RBBP6 and iso3 might compete for binding to CstF. To test this, we transfected fixed amounts of Flag-RBBP6-N with increasing amounts of HA- iso3 and performed an IP with anti-Flag antibody as above. As can be seen in Figure 4B, RBBP6 iso3 was indeed able to outcompete RBBP6-N from binding to CstF. Since RBBP6 iso3 does not bind RNA (Figure 3C) but is able to prevent CstF from interacting with RBBP6, we next examined whether adding increasing amounts of recombinant iso3 to NE would inhibit 3' cleavage. Indeed, Figure 4C shows that the purified iso3 effectively inhibited cleavage of the SVL pre-mRNA when added to HeLa NE. Inhibition of cleavage was also observed using NE prepared from HeLa cells overexpressing HA-iso3

(Figure S1E). These results support a competition model in which the relative expression levels of the two RBBP6 isoforms regulate the efficiency of 3' processing (see below, and Discussion).

RBBP6 regulates APA

We next wished to investigate the functions of RBBP6 in vivo, including the possibility that the protein plays a role in alternative polyadenylation (APA). Most human genes encode transcripts with more than one potential poly(A) site, and APA is a widespread mechanism that generates mRNA isoforms with alternative 3' ends (reviewed in Di Giammartino et al. 2011; Elkon et al. 2013; Tian and Manley 2013). Changes in the levels or activity of core polyadenylation factors have been shown to affect APA globally (e.g.,Gruber et al. 2012; Yao et al. 2012). Given the above results implicating RBBP6 in 3' processing, we wondered whether lowering its levels in cells would affect APA. To this end, we used 3' region extraction and deep sequencing (3'READS) (Hoque et al. 2013) to detect APA changes following RBBP6 KD in MCF-7 cells. (We used MCF-7 cells because, as mentioned in the introduction, RBBP6 can interact with p53 and we wanted to carry out the KD in the background of a functional p53, which is not the case in HeLa cells). Figure S2A shows a western blot following KD of RBBP6 for 48 or 72 hours. The total number of poly(A) site supporting (PASS) reads is shown in Figure S2B, and the relative abundance of a specific polyadenylated isoform was defined as the fraction of PASS reads corresponding to that isoform over all PASS reads derived from the relevant gene (see Experimental Procedures for more details). We examined APA events based on the poly(A) site types depicted in Figure 5A and found a general lengthening of 3'UTRs after siRNA treatment for 48 hours (Figure 5B). Among the upregulated isoforms, 48% of them used the last poly(A) site of the

gene (L, shown in green) after RBBP6 KD and only 10% used the first (F, shown in blue), while among the downregulated isoforms, 45% of them used the first poly(A) site and only 10% the last. This trend could also be seen when the two most abundant 3'UTR APA isoforms of each gene were compared (Figure 5C): greater than 3-fold more genes had the distal site isoform upregulated compared to those with the proximal site isoform upregulated (1,030 vs. 311). In addition, intronic APA isoforms were downregulated (Figure 5D): about 2.5-fold more genes displayed upregulation of isoforms using 3'-most exon poly(A) sites as compared to genes having upregulation of isoforms using upstream (intronic or exonic) poly(A) sites (246 vs. 99, Figure 5D). These results together indicate that there is a global shift to distal poly(A) sites in siRBBP6 cells, regardless of the poly(A) site location. siRNA treatment for 72 hours gave similar results (Figure S2C, S2D and S2E).

We also examined cis elements surrounding the poly(A) sites whose isoforms were regulated in siRBBP6-treated cells (Figure 5E). Canonical cis elements, including upstream UGUA elements, AAUAAA poly(A) sequence, and downstream UGUG elements (see Tian and Manley, 2013), were significantly associated with poly(A) sites of upregulated isoforms, indicating that strong poly(A) sites were used preferentially in siRBBP6 KD cells. Several other sequence elements were associated with poly(A) sites of downregulated isoforms (Figure 5E), whose roles in poly(A) site usage are not clear. Taken together, our genomic analysis of APA suggests that RBBP6 plays a role in global APA regulation, facilitating proximal (and weak) poly(A) site usage, consistent with its positive function in 3' end processing.

RBBP6 regulates expression of mRNAs with AU-rich 3'UTRs

We next asked whether RBBP6 KD, in addition to altering APA, might also affect transcript abundance. To this end, we used Affymetrix genome-wide exon arrays, which could be more sensitive than deep sequencing in analyzing genes expressed at low levels. We found that after 72hrs RBBP6 KD caused downregulation in expression of 3,908 genes as compared to 1,206 genes that were upregulated (Figure 5F). A similar bias in numbers of upregulated vs. downregulated genes was also observed with the 3' READS data (Figure S3A). Interestingly, the disease and biological function term "cancer" was found to be most significantly associated with downregulated genes by Ingenuity Pathway Analysis (IPA; Figure S3B), suggesting that RBBP6 plays a role in expression of cancer-related genes.

We next examined the 3'UTRs of regulated genes using the microarray data. Based on analysis of pentamers, we found that, following RBBP6 KD, pentamers rich in A and U residues were highly enriched in 3'UTRs of downregulated genes as compared to 3'UTRs of nonregulated genes (Figure 6A). AU-rich elements (AREs) are found in 3'UTRs of many mRNAs and constitute one of the most common determinants of RNA stability (Gingerich et al. 2004). AREs are defined as sequences with frequent A and U residues, the best characterized of which have a core sequence of AUUUA within a U-rich context. AUUUA in fact was one of the significant pentamers identified to be enriched in the 3'UTRs of downregulated genes (highlighted in red in Figure S3C), with a p-value of 10^{-119} (Fisher's exact test). We then compared genes with different numbers of the AUUUA motif. As shown in Figure 6B, genes with more AUUUA motifs in 3'UTRs tended to be more downregulated, further indicating that genes with AREs are downregulated in siRBBP6 cells.

We next wished to validate the microarray data. We first analyzed by RT-qPCR mRNA levels of a number of genes with AU-rich 3'UTRs that appeared to be downregulated

by RBBP6 KD in the above microarray analysis. Figure 6C shows that the six AU-rich genes analyzed, c-Jun, c-Fos, Bcas1, Rab3b, Rbl2 and Fn1, all of which have been implicated in cancer, were indeed all downregulated following RBBP6 KD. c-Jun protein levels were also reduced by RBBP6 KD (Figure 6D). Essentially identical results were obtained when a second RBBP6 siRNA was used to deplete RBBP6 (Figure S4A).

One explanation for the downregulation of the ARE-containing mRNAs was that these naturally unstable transcripts were further destabilized in the absence of RBBP6. To test this possibility, we measured the half-life of c-Jun and c-Fos mRNAs after RBBP6 KD by Actinomycin D chase (Figure S4B). No significant differences between siCNT- and siRBBP6-treated cells were observed. Similar results were obtained using a tet-inducible beta-globin mRNA-based assay in which the c-Fos ARE element was introduced in the globin 3'UTR or the whole c-Jun 3'UTR was inserted in place of the original 3'UTR. No change in stability was observed for either transcript following RBBP6 KD and subsequent removal of tet (Figure S4C).

RBBP6 KD impairs 3' end processing of ARE-containing transcripts

We next examined whether 3' end formation of the AU-rich transcripts was impaired after RBBP6 KD. For this we used RT-qPCR with primers spanning the poly(A) site of AU-rich mRNAs to detect possible 3' cleavage defects in siRBBP6- compared to siCNT- treated cells (if more than one poly(A) site was present, the most distal one was selected). Results were normalized to an internal region in each gene so that any effect that RBBP6 KD might have on transcription would not influence the results. Figure 6E shows that the AU-rich mRNAs were indeed less efficiently cleaved after siRBBP6. No significant differences were

observed for actin and GAPDH mRNAs, indicative of at least some specificity in the response to RBBP6 KD.

Defects in 3' processing are known to be coupled to exosome-mediated degradation of mRNAs (Hilleren et al. 2001; Milligan et al. 2005; Kazerouninia et al. 2010). We therefore suspected that the observed decreased accumulation of ARE-containing transcripts could be a consequence of exosomal degradation of the uncleaved transcripts. To test this idea we carried out a double KD, of RBBP6 and a catalytic subunit of the exosome, Dis3. Figure 6F shows a western blot illustrating KD efficiency and Figure 6G presents the RT-qPCR analysis, normalized to siCNT. While RBBP6 KD alone again decreased the abundance of AU-rich RNAs (compare the blue bars to the red bars), double KD of RBBP6 and Dis3 restored AU-rich RNA accumulation, with the exception of Rab3b, to levels detected in siCNT-treated cells (compare blue and green bars). The exosome has two catalytic subunits, Dis3 and Exosc10 (also known as Rrp6) that have both overlapping and specific roles in degrading distinct classes of substrates (Gudipati et al. 2012). Consistent with this, we obtained similar results for c-Fos, c-Jun and Fn1 transcripts following RBBP6/Exocs10 double KD, but not for Bcas, Rab3b and Rbl2 (Figure S4D).

Together, our results show that variations in the levels of RBBP6 can influence 3' cleavage and mRNA abundance in vivo, ARE-containing transcripts being especially sensitive.

Increased RBBP6 iso3 expression down-regulates ARE-rich mRNA expression and 3' processing

Our in vitro data provided evidence that RBBP6 iso3 competes with RBBP6 and thereby reduces 3' cleavage efficiency. If iso3 functions similarly in vivo, then iso3

overexpression should repress ARE-containing mRNA accumulation by inhibiting 3' cleavage, similar to the effects brought about by RBBP6 KD. We examined this by transfecting MCF-7 cells with HA vector alone or increasing amounts of a plasmid encoding HA-iso3, and then measuring 3' cleavage of endogenous c-Fos RNA, analogous to what was done in Figure 6E. Figure 7A shows a WB with anti-HA antibody, to visualize the transfected iso3, and anti-DWNN antibody, to detect endogenous and transfected iso3 protein levels. The amount of unprocessed c-Fos transcript indeed accumulated in a dose-dependent way relative to the increasing amount of iso3 (shown by RT-qPCR in Figure 7B). Significantly, as in the case of RBBP6 KD, this was accompanied by a dose-dependent decrease in total c-Fos RNA accumulation, as normalized to gapdh (Figure 7C). Figure S5 shows that processing of gapdh and actin transcripts was not as significantly affected by overexpression of iso3. We then confirmed by RT-qPCR that most of the AU-rich transcripts analyzed in Figure 6, but not the controls actin and gapdh, were cleaved less efficiently following expression of HA-iso3 (Figure 7D; note that inhibition was less than achieved by siRNA, which may reflect the lower transfection efficiency of the plasmid vector compared to siRNAs). Also, as expected from the defects in 3' cleavage, and consistent with the effects of KD analyzed above, accumulation of the AU-rich mRNAs was repressed by HA-iso3 expression while actin mRNA was not significantly downregulated (Figure 7E).

Discussion

In this study we used biochemical and global analyses to characterize the function of RBBP6. Our results showed that RBBP6 is essential for efficient 3' cleavage in vitro, and interacts with CstF and CPSF. These findings extend a previous proteomic study in which RBBP6 was found to associate with the 3' processing complex (Shi et al. 2009). Binding is

especially strong to CstF64, although whether this reflects direct interaction with CstF64 is not known. RBBP6 does not appear to be a core CstF subunit (Takagaki et al. 1990), but may serve to help link CstF and CPSF, as was suggested for yeast Mpe1 (Vo et al., 2001). In addition, we described a novel mechanism to regulate 3' processing, in which a truncated RBBP6 isoform produced by alternative RNA processing (iso3) competes with the functional protein to control cleavage efficiency both in vitro and in vivo. Finally, we show that RBBP6 levels can control APA, affecting accumulation of specific target transcripts. Below we discuss how these and other properties of RBBP6 contribute to 3' processing and regulation of gene expression.

Studies of RBBP6 to date have focused mostly on its RB- and p53-binding domains; this likely reflects how the protein was initially discovered (Sakai et al. 1995; Simons et al. 1997). However, these domains are found exclusively in mammalian homologues of the protein, while the first three domains are present in all eukaryotes. We show that these N-terminal domains are necessary for pre-mRNA cleavage, supporting the idea that 3' processing is the primary role of RBBP6, and that the long C-terminal domain linking the protein to cell cycle pathways was added later in evolution. In fact, the complex domain composition of RBBP6 suggests that the protein plays roles in multiple cellular events, perhaps functioning to integrate such pathways with pre-mRNA processing. Indeed, RBBP6 has been reported to have roles in cell proliferation and differentiation. For example, its expression is repressed in terminally differentiated cells, as shown by decreased RBBP6 mRNA levels in 3T3 cells that undergo the terminal step in adipocyte differentiation (Witte and Scott 1997). RBBP6 protein levels are also reduced in reversibly quiescient NIH/3T3 cells, but strongly induced again following serum stimulation (Gao et al. 2002). Moreover,

similar to other 3' processing factors, RBBP6 mRNA is upregulated after reprogramming germ cells into iPS cells (Ji and Tian 2009).

Consistent with the fact that RBBP6 expression changes according to the state of cell proliferation and differentiation, regulating the level of RBBP6 isoforms appears to be important during tumorigenesis. RBBP6 expression levels are increased in tumors such as those of the esophagus (Yoshitake et al. 2004), lung (Motadi et al. 2011) and colon (Chen et al. 2013). In contrast to the behavior of the full-length protein, also known as isoform 1 (iso1), iso3 has been shown to be downregulated in several human tumors, such as oesophageal, hepatocellular and colon cancers (Mbita et al. 2012). The association of different splice variants of a protein with normal versus tumor cells is not novel; a few well-studied examples include pyruvate kinase M, Fas and Bcl-x (reviewed in David and Manley 2010). The properties of iso1 and iso3 suggest that RBBP6 might be another example of such a protein. But our results suggest a novel mechanism for how differences in the ratio between these two isoforms could help control cell proliferation. In this model (Figure 7C), alterations in the levels of the isoforms modulates 3' processing efficiency, and thereby APA, by affecting the competition between iso1 and iso3. In cancer cells, there is less of the inhibitory iso3 and more functional iso1, which together results in enhanced 3' processing activity. Such increased activity is expected to favor use of proximal poly(A) sites, leading to shorter 3' UTRs, a known property of cancer cells that contributes to activation of certain oncogenes (Mayr and Bartel 2009). Our data further suggest this includes proteins such as c-Fos and c-Jun, as mRNAs encoding these proteins are among the strongest targets of RBBP6. It is possible that interactions between iso1 and p53 or Rb can

also influence RBBP6 activity in 3' processing, although future experiments are required to address this.

We have shown that the two RBBP6 isoforms have opposing functions in 3' processing, which provides a novel mechanism of regulating mRNA processing. Interestingly, iso1 and iso3 expression, in addition to being inversely correlated in several cancers, is also inversely regulated during differentiation of C2C12 cells, such that iso3 is up-regulated and iso1 down-regulated (Ji et al. 2009). How this alternative processing is regulated is an important question. The poly(A) site used in iso3 production is found in intron 3 and its usage might thus be determined by a competition between splicing and polyadenylation, which could be modulated in several ways. For example, it has been shown that U1 snRNP protects transcripts from premature cleavage and polyadenylation (Kaida et al. 2010) and has a role in regulating the usage of an intronic poly(A) site in CstF77 (Luo et al. 2013) and perhaps more generally in cancer (Berg et al., 2012). It is therefore possible that U1 snRNP might have a similar function in regulating usage of the intronic RBBP6 poly(A) site. Alternatively, regulation could be achieved by changes in intron 3 splicing efficiency, consistent with well-documented changes in splicing that occur during differentiation and disease (David and Manley 2010; Singh and Cooper 2012).

The DWNN domain, which constitutes the N-terminal 81 amino acids of both RBBP6 isoforms, is intriguing. NMR studies showed that this domain adopts a ubiquitin-like fold structure (Pugh et al. 2006). Although it has not yet been shown that it is indeed capable of covalently attaching to proteins, it possesses two of the conserved lysines capable of isopeptide bond formation and the di-glycine motif that are characteristic of ubiquitin-like

proteins. It is tempting to imagine that if such post-translational modification exists, it might play a role in the inhibition of 3' processing we have documented for iso3.

Our results showing global changes in APA following KD of RBBP6 support its role in 3' processing regulation. Other studies have shown that decreased expression of 3' processing factors often correlates with global 3'UTR lengthening. For example, most of the core polyadenylation factors were found to be downregulated in differentiated embryonic tissues as compared to induced pluripotent stem (iPS) cells and this correlated with global 3'UTR lengthening (Ji and Tian 2009). The same was observed during differentiation of C2C12 myoblasts into myotubes (Ji et al. 2009). Since global downregulation of 3' factors correlates with 3'UTR lengthening, we would expect that knockdown of individual 3' processing components would lead to increased usage of distal poly(A) sites, as we found after KD of RBBP6. However, the picture is more complex than this. While depletion of CstF64 (together with its paralog CstF64τ) leads to increased relative use of distal poly(A) sites (Yao et al. 2012), KD of the cleavage factor CFI-68 causes a general shortening of 3'UTRs (Gruber et al. 2012). Although CFI-68 is not a core factor of the 3' processing machinery and might function by a different mechanism, these data indicate that changes in relative abundance of a single 3'-end processing factor can modulate the length of 3'UTRs.

We also observed a general downregulation of gene expression after RBBP6 KD. To our knowledge, other examples of transcript levels after KD of a core 3' processing factor in human cells have not been reported. However, previous studies in yeast showed that generation of aberrant transcripts, either by mutating PAP (Milligan et al. 2005) or the yeast splicing factor PRP2, the homolog of human U2AF (Bousquet-Antonelli et al. 2000),

caused exosome-mediated degradation of the unprocessed RNAs and reduced levels of mRNAs in these cells. Here we show for the first time that reducing the efficiency of 3' cleavage in human cells, either by decreasing the level of RBBP6 or increasing the level of its inhibitory isoform, iso3, leads to degradation of the unprocessed transcripts by the exosome.

Why are AU-rich genes particularly downregulated by RBBP6 KD? We suggest several possibilities. In a first scenario, this specificity is mediated by binding of RBBP6 to AU-rich transcripts; in this case when RBBP6 is absent a destabilizing factor may bind to the same sequences and cause degradation of these RNAs. However, we were unable to see any differences in the half-lives of c-Jun and c-Fos mRNAs and we therefore do not favor this hypothesis. A second hypothesis is that ARE-containing transcripts have specific cis-elements around their poly(A) sites, or the poly(A) sites might all be weak, which renders them particularly susceptible to RBBP6 KD. However, extensive analysis of these poly(A) sites failed to reveal such features. Another possibility is that ARE-containing mRNAs inherently have relatively short half-lives and, consequently, differences in their expression levels would be achieved more rapidly and be more readily detected in the microarray data. If this later option is correct, then KD of any 3' processing factor could lead to preferential down regulation of ARE-containing transcripts. This hypothesis is currently difficult to evaluate because all the available data after KD of other 3' processing factors (see above for examples) was not analyzed by microarray, but by RNA sequencing techniques, which were not able to detect AU-rich transcripts that are expressed at low levels. In any event, our data indicate that ARE-containing transcripts are especially sensitive to RBBP6 (and perhaps other poly(A) factor levels). As many of these mRNAs

encode proteins implicated in cancer and cell proliferation, this suggests another mechanism, in addition to 3' UTR shortening (Mayr and Bartel, 2009), by which alterations in 3' end formation contribute to cell transformation

In conclusion, we have shown that RBBP6 is a new and functionally important component of the 3' processing machinery. Our studies also defined a new mechanism for regulating 3' processing efficiency by competition between negative and positive acting RBBP6 isoforms. Finally, we showed that alterations in RBBP6 levels in cells can affect not only APA but also the abundance of specific transcripts. These and other features of RBBP6 indicate that future studies on its mechanism of action should be informative.

Experimental Procedures

In vitro 3' processing assays

^{32}P-labeled SVL full-length or pre-cleaved RNA substrates were prepared as described previously (Ryner et al. 1989). For 3' cleavage assays, reaction mixtures consisted of 40% NE, 0.5ng labeled RNA, 0.25U RNasin (Promega), 1mM 3'dATP (Trilink), 2.5% polyvinyl alcohol (PVA), 20mM creatine phosphate (Sigma), 8mM Tris (pH 7.9), 10% glycerol, 25mM Ammonium Sulfate, 0.2mM DTT, 0.2mM PMSF. Polyadenylation assays contained the same reagents, with the omission of 3'dATP and addition of 1mM $MgCl_2$ and 1mM ATP. Reaction mixtures were incubated at 30°C for up to 90 mins, followed by proteinase K treatment, phenol/chloroform extraction, ethanol precipitation and separation on 6% urea-acrylamide gels.

Immunoprecipitation and Western blot

Following transfection, cells were collected in two packed cell volume (PCV) of lysis buffer (10mM Tris pH 7.4, 150mM NaCl or 500mM where noted, 0.5%NP40, 0.25%Sodium deoxycholate, 0.5mM EDTA and inhibitors) and lysed using a syringe. The lysate was centrifuged 15mins at maximum speed at 4 °C and 2 volumes of lysis buffer were then added to the supernatant. Pre-clearing of the supernatant was carried out for 30 min at 4 °C with IgG agarose beads and followed by IP with anti-flag M2 agarose for 3 hrs at 4 °C. Where noted, 100ug/ml of RNAseA was added for 15 min at 30°C prior to adding the antibody. Beads were then washed three times in wash buffer (10mM Tris pH7.4, 300mM NaCl, 1%NP40, 0.5% Sodium deoxycholate, 1mM EDTA and inhibitors) and IPed proteins eluted from beads using 100ug/ml 3xFLAG peptide. Proteins were resolved on SDS-PAGE and westerns were carried out using the indicated antibodies: anti-Flag antibody (SIGMA), anti-actin (SIGMA), anti-Rbbp6 (Santa Cruz), anti-Dis3 and anti-Exosc10 (Novus), anti c-Jun (Santa Cruz). Anti-CPSF100 and anti-CstF64 (Takagaki et al. 1990) were made in our laboratory and all other antibodies for 3' processing factors were from Bethyl laboratories. Anti-DWNN is a monoclonal antibody from the laboratory of Dr. David Pugh (University of the Western Cape, South Africa). As secondary antibodies we used HRP-conjugated anti-mouse or anti-rabbit (Sigma). The signal was detected using the ECL western blotting system from GE Healthcare.

Gel shift assays

Gel shift assays were carried out by incubating the indicated amounts of recombinant proteins with ^{32}P-labelled SVL RNA in a buffer containing 2ug/ml heparin, 20mM Tris pH 7.9, 0.2mM EDTA, 20% glycerol and 250mM NaCl. After 10 mins incubation at 30 °C,

sample were loaded on a 5% non-denaturing poly(A)crylamide gel. The gel was then dried and analyzed by Phosphorimager (Molecular Dynamics Storm 860).

RT-PCR

Real-time PCR was performed in 96 well plates with Maxima Sybergreen (Roche) using StepOnePlus (Applied Biosystems). RNA extraction was carried out with TRIZOL (Invitrogen) followed by DNAseI treatment (Fermentas). cDNA was produced with Maxima reverse transcriptase (Fermentas) following the manufacturer's protocol. In order to quantify the RNA expression level, cDNA was amplified using primers for the indicated genes (sequences available upon request) and qPCR data analyzed by delta delta ct method normalizing to the gapdh gene and siCNT. For calculating the relative cleavage efficiency after KD or double KD, cDNA was amplified using primers that span the poly(A) site of each of the indicated genes (the last poly(A) site was used if more than one were present) and data analyzed by normalizing to the value of internal primers for each gene and siCNT.

APA analysis for 3'READS

For 3'READS data, the relative abundance of a poly(A) isoform was defined as the fraction of PASS reads supporting the poly(A) over all PASS reads supporting the gene. To study APA regulation, we compared poly(A) isoform abundance between the RBBP6 KD sample and the siRNA control sample. For analysis of APA events in the 3'-most exons, we compared the top two poly(A) isoforms with highest abundance. For analysis of APA events in upstream regions, we compared the summed abundance for all poly(A)s in upstream regions with that for all poly(A)s in the 3'-most exon. Difference in abundance >5% and *P* value <0.05 (Fisher's exact test) were used as the cutoff to select significantly regulated APA events.

Supplemental Experimental Procedures include additional information on the methods used.

Acknowledgements

We would like to thank Dr. Orit Rozenblat-Rosen for sharing antibodies to 3' processing factors and Dr David Pough for the anti-DWNN antibody. We also would like to thank our lab members for helpful discussions. Tian lab support: GM084089. Manley lab support: GM28983.

Supplemental Experimental Procedures

Cell culture and transfections

HeLa, 293T and MCF-7 cells were cultured in DMEM with 10% fetal bovine serum. siRNA (50 nM) against RBBP6 (CGAAAGAAGAAUAUACUGA) or non-targeting control were transfected with Lipofectamine RNAimax (Invitrogen) and NE were made 72hrs post-transfection (as described in Kleiman and Manley 2001). RNA was extracted 48 or 72hrs post-transfection as indicated. If double KD was carried out, we first transfected siRNA against Exosc10 (20 nM, CAUUAAGGAUCGAAGUAAA) or against Dis3 (20nM, AGGUAGAGUUGUAGGAAUA) for 24 hours, we then tranfected siRNA against RBBP6 and waited 48 more hours.

Lipofectamine 2000 (Invitrogen) was used for transient transfection of Flag or HA tagged RBBP6 constructs into 293T cells; cells were collected 24 hrs post-transfection for IP. Lipofectamine LTX (Invitrogen) was used for transient transfection of HA-iso3 in MCF-7 cells; cells were collected 48 hrs post-transfection for westerns.

Plasmids and protein purification from bacteria

RBBP6-N and its truncations were amplified from HeLa cDNA and cloned in p3XFlag-CMV14 using HindIII and XbaI for mammalian expression or in pRSETC using XhoI and KpnI for expression from bacteria. RBBP6 iso3 was amplified from HeLa cDNA and cloned into pCMV-HA for mammalian expression using BglII and NotI or into pRSETC using XhoI and KpnI. For protein expression in *E. coli* the plasmids were transformed into BL21 bacteria and Nickel-NTA-agarose beads (Qiagen) were used for protein purification followed by a second purification with Dynabeads (Invitrogen) to obtain more pure and concentrated protein.

Affymetrix microarray

50nM RBBP6 siRNA or control siRNA was transfected in MCF7 with RNAimax (Invitrogen) for 72 hrs. RNA was purified using the RNeasy kit (Qiagen) followed by on column DNAse treatment (Qiagen).

The GeneChip WT Terminal Labeling and Controls Kit, combined with the Ambion WT Expression Kit were then used and RNA was hybridized to GeneChip Exon 1.0 ST Arrays (Affymetrix) according to standard protocols.

3'READS

Total RNA was processed by the 3' region extraction and deep sequencing (3'READS) method as described in (Hoque et al. 2013a). The reverse sequencing protocol was used, which generates reads corresponding to the antisense strand of transcript. Data analysis was carried out as previously described (Hoque et al. 2013a). Briefly, we first removed 5' adapter sequences from reads and reads with length < 15 nt after this step were discarded. We then mapped reads against the hg19 genome sequence using bowtie 2 (version

2.1.0)(Langmead and Salzberg 2012) with the following setting: "--local -5 4". We used only reads with mapping quality score (MAPQ) ≥ 10, and required mismatches to be ≤ 5% of the read. Reads with ≥ 2 unaligned Ts at the 5' end are called poly(A) site supporting (PASS) reads, which were used to identify poly(A)s. Poly(A)s located within 24 nt from each other were clustered, but each cluster did not span >48 nt. Poly(A)s mapped to the genome were further assigned to genes, using gene models defined by RefSeq, Ensembl and UCSC Known Gene databases. The 3' ends of the gene models were extended by 4 kb to include downstream poly(A)s, but the extension did not go beyond the transcription start site of the downstream gene. To reduce false poly(A)s, we further required 1) the number of PASS reads for a poly(A) was ≥ 5% of all PASS reads for the gene; and 2) detected in at least two samples. Poly(A)s were separated into different types based on the gene model.

Gene expression analysis

Raw data from the Affymetrix GeneChip Human Exon 1.0 ST Array were normalized by the RMA method in the Affymetrix Power Tools (APT) program and probe sets with detection above background (DABG) *P* value < 0.05 in at least one sample group were used for further analysis. To eliminate the potential effect of APA in 3'UTR on expression analysis of a gene, we only used probesets mapped to the coding sequence (CDS) to represent the expression level of a gene. We used t-test *P* value < 0.05 or $\log_2$(ratio) greater than 1× standard deviation of $\log_2$(ratio) of all genes to select significantly regulated genes.

For 3'READS data, all PASS reads for a gene were summed to represent the expression level of the gene. The read number of a gene was normalized to the total number of PASS reads from the sample. The resultant value, reads per million (RPM), was used to represent the expression level of a gene. To examine gene regulation, we used the set of genes

showing no expression change in the microarray data (<1% difference in probeset intensity) as reference, and compared other genes to the reference set using the Fisher's exact test. *P* value<0.01 and fold change >1.3 were used to select significantly regulated genes.

Cis-elements analysis

The 3'UTR sequence of genes were based on the last poly(A) site identified using our 3'READS libraries, or RefSeq-annotated 3' end if no reads were available. To identify cis elements associated with regulated genes, we enumerated the number of occurrence of each 5-mer in 3'UTRs of genes that were up-regulated, down-regulated, or not significantly changed based on the microarray analysis. We then used the Fisher's exact test to examine the significance of association of each 5-mer with each 3'UTR group.

3' processing complex formation assay

Cleavage reactions with SVL substrate were set up as described above using NE from siRNA treated cells. Samples were incubated for 15 mins at 30°C followed by 10 minutes incubation on ice with 5 ug/ul heparin. The RNA-protein complexes were then resolved on 1.5% low-melting-point agarose gel. The gel was dried and exposed to a phosphorimager screen.

mRNA half life of endogenous fos and jun

MCF-7 cells were transfected with 50nM siRBBP6 or siCNT for 48hrs and then treated with 5ug/ml ActinomycinD for the indicated times. Total RNA was extracted with TRIZOL, followed by DNAse treatment, reverse transcription and RT-PCR analysis as above.

mRNA half life of Tet-inducible beta globin transcript

293-tTA cells were transfected with 50nM siCNT or siRBBP6 in a 6-well plate using RNAiMAX (Invitrogen) and the medium was changed 24 hrs post-transfection. 48 hrs after KD, DNA transfection was carried out using Lipofectamine2000 (Invitrogen) with 4ug of plasmids containing either control beta-globin gene, beta-globin gene with ARE of c-Fos inserted in the 3'UTR, or the entire 3'UTR of c-Jun inserted in place of the natural beta-globin 3'UTR . 6 hrs after transfection, cells were plated in 12-well plates with 0.1ug/ul tetracycline. The following day, tetracycline was removed for 4 hours to allow transcription of the beta-globin gene. Following addition of 0.5ug/ml tetracycline (to stop transcription), cytoplasmic RNA was collected at the indicated time points using NP40 lysis buffer as above.

IPA analysis

The disease and biological function terms enriched for down-regulated genes were generated by IPA (Ingenuity® Systems, www.ingenuity.com).

References:

Barabino SM, Hubner W, Jenny A, Minvielle-Sebastia L, Keller W. 1997. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes & Development* **11**: 1703-1716.

Boucher L, Ouzounis CA, Enright AJ, Blencowe BJ. 2001. A genome-wide survey of RS domain proteins. *RNA* **7**: 1693-1701.

Bousquet-Antonelli C, Presutti C, Tollervey D. 2000. Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102**: 765-775.

Chen J, Tang H, Wu Z, Zhou C, Jiang T, Xue Y, Huang G, Yan D, Peng Z. 2013. Overexpression of RBBP6, alone or combined with mutant TP53, is predictive of poor prognosis in colon cancer. *PloS One* **8**: e66524.

Chibi M, Meyer M, Skepu A, DJ GR, Moolman-Smook JC, Pugh DJ. 2008. RBBP6 interacts with multifunctional protein YB-1 through its RING finger domain, leading to ubiquitination and proteosomal degradation of YB-1. *Journal of Molecular Biology* **384**: 908-916.

David CJ, Manley JL. 2010. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & Development* **24**: 2343-2364.

Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and consequences of alternative polyadenylation. *Molecular Cell* **43**: 853-866.

Di Giammartino DC, Shi Y, Manley JL. 2013. PARP1 represses PAP and inhibits polyadenylation during heat shock. *Molecular Cell* **49**: 7-17.

Elkon R, Ugalde AP, Agami R. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. *Nature Reviews Genetics* **14**: 496-506.

Gao S, Witte MM, Scott RE. 2002. P2P-R protein localizes to the nucleolus of interphase cells and the periphery of chromosomes in mitotic cells which show maximum P2P-R immunoreactivity. *Journal of Cellular Physiology* **191**: 145-154.

Gingerich TJ, Feige JJ, LaMarre J. 2004. AU-rich elements and the control of gene expression through regulated mRNA stability. *Animal Health Research Reviews* **5**: 49-63.

Gruber AR, Martin G, Keller W, Zavolan M. 2012. Cleavage factor Im is a key regulator of 3' UTR length. *RNA Biology* **9**: 1405-1412.

Gudipati RK, Xu Z, Lebreton A, Seraphin B, Steinmetz LM, Jacquier A, Libri D. 2012. Extensive degradation of RNA precursors by the exosome in wild-type cells. *Molecular Cell* **48**: 409-421.

Hilleren P, McCarthy T, Rosbash M, Parker R, Jensen TH. 2001. Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature* **413**: 538-542.

Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013b. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods* **10**: 133-139.

Huang P, Ma X, Zhao Y, Miao L. 2013. The C. elegans Homolog of RBBP6 (RBPL-1) Regulates Fertility through Controlling Cell Proliferation in the Germline and Nutrient Synthesis in the Intestine. *PloS One* **8**: e58736.

Ji Z, Lee JY, Pan Z, Jiang B, Tian B. 2009. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 7028-7033.

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**: e8419.

Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**: 664-668.

Kazerouninia A, Ngo B, Martinson HG. 2010. Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro. *RNA* **16**: 197-210.

Kleiman FE, Manley JL. 2001. The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. *Cell* **104**: 743-753.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Li L, Deng B, Xing G, Teng Y, Tian C, Cheng X, Yin X, Yang J, Gao X, Zhu Y et al. 2007. PACT is a negative regulator of p53 and essential for cell growth and embryonic development. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 7951-7956.

Luo W, Ji Z, Pan Z, You B, Hoque M, Li W, Gunderson SI, Tian B. 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 Genetics* **9**: e1003613.

Mandel CR, Bai Y, Tong L. 2008. Protein factors in pre-mRNA 3'-end processing. *Cellular and Molecular Life Sciences* **65**: 1099-1122.

Mather A, Rakgotho M, Ntwasa M. 2005. SNAMA, a novel protein with a DWNN domain and a RING finger-like motif: a possible role in apoptosis. *Biochimica et Biophysica Acta* **1727**: 169-176.

Mayr C, Bartel DP. 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**: 673-684.

Mbita Z, Meyer M, Skepu A, Hosie M, Rees J, Dlamini Z. 2012. De-regulation of the RBBP6 isoform 3/DWNN in human cancers. *Molecular and Cellular Biochemistry* **362**: 249-262.

Milligan L, Torchet C, Allmang C, Shipman T, Tollervey D. 2005. A nuclear surveillance pathway for mRNAs with defective polyadenylation. *Molecular and Cellular Biology* **25**: 9996-10004.

Moore MJ, Proudfoot NJ. 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**: 688-700.

Motadi LR, Bhoola KD, Dlamini Z. 2011. Expression and function of retinoblastoma binding protein 6 (RBBP6) in human lung cancer. *Immunobiology* **216**: 1065-1073.

Pugh DJ, Ab E, Faro A, Lutya PT, Hoffmann E, Rees DJ. 2006. DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways. *BMC structural biology* **6**: 1.

Ryner LC, Takagaki Y, Manley JL. 1989. Sequences downstream of AAUAAA signals affect pre-mRNA cleavage and polyadenylation in vitro both directly and indirectly. *Molecular and Cellular Biology* **9**: 1759-1771.

Saijo M, Sakai Y, Kishino T, Niikawa N, Matsuura Y, Morino K, Tamai K, Taya Y. 1995. Molecular cloning of a human protein that binds to the retinoblastoma protein and chromosomal mapping. *Genomics* **27**: 511-519.

Sakai Y, Saijo M, Coelho K, Kishino T, Niikawa N, Taya Y. 1995. cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product. *Genomics* **30**: 98-101.

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, 3rd, Frank J, Manley JL. 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Molecular Cell* **33**: 365-376.

Simons A, Melamed-Bessudo C, Wolkowicz R, Sperling J, Sperling R, Eisenbach L, Rotter V. 1997. PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb. *Oncogene* **14**: 145-155.

Singh RK, Cooper TA. 2012. Pre-mRNA splicing in disease and therapeutics. *Trends in Molecular Medicine* **18**: 472-482.

Takagaki Y, Manley JL, MacDonald CC, Wilusz J, Shenk T. 1990. A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes & Development* **4**: 2112-2120.

Tian B, Manley JL. 2013. Alternative cleavage and polyadenylation: the long and short of it. *Trends in Biochemical Sciences* **38**: 312-320.

Vo LT, Minet M, Schmitter JM, Lacroute F, Wyers F. 2001. Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. *Molecular and Cellular Biology* **21**: 8346-8356.

Witte MM, Scott RE. 1997. The proliferation potential protein-related (P2P-R) gene with domains encoding heterogeneous nuclear ribonucleoprotein association and Rb1 binding shows repressed expression during terminal differentiation. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 1212-1217.

Yao C, Biesinger J, Wan J, Weng L, Xing Y, Xie X, Shi Y. 2012. Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 18773-18778.

Yoshitake Y, Nakatsura T, Monji M, Senju S, Matsuyoshi H, Tsukamoto H, Hosaka S, Komori H, Fukuma D, Ikuta Y et al. 2004. Proliferation potential-related protein, an ideal esophageal cancer antigen for immunotherapy, identified using complementary DNA microarray analysis. *Clinical Cancer Research* **10**: 6437-6448.

Figure Legends

Figure 1. RBBP6 is an essential 3' processing factor. (A) Western blot with the indicated antibodies of HeLa NE following transfection with siRNA against RBBP6 (siRBBP6) or a non targeting sequence (siCNT). **(B)** 3' cleavage and polyadenylation assays were carried out using internally ^{32}P labeled RNA substrate and HeLa NE made after siRNA treatment. RNAs were purified, resolved by denaturing PAGE, and visualized by autoradiography. Positions of precursor and products are indicated. **(C)** Schematic diagram of full length RBBP6, RBBP6-N, ΔDWNN, DWNN only and iso3. **(D)** 3' cleavage assay as in Figure 1C but adding increasing amounts of RBBP6-N that was previously purified from E.coli.

Figure 2. RBBP6-N interacts with CstF and CPSF. (A) 293T cells were transfected with a Flag empty vector or Flag-RBBP6-N and cell extracts in a buffer containing 150mM NaCl were used to IP with anti-Flag followed by western blot with the indicated antibodies **(B)** IP was carried out as in Figure 2A but in a buffer containing 500mM NaCl **(C)** HeLa NE were used to IP endogenous CstF64 (with a polyclonal antibody) in the presence of 150mM NaCl, with or without prior incubation of the extract with 100ug/ml RNAseA for 15 min at 30°C. Western blots were carried out with anti-RBBP6 antibody and a monoclonal CstF64 antibody.

Figure 3. The DWNN is required for cleavage activity but does not bind RNA. (A) Gel shift assay with a ^{32}P-labeled RNA (SVL) and RBBP6-N purified from E.coli in the presence of one fold or ten fold concentrations of a non a non specific competitor (tRNA) or a specific competitor (cold SVL) Samples were resolved in a 5% nondenaturing poly(A)crylamide gel. The gel was dried and exposed to a PhosphorImager screen **(B)** 3' cleavage assay as in Figure 1D. 250ng of w.t. or ΔDWNN RBBP6-N purified from E.coli were added to the NE made after KD of RBBP6. The right panel shows a diagram representing the mean of three separate experiments with standard error bars. **(C)** Gel shift assays as in Figure 3A with increasing amounts of the purified indicated proteins.

Figure 4. DWNN binds CstF64 and RBBP6 iso3 competes with iso1 in binding to CstF64 and inhibits cleavage. (A) Co-IP experiment as in Figure 2A but using either w.t., ΔDWNN or DWNN-only constructs of RBBP6-N. **(B)** Co-IP as in Figure 2A but with increasing amounts of HA-tagged RBBP6iso3. **(C)** 3' cleavage assay with increasing amounts of His-RBBP6 iso3 purified from e.coli

Figure 5. RBBP6 regulates APA (A) Schematic of different poly(A) site types analyzed in this study. **(B)** Summary of up- (UP) or down- (DN) regulated APA events, based on the poly(A) type. RNA samples were processed with the 3'READS technique following KD of RBBP6 in MCF-7 cells for 48 hours. **(C)** Analysis of APA in the 3'-most exons. Significant events are colored with red (distal poly(A) isoform relatively upregulated) or blue (proximal poly(A) isoform relatively upregulated). Only two most abundant isoforms for each gene based on 3'READS data were analyzed. **(D)** Analysis of APA in upstream regions. Significant events are colored with red (3'-most poly(A) isoforms relatively upregulated) or blue (upstream region poly(A) isoforms relatively upregulated). **(E)** Significant cis

elements associated with poly(A) sites of regulated isoforms. Two regions were analyzed (-100 to -1 nt and +1 to +40 nt around the poly(A) site). Tetramers and hexamers were examined. Numbers are $-\log_{10}$(P-value) * s, where P-value was derived from the Fisher's exact test, and s is a sign indicating association with upregulation (positive sign) or with downregulation (negative sign). **(F)** Histogram for gene expression changes based on microarray data of MCF-7 cells after siRNA to RBBP6 (siRBBP6) or non-targeting sequence (siCNT); as indicated, more genes are down-regulated. pA is poly(A) site.

Figure 6. KD of RBBP6 leads to downregulation of ARE-containing transcripts. **(A)** Sequence logo of top 50 pentamers associated with the 3'UTRs of down-regulated genes. Pentamer sequences are shown in Figure S3C. **(B)** Cumulative fraction analysis of genes with different numbers of AREs in the 3'UTR. The Kolmogorov–Smirnov (KS) test *P* value for difference in data distribution between genes with AREs and those without is indicated. **(C)** RNA was extracted from MCF-7 after siRBBP6 or siCNT and RT-qPCR was used to calculate the relative amount of the indicated transcripts as normalized to siCNT and gapdh **(D)** Western blot with the indicated antibodies after KD of RBBP6 in MCF-7 cells **(E)** RNA was extracted from MCF-7 after siRBBP6 or siCNT and RT-qPCR was used to calculate the relative amount of the indicated uncleaved transcripts using primers spanning the last poly(A) site of each gene and normalizing to siCNT and an internal probe for each gene **(F)** Western blot with the indicated antibodies after KD of RBBP6 or RBBP6 and Dis3 in MCF-7 cells **(G)** RNA was extracted from MCF-7 after siCNT,siRBBP6 or siRBBP6 together with siDis3. RT-qPCR was used to calculate the relative amount of the indicated transcripts as normalized to siCNT and gapdh.

Figure 7. RBBP6 isoform3 inhibits cleavage of AU-rich mRNAs and reduces their expression level (A) MCF7 cells were transfected for 48 hours with HA or increasing amounts of HA-iso3 and cell lysates analyzed by western blots with the indicated antibodies **(B)** RNA was extracted from MCF-7 cells after transfection with an empty HA vector or increasing amounts of HA-tagged RBBP6 isoform3 (HA iso3). RT-qPCR was used to calculate the relative amount of uncleaved fos transcripts using primers spanning the last poly(A) site normalizing to an internal probe **(C)** RNA was extracted from MCF-7 cells after transfection with an empty HA vector or increasing munts of HA-iso3. RT-qPCR was used to calculate the relative amount of fos as normalized to gapdh and transfection with HA vector **(D)** RNA was extracted from MCF-7 cells after transfection with an empty HA vector or of HA-tagged RBBP6 isoform3. RT-qPCR was used to calculate the relative amount of the indicated uncleaved transcripts using primers spanning the last poly(A) site of each gene. Values were normalized to an internal probe of each gene **(E)** RNA was extracted from MCF-7 cells after transfection with an empty HA vector or HA-iso3. RT-qPCR was used to calculate the relative amount of the indicated transcripts as normalized to gapdh and transfection with HA vector **(F)** RBBP6 iso1 competes with iso3 in binding to CstF64. When iso1 is upregulated and/or iso3 is downregulated such as in cancer cells, iso1 can bind to CstF64 and pre-mRNA 3' processing functions properly; when the opposite is true, for example after KD of iso1 or overexpression of iso3, 3' cleavage is inhibited by binding of iso3 to CstF64, resulting in downregulation of gene expression, especially of ARE containing transcripts.

Supplemental Figure Legends

Supplemental Figure 1. (A) coomassie stain of the indicated proteins purified from E.coli. **(B)** HeLa NE were used to IP endogenous RBBP6. Western blots were carried out with anti-RBBP6 antibody and anti- CstF64 antibody. **(C)** Gel shift assay using ^{32}P labeled, in vitro-transcribed RNA and NE from HeLa cells that were transfected with siRNA to a non-targeting sequence (siCNT) or to RBBP6 (siRBBP6). NEs and RNA were incubated under conditions that allow pre-mRNA 3' processing for the time indicated and then loaded on a 1.5% low-melting-point agarose gel. The gel was then dried and exposed to a PhosphorImager screen. **(D)** Gel shift assay with a ^{32}P-labeled RNA (SVL) and ΔDWNN with or without increasing amounts of (iso3, left panel) or DWNN only proteins (right panel). Samples were resolved in a 5% nondenaturing poly(A)crylamide gel. The gel was dried and exposed to a PhosphorImager screen. **(E)** In vitro 3' cleavage reaction with SVL RNA and NE of HeLa cells made after transfection with an HA empty vector or increasing amounts of HA-iso3.

Supplemental Figure 2. (A) western blot with the indicated antibodies after 48 or 72 hours knockdown with siRBBP6 in MCF-7 cells **(B)** PASS read number for the 3'READS libraries. **(C)** Summary of up- (UP) or down- (DN) regulated APA events, based on the poly(A) type. RNA samples were processed with the 3'READS technique following KD of RBBP6 in MCF-7 cells for 72 hours. **(D)** Analysis of APA in the 3'-most exons after KD for 72 hours. Significant events are colored with red (distal poly(A) isoform relatively upregulated) or blue (proximal poly(A) isoform relatively upregulated). Only two most abundant isoforms for each gene based on 3'READS data were analyzed. **(E)** Analysis of APA in upstream regions after 72 hours KD. Significant events are colored with red (3'-most poly(A) isoforms relatively upregulated) or blue (upstream region poly(A) isoforms

relatively upregulated). Summary of regulated alternative polyadenylation events, based on the poly(A) type. RNA samples were processed with the 3'READS technique following KD of RBBP6 in MCF-7 cells for 72 hours. See Figure 4 for details. (x) PASS read number for the 3'READS libraries.

Supplemental Figure 3. (A) Gene expression regulation analyzed by 3'READS. Data were normalized by genes whose expression did not change based on microarray data **(B)** IPA terms associated with down-regulated genes after RBBP6 KD **(C)** Top 50 5-mers enriched for 3'UTRs of downregulated genes. Significance score (SS) is $-\log_{10}$ (P-value) determined by Fisher's exact test.

Supplemental Figure 4. (A) RNA was extracted from MCF-7 cells after siRBBP6 or siCNT and RT-qPCR was used to calculate the relative amount of the indicated transcripts as normalized to siCNT and gapdh **(B)** MCF-7 cells were subjected to siCNT or siRBBP6 and RNA samples were extracted at the indicated times following treatment with Actinomycin D. RT-qPCR was used to calculate the percentage of c-Jun (left panel) or c-Fos (right panel) mRNAs left, the values were normalized to gapdh **(C)** RNA was extracted from 293 cells stably transfected with an inducible β-globin transgene containing w.t. 3'UTR (left panel) or a 3'UTR containing the ARE of c-Fos (middle panel) or 3'UTR of c-Jun . Cells were either treated with a control siRNA or siRNA to RBBP6. To measure β-globin half-life, RNA samples were extracted at the indicated times following tet removal. **(D)** RNA was extracted from MCF-7 cells after siCNT,siRBBP6 or siRBBP6 together with the catalytic exosome subunit Exosc10. RT-qPCR was used to calculate the relative amount of the indicated transcripts as normalized to siCNT and gapdh (left panel). The right panel shows a western blot of the three conditions described above using the indicated antibodies.

Supplemental Figure 5. (A) RNA was extracted from MCF-7 cells after transfection with an empty HA vector or increasing amounts of HA-iso3. RT-qPCR was used to calculate the relative amount of uncleaved gapdh transcripts using primers spanning the poly(A) site and normalizing to an internal probe **(B)** RNA was extracted from MCF-7 cells after transfection with an empty HA vector or increasing amunts of HA-iso3. RT-qPCR was used to calculate the relative amount of actin as normalized to siCNT and gapdh.

Figure 1

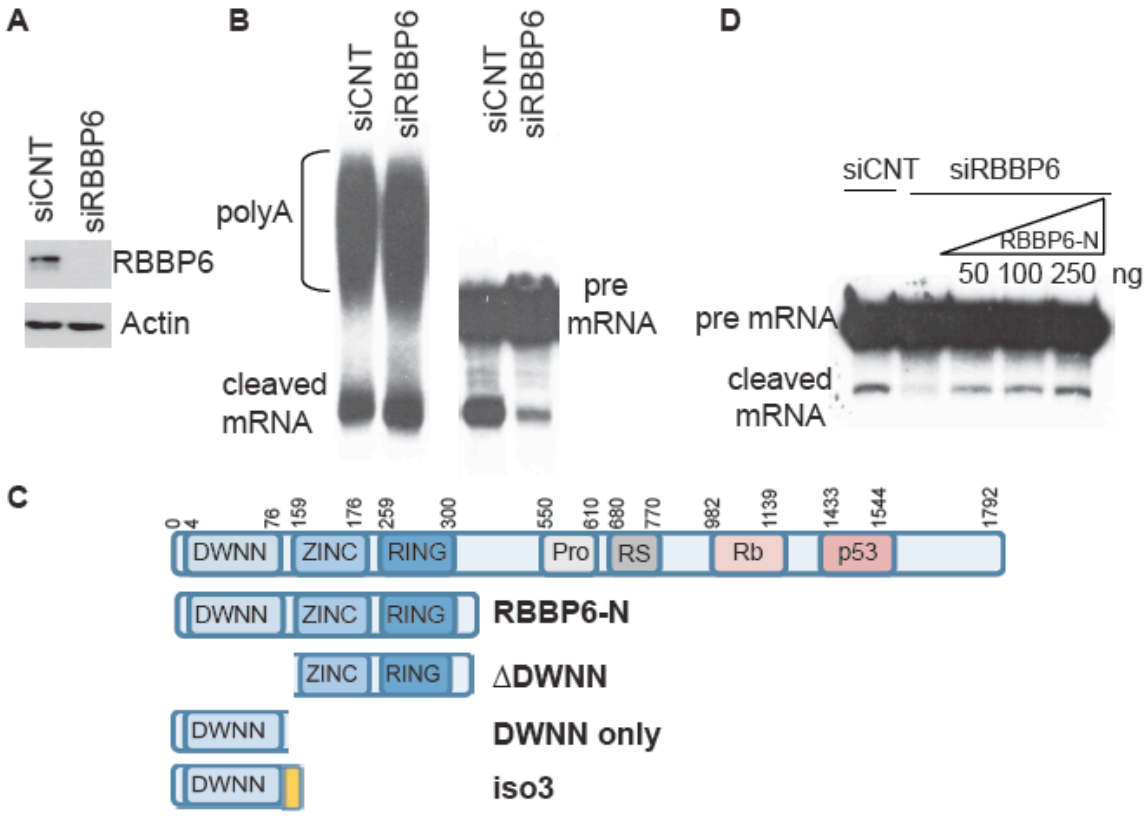

A
siCNT
siRBBP6
RBBP6
Actin
B
siCNT
siRBBP6
siCNT
siRBBP6
polyA
cleaved mRNA
pre mRNA
D
siCNT
siRBBP6
RBBP6-N
50 100 250 ng
pre mRNA
cleaved mRNA
C
0
4
76
159
176
259
300
550
610
680
770
982
1139
1433
1544
1792
DWNN
ZINC
RING
Pro
RS
Rb
p53
RBBP6-N
ΔDWNN
DWNN only
iso3

Figure 2

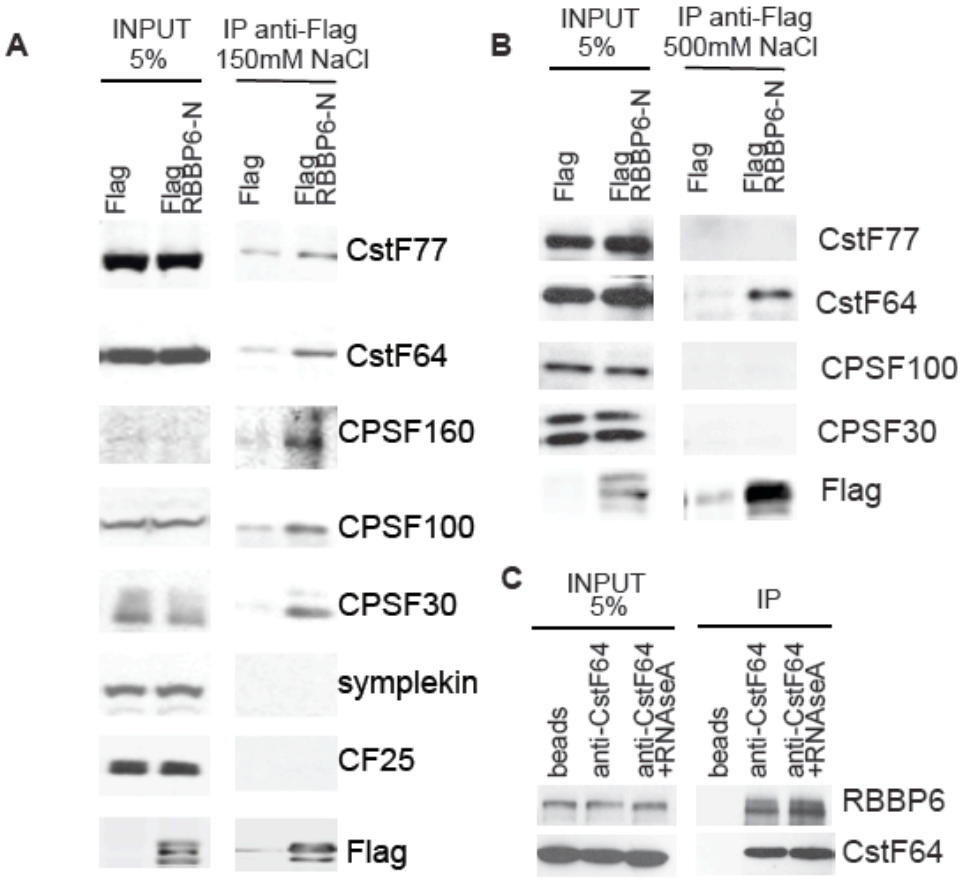

A
INPUT 5%
IP anti-Flag 150mM NaCl
Flag
Flag RBBP6-N
Flag
Flag RBBP6-N
CstF77
CstF64
CPSF160
CPSF100
CPSF30
symplekin
CF25
Flag
B
INPUT 5%
IP anti-Flag 500mM NaCl
Flag
Flag RBBP6-N
Flag
Flag RBBP6-N
CstF77
CstF64
CPSF100
CPSF30
Flag
C
INPUT 5%
IP
beads
anti-CstF64
anti-CstF64 +RNAseA
beads
anti-CstF64
anti-CstF64 +RNAseA
RBBP6
CstF64

Figure 3

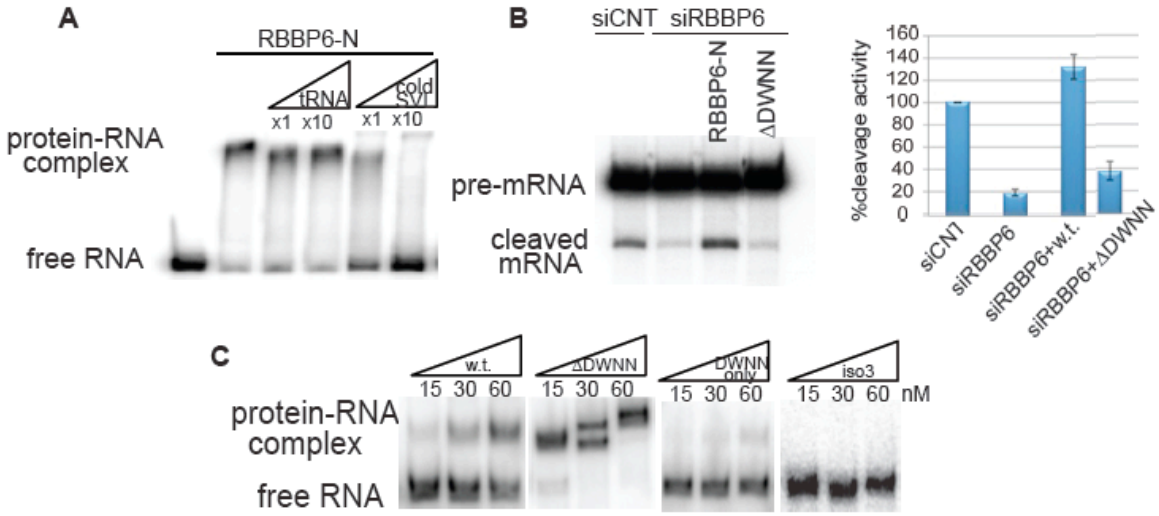

A
RBBP6-N
tRNA
cold SVT
x1 x10 x1 x10
protein-RNA complex
free RNA
B
siCNT siRBBP6
RBBP6-N
ΔDWNN
pre-mRNA
cleaved mRNA
%cleavage activity
160
140
120
100
80
60
40
20
0
siCNT
siRBBP6
siRBBP6+w.t.
siRBBP6+ΔDWNN
C
w.t.
ΔDWNN
DWNN only
iso3
15 30 60 15 30 60 15 30 60 15 30 60 nM
protein-RNA complex
free RNA

Figure 4

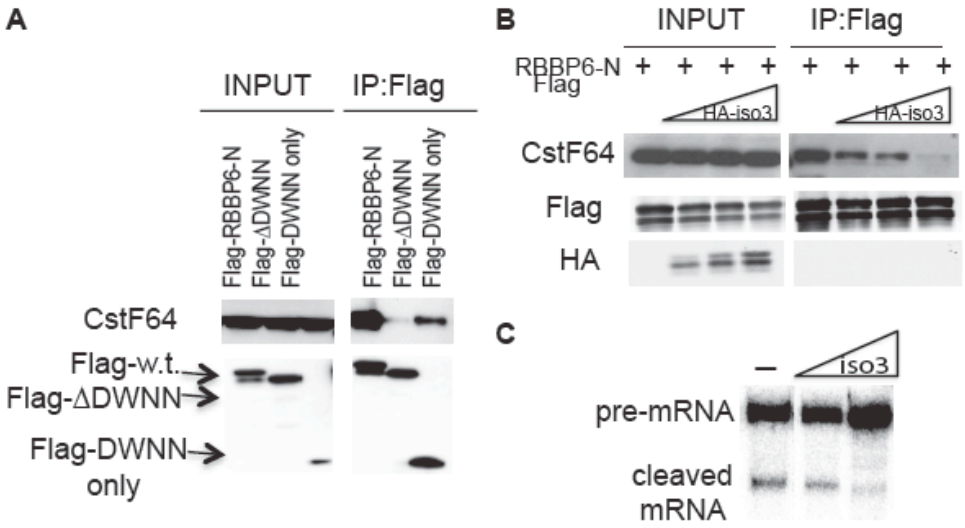

A
INPUT
IP:Flag
Flag-RBBP6-N
Flag-ΔDWNN
Flag-DWNN only
CstF64
Flag-w.t.
Flag-ΔDWNN
Flag-DWNN only
B
INPUT
IP:Flag
RBBP6-N Flag
HA-iso3
CstF64
Flag
HA
C
–
iso3
pre-mRNA
cleaved mRNA

Figure 5

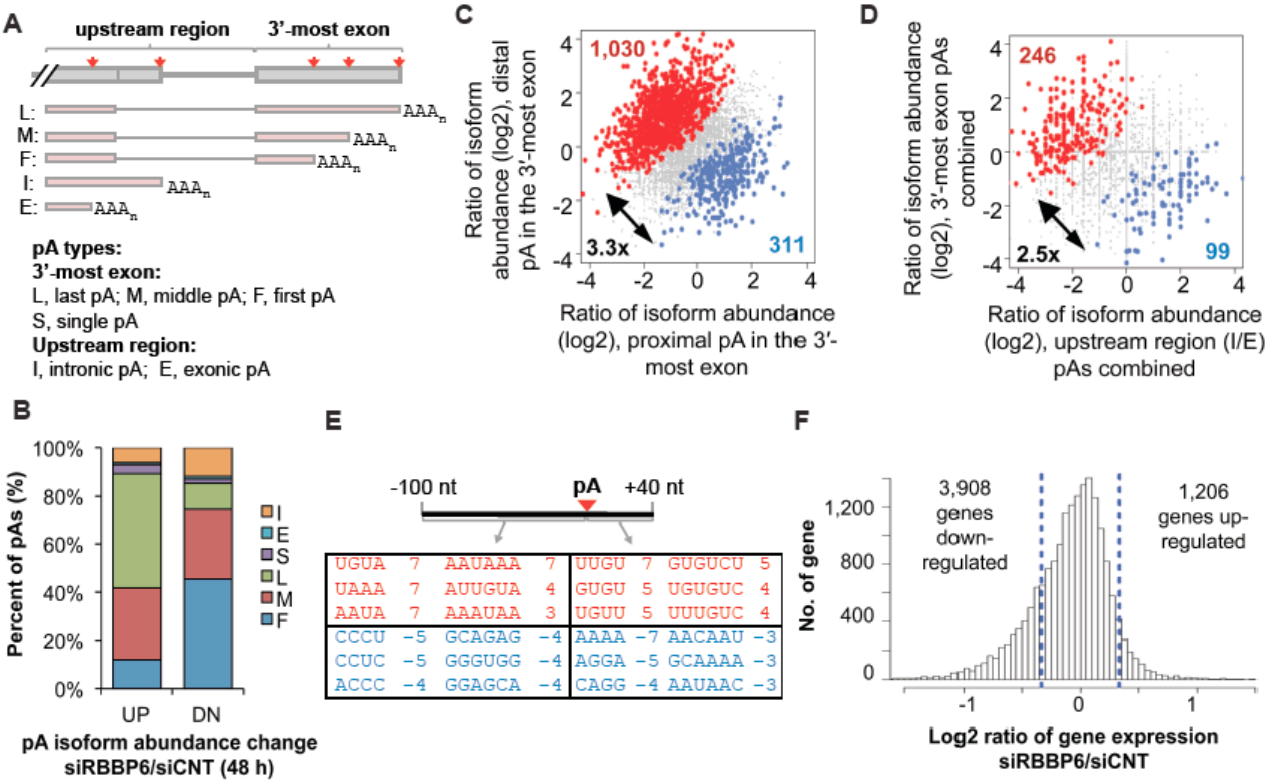

A
upstream region
3'-most exon
L:
M:
F:
I:
E:
AAAn
pA types:
3'-most exon:
L, last pA; M, middle pA; F, first pA
S, single pA
Upstream region:
I, intronic pA; E, exonic pA
C
1,030
311
3.3x
Ratio of isoform abundance (log2), distal pA in the 3'-most exon
Ratio of isoform abundance (log2), proximal pA in the 3'-most exon
D
246
99
2.5x
Ratio of isoform abundance (log2), 3'-most exon pAs combined
Ratio of isoform abundance (log2), upstream region (I/E) pAs combined
B
100%
80%
60%
40%
20%
0%
Percent of pAs (%)
UP
DN
I
E
S
L
M
F
pA isoform abundance change siRBBP6/siCNT (48 h)
E
-100 nt
pA
+40 nt
UGUA 7 AAUAAA 7
UAAA 7 AUUGUA 4
AAUA 7 AAAUAA 3
CCCU -5 GCAGAG -4
CCUC -5 GGGUGG -4
ACCC -4 GGAGCA -4
UUGU 7 GUGUCU 5
GUGU 5 UGUGUC 4
UGUU 5 UUUGUC 4
AAAA -7 AACAAU -3
AGGA -5 GCAAAA -3
CAGG -4 AAUAAC -3
F
3,908 genes down-regulated
1,206 genes up-regulated
No. of gene
1,200
800
400
0
-1
0
1
Log2 ratio of gene expression siRBBP6/siCNT

Figure 6

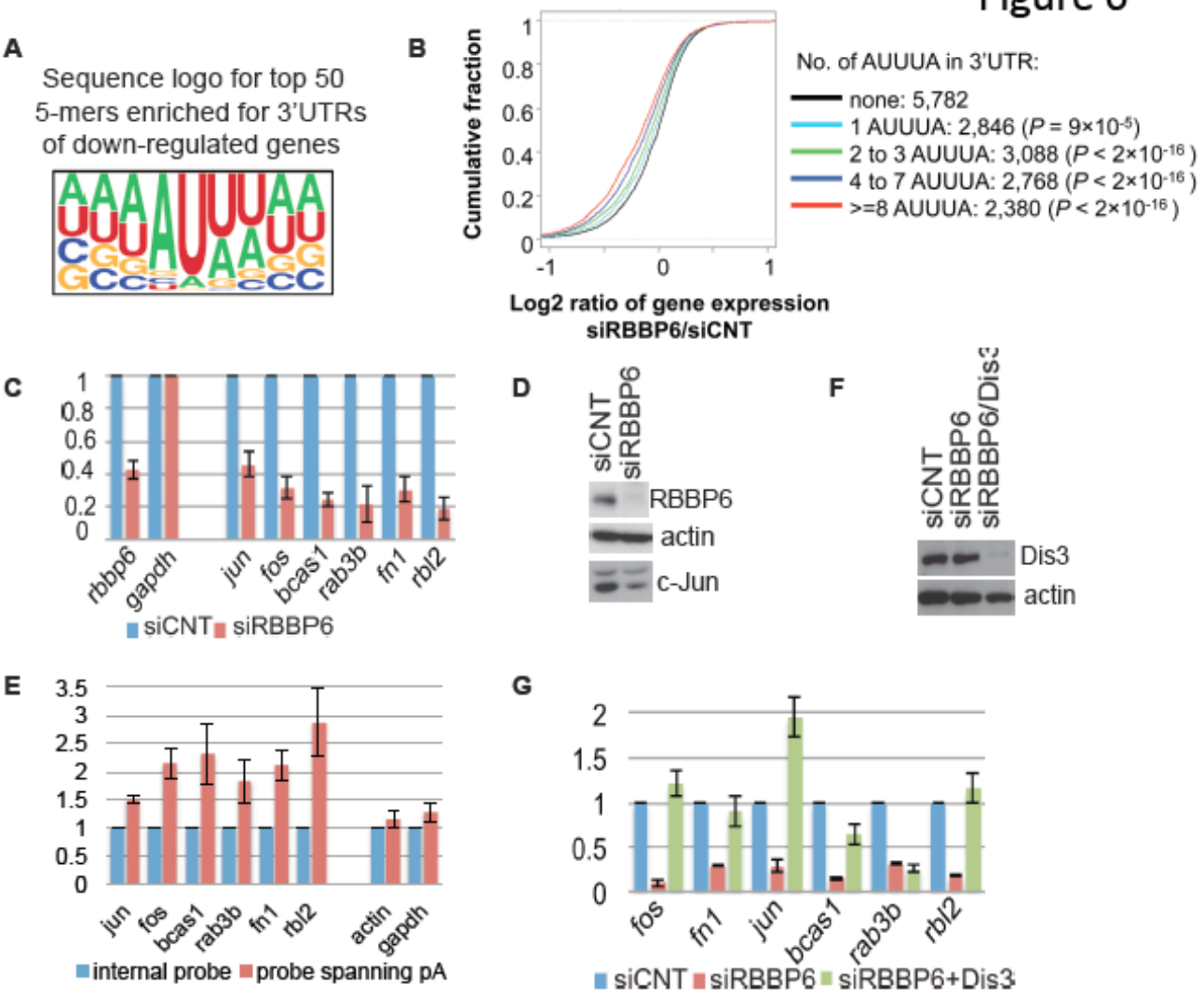

A
Sequence logo for top 50
5-mers enriched for 3'UTRs
of down-regulated genes
B
Cumulative fraction
Log2 ratio of gene expression
siRBBP6/siCNT
No. of AUUUA in 3'UTR:
none: 5,782
1 AUUUA: 2,846 (P = 9×10-5)
2 to 3 AUUUA: 3,088 (P < 2×10-16)
4 to 7 AUUUA: 2,768 (P < 2×10-16)
>=8 AUUUA: 2,380 (P < 2×10-16)
C
rbbp6
gapdh
jun
fos
bcas1
rab3b
fn1
rbl2
siCNT
siRBBP6
D
siCNT
siRBBP6
RBBP6
actin
c-Jun
F
siCNT
siRBBP6
siRBBP6/Dis3
Dis3
actin
E
jun
fos
bcas1
rab3b
fn1
rbl2
actin
gapdh
internal probe
probe spanning pA
G
fos
fn1
jun
bcas1
rab3b
rbl2
siCNT
siRBBP6
siRBBP6+Dis3

Figure 7

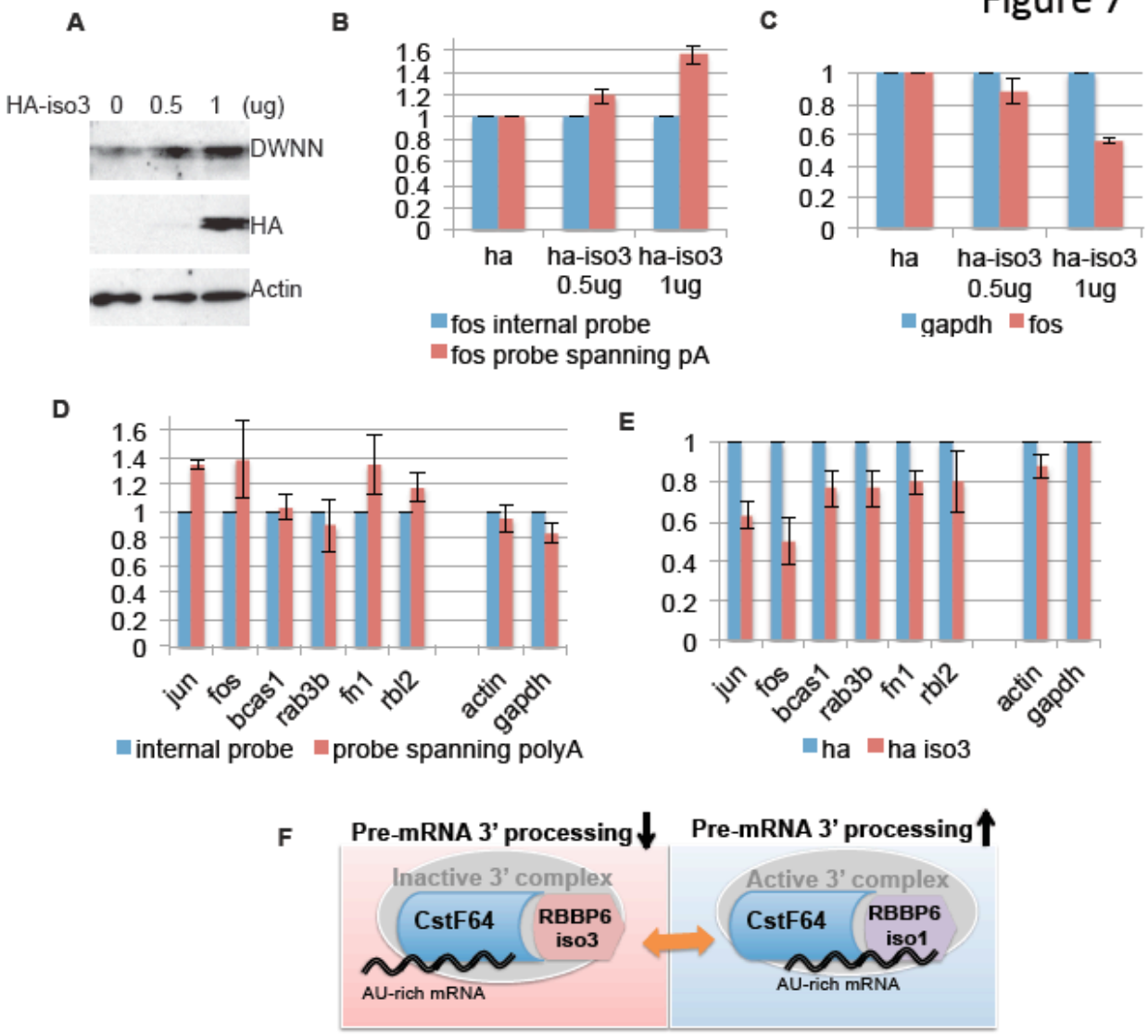
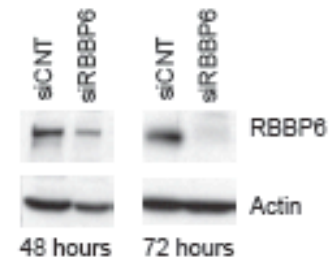
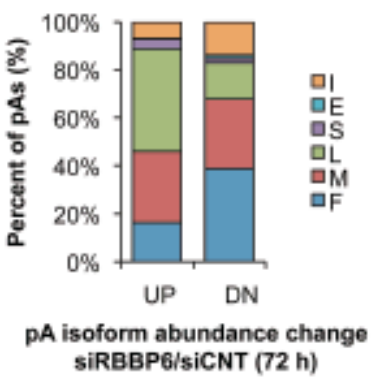

A
HA-iso3 0 0.5 1 (ug)
DWNN
HA
Actin
B
ha
ha-iso3 0.5ug
ha-iso3 1ug
fos internal probe
fos probe spanning pA
C
ha
ha-iso3 0.5ug
ha-iso3 1ug
gapdh
fos
D
jun
fos
bcas1
rab3b
fn1
rbl2
actin
gapdh
internal probe
probe spanning polyA
E
jun
fos
bcas1
rab3b
fn1
rbl2
actin
gapdh
ha
ha iso3
F
Pre-mRNA 3' processing
Pre-mRNA 3' processing
Inactive 3' complex
CstF64
RBBP6 iso3
AU-rich mRNA
Active 3' complex
CstF64
RBBP6 iso1
AU-rich mRNA

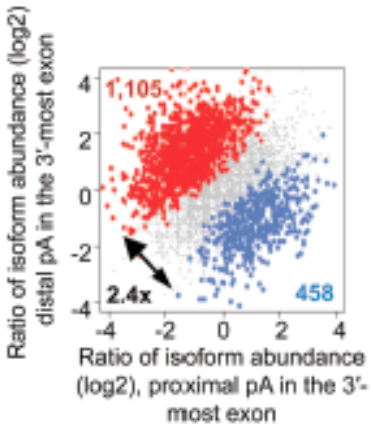
Figure S2



B

Sample	No. of PASS reads
siCNT 48h	2,478,215
siCNT 72h	1,863,263
siRBBP6 48h	2,295,175
siRBBP6 72h	2,677,613





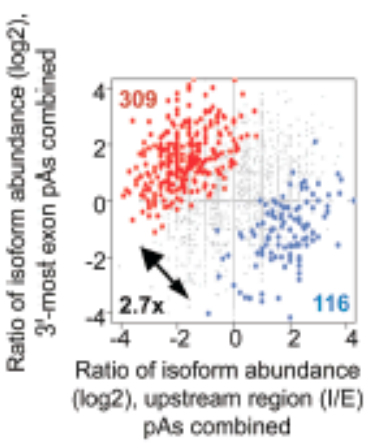
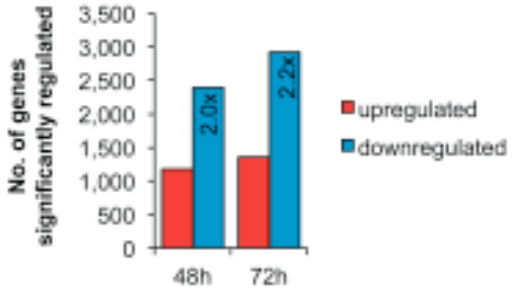


Figure S3

A



C

5-mer	SS	5-mer	SS
ATTTT	215	TATTA	107
TTTTA	213	AATTA	105
TTTAA	205	ATATA	100
AAAAT	194	AAAAA	90
TATTT	182	ATACT	80
ATATT	171	AATAA	73
TTAAA	164	CATTT	73
AATAT	159	TATAC	71
AATTT	157	AAGTA	70
TAAAA	151	AAATG	69
AAATA	144	ATACA	68
AAATT	143	ATTAA	68
ATTAT	138	TGTAT	67
TTATA	127	ATAAA	66
TTATT	124	TTGAA	66
ATAAT	123	GAAAA	66
TTTTT	122	AATGT	65
TTTAT	120	TTTAG	65
ATTTA	119	GTATA	64
TTAAT	118	AGTAT	64
TATAA	117	ATATG	63
TAAAT	115	GTATT	63
TAATT	114	TGAAA	63
TATAT	113	TATGT	61
TAATA	109	TTACT	60

Top 50 5-mers enriched for 3'UTRs of down-regulated genes (vs. no significant change)

B

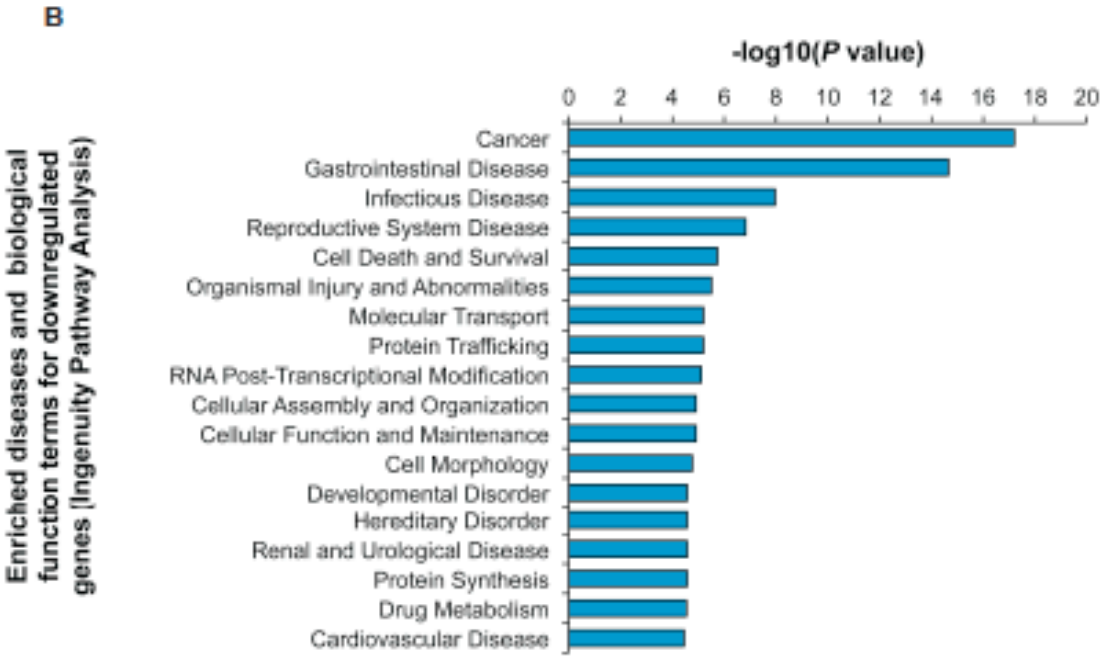


Figure S4

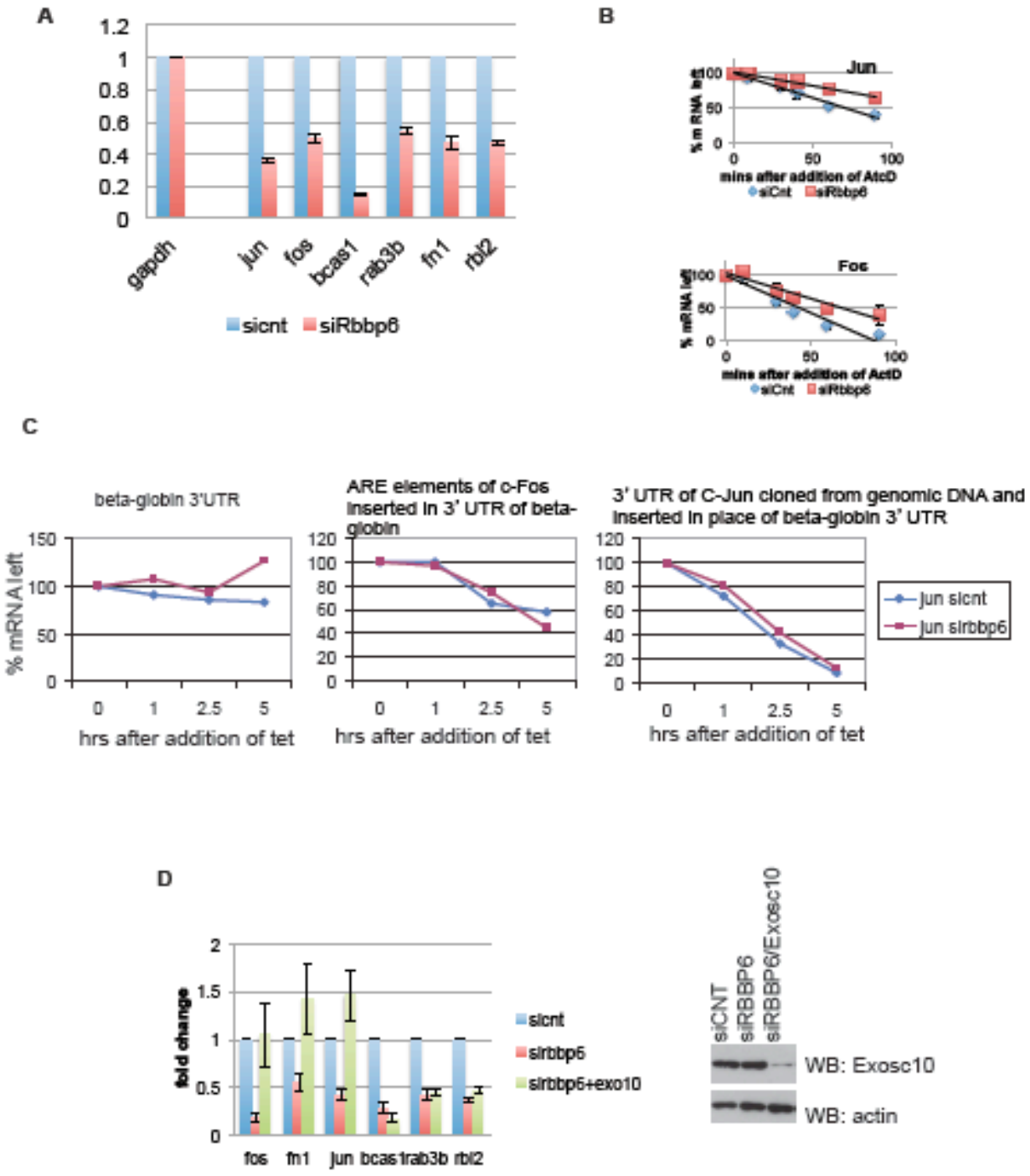

A
1.2
1
0.8
0.6
0.4
0.2
0
gapdh
jun
fos
bcas1
rab3b
fn1
rbl2
sicnt
siRbbp6
B
Jun
% m RNA left
mins after addition of AtcD
siCnt
siRbbp6
Fos
% mRNA left
mins after addition of ActD
siCnt
siRbbp6
C
beta-globin 3'UTR
% mRNA left
hrs after addition of tet
ARE elements of c-Fos inserted in 3' UTR of beta-globin
hrs after addition of tet
3' UTR of C-Jun cloned from genomic DNA and inserted in place of beta-globin 3' UTR
hrs after addition of tet
jun sicnt
jun sirbbp6
D
fold change
fos
fn1
jun
bcas1
rab3b
rbl2
sicnt
sirbbp6
sirbbp6+exo10
siCNT
siRBBP6
siRBBP6/Exosc10
WB: Exosc10
WB: actin

Figure S5

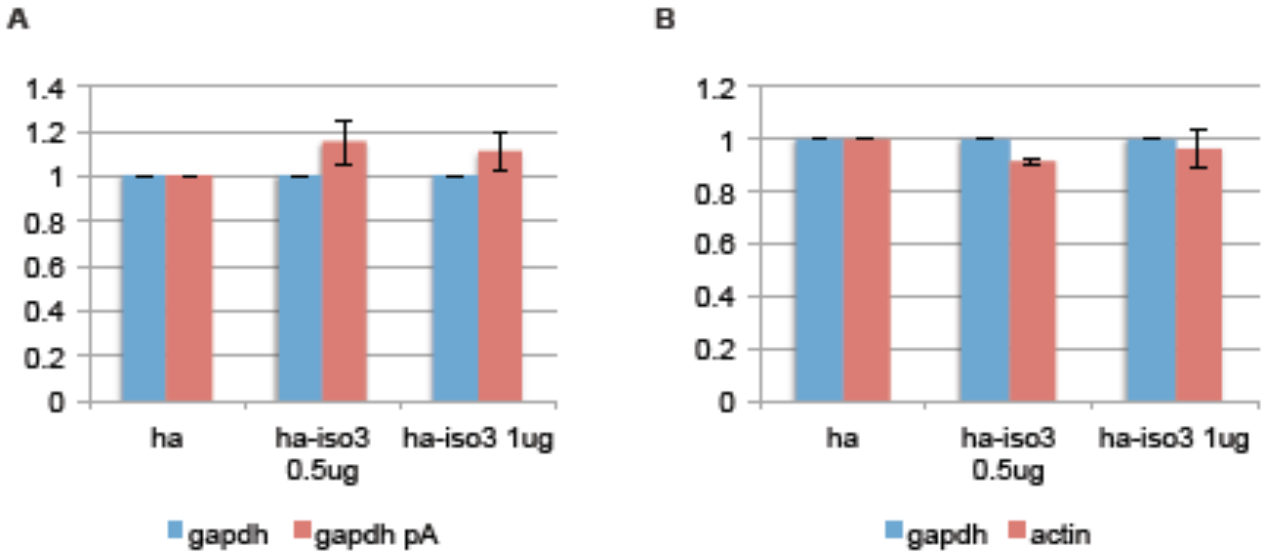
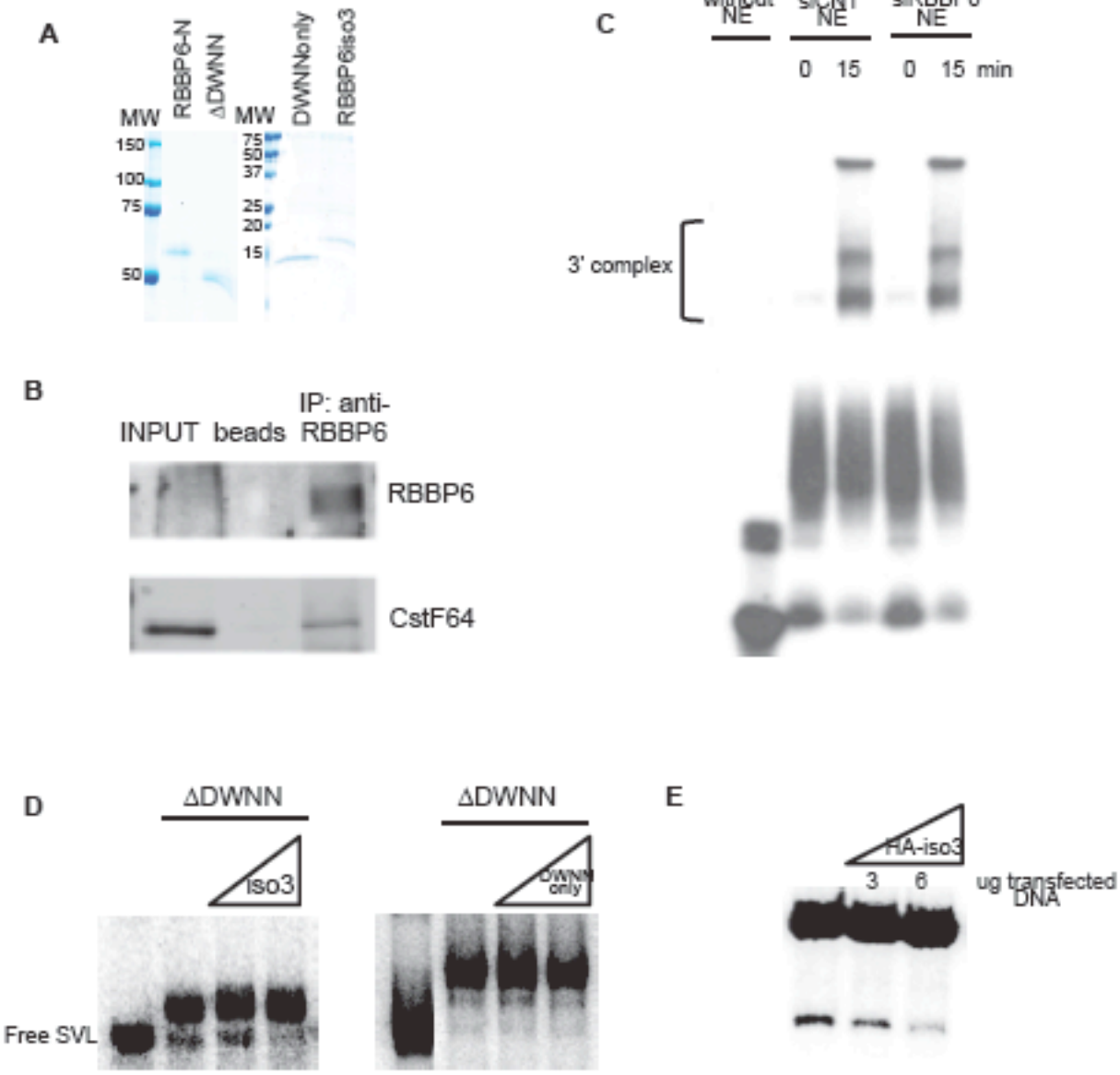

A
1.4
1.2
1
0.8
0.6
0.4
0.2
0
ha
ha-iso3 0.5ug
ha-iso3 1ug
gapdh
gapdh pA
B
1.2
1
0.8
0.6
0.4
0.2
0
ha
ha-iso3 0.5ug
ha-iso3 1ug
gapdh
actin

Figure S1


A
MW
RBBP6-N
ΔDWNN
MW
DWNNonly
RBBP6iso3
150
100
75
50
75
50
37
25
20
15
B
INPUT
beads
IP: anti-RBBP6
RBBP6
CstF64
C
without NE
siCNT NE
siRBBP6 NE
0
15
0
15
min
3' complex
D
ΔDWNN
iso3
ΔDWNN
DWNN only
Free SVL
E
HA-iso3
3
6
ug transfected DNA

APPENDIX

Molecular architecture of the human pre-mRNA 3' processing complex

The paper included in this appendix was published in 2009 in *Molecular Cell* (33:365-76) and represents the starting point for the works presented in chapters three and four of this dissertation.

Molecular Cell
Article

Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex

Yongsheng Shi,[1,5] Dafne Campigli Di Giammartino,[1] Derek Taylor,[2] Ali Sarkeshik,[3] William J. Rice,[4] John R. Yates III,[3] Joachim Frank,[2] and James L. Manley[1,*]
[1]Department of Biological Sciences, Columbia University, New York, NY 10027, USA
[2]Howard Hughes Medical Institute, Wadsworth Center, Albany, NY 12201-0509, USA
[3]The Scripps Research Institute, Department of Chemical Physiology, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
[4]New York Structural Biology Center, 89 Convent Avenue, New York, NY 10027-7556, USA
[5]Present address: Department of Microbiology and Molecular Genetics, School of Medicine, University of California, Irvine, Irvine, CA 92697, USA
*Correspondence: jlm2@columbia.edu
DOI 10.1016/j.molcel.2008.12.028

SUMMARY

Pre-mRNA 3′ end formation is an essential step in eukaryotic gene expression. Over half of human genes produce alternatively polyadenylated mRNAs, suggesting that regulated polyadenylation is an important mechanism for posttranscriptional gene control. Although a number of mammalian mRNA 3′ processing factors have been identified, the full protein composition of the 3′ processing machinery has not been determined, and its structure is unknown. Here we report the purification and subsequent proteomic and structural characterization of human mRNA 3′ processing complexes. Remarkably, the purified 3′ processing complex contains ~85 proteins, including known and new core 3′ processing factors and over 50 proteins that may mediate crosstalk with other processes. Electron microscopic analyses show that the core 3′ processing complex has a distinct "kidney" shape and is ~250 Å in length. Together, our data has revealed the complexity and molecular architecture of the pre-mRNA 3′ processing complex.

INTRODUCTION

Polyadenylation is a nearly universal step in eukaryotic gene expression. Poly(A) tails have profound influence on the stability, export, and translation efficiency of mRNAs (Colgan and Manley, 1997; Zhao et al., 1999). In mammals, biochemical studies have shown that pre-mRNA 3′ processing requires four multisubunit protein complexes—CPSF, CstF, CF I, and CF II—in addition to the single subunit poly(A) polymerase (PAP) (Takagaki et al., 1989; reviewed by Colgan and Manley, 1997; Mandel et al., 2008; Zhao et al., 1999). With the aid of RNA polymerase (RNAP) II (McCracken et al., 1997; Hirose and Manley, 1998), these and other factors assemble onto the nascent pre-mRNA to form a macromolecular complex in which the 3′ processing reactions take place. Despite significant divergence in the *cis*-elements required for 3′ processing between yeast and mammalian mRNAs, most 3′ processing factors are conserved. Interestingly, although 20 polyadenylation factors have been identified in yeast, only 15 have been found in mammals, suggesting either that the yeast machinery is more complex than its mammalian counterpart or that more mammalian 3′ processing factors remain to be discovered.

Most of our knowledge of 3′ processing has been based on biochemical studies of individual factors. In comparison, the molecular architecture and dynamics of the 3′ processing complex remain poorly understood. First, it is not clear what proteins, in addition to the known 3′ processing factors, constitute the functional 3′ processing complex, and in what stoichiometry. Second, indirect evidence suggests that the 3′ processing complex is dynamic and structural and/or compositional rearrangements may occur during the reactions. For example, although all 3′ processing factors are required for cleavage, CPSF and PAP are enough to reconstitute specific polyadenylation on precleaved RNAs (Colgan and Manley, 1997; Zhao et al., 1999). Based on these observations, it is possible that CstF, CF I, and CF II dissociate from the 3′ processing complex after cleavage. Finally, structural information is critical for understanding the detailed mechanisms of 3′ processing. Currently, crystal structures are available for several individual 3′ processing factors (Deo et al., 1999; Bard et al., 2000; Martin et al., 2000; Meinhart and Cramer, 2004; Mandel et al., 2006; Perez-Canadillas, 2006; Bai et al., 2007; Legrand et al., 2007; Noble et al., 2007; Qu et al., 2007; Coseno et al., 2008; Grant et al., 2008; Meinke et al., 2008), but the structure of the functional 3′ processing complex is unknown. Studies of 3′ processing complexes have been hampered by the lack of a method for purifying such complexes in their intact and functional form.

Accumulating evidence suggests that all steps of gene expression are highly coordinated (Hirose and Manley, 2000; Maniatis and Reed, 2002; Proudfoot et al., 2002; Bentley, 2005). For example, 3′ processing is necessary both for transcription termination and for the export of mRNAs, although the exact mechanisms are not fully understood. The coupling of different steps in gene expression is often mediated by physical interactions among factors involved in seemingly distinct processes. It has been shown that the 3′ processing factor

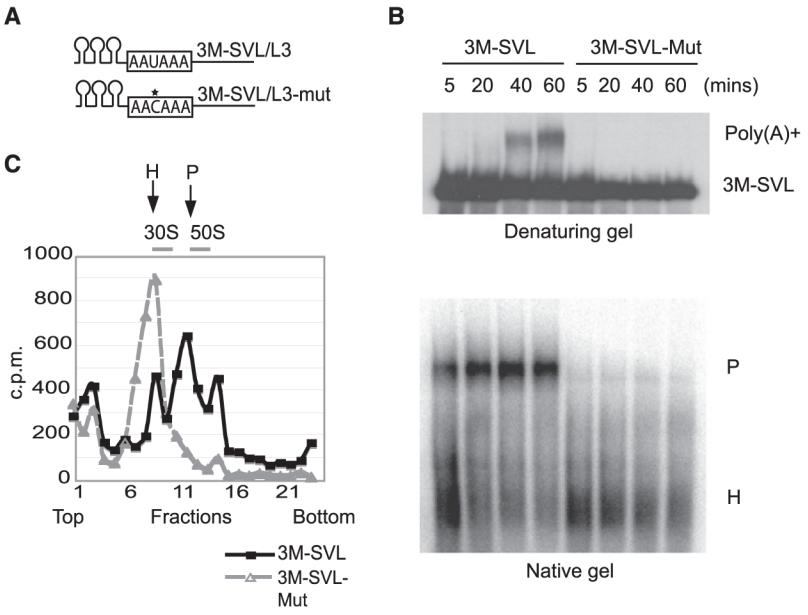


Figure 1. Characterization of RNA Substrates

(A) A schematic drawing of the RNA substrates. 3MS2 tags (hairpins) and the AAUAAA element in wild-type substrates and AACAAA in mutant substrates (boxes) are shown. The site of the single nucleotide change is marked with an asterisk.

(B) Comparison of the wild-type and mutant RNA substrates in polyadenylation and complex formation assays. A time course of polyadenylation reactions was analyzed on a denaturing 6% gel (top) or on a 1.5% native agarose gel (bottom) and visualized by phosphorimaging.

(C) RNA profiles (measured by radioactivity) on glycerol gradients. The peaks corresponding to the 30S/H and 50S/P complexes and the peak positions of *E. coli* 30S and 50S ribosomes from a parallel gradient are marked.

CPSF is associated with the transcription machinery as early as in the preinitiation complex through a direct interaction with TFIID (Dantonel et al., 1997), and both CPSF and CstF can remain associated with RNAP II throughout the coding region (Venkataraman et al., 2005; Glover-Cutter et al., 2008). In addition, interactions between CPSF and a component of the U2 snRNP, SF3b, are important for coupling between splicing and 3′ processing (Kyburz et al., 2006). A comprehensive characterization of the crosstalk between 3′ processing and other cellular processes will be important for better understanding gene regulation on a systems level.

In this study, we purified the human 3′ processing complex and determined its protein composition. Remarkably, the purified complex contains ~85 proteins. In addition to known polyadenylation factors, we identified new essential 3′ processing factors and over 50 proteins that may mediate coupling with other cellular processes. We also visualized the core 3′ processing complex using electron microscopy for the first time and describe its basic features.

RESULTS AND DISCUSSION

Purification of Functional Human Pre-mRNA 3′ Processing Complex

To purify the human pre-mRNA 3′ processing complex, we adopted an RNA-tagging strategy used previously to purify spliceosomal complexes (Jurica et al., 2002; Zhou et al., 2002; Deckert et al., 2006). Briefly, SV40 late (SVL) and adenovirus L3 pre-mRNAs, two commonly used substrates for in vitro 3′ processing analyses (Takagaki et al., 1988), were fused at their 5′ ends to 3 copies of the hairpin that specifically binds to the bacteriophage coat protein MS2 (3M-SVL and 3M-L3, Figure 1A). As controls, we used mutant RNA substrates with single point mutations (U to C) in the highly conserved AAUAAA sequence (3M-SVL-mut and 3M-L3-mut, Figure 1A). In vitro 3′ processing assays showed that, as expected, this single nucleotide substitution completely abolished cleavage (Figure S1 available online) and polyadenylation (Figure 1B, upper panel). Native gel analyses showed that 3′ processing complexes (P complexes) assembled efficiently on the wild-type RNAs, whereas the mutant RNAs were found almost exclusively in faster-migrating heterogeneous complexes (H complex, Figure 1B, lower panel). Complexes assembled on the wild-type and mutant substrates were further analyzed by glycerol gradient sedimentation (Figure 1C). The RNA distribution profile along the gradient showed that the mutant RNAs were concentrated in a peak of ~30S that corresponds to the H complex. Although a small portion of wild-type RNAs was also present in the same peak, the majority was found in an ~50S peak that corresponds to the 3′ processing complexes (see below).

To purify assembled 3′ processing complexes, RNA substrates were first bound to the adaptor protein MBP-MS2 (MBP is maltose-binding protein) and then incubated with HeLa nuclear extract (NE) under polyadenylation conditions (with ATP) to allow assembly of the complexes. Reaction mixtures were then fractionated by glycerol gradient sedimentation as described above. The 30S and 50S fractions were then used for affinity purification with amylose beads. Analysis of the RNAs in the purified complexes showed that the mutant RNA was exclusively found in the 30S/H complexes (Figure 2A, bottom panel). Although some wild-type RNA was also detected in this peak, the majority of it was found in the 50S/P complexes (Figure 2A, top panel), consistent with the glycerol gradient profile (Figure 1C). Silver staining of the eluted complexes revealed that a large number of proteins were specifically purified with wild-type substrate in the 50S/P complexes (Figure 2B). The protein profile of the purified 50S/P complex was distinct from that of the purified 30S/H complex (Figure 2B, compare the left and right panel), suggesting that P complexes were effectively separated from the H complexes. Western blotting showed that all 3′ processing factors tested, including CPSF73, CstF64, and symplekin, were specifically detected in the P complexes, but not in the H complexes or with mutant RNA substrates (Figure 2C). In contrast, hnRNP A1, a protein commonly found in H complexes, was highly enriched in the H complex assembled on the mutant substrate. We conclude

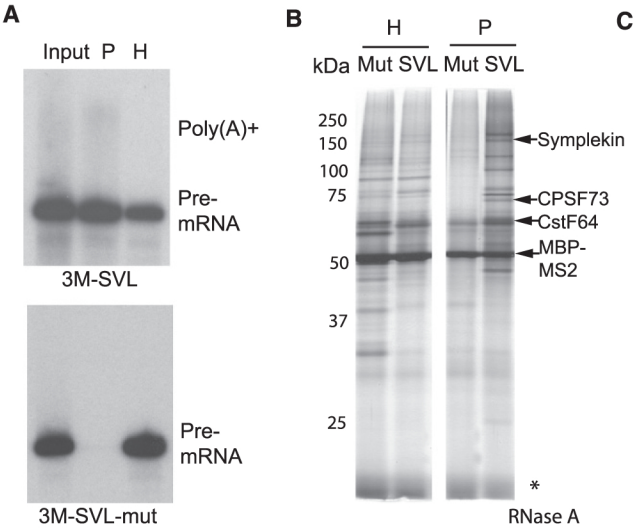


Figure 2. Purification of 3′ Processing Complexes

(A) RNAs from input and purified 30S/H and 50S/P complexes were purified, resolved, and visualized as described above.

(B and C) Proteins from the purified 30S/H and 50S/P complexes were resolved by SDS-PAGE and analyzed by silver staining (MBP-MS2 and several western-confirmed bands are labeled) (B) and western blotting (C).

that we have successfully purified 3′ processing complexes. It is important to point out that under the conditions used there was a significant time window (~20 min) during which the 3′ processing machinery was fully assembled (Figure 1B, lower panel, from 20 min time point to 40 min), but no significant 3′ processing had occurred (Figure 1B, top panel). Since we purified the 3′ processing complexes from this time window, the vast majority of the purified complexes were in the precleavage stage.

We next wished to determine whether the purified 3′ processing complexes were functional. To this end, we assembled 3′ processing complexes on the 3M-SVL RNA as above. Following purification, the complexes were tested for cleavage activity (in the presence of 3′ dATP, a polyadenylation inhibitor). No significant cleavage was observed with the purified complexes alone (Figure 3, lane 2), possibly indicating that one or more factors were missing or limiting. Indeed, proteomic analyses (see below) indicated that the CF II component Pcf11 was detected at substoichiometric levels and the other CF II subunit, Clp1, was completely missing. We therefore tested whether supplementing the purified complexes with CF components could allow cleavage to take place. To this end, we added partially purified cleavage factor complex (CF) that contains both CF I and II (Takagaki et al., 1989) to the purified 3′ processing complexes, and indeed, cleavage products were now detected (Figure 3, lanes 3–5). Therefore, we conclude that the purified 3′ processing complexes were functional in this complementation assay.

It is notable that other highly purified RNA processing complexes, such as the spliceosome, are also not active on their own but can be activated in complementation experiments analogous to the one described above (Jurica et al., 2002; Zhou et al., 2002; Deckert et al., 2006). Due to the dynamic nature of such complexes, certain factors may be missing at any given time point.

Proteomic Analyses of the Purified 3′ Processing Complexes

For proteomic analyses, we purified 3′ processing complexes assembled on the two aforementioned substrates, 3M-SVL and 3M-L3. For comparison, we also purified the H complexes assembled on 3M-SVL-mut RNAs. The protein composition of each complex was determined using the Multidimensional Protein Identification Technology (MudPIT; see Link et al., 1999 and the Experimental Procedures). Table 1 lists proteins found in either the SVL and L3 complexes, but only those found in both complexes were considered to be components of the 3′ processing complex.

The purified 3′ processing complexes contained ~85 proteins, including nearly all previously known 3′ processing factors. The only exception was Clp1, a component of the CF II complex (de Vries et al., 2000), which plays an unknown role in 3′ processing. Pcf11, another subunit of CF II, also seemed to be present at low levels, as only a small number of unique peptides from this protein were detected. These observations indicate that the association between CF II and the core 3′ processing complex may be weak and/or transient. Interestingly, instead of the canonical PAP (Lingner et al., 1991; Raabe et al., 1991), the

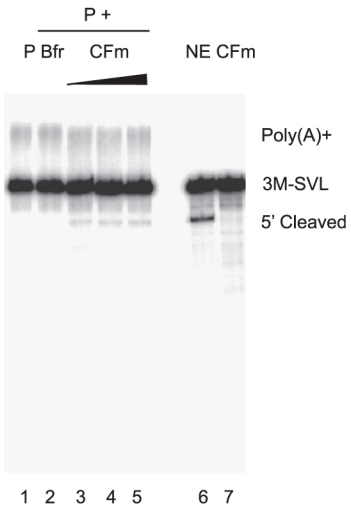


Figure 3. Purified 3′ Processing Complexes Are Functional in a Complementation Assay

Purified 3M-SVL P complexes alone (P, lane 1), supplemented with buffer D (Bfr, lane 2) or increasing amounts of CF fraction (CF, lanes 3–5), were incubated under cleavage conditions. Standard cleavage assays with NE and CF are shown as controls.

related neo-PAP/PAPOLG (Kyriakopoulou et al., 2001; Topalian et al., 2001) was the sole PAP detected. The reason for this is unclear but is consistent with many studies indicating that PAP is not tightly associated with other processing factors (e.g., Takagaki et al., 1988) and may reflect a dynamic association between PAP and the core 3′ processing complex and/or functional redundancy between the two poly(A) polymerases.

The complexes described above were assembled under polyadenylation conditions (with ATP). For comparison, we also purified 3′ processing complexes assembled under cleavage conditions (in the presence of the polyadenylation inhibitor 3′-dATP and without ATP) and analyzed them by mass spectrometry (Figure S2 and Table S1). Again, most RNAs within the purified complexes were unprocessed (Figure S2), indicating the majority of the purified complexes were in the precleavage stage. Comparison between Table 1 and Table S1 shows that the protein compositions of the 3′ processing complexes assembled under the two conditions were highly similar. For example, under both conditions, all subunits of CPSF, CstF, and CF I were identified, while CF II components were either detected only by a small number of unique peptides or entirely missing. In addition, many other factors, including WDR33, PP1, Rbbp6, and CstF64 tau (see below), were identified in both analyses. These observations suggest that the assembly conditions have very little effect on the general protein composition of the precleavage 3′ processing complexes.

Characterization of New 3′ Processing Factors

Three proteins identified in our study not previously implicated in mRNA 3′ processing in mammals (WDR33, Rbbp6, and PP1) are known or putative homologs of yeast 3′ processing factors. WDR33/WDC146 is a WD40 repeat-containing protein and is the putative mammalian homolog of the yeast 3′ processing factor Pfs2 (Ohnacker et al., 2000). To characterize its potential functions in 3′ processing, we first tested whether WDR33 interacts with the CPSF complex since Pfs2 binds strongly to Ysh1, the yeast homolog of CPSF73 (Ohnacker et al., 2000). To this end, we established a stable cell line expressing Flag-tagged CPSF73 and purified CPSF73 and associated proteins by immunoprecipitation (IP) (Figure 4A). We analyzed the CPSF73-containing complexes by mass spectrometry, and the identified proteins are listed in Table S2 next to those found in the 3′ processing complexes. All the known subunits of the CPSF complex (CPSF-160, -100, -73 and -30, and Fip1) were identified. Although not detected in CPSF previously purified through multiple chromatographic steps (Bienroth et al., 1991; Murthy and Manley, 1992), symplekin was detected at close to stochiometric levels, consistent with previous studies that it associates with CPSF as well as CstF (Takagaki and Manley, 2000). Strikingly, WDR33 was also present in the CPSF complex, as indicated by the large number of unique peptides detected (Figure S3 and Table S2) and confirmed by western blotting (Figure 4B). The WDR33 band partially overlapped the CPSF160 band on SDS-PAGE (Figure 4A), perhaps explaining in part why WDR33 previously escaped detection. Gel filtration analysis of the purified CPSF complex showed that WDR33 coeluted with CPSF (Figure 4C).

We next immunodepleted WDR33 from NE to determine whether WDR33 is required for 3′ processing. Quantitative western analyses showed that ~95% of WDR33 was removed (Figure 4D). About two-thirds of CPSF73 was also codepleted, indicating that the majority of CPSF73 is associated with WDR33. CstF64 levels were reduced by about a third while the level of the phosphatase PP1 was not significantly affected. When WDR33-depleted NE was used in 3′ processing assays, both cleavage (Figure 4E) and polyadenylation (Figure S4) were essentially abolished. Add-back of the immunopurified CPSF complex restored cleavage (Figure 4E). Together, these data indicate that WDR33, despite the fact that it had not been previously identified, is a bona fide component of the CPSF complex and suggest that it plays an essential role in mammalian 3′ processing.

Pfs2 was initially suggested to be the yeast equivalent of mammalian CstF50, another WD40 repeat protein with which it shares limited similarity (Ohnacker et al., 2000). This was consistent with the fact that the yeast CstF equivalent, CF IA, lacks a CstF50 homolog (Zhao et al., 1999). However, our data now indicate that the human 3′ processing machinery contains two WD40 proteins, one in CPSF and another in CstF. It is noteworthy that the plant PFS2/WDR33 homolog, FY, has been implicated as playing an important role during floral transition (Simpson et al., 2003).

Rbbp6/PACT is the putative homolog of the yeast 3′ processing factor Mpe1 (Vo et al., 2001). Interestingly, Rbbp6 was originally identified as a p53- and Rb-binding protein, playing important roles in apoptosis, cell-cycle, and p53 regulation (Sakai et al., 1995; Simons et al., 1997). Rbbp6 is significantly larger than Mpe1 and contains additional domains, including an arginine/serine-rich (RS) domain that is found in many splicing factors and a RING-finger-related domain (Pugh et al., 2006). Although we have not directly tested its role in processing, it is likely that Rbbp6, like its yeast counterpart Mpe1, functions in 3′ processing. It is possible that Rbbp6 may link mRNA 3′ end formation to the Rb/p53 pathways and tumorigenesis.

Our results show that the serine/threonine phosphatase PP1 and its regulator PNUTS are components of the human 3′ processing complex (Table 1). The PP1 homolog in yeast, Glc7, is a known 3′ processing factor, and its phosphatase activity is specifically required for polyadenylation, but not for cleavage (He and Moore, 2005). To test if PP1 is involved in mammalian 3′ processing, we depleted PP1 and the related PP2A family phosphatases using microcystin (MC)-conjugated beads (Figure 5A). MC is a specific small-molecule inhibitor of the PP1/2A phosphatases, and we have shown previously that MC-conjugated beads can be used to efficiently deplete these phosphatases from NE (Shi et al., 2006). When mock- or MC-treated NE were used in standard in vitro 3′ processing assays, similar levels of cleaved products were observed (Figure 5B). Polyadenylation, however, was significantly reduced in MC-treated NE, and add-back of recombinant PP1 restored polyadenylation (Figure 5C), suggesting that PP1 is specifically required for polyadenylation. Therefore, dephosphorylation by PP1 is an evolutionarily conserved step in 3′ processing.

We also found CstF64 tau in our purified 3′ processing complex. CstF64 tau is highly homologous to CstF64 and

Table 1. Protein Composition of the Human Pre-mRNA 3′ Processing Complex

Protein Name	Accession No.	Yeast homolog	Motifs/Notes on Functions	Cal. Mass	No. of peptides	
					L3	SVL
Known Polyadenylation Factors						
CPSF complex						
CPSF160 (CPSF1)	NP_037423	CFT1	SFT1	160822	80	72
CPSF100 (CPSF2)	NP_059133.1	CFT2	β-CASP	88487	59	52
CPSF73 (CPSF3)	NP_057291	YSH1	β-CASP	77486	42	22
CPSF30 (CPSF4)	NP_006684	YTH1	Zinc finger	30124	18	13
hFip1	NP_112179	FIP1		66526	23	19
CstF complex						
CstF77 (CSTF3)	NP_001317	RNA14	HAT	82922	60	54
CstF64 (CSTF2)	NP_001316	RNA15	RRM	60959	31	22
CstF50 (CSTF1)	NP_001315		WD repeats	48358	21	18
CF Im complex						
CF Im 68 (CPSF6)	NP_008938		RRM	59209	15	10
CF Im 59	NP_079087		RRM	52050	15	12
Cf Im 25 (CPSF5)	NP_008937			26227	26	21
CF IIm complex						
Pcf11	NP_056969.2	PCF11	CID	173050	5	3
Other Known Polyadenylation Factors						
Symplekin	NP_004810	PTA1		126500	51	39
Symplekin variant	BAE06092.1			118793	48	38
PAPOLG	NP_075045	PAP	RRM	82803	2	3
PABPC1	NP_002559.2	PAB1	RRM	70671	27	33
PABPC4	NP_003810.1		RRM	70783	20	23
PABPN1	NP_004634		RRM	31749	14	12
(Putative) Homologs of Yeast Polyadenylation Factors						
CstF64 tau (CSTF2T)	NP_056050.1	SPAC644.16 (*S. pombe*)	RRM	64436	21	16
WDR33	NP_060853	PFS2	WD repeats	145921	49	45
RBBP6	NP_008841	MPE1	RS, DWNN	201563	4	3
PP1 alpha	NP_002699.1	GLC7	Phosphatase	37512	2	7
PP1 beta	NP_002700	GLC7	Phosphatase	37187	2	5
DNA-Damage-Response Factors						
DNA-PK	NP_008835		Kinase	469093	21	15
Ku 70	NP_001460	YKU70	DNA helicase	69843	4	7
Ku 86/XRCC5	NP_066964	YKU86	DNA helicase	82705	6	13
PARP-1*	NP_001609			113135	5	
RNAP II and Associated Factors						
Pol II large subunit Rpb1	NP_000928.1	RPB1	Polymerase	217204	2	2
Pol II B*	NP_000929.1	RPB2		133896		2
Rpb5 (Pol II E)*	NP_002686.2	RPB5		24611	2	
Rpb11*	AF468111_1	RPB11		17208		2
Transcription factors						
TF II, I	NP_127492.1			112416	8	2
TAF15	AAH46099.1		RRM, zf-RanBP	52061	6	5
Integrator complex						
INTS2	NP_065799.1			134346	2	2
INTS3	NP_075391.3			118013	4	9
INTS4	NP_291025.3			108171	2	3

(*Continued on next page*)

Table 1. *Continued*

Protein Name	Accession No.	Yeast homolog	Motifs/Notes on Functions	Cal. Mass	No. of peptides L3	No. of peptides SVL
Known Polyadenylation Factors						
INTS6	NP_036273.1			100390	2	3
INTS7	NP_056249.1			106834	2	2
INTS8*	NP_060334.2			113088		3
INTS9*	NP_060720.1		YSH1 homology	73815		4
INTS10*	NP_060612.2			82236		2
PAF complex						
Cdc73	AAH14351.2	CDC73		46675	2	2
FACT complex						
FACT complex large subunit	NP_009123	SPT16		119914	4	4
Splicing Factors						
p54nrb	NP_031389.3			54232	11	3
PSF	NP_005057.1			76150	10	4
PUF60	NP_510965.1		RRMs	59876	3	9
UAP56	NP_004631.1	SUB2	RNA Helicase	48991	12	9
SF1	NP_004621.2	BBP		68330	5	5
SF3A, subunit 1	NP_005868.1	PRP21	Surp	88886	9	15
SF3A, subunit 2	NP_009096.2	PRP11		49256	2	4
SF3A60	CAA57388	PRP9		58777	3	18
SF3B, subunit 1	NP_036565.2	HSH155		145830	16	34
SF3B, subunit 2	NP_006833.2	CUS1		100228	2	21
SF3B, subunit 4	NP_005841.1	HSH49		44386	3	5
SF3B, 14 kDa*	NP_057131.1	SNU17		14585		5
Prpf38b	NP_060531.1			64441	2	8
SRm300	NP_057417.2		RS	299676	13	15
U1 70K	NP_003080.2	SNP1	RRM	51557	7	7
U2AF65	NP_009210.1	MUD2	RRM	53501	6	11
U2AF35	NP_006749.1		RRM	27872	2	3
PRP19	NP_055317.1	PRP19	RING, WD	55181	5	3
Exosome						
SKIV2L2 (hMTR4)	NP_056175.2	MTR4	DEXD	117933	29	28
Translation Factors						
eEF1-alpha	NP_001393.1	TEF1/2	GTPase	50141	14	12
eEF1-gamma	NP_001395.1	TEF4	GST	50119	5	4
eIF2A	NP_114414.2	YGR054W		64990	4	4
eIF 4A	NP_001407.1	TIF1	Helicase	46154	12	6
eIF3 gamma	NP_003747.1		JAB/MPN	39930	5	6
eIF3S2	NP_003748.1	TIF34	WD	36502	2	2
eIF3S5	NP_003745.1		PCI	37564	3	2
eIF3S6	NP_001559.1			52221	2	4
eIF3S9 eta	NP_003742.2		RRM	92482	7	7
eIF3A	NP_003741.1	RPG1		166569	16	9
eIF4G1	NP_886553.2	TIF4631		175460	2	2
RACK1/GNB2L1/lung cancer oncogene 7	NP_006089.1	ASC1	WD	35077	5	5
Factors with Known Motifs						
RNA binding motif protein 7	NP_057174.1		RRM	30503	4	4

Table 1. *Continued*

Protein Name	Accession No.	Yeast homolog	Motifs/Notes on Functions	Cal. Mass	No. of peptides L3	No. of peptides SVL
Known Polyadenylation Factors						
RNA binding motif protein 9/ Ataxin-binding protein/Fox-1	NP_055124		RRM	39515	2	7
RNA binding motif protein 25	NP_067062.1		RRM	100186	4	16
RNA binding motif protein 39, isoform a	NP_909122.1		RRM	59380	2	4
RNA binding motif protein 39, isoform b	NP_004893.1		RRM	58657	2	4
DEAH box polypeptide 3	NP_001347.3	DBP1	RNA helicase	7244	13	9
DEAH box polypeptide 5	NP_004387.1	DBP2	RNA helicase	69148	4	5
DEAH box polypeptide 9	NP_001348.2		RNA helicase	140958	32	3
DEAH box polypeptide 15	NP_001349.2	PRP43	RNA helicase	90933	3	17
DEAH box polypeptide 36	NP_065916.1		RNA helicase	114776	12	27
DEAH box polypeptide 39	NP_005795.2		RNA helicase	49130	8	5
DEAD box polypeptide 6	NP_004388.1	DHH1	RNA helicase	82432	3	2
DEAD box polypeptide 17 (p82)	NP_006377.2		RNA helicase	80273	6	4
Gemin3 (DEAD box polypeptide 20)	NP_009135.3		RNA helicase	92213	4	2
DEAD box polypeptide 23	NP_004809.2	PRP28 (*S. Pombe*)	RNA helicase	95583	4	5
ZCCHC8	NP_060082		Zinc finger	79375	15	16
EBNA2 coactivator, p100/SND1	NP_055205.2		Staphylococcal nuclease domain, Tudor	101997	12	2
Interferon-induced protein with tetratricopeptide repeats 1 isform 2	NP_001539.3		TPR	55360	9	7
Interferon-induced protein with tetratricopeptide repeats 3	NP_001540		TPR	55985	8	5
Other Factors						
G3BP/Ras-GTPase-activating protein SH3-domain-binding protein	NP_005745.1		Endoribonuclease	52164	5	2
JUP protein/ gamma-catenin	NP_002221.1			81745	5	3
PNUTS (PPP1R10)	NP_002705.2		PP1 regulator	99058	6	9

Components of multi-subunit complexes that are present in only one purified complex are listed and are indicated with asterisks (*). Proteins that are abundant in H complexes were designated H complex components and not listed here.

reportedly found only in testis and not expressed in HeLa cells (Wallace et al., 1999). To determine whether CstF64 tau is a component of the CstF complex, we established a HEK293 cell line stably expressing Flag-tagged CstF77 and purified the CstF complex by IP (Figure S5A). Results of mass spectrometry analyses of purified CstF are listed in Table S2, and CstF64 tau was indeed detected (Figure S5B and Table S2). We suspect that CstF64 tau may be a general component of the CstF complex. An intriguing possibility is that CstF complexes may contain either CstF64 or CstF64 tau, and their functions may be partially redundant. This is consistent with earlier observations that an ~90% reduction in CstF64 levels had no significant effect on cell growth (Takagaki and Manley, 1998) and that CstF may function as a dimer in 3′ processing (Bai et al., 2007).

Accumulating evidence suggests that all steps of gene expression are highly coordinated, and coupling of different steps is often mediated by physical interactions among factors involved in seemingly distinct processes (Hirose and Manley, 2000; Maniatis and Reed, 2002; Proudfoot et al., 2002; Bentley, 2005). Consistent with this theme, we identified in our purified 3′ processing complex a large number of proteins that have known or putative functions in transcription and splicing, both of which are known to be connected to 3′ processing. In fact, splicing factors detected in our complexes, such as SF3b, U2AF, and U1-70K, have been shown to associate with specific 3′ processing factors and mediate crosstalk between splicing and polyadenylation (Gunderson et al., 1998; Vagner et al., 2000; Kyburz et al., 2006). We detected RNAP II in the purified 3′ processing complex, consistent with earlier findings that RNAP II is necessary for efficient 3′ cleavage in vitro (Hirose and Manley, 1998). In addition, we identified the PAF complex, a RNAP II-associated transcription elongation factor that was recently shown to function in 3′ end formation of polyadenylated mRNAs in yeast (Penheiter et al., 2005). Most subunits of another RNAP II-associated complex, the Integrator, were also found in the 3′ processing complexes. The Integrator complex was recently shown to

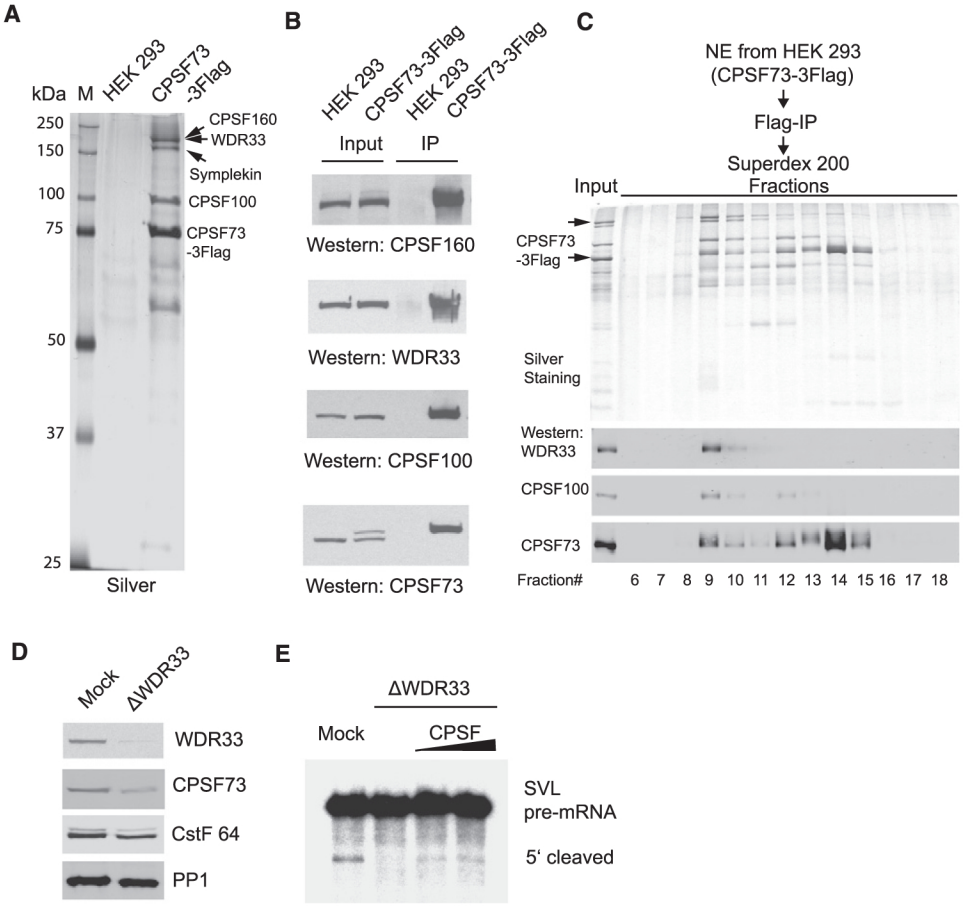


Figure 4. WDR33 Is a Bona Fide Component of the CPSF Complex

(A and B) Characterization of the immunopurified CPSF73 complex. Proteins from Flag-IP of NE from HEK293 or the CPSF73-3Flag cell line were resolved by SDS-PAGE and stained with silver (A) or analyzed by western blotting using specific antibodies (B). Protein bands in (A) whose identities were confirmed by western are labeled.

(C) Immunopurified CPSF complexes were analyzed by gel filtration using a Superdex-200 column. Proteins from input and each fraction were analyzed by SDS-PAGE and silver staining (top panel) and western (bottom panel). WDR33 is marked by an arrow.

(D) Western analyses of the mock-treated (mock) or WDR33-depleted (ΔWDR33) NE. Intensities of individual bands were quantified using the Odyssey Infrared scanner (Li-Cor).

(E) WDR33 is essential for 3′ processing. Mock, ΔWDR33 NE alone, or ΔWDR33 supplemented with immunopurified CPSF were used in standard cleavage assays. RNAs were resolved and visualized as described above.

function in the 3′ processing of snRNAs, and two of its subunits, Ints9 and Ints11, display sequence homology with CPSF 100 and 73, respectively (Baillat et al., 2005). The integrator subunits INTS8, -9, and -10 were missing in the L3 complex. The reason for their absence is unclear but may be due to slightly lower levels of the Integrator complex in the L3 complexes or the association between these subunits, and the rest of the complex might be weak and/or transient. It is currently unclear what, if any, role the Integrator might play in the 3′ processing of pre-mRNAs.

Our study also identified a number of factors that may mediate unexpected connections between 3′ processing and other cellular processes. For example, we found that the DNA-activated protein kinase complex (DNA-PKcs/Ku70/Ku86), well-studied for its functions in DNA-damage repair, is associated with the 3′ processing complex (Table 1). This is potentially similar to transcription where several DNA repair factors, such as XPB and XPD, are also essential transcription factors as components of the TFII H (Drapkin et al., 1994). These results are also consistent with previous studies showing that 3′ processing is connected to DNA-damage response (Kleiman and Manley, 2001; Mirkin et al., 2008). Another intriguing factor associated with the 3′ processing complex was the translation elongation factor and GTPase eEF1 alpha. Interestingly, Tef1, the yeast homolog of eEF1 alpha, copurifies with the yeast 3′ processing factor CF I (Gross and Moore, 2001). It will be of interest to examine what, if any, roles these and other factors identified in our proteomic analyses play in 3′ processing.

A comparison between our proteomic analyses of the 3′ processing complexes and previously studies of the spliceosome (Jurica et al., 2002; Zhou et al., 2002; Deckert et al., 2006) revealed both similarities and differences. One common

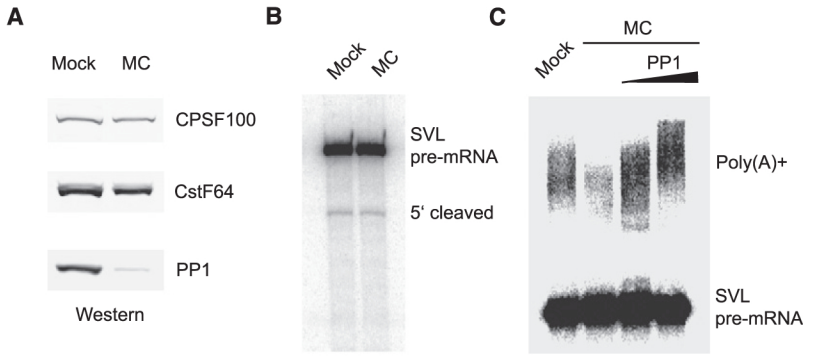


Figure 5. PP1 Is Required for Polyadenylation, but Not for Cleavage

(A) Western analyses of mock-treated (Mock) or microcystin-treated (MC) NE.

(B) Depletion of PP1/2A phosphatases has no effect on cleavage. Mock and MC NE were used in standard in vitro cleavage assays with SVL as the substrate. Pre-mRNA and the 5′ cleaved products are marked.

(C) PP1 is required for polyadenylation. Mock, MC NE alone, or MC NE supplemented with recombinant PP1 were used in standard polyadenylation assays with SVL as the substrate. Pre-mRNA and poly(A)$^+$ RNAs are marked.

theme is that all of these studies strongly support a link between splicing and 3′ processing. An interesting difference is that although the spliceosome includes a number of factors that have been implicated in transcription (e.g., TREX and TAT-SF1) (Zhou et al., 2002), the 3′ processing complex contains RNAP II and RNAP II-associated factors, such as the Integrator and PAF (Table 1). Therefore, splicing and 3′ processing are both connected to transcription but probably by different factors.

Structural Analyses of 3′ Processing Complexes by EM

We next wished to begin to analyze the structure of the 3′ processing complex. To this end, the isolated complexes were processed following the GraFix method (Kastner et al., 2008). Briefly, the 3′ processing complexes purifed as described above were subject to a second glycerol gradient sedimentation during which the complexes are centrifuged into increasing concentration of the fixation reagent glutaraldehyde. This step serves to further purify and gently fix the complexes to preserve their integrity and has been successfully used, for example, to process spliceosomes for electron microscopic (EM) analyses (Deckert et al., 2006; Behzadnia et al., 2007). During this second glycerol gradient sedimentation, the 3′ processing complexes were found again in an ~50S peak (data not shown), indicating that structural integrity of the complex was preserved throughout the purification. The peak fraction was negatively stained with uranyl formate using the carbon-sandwich method (Radermacher et al., 1987) and analyzed by EM.

A typical raw image of the 3′ processing complex shows monodisperse particles of similar sizes (Figure 6A). We also obtained tilted images of the particles to confirm that they were fully sandwiched between carbon membranes and that the staining was homogeneous (Figure S6). Images of representative particles show a distinct "kidney" shape, slightly elongated and bent (Figures 6B). There appears to be a central cavity surrounded by two or more peripheral densities. The maximum dimension of the complex is ~250 Å, which is consistent with its ~50S sedimentation coefficient.

3,671 molecular images were collected for image processing using both SPIDER (Frank, 1996) and EMAN (Ludtke et al., 1999). After reference-free alignment and classification, 50 (using SPIDER, shown in Figure S7) and 47 (using EMAN, shown in Figure S8) two-dimensional class averages were obtained. Although most of the class averages have defined edges and consistent sizes, they seem to lack strong internal features. This could potentially be due to heterogeneity among particles caused by the dynamic nature of this complex and/or the presence of substoichiometric factors. Another possible explanation is that classification of negative stain images focuses on the shape of the boundary so that small changes in orientation will not affect the class designation yet can result in changes of internal features seen in projection. As a result, internal features of class averages can be blurred. Nonetheless, our results have provided a first view of the polyadenylation complex and revealed its general structural features.

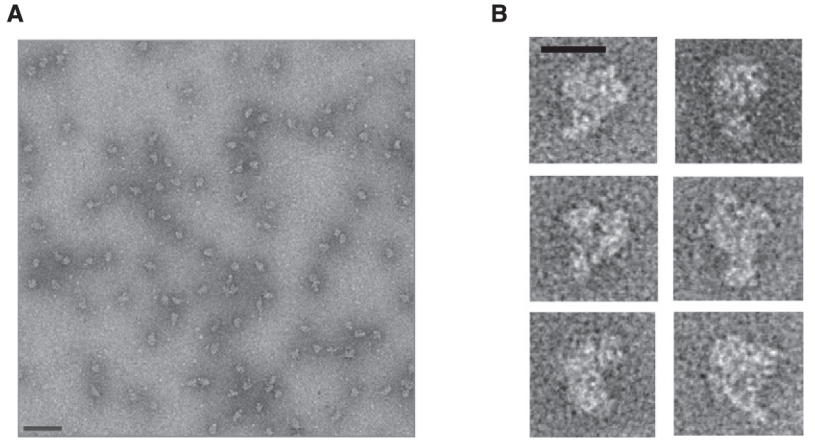


Figure 6. EM Analyses of the Purified 3′ Processing Complex

(A) A typical CCD micrograph of the negatively stained purified 3′ processing complexes. Bar, 100 nm.

(B) A gallery of representative particles. Bar, 20 nm.

Given the seemingly simple nature of the polyadenylation reaction, it is remarkable that it involves such a large complex. The size of the 3′ processing complex is close to that of the bacterial ribosome large subunit (Radermacher et al., 1987) and the spliceosomal A complex (Behzadnia et al., 2007). The major components of our purified complexes, such as CPSF, symplekin, CstF, and CF I, likely constitute the "core" of the 3′ processing complex seen in the class averages, as their collective molecular weight (~1.1 MDa) is already close to that of the bacterial ribosome large subunit (1.5 MDa) (Radermacher et al., 1987). These core factors, at least one of which, CstF, may be present as a dimer (Bai et al., 2007), likely correspond to some of the observed major densities. For the rest of the factors identified in the proteomics (total molecular weight ~7 MDa), the majority are present at substoichiometric levels and likely contribute to the heterogeneity observed among particles. The structure described here, we believe, corresponds to the core polyadenylation machinery.

In this study, we purified functional human pre-mRNA 3′ processing complexes and determined their protein composition. We detected all but one known 3′ processing factor and identified several new and potentially essential ones. We identified a number of proteins involved in other cellular processes, expanding the view that 3′ processing is integrated with other cellular events. We also visualized the 3′ processing complex for the first time and characterized its basic structural features. Together, our study has provided critical insights into the molecular composition and the structure of the 3′ processing complex, revealing a molecular architecture that is much more complex than previously expected.

EXPERIMENTAL PROCEDURES

Antibodies

Anti-CPSF160, -100, -73, and WDR33 were kindly provided by Orit Rosenblatt and Bethyl Laboratories; anti-CstF64 6A9 was described previously (Takagaki et al., 1990); anti-symplekin was from BD Biosciences; and anti-hnRNP A1 was from ImmuQuest.

In Vitro 3′ Processing Assays

Constructs used in this study were derived from pG3SVL-A and pG3L3-A, which contain the SV40 late site and adenovirus 2 L3 poly(A) site, respectively (Takagaki et al., 1988). 3 MS2-binding sequences were as described previously (Zhou et al., 2002) and were inserted between Acc I and Xba I sites before the SVL and L3 sequences. ^{32}P-labeled pre-mRNAs were prepared with SP6 RNA polymerase (Promega) from linearized plasmids. Polyadenylation reactions typically contain 8 pmol radiolabled RNA/ml reaction, 40% NE, 8.8 mM HEPES (pH 7.9), 44 mM KCl, 0.4 mM DTT, 0.7 mM $MgCl_2$, 1 mM ATP, and 20 mM creatine phosphate. In cleavage reactions, ATP was omitted and 0.2 mM 3′ dATP (Sigma), 2.5% PVA, and 40 mM creatine phosphate were added.

Purification of 3′ Processing Complexes

Radiolabeled RNA substrates were incubated with 50 molar excess of MBP-MS2 adaptor protein for 30 min on ice. Then, the other ingredients of the polyadenylation reaction were added, and the reactions were incubated in 200 μl aliquots at 30°C for 40 min or otherwise specified time. The reactions were chilled and loaded onto 11 ml 10%–30% glycerol gradients (20 mM HEPES [pH 7.9], 100 mM KCl, 0.1 mM EDTA, and 1 mM DTT). The gradients were centrifuged at 22,000 rpm for 16 hr in a SW41 rotor, and then 500 μl fractions were manually collected from the top to the bottom. Radioactivity of each fraction was measured using a liquid scintillation counter. Peak fractions were pooled and mixed with amylose beads for 1 hr at 4°C. After washing with wash buffer (20 mM HEPES [pH 7.9], 100 mM KCl, 0.1 mM EDTA, 1 mM DTT), the complexes were eluted in wash buffer plus 12 mM maltose. For mass spectrometry analyses, the eluted complexes were treated with RNase A and precipitated before analyses. For purification of complexes assembled on mutant substrates and cleavage complexes, reaction mixtures were loaded onto a Sephacryl S-400 size-exclusion column, and RNP-containing fractions were pooled and used for affinity purification with amylose beads (Jurica et al., 2002).

Proteomic Analyses of Purified 3′ Processing Complex Using Multidimensional Protein Identification Technology

Precipitated protein preparations were dissolved in digestion buffer, digested by trypsin, and analyzed by LC/LC/MS/MS according to published protocols (Link et al., 1999). MS/MS spectra obtained were analyzed by SEQUEST using a nonredundant NCBI protein database. The SEQUEST outputs were then analyzed by DTASelect (version 2.0) program. The type of digestion method used was specified (-trypstat for tryptic digests) so as to specifically filter for peptides with trypsin specificity. A user-specified false positive rate was used to dynamically set XCorr and DeltaCN thresholds through quadratic discriminant analysis. This data set was then further filtered to remove contaminants (i.e., keratin) through the use of Contrast (version 2.0). A minimum of two peptides and half tryptic status (−p 2–y 1) were set in the Contrast.params file. For analyses of the cleavage complexes, mininum peptide number was set to 1.

Immunopurification of CPSF73- and CstF77-Associated Proteins

The plasmids CPSF73-3Flag-pCMV14 and Flag-CstF77-pCDNA3.1 were transfected into HEK293 cells, and stable transfectants were selected using G418 (Invitrogen). NE was made from these stable cell lines using standard protocol, and IP was performed using M2 beads (Sigma). For functional analyses, eluted proteins were concentrated using Centricon Y-30 (Millipore) and directly used in in vitro 3′ processing assays.

Immunodepletion of WDR33

100 μl NE was diluted with equal volume of Buffer D, and NP-40 was added to final concentration of 0.1%. The diluted NE was then mixed with either protein G-agarose (mock) or with anti-WDR33-conjugated protein G-agarose (ΔWDR33) for 2 hr at 4°C. The depletion efficiency was measured by quantitative western blotting using the Odyssey infrared scanner (Li-Cor).

Electron Microscopy

A 9 ml polyadenylation reaction was used for purification as described above. Following affinity purification, the complexes were eluted in 300 μl and further treated using the GraFix method (Kastner et al., 2008). Briefly, the eluted complexes were loaded on a 4 ml 10%–30% glycerol and 0%–0.1% glutaraldehyde gradient and centrifuged at 51,000 rpm for 2.5 hr in an SW55Ti rotor. Afterwards, 180 μl fractions were taken manually from the top. Negative staining was performed using the carbon-sandwich method (Radermacher et al., 1987). Images were acquired on a JEOL JEM2100F electron microscope operating at 200 kV. Imaging was performed at a set magnification of 30,000× under low-dose conditions at an underfocus of 2.3 microns, and images were collected on a Tietz 224HD 2Kx2K CCD camera with 24 micron pixel size. The calibrated pixel size was 5.11 A/pixel. Images were processed using the standard SPIDER protocol (Frank, 1996) or the refine2d.py script from EMAN (Ludtke et al., 1999).

SUPPLEMENTAL DATA

The Supplemental Data include eight figures and two tables and can be found with this article online at http://www.cell.com/molecular-cell/supplemental/S1097-2765(09)00025-2.

ACKNOWLEDGMENTS

We thank Drs. M. Jurica, J. Vilardell, and O. Rosenblatt for providing reagents; K.D. Derr and Dr. R. Diaz at the New York Structural Biology Center for technical assistance; Drs. A. Tzagoloff and R. Gonzalez for sharing equipment;

B. Reddy for help in the early stage of this study; Dr. V. Vathantham for providing DNA constructs; and other members of the Manley lab for helpful discussions. This work was supported by NIH grants GM028983 to J.L.M, P41 RR011823 YRC grant to J.Y, and R37 GM29169 and GM55440 to J.F. J.F. is a Howard Hughes Medical Institute investigator.

Received: August 22, 2008
Revised: October 24, 2008
Accepted: December 12, 2008
Published: February 12, 2009

REFERENCES

Bai, Y., Auperin, T.C., Chou, C.Y., Chang, G.G., Manley, J.L., and Tong, L. (2007). Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. Mol. Cell *25*, 863–875.

Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell *123*, 265–276.

Bard, J., Zhelkovsky, A.M., Helmling, S., Earnest, T.N., Moore, C.L., and Bohm, A. (2000). Structure of yeast poly(A) polymerase alone and in complex with 3′-dATP. Science *289*, 1346–1349.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J. *26*, 1737–1748.

Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr. Opin. Cell Biol. *17*, 251–256.

Bienroth, S., Wahle, E., Suter-Crazzolara, C., and Keller, W. (1991). Purification of the cleavage and polyadenylation factor involved in the 3′-processing of messenger RNA precursors. J. Biol. Chem. *266*, 19768–19776.

Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev. *11*, 2755–2766.

Coseno, M., Martin, G., Berger, C., Gilmartin, G., Keller, W., and Doublie, S. (2008). Crystal structure of the 25 kDa subunit of human cleavage factor Im. Nucleic Acids Res. *36*, 3474–3483.

Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. Nature *389*, 399–402.

de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H., and Keller, W. (2000). Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J. *19*, 5895–5904.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Mol. Cell. Biol. *26*, 5528–5543.

Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell *98*, 835–845.

Drapkin, R., Sancar, A., and Reinberg, D. (1994). Where transcription meets repair. Cell *77*, 9–12.

Frank, J. (1996). Three-dimensional electron microscopy of macromolecular assemblies (San Diego, CA: Academic Press).

Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat. Struct. Mol. Biol. *15*, 71–78.

Grant, R.P., Marshall, N.J., Yang, J.C., Fasken, M.B., Kelly, S.M., Harreman, M.T., Neuhaus, D., Corbett, A.H., and Stewart, M. (2008). Structure of the N-terminal Mlp1-binding domain of the Saccharomyces cerevisiae mRNA-binding protein, Nab2. J. Mol. Biol. *376*, 1048–1059.

Gross, S., and Moore, C. (2001). Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. Proc. Natl. Acad. Sci. USA *98*, 6080–6085.

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.

He, X., and Moore, C. (2005). Regulation of yeast mRNA 3′ end processing by phosphorylation. Mol. Cell *19*, 619–629.

Hirose, Y., and Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93–96.

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

Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N., and Moore, M.J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA *8*, 426–439.

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., et al. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. Nat. Methods *5*, 53–55.

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.

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.

Kyriakopoulou, C.B., Nordvarg, H., and Virtanen, A. (2001). A novel nuclear human poly(A) polymerase (PAP), PAP gamma. J. Biol. Chem. *276*, 33504–33511.

Legrand, P., Pinaud, N., Minvielle-Sebastia, L., and Fribourg, S. (2007). The structure of the CstF-77 homodimer provides insights into CstF assembly. Nucleic Acids Res. *35*, 4515–4522.

Lingner, J., Kellermann, J., and Keller, W. (1991). Cloning and expression of the essential gene for poly(A) polymerase from S. cerevisiae. Nature *354*, 496–498.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R., III. (1999). Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. *17*, 676–682.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82–97.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3′-end processing. Cell. Mol. Life Sci. *65*, 1099–1122.

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

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499–506.

Martin, G., Keller, W., and Doublie, S. (2000). Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. EMBO J. *19*, 41 93–4203.

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

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature *430*, 223–226.

Meinke, G., Ezeokonkwo, C., Balbo, P., Stafford, W., Moore, C., and Bohm, A. (2008). Structure of yeast poly(A) polymerase in complex with a peptide from Fip1, an intrinsically disordered protein. Biochemistry *47*, 6859–6869.

Mirkin, N., Fonseca, D., Mohammed, S., Cevher, M.A., Manley, J.L., and Kleiman, F.E. (2008). The 3′ processing factor CstF functions in the DNA repair response. Nucleic Acids Res. *36*, 1792–1804.

Murthy, K.G., and Manley, J.L. (1992). Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. J. Biol. Chem. *267*, 14804–14811.

Noble, C.G., Beuth, B., and Taylor, I.A. (2007). Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. Nucleic Acids Res. *35*, 87–99.

Ohnacker, M., Barabino, S.M., Preker, P.J., and Keller, W. (2000). The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3′-end-processing complex. EMBO J. *19*, 37–47.

Penheiter, K.L., Washburn, T.M., Porter, S.E., Hoffman, M.G., and Jaehning, J.A. (2005). A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. Mol. Cell *20*, 213–223.

Perez-Canadillas, J.M. (2006). Grabbing the message: structural basis of mRNA 3′UTR recognition by Hrp1. EMBO J. *25*, 3167–3178.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501–512.

Pugh, D.J., Ab, E., Faro, A., Lutya, P.T., Hoffmann, E., and Rees, D.J. (2006). DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways. BMC Struct. Biol. *6*, 1.

Qu, X., Perez-Canadillas, J.M., Agrawal, S., De Baecke, J., Cheng, H., Varani, G., and Moore, C. (2007). The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3′-end processing. J. Biol. Chem. *282*, 2101–2115.

Raabe, T., Bollum, F.J., and Manley, J.L. (1991). Primary structure and expression of bovine poly(A) polymerase. Nature *353*, 229–234.

Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1987). Three-dimensional structure of the large ribosomal subunit from Escherichia coli. EMBO J. *6*, 1107–1114.

Sakai, Y., Saijo, M., Coelho, K., Kishino, T., Niikawa, N., and Taya, Y. (1995). cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product. Genomics *30*, 98–101.

Shi, Y., Reddy, B., and Manley, J.L. (2006). PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. Mol. Cell *23*, 819–829.

Simons, A., Melamed-Bessudo, C., Wolkowicz, R., Sperling, J., Sperling, R., Eisenbach, L., and Rotter, V. (1997). PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb. Oncogene *14*, 145–155.

Simpson, G.G., Dijkwel, P.P., Quesada, V., Henderson, I., and Dean, C. (2003). FY is an RNA 3′ end-processing factor that interacts with FCA to control the Arabidopsis floral transition. Cell *113*, 777–787.

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.

Takagaki, Y., and Manley, J.L. (2000). Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol. Cell. Biol. *20*, 1515–1525.

Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J., and Shenk, T. (1990). A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. Genes Dev. *4*, 2112–2120.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1988). Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell *52*, 731–742.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1989). Four factors are required for 3′-end cleavage of pre-mRNAs. Genes Dev. *3*, 1711–1724.

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.

Vagner, S., Vagner, C., and Mattaj, I.W. (2000). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3′-end processing and splicing. Genes Dev. *14*, 403–413.

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

Vo, L.T., Minet, M., Schmitter, J.M., Lacroute, F., and Wyers, F. (2001). Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. Mol. Cell. Biol. *21*, 8346–8356.

Wallace, A.M., Dass, B., Ravnik, S.E., Tonk, V., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., and MacDonald, C.C. (1999). Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells. Proc. Natl. Acad. Sci. USA *96*, 6763–6768.

Zhao, J., Hyman, L., and 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*, 405–445.

Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182–185.

Molecular Cell, Volume *33*

Supplemental Data

Molecular Architecture of the Human

Pre-mRNA 3′ Processing Complex

Yongsheng Shi, Dafne Campigli Di Giammartino, Derek Taylor, Ali Sarkeshik, William J. Rice, John R. Yates III, Joachim Frank, and James L. Manley

Figure S1. Mutant RNA substrates are defective in 3’ cleavage. 3M-SVL and 3M-SVL-mut RNA substrates were used in cleavage assays. Purified RNAs were resolved by 6% denaturing gel and visualized by phosphorimagery. Pre-mRNA and the 5’ cleaved product were marked.

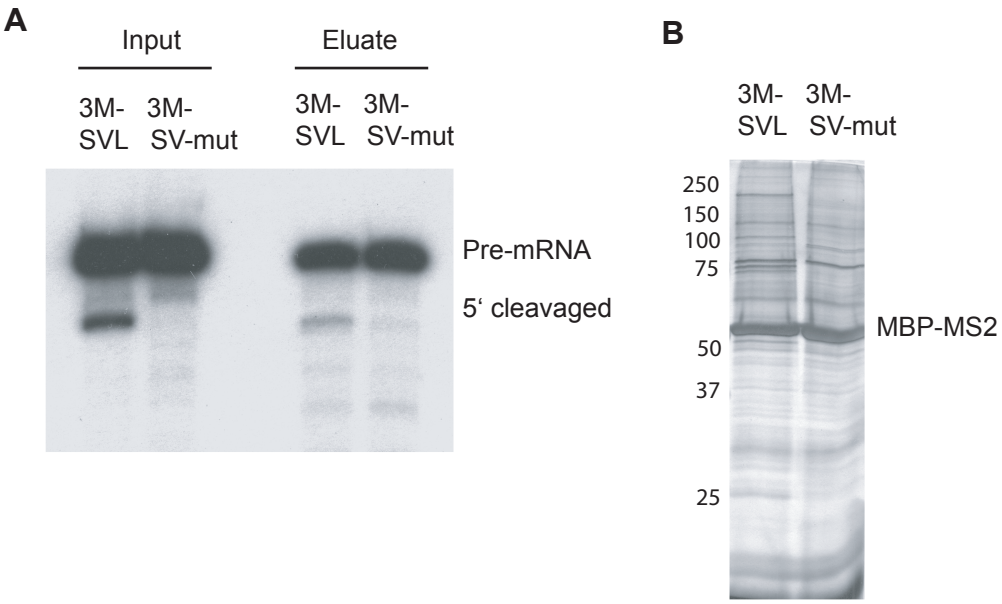


Figure S2. Purification of 3’ processing complexes assembled under cleavage conditions.
A. 3M-SVL or mutant (3M-SVL mut) RNAs were isolated from input cleavage reaction mixtures (input) or from eluted complexes after affinity purification (eluate), resolved on a 6% denaturing gel, and visualized by using a phosphorimager. Pre-mRNA and the 5’ cleaved products were marked. B. Proteins in complexes assembed on 3M-SVL or mutant (3M-SVL mut) RNAs after affinity purification were resolved on SDS-PAGE and visualized by silver staining. The position of the MBP-MS2 protein is marked.

Figure S3. Unique peptides from WDR33 detected in mass spectrometry analyses of the CPSF73 complex.

>gi|56243590|ref|NP_060853.3| WD repeat domain 33 isoform 1 [Homo sapiens]

MATEIGSPPRFFHMPRFQHQAPRQLFY**KRPDFAQQQAMQQLTFD**GKRMRKAVN**RKTIDYN**
PSVIKYLENRIWQRDQRDM**RAIQPDAGYYNDLVPPIGMLNNPMNAVTTKF**VRTSTNKV**KC**
PVFVVRWTPEGR**RLVTGASSGEFTLWNGLTFNFETILQAHDSPVRA**MTWSHNDMWMLTAD
HGGYV**KYWQSNMNNVKM**FQAHKEAIREASFSPTDNKFATCSDDGTVRIWDFLRCHEERIL
RGHGADVKCVDWHPTKGLVVSGS**KDSQQPIKFWDPKT**GQSLATLHAHKNTVMEV**KLNLNG**
NWLLTASRDHLCKLFDI**RNLKEELQVFRG**HK**KEATAVAWHPVHEGLFASGGSDG**SLLFWH
VGVEKEVGGMEMAHEGMIWSLAWHPLG**HILCSGSNDHTSKF**WTRNRPGDKM**RDRYNLNLL**
PGMSEDGVEYDDLEPNSLAVIPGMGIPEQLKLAMEQEQMGKDESNEIEMTIPGLDWGMEE
VMQKDQKKVPQKKVP**YAKPIPAQFQQAWMQNKVPIPAPNEVLNDRK**EDIKLEEK**KKTQAE**
IEQEMATLQYTNPQLLEQLKIERLAQKQVEQIQPPPSSGTPLLGPQPFPGQGPMSQIPQG
FQQPHPSQQMPMNMAQMGPPGPQGQFRPPGPQGQMGPQGPPLHQGGGGPQGFMGPQGPQG
PPQGLPRPQDMHGPQGMQRHPGPHGPLGPQGPPGPQGSSGPQGHMGPQGPPGPQGHIGPQ
GPPGPQGHLGPQGPPGTQGMQGPPGPRGMQGPPHPHGIQGGPGSQGIQGPVSQGPLMGLN
PRGMQGPPGP**RENQGPAPQGMIMGHPPQEMRG**PHPPGGLLGHGPQEMRGPQEI**RGMQGPP**
PQGSMLGPPQELRGPPGSQSQQGPPQGSLGPPPQGGMQGPPGPQGQQNPARGPHPSQGPI
PFQQQ**KTPLLGDGPRAPFNQEGQSTGPPPLIPGLGQQGAQGRIPPLNPGQGPGPNKG**DSR
GPPNHHMGPMSERRHEQSGGPEHGPERGPFRGGQDCRGPPDRRGPHPDFPDDFSRPDDFH
PDKRFGHRLREFEGRGGPLPQEEKWRRGGPGPPFPPDHREFSEGDGRGAA**RGPPGAWEGR**
RPGDERFPRDPEDPRFRGRREESFRRGAPPRHEGRAPP**RGRDGFPGPEDFGPEENFDASE**
EAARGRDLRGRGRGTPRGGR**KGLLPTPDEFPRF**EGGRKPDSWDGNREPGPGHEHFRDTPR
PDHPPHDGHSPASRE**RSSSLQGMDMASLPPRK**RPWHDGPGTSEH**REMEAPGGPSEDRG**GK
GRGGPGPAQRVPKSGRSSSLDGEHHDGYHRDEPFGGPPGSGTPSRGGRSGSNWGRGSNMN
SGPPRRGASRGGGRGR

*Sequences covered by unique peptides detected by mass spectrometry were underlined.

*Total number of unique peptides detected: 55

*Total number of spectrum: 521

*Coverage: 38.4%

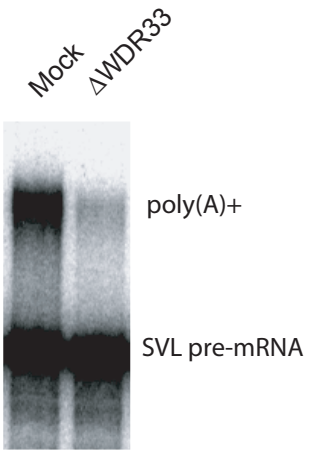
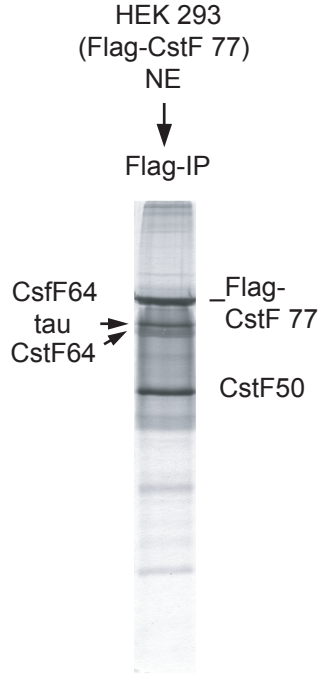


Figure S4. Depletion of WDR33 abolishes polyadenylation.
Mock-depleted (mock) and WDR33-depleted (ΔWDR33) NE were used in polyadenylation assays with SVL substrate. Purified RNAs were resolved on a 6% denaturing gel and visualizedusing a Phosphorimager. Pre-mRNA and poly(A)+ RNAs are marked.

A



B

>gi|14149675|ref|NP_056050.1| CstF 64, tau [Homo sapiens]

MSSLAVRDPAMDRSL**RSVFVGNIPYEATEEQLKDIFSEVGSVVSFRL**
VYDRETGKP**KGYGFCEYQDQETALSAMRN**LNGREFSGRALRVDNAAS
EKNKEEL**KSLGPAAPIIDSPYGDPIDPEDAPESITRAVASLPPEQMF**
ELMKQMKLCVQNSHQEA**RNMLLQNPQLAYALLQAQVVMRIMDPEIAL**
KILHRKIHVTPLIPG**KSQSVSVSGPGPGPGPGLCPGPNVLLNQQNPP**
APQPQHLARRPV**KDIPPLMQTPIQGGIPAPGPIPAAVPGAGPGSLTP**
GGAMQPQLGMPGVGPVPLERGQVQMSDPRAPIP**RGPVTPGGLPPRGL**
LGDAPNDPRGGTLLSVTGEVEPRGYLGPPHQGPPMHHASGHDTRGPS
SHEMRGGPLGDPRLLIGEPRGPMIDQRGLPMDGRGGRDSRAMET**RAM**
ETEVLETRVMER**RGMETCAMETRG**MEARGMDARGLEMRGPVPSS**RGP**
MTGGIQGPGPINIGAGGPPQGPRQVPGISGVGNPGAGMQGTGIQGTG
MQGAGIQGGGMQGAGIQGVSIQGGGIQGGGIQGAS**KQGGSQPSSFSP**
GQSQVTPQDQEKAALIMQVLQLTADQIAMLPPEQRQSILILKEQIQK
STGAS

*Sequences covered by unique peptides detected by mass spectrometry were underlined.
*Total number of unique peptides detected: 50
*Total number of spectrum: 1013
*Coverage: 56.3%

Figure S5. CstF64 tau is a component of the CstF complex.
(A) Immuno-purification of the CstF complex. NE was made from a stable HEK293 cell line expressing Flag-CstF77, and IP was performed using anti-Flag antibodies. Purified proteins were resolved and stained with silver. CstF components are marked.
(B) Mass spectrometry analyses of the CstF complexdetect CstF64 tau.

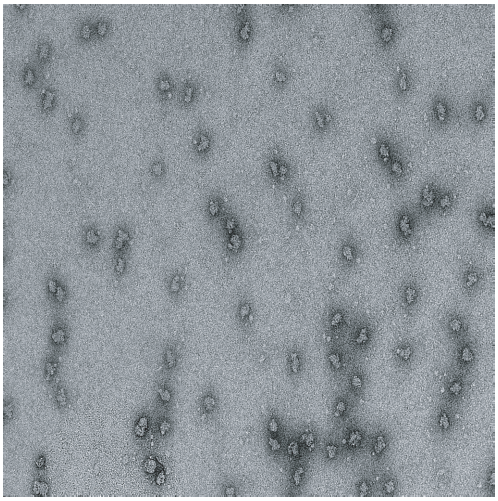


Figure S6. Tilted image of the negatively stained 3' processing complexes. A 50° tilted image of the negatively stainedparticles. The tilted images indicate that the sample is fully sandwiched between carbon membranes, and the staining seems homogeneous.

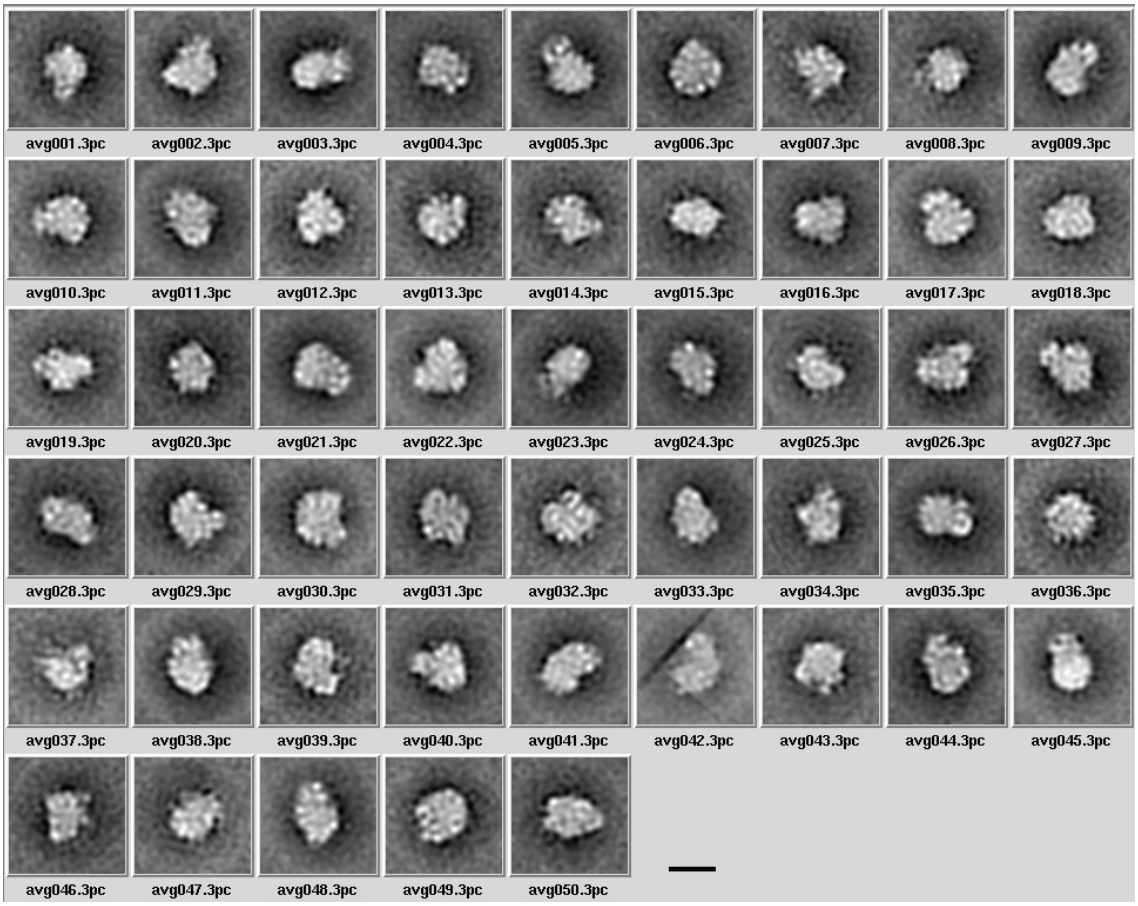


Figure S7. Class average images of the 3' processing complex using SPIDER. 3,671 images of negatively stained single particles were classified into 50 groups and class averages were obtained after reference-free alignment using SPIDER. Scale bar, 20nm.

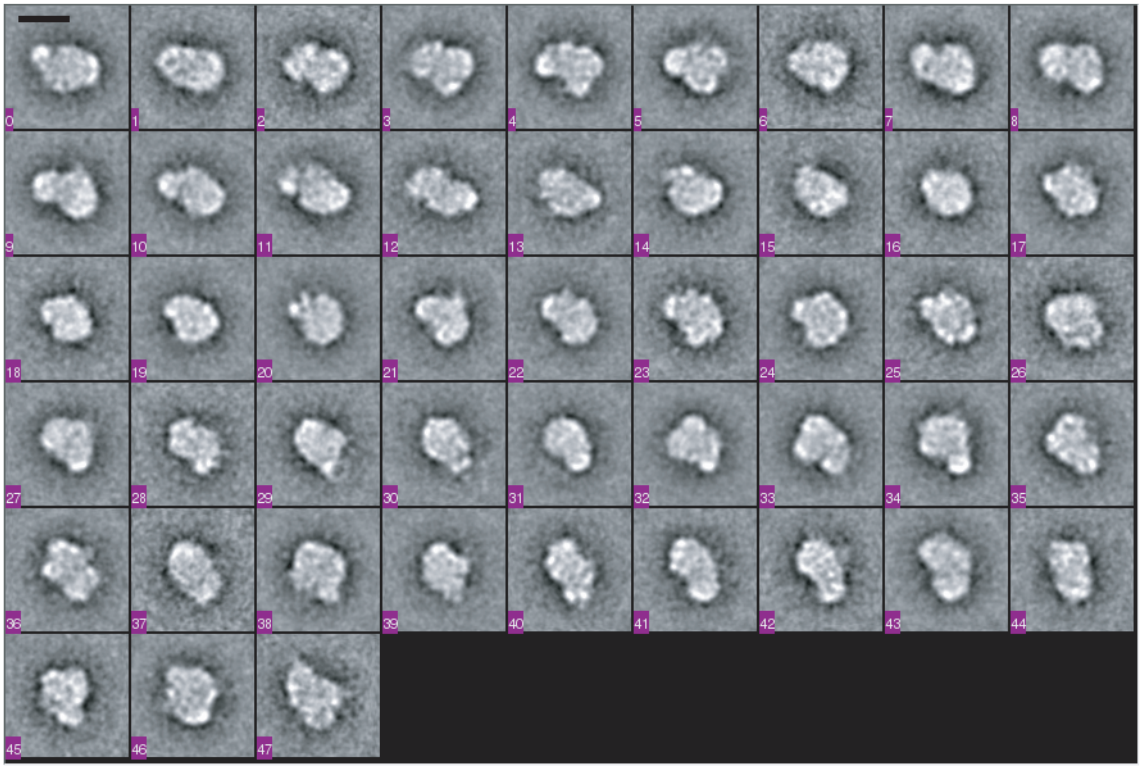


Figure S8. Class average images of the 3' processing complex obtained with EMAN. 3,671 images of negatively stained single particles were classified into 47 groups (~40-110 particles in each group) and class averages were obtained after reference-free alignment using EMAN. Scale bar, 10nm.

Table S1. Protein composition of the human pre-mRNA Cleavage complexes

Protein Name	Accession #	Mol. Weight	# peptides	
			L3	SVL
CPSF160	NP_037423	160822	18	17
CPSF100	gi 51338827	88487	15	9
CPSF73	NP_057291	77486	5	3
CPSF30	NP_006684	30124	1	0
hFip1	NP_112179	66526	12	12
CstF77	NP_001317	82922	16	16
CstF64	NP_001316	60959	12	10
CstF50	NP_001315	48358	5	6
CF Im 25	NP_008937	26227	14	7
CF Im 59	NP_079087	52050	11	5
CF Im 68	NP_008938	59209	12	7
Symplekin	NP_004810	126500	11	3
PABP 1	NP_002559.1	70324	5	1
WDR33	NP_060853	145921	10	7
RBBP6	NP_008841	201563	6	0
PP1 beta	NP_002700	37187	1	3
DNA topoIIalpha	NP_001058	174384	4	5
PARP1	NP_001609	113135	5	2
DNA-PK	NP_008835	469093	14	8
Ku 70	NP_001460	69843	4	2
Ku 86	NP_066964	82705	5	3
MDC1	NP_055456	226643	1	1
THO complex subunit 4	NP_005773	26757	3	4
THO complex subunit 6	AAH03118	32891	1	1
FACT complex large subunit	NP_003137	119914	2	2
NELF B subunit (BRCA1 coactivator)	NP_060853	61640	2	9
NELF E subunit	NP_002895.3	43240	1	1
BAF53a	NP_004292	47461	1	2
BRG1/SMARCA4	NP_003063	184644	1	1
BAF170	NP_003066	132879	1	0
BAF155	NP_003065	122753	1	0
BAF60b/SMARCAD	NP_003068	54945	1	0
Enhancer of rudimentary	NP_004441	12259	1	2

homolog (repressor)				
Bre1A	NP_149974	113977	1	3
U1-70K	NP_003080	70082	2	1
U4/U6.U5 tri-snRNP associated protein 1	NP_055317	55181	4	2
U4/U6.U5 tri-snRNP associated protein 2	NP_005137	90255	1	2
SF3b155 (SAP155)	NP_036565	145815	6	3
P54/NRB	NP_004759	53542	1	2
ASF/SF2	NP_008855	27745	1	2
SR-A1	NP_067051	139296	1	?
SRp38	NP_473357	31345	1	1
SRrp86 (SR12)	NP_631907	71650	1	2
SRm300	NP_057417	299676	6	3
(KSRP) KH-type splicing regulatory protein (also mRNA turnover)	NP_003676	73161	2	2
hnRNPA0	NP_006796.1	30841	4	1
hnRNP A3	NP_005749.1	29357	4	2
hnRNP U-like (E1B5-associate protein 5)	NP_008971.2	95739	3	2
hnRNP D	NP_005454.1	46437	2	1
RNA helicase A (DEAD/H box-9)	NP_001348	142069	7	1
DEAD/H box-39	NP_005795.2	49130	2	1
DEAD/H box-49	NP_-61943	54226	1	1
DEAH box-8	NP_004932	139314	1	1
DEAD box-21	NP_004719	87344	2	2
DEAD/H box-15	NP_001349	92829	2	2
DEAD box 42	NP_987095	102975	1	1
Exosome 10		100831	3	1
RRP4		32789	0	1
RRP40		29441	0	1
RRP41		26252	2	0
RRP42		31835	1	0
RRP43		30040	1	0
EEF1-alpha	NP_001393	50141	3	2
EEF1-epsilon	NP_004271.1	37974	1	1
EIF 4B	NP_001408	69224	4	2
40S ribosomal protein S5	NP_001000.2	22745	1	1
40S ribosomal	NP_001008.1	17091	1	2

protein S13				
40S ribosomal protein S24	NP_148982.1	15069	1	1
40S ribosomal protein S26	NP_001020.2	13015	1	1
Similar to 60S ribosomal protein L7a	NP_000963.1	29996	1	1
60S ribosomal protein L8	NP_000964.1	28025	2	2
60S ribosomal protein L9	NP_000652.2	21863	2	1
60S ribosomal protein L24	NP_000977.1	17779	2	1
60S ribosomal protein L27	NP_000979.1	15798	1	1
zinc finger CCHC domain-containing protein 8 (ZCCHC8)	NP_060082	79375	8	3
Scaffold attachment factor B	NP_002958.2	102768	3	2
TAR DNA-binding protein	NP_031401.1	33730	3	2
Lamin-A/C	NP_005563.1	74140	3	2
Clathrin heavy chain	NP_004850	191613	3	5
Lamina-associated polypeptide 2 alpha (Thymopoietin)	NP_003267.1	75361	3	2
Phenyalanyl-tRNA synthetase beta	NP_00678.2	66130	1	1

*Components of multi-subunit complexes that are present in only one purified complexe are listed and lightly shaded.

Table S2: Protein compositions of the CPSF and CstF subcomplexes, and comparison with that of the human pre-mRNA 3' processing complex.

Protein Name	Accession #	Yeast homolog	Motifs	Cal. Mass	# of peptides			
					L3	SVL	CPSF73-3Flag	Flag-CstF77
KNOWN POLYADENYLATION FACTORS								
CPSF Complex								
CPSF160 (CPSF1)	NP_037423	CFT1	SFT1	160822	80	72	87	22
CPSF100 (CPSF2)	NP_059133.1	CFT2	β-CASP	88487	59	52	83	24
CPSF73 (CPSF3)	NP_057291	YSH1	β-CASP	77486	42	22	81	9
CPSF30 (CPSF4)	NP_006684	YTH1	Zinc finger	30124	18	13	19	2
hFip1	NP_112179	FIP1		66526	23	19	35	13
CstF Complex								
CstF77 (CSTF3)	NP_001317	RNA14	HAT	82922	60	54	3	118
CstF64 (CSTF2)	NP_001316	RNA15	RRM	60959	31	22	11	45
CstF50 (CSTF1)	NP_001315		WD repeats	48358	21	18	0	48
CF Im Complex								
CF Im 68 (CPSF6)	NP_008938		RRM	59209	15	10	0	0

CF Im 59	NP_079087		RRM	52050	15	12	0	0
Cf Im 25 (CPSF5)	NP_008937			26227	26	21	0	0
CF IIm Complex								
Pcf11	NP_056969.2	PCF11	CID	173050	5	3	0	0
Other known polyadenylation factors								
Symplekin	NP_004810	PTA1		126500	51	39	31	0
Symplekin variant	BAE06092.1			118793	48	38	33	0
PAPOLG	NP_075045	PAP	RRM	82803	2	3	0	0
PABPC1	NP_002559.2	PAB1	RRM	70671	27	33	3	0
PABPC4	NP_003810.1		RRM	70783	20	23	3	0
PABPN1	NP_004634		RRM	31749	14	12	0	0
(Putative) homologues of yeast polyadenylation factors								
CstF64 tau (CSTF2T)	NP_056050.1	SPAC644.16 (S. pombe)	RRM	64436	21	16	8	50
WDR33	NP_060853	PFS2	WD repeats	145921	49	45	55	10
RBBP6	NP_008841	MPE1	RS, DWNN	201563	4	3	0	0
PP1 alpha	NP_002699.1	GLC7	Phosphatase	37512	2	7	2	0
PP1 beta	NP_002700	GLC7	Phosphatase	37187	2	5	0	0
DNA Damage Response Factors								
DNA-PK	NP_008835		Kinase	469093	21	15	0	0
Ku 70	NP_001460	YKU70	DNA helicase	69843	4	7	0	0
Ku 86/XRCC5	NP_066964	YKU86	DNA helicase	82705	6	13	0	0

PARP-1	NP_001609			113135	5		0	0
RNAP II and associated factors								
Pol II large subunit Rpb1	NP_000928.1	RPB1	Polymerase	217204	2	2	2	0
Pol II B	NP_000929.1	RPB2		133896		2	4	0
Rpb5 (Pol II E)	NP_002686.2	RPB5		24611	2		0	0
Rpb11	AF468111_1	RPB11		17208		2	0	0
Transcription Factors								
TF II, I	NP_127492.1			112416	8	2	0	0
TAF15	AAH46099.1		RRM, zf-RanBP	52061	6	5	0	0
Integrator Complex								
INTS2	NP_065799.1			134346	2	2	0	0
INTS3	NP_075391.3			118013	4	9	0	0
INTS4	NP_291025.3			108171	2	3	0	0
INTS6	NP_036273.1			100390	2	3	0	0
INTS7	NP_056249.1			106834	2	2	0	0
INTS8	NP_060334.2			113088		3	0	0
INTS9	NP_060720.1		YSH1 homology	73815		4	0	0
INTS10	NP_060612.2			82236		2	0	0
PAF Complex								
Cdc73	AAH14351.2	CDC73		46675	2	2	0	0
FACT Complex								

FACT complex large subunit	NP_009123	SPT16		119914	4	4	0	0
Splicing Factors								
p54nrb	NP_031389.3			54232	11	3	2	0
PSF	NP_005057.1			76150	10	4	3	0
PUF60	NP_510965.1		RRMs	59876	3	9	0	0
UAP56	NP_004631.1	SUB2	RNA Helicase	48991	12	9	3	0
SF1	NP_004621.2	BBP		68330	5	5	0	0
SF3A, subunit 1	NP_005868.1	PRP21	Surp	88886	9	15	0	0
SF3A, subunit 2	NP_009096.2	PRP11		49256	2	4	0	0
SF3A60	CAA57388	PRP9		58777	3	18	0	0
SF3B, subunit 1	NP_036565.2	HSH155		145830	16	34	5	0
SF3B, subunit 2	NP_006833.2	CUS1		100228	2	21	3	0
SF3B, subunit 4	NP_005841.1	HSH49		44386	3	5	0	0
SF3B, 14 kDa	NP_057131.1	SNU17		14585		5	0	0
Prpf38b	NP_060531.1			64441	2	8	0	0
SRm300	NP_057417.2		RS	299676	13	15	0	0
U1 70K	NP_003080.2	SNP1	RRM	51557	7	7	3	0
U2AF65	NP_009210.1	MUD2	RRM	53501	6	11	0	0
U2AF35	NP_006749.1		RRM	27872	2	3	0	0
PRP19	NP_055317.1	PRP19	RING, WD	55181	5	3	6	0
Exosome								
SKIV2L2 (hMTR4)	NP_056175.2	MTR4	DEXD	117933	29	28	0	0
Translation Factors								

eEF1-alpha	NP_001393.1	TEF1/2	GTPase	50141	14	12	5	0
eEF1-gamma	NP_001395.1	TEF4	GST	50119	5	4	0	0
eIF2A	NP_114414.2	YGR054W		64990	4	4	0	0
eIF 4A	NP_001407.1	TIF1	Helicase	46154	12	6	0	0
eIF3 gamma	NP_003747.1		JAB/MPN	39930	5	6	0	0
eIF3S2	NP_003748.1	TIF34	WD	36502	2	2	0	0
eIF3S5	NP_003745.1		PCI	37564	3	2	0	0
eIF3S6	NP_001559.1			52221	2	4	0	0
eIF3S9 eta	NP_003742.2		RRM	92482	7	7	0	0
eIF3A	NP_003741.1	RPG1		166569	16	9	0	0
eIF4G1	NP_886553.2	TIF4631		175460	2	2	0	0
RACK1/GNB2L1/lung cancer oncogene 7	NP_006089.1	ASC1	WD	35077	5	5	0	0
Factors with known motifs								
RNA binding motif protein 7	NP_057174.1		RRM	30503	4	4	0	0
RNA binding motif protein 9/Ataxin-binding protein/Fox-1	NP_055124		RRM	39515	2	7	0	0
RNA binding motif protein 25	NP_067062.1		RRM	100186	4	16	0	0
RNA binding motif protein 39, isoform a	NP_909122.1		RRM	59380	2	4	2	0
RNA binding motif protein 39, isoform b	NP_004893.1		RRM	58657	2	4	2	0

DEAH box polypeptide 3	NP_001347.3	DBP1	RNA helicase	7244	13	9	0	0
DEAH box polypeptide 5	NP_004387.1	DBP2	RNA helicase	69148	4	5	0	0
DEAH box polypeptide 9	NP_001348.2		RNA helicase	140958	32	3	0	0
DEAH box polypeptide 15	NP_001349.2	PRP43	RNA helicase	90933	3	17	0	0
DEAH box polypeptide 36	NP_065916.1		RNA helicase	114776	12	27	0	0
DEAH box polypeptide 39	NP_005795.2		RNA helicase	49130	8	5	0	0
DEAD box polypeptide 6	NP_004388.1	DHH1	RNA helicase	82432	3	2	0	0
DEAD box polypeptide 17 (p82)	NP_006377.2		RNA helicase	80273	6	4	0	0
Gemin3 (DEAD box polypeptide 20)	NP_009135.3		RNA helicase	92213	4	2	0	0
DEAD box polypeptide 23	NP_004809.2	PRP28 (S. Pombe)	RNA helicase	95583	4	5	0	0
ZCCHC8	NP_060082		Zinc finger	79375	15	16	0	0
EBNA2 coactivator, p100/SND1	NP_055205.2		Staphylococcal nuclease domain, Tudor	101997	12	2	0	0
Interferon-induced protein with tetratricopeptide repeats 1 isform 2	NP_001539.3		TPR	55360	9	7	0	0

Interferon-induced protein with tetratricopeptide repeats 3	NP_001540		TPR	55985	8	5	0	0
Other Factors								
G3BP/Ras-GTPase-activating protein SH3-domain-binding protein	NP_005745.1		Endoribonuclease	52164	5	2	0	0
JUP protein/ gamma-catenin	NP_002221.1			81745	5	3	0	0
PNUTS (PPP1R10)	NP_002705.2		PP1 regulator	99058	6	9	0	0