Title

The Landscape of RNAs Bound by Poly(A) Binding Proteins PABPN and PABPC

Author

Nicholson-Shaw, Angela

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

The Landscape of RNAs Bound by Poly(A) Binding Proteins PABPN and PABPC

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Biology

by

Angela Nicholson-Shaw

Committee in charge:

Professor Amy Pasquinelli, Chair
Professor Susan Ackerman
Professor Heidi Cook-Andersen
Professor Jens Lykke-Andersen
Professor Eugene Yeo

2021

The dissertation of Angela Nicholson-Shaw is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To all the bright, young females who are in search of their path. May every door be opened to you and may you have mentors to show you those doors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my doctoral advisor, Amy Pasquinelli, for being a supportive mentor who celebrated even the small accomplishments with me, and who pushed me to develop a level of independence and critical thinking that would not have been otherwise possible.

All members of the Pasquinelli lab made this journey much more enjoyable. Both inside and outside the lab, the conversations and companionship we share has helped make the lab a happy place to work.

I would like to thank the friends that I have made throughout my PhD, especially those who joined the program in the same year as I did. So many of us formed a quick bond early on and have supported each other through every challenge life throws at you, far beyond just the difficulties of doctoral work. A special thanks goes to “da ladies,” who have been a sounding board for every type of success and failure in life. Emily, I somehow knew immediately in New York that I should hang on to you. Your honesty and compassion are invaluable to me. Laura, we joined the lab at the same time and it’s hard to imagine this voyage without you there. I’m thankful to have you to share the ups and downs of bench work.

The beginning of this journey would not have been possible without the guidance of Dr. Behzad Varamini. You saw my potential before I did, and I honestly do not believe I would have applied to summer research opportunities or to graduate programs without your belief in me and your constant expectation that I better bring my next application draft by your office soon.

A big thanks goes to all my family for supporting me throughout my PhD. My parents, Rita and Scott, taught me to enjoy the process of learning early on, both inside the classroom and out. They have whole-heartedly supported me no matter what career path I chose to take, including the twists and turns. Jessica and Mike, thanks for always providing a listening ear or a drink and a board game, depending on what was needed.

I would like to express my incredible gratitude for my partner, Tim Nicholson-Shaw. You are a tremendous listener and an innovative scientist. The completion of this dissertation would not have been possible without you.

Chapter 2, in full, is a reprint of material as it occurs in *Trends in Cell Biology,* "Tales of Detailed Poly(A) Tails," Nicholson, A.L., Pasquinelli, A. E., 2019. I was the primary author.

Chapter 3, in full, is a reprint of material as it is prepared for submission to *NSMB*, "Distinct Transcripts and Isoforms Favor the Nuclear and Cytoplasmic Poly(A) Binding Proteins (PABPs)," Nicholson-Shaw, A.L., Kofman, E., Yeo, G.W., and Pasquinelli A.E., 2021. I was the primary author.

Chapter 4, section 4.4, contains unpublished material coauthored with Ian Nicastro and Amy Pasquinelli. I was the primary author of this material.

VITA

2006- 2008	Associate Degree Allan Hancock College, Santa Maria, CA
2012- 2015	Bachelor of Science in Biochemistry Biola University, La Mirada, CA
2017- 2020	National Science Foundation – Graduate Research Program Fellow
2015- 2021	Doctor of Philosophy in Biology University of California San Diego

PUBLICATIONS

Azoubel Lima S, Chipman LB, **Nicholson AL,** Chen YH, Yee BA, Yeo GW, Coller J and Pasquinelli, AE. Short poly(A) tails are a conserved feature of highly expressed genes. *Nature Structural and Molecular Biology*. 2017.

Nicholson AL, and Pasquinelli AE. Tales of Detailed Poly(A) Tails. *Trends in Cell Biology*. 2019.

Schisknis, EC, **Nicholson AL,** Modena MS, Pule MN, Arribere JA, and Pasquinelli AE. Auxin-independent depletion of degron-tagged proteins by TIR1. *Biology Micropub*. 2020.

ABSTRACT OF THE DISSERTATION

The landscape of RNAs bound by poly(A) binding proteins PABPN and PABPC

by

Angela Nicholson-Shaw

Doctor of Philosophy in Biology

University of California San Diego, 2021

Professor Amy Pasquinelli, Chair

Nearly all eukaryotic RNAs have a continual tract of polyadenosine at their end, called a poly(A) tail. Unlike the main portion of the RNA, this addition is added after transcription has occurred and is not genomically encoded. Poly(A) tails have been implicated in a wide range of roles such as protecting the RNA and facilitating translation.

In a cellular environment, RNAs are covered with an array of proteins that influence how an RNA is used and ultimately determine its fate. The poly(A) tail is bound

primarily by the nuclear and cytoplasmic poly(A) binding proteins, PABPN and PABPC. These proteins play integral roles in the creation and subsequent protection of the poly(A) tail, partnering with numerous other proteins along the way.

Chapter 2 summarizes recent discoveries and technological advances in understanding poly(A) tail dynamics. Sequencing of poly(A) tails is challenging due to their repetitive nature, but advances made in the last ten years have begun to make this possible. This has enabled genome-wide approaches to understanding native poly(A) tail lengths in numerous organisms, leading to profound discoveries about the range of poly(A) tail sizes that can exist on an RNA.

Chapter 3 presents an in-depth study of poly(A) binding proteins PABPN and PABPC, and the RNAs that they are bound to at steady state in human cells. Although PABPN and PABPC are likely binding to nearly all RNAs at some point in their life cycle, they each have a unique enrichment and depletion binding profile. PABPC tends to be enriched with well-translated RNAs with long half-lives and shorter poly(A) tails. PABPN tends to be enriched with poorly translated RNAs with longer poly(A) tails.

To advance these types of studies in another organism, I have developed tools and approaches for investigating the roles of PABPN, PABPC, and the poly(A) tail in *Caenorhabditis elegans*. In chapter four, I will discuss these resources. This includes the introduction of Nanopore long-read sequencing to our lab, and details how this was used to look at poly(A) tail length in *C. elegans*.

Overall, this work deepens our understanding of the roles of PABPN and PABPC throughout the lifetime of an RNA.

Chapter 1

Introduction

1.1 Transcription, Splicing, Cleavage and Polyadenylation of RNA

All gene expression arises from copying sections of our permanent genetic information, our DNA, into transient RNA molecules that are able to travel throughout the cell. Although our DNA is stowed tightly in the nucleus of a cell, RNA can take this information and deliver it in a useful form, responding to the needs of an organism.

From the very beginning of its existence, an RNA quickly becomes coated with a myriad of proteins, forming a ribonucleoprotein (RNP) complex (Singh et al. 2015). More than 1,400 partner proteins have been identified by this point, with the number still growing (Singh et al. 2015; Hentze et al. 2018; Mallam et al. 2019). Each protein has a unique function and can dictate what will happen to that RNA. For example, some proteins contain localization information that will take the RNA to a particular organelle within the cell, acting as a taxi driver to get the RNA to the right destination. Other proteins are involved in decay and will lead to destruction of the RNA that they are bound to. Some proteins are responsible for acting on the RNA itself, such as mutating the sequence that is in that RNA. Viewing an RNA and its associated proteins as a dynamic and complex RNP structure gives a much more accurate picture of what is happening inside a cell.

The creation of a eukaryotic RNA consists of numerous interconnected steps, each one facilitated by specific proteins. Transcription and subsequent pre-mRNA processing (adding a 5' cap, splicing of introns, 3' cleavage and for the majority of RNAs, polyadenylation) requires coordination of huge nuclear complexes. These steps are vital to produce a mature RNA that will go on to serve its intended function in the cell.

Capping is the first modification that occurs on an RNA. When RNA polymerase II begins transcription of a gene, a cap is added after the first 25-30 nucleotides have been incorporated and consists of a 7-methylguanosine moiety that is linked to the growing RNA by an unusual 5'-5' triphosphate bridge (Shatkin 1976; Moteki and Price 2002). This cap serves a range of biological functions such as protecting the 5' end of the RNA from 5'-3' exonucleolytic decay and facilitating RNA splicing, polyadenylation, and export of the RNA into the cytoplasm (Topisirovic et al. 2011). In the cytoplasm, the cap also serves to recruit translation initiation factors (Merrick 2004).

Most protein-coding genes, as well as some non-coding genes, contain intervening intronic sequences which need to be spliced out so that the neighboring exons can be ligated together. On average, protein coding genes have about 11 exons and 10 introns, but great variation exists (Piovesan et al. 2019). The length of introns is far greater than that of exons. Average human intron length is 6938 base-pairs (bp) but only 311 bp for exons (Piovesan et al. 2019). Growing evidence supports the idea that splicing largely occurs co-transcriptionally (Ameur et al. 2011; Khodor et al. 2011; Oesterreich et al. 2016; Osheim, O.L. Miller, and Beyer 1985). Intron boundaries are recognized by the spliceosome as the RNA is being transcribed by RNA polymerase. Introns are excised

and the resultant ligated exons create an open reading frame (ORF) for protein coding genes, which will be later decoded by ribosomes during the translation process.

Throughout transcription, the nascent RNA is physically connected to its blueprint DNA by RNA polymerase (Figure 1.1). In order to detach from RNA polymerase, the RNA undergoes a two-step process: endonucleolytic cleavage followed by polyadenylation. Selecting the site of cleavage involves recognition of a hexamer element termed the poly(A) site (PAS), which has a consensus sequence of AAUAAA (Tian and Graber 2012; Proudfoot 2011). The PAS is identified by Cleavage and Polyadenylation Specificity Factor (CPSF). Additionally, a downstream GU- or U-rich element is recognized by Cleavage Stimulating Factor (CSTF). Collectively, these protein complexes help guide cleavage to occur between those two sites. After cleavage has occurred, poly(A) polymerase (PAP) begins synthesizing a poly-adenosine (poly(A)) tail on the 3' end of the RNA. After PAP has synthesized the first 11-14 nucleotides of the poly(A) tail, Poly(A) Binding Protein Nuclear (PABPN) is now able to bind the growing tail, which causes PAP to switch to processive synthesis, rapidly completing the creation of the tail (Bienroth, Keller, and Wahle 1993; Wahle 1991). A full-length poly(A) tail is thought to be around 200-250 nucleotides (nt) long when it is first made (Diana Sheiness and Darnell 1973; Brawerman 1981; Michael D Sheets and Wickens 1989).

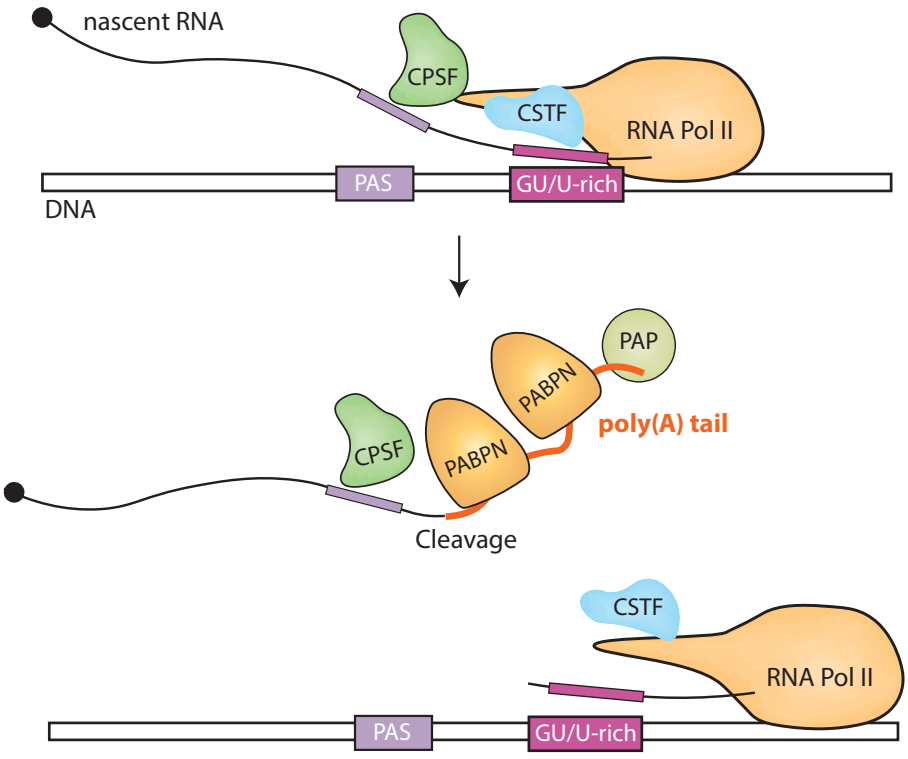


Figure 1.1: Cleavage and Polyadenylation of RNA. Endonucleolytic cleavage releases the nascent RNA from being tethered to RNA Polymerase II and the DNA. Polyadenylation by PAP occurs at the 3' end following cleavage. CPSF = Cleavage and Polyadenylation Specificity Factor. PAS = Poly(A) Site. CSTF = Cleavage Stimulating Factor. PABPN = Poly(A) Binding Protein Nuclear. PAP = Poly(A) Polymerase

After these nuclear processing steps are complete, an RNA is considered mature and ready for export into the cytoplasm if that is its final destination. The nuclear membrane is covered with nuclear pore complexes (NPCs) that selectively allow cargo in and out of the nucleus. Mature RNAs are actively recruited and channeled through NPCs into the cytoplasm.

1.2 Dynamic RNP remodeling on RNAs

When an RNA arrives in the cytoplasm, it now must undergo many changes regarding the proteins that are bound. For a messenger RNA (mRNA), RNP remodeling is what allows that message to be recognized as a substrate for translation and decoded into a protein.

In the nucleus, the 5' methylguanosine cap is bound by the cap-binding complex (CBC), which is replaced by eIF4F in the cytoplasm (Figure 1.2). The eIF4F complex consists of the cap-binding subunit eIF4E which is bound to a scaffolding protein, eIF4G, that is able to recruit additional subunits including eIF4A, an RNA helicase that assists in unwinding of the substrate RNA (Topisirovic et al. 2011). The vast majority of eukaryotic RNAs are translated though a cap-dependent mechanism, although select RNAs can begin translation through an internal ribosome entry site (IRES) (Spriggs, Bushell, and Willis 2010). Cap-dependent translation requires eIF4F due to eIF4G binding of eIF3 (Richter and Sonenberg 2005). eIF3 recruits the 43S pre-initiation complex which will scan across the 5' untranslated region (UTR) until it arrives at the start codon (typically AUG) and begins translation. eIF4G is also able to bind to the cytoplasmic poly(A) binding protein (PABPC), thus facilitating interactions both with initiation at the 5' end of the RNA and the poly(A) tail at the 3' end (Imataka, Gradi, and Sonenberg 1998). These interactions are thought to promote translation through a 'closed-loop' structure where the RNA forms an end-to-end circle due to these protein interactions at each end (Gallie 1991; Preiss and Hentze 1998; Tarun and Sachs 1995; Tarun et al. 1997; Imataka, Gradi, and Sonenberg 1998). However, recent studies reevaluating the closed-loop have suggested that although it can form, this may not be

the predominant state of all translating RNAs (Archer et al. 2015; Gilbert and Thompson 2016).

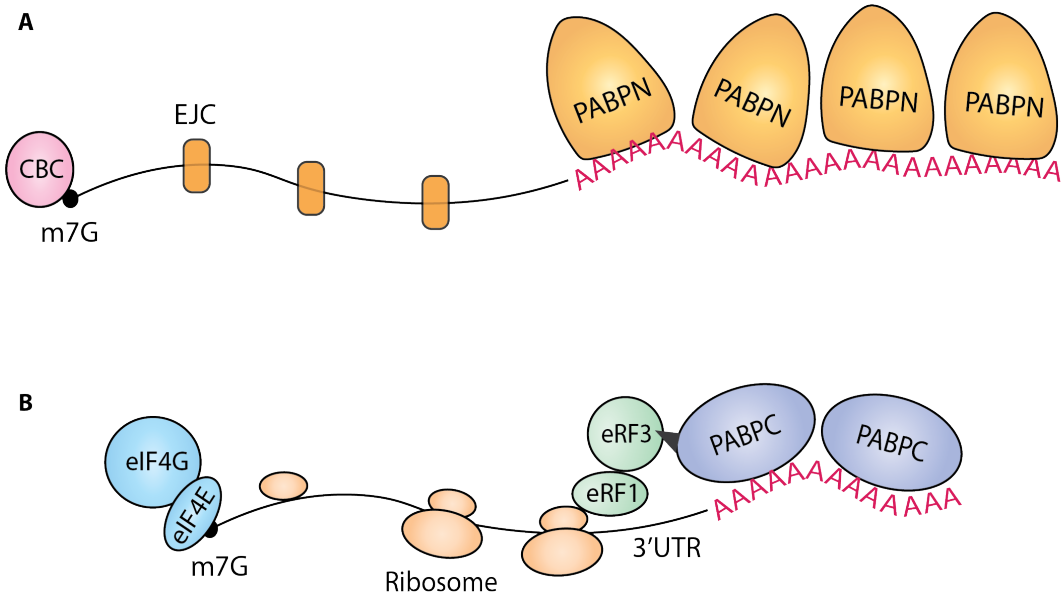


Figure 1.2: RNP remodeling. A) A simplified mRNA when it has just arrived in the cytoplasm following export from the nucleus. B) An mRNA after it has undergone substantial mRNP remodeling and now is being translated in the cytoplasm. CBC = Cap binding complex. EJC = Exon-junction complex.

Inherent in their naming, PABPN and PABPC are predominantly nuclear and cytoplasmic, respectively. PABPN was first identified as a factor important for polyadenylation of nascent RNA, binding to the first few nucleotides of the poly(A) tail and allowing PAP to rapidly synthesize a full length tail (Wahle 1991; Bienroth, Keller, and Wahle 1993; Krause et al. 1994). The initial discovery of PABPC was as a protein highly abundant in the cytoplasm (Blobel 1972). More recent evidence has shown that PABPN and PABPC are shuttling proteins, existing in lower abundance in the opposite compartment as well (Afonina, Stauber, and Pavlakis 1998; Calado and Carmo-Fonseca

2000; Woods et al. 2002; Hosoda, Lejeune, and Maquat 2006). Most RNAs in the cytoplasm ultimately have PABPC coating their poly(A) tail, but the details regarding how this transition from PABPN to PABPC occurs are unknown. PABPC binds to multiple translation factors including the aforementioned initiation factors as well as the release factor eRF3 which, along with eRF1, allows the ribosome to stop translation and release the nascent peptide (Figure 1.2). Translation itself has been suggested as a potential turning point for the transition between PABPN and PABPC (Sato and Maquat 2009).

Another main area of remodeling is splicing-dependent proteins. The splicing machinery leaves behind proteins during the process of ligating exons together, 'marking' where splicing previously occurred (Le Hir et al. 2000). Termed the exon-junction complex (EJC), these proteins stably bind about 20-24 nucleotides upstream of an exon-exon junction and travel with the RNA during transport into the cytoplasm. Once in the cytoplasm, the EJC will normally be displaced during the first round of translation and recycled back to the nucleus (Lejeune et al. 2002; Dostie and Dreyfuss 2002). If the EJC is not successfully displaced, such as during premature translation termination which occurs in nonsense-mediated decay (NMD), the EJC serves as a beacon to initiate decay processes (Kervestin and Jacobson 2012). The continued presence of the EJC signals to the cell that this is a problematic RNA that could create aberrant protein products whose presence may cause deleterious effects.

1.3 Steady State Poly(A) Tail Length

Although poly(A) tails are created around 200-250 nt long, they do not remain at this length throughout their lifetime. Early studies into poly(A) tail length recognized that broad scale shortening in the cytoplasm was occurring, but the purpose and cause of this deadenylation was unclear (D Sheiness, Puckett, and Darnell 1975; Greenberg and Perry 1972; Diana Sheiness and Darnell 1973; Palatnik, Storti, and Jacobson 1979). Almost 50 years later, scientists are still investigating the exact cause and effect of tail shortening, but many key players and connected processes have been discovered. Importantly, it has been established that shortening of the poly(A) tail does not only occur as a part of rapid decay. Many RNAs exist at steady state at discrete poly(A) tail lengths, still being translated and functioning as a mature RNA.

Recent work in our lab, in which I was a co-author, revealed that the median length of poly(A) tails in *Caenorhabditis elegans* is 57 nt, much shorter than their initial long tail (Lima et al. 2017). Additionally, we found that these relatively short-tailed transcripts tend to be the most highly expressed, well-translated genes. Using published poly(A) tail length data in human, mouse and yeast, we discovered that this is a conserved relationship and not specific to worms. This finding seems to contradict the dogma that longer tails promote mRNA stability and translation. Instead, this suggests that efficiently translated genes accumulate at steady state at an optimal shortened length, which we term pruning.

1.4 Deadenylation and Decay

The expression level of any given gene is directly affected by transcriptional rate and decay rate. RNA turnover is an important post-transcriptional process that regulates gene expression. Deadenylation is thought to be the rate-limiting step in RNA decay, positioning the poly(A) tail as a key player in RNA degradation. Once deadenylation removes the poly(A) tail, the transcript can now undergo decapping (removal of the 5' cap) and 5'-3' degradation, or 3'-5' decay by the exosome (Mugridge, Coller, and Gross 2018). Additionally, considering recent discoveries that transcripts with pruned poly(A) tails accumulate at steady state, deadenylation also plays an important role in modifying RNA outside of the decay process.

The main cytoplasmic deadenylases are Pan2-Pan3 (Wolf and Passmore 2014) and the Ccr4-Caf1-CNOT1 complex (Collart 2016; Tucker et al. 2001). It was originally thought that these two complexes acted sequentially on RNA, with Pan2-Pan3 acting first to trim longer tails. However, new data shows that there is some substrate specificity for the deadenylases, and that the two enzymatic subunits of the Ccr4-Caf1-NOT1 complex act differently depending on the presence of PABPC (Webster et al. 2018; Sun et al. 2013). Ccr4 can deadenylate PABPC-bound RNAs and even remove PABPC from protecting the tail. Caf1 on the other hand is only able to deadenylate exposed adenosines that are not covered by PABPC. These new findings are covered in further detail in chapter 2 section 6.

Chapter 2

Tales of Detailed Poly(A) Tails

2.1 Highlights

- Poly(A) tails are dynamic additions to mRNA that play an important role in gene expression.
- Cutting-edge sequencing methods provide new insights into poly(A) tail size and composition on individual mRNAs.
- In somatic cells, highly expressed mRNAs accumulate with relatively short poly(A) tails.
- PABP facilitates the activities of deadenylases while also protecting the poly(A) tail in a manner linked to translation.

2.2 Abstract

Poly(A) tails are non-templated additions of adenosines at the 3' ends of most eukaryotic mRNAs. In the nucleus, these RNAs are co-transcriptionally cleaved at a poly(A) site and then polyadenylated before being exported to the cytoplasm. In the cytoplasm, poly(A) tails play pivotal roles in the translation and stability of the mRNA. One challenge in studying poly(A) tails is that they are difficult to sequence and accurately measure. However, recent advances in sequencing technology, computational algorithms, and other assays have enabled a more detailed look at poly(A) tail length

genome-wide throughout many developmental stages and organisms. With the help of these advances, our understanding of poly(A) tail length has evolved over the past 5 years with the recognition that highly expressed genes can have short poly(A) tails and the elucidation of the seemingly contradictory roles for poly(A)-binding protein (PABP) in facilitating both protection and deadenylation.

2.3 Poly(A) Tails Are a Dynamic and Important Modification of RNA

The first reports of a repetitive poly(A) stretch found on RNA came in the early 1970s (Darnell, Wall, and Tushinski 1971; S. Y. Lee, Mendecki, and Brawerman 1971; Edmonds, Vaughan, and Nakazato 1971). At the time, little was known about how or why RNAs had a poly(A) tail, but already there was speculation that it could be 'a signal related most probably to translation or mRNA formation and transport' (Darnell, Wall, and Tushinski 1971). Over the following years these predictions were validated and the dynamic regulation of poly(A) tails became more apparent. Numerous polymerases and deadenylases (see Glossary) have been identified that are important for modulating tail length (Goldstrohm and Wickens 2008; Schmidt and Norbury 2010; Laishram 2014). The poly(A) tail was beginning to reveal itself as a key player in mRNA post-transcriptional regulation, much more than merely an afterthought.

The creation of a poly(A) tail on newly synthesized RNAs involves the cooperation of many proteins and sequence elements. Almost all metazoan mRNAs contain a

polyadenylation signal (PAS), which has the canonical sequence AAUAAA or a close variant. This PAS as well as a downstream GU- or U-rich sequence guide the formation of the poly(A) tail by recruiting multiple protein complexes that are involved in initial 3'-end processing (Proudfoot 2011; Tian and Graber 2012). Other sequence elements can modulate the efficiency or exact location of polyadenylation. The position where polyadenylation takes place is not decided by the RNA polymerase terminating the pre-mRNA; instead, cleavage of the RNA occurs cotranscriptionally, within 10–30 nt downstream of the PAS, and poly(A) polymerase (PAP) then adds the poly(A) tail. Once 11–14 adenosines have been added, nuclear poly(A)-binding protein (PABPN) is able to bind to the growing poly(A) tail (Meyer, Urbanke, and Wahle 2002). This binding allows PAP to transition from distributive synthesis to processive synthesis, and PAP is then able to rapidly synthesize a full-length poly(A) tail, which is thought to be ~200–250 nt in length in metazoans (Brawerman 1981; Diana Sheiness and Darnell 1973). Several polyadenylation studies have contributed to our understanding of this transition and the key proteins involved, but it has been difficult to capture this initial moment of polyadenylation in the context of an intact whole organism (Sawicki, Jelinek, and Darnell 1977; Wahle 1995). Whether all transcripts are 'fully' polyadenylated to ~250 adenosines in all tissues is not entirely clear. For example, some genes have been found to include a poly(A)-limiting element (PLE) which acts to restrict the initial length of the poly(A) tail on the pre-mRNA to less than 20 nt (Gu, Das Gupta, and Schoenberg 1999). Nearly all mRNAs undergo cleavage and polyadenylation to some extent. The known metazoan exceptions are the replication-dependent histone protein mRNAs which terminate in a stem-loop structure. Some non-coding RNAs, such as several long non-coding RNAs

(lncRNAs) and a few small non-coding RNAs, have also been found to contain poly(A) tails (Cai, Hagedorn, and Cullen 2004).

Cleavage and polyadenylation are thought to be necessary for proper export of an mRNA from the nucleus into the cytoplasm. Once in the cytoplasm, the poly(A) tail is predominantly coated with the cytoplasmic poly(A)-binding protein (PABPC). Little is known about the transition between PABPN and PABPC on the poly(A) tail. Although both are shuttling proteins that can move between the cytoplasm and nucleus, how and when they complete their trade-off is poorly understood. The protein landscape of a newly synthesized transcript is very different than that of an actively translating mRNA in the cytoplasm, and remodeling of many proteins must take place; for PABPN and PABPC, the first round of translation seems to promote this transformation (Sato and Maquat 2009; Singh et al. 2015). The exchange between PABPN and PABPC could also be influenced by nuclear export or through passive remodeling in the cytoplasm. PABPN lacks any repeating footprint pattern, but PABPC repetitively coats the poly(A) tail with a footprint of ~20–30 adenosine nucleotides (Smith et al. 1997; Baer, Kornberg, and Supp 1983). PABPC facilitates a host of interactions important for translation and stability including binding to initiation factor eIF4G of the eIF4F cap-binding complex, as well as the translation termination factor eRF3 (Hoshino, Hosoda, et al. 1999; Hoshino, Imai, et al. 1999; Tarun et al. 1997; Jacobson and Favreau 1983; Wigington et al. 2014). Through these interactions, PABPC and the poly(A) tail are able to synergistically promote translation. Interestingly, PABPC has been implicated in mRNA protection and stability, as well as in the recruitment of deadenylases – a seemingly contradictory role that has only begun to be elucidated. The complex landscape of poly(A) tail-binding interactions is

further complicated by the fact that there are other factors such as La proteins that can bind to the poly(A) tail and to PABPC to potentially modulate translation (Maraia et al. 2017; Vinayak et al. 2018). The journey from initial biogenesis to decay contains many multifaceted relationships between the poly(A) tail and various protein factors.

In many instances, the poly(A) tail serves as a gatekeeper to protect the mRNA. Coming from the 3' end, an enzyme looking to degrade an mRNA must chew through the whole poly(A) tail before affecting the protein-coding region. At the 5' end, the mRNA is protected by its cap, and this cap needs to be specifically removed before degradation can initiate there. In addition, deadenylation generally occurs before decapping. This positions the poly(A) tail at a key threshold in mRNA decay. Given this role, the traditional view for many decades was that a longer tail meant more protection, greater stability, and overall a positive influence on translation (Goldstrohm and Wickens 2008; Weill et al. 2012; Jalkanen, Coleman, and Wilusz 2014). Although these roles in decay are a key part of the life of a poly(A) tail, recent research has shown that this is a simplified version of a fuller story.

2.4 Quantitative Measurements of Poly(A) Tail Length

Much of the early work studying poly(A) tail length relied on reporter genes or single-gene analysis, or was conducted in a specific cellular context such as embryogenesis. Although these studies provided insight into some of the factors controlling length, we were still missing the full story of poly (A) tail-length dynamics on a genome-wide scale in multiple cellular contexts. With the introduction of high-

throughput sequencing, many fields within RNA biology took great leaps forward in understanding transcriptome-level events. However, poly(A) sequencing was still not possible with standard RNA-seq protocols primarily due to the difficulty of reading homopolymeric sequences (Quail et al. 2012). Current sequencers are largely unable to accurately call multiple adenosines in a row. Despite these challenges, creative approaches were developed to circumvent this issue.

Chang et al. developed a solution termed TAIL-seq that included both an experimental procedure as well as software that uses a machine-learning model to accurately measure poly(A) tails (Figure 1A) (Chang et al. 2014). In brief, their experimental method involves depletion of ribosomal RNA (rRNA) and size selection against other small noncoding RNAs (tRNA, snRNA, snoRNA, and miRNA). Because noncoding RNAs make up the vast majority of the cellular RNA, these steps allow enrichment of the library for RNAs of interest. A 3' biotin adaptor is ligated to the RNA and RNase T1 is used to partially digest the RNA selectively after guanosine, leaving the poly(A) tail intact. The ligated RNAs are isolated with streptavidin beads and gel purified before a 5' adaptor is added. These libraries undergo paired-end sequencing on the Illumina platform (MiSeq or HiSeq instruments). The first read of 51 nt is used for genome mapping, while the second read of 231 nt is used to determine the 30-end sequence. The fluorescence intensity files are then reanalyzed using their Tailseeker software to more accurately assess the length of the poly(A) tail as well as determine any non-A residues present in the tail.

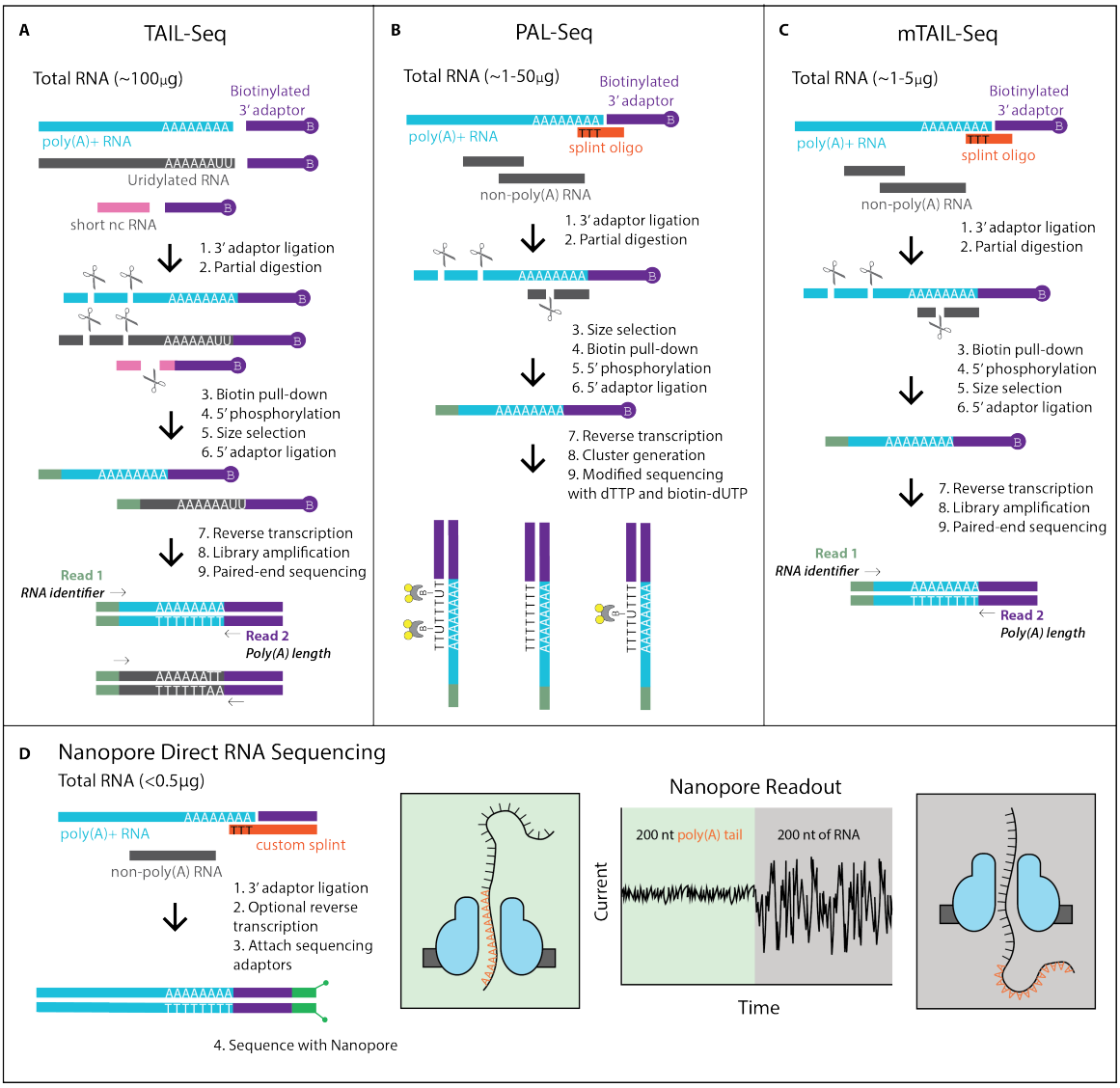


Figure 2.1: Comparison of Different Sequencing Methods for Reading Poly(A) Tails. (A) TAIL-seq is able to capture the 30 end of any RNA, and therefore gives a readout of both poly(A) tail length as well as other modifications such as uridylation through an innovative Tailseeker algorithm. (B) PAL-seq uses a splint oligonucleotide to preferentially capture polyadenylated RNAs, thus bypassing rRNA removal. Biotin-labeled dUTP marks each cluster in proportion to the length of the tail. (C) mTAIL- seq uses the splint oligonucleotide approach to reduce the amount of starting material needed, and uses the Tailseeker software to read poly(A) tail length. (D) Nanopore technology is a new way to sequence that can be used to directly sequence RNA or cDNA with minimal library preparation being needed. The nucleic acid travels through the nanopore at a constant rate; therefore, the dwell time of the poly(A) tail in the nanopore correlates with its length. Abbreviations: B, biotin; nc, non-coding.

Another method was developed by Subtelny et al. in the same year, called PAL-seq [poly(A) tail length sequencing] (Figure 1B) (Subtelny et al. 2014). To enrich for polyadenylated species and remove non-coding RNAs, this protocol relies on a ligation step with a DNA splint oligonucleotide that bridges the last part of the poly(A) tail with a sequence that matches the 3' biotin adaptor, therefore preferentially adding the adaptor only to RNAs that contain a poly(A) tail. Another key difference in this protocol is a modification performed during the sequencing, which is per- formed on the Genome Analyzer (originally made by Solexa and acquired by Illumina). During sequencing, a mixture of dTTP and biotin-conjugated dUTP is introduced, and each cluster is thus marked with an amount of biotin proportional to the length of the tail. Finally, fluorophore-tagged streptavidin is introduced, which reports on the amount of biotin included in each poly(A) tract, and therefore provides a way to determine tail length.

Both of these innovative technologies have expanded our knowledge of poly(A) tail length at a genome-wide scale. Each has its own pros and cons that will vary depending on the application. The original TAIL-seq protocol requires a large amount of starting material, on the order of 100 mg of RNA, but because the 3' ligation is not biased in any way, it is possible to capture other nucleotides present in the tail such as uridylation or guanylation events. On the other hand, PAL-seq requires much less starting RNA and bypasses costly (and sometimes ineffective) rRNA depletion steps owing to the splint ligation step, but this results in capturing tails that only have adenosines present at their most 3' end. The final quantification of poly(A) tail length is reliant on the random distribution of biotin-tagged uridine and may not be as accurate as determining the identity

of each nucleotide, as in the Tailseeker program. In addition, PAL-seq requires the user to modify the standard sequencing workflow of a Genome Analyzer.

Further advances have been made that combine key components of these two protocols. Two different laboratories devised methods that took advantage of the splint oligonucleotide approach during the ligation step and combined that with the powerful Tailseeker algorithm and ease of Illumina sequencing (Lim et al. 2016; Lima et al. 2017). Because these approaches were very similar to one another, they are both termed mTAIL-seq (Figure 1C).

Although still in early stages, an exciting new avenue for poly(A) tail length measurement is through the use of Nanopore technology (Figure 1D) (Garalde et al. 2018). Nanopore still faces the same difficulty as other sequencing platforms in accurately calling long stretches of homopolymeric sequence. However, because the nucleic acid being read is pulled through the pore at a constant rate, the amount of time that the poly(A) tail spends going through the pore correlates with its length. In addition, because Nanopore technology can directly sequence RNA (and cDNA with minimal library preparation), this method can bypass much of the time and cost associated with preparing a traditional library. Furthermore, this can completely bypass PCR amplification, thus eliminating any potential biases introduced at that step. Although this dwell-time readout does not report on whether any other nucleotides are present in the tail in addition to adenosines, this method can be used to proceed from experimental condition to actual tail length readout in a very short time. Another new option for inferring poly(A) tail length is TED-seq, which relies on precise size selection of libraries such that the tail length can be deduced by subtracting the distance from the mapped 5’ end of the read to the expected 3’

cleavage/polyadenylation site from the selected fragment size (Woo et al. 2018). Although direct tail measurement is not possible, this allows researchers to estimate tail length without complex sequencing methods.

2.5 Poly(A) Tails: Connections to Expression and Translation

For many years it was thought that, for most transcripts, a long poly(A) tail would protect the mRNA from decay and degradation. As with many areas of biology, there seems to be greater nuance involved in poly(A) tail length control than was previously appreciated. One of the first clues came with the discovery that, contrary to earlier thoughts that most tails would stay long, several transcripts had much shorter poly(A) tails than expected (Meijer et al. 2007; Choi and Hagedorn 2003). These studies did not map a specific mRNA with an exact tail length, but instead looked at pools of RNAs that were either not captured by earlier studies that used oligo(dT) beads, or by preferentially eluting RNA from oligo(dT) beads to produce short-tailed and long-tailed fractions. Some very stable transcripts (e.g., encoding beta-actin) were shown to have a short poly(A) tail of less than 30 nt. Overall many more transcripts were found to have short poly(A) tails than was previously expected. This opened up several new inquiries: do specific types of mRNAs have short poly(A) tails? How could a stable transcript have a short tail? Does this serve a biological purpose? Subsequent genome-wide studies began to address these questions. With the implementation of new sequencing methods, the landscape of poly(A) tail length became more clear. Not only were many short-tailed

species present, but overall median tail lengths were seen to be in the range of 50–100 adenosines for nearly all species studied: human, *Drosophila*, mouse, and Caenorhabditis elegans (Chang et al. 2014; Subtelny et al. 2014; Lima et al. 2017). Yeast was the only exception, with a median tail length of ~33 adenosines, but this was actually not strikingly different considering that yeast poly(A) tails were already known to be limited in their initial lengths to ~90 nt (Brown and Sachs 1998). It was becoming clear that a shorter tail is not always linked to destabilization and decay.

These median values reflected much shorter tails than expected, but there still was a broad range of sizes found in each organism. Investigating whether specific gene transcripts were enriched for short or long poly(A) tails was actually a more complex undertaking than one might expect. For one, the depth obtained in each sequencing study greatly influences the conclusions reached, especially for transcripts expressed at a low level. If only a handful of reads are captured for transcripts of a particular gene, this median may or may not reflect the entire pool of poly(A) lengths for that gene at a given time. To assess the poly(A) tail-length profile at a truly transcriptome-wide level, sequencing methods must capture even poorly expressed transcripts at their various tail lengths. Second, the use of words such as ‘short’ or ‘long’ to describe tails in the literature must be carefully considered because these words are limited to describing differences within the total length spectrum that an experiment was able to capture. What may be described as ‘long’ in one context could perhaps fall within the ‘short’ category according to another researcher, particularly given recent changes and advances in the field.

By analyzing recent poly(A) datasets relative to their own median lengths, and taking into account the greatest number of reads possible for each gene transcript, it was determined that short poly(A) tails are associated with highly expressed, well-translated genes (Lima et al. 2017). Longer poly(A) tails are associated with transcripts of lower abundance and poor translation (Figure 2). This was a surprising finding and counters the idea that a long tail is universally better for protection and translation. These shortened tails do not appear to be on their way to decay because they accumulate at steady-state at discrete lengths. Furthermore, there was an inverse correlation between poly(A) tail length and half-life of a transcript. This is consistent with early reports which found that short poly(A) tails were associated with the most stable mRNAs in vegetatively growing *Dictyostelium discoideum* cells (Palatnik et al. 1980; Palatnik, Storti, and Jacobson 1979).

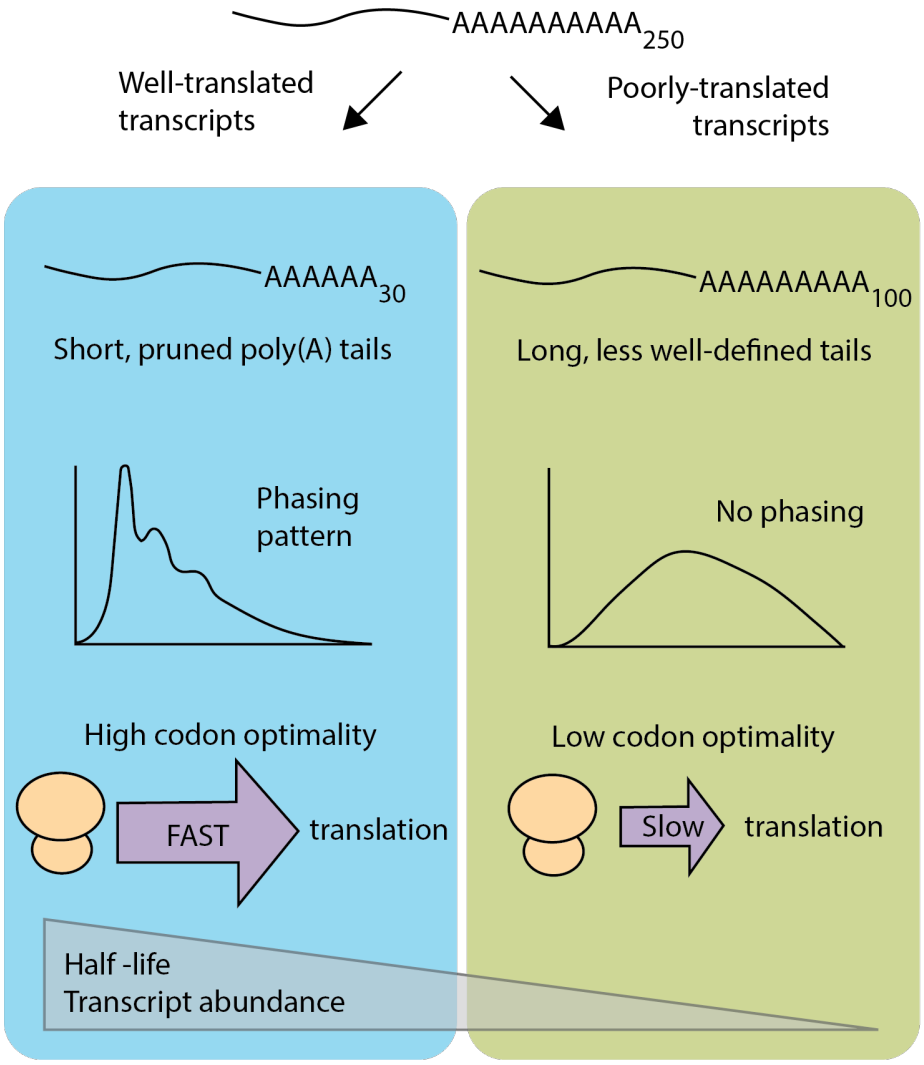


Figure 2.2: Short Poly(A) Tails Are Associated with Highly Expressed, Well-Translated Transcripts. These shortened tails occur at discrete lengths that have a phasing pattern matching the footprint of serial binding of cytoplasmic poly(A)-binding protein. Longer tails do not show this phasing and have less well-defined tails. In somatic cells, short-tailed transcripts tend to have high codon optimality and long half-lives.

Another interesting feature found in some poly(A) datasets is that there is a phasing pattern to poly(A) tail length wherein there is a greater enrichment for lengths that would be expected to occur with serial binding of PABPC (Lima et al. 2017; Yi et al. 2018). This enrichment for tails at these footprint lengths suggests that unprotected

adenosines are likely to be quickly removed. Intriguingly, this phasing distribution is only found on mRNA poly(A) tails, and not on other polyadenylated but untranslated species such as lncRNAs [33]. Furthermore, phasing is seen most clearly on highly expressed and well-translated transcripts. This suggests that the shortening of poly(A) tails to discrete lengths, termed pruning, is a process linked to translational activity.

Thus far, the available sequencing datasets have measured poly(A) tail lengths only at steady-state. Pulse-chase time-course experiments will provide a new level of information regarding the dynamics of poly(A) tail pruning. Deadenylases may remove adenosine nucleotides from all transcripts, regardless of translation status or half-life, but are blocked from fully deadenylating particular transcripts because efficient translation has resulted in closed-loop formation that protects the transcript. Tails might be deadenylated at the same rate until the deadenylase reaches the critical point of pruning, stopping at the most proximal PABP. However, another possibility is that a sensor recognizes high translation and recruits pruning factors. In this scenario, the tails on well-translated transcripts would be shortened faster, thus resulting in short tails primarily being detected at steady-state. Future time-course experiments would address these kinetic questions.

These recent studies linking short poly(A) tails with high expression and translation have all been conducted outside the embryonic context, where it has long been known that a very different polyadenylation landscape is at play. During oocyte maturation and early embryonic development, selective cytoplasmic polyadenylation lengthens the tails of particular mRNAs, and this actually increases translation, thereby reactivating silenced transcripts (Subtelny et al. 2014; Lim et al. 2016; M D Sheets et al.

1994; Eichhorn et al. 2016; Bazzini et al. 2016). Cytoplasmic polyadenylation has also been found to activate some neuronal transcripts (Udagawa et al. 2012). These seem to be scenarios that are specific for a particular cellular or developmental context.

On top of length as a way to modulate the tail, the composition of poly(A) tails can also be variable. Uridylation typically occurs on very short tails (less than 25 nt) and is a way to mark a transcript for decay. PABPC binding to the tail inhibits uridylation, and miRNA targeting induces it (Rissland, Mikulasova, and Norbury 2007; Rissland and Norbury 2009; Lim et al. 2014; Morgan et al. 2017). Guanylation is found selectively on longer poly(A) tails and can stall deadenylation by the Ccr4–Not complex, thus delaying decay (Chang et al. 2014; Lim et al. 2018). Cytosine addition has also been seen on poly(A) tails (although less frequently), but its biological function has not yet been characterized (Chang et al. 2014; Lim et al. 2018).

2.6 Dual Roles of PABP

The poly(A) tail facilitates numerous interactions between mRNA and proteins. Many of these interactions occur through PABPs which bind to the poly(A) tail with high affinity (Görlach, Burd, and Dreyfuss 1994; Sachs, Davis, and Kornberg 1987). A definitive role for PABPC has remained obscure owing to the seemingly conflicting roles it plays in gene expression. On the one hand, it is able to promote deadenylation by direct binding to the deadenylase complexes Pan2–Pan3 and Ccr4–Not (Mangus et al. 2004; Funakoshi et al. 2007; Uchida, Hoshino, and Katada 2004; Webster et al. 2018). However, it is most commonly recognized as a protein that directly binds to and protects the poly(A)

tail from degradation (Bernstein, Peltz, and Ross 1989; Wang et al. 1999). Moreover, through its interaction with 5' cap-binding factors as well as translation termination factor eRF3, PABPC is thought to promote translation. Nevertheless, as a protein that can recruit deadenylases, it is also involved in the decay and downregulation of gene transcripts. These pleiotropic effects have necessitated careful inquiry to delineate the parameters of each of these roles.

One major breakthrough came in defining the deadenylase activity of the Ccr4 and Caf1 enzymes, components of the larger Ccr4–Not complex, in the presence of PABPC. Two different groups studying yeast and human deadenylases simultaneously found that binding by PABPC (Pab1 in yeast) does not block Ccr4 activity and, in fact, Ccr4 can release PABPC from the tail and continue deadenylation (Yi et al. 2018; Webster et al. 2018). By contrast, Caf1 can only remove adenosines outside the protective footprint of PABPC (Figure 3A). PABPC still retains a central role in recruitment, however, because depletion of Pab1 resulted in much slower deadenylation rates by the Ccr4–Not complex overall (Webster et al. 2018). These distinct functional roles suggest that the amount of PABPC on the tail could result in varying tail lengths.

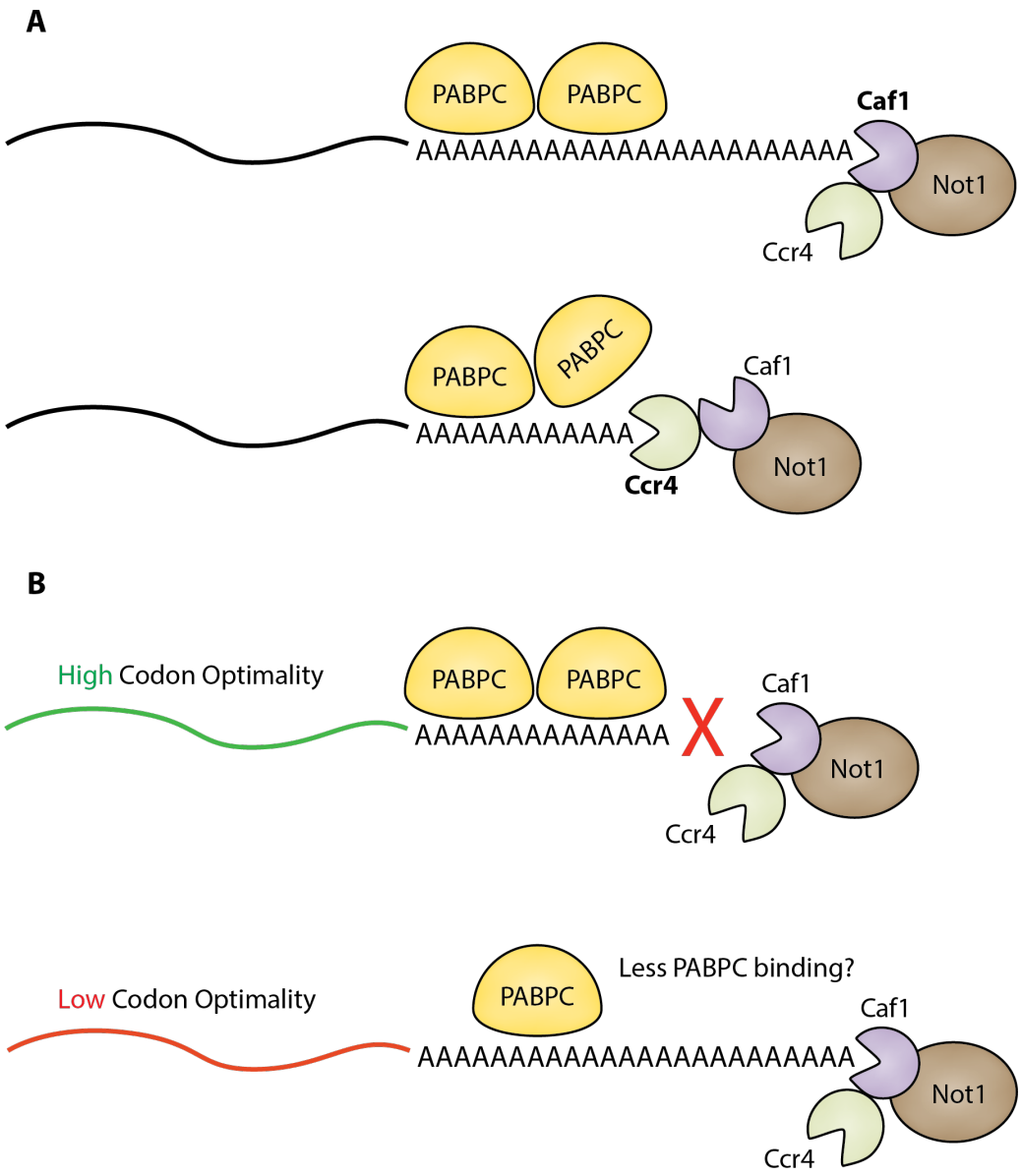


Figure 2.3: Differential Activities of the Ccr4 and Caf1 Deadenylases. (A) Caf1 is able to deadenylate portions of the poly(A) tail that are not tightly bound by cytoplasmic poly(A)-binding protein (PABPC) but halts once it reaches PABPC. On the other hand, Ccr4 is able to displace PABPC from the poly(A) tail and continue deadenylation. Ccr4 is also able to act on poly(A) stretches that are not bound by PABPC. (B) Caf1 preferentially accelerates deadenylation of low codon-optimality transcripts. Ccr4 is able to act on both substrates, but high codon-optimality transcripts seem to rely solely on Ccr4- mediated deadenylation. This may be due to differences in PABPC occupancy on the tails of transcripts with higher (more PABPC) and lower (less PABPC) levels of optimal codons.

Interestingly, deadenylation patterns of Ccr4 and Caf1 were also shown to be influenced by codon optimality, a proxy for translation status (Webster et al. 2018; Hanson and Coller 2017). Using a reporter system, transcripts with lower codon optimality were deadenylated more rapidly than counterparts with higher optimality. In addition, depletion of Caf1 preferentially influenced the deadenylation rate of the low codon-optimality reporter, suggesting that its poly(A) tail might have less PABPC bound, allowing Caf1 to be more active on poorly compared to highly translated transcripts (Figure 3B).

These studies again situate translation and poly(A) tail length in close contact with one another, with PABPC as a main player modulating this relationship. It remains unclear how some well-translated transcripts might end up with more PABPC coating their tail than others. The mechanistic work carried out to characterize PABPC has revealed non-identical roles for the four RNA-binding domains, and it has been suggested that the arrangement of PABPC on the tail may change in response to deadenylation (Webster et al. 2018). Whether this shifting of PABPC influences translation or decay is not understood. Further investigation into the multipronged roles of PABPC in gene expression will provide answers in this area.

2.7 Concluding Remarks

Given the importance of fine-tuned gene expression across all domains of biology, it may come as no surprise that the poly(A) tail is not a passive bystander in this process. From its initial biogenesis to dynamic control in the cytoplasm and ultimately decay, the length of the poly(A) tail continues to show itself as an important player in processes as

central as translation and mRNA stability. Elucidating the cause and effect of these relationships will be difficult but of ultimate importance in understanding these complex interactions (see Outstanding Questions).

As these types of central questions are addressed, the role of the poly(A) tail in more specific contexts, such as mediating miRNA target regulation or consequential alterations in disease, may become clearer. With the advent of new technology able to read poly(A) tail size more accurately than ever before, a more coherent and, at times, surprising picture is developing to explain how an addition present on almost all mRNAs is dynamically controlled to exert a vast influence on gene expression.

2.8 Outstanding Questions

- What mechanisms control how different genes accumulate transcripts with distinct poly(A) tail lengths at steady-state?
- How is translation coupled to poly(A) tail length? Is translation itself the driver behind pruning tails from their initial long length to a shorter steady-state size? Or are other processes involved first, and then translation takes advantage of this new piece of information on the mRNA?
- What is the rate of poly(A) tail pruning? Are highly expressed transcripts shortened more rapidly, or does their stability allow them to accumulate to a greater extent with pruned tails?

- What is the relationship between tail length and other RNA-binding proteins? PABPC has already been implicated, but other factors may be directly involved, such as eukaryotic initiation factors, argonaute proteins, quality control factors, and more.

2.9 Glossary

Biotin: a molecule frequently used to tag or label a nucleic acid or protein of interest. Biotin binds to streptavidin with very high affinity, and therefore this interaction can be used to pull out a molecule of interest, separating it from other molecules that do not have this biotin tag.

Cap: the 5- methyl guanosine cap is added to RNA as part of initial processing into a mature mRNA. It protects the transcript from degradation on the 50 side. The cap is added to the growing RNA cotranscriptionally and is important for export from the nucleus into the cytoplasm, mRNA stability, and translation initiation. Deadenylases: enzymes that catalyze the removal of adenosines. There are several different known deadenylases that can act on the poly(A) tail for both pruning and decay.

Distributive: referring to the activity of an enzyme, a distributive process is one in which the enzyme dissociates from its substrate frequently after a catalytic event. This is in contrast to a processive enzyme, defined below.

Homopolymeric: a repeating sequence of the same nucleotide. The poly(A) tail is a homopolymeric sequence because it contains all adenosine nucleotides. Homopolymeric sequences are notoriously difficult to sequence with current technology.

Nanopore: a type of sequencing that enables minimal library preparation and long sequencing reads. Protein nanopores are embedded on a membrane and an ionic current is passed through the nanopore. When a strand of nucleic acid goes through the pore, the current changes depending on which base is going through the pore, therefore giving a readout of G, C, T or A.

Poly(A) polymerase (PAP): canonical PAP (also called PAPa) is the enzyme responsible for adding the poly(A) tail to the newly made RNA. Other nuclear and cytoplasmic polymerases have also been discovered which have more specialized functions that differ from the general function of canonical PAP.

Polyadenylation signal (PAS): a sequence element directing where cleavage and polyadenylation should occur for most RNA polymerase II transcripts. Because cleavage then occurs downstream of this element, it remains in the mature mRNA. The canonical sequence AAUAAA is found in most mammalian transcripts, with some minor variability. There can be greater variability in this signal sequence in other eukaryotes.

Processive: a processive enzyme can catalyze multiple reactions after a single substrate–enzyme encounter. This results in continual activity because, once an enzyme becomes associated with the correct substrate, it can continue its activity until complete.

2.10 Acknowledgments

We thank Jens Lykke-Andersen and members of the laboratory of A.E.P. for suggestions and critical reading of the manuscript. Support for this work was from the UCSD Cellular and Molecular Genetics Training Program through an institutional grant

from the National Institute of General Medicine (T32 GM007240) and a National Science Foundation Graduate Research Fellowship (DGE-1650112) to A.L.N., and by a grant from the National Institutes of Health (R35 GM127012) to A.E.P.

Chapter 2, in full, is a reprint of material as it occurs in *Trends in Cell Biology,* "Tales of Detailed Poly(A) Tails," Nicholson, A.L., Pasquinelli, A. E., 2019. I was the primary author.

Chapter 3

Distinct Transcripts and Isoforms Favor the Nuclear and Cytoplasmic Poly(A) Binding Proteins (PABPs)

3.1 Introduction

The vast majority of eukaryotic RNAs have a 3' poly(A) tail added to them co-transcriptionally. The poly(A) tail serves a wide range of functions- promoting nuclear export (Fuke and Ohno 2008), protecting the RNA from exonucleases and degradation (Jalkanen, Coleman, and Wilusz 2014; Goldstrohm and Wickens 2008), and enhancing translation (Weill et al. 2012; Jalkanen, Coleman, and Wilusz 2014; Sachs and Varani 2000). These roles are largely mediated through RNA-binding proteins. The nuclear and cytoplasmic poly(A) binding proteins (PABPs), PABPN and PABPC, are the main poly(A) binding proteins and have distinct roles when they are bound to the tail (Mangus, Evans, and Jacobson 2003).

PABPN1 (hereafter referred to as PABPN) is involved in the initial creation of the poly(A) tail in the nucleus. Once cleavage occurs near the poly(A) site (PAS) of a nascent RNA, poly(A) polymerase (PAP) begins synthesizing the tail (Proudfoot 2011; Tian and Graber 2012). After PAP has added 11-14 adenosines, PABPN is then able to bind to the growing tail, which causes PAP to switch to processive synthesis, rapidly completing creation of the tail (Bienroth, Keller, and Wahle 1993; Wahle 1991). Cleavage and polyadenylation can occur either before or after splicing is fully complete (Muniz, Davidson, and West 2015; Rigo and Martinson 2009). PABPN-oligo(A) binding equilibrium measurements determined that the minimum tail length that PABPN can bind is 11 adenosines and the site size covered by PABPN is 11-15 adenosines (Meyer, Urbanke, and Wahle 2002; Nemeth et al. 1995). Polyadenylation proceeds until the tail has ~200-250 adenosines (Diana Sheiness and Darnell 1973; Brawerman 1981; Michael D Sheets and Wickens 1989). The exact mechanism of tail length control is not clear, but PABPN forms a 21 nm spherical particle with poly(A) that is thought to serve as a molecular ruler (Keller et al. 2000). Although PABPN can contiguously bind along the poly(A) tail, PABPN-PABPN interactions show weak cooperativity (Keller et al. 2000; Meyer, Urbanke, and Wahle 2002). When the RNA is fully mature, it can be exported out to the cytoplasm.

Once in the cytoplasm, PABPC is the predominant protein coating poly(A) tails. The transition from having PABPN on the tail to PABPC is not well understood. Reporter studies have suggested that translation may facilitate this transition (Sato and Maquat 2009). PABPC facilitates numerous protein interactions that promote translation and stability such as binding to the translation initiation factor eIF4G and the translation

termination factor eRF3 (Uchida et al. 2002; Jacobson and Favreau 1983; Hoshino, Imai, et al. 1999; Hoshino, Hosoda, et al. 1999; Tarun et al. 1997; Wigington et al. 2014). PABPC can sequentially bind the poly(A) tail and has a footprint of ~20-30 adenosine nucleotides (Smith et al. 1997; Baer, Kornberg, and Supp 1983), needing a minimum of 11 or 12 adenosines in order to bind (Deo et al. 1999).

Humans have one nuclear PABP, PABPN1, and five cytoplasmic PABP proteins (Gorgoni and Gray 2004; Guzeloglu-Kayisli et al. 2008; Mangus, Evans, and Jacobson 2003). The major PABPC present in somatic tissues is PABPC1. The four remaining proteins are testis-specific PABP (tPABP, also termed PABPC3), embryonic PABP (ePAB), X-linked PABP (PABPC5), and a PABP discovered as an inducible protein in stimulated T-cells (PABPC4, also known as iPABP) (Blanco et al. 2001; Feral, Guellaën, and Pawlak 2001; Guzeloglu-Kayisli et al. 2008; Kini, Kong, and Liebhaber 2014; Kleene et al. 1998; Yang, Duckett, and Lindsten 1995). Since PABPC1 is the most abundant cytoplasmic PABP, we focused our studies on PABPC1 and it is referred to throughout the rest of the text as PABPC. Estimates of cellular abundance of PABPN are 2-3.4 x 10^6 molecules per cell and for PABPC, 8 x 10^6 molecules (Krause et al. 1994; Görlach, Burd, and Dreyfuss 1994). Affinity for the poly(A) tail as measured through dissociation constants are similar: 2nM for PABPN (Wahle et al. 1993) and 0.69 nM-7 nM range for PABPC (Sawazaki et al. 2018; Görlach, Burd, and Dreyfuss 1994; Kühn and Pieler 1996). Despite their shared affinity for the poly(A) tail, PABPN and PABPC are structurally and functionally distinct from one other. In addition to their predominant localization in the nucleus and cytoplasm, PABPN and PABPC are shuttling proteins, existing in lower abundance in the opposite compartment as well (Afonina, Stauber, and Pavlakis 1998;

Calado and Carmo-Fonseca 2000; Woods et al. 2002; Hosoda, Lejeune, and Maquat 2006). The presence of one PABP on the poly(A) tail is not known to exclude the other and therefore PABPN and PABPC may exist together on a single RNA (Hosoda, Lejeune, and Maquat 2006).

Although poly(A) tails can reach a full-length size of 200 adenosines, their length at steady state is much shorter. Early bulk poly(A) studies revealed shorter cytoplasmic lengths (Diana Sheiness and Darnell 1973; Greenberg and Perry 1972). With the advent of high-throughput sequencing, a few genome-wide studies have confirmed this in human cells (Subtelny et al. 2014; Chang et al. 2014; Workman et al. 2019; Eisen et al. 2020; Nicholson and Pasquinelli 2019). However, sequencing of poly(A) tails when they are first made has been limited, leaving the question of whether all tails on all transcripts follow these model patterns.

In the present study, we provide in depth analysis of the transcripts associated with PABPN and PABPC at steady state in human cells. We find that distinct sets of messenger RNA (mRNA) and non-coding RNA (ncRNA) predominantly bind PABPN or PABPC. Using Ribo-STAMP (Brannan et al. 2021), we present evidence that when transcripts are still associated with PABPN they make contact with the ribosome but less frequently than when they are bound by PABPC. Additionally, the RNA binding profile of PABPC depends on active translation, as we observed reduced association of new transcripts with PABPC in the presence of translation inhibitors. Through Nanopore direct RNA sequencing, we show that distinct isoforms of RNA, differentiated by poly(A) tail length and intron presence, are bound to PABPN or PABPC. Overall, our results capture the broad landscape of RNAs that associate with PABPN and PABPC, providing new insights

and confirming some aspects of the model for PABPN binding nascent polyadenylated transcripts and the accumulation of well-translated mRNAs containing pruned poly(A) tails with PABPC in the cytoplasm.

3.2 Results

3.2.1 PABPN and PABPC have distinct RNA binding profiles in human cells

To identify the RNAs associated with PABPN and PABPC (PABPC1) at steady state, we performed RNA Immunoprecipitation (RIP) assays using a single total cell extract from human cells. To preserve native RNA-protein interactions, HEK293T cells were crosslinked with formaldehyde and endogenous PABPN and PABPC were immunoprecipitated along with their associated RNAs in three independent biological replicates (Supplemental Figures 3.6A – 3.6C). Formaldehyde crosslinking was chosen so that reverse crosslinking could be performed and the intact RNA could be extracted and used for RNA sequencing (RNA-seq).

Since it is known that some splicing is completed after polyadenylation, intronic reads were expected in the sequencing results (Muniz, Davidson, and West 2015). As typical RNA sequencing (RNA-seq) analysis pipelines only quantify exonic reads, we created a pipeline that allowed quantification of reads coming from both intronic and exonic fragments. Particularly in intronic regions, current human annotation files have many overlapping genes. This presented a problem for properly annotating reads in these regions (Zerbino, Frankish, and Flicek 2020; S. Lee et al. 2020). Arbitrarily changing

typical pipeline parameters to select one gene over another can unfairly bias towards exonic over intronic sections, or long genes over short genes. Additionally, there are genes present in the NCBI annotation that are not present in the gencode annotation (Frankish et al. 2019; O'Leary et al. 2016). After discovering that our sequencing contained reads mapping to the genomic areas of these missing genes, we extracted the 672 missing genes from the NCBI annotation and manually added them to the gencode annotation. To address the issue of overlapping genes, we removed from the annotation file any regions that overlapped with two or more genes (Supp. Figure 3.6D). Although this functionally reduces the amount of annotated genomic space, it prevents the possibility of false positives due to mis-annotation at these regions. Furthermore, transcripts that are tagged as "readthrough_transcripts" by gencode were removed as these do not represent the predominantly expressed transcript. Their presence in the annotation file causes large regions to appear overlapping when, in fact, this is a rarer event that should not preclude the ability to annotate the two individual genes comprising this annotated readthrough event (Supp. Figure 3.6E).

After visually inspecting the resulting reads (by converting to BigWig format and viewing on the UCSC genome browser), we found that some genes showed evidence of improper termination and continuous downstream transcription, reminiscent of DoGs characterized previously (Vilborg et al. 2015; Rosa-Mercado et al. 2021; Schreiner et al. 2019). For example, the protein-coding gene TNFRSF13C was marked as enriched in our PABPN IP dataset, but showed reads spanning the upstream intergenic region reaching the neighboring gene CENPM (Supp. Figure 3.6F). The reads covering TNFRSF13C likely originated from transcripts that failed to be properly terminated from the CENPM gene.

These downstream transcripts likely do not represent functional coding gene products. We filtered these genes from all analysis by quantifying the depth and breadth of reads present 2000 base pairs (bp) upstream of each annotated gene. Any genes that had coverage across 70% or more of this upstream region and an upstream expression level equivalent to 20% or more of the downstream gene were removed from results tables. This consisted of 413 genes in the input condition, 507 genes in the PABPN IP, and 124 genes in the PABPC IP (Table S1), indicating that this type of transcription without proper termination may be more prevalent and not limited to stress conditions.

After filtering, over 15,000 genes were detected with at least a TPM (Transcript Per kilobase Million) value of 1 in all conditions. To identify transcripts that were enriched or depleted from PABPN or PABPC at steady state, we used DESeq2 to compare the abundance of transcripts in IP conditions compared to input (Love, Huber, and Anders 2014). Generating a principal component analysis (PCA) plot showed that our three replicates were highly reproducible and IP conditions resulted in substantial differences, clustering separately from both input and the other IP condition (Supp. Figure 3.6G). Many protein-coding genes (PCGs) and non-coding genes were enriched and depleted at steady state in both IPs (Figures 3.1A-3.1D), indicating that there may be unique characteristics of an RNA that cause it to be preferentially associated with a particular PABP. Using cut-offs of $\log_2$FoldChange > 0.5, padj <= 0.01, and baseMean > 50, there were 2716 genes detected as significantly enriched with PABPN, and 5703 genes enriched with PABPC compared to input (Supp. Figure 3.6H, Supp. Table 2). 1113 of these genes were enriched in both conditions (Supp. Figure 3.6H, Supp Table 3). Using a cut-off of $\log_2$FoldChange < -0.5, padj <= 0.01, and baseMean > 50 for depletion, 2260 genes were detected as

significantly depleted from PABPN and 4802 genes were significantly depleted from PABPC (Supp Table 2), with 412 of those genes being depleted from both conditions (Supp. Figure 3.6I; Supp Table 3).

3.2.2 Genes lacking a terminal poly(A) tail are depleted from PABPN and PABPC RIPs

Under our stringent RIP conditions, genes that are known to lack a poly(A) tail were depleted from the PABPN and PABPC RIP compared to input sequencing datasets. Replication-dependent histone genes, which terminate in a unique stem-loop structure, were depleted from both IPs but robustly detected in the input sample (Figures 3.1A and 3.1B). Of the 60 PCGs that are significantly depleted from both RIPs by at least two-fold, 48 are histone genes, 2 are mitochondrial-encoded genes, and the remaining 10 are either lowly expressed or not as robustly depleted as the histone genes (Supp Table 3). Of the 24 lncRNAs that are two-fold depleted, 19 are from classes that would not be expected to contain a poly(A) tail, such as genes transcribed by RNA Polymerase III (Pol III). The nuclear polyadenylation complex is known to be associated with RNA polymerase II (Hirose and Manley 2000, Bentley 2005) and therefore Pol III-transcribed genes are typically not polyadenylated. Many Pol III-transcribed genes were depleted from both RIPs but detected in the input condition, such as RNAse P RNA and Y RNA (Supp Table 3). The remaining 5 lncRNAs that were robustly depleted from both RIPs are novel transcripts and their depletion suggests that they may harbor non-polyadenylated 3' ends.

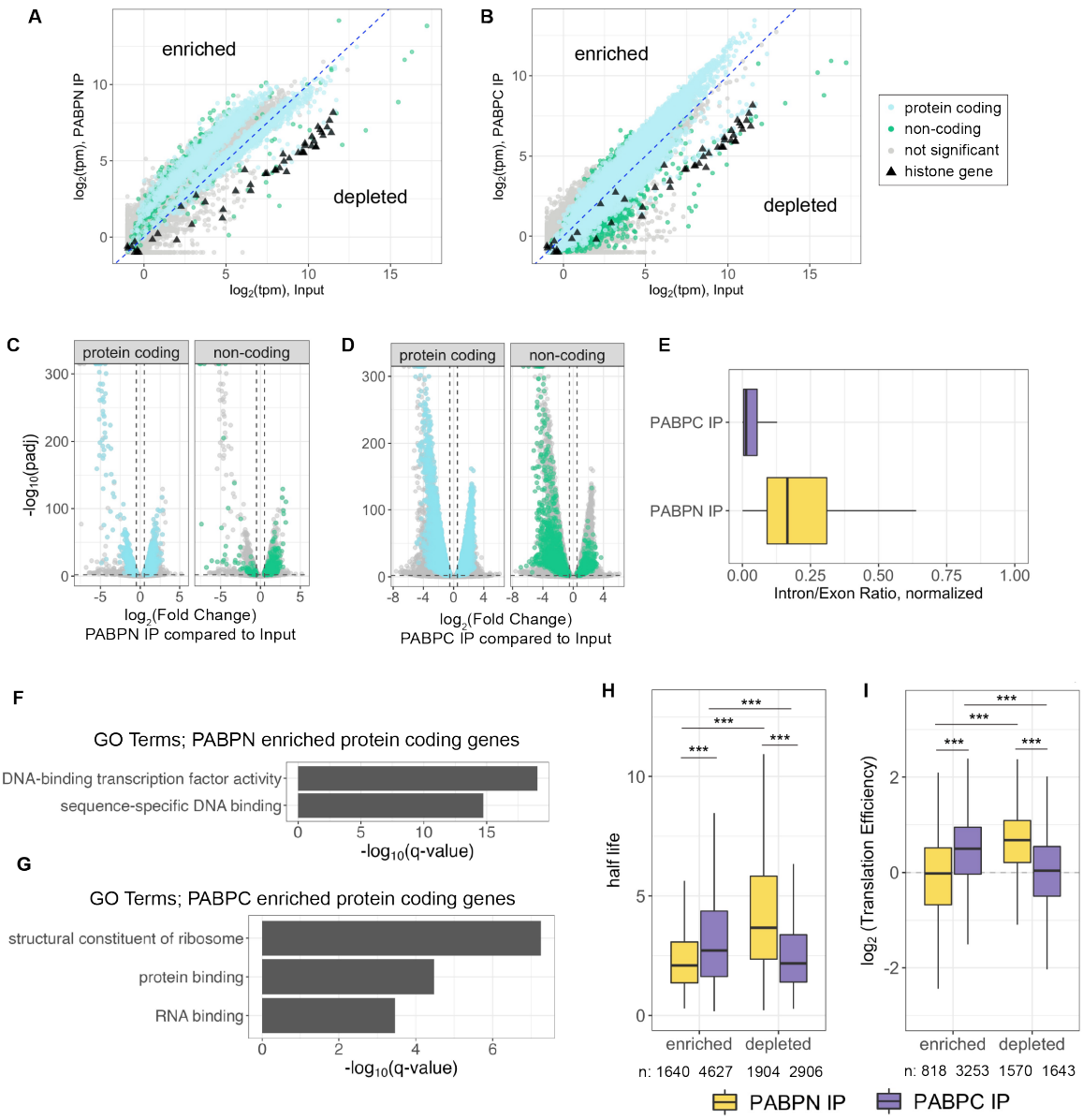
3.2.3 PABPN binds to RNAs before splicing is complete

For some introns, PABPN binding to the poly(A) tail promotes splicing. Hence, some transcripts undergo polyadenylation before intron removal has been completed (Muniz, Davidson, and West 2015; Rigo and Martinson 2009). This suggests that PABPN RIP sequencing datasets could contain pre-mRNAs. To investigate if either PABP binds to pre-mRNAs, we looked at the presence of intronic sequences in the two IPs. For each gene, intronic and exonic reads were normalized to the length of the intronic or exonic region, respectively. Calculating the ratio between the two gives a value of zero if a gene is completely spliced and a ratio of 1 if normalized exonic and intronic reads are equal, thus indicating a completely unspliced gene. As predicted, transcripts associated with PABPN had a higher intron/exon ratio than when they were associated with PABPC (median PABPN ratio: 0.17, median PABPC ratio: 0.01) indicating that PABPN binds to pre-mRNA before splicing is complete (Figure 3.1E).

Previous reports suggest that polyadenylation preceding completion of splicing may only occur for a small subset of genes. For example, upon siRNA depletion of PABPN followed by RNA-seq, only 226 genes were seen to be >2-fold misregulated and later studies confirmed that some of these genes had difficulty splicing in the absence of PABPN (Beaulieu et al. 2012; Muniz, Davidson, and West 2015). To examine if the intron ratios shown in Figure 3.1E were coming from a small subset of genes, we asked how many genes had intron representation of at least 1 TPM, only considering genes that had a genomically encoded intron. Surprisingly, in PABPN RIPs, 80% of genes are represented by intronic reads. In contrast, PABPC RIPs only contain intronic reads for 12% of genes.

For the 20% of genes that were not represented by an intronic read in the PABPN RIP dataset, their median total TPM was only 3, suggesting that the absence of intron representation for these genes may be due to their low abundance. While these results show that intron-containing transcripts for the majority of genes associate with PABPN, it is unclear if binding is dependent on the presence of a poly(A) tail or is a co-transcriptional event where PABPN is in the vicinity of the transcribing polymerase complex.

Figure 3.1: RNA Immunoprecipitation of PABPN and PABPC reveals distinct enrichment profiles. (A and B) Transcripts Per Kilobase Million (TPM) of Input compared to TPM of PABPN IP (A) or PABPC IP (B) from Illumina RNA-Seq experiments. ▲ indicates a canonical histone gene as annotated in HistoneDB 2.0 (Draizen). Blue dashed line is an overlaid 1:1 line. Significant genes enriched or depleted were determined by comparison to input samples using DESeq2 (Love et al., 2014), using cut-offs of $\log_2$FoldChange 0.5/-0.5, padj <= 0.01, and baseMean > 50. Non-significant genes are colored in grey. (C and D) Volcano plots showing enrichment and depletion of protein-coding and non-coding genes in PABPN IP vs Input (C) or PABPC IP vs Input (D). Grey dashed lines indicate significance cut-offs of padj <= 0.01 and $\log_2$FoldChange > 0.5 or < -0.5. (E) Intron presence in PABPN and PABPC IPs, analyzed by normalizing each exon or intron to their respective length and then comparing the ratio of intron reads/exon reads for each gene. A value of 1 would indicate completely unspliced, and a value of 0 indicates fully spliced. Genes used for calculation had at least a TPM of 1 in both IP conditions. (F) Gene ontology (GO) molecular function enrichment analysis of protein coding genes significantly enriched in PABPN IP compared to Input (n = 1893) using PANTHER. Significance cut-offs used are baseMean > 50, padj <= 0.01 and log2FC > 0.5. Reference list used was the protein coding genes that were detected with a baseMean greater than 50 overall (n = 12608). (G) Gene ontology (GO) molecular function enrichment analysis of protein coding genes significantly enriched in PABPC IP compared to Input (n = 5113) using PANTHER. Significance cut-offs used are baseMean > 50, padj <= 0.01 and log2FC > 0.5. Reference list used was the protein coding genes that were detected with a baseMean greater than 50 overall (n = 12398). (H) Protein coding genes determined to be enriched or depleted in PABPN IP or PABPC IP were grouped and compared to published half-life values (Lugowski, 2018). Significant differences in the cumulative distributions attributable to enrichment or depletion with PABPN or PABPC are indicated: ***p < 0.001; two tailed Kolmogorov-Smirnov test. Number of genes in each boxplot are displayed at the base of the graph. (I) Groupings were compared to published translation efficiency data as determined by ribosome profiling (Subtelny, 2014). Otherwise, this panel is the same as in (H).


A
enriched
depleted
log2(tpm), PABPN IP
log2(tpm), Input
B
enriched
depleted
log2(tpm), PABPC IP
log2(tpm), Input
protein coding
non-coding
not significant
histone gene
C
protein coding
non-coding
-log10(padj)
log2(Fold Change)
PABPN IP compared to Input
D
protein coding
non-coding
log2(Fold Change)
PABPC IP compared to Input
E
PABPC IP
PABPN IP
Intron/Exon Ratio, normalized
F
GO Terms; PABPN enriched protein coding genes
DNA-binding transcription factor activity
sequence-specific DNA binding
-log10(q-value)
G
GO Terms; PABPC enriched protein coding genes
structural constituent of ribosome
protein binding
RNA binding
-log10(q-value)
H
half life
enriched
depleted
n: 1640 4627 1904 2906
I
log2 (Translation Efficiency)
enriched
depleted
n: 818 3253 1570 1643
PABPN IP
PABPC IP

3.2.4 Transcripts of genes enriched in PABPC RIPs tend to be long-lived and well-translated

Given the extent of non-overlapping RNA binding profiles (Supp Figures 3.6H and 3.6I), we next asked if there were distinguishing characteristics of the PCGs that were enriched and depleted with each PABP. To look at the molecular functions of the PCGs enriched in each IP, we used gene ontology (GO) enrichment analysis, looking at statistical overrepresentation through PantherDB (Mi et al. 2021). PCGs enriched with PABPN are involved in DNA-binding, whereas those enriched with PABPC show strong enrichment for ribosomal proteins, protein binding, and RNA binding proteins (Figures 3.1F and 3.1G). Using published half-life values (Lugowski, Nicholson, and Rissland 2018), we find that genes enriched with PABPC tend to be more stable than those enriched with PABPN, whereas genes depleted from PABPN had longer half-lives than those depleted from PABPC (Figure 3.1H, Supp. Table 4). Using published translation efficiency (TE) data from ribosome profiling (Subtelny et al. 2014), we find that PCGs enriched with PABPC and depleted from PABPN have a high TE value, indicating that they are well-translated (Figure 3.1I; Supp. Table 4). These correlations align well with what is known regarding PABPC facilitating translation and binding to multiple translation factors (Osawa et al. 2012; Sachs and Varani 2000; Borman, Michel, and Kean 2000). By comparing to coding sequence (CDS) length, we find a slight positive correlation between length and degree of enrichment with PABPN PCGs (spearman 0.18, $p<2e-16$) and a substantial negative correlation between length and degree of enrichment with PABPC PCGs (spearman -0.59, $p <2e-16$) (Supp. Table 4). Genes that have been identified as well-translated and highly

expressed, such as housekeeping genes, are known to be more compact and have a shorter CDS length than other genes, consistent with our correlations for highly enriched PABPC PCGs (Eisenberg and Levanon 2003). Altogether, these analyses indicate that transcripts of genes associated with PABPN and PABPC differ in their average stability, translation efficiency, and coding potential.

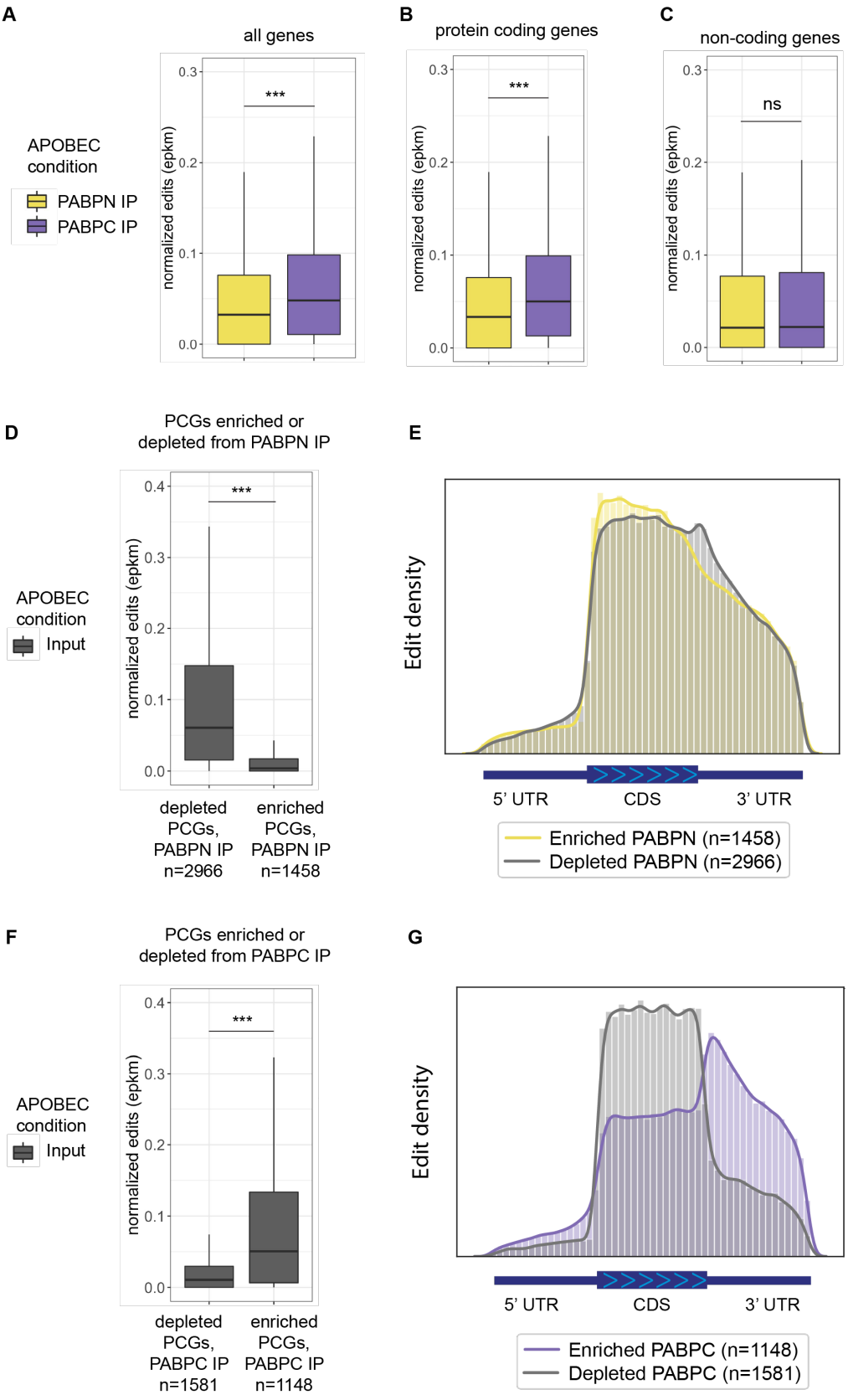
3.2.5 Ribosome contacts are higher on PCG transcripts associated with PABPC

To characterize the translation status of transcripts specifically while they are associated with PABPN or PABPC, we turned to recently developed technology that utilizes the RNA-editing enzyme APOBEC1 fused to the small ribosomal subunit RPS2, called Ribo-STAMP (Surveying Targets by APOBEC-Mediated Profiling) (Brannan et al. 2021). In this system, when RPS2 is associated with an RNA (either during scanning or when complexed with the large ribosomal subunit for translation), APOBEC1 can edit that RNA in regions that are proximal to RPS2, resulting in a C to U edit that can be readily detected as a mutation by Illumina sequencing. A higher level of editing suggests a more well-translated substrate. Using stable HEK293T Ribo-STAMP cell lines generated by lentiviral integration, we performed PABPN and PABPC RIPs followed by RNA-seq. Because we are investigating the cytoplasmic process of translation, cells were fractionated and only the cytoplasmic portion was used for PABPC and PABPN RIPs. The APOBEC-RPS2 or APOBEC-control constructs were induced for 24 hours, the shortest time period published, at the lowest doxycycline concentration, in order to avoid unintended cellular

effects of editing. Principal component analysis (PCA) indicated high reproducibility for the three biological replicates and no obvious effect of the APOBEC constructs on separate clustering of each IP and the input samples (Supp. Figure 3.7A).

Coupling Ribo-STAMP with PABPN and PABPC RIPs enabled us to study the translation status of transcripts for a given gene while associated with these different poly(A) binding proteins. We found that when transcripts are associated with PABPC, they tend to exhibit much higher levels of editing compared to when they are associated with PABPN (Figure 3.2A). Despite a higher degree of editing present on PABPC bound transcripts, many edits were detected above background on transcripts immunoprecipitated with PABPN (Figure 3.2A). The editing that we detect when PCG transcripts are with PABPN suggests that translation is occurring while RNAs are still associated with PABPN in the cytoplasm. The PABPN RIP can potentially pull down transcripts that are coated entirely with PABPN on the poly(A) tail as well as transcripts that have both PABPN and PABPC on their poly(A) tail, as these two states are not known to be mutually exclusive (Hosoda, Lejeune, and Maquat 2006). In either case, the higher level of editing on transcripts immunoprecipitated with PABPC suggests that the association of PABPN with translating RNAs is usually transient. The significant difference in editing due to a transcript being with PABPC remained evident when only protein-coding genes were considered (Figure 3.2B). However, non-coding transcripts received similarly low levels of editing when bound to either PABP (Figure 3.2C). Thus, association with PABPC alone is not sufficient for supporting ribosome occupancy.

Figure 3.2: Transcripts bound to PABPC associate with the ribosome. (A) Boxplot of edit scores (>= 0.5 confidence score) for all transcripts in PABPC IP or PABPN IP. Genes displayed have a TPM >= 5 in both IP conditions (n=10,391). ***p < 0.001; two tailed Kolmogorov-Smirnov test. (B) Boxplot of edit scores (>= 0.5 confidence score) for protein-coding genes (TPM >= 5, n = 9607) when they are associated with PABPN or when associated with PABPC. ***p < 0.001; two tailed Kolmogorov-Smirnov test. (C) Boxplot of edit scores (>= 0.5 confidence score) for non-coding genes (TPM >= 5, n = 784) when they are associated with PABPN or when associated with PABPC. Significance calculated with two tailed Kolmogorov-Smirnov test. (D) Boxplot of edit scores (>= 0.5 confidence score) from cytoplasmic input APOBEC condition for transcripts of protein-coding genes depleted or enriched with PABPN (TPM >= 5, n depleted = 2966, n enriched = 1458). ***p < 0.001; two tailed Kolmogorov-Smirnov test. Significance was determined by DESeq2 with cut-offs of log2FoldChange > 0.5 (enriched) or < -0.5 (depleted), baseMean > 50 and padj <=0.01. (E) Metagene plot showing edit (>= 0.5 confidence score) distribution for transcripts from protein-coding genes enriched or depleted from PABPN across 5' UTR, CDS, and 3' UTR gene regions, when they were associated with PABPN. (F) Boxplot of edit scores (>= 0.5 confidence score) from cytoplasmic input APOBEC condition for transcripts of protein-coding genes depleted or enriched with PABPC (TPM >= 5, n depleted = 1581, n enriched = 1148). ***p < 0.001; two tailed Kolmogorov-Smirnov test. Significance was determined by DESeq2 with cut-offs of log2FoldChange > 0.5 (enriched) or < -0.5 (depleted), baseMean > 50 and padj <=0.01. (G) Metagene plot showing edit (>= 0.5 confidence score) distribution for transcripts from protein-coding genes enriched or depleted from PABPC across 5' UTR, CDS, and 3' UTR gene regions, when they were associated with PABPC.


A
all genes
B
protein coding genes
C
non-coding genes
APOBEC condition
PABPN IP
PABPC IP
normalized edits (epkm)
0.0
0.1
0.2
0.3
***
ns
D
PCGs enriched or depleted from PABPN IP
APOBEC condition
Input
0.4
depleted PCGs, PABPN IP n=2966
enriched PCGs, PABPN IP n=1458
E
Edit density
5' UTR
CDS
3' UTR
Enriched PABPN (n=1458)
Depleted PABPN (n=2966)
F
PCGs enriched or depleted from PABPC IP
depleted PCGs, PABPC IP n=1581
enriched PCGs, PABPC IP n=1148
G
Enriched PABPC (n=1148)
Depleted PABPC (n=1581)

We next analyzed the levels and patterns of Ribo-STAMP edits on transcripts of genes enriched and depleted in the PABP IPs. Because this Ribo-STAMP experiment was focused on PABPN and PABPC in the cytoplasm, enrichment and depletion in each IP compared to the total cytoplasmic input was calculated using DESeq2. Comparisons were first made between the APOBEC-RPS2 and APOBEC-Control cell lines to determine if inducing editing with APOBEC-RPS2 altered the binding profiles of PABPN and PABPC. However, zero genes were found to be differentially enriched or depleted due to APOBEC-RPS2 induction, and therefore all samples were used together to determine enrichment and depletion for these cytoplasmic samples (Supp. Figure 3.7A). Using editing status in the input condition, genes depleted from PABPN have higher editing levels than those that are enriched with PABPN (Figure 3.2D). By plotting the position of the edits across a composite PCG, we observed a slight bias for edits towards the beginning of the coding sequence (CDS) in genes enriched with PABPN and a slight bias for edits closer to the end of the CDS in genes that were depleted from the PABPN IP (Figure 3.2E). Genes found to be enriched with PABPC had higher editing levels than those that were depleted from PABPC (Figure 3.2F). The pattern of edits was strikingly distinct on the transcripts for genes enriched and depleted in the PABPC IPs. Genes enriched with PABPC exhibited a peak of edits near the stop codon and high levels of editing throughout the 3' UTR, comparable to the density seen in the CDS (Figure 3.2G). In contrast, genes depleted in the PABPC IP accumulated edits primarily in the CDS with reduced levels in the 3'UTR. Considering the previous positive correlation between PABPC-enrichment and highly translated genes (Figure 3.1I and Supp Table S4), we asked if this metagene profile was

characteristic of well-translated genes. Graphing the top quartile of highest Ribo-STAMP edited genes revealed higher editing in the 3' UTR compared to genes with the lowest quartile of editing (Supp. Figure 3.7B), suggesting that high levels of translation may contribute to this distribution of edits along the 3' UTR. Overall, these results indicate that the occupancy and access to a transcript by the ribosome is influenced by its PABP-bound state.

3.2.6 The RNA binding profile of PABPC depends on active translation

Based on the correlations we saw linking translation and PABPC enrichment, we asked if active translation has a role in determining the transcripts bound to PABPC. In support of a model where translation facilitates the transition of a transcript from being bound by PABPN to being bound by PABPC, studies using an iron-response element reporter to control translation of a B-globin construct found a decrease in PABPC association and an increase in PABPN association when translation was halted (Sato and Maquat 2009). To investigate if this occurs on a global scale for endogenous genes, translation was blocked with the inhibitor drugs cycloheximide and puromycin for four hours, followed by PABPC RIPs (Supp. Figure 3.8A). Inhibition was confirmed by ^{35}S labeling of newly made proteins (Supp. Figures 3.8B and 3.8C). This brief treatment period was used so as to minimize global stress responses and instead capture any changes in the relative pools of RNA bound to PABPC. DESeq2 was used to calculate a ratio of change for PABPC association, comparing the baseline binding of a transcript in control

conditions to the ratio of that transcript following translation inhibition, thus giving a metric for how translation inhibition affected enrichment or depletion. We found that when translation was inhibited by cycloheximide or puromycin, hundreds of genes changed in their relative level of enrichment or depletion with PABPC (Figure 3.3A and 3.3B). Thus, active translation contributes to the binding profiles of PABPC.

Since transcripts initially bind PABPN and then transition to PABPC, and the reverse binding order is not known to occur, changes in PABP binding profiles are likely influenced by changes in RNA abundance. Compared to control (DMSO) conditions, 307 genes increased by at least two-fold, and 113 genes decreased in expression by at least-two fold in the presence of cycloheximide (Supp Table 5); comparable numbers of gene expression changes were induced by puromycin treatment (Supp Table 5). By comparing abundance values to $\log_2$FoldChange values for PABPC IP, we found that upon translation inhibition many of the genes with reduced PABPC binding increased in abundance (Figure 3.3A and 3.3B). An apparent increase in gene expression could result if addition of the translation inhibitor upregulated its transcription or caused large scale decreased expression of other genes. Since only 113 genes decreased in abundance by two-fold during the four-hour treatment with cycloheximide or puromycin, the former likely explains most of the increases in gene expression. To look more directly at transcripts that are increasing at the transcriptional level, we calculated the change in intronic reads for each gene in input control conditions compared to input translation inhibition conditions. Using this metric, we see the same correlation: genes that show an increase in intronic reads are more likely to be depleted from PABPC following translation inhibition, both for cycloheximide (Figure 3.3C, spearman corr. -0.31, $p<2e-16$) and puromycin treatment

(Figure 3.3D, spearman corr. -0.59, p<2e-16). These results suggest that active translation plays a role in facilitating the association of PABPC with newly made transcripts.

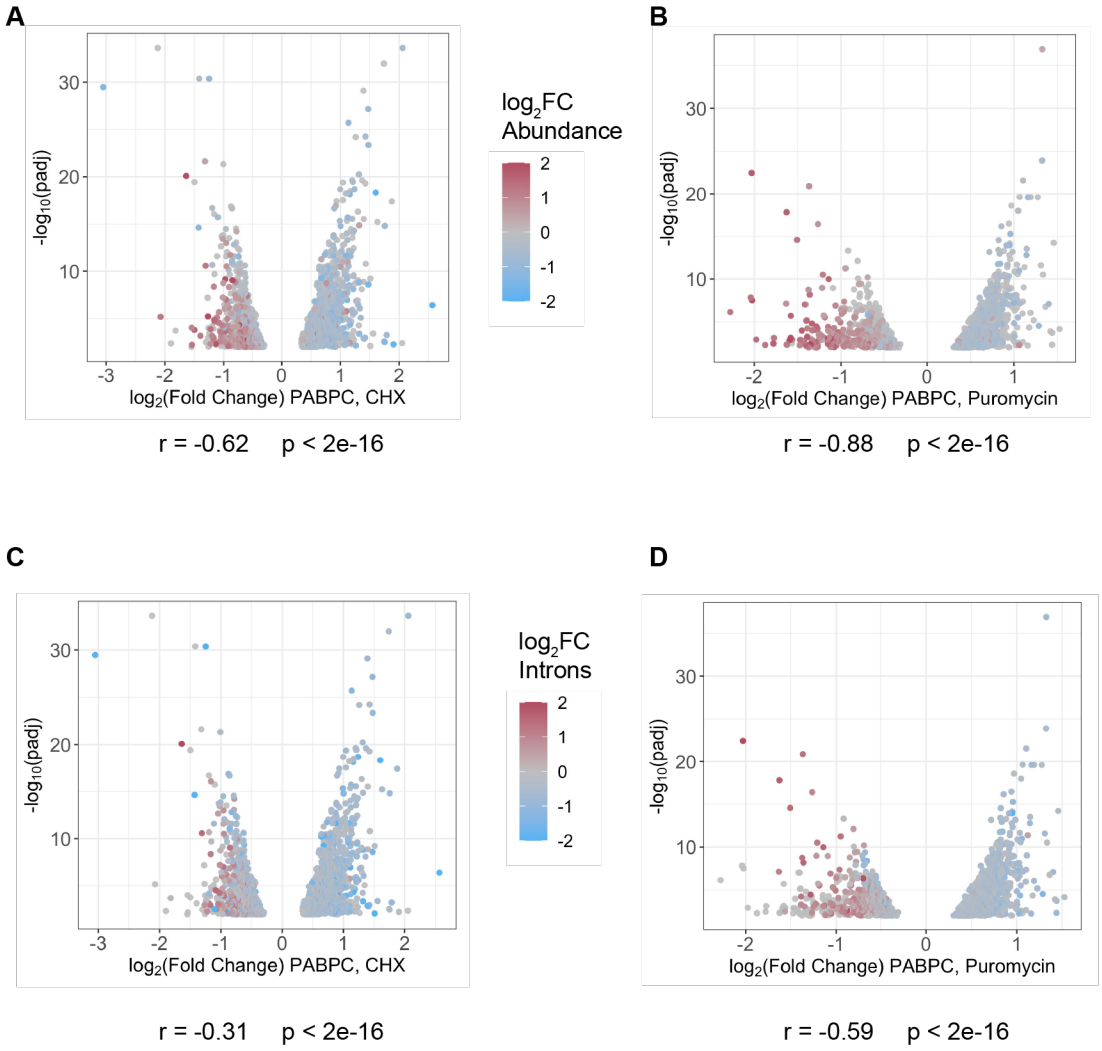


Figure 3.3: Translation influences the association of transcripts with PABPC. (A and B) Volcano plot of changes in enrichment or depletion with PABPC following A) cycloheximide treatment or B) puromycin treatment. Only significant genes are shown: baseMean > 50 and padj <= 0.01. Colors indicate whether that transcript increased (red) or decreased (blue) in overall abundance in input treatment vs. input control conditions due to the four-hour translation inhibition treatment. (C and D) Volcano plot of changes in enrichment or depletion with PABPC following C) cycloheximide treatment or D) puromycin treatment. Only significant genes are shown: baseMean > 50 and padj <= 0.01. Colors signify an increase (red) or decrease (blue) in $\log_2$ fold change of intronic reads in input treatment vs. input control conditions, indicating an increase or decrease in transcription.

3.2.7. Enrichment with PABPN or PABPC is related to cellular localization

Considering the predominant localization of PABPN in the nucleus and PABPC in the cytoplasm (Krause et al. 1994; Blobel 1972), we predicted that RNAs enriched with each protein would similarly partition. First, we plotted the TPM values for transcripts of genes when associated with PABPN versus when associated with PABPC, which revealed a strong bias of many genes for associating with one protein more than the other. To determine if these profiles correlate to biased localization of an RNA either in the nucleus or the cytoplasm, we performed subcellular nuclear and cytoplasmic fractionation and sequenced the RNAs in each compartment (Supp. Figures 3.9A and 3.9B). By comparing relative RNA localization to the abundance of that RNA in each of the RIPs, we observed that RNA localization largely reflects enrichment with PABPN or PABPC (Figure 3.4A). Genes that are highly cytoplasmic tend to be more associated with PABPC than with PABPN, and genes that are more nuclear restricted tend to be enriched with PABPN. As expected, the nuclear localized non-coding RNAs XIST, NEAT1 and MALAT1 show a much greater association with PABPN (Figure 3.4A). In contrast, highly stable transcripts, such as those encoding ribosomal proteins, predominantly exist in the cytoplasm bound to PABPC (Figure 3.4A).

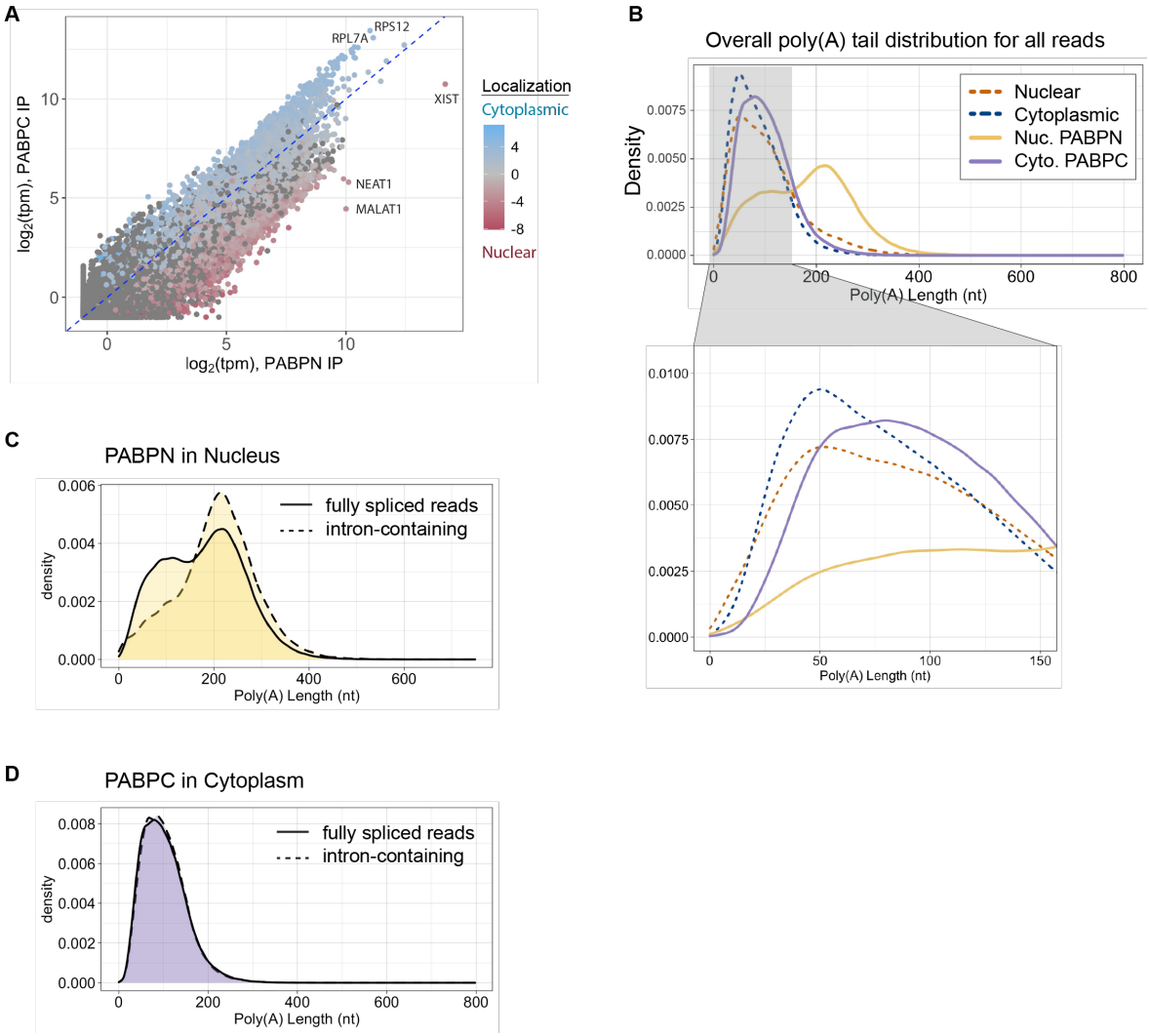


Figure 3.4: Incompletely spliced transcripts with longer poly(A) tails are associated with PABPN (A) TPM of PABPN IP compared to TPM of PABPC IP, colored by degree of enrichment in cytoplasmic (blue) or nuclear (red) fractions. Select abundant genes highly enriched in the nucleus or cytoplasm are labeled by name. (B) Density plot showing overall poly(A) tail length distribution of all nuclear-encoded genes detected by Nanopore direct RNA sequencing. Dashed lines indicate total cytoplasm and total nuclear fraction, and solid lines indicate the two IP conditions. Region from 0-150 nt is zoomed in on in the bottom panel to see the differences in distributions for shorter poly(A) tails. Number of reads in each density plot line are as follows: Nuclear 585,914; Cytoplasm 1,039,310; Nuclear PABPN 300,565; Cytoplasm PABPC 1,047,845. **(**C) Density plot of poly(A) tail length for reads that contained one or more introns (dashed line, n = 29,675) or no introns (solid line, n = 250,115) while with PABPN in the nucleus. **(**D) Density plot of poly(A) tail length for reads that contained one or more introns (dashed line, n = 63,164) or no introns (solid line, n=877,256) while with PABPC in the cytoplasm.

3.2.8 Poly(A) tails are longer when a transcript is associated with PABPN

Previous poly(A) tail analyses concluded that tails are longest in the nucleus and undergo shortening over time once they are in the cytoplasm (Diana Sheiness and Darnell 1973; Palatnik, Storti, and Jacobson 1979). The degree of difference reported in these early studies for nuclear and cytoplasmic steady-state poly(A) tail sizes was relatively minimal. For instance, work in HeLa cells showed a 30 nt difference between the predominant peak of tail sizes found in the nucleus versus the cytoplasm (Diana Sheiness and Darnell 1973). More recent work using 5-ethynyl uridine (5EU) time-course labeling in mouse 3T3 cells was able to capture a greater difference between newly made and steady state tails, potentially due to this specific time-course labeling and high-throughput sequencing (Eisen et al. 2020). Although they note great inter- and intragenic variation for tail length, they found the peak of distributions for newly made tails centers around 175 nt long, while steady state centered around 100 nt (Eisen et al. 2020). Considering the model of sequential binding of PABPN to nascent transcripts followed by PABPC as they undergo translation, we asked if the RNAs bound by these proteins had different poly(A) tail lengths.

Using subcellular fractionated extracts, we performed RIP of PABPN from the nucleus and PABPC from the cytoplasm and used Nanopore direct RNA sequencing to look at poly(A) tail length on transcripts when they were associated with these proteins (Supp. Figures 3.9A and 3.9B). Direct RNA sequencing has the distinct advantage of being able to sequence the entire transcript as well as using pore dwell time to infer poly(A) tail

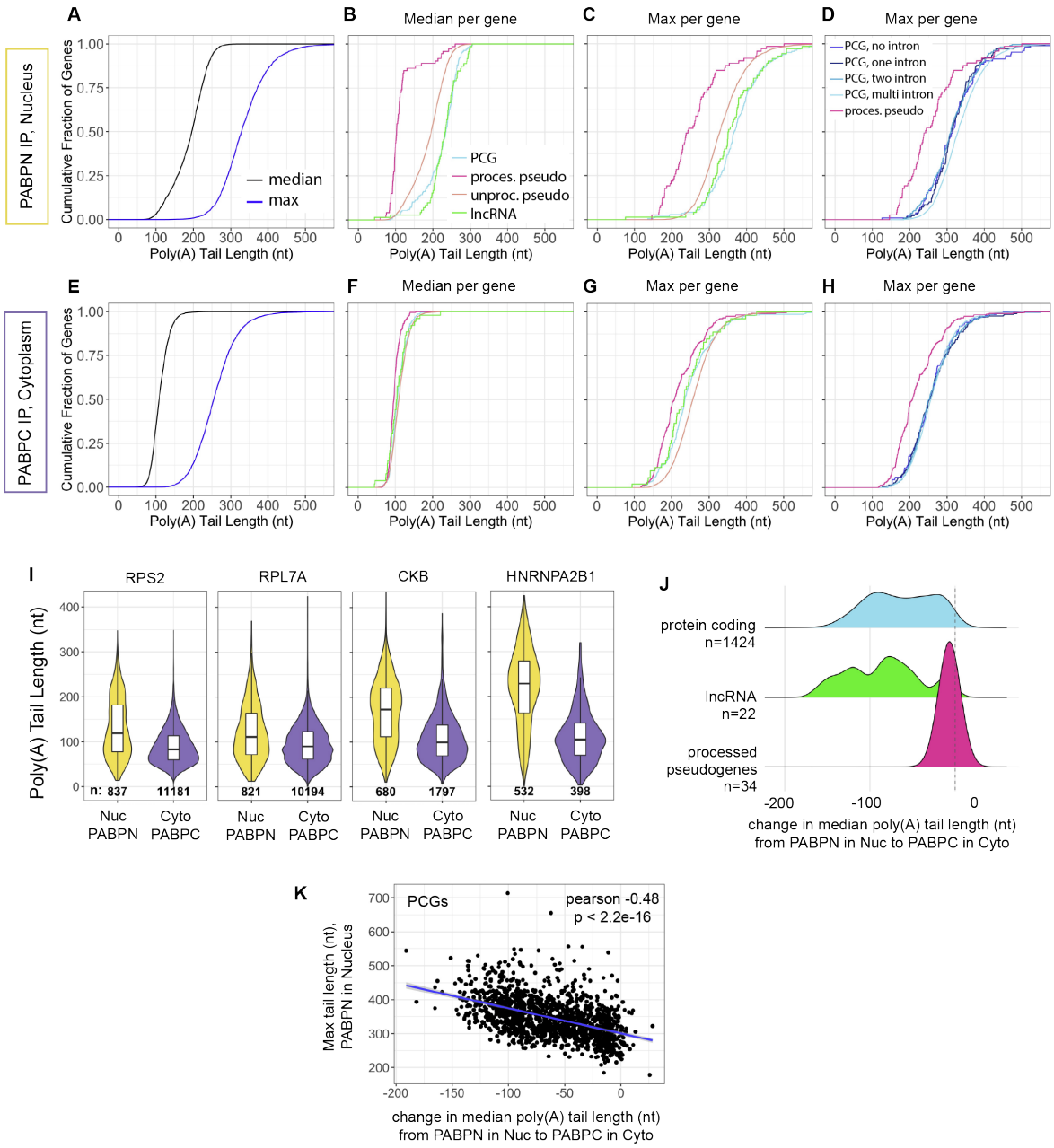
length (Garalde et al. 2018; Workman et al. 2019). Consistent with previous reports (Diana Sheiness and Darnell 1973; Palatnik, Storti, and Jacobson 1979; Eisen et al. 2020), a greater fraction of long poly(A) tail reads were observed for total nuclear versus cytoplasmic RNAs (Figure 3.4B). Close examination of the shortest detected tail sizes showed that tail sizes less than 10 nt were detected in the total nuclear and cytoplasmic RNA but were largely nonexistent in the IP samples. This observation is consistent with a minimal poly(A) tail size greater than 10 being needed for stable association with PABPN or PABPC *in vivo*. Most striking, though, was the difference in tail lengths of RNAs associated with PABPN in the nucleus compared to all other samples, including the total nuclear fraction (Figure 3.4B). These results suggest that association with PABPN may be a prerequisite for the maintenance of poly(A) tail sizes over 200 nt.

Since we detected intronic reads associated with PABPN and, to a much lesser extent, PABPC (Figure 3.1E), we took advantage of the full length reads generated by nanopore direct RNA sequencing to further examine the pre-mRNAs bound by these proteins. When reads were separated based on whether they contained introns or not, we observed distinct poly(A) tail size profiles for the PABPN bound transcripts. Transcripts that still had at least one intron present tended to have a relatively uniform longer poly(A) tail distribution that centered around 230 nt (Figure 3.4C). In contrast, fully spliced transcripts produced a peak around 230 nt as well as a broad shoulder of shorter tail sizes (Figure 3.4C). To investigate whether this was due to an inherent difference in the types of genes in these categories, we only analyzed genes that had a representative intron-containing read and observed comparable profiles (Supp. Figure 3.9C). These patterns based on intron presence likely account for the phased profile of all tail sizes on transcripts

bound to PABPN seen in Figure 3.4B. The retention of an intron in the PABPC-bound transcripts did not seem to influence poly(A) tail length, as fully spliced and intron-containing transcripts exhibited nearly identical distributions that centered around 80 nt (Figure 3.4D). Taken together, these analyses show that in the nucleus, PABPN can associate with pre-mRNAs and that poly(A) tail sizes can reach maximum lengths prior to completion of splicing.

To determine if the dramatic difference in poly(A) tail sizes detected on transcripts associated with PABPN versus PABPC reflected the entire population or more limited sets of abundant transcripts (Figure 3.4B), we calculated the median and maximum tail lengths for each gene (Figures 3.5A-3.5H). Transcripts for 46% of all genes bound to PABPN had median poly(A) tail sizes over 200 nucleotides long, with a range of 43-325 nt (Figure 3.5A). In contrast, the median for all genes with transcripts bound to PABPC was 108 nt, with a range of 42-239 nt (Figure 3.5E). Overall, most genes exhibited much longer poly(A) tail sizes on transcripts bound to PABPN compared to PABPC. This is exemplified in Figure 5I where we plotted the tail length for all transcripts of a given gene associated with PABPN or PABPC (Figure 3.5I). The transcripts for the ribosome protein-encoding genes, RPS2 and RPL7A, are among the most abundant detected in RIPs for both PABPN and PABPC and exhibit distinct poly(A) tail size profiles when associated with each protein. The median and average tail lengths are shorter on transcripts associated with PABPC (Figure 3.5I). This difference is exacerbated for genes with high numbers of transcripts bound to PABPN, such as CKB and HNRNPAB1 (Figure 3.5I).

Figure 3.5 Poly(A) tail size differs depending on whether a transcript is associated with PABPN or PABPC. (A) Cumulative plot showing median and maximum poly(A) tail length for each gene that had at least 10 reads in PABPN IP from the nucleus (n genes = 7343). Graphing area for all cumulative plots has been limited to 0-550 nt. **(**B) Cumulative plot showing median tail length detected in PABPN IP in the nucleus, separated by gene biotype, for genes that had at least 10 reads. PCG n = 7025, processed pseudogene n = 73, unprocessed pseudogene n = 72, lncRNA n = 168. **(**C) Cumulative plot showing maximum tail length detected in PABPN IP in the nucleus, separated by gene biotype, for genes that had at least 10 reads. PCG n = 7025, processed pseudogene n = 73, unprocessed pseudogene n = 72, lncRNA n = 168. **(**D) Cumulative plot showing maximum tail length detected in PABPN IP in the nucleus for protein coding genes that contain none, one, two or multiple introns, as well as processed pseudogenes, for genes that had at least 10 reads. PCG no introns n = 104, PCG 1 intron n = 140, PCG 2 introns n = 217, PCG 3+ introns n = 6563, processed pseudogene n = 72. **(**E) Cumulative plot showing median and maximum poly(A) tail length for each gene that had at least 10 reads in PABPC IP from the cytoplasm (n genes = 7736). **(**F) Cumulative plot showing median tail length detected in PABPC IP from the cytoplasm, separated by gene biotype, for genes that had at least 10 reads. PCG n = 7209, processed pseudogene n = 236, unprocessed pseudogene n = 52, lncRNA n = 233. **(**G) Cumulative plot showing maximum tail length detected in PABPC IP from the cytoplasm, separated by gene biotype, for genes that had at least 10 reads. PCG n = 7209, processed pseudogene n = 236, unprocessed pseudogene n = 52, lncRNA n = 233. **(**H) Cumulative plot showing maximum tail length detected in PABPC IP from the cytoplasm for protein coding genes that contain none, one, two or multiple introns, as well as processed pseudogenes, for genes that had at least 10 reads. PCG no introns n = 134, PCG 1 intron n = 209, PCG 2 introns n = 308, PCG 3+ introns n = 6557, processed pseudogene n = 236. **(**I) Violin plots of individual genes and their poly(A) tail distributions when associated with PABPN in the nucleus and when associated with PABPC in the cytoplasm. RPS2 and RPL7A are among the top most abundant genes in both PABPN IP from the nucleus and PABPC IP from the cytoplasm. CKB and HNRNPA2B1 are among the top most abundant with PABPN in the nucleus. Number of reads for each violin are shown in black text at the base of the violin. White boxplots are inlaid within the violin, indicating the median and upper and lower quartiles. The lines extending out from the central box indicate the minimum and maximum value in that dataset. **(**J) Density plots of the change in median poly(A) tail length when a transcript is with PABPN in the nucleus compared to PABPC in the cytoplasm, separated by gene biotype. A negative change indicates that the poly(A) tail was shorter when associated with PABPC in the cytoplasm. Genes must have been represented by at least 35 reads in each IP to be displayed. n (protein coding genes) = 1424, n (lncRNAs) = 22, n (processed pseudogenes) = 34. Unprocessed pseudogenes did not have enough reads to pass cut-offs and are thus not displayed here. **(**K) Change in median poly(A) tail length of all protein coding genes that were displayed in blue in (J), compared to their maximum detected poly(A) tail length when associated with PABPN in the nucleus. Blue line is a best fit line using a linear model.


A
B
Median per gene
C
Max per gene
D
Max per gene
PABPN IP, Nucleus
Cumulative Fraction of Genes
median
max
PCG
proces. pseudo
unproc. pseudo
lncRNA
PCG, no intron
PCG, one intron
PCG, two intron
PCG, multi intron
proces. pseudo
Poly(A) Tail Length (nt)
E
F
Median per gene
G
Max per gene
H
Max per gene
PABPC IP, Cytoplasm
I
RPS2
RPL7A
CKB
HNRNPA2B1
Poly(A) Tail Length (nt)
n: 837
11181
821
10194
680
1797
532
398
Nuc PABPN
Cyto PABPC
J
protein coding n=1424
lncRNA n=22
processed pseudogenes n=34
change in median poly(A) tail length (nt) from PABPN in Nuc to PABPC in Cyto
K
PCGs
pearson -0.48
p < 2.2e-16
Max tail length (nt), PABPN in Nucleus
change in median poly(A) tail length (nt) from PABPN in Nuc to PABPC in Cyto

3.2.9 Transcripts with poly(A) tails longer than 200 nt are bound by PABPN and PABPC

While the median poly(A) tail size largely differed when transcripts for a given gene were associated with PABPN (196 nt) versus PABPC (108 nt), much longer tailed representatives were isolated with both PABPs. The maximum poly(A) tail size was typically over 200 nucleotides for transcripts of a given gene when associated with PABPN or, unexpectedly, PABPC (Figures 3.5A and 3.5E). The maximum tail length for genes detected with at least 10 reads in the PABPN IP was greater than 200 nt for more than 99% of genes (Figure 3.5A). This is consistent with earlier *in vitro* polyadenylation and radioactive studies, which demonstrated that the majority of newly-made poly(A) tails are around 200 adenosines long (Brawerman 1981; Diana Sheiness and Darnell 1973). With PABPC in the cytoplasm, more than 86% of genes have a transcript with a tail length greater than 200 nt, suggesting that shortening of the poly(A) tail may occur after the RNA reaches the cytoplasm and acquires PABPC on its poly(A) tail (Figure 3.5E).

When the median and maximum poly(A) tail length values were separated by gene biotype, processed pseudogenes stood out as a class exhibiting much shorter tail lengths than other biotypes when associated with PABPN (Figures 3.5B and 3.5C). This difference was much less evident when transcripts for processed pseudogenes were associated with PABPC (Figures 3.5F and 3.5G). Processed pseudogenes typically lack introns as they are thought to have undergone retrotransposition from a mature RNA, rather than a duplication event which occurs with unprocessed pseudogenes (Kazazian 2014). Therefore, we asked if lacking introns might be related to the shorter median and maximum tail lengths

observed for processed pseudogenes by separating PCGs into those with no encoded introns, one intron, two introns or multiple introns (3 or more encoded in the genome). None of these categories match the distinctly shorter length of processed pseudogenes with PABPN, indicating that the presence of introns and a need for splicing is not necessary for acquiring a maximum tail length greater than 200 nt (Figures 3.5D and 3.5H).

For all gene biotypes, the median poly(A) tail lengths were substantially shorter for transcripts of a given gene bound by PABPC compared to PABPN and different genes showed different degrees of shortening (Figures 3.5I, Supp. Figures 3.10A and 3.10B). The extent of shortening is stunted for processed pseudogenes due to their shorter initial poly(A) tail lengths (Figure 3.5J). The initial poly(A) tail size, inferred as the maximum tail length for a gene bound by PABPN, varied greatly for protein coding genes (range: 125-713 nt) and the longer the initial tail length, the greater shortening that gene underwent after association with PABPC (Figure 3.5K). This difference suggests that substantial pruning of the poly(A) tail occurs as a transcript transitions from binding PABPN to PABPC. Altogether, these analyses show that distinct populations of RNAs, differentiated by splicing status and poly(A) tail length, are present with PABPN versus PABPC.

3.3 Discussion

The prevailing model of PABP association suggests that for the majority of RNAs, PABPN binds nascent RNAs in the initial stages of polyadenylation and facilitates creation of the poly(A) tail. Upon transport to the cytoplasm, PABPN is replaced by PABPC on the poly(A) tail, which promotes translation and stability of the RNA. Here, we examine

several aspects of this model by performing comprehensive and direct RNA sequencing of RNAs bound by PABPN and PABPC in human cells. While almost all known polyadenylated transcripts were detected in both the PABPN and PABPC IPs, the relative amount differed substantially for many RNAs. Consistent with the model, transcripts enriched with PABPN tended to have biased nuclear localization and hallmarks of nascent RNAs, such as intronic reads and long poly(A) tails. In contrast, transcripts enriched with PABPC tended to be more efficiently translated and longer-lived with pruned poly(A) tails. Moreover, our analyses raised new considerations for timing and roles of PABPN and PABPC in their association with a given RNA from its synthesis to its functional state.

Upon investigating whether active translation is necessary for the binding profile seen with PABPC, we identified many changes in RNA enrichment following translation inhibition. These findings are in line with previous reports showing that inhibiting translation of a reporter altered its binding to PABPC (Hosoda, Lejeune, and Maquat 2006). If nuclear export and the move to the cytoplasm alone allows the transition to PABPC, then we would not anticipate much change in PABPC enrichment following translation inhibition. This type of model might suggest that binding kinetics and the abundance of PABPC in the cytoplasm causes this transition. However, the changes in enrichment that we saw following translation inhibition suggest that translation can affect the binding of transcripts to PABPC in the cytoplasm. Because our studies are a snapshot of PABPC binding at a single time point, we cannot determine whether transcripts that become more depleted from PABPC because of the translation inhibition conditions were made so recently that they have not left the nucleus, thus implicating transcription as the driving factor behind the changes we are seeing. Although it is difficult to measure the

kinetics of nuclear export, estimates from several studies put this in the range of a few minutes to possibly up to 30 minutes (Siebrasse, Kaminski, and Kubitscheck 2012; Janicki et al. 2004). This suggests that our four-hour translation inhibition gives substantial time for transcription and subsequent export out to the cytoplasm, indicating that the depletion from PABPC is less likely to be the result of new transcripts not having sufficient time to be exported and engage in this transition to PABPC.

Recent genome-wide studies have discovered that steady state poly(A) tail length is not uniform for all transcripts of a given gene and median tail sizes for each gene vary considerably (Webster et al. 2018; Yi et al. 2018; Subtelny et al. 2014; Lima et al. 2017; Bartel and Xiang 2021). Most of these studies have captured a singular distribution of poly(A) size for each gene, taking a snapshot of the entire cell at once. Eisen et al. performed a time-course to investigate cytoplasmic poly(A) tail lengths, showing that greater variation can be captured this way (Eisen et al. 2020). Here we show that part of this variation results from distinct poly(A) tail profiles of RNAs associated with PABPN versus PABPC. By isolating each endogenous protein from the same cell extract, we were able to make direct comparisons between PABPN- and PABPC-bound transcripts. Consistent with the time course study showing longer poly(A) tails on newly synthesized transcripts (Eisen et al. 2020), we found that RNAs bound by PABPN tended to have much longer poly(A) tails than those bound by PABPC. The striking difference in poly(A) tail profiles for RNAs isolated with PABPN compared to nuclear extract suggests that PABPN IP selects for a subset of nuclear RNAs that are obscured in the total nuclear sample. From this population of RNAs captured by PABPN, we were able to determine that virtually all polyadenylated genes achieve maximal poly(A) tail lengths of at least 200 nucleotides and

many can be hundreds of nucleotides longer. Illumina-based sequencing approaches measuring the poly(A) tail are unable to accurately quantify tails beyond 230 nt (Chang et al. 2014; Lima et al. 2017; Lim et al. 2016). Previously published Nanopore sequencing also detected poly(A) tails up to ~600 adenosines in human cells (Workman et al. 2019). The functional consequences of having a poly(A) tail longer than 200-300 nt have not been investigated and may have implications for downstream interactions.

Interestingly, maximum tail lengths greater than 200 nt were found for the majority of genes whether a transcript was associated with PABPN or PABPC (Figures 3.5A and 3.5E). This supports a model where tails can reach a length of at least 200 adenosines when they are associated with PABPN in the nucleus and subsequent shortening of the RNA, termed pruning, largely occurs after a transcript is bound by PABPC in the cytoplasm. The functional implications of this process and why a cell would expend energy to produce a long poly(A) tail just to shorten it later is not understood. Evidence supporting tail shortening exists in many organisms, suggesting that pruning may be a coordinated, conserved process (Subtelny et al. 2014; Lima et al. 2017). Recent studies of deadenylases that delineated their precise functions in the presence of PABPC have provided insights into some of the players (Webster et al. 2018; Yi et al. 2018), but many mechanistic details remain to be discovered. Furthermore, our RIP experiments differentiate transcripts based on PABP partners but for a given PABP, we cannot distinguish between RNAs of different ages. Future time-course studies could elucidate whether the tail length variation observed within the PABPC dataset or the PABPN dataset is due to aging of the RNA or intragenic variation.

Overall, our findings provide a genome-wide view of the identities, splicing status, ribosome occupancy, and polyadenylation state of RNAs that preferentially associate with PABPN or PABPC in human cells. A broad implication of this work is that most RNAs exist as a heterogenous pool, partly distinguished by being bound to PABPN or PABPN, and thus, may be differentially susceptible to specific post-transcriptional regulatory mechanisms. Read-outs of regulation such as changes in steady state mRNA levels or poly(A) tail length may gain sensitivity is specific PABP-bound transcripts are considered instead of the entire cell population.

3.4 Methods

Cell Culture

HEK293T cells were cultured as recommended by the manufacturer, in Dulbecco's modified Eagle medium (DMEM, Gibco #11965092) supplemented with 10% fetal bovine serum (Gibco #10437-028) and 1% penicillin/streptomycin (Gibco).

Formaldehyde Crosslinking

10% formaldehyde stock solution was made by heating paraformaldehyde to crack it and storing single-use aliquots at -20°C until ready to use. Cells were washed twice on the plate with phosphate-buffered saline (PBS) and then collected into conical tubes with PBS and spun down. Cell pellets were resuspended in 0.1% formaldehyde in PBS and incubated at room temperature for 10 minutes. Glycine was added to a final concentration of 0.17 mM to quench the formaldehyde and incubated for 5 min. Cells were then spun down and

washed with PBS two times before pellets were snap frozen on liquid nitrogen for storage at -80°C.

siRNA transfection

Knockdowns to validate antibodies used in immunoprecipitations were performed with 200uM small interfering (si)RNAs targeting either luciferase (control), PABPN or PABPC, using siLentFect (Bio-Rad) transfection reagent according to manufacturer's recommendations at 72 and 24 hours before harvest. siRNA duplex oligos ordered from Dharmacon were as follows: siLuciferase: CGUACGCGGAAUACUUCGAUU; siPABPN: GUAGAGAAGCAGAUGAAUA; and siPABPC: GAAAGGAGCUCAAUGGAAA. siPABPC and siPABPN were previously validated and published (Y. J. Lee and Glaunsinger 2009).

RNA Immunoprecipitation

Cell pellets were resuspended in ice-cold RIPA buffer (Thermo Fisher) supplemented with 40U/mL rRNAseIn Plus (Promega), 0.5mM DTT, 5mM EDTA, and 1 tablet/25mL mini cOmplete protease inhibitor cocktail (Roche). Cells were sonicated three times at 8 watts: 20 seconds on, 2 minutes off, in an ice bath. Lysates were spun down and Protein G Dynabeads (Invitrogen) were prepared to pre-clear the lysates by washing with RIPA buffer twice. Pre-clearing was performed on the nutator for 30 minutes at 4°C. After this, pre-cleared lysates were incubated with 5 ug antibody for 2 hours at 4°C on nutator (anti-PABPC, ab21060 Abcam; anti-PABPN, EP3000Y ab75855 Abcam, isotype control anti-

rabbit IgG monoclonal ab172730 Abcam; isotype control anti-rabbit IgG polyclonal ab171870 Abcam). Protein G Dynabeads were again prepared by washing with RIPA buffer and 100uL slurry was added per IP, rotating on nutator for 1 hour at 4°C. IPs were washed four times with supplemented RIPA buffer. Final beads were eluted in elution buffer (50mM Tris-HCl pH 7.2, 5mM EDTA, 1% SDS, 10mM DTT) and proteinase K (NEB). Reverse crosslinking was performed on the thermomixer shaking at 1200 rpm, first at 60°C for 20 minutes to allow proteinase K to work, and then at 70°C for an additional 25 minutes. RNA was isolated using a standard Trizol (Life Technologies) RNA extraction.

Illumina Library Prep

For total cell lysate IPs, cDNA libraries from three independent biological replicates were prepared from 1ug RNA using Illumina ribodepleted RNA stranded kit. Libraries were sequenced on the Illumina HiSeq 4000, single-end reads (75 nucleotides).

For nuclear and cytoplasmic fractionation, cDNA libraries from two independent biological replicates were prepared from 400ng RNA using Illumina ribodepleted RNA stranded kit. Libraries were sequenced on the Illumina NovaSeq 6000, paired-end reads (100 nucleotides).

For APOBEC-RPS2 IPs, cDNA librares from three independent biological replicates were prepared from 200-300ng RNA using Illumina ribodepleted RNA stranded kit. Libraries were sequenced on the Illumina NovaSeq 6000, paired-end reads (100 nucleotides).

Illumina RNA-Seq Analysis

For total lysate IPs, libraries were at least 44 million reads per sample, with an average of 55 million reads. Average percent of uniquely mapped reads was 88%.

For fractionations, libraries were at least 28 million reads per sample, with an average of 29 million reads. Average percent of uniquely mapped reads was 82%.

For APOBEC experiments, libraries were at least 26 million reads per sample, with an average of 32 million reads. Average percent of uniquely mapped reads was 77%.

All reads were aligned to the human genome hg38 primary assembly using STAR. Bam files were sorted and indexed using samtools. featureCounts version 2.0.2 was used to annotate reads, using the flag --minOverlap 20 and a custom GTF file derived from gencode v34 as described in results (Liao, Smyth, and Shi 2014). Differential expression was calculated using DESeq2 (Love, Huber, and Anders 2014).

Due to the contribution of intronic reads to our datasets, calculation of TPM (Transcript per kilobase Million) values was performed by separately determining TPMs for exonic regions and intronic regions and then summing together. This was important in order to not skew values by using the full length of the unspliced gene, as introns are very long in comparison to exons. Additionally, for genes that had a large number of intronic reads,

using the spliced exonic length would also not be appropriate, as that gene would appear to be more highly expressed than it was.

Upstream transcription problem genes

Upstream intergenic regions were extracted using bedtools flank -l 2000 -r 0 -s. Intervening upstream genes that had any overlap with this region were removed with bedtools subtract.

Reads that did not align to intronic or exonic regions were extracted from bam files with fgrep. Coverage across the 2000 bp upstream region was determined with bedtools coverage -S -split. A TPM value was determined for this upstream region and compared to the TPM of the adjacent downstream gene. If a ratio of 20% of the signal (determined by TPM) was present in the upstream region, and a breadth of 70% of that region was covered, that gene was determined to have significant enough upstream transcription so as to not be a reliable functional coding product, and was removed from subsequent analysis. Genes that showed this pattern in either the input or IP condition were both removed from that IP analysis.

Western Blotting

Western blotting was performed as previously described (Zisoulis, Van Wynsberghe) using the following antibodies: Calreticulin, Cell Signaling 2891; Histone H3, abcam ab1791; U1 snRNP 70, Santa Cruz sc-390899; Pol-II, abcam ab5408; PABPC, abcam ab21060; PABPN, abcam ab75855; GAPDH proteintech 60004-1-Ig; Actin, MP Biomedicals

0869100-CF; Tubulin, Sigma T9026. Western blots were visualized using an Odyssey Fc imaging system (LI-COR).

Fractionation

Subcellular fractionation protocol was adapted from Gagnon, Li, Janowski and Corey paper with slight modifications to buffer recipes (Gagnon et al. 2014). Section B, "Preparation of cytoplasmic, total nuclear, nucleoplasmic and chromatin fractions for biochemical assays" was followed for total nuclear and cytoplasmic fractions. Changes to buffer recipes consisted of eliminating $MgCl_2$ and adding EDTA, which helps prevent deadenylation from occurring post-lysis. Hypotonic lysis buffer consisted of 10 mM Tris, pH 7.5, 10 mM NaCl, 1 mM EDTA, 0.3% NP-40 (vol/vol), and 10% glycerol (vol/vol). Nuclear Lysis buffer consisted of 20 mM Tris, pH 7.5, 150 mM KCl, 1 mM EDTA, 0.3% NP-40 (vol/vol), and 10% glycerol (vol/vol).

Nanopore Direct RNA Sequencing

RNA was prepared for nanopore direct RNA sequencing following the Oxford Nanopore Technologies (ONT) SQK-RNA002 kit protocol, including the optional reverse transcription step recommended by ONT. RNA was sequenced in-house on the minION platform using ONT R9.4.1 flow cells. Total reads (in millions) passing default filters were Cytoplasm: 2.5, Nuclear: 1.8, PABPC IP in Cytoplasm: 1.9, PABPN IP in Nucleus: 0.9

Nanopore Direct RNA Analysis

Direct RNA reads were basecalled in real time with the minKNOW software using guppy. Reads were mapped to the genome with minimap2 (v2.15) using the flags -ax splice -uf -k14 --secondary=no.

A minimap2 MAPQ score of 0 indicates multi-mapping and thus all alignments with a MAPQ score of 0 were filtered out from bam files, as well as any supplementary alignments, using samtools view -bq 1 -F 2048. Minimap2 does not currently output the typical "NH" flag in bam files which indicates number of mappings per read, therefore when using featureCounts to annotate these reads, if you do not want to count multi-mappers, you must filter your bam file first because featureCounts normally would use this NH flag to determine multi-mapping. Filtered bams were used with featureCounts version 2.0.2 to annotate reads, using the flags --minOverlap 20 and -L for long read mode.

After annotation, any reads that were not from nuclear-encoded genes and mapped to the mitochondrial genome were removed from subsequent analysis and graphing.

Poly(A) Length Estimation

We used the nanopolish-polya pipeline to estimate poly(A) tail lengths from basecalled reads. Reads were then filtered based on their QC tag, removing any reads that had QC tags of "READ_FAILED_LOAD", "SUFFCLIP", or "NOREGION."

Ribo-STAMP editing

For stable cell STAMP-fusion protein expression, cells were induced with 50ng/ml doxycycline in DMEM for 24 hours. APOBEC-RPS2 cell lines were used for the experimental condition and APOBEC-Control cell lines were used for control.

Edit distribution, EPKM and e score method details

EPKM values were determined as in (Brannan et al. 2021) and metagene plots were created using metaPlotR and seaborn.

Translation Inhibition

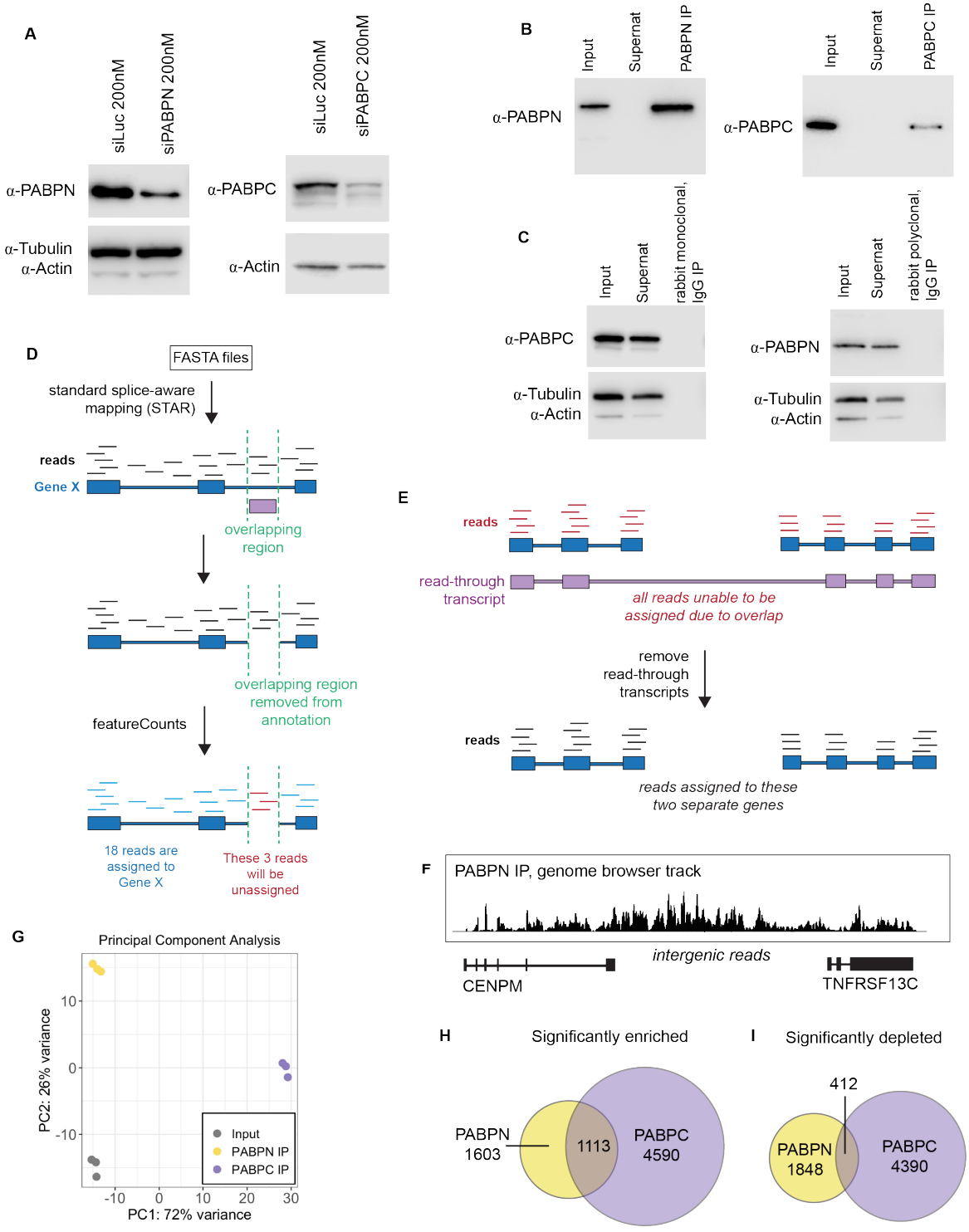
Cells were grown to 85% confluency before fresh DMEM media was added containing 25ug/mL cycloheximide, 5ug/mL puromycin or 0.1% DMSO for 4 hours. Cells were collected following the same formaldehyde crosslinking protocol detailed above.

^{35}S Labeling

After 3 hours and 15 minutes of translation inhibition treatment, cells were washed twice with starvation media (DMEM without methione, cysteine or glutamine; Gibco #21013024) and then incubated with starvation media plus translation inhibitors to starve cells of these amino acids. After 15 minutes of starvation media, EXPRESS S35 protein labeling mix (PerkinElmer) containing radioactive methionine and cysteine was added to cells at a concentration of 10uCi/mL media. Labeling was allowed to proceed for 30 minutes. Cells were washed with PBS and collected for analysis by SDS-PAGE. Radioactivity was detected with a phosphoscreen and total protein content was assessed by Coomassie (Simply Blue SafeStain, Invitrogen).

3.5 Supplemental Figures

Figure 3.6: Validation of PABPN and PABPC RIPs and Annotation Pipeline (A) Western blots showing that knockdown of PABPN or PABPC with siRNAs shows a subsequent reduction in the protein band that is targeted by the antibodies used for immunoprecipitations. Tubulin and actin are used as loading controls. (B) Western blots showing representative IPs pulling down PABPN or PABPC from total cell lysate. (C) Western blots showing matched IgG isotype control antibodies that were also used for pulldown. The RNA isolated from these RIPs was so minimal that it could not be prepared for sequencing, suggesting that non-specific RNAs were not binding significantly. (D) Schematic of part of the pipeline developed to analyze RNA sequencing experiments. Any portion of the annotation file that had two or more genes overlapping the exact same genomic space was removed to avoid false positives. (E) Schematic showing "readthrough_transcript" annotations and how they block any assignment of reads to the two genes that comprise that readthrough event. Readthrough_transcripts were removed from the annotation file. (F) Genome browser track read pileup for PABPN RIP showing reads that suggest failure to properly terminate transcription after the CENPM protein-coding gene, resulting in intergenic reads until reaching the neighboring gene, TNFRSF13C. (G) Principal Component Analysis (PCA) plot from Illumina RNA sequencing of total input, PABPC IP and PABPN IP, three independent biological replicates. (H) Venn diagrams showing genes considered significantly enriched in PABPN IP or PABPC IP compared to input with the overlap showing the genes that were enriched in both. Cut-offs used were baseMean > 50, padj <= 0.01 and $\log_2$FoldChange > 0.5. (I) Venn diagrams showing genes considered significantly depleted in PABPN IP or PABPC IP compared to input with the overlap showing the genes that were enriched in both. Cut-offs used were baseMean > 50, padj <= 0.01 and $\log_2$FoldChange < -0.5.


A
siLuc 200nM
siPABPN 200nM
siLuc 200nM
siPABPC 200nM
α-PABPN
α-PABPC
α-Tubulin
α-Actin
α-Actin
B
Input
Supernat
PABPN IP
α-PABPN
Input
Supernat
PABPC IP
α-PABPC
C
Input
Supernat
rabbit monoclonal, IgG IP
α-PABPC
α-Tubulin
α-Actin
Input
Supernat
rabbit polyclonal, IgG IP
α-PABPN
α-Tubulin
α-Actin
D
FASTA files
standard splice-aware mapping (STAR)
reads
Gene X
overlapping region
overlapping region removed from annotation
featureCounts
18 reads are assigned to Gene X
These 3 reads will be unassigned
E
reads
read-through transcript
all reads unable to be assigned due to overlap
remove read-through transcripts
reads
reads assigned to these two separate genes
F
PABPN IP, genome browser track
intergenic reads
CENPM
TNFRSF13C
G
Principal Component Analysis
PC2: 26% variance
PC1: 72% variance
Input
PABPN IP
PABPC IP
H
Significantly enriched
PABPN 1603
1113
PABPC 4590
I
Significantly depleted
412
PABPN 1848
PABPC 4390

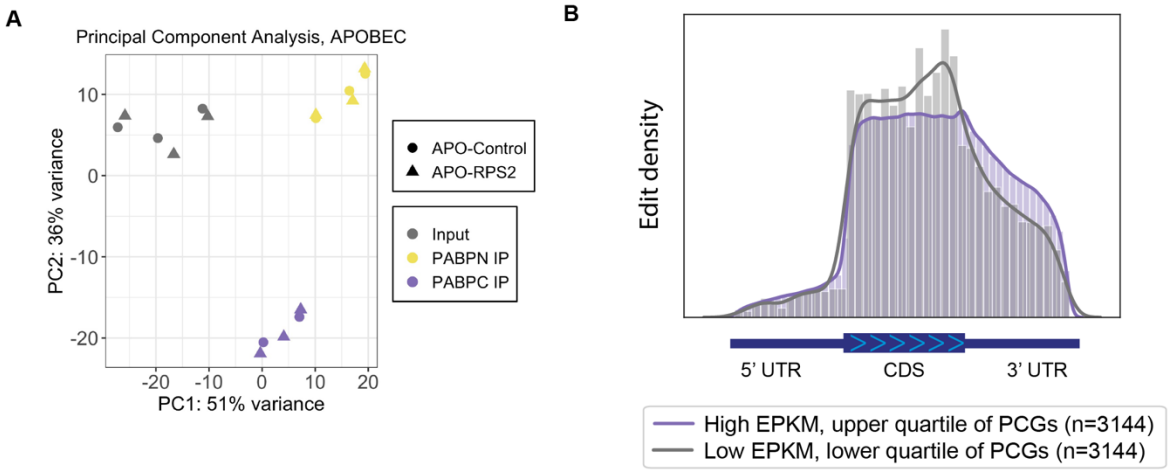


Figure 3.7: Ribosome contact with transcripts inferred by Ribo-STAMP. (A) Principal Component Analysis plot from Illumina RNA sequencing of APOBEC-Control and APOBEC-RPS2 cell lines. (B) Metagene plot showing edit (>=0.5 confidence score) distribution for transcripts of protein-coding genes that were in the upper quartile for editing (highly translated) compared to the lowest quartile for editing (lowly translated).

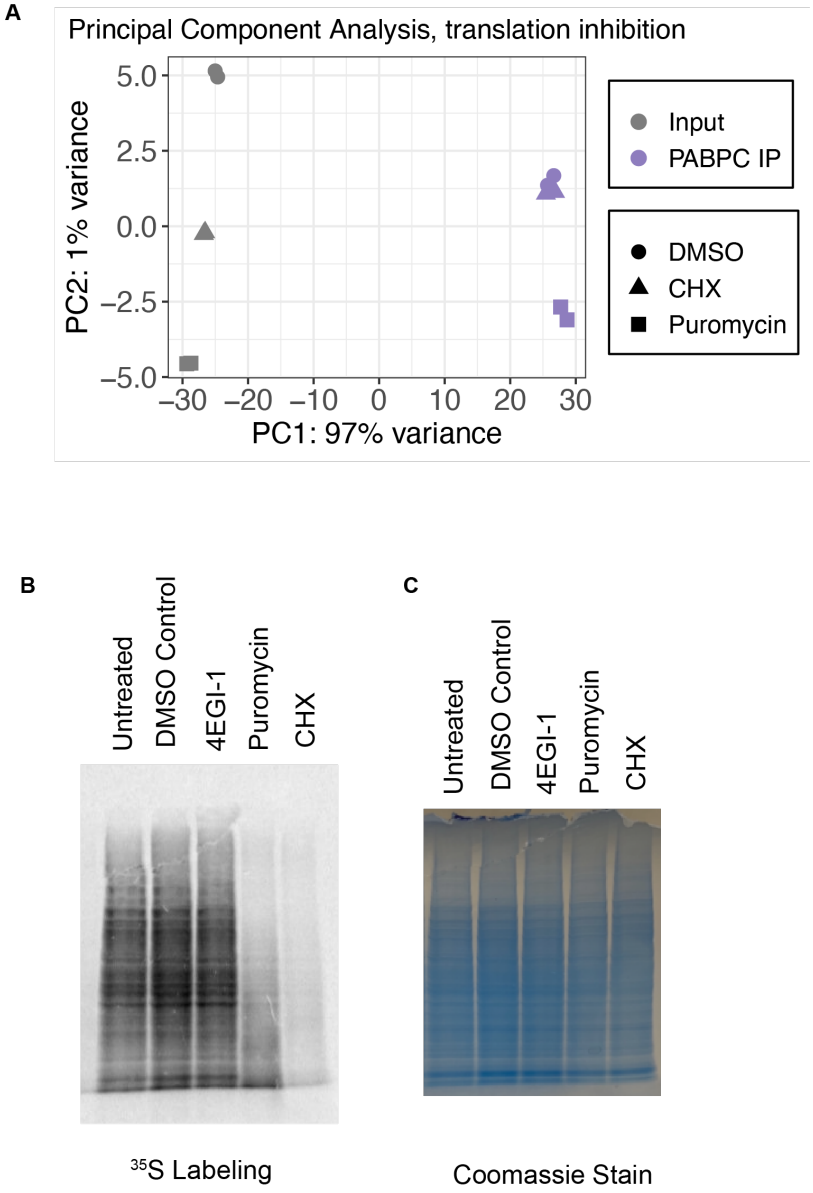


Figure 3.8: Validation of translation inhibition conditions. (A) Principal Component Analysis (PCA) plot from RNA sequencing of total input and PABPC IP from control DMSO, cycloheximide and puromycin experiments, three biological replicates. (B) Autoradiograph of ^{35}S-labeling of HEK293T cells during translation inhibition conditions after running on SDS-PAGE. 4EGI-1, puromycin, and cycloheximide (CHX) drugs were used for translation inhibition. Since puromycin and cycloheximide showed the greatest changes in protein level over the four-hour treatment time period, we moved forward with those two drugs for translation inhibition experiments. (C) Coomassie staining of the gel in (A), showing total protein levels present in each lane.

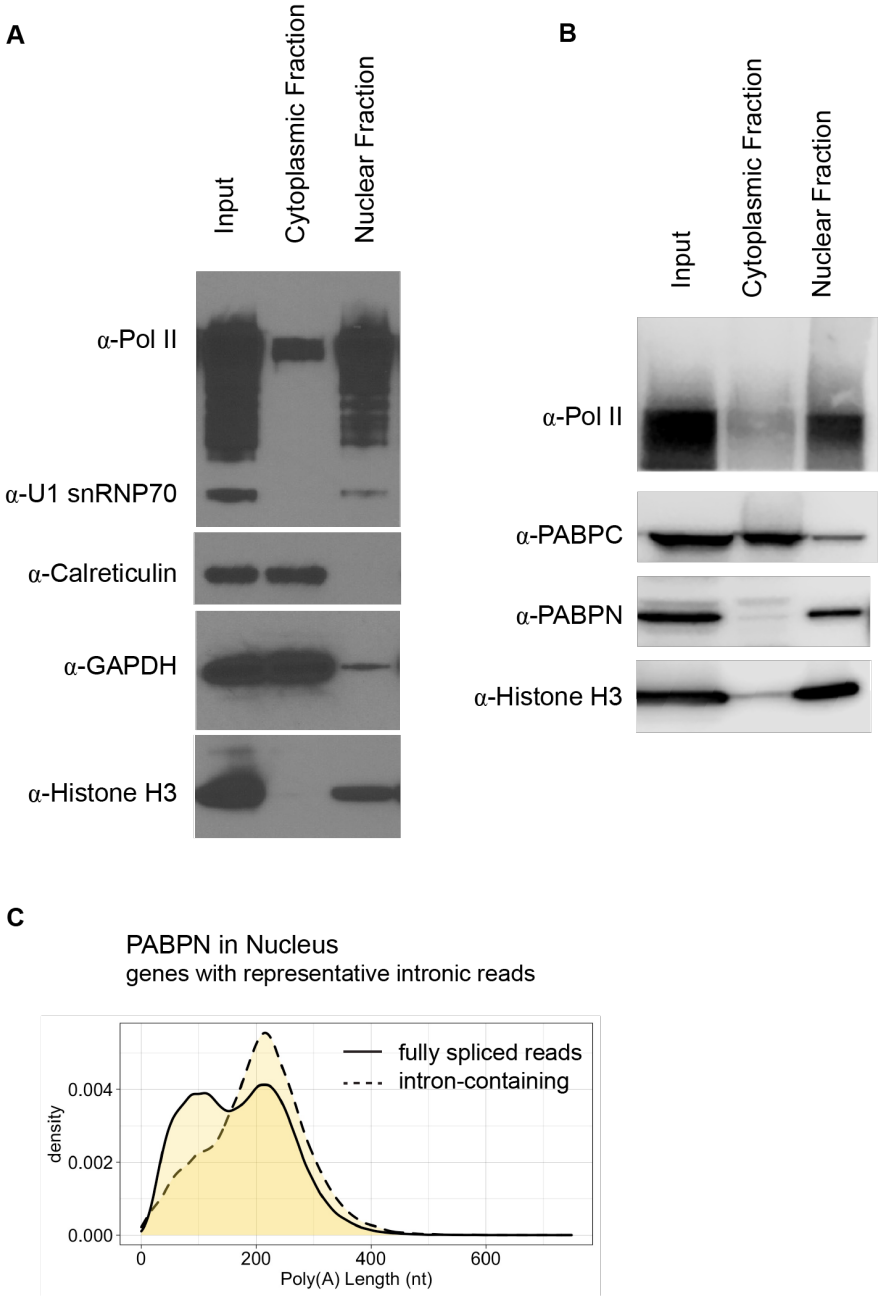


Figure 3.9 (A) Representative western blot of subcellular fractionation. (B) Western blot showing the nuclear and cytoplasmic distribution of PABPN and PABPC (C) Density plot of poly(A) tail length for reads that contained one or more introns (dashed line, n = 29,675) or no introns (solid line, n = 250,115) while with PABPN in the nucleus, only for genes that had an intron-containing representative.

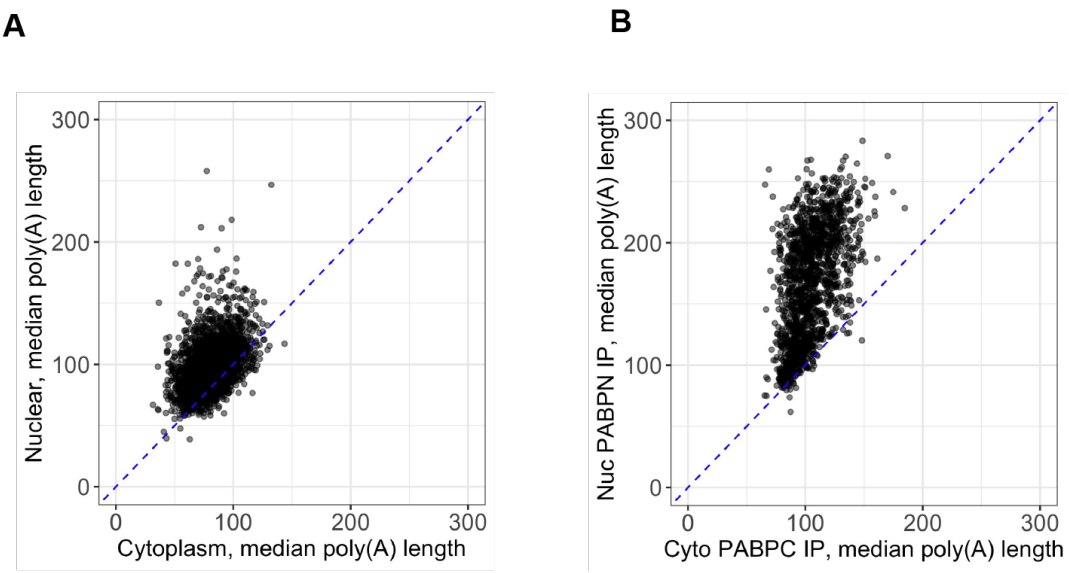


Figure 3.10: Poly(A) tails are longer on transcripts in the nucleus and with PABPN. (A) Median poly(A) tail length (per gene) in the cytoplasm compared to median poly(A) tail length in the nucleus. Genes must have been represented by at least 35 reads in each condition to be displayed. (B) Median poly(A) tail length (per gene) in PABPN IP from the nucleus compared to median poly(A) tail length in PABPC IP from the cytoplasm. Genes must have been represented by at least 35 reads in each condition to be displayed.

3.6 Acknowledgements

Chapter 3, in full, is a reprint of material as it is prepared for submission to *NSMB*, "Distinct Transcripts and Isoforms Favor the Nuclear and Cytoplasmic Poly(A) Binding Proteins (PABPs)," Nicholson-Shaw, A.L., Kofman, E., Yeo, G.W., and Pasquinelli A.E., 2021. I was the primary author.

Chapter 4

Development of Tools for Studying PABPs and poly(A) tails in *Caenorhabditis elegans*

4.1 Introduction

Poly(A) binding proteins and their cellular functions are largely conserved across eukaryotes. Most eukaryotes have multiple genes encoding PABPC and PABPN. *Caenorhabditis elegans* represent a relatively simple system with two genes for PABPC- the predominant one being *pab-1* and a lower expressed *pab-2*. They have a single PABPN- *pabp-2*. Our lab characterized the steady state poly(A) tail length of *C. elegans*, with the general principles found to be very similar to that of human cells, mouse, and yeast (Lima et al. 2017). Using *C. elegans* to study PABPs and poly(A) tails has the advantage of being able to answer questions in the context of an intact organism. If there are any developmental or tissue specific differences, these typically cannot be observed in a cell culture system. Investigating how an intact system behaves in regard to complex

cellular processes such as polyadenylation and translation provides unique *in vivo* insights.

4.2 CRISPR strains endogenously tagging *pabp-2*, *pab-1* and *pab-2*

4.2.1 Advantages of CRISPR

A continual challenge in the *C. elegans* field is the lack of commercially available antibodies to study proteins of interest. With the recent advances in CRISPR technology, strains can now be made to endogenously tag proteins of interest, forgoing the need for specific antibodies to each protein. This enables study of endogenous protein dynamics through Western blots and use of biochemical techniques such as immunoprecipitations to understand associated RNAs and proteins. This has great advantages over previous methods which relied on exogenously expressed plasmid constructs or transgenes generated by MosSCI insertion (Frøkjaer-Jensen et al. 2008). Exogenous expression may or may not recapitulate the circumstances of endogenous expression and results can be harder to interpret. Here I describe a set of strains that I generated to explore the role of cytoplasmic and nuclear poly(A) binding proteins in *C. elegans*. Although PAB-1 is considered the predominant PABPC in *C. elegans*, I set out to tag both of them as the literature is unclear as to the contributions of PAB-1 vs. PAB-2. By crossing the two PABPC strains that I created, this allows for questions to be asked of PABPC function overall without biasing towards only looking at the function of the more highly expressed PAB-1.

4.2.2 Strain Creation

Table 4.1: CRISPR Strains Created

Strain Name	Gene Tagged	Tag Added	'Silent' mutations made?
PQ599	*pabp-2* (PABPN)	GFP + 3xFLAG, N-terminal	Yes, 4
PQ606	*pab-1* (main PABPC)	mKate2 + 3x FLAG, N-terminal	Yes, 5
PQ611	*pab-2* (other PABPC)	mKate2 + 3x FLAG, N-terminal	Yes, 1

4.2.2.1 *pabp-2* tagged with GFP and 3xFLAG

To endogenously tag *pabp-2*, I followed a published protocol (Dickinson et al. 2015). To make the PQ599 strain (*pabp-2(ap434 [gfp::3xFLAG::pabp-2] I)*), young adult wildtype worms (N2) were injected with the following plasmids:

20ng/uL homologous repair template (pAN05)

50ng/uL Cas9-sgRNA construct (pAN02)

10ng/uL pGH8

5ng/uL pCFJ104

2.5ng/uL pCFJ90

All plasmids that were used for microinjection were purified using the Purelink HQ mini plasmid purification kit, including the optional guanidine HCl wash step. The homology arms were 616 and 622 bp. One side was able to be amplified from genomic

DNA and the other side was generated using IDT GeneBlocks, due to silent mutations that needed to be introduced to abolish the guide RNA cut site following recombination.

The Cas9-sgRNA construct was made by adding the guide RNA sequence into the pDD162 (Addgene #47549) construct using NEB's Q5 site-directed mutagenesis kit, following detailed online protocols written by Dan Dickinson, "Protocol for cloning SEC-based repair templates using Gibson assembly and ccdB negative selection" (wormcas9hr.weebly.com). The guide RNA sequence used for pabp-2 was: 5'CTCAACATTGAGGATATGAC3'. GuideScan was used to assist in selecting a guide RNA (http://guidescan.com).

Isolation of recombinants was performed as described in Dickinson et al. 2015 (Dickinson et al. 2015). Briefly, injected worms were grown at 25°C for 3 days. Plates were then flooded with 250ug/mL hygromycin and returned to 25°C for 3 days. Worms that were Hyg resistant, non-glowing rollers were singled to individual plates. L1 larvae from these worms were moved to new plates and heat shocked to remove the self-excising cassette. Worms that no longer showed a rol phenotype were PCR screened for integration of gfp::3xFLAG at the N-terminus of the pabp-2 locus. Insertion was additionally confirmed by Sanger sequencing. A successful integrant was backcrossed 4x to N2 to generate PQ599.

4.2.2.2 *pab-1* tagged with RFP and 3xFLAG

To make the PQ606 strain (*pab-1(ap436[mKate2/rfp::3xFLAG::pab-1] I)*), young adult wildtype worms (N2) were injected following methods described in Dokshin

et al. 2018, and included 5µg Cas9 protein, 2µg tracrRNA, 1.12µg crRNA, 800ng pRF4::rol-6 plasmid and 4µg of a dsDNA donor cocktail (Dokshin et al. 2018). Homology arms were 650bp and 635bp long. The dsDNA donor had longer homology arms than what is recommended in Dokshin et al (only 120bp) due to reagents that I already had in the lab from a plasmid I previously generated for another use. The guide RNA was selected through the help of GuideScan and had the following sequence: 5'GGCGCTGCTGCTCCTCAACC3'.

Cas9 protein, tracrRNA, and crRNA were ordered from IDT. Worms were grown at 20°C. 4-5 days later, F1 rollers were singled onto new plates. F2s were screened for mKate2 and then lysed and PCR screened for integration of mKate2::3xFLAG at the N-terminus of the pab-1 locus. Insertion was additionally confirmed by Sanger sequencing. A successful integrant was backcrossed 6x to N2 to generate PQ606.

4.2.2.3 *pab-2* tagged with RFP and 3xFLAG

To make the PQ611 strain (*pab-2(ap438[mKate2/rfp::3xFLAG::pab-2] X)*), young adult wildtype worms (N2) were injected following methods described in Dokshin et al. 2018, and included 5µg Cas9 protein, 2µg tracrRNA, 1.12µg crRNA, 250ng pRF4::rol-6 plasmid and 4µg of a dsDNA donor cocktail. Homology arms were 120 bp long. Cas9 protein, tracrRNA, and crRNA were ordered from IDT. Guide RNA sequence was: 5'GAGTCATTGGCGGTGGAACT3'. Worms were grown at 25°C. 3 days later, F1 rollers were singled onto new plates. F2s were screened for mKate2 and then lysed and PCR screened for integration of mKate2::3xFLAG at the N-terminus of the *pab-2*

locus. Insertion was confirmed by Sanger sequencing. A successful integrant was backcrossed 6x to N2 to generate PQ611.

4.2.2.4 Silent Mutations Introduced

To prevent the guide RNA from re-cutting once my homologous repair template was inserted, mutations had to be made. Mutation within the PAM sequence is ideal, but not always possible. For generation of PQ611 tagging *pab-2*, I was able to engineer a single mutation in the PAM site. If the mutation can be placed in the PAM site, one nucleotide is enough to prevent re-cutting, but in cases where this is not possible, it is recommended to introduce 4-5 mutations within the overall guide RNA sequence. PAM site mutation wasn't possible for the other two strains due to where the guide RNA was cutting within the ORF, therefore PQ599 and PQ606 contain mutations at 4 or 5 nucleotides within the guide RNA site. Specific nucleotide changes were selected so that the codon remained as close in optimality/usage as possible in *C. elegans* (Duret 2000). Mutations made were as follows, with changes indicated by a capital letter:

pabp-2 PQ599 sequence: (starting from start codon)

5'atgagcgataacgatattatcgacgatgatgttctTaacatCgaAgaC3'

pab-1 PQ606 sequence: (starting from start codon)

5'atggaaatgaacgtcgctgctcccgctgccgccgttgctggTgcCgctgcCccGcaG3'

pab-2 PQ611 sequence: (starting from start codon)

5’atggcT3’

4.2.3 Localization of PABPN and PABPC in *C. elegans*

The endogenous fluorescent tags added to these strains enable simple microscopy studies to look at localization of these proteins both in terms of cellular and tissue distribution. The following micrographs were taken using PQ599 (GFP tagged PABPN (*pabp-2*)) and PQ606 (RFP tagged PABPC (*pab-1*)). As expected, PQ599 animals reveal that PABPN is predominately localized to the nucleus and does not show cell-type specificity. PABPN is well expressed across all life stages imaged (Figure 4.1). PQ606 animals show cytoplasmic localization of PABPC and ubiquitous expression across all tissues, however there is greater expression in the gonad (Figure 4.2). *pab-1* depletion in worms has been shown to arrest germline development and previous immunostaining experiments also saw high levels of PAB-1 in the adult gonad and embryos (Ko, Kawasaki, and Shim 2013), thus this aligns with previously published data.

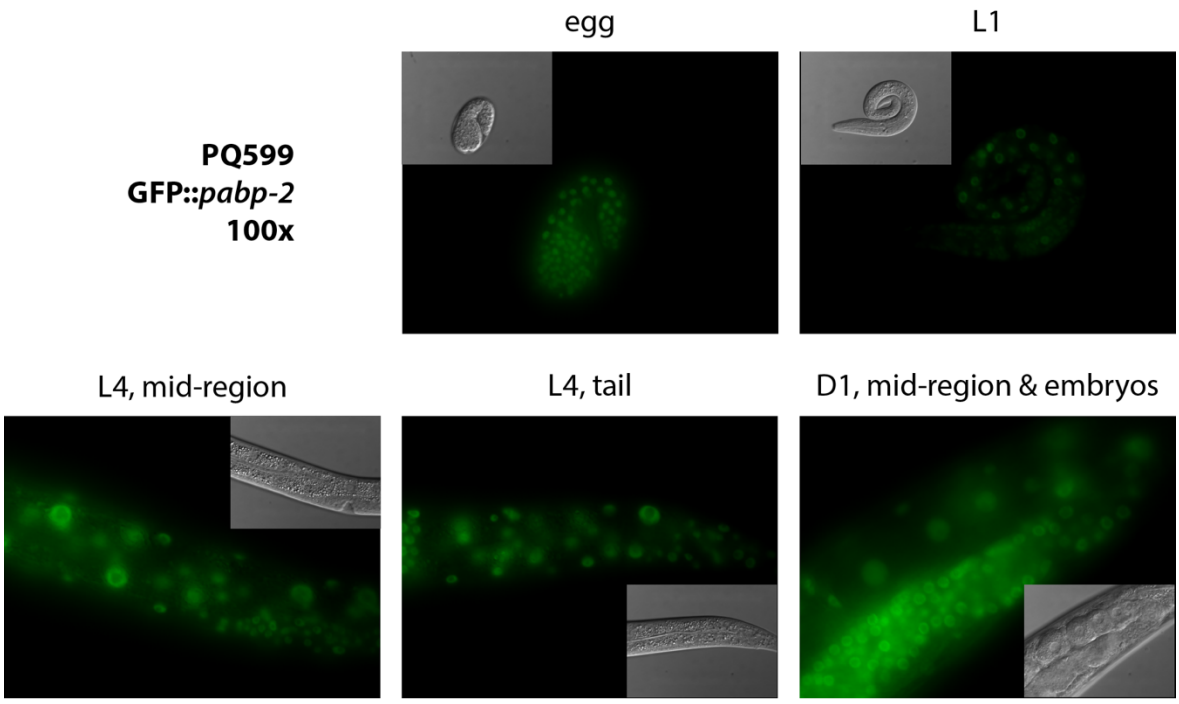


Figure 4.1: Fluorescent microscopy of PQ599 worms, GFP::*pabp-2* (PABPN).

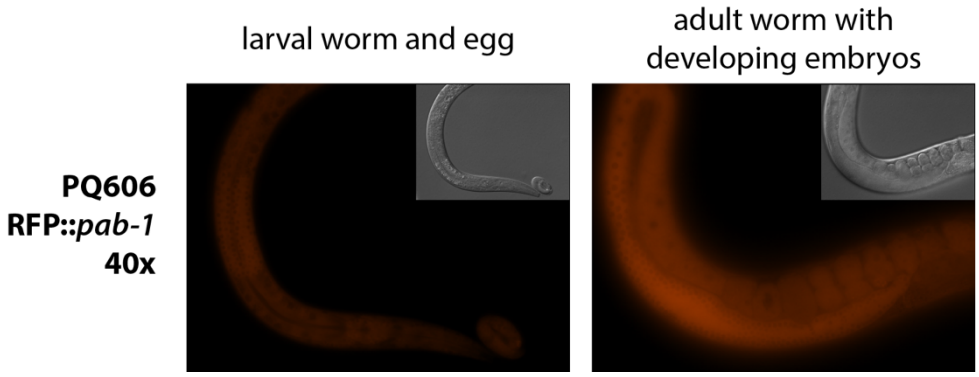


Figure 4.2: Fluorescent microscopy of PQ606 worms, RFP::*pab-1* (PABPC).

4.3 Polysome Profiling

4.3.1 Translation and Poly(A) Tails

PABPC binds to translation initiation factors and ribosome release factors, connecting it to both the 5' and 3' UTR (Imataka, Gradi, and Sonenberg 1998; Uchida et al. 2002; Hoshino, Imai, et al. 1999). Both PABPC and the poly(A) tail have been shown to stimulate translation, even apart from being directly connected to the translating RNA. Furthermore, work from our lab and others has suggested a connection between poly(A) tail size and translation. At this point, it is unclear whether the translation process directly changes poly(A) tail length, or whether poly(A) tail length regulates translation. The timing and consequences of poly(A) tail dynamics over the course of an mRNA's lifetime remain to be elucidated. Future studies delineating how these players are connected will be important to answer these outstanding questions.

Polysome profiling allows for a unique glimpse into the translation status of an RNA. By separating RNAs based on the number of associated ribosomes, levels of translation can be inferred. This differs from ribosome profiling, which looks at an isolated sequence that was bound by a ribosome. With polysome profiling, you can isolate the entire RNA and know whether it had one ribosome bound or six ribosomes bound, suggesting lesser or greater translation of that species. By purifying the entire RNA, there is potential to study numerous factors that contributed to that particular translational state, such as intron retention or even potentially the poly(A) tail length that existed on an RNA. Additionally, poly(A) binding proteins and their binding partners could be manipulated to determine if these influence translation of specific RNAs. However, many methods that could be applied here such as long-read sequencing or poly(A) tail sequencing require large amounts of starting material, beyond what is typically isolated from a polysome experiment. Here I present an adapted polysome

profiling method in *C. elegans* to produce large quantities of RNA from the pooled fractions.

4.3.2 Experimental Procedures

N2 worms were grown at 25°C for 29 hours, to the L4 stage. Worms were collected and lysed following Ding et al. 2009 (Ding and Grosshans 2009). Briefly, worms were washed three times with cold M9 supplemented with 1mM cycloheximide and once with lysis buffer (recipe below) without RNAseIn or PTE/DOC (polyoxyethylene-10-tridecylether/sodium deoxycholate monohydrate).Worms were spun in picofuge and buffer was aspirated off before snap freezing and storing at -80C until lysis.

Worms were lysed using a liquid nitrogen/mortar and pestle method. Lysis buffer consisted of 20mM Tris pH 8.3, 140mM KCl, 1.5mM MgCl2, 0.5% Nonidet P40, 2% PTE (polyoxyethylene-10-tridecylether), 1% DOC (sodium deoxycholate monohydrate), 1mM DTT, 1mM cycloheximide, and 0.4U/uL RNAseIn. Lysates were cleared by centrifugation, 15 minutes at 13,200 rpm at 4°C. Cleared lysates were stored in -80C until the next day.

Sucrose cushions were prepared on a Biocomp gradient master. Cellular transcripts were separated on sucrose cushions based on density (110,000 x g for 3 hr at 4°C) and were dispensed via fractionation as twenty-two individual aliquots (0.5mL/each) spanning the 10%–50% sucrose range. To each fraction, 1mL cold ethanol

was added and precipitation occurred overnight in -80°C. This precipitation step with ethanol was key to being able to extract large quantities of RNA from the fractions.

The next day, fractions were spun down for 30 minutes at 13,200 rpm at 4°C and the pellet was resuspended in 350uL LET + 1% SDS (25mM Tris pH 8, 100mM LiCl, 20mM EDTA, 1% SDS). This resuspension buffer gave cleaner RNA than any other method tested, as determined by Nanodrop absorbance spectra after completing isolation. Pipet and vortex to resuspend as thoroughly as possible. At this point, fractions were then pooled as needed by moving the resuspended material onto the next pellet and beginning resuspension again. I targeted five final categories: free RNA, scanning subunit, 80S, light polysomes and heavy polysomes (Figure 4.3). Once the pellets were resuspended, 1mL of Trizol LS reagent was added and the Zymo Direct-zol RNA Miniprep Kit was used for extraction, including on-column DNAse treatment if needed.

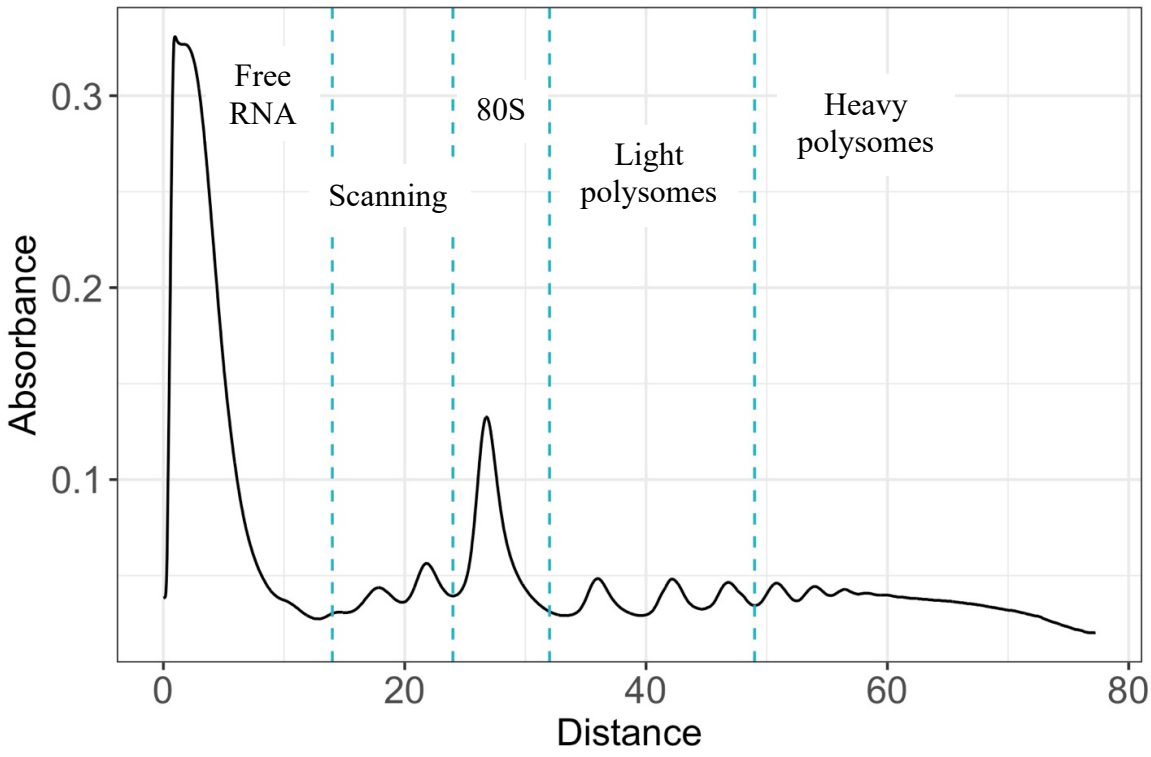


Figure 4.3: Representative absorbance spectrum of polysome profiling. Blue dashed lines indicate separation between which fractions were selected for each category representing distinct biological situations: free RNA, scanning subunit, 80S, light polysomes and heavy polysomes.

4.3.3 Downstream Experimental Considerations

I initially developed this polysome profiling protocol in order to use the RNA in TAIL-seq, which requires 5-20ug of starting RNA. Although the TAIL-seq libraries were successfully prepared, the resulting poly(A) tail profiles revealed uniformly short tails, clustering around 30nt for all samples sequenced. Although the RNA is separated into specific biologically relevant fractions, together this ultimately should comprise the majority of RNA in the cell, therefore it did not make sense that we would no longer detect any long tails, as we had previously seen many long tails with our total RNA

TAIL-seq results in *C. elegans* (Lima et al. 2017). This suggested that during the polysome profiling protocol itself, poly(A) tails were being deadenylated. One method to combat deadenylation occurring post-lysis is to add EDTA to all buffers. However, EDTA dissociates polysomes and therefore could not be used in this method.

The TAIL-seq protocol as well as the resulting analysis pipeline, tailseeker, are very challenging methods that I and others have troubleshot on many occasions, with varied success. I cannot rule out the possibility that TAIL-seq or tailseeker could have caused this shortened tail result, and perhaps it would not be seen with another poly(A) tail analysis method. However, in conversation with another lab, we learned that they also had difficulty with sequencing poly(A) tails from polysome fractions and saw many shortened tails. They also used TAIL-seq, but it is relatively unlikely that they encountered the same errors in a separate experiment by a different researcher. This suggests that testing and optimization may need to occur before employing a different poly(A) tail sequencing method while using polysome profiling.

4.4 Nanopore Direct RNA Sequencing

4.4.1 Unique Benefits of Nanopore Long-Read Technology

Next generation sequencing (NGS) technology revolutionized our understanding of the genome and transcriptome. Due to the high-throughput nature of sequencing by synthesis, researchers could gain vast amounts of genome-wide information which simply wasn't possible before with the previous Sanger sequencing methods (Bentley et

al. 2008). NGS relies on short reads, with typical sequencing runs reading around 50-150 bases per read fragment. After sequencing, complex algorithms piece together this information to understand what may have been the original DNA or RNA sequence that a read came from (Metzker 2010). While NGS has had immeasurable positive impacts on science and human health, there has been a recent trend towards developing long read sequencing technology (Reuter, Spacek, and Snyder 2015; van Dijk et al. 2018).

Long read sequencing has the distinct advantage of minimizing the guesswork associated with piecing small read fragments back together ("The Long View on Sequencing" 2018). In the context of RNA, the entire RNA can be sequenced, thus providing valuable information about the exact isoform that was present in the cell.

Oxford Nanopore Technologies (ONT) is one company that has developed a long-read sequencing platform. Their system relies on an array of small "nanopores" embedded in a membrane on a flow cell. An electrode is able to read the changing electric current as a molecule (such as DNA or RNA) passes through each nanopore. This current reading is decoded using basecalling algorithms to determine the nucleic acid sequence in real-time. In addition to being able to provide full-length isoform-specific information, Nanopore sequencing allows direct sequencing of the RNA itself, thus bypassing typical PCR steps involved in sequencing, which have the potential to introduce bias and duplicates (Garalde et al. 2018). Direct RNA sequencing can even give information about modifications on an RNA, such as m^6A. Finally, and most pertinent to my dissertation, direct RNA sequencing provides a way to infer the length of the poly(A) tail (Workman et al. 2019). Because the entire original RNA is sequenced, this includes the poly(A) tail itself. Although homopolymeric sequences are still difficult

to read even with long-read sequencing, you can infer the length of the sequence based on the dwell-time within the nanopore. The motor protein that pulls the nucleic acid through the nanopore does so at a constant rate. Therefore, timing is directly related to sequence length, and thus the time it took to sequence the poly(A) tail will equate to the length of the tail.

Nanopore sequencing, and in particular direct RNA sequencing, lacks the depth that is available with NGS. Therefore, it is often used as a way to obtain complementary information that would not be available solely with NGS. Nanopore sequencing can answer many questions that simply could not be asked with NGS experiments.

In order to open up new avenues for these types of questions, I pioneered the use of the Nanopore in our lab. The minION sequencer from Nanopore is only a few inches long and directly connects to a computer via USB, therefore it is quite simple to set up this small machine and use in a single lab. As long-read sequencing becomes more common, basecalling and other downstream analysis becomes more widespread, with significant improvements happening even in the last few years since I first started using the Nanopore. The coming years will likely see increases in accuracy and read depth as well.

4.4.2 *C. elegans* Poly(A) Tail Length as Determined by Nanopore Sequencing

Poly(A) tails as determined by Nanopore sequencing are strikingly similar to our previous studies using Illumina sequencing (Lima et al. 2017). Using three replicates of

larval stage L4 worms, we performed Direct RNA sequencing on the Nanopore. The Nanopolish-polya software (https://github.com/jts/nanopolish) was used to estimate poly(A) tail length. We removed any genes that were not nuclear-encoded, as these have a different polyadenylation system. The overall distribution of all reads shows the same peak at 34 nt for both methods, and an abundance of shorter tails (Figure 4.4).

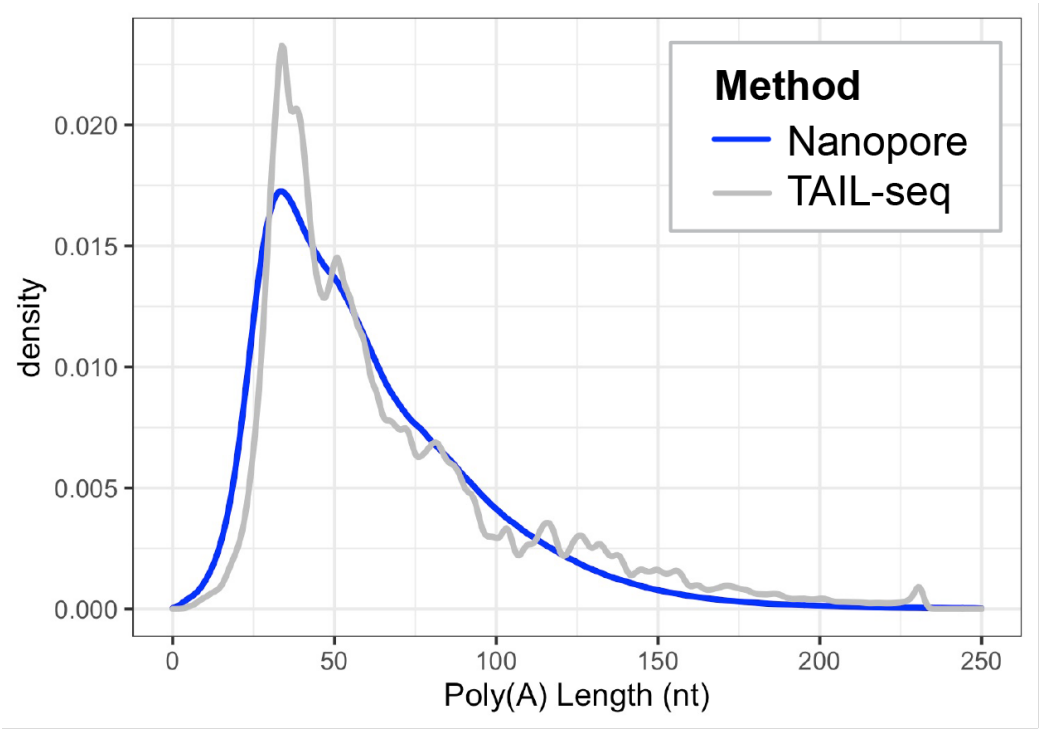


Figure 4.4: Total Poly(A) Tail Distribution for Nanopore and TAIL-seq. Density plots showing all reads as determined by Nanopore (blue line) or TAIL-seq (grey line).

Additionally, Nanopore validates the conclusion that transcripts with short tails are highly abundant, well-translated genes. Using violin plots to look at the same genes that were published in our TAIL-seq paper in *C. elegans*, similar profiles are shown (Figure 4.5). The most highly abundant, efficiently translated genes, such as ribosomal

protein genes, have discrete poly(A) tail lengths that are much shorter than other less abundant species.

4.4.3 Acknowledgements

Chapter 4, section 4.4, contains unpublished material coauthored with Ian Nicastro and Amy Pasquinelli. I was the primary author of this material.

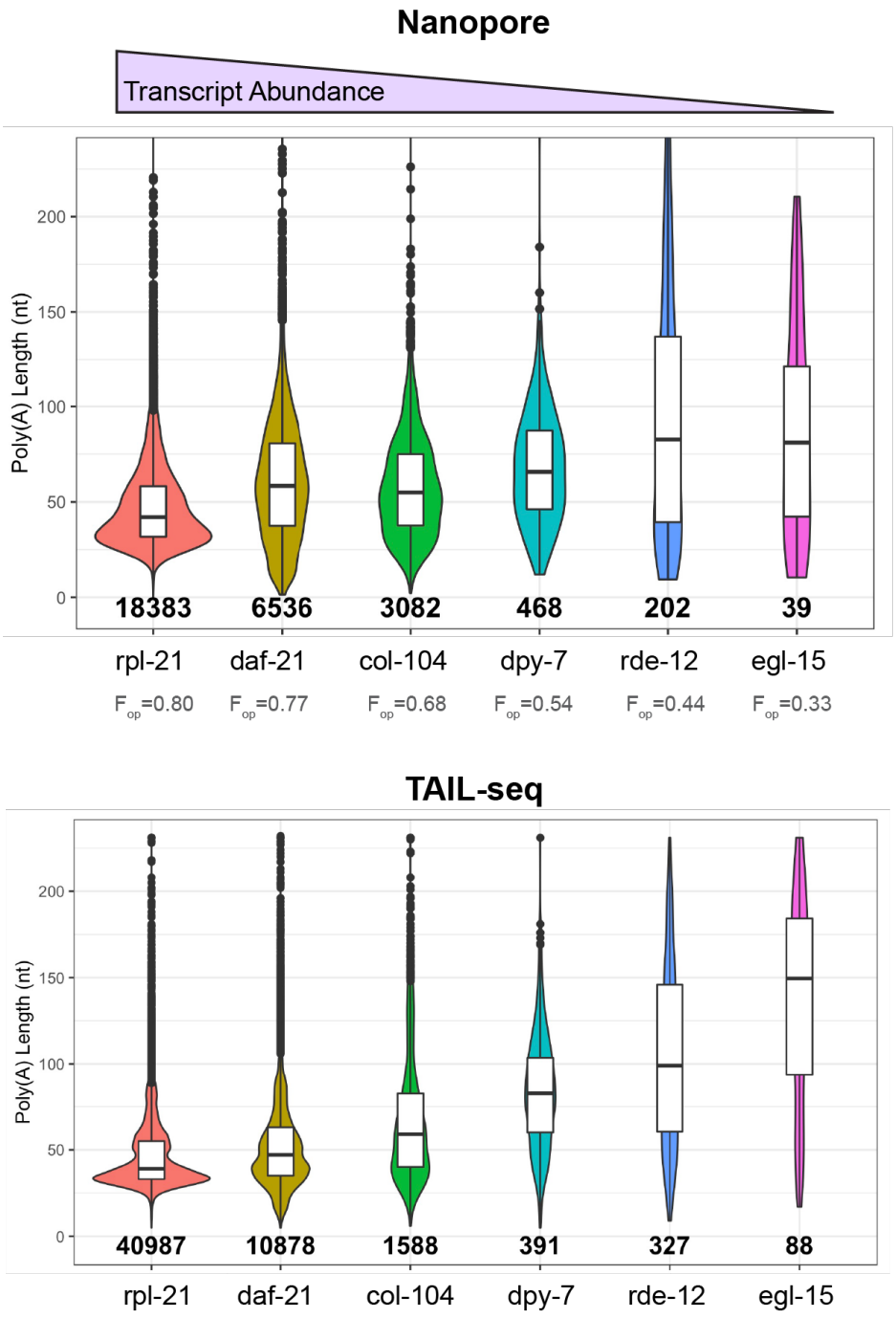


Figure 4.5: Highly expressed mRNAs have short poly(A) tails. Violin distribution plots with inlaid box plots of all tail length measurements for genes with different frequencies of optimal codons (Fop) and abundance levels. The top graph shows poly(A) tail length as determined by Nanopore and the bottom as determined by TAIL-seq (Lima et al. 2017). Number of reads for each gene are displayed at the base of the violin plot.

Chapter 5

Conclusions

5.1 PABPN and PABPC's involvement throughout the lifetime of an mRNA

Since the discovery of the poly(A) tail, a myriad of roles have been connected to this 3' addition, from nuclear export to RNA stability and translation. Many of these important functions are mediated through the nuclear and cytoplasmic poly(A) binding proteins, PABPN and PABPC. Several decades of accumulating evidence has brought us to the model we have today that follows the trajectory of an RNA's life cycle. PABPN is thought to facilitate the creation of all poly(A) tails and accompany an RNA throughout its time in the nucleus (Bienroth, Keller, and Wahle 1993; Wahle 1991). Since the vast majority of Pol-II-transcribed RNAs receive a poly(A) tail, PABPN is presumed to interact with nearly all RNAs. After an RNA reaches the cytoplasm, PABPC takes over the role of poly(A) binding partner and mediates interactions with translation factors and even deadenylases (Uchida et al. 2002; Jacobson and Favreau 1983; Hoshino, Imai, et al. 1999; Hoshino, Hosoda, et al. 1999; Tarun et al. 1997; Wigington et al. 2014). PABPC remains with the RNA for the duration of its time in the cytoplasm, up until the tail no longer has

enough adenosines to allow for binding (Yi et al. 2018; Goldstrohm and Wickens 2008; Webster et al. 2018). This brings the RNA to the end of its life cycle, where decapping rapidly follows full deadenylation, and the RNA can be degraded. An overwhelming amount of research has enabled us to have this much knowledge into a nuanced and multi-step process.

Considering the bulk of evidence that has brought us to this model, these ideas likely represent the prevailing state for a typical protein-coding gene. However, there have been relatively few genome-wide studies looking at the details of this model. Does this sequential timeline happen for all genes that are polyadenylated? Are there additional check points that are facilitated by PABPN and PABPC for a mature mRNA? Although PABPC's physical connection to translation is beneficial for protein-coding genes, how do these interactions affect non-coding genes that are bound by PABPC? As with nearly all complex biological processes, there remain outstanding questions that would inform this model, filling in the missing details.

5.2 Further studies

Although PABPN and PABPC predominantly exist in the nucleus and cytoplasm, respectively, they are also shuttling proteins. Scenarios such as viral infection, stress, and UV irradiation have been shown to change the localization of PABPC to be more nuclear (Kumar and Glaunsinger 2010; Burgess et al. 2011; Y. J. Lee and Glaunsinger 2009). Experiments blocking nuclear export resulted in accumulation of PABPC in the nucleus, suggesting that PABPC export and RNA export are linked (Burgess et al. 2011). PABPC

and poly(A) polymerase co-immunoprecipitate, suggesting that PABPC may be involved close to the initial creation of the tail for some RNAs (Hosoda, Lejeune, and Maquat 2006). It is currently unclear if this is a functional interaction that affects polyadenylation or not. Regarding PABPN, our Ribo-STAMP experiments suggest that translation is occurring while protein-coding transcripts are still associated with PABPN in the cytoplasm (Figure 3.2B). Other studies have found PABPN and PABPC co-existing on a single RNA (Hosoda, Lejeune, and Maquat 2006). These findings all point to a specific role for PABPN in the cytoplasm and PABPC in the nucleus. However, little is known about PABPN and PABPC's role in the opposite compartment under normal, stress-free conditions.

Considering the prevailing model where PABPN binds to the poly(A) tail when it is first made and that PABPC binds to the RNA later in when it is fully matured and exported, this raises the question of how PABPN and PABPC can play a role in their opposing compartments. One scenario may be that PABPN only remains bound to a select few RNAs throughout their life in the cytoplasm. In this case, high-throughput sequencing and reporter studies could elucidate the characteristics of these RNAs which encourage continued PABPN binding. Another scenario may be that PABPN travels with transcripts as they are exported out into the cytoplasm and a specific cytoplasmic process, such as translation, causes PABPN to be removed.

In all current versions of our poly(A) tail model, the transition from PABPN to PABPC is assumed to occur for the vast majority of RNAs. However, how this transition occurs, where in the cell, and what causes it are still yet to be determined. Translation has emerged as a possible player, and even if this is the case, nothing is known about the exact

mechanism of how this contributes to the transition (Hosoda, Lejeune, and Maquat 2006; Lima et al. 2017). Further experiments elucidating how this process occurs will give us insight into where this process might go wrong. Questions regarding this transition will also help researchers gain insight into when and why PABPs exist in their opposite compartment. RNAs are rarely not covered by proteins (Singh et al. 2015), so a scenario where PABPN comes off before PABPC is readily available to bind seems unlikely. This implies that PABPN and PABPC may need to be in the same compartment to complete this trade-off, at least temporarily, before shuttling back.

Translation has been connected to poly(A) tails through the discovery that well-translated substrates have shorter poly(A) tails that can accommodate 1-2 PABPCs in somatic cells (Subtelny et al. 2014; Lima et al. 2017). This was a rather surprising finding, challenging the long-held paradigm where long poly(A) tails promote stability and translation (Goldstrohm and Wickens 2008; Weill et al. 2012; Jalkanen, Coleman, and Wilusz 2014). A defined mechanistic connection between poly(A) tails and translation efficiency has not been demonstrated before; thus far the association has only been investigated through trends found in large data sets. It remains to be determined whether translation itself facilitates tail shortening, or whether tail shortening subsequently improves translation efficiency. Reporter studies blocking translation and assaying tail length could provide insight into the mechanism of this process. Defining this connection will be key to understanding the connection between translation and poly(A) tail length.

Appendix: Grad Slam Script

In 2019, I competed in UCSD's Grad Slam, a TED-style speaking competition aimed at sharing your graduate research with a non-specialist audience. Through several rounds, UCSD picks one "campus champion" to represent them at the state-wide finals in San Francisco. I was selected as the winner in 2019, and here I have appended my script. The video can be accessed at: https://www.youtube.com/watch?v=TIn2M5r8c6E

RNA Regulation: The Tail Wagging the Dogma

Inside each one of our cells, millions of particles are constantly bustling with activity. Nearly all of this activity is organized by a molecule called RNA, shown here as yellow dots. Each piece of RNA is an individual written message that your genome uses to transmit information out into your working cell. This message system directs everything from production of muscle fibers to protecting you from viruses.

When an RNA first gets made, your cell synthesizes just the main message portion. After that, a "tail" gets added to each RNA at a predetermined long length. Because of the protective role that the tail plays, long has traditionally been viewed as better and it was assumed that this was the end of the story. So it got written down in a textbook in this way (RNAs have long tails) and sometimes once something is written down in a textbook, it

doesn't get questioned anymore. But sometimes, you find something totally unexpected in your research.

My lab has recently found that while RNAs might start out like this, they wind up with a myriad of tail sizes, and that short tails are actually the most common. But why would a cell bother making the tail long to begin with, just to shorten it later on? By following these short-tailed RNAs, I've discovered that they actually play a very specific role in signaling to the cell that this particular message should be communicated quickly and effectively. By allowing for long tails to be on some RNAs and short tails on others, tail length can now be a key part of differentiating each message, being specific to what the cell needs. This finding turns the textbook dogma on its head.

Now, this is a very small molecule and I'm talking about even smaller changes in the tail region, but these findings implicate the tail as a vital part of the message. For an example of how important this is, I want to talk about Fragile X Syndrome, which is a major cause of intellectual disability. Recently it was discovered that patients with Fragile X have RNAs with different tail lengths than those of a non-effected person. As you'll remember, the prevailing view has been that tails are long and don't have any importance for how the message is communicated, so how could they possibly be involved in disease? Based on my work, we now know that tail length can have drastic effects on development, so working to understand how tails are mis-regulated in Fragile X may provide us with new treatments for this disease. This seemingly tiny change on a tiny molecule can affect

how the entire human body is functioning. This is just one example where my discoveries could give us a new lens to understand complicated diseases.

On top of the scientific implications, this work speaks to the importance of re-evaluating long held beliefs in any field. Scientists once thought that the tail of an RNA was a passive bystander in regards to the message being delivered. Turns out it may be more of a case of the tail wagging the dog. Staying open and curious in our research may lead to some big breakthroughs in health in the coming decade.

References

Afonina, E, R Stauber, and G N Pavlakis. 1998. “The Human Poly(A)-Binding Protein 1 Shuttles between the Nucleus and the Cytoplasm.” *The Journal of Biological Chemistry* 273 (21): 13015–21. https://doi.org/10.1074/jbc.273.21.13015.

Ameur, Adam, Ammar Zaghlool, Jonatan Halvardson, Anna Wetterbom, Ulf Gyllensten, Lucia Cavelier, and Lars Feuk. 2011. “Total RNA Sequencing Reveals Nascent Transcription and Widespread Co-Transcriptional Splicing in the Human Brain.” *Nature Structural & Molecular Biology* 18 (12): 1435–40. https://doi.org/10.1038/nsmb.2143.

Archer, Stuart K, Nikolay E Shirokikh, Claus V Hallwirth, Traude H Beilharz, and Thomas Preiss. 2015. “Probing the Closed-Loop Model of MRNA Translation in Living Cells.” *RNA Biology* 12 (3). https://doi.org/10.1080/15476286.2015.1017242.

Baer, Bradford W, Roger D Kornberg, and | Supp. 1983. “The Protein Responsible for the Repeating Strcuture of Cytoplasmic Poly(A)-Ribonucleoprotein.” *Journal of Cell Biology* 96: 717–21. https://doi.org/10.1083/jcb.96.3.717.

Bartel, David P, and Kehui Xiang. 2021. “The Molecular Basis of Coupling between Poly(A)-Tail Length and Translational Efficiency.” *ELife*, 2021.01.18.427055. http://biorxiv.org/content/early/2021/01/19/2021.01.18.427055.abstract.

Bazzini, Ariel A, Florencia del Viso, Miguel A Moreno-Mateos, Timothy G Johnstone, Charles E Vejnar, Yidan Qin, Jun Yao, Mustafa K Khokha, and Antonio J Giraldez. 2016. “Codon Identity Regulates MRNA Stability and Translation Efficiency during the Maternal-to-zygotic Transition.” *The EMBO Journal* 35 (19): 2087–2103. https://doi.org/10.15252/embj.201694699.

Beaulieu, Yves B., Claudia L. Kleinman, Anne Marie Landry-Voyer, Jacek Majewski, and François Bachand. 2012. “Polyadenylation-Dependent Control of Long Noncoding RNA Expression by the Poly(A)-Binding Protein Nuclear 1.” *PLoS Genetics* 8 (11). https://doi.org/10.1371/journal.pgen.1003078.

Bentley, David R, Shankar Balasubramanian, Harold P Swerdlow, Geoffrey P Smith, John Milton, Clive G Brown, Kevin P Hall, et al. 2008. “Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry.” *Nature* 456 (7218): 53–59. https://doi.org/10.1038/nature07517.

Bernstein, P, S W Peltz, and J Ross. 1989. “The Poly(A)-Poly(A)-Binding Protein Complex Is a Major Determinant of MRNA Stability in Vitro.” *Molecular and Cellular Biology* 9 (2): 659–70. http://www.ncbi.nlm.nih.gov/pubmed/2565532.

Bienroth, Silke, Walter Keller, and Elmar Wahle. 1993. “Assembly of a Processive Messenger RNA Polyadenylation Complex.” *The EMBO Journal* 12 (2): 585–94. http://www.ncbi.nlm.nih.gov/pubmed/8440247.

Blanco, P, C A Sargent, C A Boucher, G Howell, M Ross, and N A Affara. 2001. “A Novel Poly(A)-Binding Protein Gene (PABPC5) Maps to an X-Specific Subinterval in the Xq21.3/Yp11.2 Homology Block of the Human Sex Chromosomes.” *Genomics* 74 (1): 1–11. https://doi.org/10.1006/geno.2001.6530.

Blobel, Günter. 1972. “Protein Tightly Bound to Globin MRNA.” *Biochemical and Biophysical Research Communications* 47 (1): 88–95. https://doi.org/10.1016/S0006-291X(72)80014-7.

Borman, A M, Y M Michel, and K M Kean. 2000. “Biochemical Characterisation of Cap-Poly(A) Synergy in Rabbit Reticulocyte Lysates: The EIF4G-PABP Interaction Increases the Functional Affinity of EIF4E for the Capped MRNA 5’-End.” *Nucleic Acids Research* 28 (21): 4068–75. https://doi.org/10.1093/nar/28.21.4068.

Brannan, Kristopher W., Isaac A. Chaim, Ryan J. Marina, Brian A. Yee, Eric R. Kofman, Daniel A. Lorenz, Pratibha Jagannatha, et al. 2021. “Robust Single-Cell Discovery of RNA Targets of RNA-Binding Proteins and Ribosomes.” *Nature Methods* 18 (5): 507–19. https://doi.org/10.1038/s41592-021-01128-0.

Brawerman, George. 1981. “The Role of the Poly(a) Sequence in Mammalian Messenger RNA.” *Critical Reviews in Biochemistry and Molecular Biology* 10 (1): 1–38. https://doi.org/10.3109/10409238109114634.

Brown, C E, and A B Sachs. 1998. “Poly(A) Tail Length Control in Saccharomyces Cerevisiae Occurs by Message-Specific Deadenylation.” *Molecular and Cellular Biology* 18 (11): 6548–59. http://www.ncbi.nlm.nih.gov/pubmed/9774670.

Burgess, H. M., W. A. Richardson, R. C. Anderson, C. Salaun, S. V. Graham, and N. K. Gray. 2011. “Nuclear Relocalisation of Cytoplasmic Poly(A)-Binding Proteins PABP1 and PABP4 in Response to UV Irradiation Reveals MRNA-Dependent Export of Metazoan PABPs.” *Journal of Cell Science* 124 (19): 3344–55. https://doi.org/10.1242/jcs.087692.

Cai, Xuezhong, Curt H. Hagedorn, and Bryan R. Cullen. 2004. “Human MicroRNAs Are Processed from Capped, Polyadenylated Transcripts That Can Also Function as MRNAs.” *RNA* 10 (12): 1957–66. https://doi.org/10.1261/rna.7135204.

Calado, Angelo, and Maria Carmo-Fonseca. 2000. “Localization of Poly(A)-Binding Protein 2 (PABP2) in Nuclear Speckles Is Independent of Import into the Nucleus and Requires Binding to Poly(A) RNA.” *Journal of Cell Science* 113 (12): 2309–18. https://jcs.biologists.org/content/113/12/2309.long.

Chang, Hyeshik, Jaechul Lim, Minju Ha, and V. Narry Kim. 2014. “TAIL-Seq: Genome-Wide Determination of Poly(A) Tail Length and 3’ End Modifications.” *Molecular Cell* 53 (6): 1044–52. https://doi.org/10.1016/j.molcel.2014.02.007.

Choi, Youkyung Hwang, and Curt H Hagedorn. 2003. “Purifying MRNAs with a High-Affinity EIF4E Mutant Identifies the Short 3’ Poly(A) End Phenotype.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (12): 7033–38. https://doi.org/10.1073/pnas.1232347100.

Collart, Martine A. 2016. “The Ccr4-Not Complex Is a Key Regulator of Eukaryotic Gene Expression.” *Wiley Interdisciplinary Reviews. RNA* 7 (4): 438–54. https://doi.org/10.1002/wrna.1332.

Darnell, J E, R Wall, and R J Tushinski. 1971. “An Adenylic Acid-Rich Sequence in Messenger RNA of HeLa Cells and Its Possible Relationship to Reiterated Sites in DNA.” *Proceedings of the National Academy of Sciences of the United States of America* 68 (6): 1321–25. https://doi.org/10.1073/pnas.68.6.1321.

Deo, R C, J B Bonanno, N Sonenberg, and S K Burley. 1999. “Recognition of Polyadenylate RNA by the Poly(A)-Binding Protein.” *Cell* 98 (6): 835–45. https://doi.org/10.1016/s0092-8674(00)81517-2.

Dickinson, Daniel J, Ariel M Pani, Jennifer K Heppert, Christopher D Higgins, and Bob Goldstein. 2015. “Streamlined Genome Engineering with a Self-Excising Drug Selection Cassette.” *Genetics* 200 (4): 1035–49. https://doi.org/10.1534/genetics.115.178335.

Dijk, Erwin L. van, Yan Jaszczyszyn, Delphine Naquin, and Claude Thermes. 2018. “The Third Revolution in Sequencing Technology.” *Trends in Genetics* 34 (9): 666–81. https://doi.org/10.1016/J.TIG.2018.05.008.

Ding, Xavier C, and Helge Grosshans. 2009. “Repression of C. Elegans MicroRNA Targets at the Initiation Level of Translation Requires GW182 Proteins.” *The EMBO Journal* 28 (3): 213–22. https://doi.org/10.1038/emboj.2008.275.

Dokshin, Gregoriy A., Krishna S. Ghanta, Katherine M. Piscopo, and Craig C. Mello. 2018. “Robust Genome Editing with Short Single-Stranded” 210 (November): 781–87.

Dostie, Josée, and Gideon Dreyfuss. 2002. “Translation Is Required to Remove Y14 from MRNAs in the Cytoplasm.” *Current Biology* 12 (13): 1060–67. https://doi.org/10.1016/S0960-9822(02)00902-8.

Duret, Laurent. 2000. “TRNA Gene Number and Codon Usage in the C. Elegans Genome

Are Co-Adapted for Optimal Translation of Highly Expressed Genes." *Trends in Genetics* 16 (7): 287–89. https://doi.org/10.1016/S0168-9525(00)02041-2.

Edmonds, M, M H Vaughan, and H Nakazato. 1971. "Polyadenylic Acid Sequences in the Heterogeneous Nuclear RNA and Rapidly-Labeled Polyribosomal RNA of HeLa Cells: Possible Evidence for a Precursor Relationship." *Proceedings of the National Academy of Sciences of the United States of America* 68 (6): 1336–40. https://doi.org/10.1073/pnas.68.6.1336.

Eichhorn, Stephen W., Alexander O. Subtelny, Iva Kronja, Jamie C. Kwasnieski, Terry L. Orr-Weaver, and David P. Bartel. 2016. "MRNA Poly(A)-Tail Changes Specified by Deadenylation Broadly Reshape Translation in Drosophila Oocytes and Early Embryos." *ELife* 5: 1–24. https://doi.org/10.7554/eLife.16955.

Eisen, Timothy J, Stephen W Eichhorn, Alexander O Subtelny, Kathy S Lin, Sean E McGeary, Sumeet Gupta, and David P. Bartel. 2020. "The Dynamics of Cytoplasmic MRNA Metabolism." *Molecular Cell*, 1–14. https://doi.org/10.1101/763599.

Eisenberg, Eli, and Erez Y Levanon. 2003. "Human Housekeeping Genes Are Compact." *Trends in Genetics : TIG* 19 (7): 362–65. https://doi.org/10.1016/S0168-9525(03)00140-9.

Feral, C., G. Guellaën, and A. Pawlak. 2001. "Human Testis Expresses a Specific Poly(A)-Binding Protein." *Nucleic Acids Research* 29 (9): 1872–83. https://doi.org/10.1093/nar/29.9.1872.

Frankish, Adam, Mark Diekhans, Anne-Maud Ferreira, Rory Johnson, Irwin Jungreis, Jane Loveland, Jonathan M Mudge, et al. 2019. "GENCODE Reference Annotation for the Human and Mouse Genomes." *Nucleic Acids Research* 47 (D1): D766–73. https://doi.org/10.1093/nar/gky955.

Frøkjaer-Jensen, Christian, M Wayne Davis, Christopher E Hopkins, Blake J Newman, Jason M Thummel, Søren-Peter Olesen, Morten Grunnet, and Erik M Jorgensen. 2008. "Single-Copy Insertion of Transgenes in Caenorhabditis Elegans." *Nature Genetics* 40 (11): 1375–83. https://doi.org/10.1038/ng.248.

Fuke, Hiroyuki, and Mutsuhito Ohno. 2008. "Role of Poly (A) Tail as an Identity Element for MRNA Nuclear Export." *Nucleic Acids Research* 36 (3): 1037–49. https://doi.org/10.1093/nar/gkm1120.

Funakoshi, Yuji, Yusuke Doi, Nao Hosoda, Naoyuki Uchida, Masanori Osawa, Ichio Shimada, Masafumi Tsujimoto, Tsutomu Suzuki, Toshiaki Katada, and Shin-ichi Hoshino. 2007. "Mechanism of MRNA Deadenylation: Evidence for a Molecular Interplay between Translation Termination Factor ERF3 and MRNA Deadenylases." *Genes & Development* 21 (23): 3135–48. https://doi.org/10.1101/gad.1597707.

Gagnon, Keith T, Liande Li, Bethany A Janowski, and David R Corey. 2014. “Analysis of Nuclear RNA Interference in Human Cells by Subcellular Fractionation and Argonaute Loading.” *Nature Protocols* 9 (9): 2045–60. https://doi.org/10.1038/nprot.2014.135.

Gallie, D R. 1991. “The Cap and Poly(A) Tail Function Synergistically to Regulate MRNA Translational Efficiency.” *Genes & Development* 5 (11): 2108–16. https://doi.org/10.1101/gad.5.11.2108.

Garalde, Daniel R, Elizabeth A Snell, Daniel Jachimowicz, Botond Sipos, Joseph H Lloyd, Mark Bruce, Nadia Pantic, et al. 2018. “Highly Parallel Direct RNA Sequencing on an Array of Nanopores.” *Nature Methods* 15 (3): 201–6. https://doi.org/10.1038/nmeth.4577.

Gilbert, WV., and M. Thompson. 2016. “MRNA Length-Sensing in Eukaryotic Translation: Reconsidering the ‘Closed Loop’ and Its Implications for Translational Control.” *Current Genetics*. https://doi.org/10.1007/s00294-016-0674-3.

Goldstrohm, Aaron C., and Marvin Wickens. 2008. “Multifunctional Deadenylase Complexes Diversify MRNA Control.” *Nature Reviews Molecular Cell Biology* 9 (4): 337–44. https://doi.org/10.1038/nrm2370.

Gorgoni, Barbara, and Nicola K Gray. 2004. “The Roles of Cytoplasmic Poly(A)-Binding Proteins in Regulating Gene Expression: A Developmental Perspective.” *Briefings in Functional Genomics and Proteomics* 3 (2): 125–41. https://academic.oup.com/bfg/article/3/2/125/251394.

Görlach, Matthias, Christopher G. Burd, and Gideon Dreyfuss. 1994. “The MRNA Poly(A)-Binding Protein: Localization, Abundance, and RNA-Binding Specificity.” *Experimental Cell Research* 211 (2): 400–407. https://doi.org/10.1006/EXCR.1994.1104.

Greenberg, Jay R., and Robert P. Perry. 1972. “The Isolation and Characterization of Steady-State Labeled Messenger RNA from L-Cells.” *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 287 (2): 361–66. https://doi.org/10.1016/0005-2787(72)90386-3.

Gu, H, J Das Gupta, and D R Schoenberg. 1999. “The Poly(A)-Limiting Element Is a Conserved Cis-Acting Sequence That Regulates Poly(A) Tail Length on Nuclear Pre-MRNAs.” *Proceedings of the National Academy of Sciences of the United States of America* 96 (16): 8943–48. https://doi.org/10.1073/PNAS.96.16.8943.

Guzeloglu-Kayisli, Ozlem, Samuel Pauli, Habibe Demir, Maria D. Lalioti, Denny Sakkas, and Emre Seli. 2008. “Identification and Characterization of Human Embryonic

Poly(A) Binding Protein (EPAB).” *MHR: Basic Science of Reproductive Medicine* 14 (10): 581–88. https://doi.org/10.1093/molehr/gan047.

Hanson, Gavin, and Jeff Coller. 2017. “Codon Optimality, Bias and Usage in Translation and MRNA Decay.” *Nature Reviews Molecular Cell Biology* 19 (1): 20–30. https://doi.org/10.1038/nrm.2017.91.

Hentze, Matthias W., Alfredo Castello, Thomas Schwarzl, and Thomas Preiss. 2018. “A Brave New World of RNA-Binding Proteins.” *Nature Reviews Molecular Cell Biology* 19 (5): 327–41. https://doi.org/10.1038/nrm.2017.130.

Hir, H Le, E Izaurralde, L E Maquat, and M J Moore. 2000. “The Spliceosome Deposits Multiple Proteins 20-24 Nucleotides Upstream of MRNA Exon-Exon Junctions.” *The EMBO Journal* 19 (24): 6860–69. https://doi.org/10.1093/emboj/19.24.6860.

Hoshino, S, N Hosoda, Y Araki, T Kobayashi, N Uchida, Y Funakoshi, and T Katada. 1999. “Novel Function of the Eukaryotic Polypeptide-Chain Releasing Factor 3 (ERF3/GSPT) in the MRNA Degradation Pathway.” *Biochemistry. Biokhimiia* 64 (12): 1367–72. http://www.ncbi.nlm.nih.gov/pubmed/10648960.

Hoshino, S, M Imai, T Kobayashi, N Uchida, and T Katada. 1999. “The Eukaryotic Polypeptide Chain Releasing Factor (ERF3/GSPT) Carrying the Translation Termination Signal to the 3’-Poly(A) Tail of MRNA. Direct Association of Erf3/GSPT with Polyadenylate-Binding Protein.” *The Journal of Biological Chemistry* 274 (24): 16677–80. http://www.ncbi.nlm.nih.gov/pubmed/10358005.

Hosoda, Nao, Fabrice Lejeune, and Lynne E Maquat. 2006. “Evidence That Poly (A) Binding Protein C1 Binds Nuclear Pre-MRNA Poly (A) Tails.” *Molecular and Cellular Biology* 26 (8): 3085–97. https://doi.org/10.1128/MCB.26.8.3085.

Imataka, H, A Gradi, and N Sonenberg. 1998. “A Newly Identified N-Terminal Amino Acid Sequence of Human EIF4G Binds Poly(A)-Binding Protein and Functions in Poly(A)-Dependent Translation.” *The EMBO Journal* 17 (24): 7480–89. https://doi.org/10.1093/emboj/17.24.7480.

Jacobson, A, and M Favreau. 1983. “Possible Involvement of Poly(A) in Protein Synthesis.” *Nucleic Acids Research* 11 (18): 6353–68. http://www.ncbi.nlm.nih.gov/pubmed/6137807.

Jalkanen, Aimee L, Stephen J Coleman, and Jeffrey Wilusz. 2014. “Determinants and Implications of MRNA Poly(A) Tail Size--Does This Protein Make My Tail Look Big?” *Seminars in Cell & Developmental Biology* 34 (October): 24–32. https://doi.org/10.1016/j.semcdb.2014.05.018.

Janicki, Susan M, Toshiro Tsukamoto, Simone E Salghetti, William P Tansey, Ravi

Sachidanandam, Kannanganattu V Prasanth, Thomas Ried, et al. 2004. “From Silencing to Gene Expression: Real-Time Analysis in Single Cells.” *Cell* 116 (5): 683–98. https://doi.org/10.1016/s0092-8674(04)00171-0.

Kazazian, Haig H. 2014. “Processed Pseudogene Insertions in Somatic Cells.” *Mobile DNA* 5 (1): 20. https://doi.org/10.1186/1759-8753-5-20.

Keller, Rebecca W., Uwe Kühn, Mateo Aragón, Larissa Bornikova, Elmar Wahle, and David G. Bear. 2000. “The Nuclear Poly(A) Binding Protein, PABP2, Forms an Oligomeric Particle Covering the Length of the Poly(A) Tail.” *Journal of Molecular Biology* 297 (3): 569–83. https://doi.org/10.1006/JMBI.2000.3572.

Kervestin, Stephanie, and Allan Jacobson. 2012. “NMD: A Multifaceted Response to Premature Translational Termination.” *Nature Reviews. Molecular Cell Biology* 13 (11): 700–712. https://doi.org/10.1038/nrm3454.

Khodor, Yevgenia L, Joseph Rodriguez, Katharine C Abruzzi, Chih-Hang Anthony Tang, Michael T Marr, and Michael Rosbash. 2011. “Nascent-Seq Indicates Widespread Cotranscriptional Pre-MRNA Splicing in Drosophila.” *Genes & Development* 25 (23): 2502–12. https://doi.org/10.1101/gad.178962.111.

Kini, Hemant K., Jian Kong, and Stephen A. Liebhaber. 2014. “Cytoplasmic Poly(A) Binding Protein C4 Serves a Critical Role in Erythroid Differentiation.” *Molecular and Cellular Biology* 34 (7): 1300–1309. https://doi.org/10.1128/MCB.01683-13.

Kleene, K C, E Mulligan, D Steiger, K Donohue, and M A Mastrangelo. 1998. “The Mouse Gene Encoding the Testis-Specific Isoform of Poly(A) Binding Protein (Pabp2) Is an Expressed Retroposon: Intimations That Gene Expression in Spermatogenic Cells Facilitates the Creation of New Genes.” *Journal of Molecular Evolution* 47 (3): 275–81. https://doi.org/10.1007/pl00006385.

Ko, Sunhee, Ichiro Kawasaki, and Yhong Hee Shim. 2013. “PAB-1, a Caenorhabditis Elegans Poly(A)-Binding Protein, Regulates MRNA Metabolism in Germline by Interacting with CGH-1 and CAR-1.” *PLoS ONE* 8 (12). https://doi.org/10.1371/journal.pone.0084798.

Krause, Sabine, Stan Fakan, Karsten Weis, and Elmar Wahle. 1994. “Immunodetection of Poly(A) Binding Protein II in the Cell Nucleus.” *Experimental Cell Research* 214 (1): 75–82. https://doi.org/10.1006/excr.1994.1235.

Kühn, U, and T Pieler. 1996. “Xenopus Poly(A) Binding Protein: Functional Domains in RNA Binding and Protein-Protein Interaction.” *Journal of Molecular Biology* 256 (1): 20–30. https://doi.org/10.1006/jmbi.1996.0065.

Kumar, G Renuka, and Britt A Glaunsinger. 2010. “Nuclear Import of Cytoplasmic

Poly(A) Binding Protein Restricts Gene Expression via Hyperadenylation and Nuclear Retention of MRNA." *Molecular and Cellular Biology* 30 (21): 4996–5008. https://doi.org/10.1128/MCB.00600-10.

Laishram, Rakesh S. 2014. "Poly(A) Polymerase (PAP) Diversity in Gene Expression – Star-PAP vs Canonical PAP." *FEBS Letters* 588 (14): 2185–97. https://doi.org/10.1016/J.FEBSLET.2014.05.029.

Lee, S Y, J Mendecki, and G Brawerman. 1971. "A Polynucleotide Segment Rich in Adenylic Acid in the Rapidly-Labeled Polyribosomal RNA Component of Mouse Sarcoma 180 Ascites Cells." *Proceedings of the National Academy of Sciences of the United States of America* 68 (6): 1331–35. https://doi.org/10.1073/pnas.68.6.1331.

Lee, Stuart, Albert Y Zhang, Shian Su, Ashley P Ng, Aliaksei Z Holik, Marie-Liesse Asselin-Labat, Matthew E Ritchie, and Charity W Law. 2020. "Covering All Your Bases: Incorporating Intron Signal from RNA-Seq Data." *NAR Genomics and Bioinformatics* 2 (3). https://doi.org/10.1093/nargab/lqaa073.

Lee, Yeon J., and Britt A. Glaunsinger. 2009. "Aberrant Herpesvirus-Induced Polyadenylation Correlates With Cellular Messenger RNA Destruction." Edited by Bill Sugden. *PLoS Biology* 7 (5): e1000107. https://doi.org/10.1371/journal.pbio.1000107.

Lejeune, Fabrice, Yasuhito Ishigaki, Xiaojie Li, and Lynne E Maquat. 2002. "The Exon Junction Complex Is Detected on CBP80-Bound but Not EIF4E-Bound MRNA in Mammalian Cells: Dynamics of MRNP Remodeling." *The EMBO Journal* 21 (13): 3536–45. https://doi.org/10.1093/emboj/cdf345.

Liao, Yang, Gordon K Smyth, and Wei Shi. 2014. "FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." *Bioinformatics (Oxford, England)* 30 (7): 923–30. https://doi.org/10.1093/bioinformatics/btt656.

Lim, Jaechul, Minju Ha, Hyeshik Chang, S. Chul Kwon, Dhirendra K. Simanshu, Dinshaw J. Patel, and V. Narry Kim. 2014. "Uridylation by TUT4 and TUT7 Marks MRNA for Degradation." *Cell* 159 (6): 1365–76. https://doi.org/10.1016/J.CELL.2014.10.055.

Lim, Jaechul, Dongwan Kim, Young-Suk Lee, Minju Ha, Mihye Lee, Jinah Yeo, Hyeshik Chang, Jaewon Song, Kwangseog Ahn, and V. Narry Kim. 2018. "Mixed Tailing by TENT4A and TENT4B Shields MRNA from Rapid Deadenylation." *Science (New York, N.Y.)* 361 (6403): 701–4. https://doi.org/10.1126/science.aam5794.

Lim, Jaechul, Mihye Lee, Ahyeon Son, Hyeshik Chang, and V. Narry Kim. 2016. "MTAIL-Seq Reveals Dynamic Poly(A) Tail Regulation in Oocyte-to-Embryo

Development." *Genes & Development* 30 (14): 1671–82. https://doi.org/10.1101/gad.284802.116.

Lima, Sarah Azoubel, Laura B Chipman, Angela L Nicholson, Ying-Hsin Chen, Brian A Yee, Gene W Yeo, Jeff Coller, and Amy E Pasquinelli. 2017. "Short Poly(A) Tails Are a Conserved Feature of Highly Expressed Genes." *Nature Structural & Molecular Biology* 24 (12): 1057–63. https://doi.org/10.1038/nsmb.3499.

Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Lugowski, Andrew, Beth Nicholson, and Olivia S Rissland. 2018. "DRUID: A Pipeline for Transcriptome-Wide Measurements of MRNA Stability." *RNA (New York, N.Y.)* 24 (5): 623–32. https://doi.org/10.1261/rna.062877.117.

Mallam, Anna L., Wisath Sae-Lee, Jeffrey M. Schaub, Fan Tu, Anna Battenhouse, Yu Jin Jang, Jonghwan Kim, et al. 2019. "Systematic Discovery of Endogenous Human Ribonucleoprotein Complexes." *Cell Reports* 29 (5): 1351-1368.e5. https://doi.org/10.1016/j.celrep.2019.09.060.

Mangus, David A, Matthew C Evans, Nathan S Agrin, Mandy Smith, Preetam Gongidi, and Allan Jacobson. 2004. "Positive and Negative Regulation of Poly (A) Nuclease." *Molecular and Cellular Biology* 24 (12): 5521–33. https://doi.org/10.1128/MCB.24.12.5521.

Mangus, David A, Matthew C Evans, and Allan Jacobson. 2003. "Poly(A)-Binding Proteins: Multifunctional Scaffolds for the Post-Transcriptional Control of Gene Expression." *Genome Biology* 4 (7): 223. https://doi.org/10.1186/gb-2003-4-7-223.

Maraia, Richard J., Sandy Mattijssen, Isabel Cruz-Gallardo, and Maria R. Conte. 2017. "The La and Related RNA-Binding Proteins (LARPs): Structures, Functions, and Evolving Perspectives." *Wiley Interdisciplinary Reviews: RNA* 8 (6): e1430. https://doi.org/10.1002/wrna.1430.

Meijer, Hedda A, Martin Bushell, Kirsti Hill, Timothy W Gant, Anne E Willis, Peter Jones, and Cornelia H de Moor. 2007. "A Novel Method for Poly(A) Fractionation Reveals a Large Population of MRNAs with a Short Poly(A) Tail in Mammalian Cells." *Nucleic Acids Research* 35 (19): e132. https://doi.org/10.1093/nar/gkm830.

Merrick, William C. 2004. "Cap-Dependent and Cap-Independent Translation in Eukaryotic Systems." *Gene* 332 (May): 1–11. https://doi.org/10.1016/j.gene.2004.02.051.

Metzker, Michael L. 2010. "Sequencing Technologies — the next Generation." *Nature*

Reviews Genetics 11 (1): 31–46. https://doi.org/10.1038/nrg2626.

Meyer, Sylke, Claus Urbanke, and Elmar Wahle. 2002. "Equilibrium Studies on the Association of the Nuclear Poly(A) Binding Protein with Poly(A) of Different Lengths." *Biochemistry* 41 (19): 6082–89. https://doi.org/10.1021/bi0160866.

Mi, Huaiyu, Dustin Ebert, Anushya Muruganujan, Caitlin Mills, Laurent-Philippe Albou, Tremayne Mushayamaha, and Paul D Thomas. 2021. "PANTHER Version 16: A Revised Family Classification, Tree-Based Classification Tool, Enhancer Regions and Extensive API." *Nucleic Acids Research* 49 (D1): D394–403. https://doi.org/10.1093/nar/gkaa1106.

Morgan, Marcos, Christian Much, Monica DiGiacomo, Chiara Azzi, Ivayla Ivanova, Dimitrios M. Vitsios, Jelena Pistolic, et al. 2017. "MRNA 3′ Uridylation and Poly(A) Tail Length Sculpt the Mammalian Maternal Transcriptome." *Nature* 548 (7667): 347–51. https://doi.org/10.1038/nature23318.

Moteki, Shin, and David Price. 2002. "Functional Coupling of Capping and Transcription of MRNA." *Molecular Cell* 10 (3): 599–609. https://doi.org/10.1016/S1097-2765(02)00660-3.

Mugridge, Jeffrey S., Jeff Coller, and John D. Gross. 2018. "Structural and Molecular Mechanisms for the Control of Eukaryotic 5′–3′ MRNA Decay." *Nature Structural & Molecular Biology* 25 (12): 1077–85. https://doi.org/10.1038/s41594-018-0164-z.

Muniz, Lisa, Lee Davidson, and Steven West. 2015. "Poly(A) Polymerase and the Nuclear Poly(A) Binding Protein, PABPN1, Coordinate the Splicing and Degradation of a Subset of Human Pre-MRNAs." *Molecular and Cellular Biology* 35 (13): 2218–30. https://doi.org/10.1128/MCB.00123-15.

Nemeth, A, S Krause, D Blank, A Jenny, P Jenö, A Lustig, and E Wahle. 1995. "Isolation of Genomic and CDNA Clones Encoding Bovine Poly(A) Binding Protein II." *Nucleic Acids Research* 23 (20): 4034–41. https://doi.org/10.1093/nar/23.20.4034.

Nicholson, Angela L, and Amy E Pasquinelli. 2019. "Tales of Detailed Poly(A) Tails." *Trends in Cell Biology* 29 (3): 191–200. https://doi.org/10.1016/j.tcb.2018.11.002.

O'Leary, Nuala A, Mathew W Wright, J Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, et al. 2016. "Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation." *Nucleic Acids Research* 44 (D1): D733-45. https://doi.org/10.1093/nar/gkv1189.

Oesterreich, Fernando Carrillo, Lydia Herzel, Korinna Straube, Katja Hujer, Jonathon Howard, and Karla M Neugebauer. 2016. "Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II." *Cell* 165 (2): 372–81.

https://doi.org/10.1016/j.cell.2016.02.045.

Osawa, Masanori, Nao Hosoda, Tamiji Nakanishi, Naoyuki Uchida, Tomomi Kimura, Shunsuke Imai, Asako Machiyama, Toshiaki Katada, Shin-ichi Hoshino, and Ichio Shimada. 2012. "Biological Role of the Two Overlapping Poly(A)-Binding Protein Interacting Motifs 2 (PAM2) of Eukaryotic Releasing Factor ERF3 in MRNA Decay." *RNA (New York, N.Y.)* 18 (11): 1957–67. https://doi.org/10.1261/rna.035311.112.

Osheim, Yvonne N., Jr. O.L. Miller, and Ann L. Beyer. 1985. "RNP Particles at Splice Junction Sequences on Drosophila Chorion Transcripts." *Cell* 43 (1): 143–51. https://doi.org/10.1016/0092-8674(85)90019-4.

Palatnik, Carl Mathew, Robert V. Storti, and Allan Jacobson. 1979. "Fractionation and Functional Analysis of Newly Synthesized and Decaying Messenger RNAs from Vegetative Cells of Dictyostelium Discoideum." *Journal of Molecular Biology* 128 (3): 371–95. https://doi.org/10.1016/0022-2836(79)90093-7.

Palatnik, Carl Mathew, Robert V Storti, Anne K Capone, and Allan Jacobson. 1980. "Messenger RNA Stability in Dictyostelium Discoideum Does Poly (A) Have a Regulatory Role ?," 99–118.

Piovesan, Allison, Francesca Antonaros, Lorenza Vitale, Pierluigi Strippoli, Maria Chiara Pelleri, and Maria Caracausi. 2019. "Human Protein-Coding Genes and Gene Feature Statistics in 2019." *BMC Research Notes* 12 (1): 315. https://doi.org/10.1186/s13104-019-4343-8.

Preiss, Thomas, and Matthias W. Hentze. 1998. "Dual Function of the Messenger RNA Cap Structure in Poly(A)-Tail-Promoted Translation in Yeast." *Nature* 392 (6675): 516–20. https://doi.org/10.1038/33192.

Proudfoot, Nick J. 2011. "Ending the Message: Poly(A) Signals Then and Now." *Genes & Development* 25 (17): 1770–82. https://doi.org/10.1101/gad.17268411.

Quail, Michael A, Miriam Smith, Paul Coupland, Thomas D Otto, Simon R Harris, Thomas R Connor, Anna Bertoni, Harold P Swerdlow, and Yong Gu. 2012. "A Tale of Three next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq Sequencers." *BMC Genomics* 13 (July): 341. https://doi.org/10.1186/1471-2164-13-341.

Reuter, Jason A, Damek V Spacek, and Michael P Snyder. 2015. "High-Throughput Sequencing Technologies." *Molecular Cell* 58 (4): 586–97. https://doi.org/10.1016/j.molcel.2015.05.004.

Richter, Joel D., and Nahum Sonenberg. 2005. "Regulation of Cap-Dependent Translation

by EIF4E Inhibitory Proteins." *Nature* 433 (7025): 477–80. https://doi.org/10.1038/nature03205.

Rigo, Frank, and Harold G Martinson. 2009. "Polyadenylation Releases MRNA from RNA Polymerase II in a Process That Is Licensed by Splicing." *RNA (New York, N.Y.)* 15 (5): 823–36. https://doi.org/10.1261/rna.1409209.

Rissland, Olivia S, Andrea Mikulasova, and Chris J Norbury. 2007. "Efficient RNA Polyuridylation by Noncanonical Poly(A) Polymerases." *Molecular and Cellular Biology* 27 (10): 3612–24. https://doi.org/10.1128/MCB.02209-06.

Rissland, Olivia S, and Chris J Norbury. 2009. "Decapping Is Preceded by 3′ Uridylation in a Novel Pathway of Bulk MRNA Turnover." *Nature Structural & Molecular Biology* 16 (6): 616–23. https://doi.org/10.1038/nsmb.1601.

Rosa-Mercado, Nicolle A, Joshua T Zimmer, Maria Apostolidi, Jesse Rinehart, Matthew D Simon, and Joan A Steitz. 2021. "Hyperosmotic Stress Alters the RNA Polymerase II Interactome and Induces Readthrough Transcription despite Widespread Transcriptional Repression." *Molecular Cell* 81 (3): 502-513.e4. https://doi.org/10.1016/j.molcel.2020.12.002.

Sachs, Alan B, Ronald W Davis, and Roger D Kornberg. 1987. "A Single Domain of Yeast Poly(A)-Binding Protein Is Necessary and Sufficient for RNA Binding and Cell Viability." *Molecular and Cellular Biology* 7 (9): 3268–76. https://mcb.asm.org/content/mcb/7/9/3268.full.pdf.

Sachs, Alan B, and Gabriele Varani. 2000. "Eukaryotic Translation Initiation: There Are (at Least) Two Sides to Every Story." *Nature Structural Biology* 7 (5). http://structbio.nature.com.

Sato, Hanae, and Lynne E Maquat. 2009. "Remodeling of the Pioneer Translation Initiation Complex Involves Translation and the Karyopherin Importin Beta." *Genes & Development* 23 (21): 2537–50. https://doi.org/10.1101/gad.1817109.

Sawazaki, Ryoichi, Shunsuke Imai, Mariko Yokogawa, Nao Hosoda, Shin-ichi Hoshino, Muneyo Mio, Kazuhiro Mio, Ichio Shimada, and Masanori Osawa. 2018. "Characterization of the Multimeric Structure of Poly(A)-Binding Protein on a Poly(A) Tail." *Scientific Reports* 8 (1): 1455. https://doi.org/10.1038/s41598-018-19659-6.

Sawicki, SG, W. Jelinek, and JE Darnell. 1977. "3 ' Terminal Addition to HeLa Cell Nuclear and Cytoplasmic Poly(A)." *Journal of Molecular Biology* 113 (1): 219–235. http://www.sciencedirect.com/science/article/pii/0022283677900511.

Schmidt, Marie Joëlle, and Chris J. Norbury. 2010. "Polyadenylation and beyond:

Emerging Roles for Noncanonical Poly(A) Polymerases." *Wiley Interdisciplinary Reviews: RNA* 1 (1): 142–51. https://doi.org/10.1002/wrna.16.

Schreiner, William P, Delaney C Pagliuso, Jacob M Garrigues, Jerry S Chen, Antti P Aalto, and Amy E Pasquinelli. 2019. "Remodeling of the Caenorhabditis Elegans Non-Coding RNA Transcriptome by Heat Shock." *Nucleic Acids Research* 47 (18): 9829–41. https://doi.org/10.1093/nar/gkz693.

Shatkin, A J. 1976. "Capping of Eucaryotic MRNAs." *Cell* 9 (4 PT 2): 645–53. https://doi.org/10.1016/0092-8674(76)90128-8.

Sheets, M D, C A Fox, T Hunt, G Vande Woude, and M Wickens. 1994. "The 3'-Untranslated Regions of c-Mos and Cyclin MRNAs Stimulate Translation by Regulating Cytoplasmic Polyadenylation." *Genes & Development* 8 (8): 926–38. https://doi.org/10.1101/GAD.8.8.926.

Sheets, Michael D, and Marvin Wickens. 1989. "Two Phases in the Addition of a Poly(A) Tail." *Genes and Development*, no. 3: 1401–12. http://genesdev.cshlp.org/content/3/9/1401.full.pdf.

Sheiness, D, L Puckett, and J E Darnell. 1975. "Possible Relationship of Poly(A) Shortening to MRNA Turnover." *Proceedings of the National Academy of Sciences of the United States of America* 72 (3): 1077–81. https://doi.org/10.1073/pnas.72.3.1077.

Sheiness, Diana, and J E Darnell. 1973. "Polyadenylic Acid Segment in MRNA Becomes Shorter with Age." *Nature New Biology* 241: 265–68. https://doi.org/https://doi.org/10.1038/newbio241265a0.

Siebrasse, Jan Peter, Tim Kaminski, and Ulrich Kubitscheck. 2012. "Nuclear Export of Single Native MRNA Molecules Observed by Light Sheet Fluorescence Microscopy." *Pnas* 109 (24): 9426–31. https://doi.org/10.1073/pnas.1201781109/-/DCSupplemental.

Singh, Guramrit, Gabriel Pratt, Gene W. Yeo, and Melissa J. Moore. 2015. "The Clothes Make the MRNA: Past and Present Trends in MRNP Fashion." *Annual Review of Biochemistry* 84 (1): 325–54. https://doi.org/10.1146/annurev-biochem-080111-092106.

Smith, Bettye L., Daniel R. Gallie, Hanh Le, and Paul K. Hansma. 1997. "Visualization of Poly(A)-Binding Protein Complex Formation with Poly(A) RNA Using Atomic Force Microscopy." *Journal of Structural Biology* 119 (2): 109–17. https://doi.org/10.1006/jsbi.1997.3864.

Spriggs, Keith A, Martin Bushell, and Anne E Willis. 2010. "Translational Regulation of Gene Expression during Conditions of Cell Stress." *Molecular Cell* 40 (2): 228–37.

https://doi.org/10.1016/j.molcel.2010.09.028.

Subtelny, Alexander O., Stephen W. Eichhorn, Grace R. Chen, Hazel Sive, and David P. Bartel. 2014. “Poly(A)-Tail Profiling Reveals an Embryonic Switch in Translational Control.” *Nature* 508 (7494): 66–71. https://doi.org/10.1038/nature13007.

Sun, Mai, Björn Schwalb, Nicole Pirkl, Kerstin C Maier, Arne Schenk, Henrik Failmezger, Achim Tresch, and Patrick Cramer. 2013. “Global Analysis of Eukaryotic MRNA Degradation Reveals Xrn1-Dependent Buffering of Transcript Levels.” *Molecular Cell* 52 (1): 52–62. https://doi.org/10.1016/j.molcel.2013.09.010.

Tarun, S Z, and A B Sachs. 1995. “A Common Function for MRNA 5’ and 3’ Ends in Translation Initiation in Yeast.” *Genes & Development* 9 (23): 2997–3007. https://doi.org/10.1101/gad.9.23.2997.

Tarun, S Z, S E Wells, J A Deardorff, and A B Sachs. 1997. “Translation Initiation Factor EIF4G Mediates in Vitro Poly(A) Tail-Dependent Translation.” *Proceedings of the National Academy of Sciences of the United States of America* 94 (17): 9046–51. https://doi.org/10.1073/PNAS.94.17.9046.

“The Long View on Sequencing.” 2018. *Nature Biotechnology* 36 (4): 287–287. https://doi.org/10.1038/nbt.4125.

Tian, Bin, and Joel H Graber. 2012. “Signals for Pre-MRNA Cleavage and Polyadenylation.” *Wiley Interdisciplinary Reviews. RNA* 3 (3): 385–96. https://doi.org/10.1002/wrna.116.

Topisirovic, Ivan, Yuri V. Svitkin, Nahum Sonenberg, and Aaron J. Shatkin. 2011. “Cap and Cap-Binding Proteins in the Control of Gene Expression.” *Wiley Interdisciplinary Reviews: RNA* 2 (2): 277–98. https://doi.org/10.1002/wrna.52.

Tucker, M, M A Valencia-Sanchez, R R Staples, J Chen, C L Denis, and R Parker. 2001. “The Transcription Factor Associated Ccr4 and Caf1 Proteins Are Components of the Major Cytoplasmic MRNA Deadenylase in Saccharomyces Cerevisiae.” *Cell* 104 (3): 377–86. http://www.ncbi.nlm.nih.gov/pubmed/11239395.

Uchida, Naoyuki, Shin-Ichi Hoshino, Hiroaki Imataka, Nahum Sonenberg, and Toshiaki Katada. 2002. “A Novel Role of the Mammalian GSPT/ERF3 Associating with Poly(A)-Binding Protein in Cap/Poly(A)-Dependent Translation.” *The Journal of Biological Chemistry* 277 (52): 50286–92. https://doi.org/10.1074/jbc.M203029200.

Uchida, Naoyuki, Shin-ichi Hoshino, and Toshiaki Katada. 2004. “Identification of a Human Cytoplasmic Poly(A) Nuclease Complex Stimulated by Poly(A)-Binding Protein.” *Journal of Biological Chemistry* 279 (2): 1383–91. https://doi.org/10.1074/jbc.M309125200.

Udagawa, Tsuyoshi, Sharon A Swanger, Koichi Takeuchi, Jong Heon Kim, Vijayalaxmi Nalavadi, Jihae Shin, Lori J Lorenz, R Suzanne Zukin, Gary J Bassell, and Joel D Richter. 2012. "Bidirectional Control of MRNA Translation and Synaptic Plasticity by the Cytoplasmic Polyadenylation Complex." *Molecular Cell* 47 (2): 253–66. https://doi.org/10.1016/j.molcel.2012.05.016.

Vilborg, Anna, Maria C. Passarelli, Therese A. Yario, Kazimierz T. Tycowski, and Joan A. Steitz. 2015. "Widespread Inducible Transcription Downstream of Human Genes." *Molecular Cell* 59 (3): 449–61. https://doi.org/10.1016/J.MOLCEL.2015.06.016.

Vinayak, Jyotsna, Stefano A Marrella, Rawaa H Hussain, Leonid Rozenfeld, Karine Solomon, and Mark A Bayfield. 2018. "Human La Binds MRNAs through Contacts to the Poly(A) Tail." *Nucleic Acids Research* 46 (8): 4228–40. https://doi.org/10.1093/nar/gky090.

Wahle, Elmar. 1991. "A Novel Poly (A) -Binding Protein Acts As a Specificity Factor in the Second Phase of Messenger RNA Polyadenylation." *Cell* 66: 759–68.

Wahle, Elmar. 1995. "Poly(A) Tail Length Control Is Caused by Termination of Processive Synthesis." *Journal of Biological Chemistry* 270 (6): 2800–2808. http://www.jbc.org/content/270/6/2800.full.pdf.

Wahle, Elmar, Ariel Lustig, Paul Jeno, and Patrik Maurer. 1993. "Mammalian Poly(A)-Binding Protein II." *Journal of Biological Chemistry* 268 (4): 2937–45. https://reader.elsevier.com/reader/sd/pii/S0021925818538643?token=38A69027BEF9474A1F3C46B85CE5161C71D0B2C070A5F4AF8A6565A99D931E89E6825631456B285C359613EB90D39D17&originRegion=us-east-1&originCreation=20210822213746.

Wang, Z, N Day, P Trifillis, and M Kiledjian. 1999. "An MRNA Stability Complex Functions with Poly(A)-Binding Protein to Stabilize MRNA in Vitro." *Molecular and Cellular Biology* 19 (7): 4552–60.

Webster, Michael W, Ying-Hsin Chen, James A W Stowell, Najwa Alhusaini, Thomas Sweet, Brenton R Graveley, Jeff Coller, and Lori A Passmore. 2018. "MRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases." *Molecular Cell* 70 (6): 1089-1100.e8. https://doi.org/10.1016/j.molcel.2018.05.033.

Weill, Laure, Eulàlia Belloc, Felice-Alessio Bava, and Raúl Méndez. 2012. "Translational Control by Changes in Poly(A) Tail Length: Recycling MRNAs." *Nature Structural & Molecular Biology* 19 (6): 577–85. https://doi.org/10.1038/nsmb.2311.

Wigington, Callie P, Kathryn R Williams, Michael P Meers, Gary J Bassell, and Anita H

Corbett. 2014. “Poly(A) RNA-Binding Proteins and Polyadenosine RNA: New Members and Novel Functions.” *Wiley Interdisciplinary Reviews. RNA* 5 (5): 601–22. https://doi.org/10.1002/wrna.1233.

Wolf, Jana, and Lori A Passmore. 2014. “MRNA Deadenylation by Pan2-Pan3.” *Biochemical Society Transactions* 42 (1): 184–87. https://doi.org/10.1042/BST20130211.

Woo, Yu Mi, Yeonui Kwak, Sim Namkoong, Katla Kristjánsdóttir, Seung Ha Lee, Jun Hee Lee, and Hojoong Kwak. 2018. “TED-Seq Identifies the Dynamics of Poly(A) Length during ER Stress.” *Cell Reports* 24 (13): 3630-3641.e7. https://doi.org/10.1016/j.celrep.2018.08.084.

Woods, Alison J, Marnie S Roberts, Jyoti Choudhary, Simon T Barry, Yuichi Mazaki, Hisataka Sabe, Simon J Morley, David R Critchley, and Jim C Norman. 2002. “Paxillin Associates with Poly(A)-Binding Protein 1 at the Dense Endoplasmic Reticulum and the Leading Edge of Migrating Cells.” *The Journal of Biological Chemistry* 277 (8): 6428–37. https://doi.org/10.1074/jbc.M109446200.

Workman, Rachael E, Alison D Tang, Paul S Tang, Miten Jain, John R Tyson, Roham Razaghi, Philip C Zuzarte, et al. 2019. “Nanopore Native RNA Sequencing of a Human Poly(A) Transcriptome.” *Nature Methods* 16 (12): 1297–1305. https://doi.org/10.1038/s41592-019-0617-2.

Yang, H, C S Duckett, and T Lindsten. 1995. “IPABP, an Inducible Poly(A)-Binding Protein Detected in Activated Human T Cells.” *Molecular and Cellular Biology* 15 (12): 6770–76. https://doi.org/10.1128/MCB.15.12.6770.

Yi, Hyerim, Joha Park, Minju Ha, Jaechul Lim, Hyeshik Chang, and V. Narry Kim. 2018. “PABP Cooperates with the CCR4-NOT Complex to Promote MRNA Deadenylation and Block Precocious Decay.” *Molecular Cell* 70 (6): 1081-1088.e5. https://doi.org/10.1016/j.molcel.2018.05.009.

Zerbino, Daniel R., Adam Frankish, and Paul Flicek. 2020. “Progress, Challenges, and Surprises in Annotating the Human Genome.” *Annual Review of Genomics and Human Genetics* 21: 55–79. https://doi.org/10.1146/annurev-genom-121119-083418.