UNIVERSITY OF CALIFORNIA, SAN DIEGO

Chasing tails: a link between mRNA poly(A) tail length and translation

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

in

Biology

by

Sarah Azoubel Lima

Committee in charge:

Professor Amy E. Pasquinelli, Chair
Professor Jens Lykke-Andersen
Professor Emily Troemel
Professor Steven Wasserman
Professor Eugene Yeo

2016

The Dissertation of Sarah Azoubel Lima is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2016

DEDICATION

To Paulo,

for his unwavering support.

EPIGRAPH

É pau, é pedra, é o fim do caminho

É um resto de toco, é um pouco sozinho

É um passo, é uma ponte, é um sapo, é uma rã

É um belo horizonte, é uma febre terçã

São as águas de março fechando o verão

É a promessa de vida no teu coração

—Antonio Carlos Jobim

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Professor Amy E. Pasquinelli for her constant mentoring and support during my time in her laboratory. Amy' dedication and scientific brilliance are accompanied by a thoughtful dose of understanding and guidance, from which I benefited more times than I can count. I would also like to acknowledge my committee members, Professor Jens Lykke-Andersen, Professor Emily Troemel, Professor Gene Yeo, and Professor Steven Wasserman for propelling me forward with their helpful and intelligent considerations about this work. In addition, I would like to thank my former P.I., Professor Carmen V. Ferreira, who always encouraged me to pursue my scientific aspirations.

Thanks to my fellow lab members, who made for a very productive and happy work environment, including Emily, Vanessa, Antti, Ian and Will. My time here would not have been nearly as enjoyable without our lively daily interactions. Thanks to Laura and Angela, who are great new additions to our lab and did not hesitate to jump in and help with the final experiments of my project. I would like to specially thank James, we started together and have continued to help and support each other at every step of our Ph.Ds. I think I can safely say he helped me more: it did not matter if I needed to learn new coding, troubleshoot experiments, pick color schemes, format this very thesis, or have a beer after a tough day, James was there.

I want to thank the friends I made in San Diego for making my time here so much richer. I am waiting for your visit. I also have to specially thank my friends Raissa, Nani and Pirro, who were not always in the same continent but are always in my thoughts, in my heart, and in my WhatsApp.

I would like to thank my family, who has given me incredible love and support.

My parents valued our education above all and have stood by my side in everything I chose to do. Finally, I would like to thank Paulo. We started a life together in San Diego, and I could not ask for a more loving, understanding, and encouraging partner.

Chapter 2, in full, is a reprint of the material as it appears in Advances in Experimental Medicine and Biology by Sarah Azoubel Lima and Amy E. Pasquinelli, Springer, 2014. I was the primary author.

Chapter 3, in full, is a reprint of the submitted manuscript "Efficient translation promotes poly(A) pruning" by Sarah Azoubel Lima and Amy E. Pasquinelli. I was the primary author.

VITA

2005-2008	B. S. in Biological Sciences State University of Campinas (UNICAMP), Brazil
2009-2011	M. S. in Biochemistry State University of Campinas (UNICAMP), Brazil
2011-2016	Ph. D. in Biology University of California, San Diego

AWARDS AND HONORS

2009	FAPESP M. S. Fellowship, Brazil
2013	HHMI International Pre-doctoral Fellowship
2016	Best Poster Award, RNA Society Meeting (Kyoto, Japan)

PUBLICATIONS

Souza, A.C.S., Azoubel, S., Queiroz, K.C.S., Peppelenbosch, M.P., and Ferreira, C. V., "From immune response to cancer: a spot on the low molecular weight protein tyrosine phosphatase", *Cell. Mol. Life Sci.*, 66, 2009.

Ourique, A.F., Azoubel, S., Ferreira, C. V, Silva, C.B., Marchiori, M.C.L., Pohlmann, A.R., Guterres, S.S., and Beck, R.C.R., "Lipid-core nanocapsules as a nanomedicine for parenteral administration of tretinoin: development and in vitro antitumor activity on human myeloid leukaemia cells", *J. Biomed. Nanotechno.*, 6, 2010.

Lima, S.A., and Pasquinelli, A.E.,"Identification of miRNAs and their targets in *C. elegans*", *Adv. Exp. Med. Biol.*, 825, 2014.

Lima, S.A., and Pasquinelli, A.E.,"Efficient translation promotes poly(A) pruning", *in revision*, 2016.

ABSTRACT OF THE DISSERTATION

Chasing tails: a link between mRNA poly(A) tail length and translation

by

Sarah Azoubel Lima

Doctor of Philosophy in Biology

University of California, San Diego, 2016

Professor Amy E. Pasquinelli, Chair

Every step in the existence of a messenger mRNA is carefully regulated. I explored the impact of a feature present in the great majority of eukaryotic mRNAs: the poly(A) tail. Poly(A) tails are important elements in mRNA maturation, translation and stability, but only recently the tail lengths of mRNAs have been revealed on a transcriptome wide scale. I have adapted the poly(A) sequencing method TAIL-seq and developed aTAIL-seq (adapted TAIL-seq). Chapter 4 contains detailed protocols of aTAIL-seq and other methods of poly(A) analysis.

In this work, I have applied aTAIL-seq to measure the poly(A) tails of

Caenorhabditis elegans. The nematode *C. elegans* is an model organism that has been used to identify multiple conserved biological processes, including the discovery of microRNAs (reviewed in Chapter 2).

In Chapter 3, I explore the relationship between the poly(A) tail, translation and stability in *C. elegans* and other eukaryotes. Traditionally, long tails have been thought to enhance translation and stability of mRNAs. However, we found that the most abundant types of mRNAs, such as those encoding ribosomal proteins, have the shortest tail lengths, while the least abundant, such as mRNAs for transcription factors, have the longest tails. This difference is related to translation efficiency, as genes enriched for optimal codons and ribosomal occupancy have the shortest median tails. These results suggest that translation promotes poly(A) tail shortening, an idea supported by our observation that non-coding RNAs carry long poly(A) tails. We find that, in general, the most abundant and well-translated mRNAs have the shortest median poly(A) tail lengths. While this study seems to contradict the dogma that deadenylation is associated with translational inhibition and mRNA decay, it actually points to a mechanism where well expressed mRNAs undergo pruning of their poly(A) tails to lengths that accommodate 1-2 poly(A) binding proteins (PABPs). This may be an optimal size for PABP to engage in protective and translational functions. Overall, our findings suggest that translation regulates pruning of poly(A) tails, either by actively promoting their shortening or by stabilizing the short-tailed mRNAs. This work changes our understanding of the regulation of poly(A) tail length and gene expression.

Chapter 1

Introduction

1.1 The complex regulation of mRNAs

Every step in the existence of a messenger mRNA is carefully regulated: from its transcription and processing in the nucleus, to its transport to the cytoplasm, to its translation by ribosomes, and finally its decay. These tightly orchestrated processes screen out errors in the message and ensure that genes are robustly expressed at proper intensities.

In this work, I explored the impact of an important element present in the great majority of eukaryotic mRNAs: the poly(A) tail. This introduction is an overview of how the poly(A) tail is generated and its involvement in the translation and decay of mRNAs.

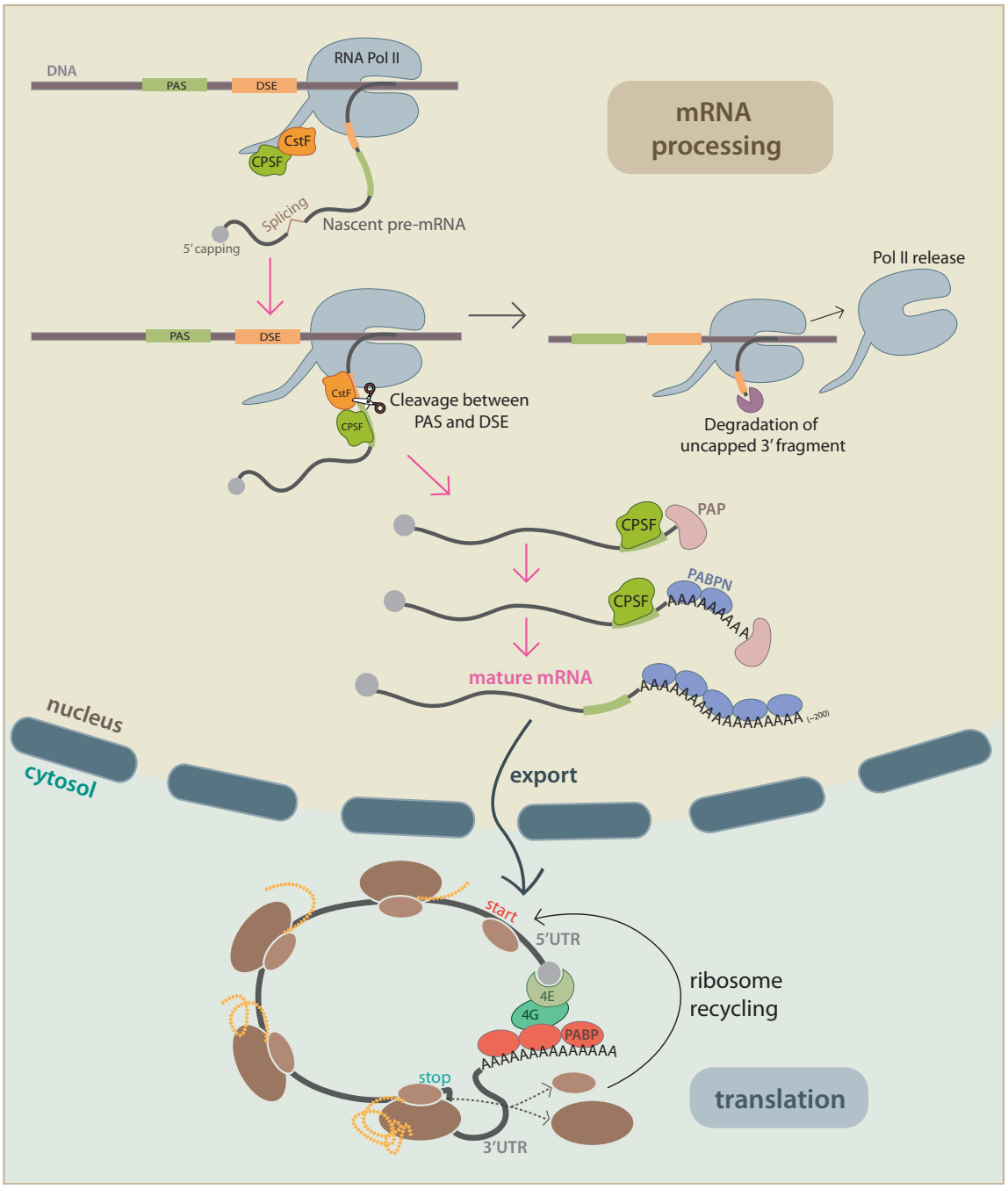
1.2 mRNA transcription and 3' processing

In the nucleus, mRNAs start out as precursor mRNAs or pre-mRNAs. Pre-mRNAs are transcribed from DNA by RNA polymerase II (Pol II), and go through a co-transcriptional maturation process in which introns are removed through splicing, a 7-methylguanylate cap (m7G) cap is added to the 5' end and the 3' is cleaved and polyadenylated (Figure 1.1) (Bentley, 2014). Cleavage and polyadenylation of mRNAs is a process orchestrated by multiple RNA-binding proteins (RBPs) and cis-acting sequence elements in the message. In particular, polyadenylation signals (PAS) present in the 3'UTRs of mRNAs are crucial in effectively promoting efficient 3' end processing and transcription termination. The canonical PAS is the hexamer AAUAAA, which seems to be present in 50-60% of mammalian mRNAs and ~40% of *Caenorhabiditis elegans* mRNAs (Macdonald and Redondo, 2002; Tian and Graber, 2012). AAUAAA is a motif conserved in metazoans and yeast, however sequence variants can also trigger polyadenylation (with some decrease in efficiency) and some polyadenylated transcripts do not possess a recognizable PAS. Another important supporting sequence element is the U or GU-rich downstream sequence element (DSE), located 30-45 nucleotides (nt) downstream from the PAS. As illustrated in Figure 1.1, the PAS and DSE act as recruiting elements for the cleavage and polyadenylation factors. While the PAS is the anchoring point for the cleavage and polyadenylation specificity factor (CPSF), the DSE associates with the cleavage stimulation factor (CstF). The relative position between the two complexes promotes the cleavage of pre-mRNA, which occurs between the two sequence elements, typically ~20nt downstream from the PAS. The CPSF complex contains the endonuclease responsive for cleavage, ensuring dependence on the

presence of the PAS (Eckmann et al., 2011).

After cleavage, a canonical poly(A) polymerase (PAP), which is also recruited by the CPSF, adds untemplated adenosines to the 3' end of the nascent mRNA, while the nuclear poly(A) binding protein (PABPN) starts coating the new poly(A) tail. Because CPSF and PABPN act together to greatly stimulate PAP's enzymatic activity, it is thought that loss of contact between PAP and CPSF is the cause for tails reaching a maximum length of 200-250nt in metazoans and 70-90nt in yeast. Once the growing poly(A) tail reaches these lengths, PABN may be incapable of maintaining the structure necessary for the maintenance of contact between PAP and CPSF, causing a sharp drop in processivity, which virtually stops tail elongation (Eckmann et al., 2011).

Figure 1.1: mRNA synthesis, processing and translation. mRNAs are transcribed by RNA Pol II in the nucleus. Nascent transcripts go through a co-transcriptional maturation process that involves 5' capping, splicing out introns, and 3' cleavage and polyadenylation. Two RNA-binding complexes are responsible for promoting cleavage and polyadenylation. The CPSF (cleavage and polyadenylation specificity factor) recognizes the PAS and the CstF (cleavage stimulation factor) interacts with the DSE. Once both of these factors are docked on the mRNA, an endonuclease in the CPSF cleaves the message between the two sequence elements. The CPSF then recruits PAP (poly(A) polymerase) to extend a tail of untemplated adenosines to the end of the mRNA. The budding tail is coated with the nuclear poly(A) binding protein (PABPN). After completing processing, mature transcripts are exported to the cytoplasm, where PABPN is replaced with its cytoplasmic counterpart, PABP. During translation, PABP interacts with the cap-bound initiation factor eIF4G, enhancing translation efficiency and ribosome recycling by promoting the circularization of the message.


DNA
PAS
DSE
RNA Pol II
CstF
CPSF
Splicing
5′ capping
Nascent pre-mRNA
mRNA processing
Pol II release
Cleavage between PAS and DSE
Degradation of uncapped 3′ fragment
PAP
PABPN
AAAAAAAAA
mature mRNA
AAAAAAAAAAAAAAAAAA (~200)
nucleus
cytosol
export
start
5′UTR
4E
4G
PABP
AAAAAAAAAAAAAA
ribosome recycling
stop
3′UTR
translation

1.3 mRNA export and translation

After complete maturation, mRNAs are exported into the cytoplasm. In metazoans, splicing of the new transcript plays a role in recruiting export factors that will guide the mRNA through the nuclear pore and into the cytosol (Wickramasinghe and Laskey, 2015). Although splicing seems to be the primary driver of export, loss of PABN can also lead to nuclear accumulation of mRNAs (Apponi et al., 2010). The coupling of maturation and export is an important fail-safe to prevent export of defective or immature mRNAs.

Once in the cytoplasm, mRNAs can fulfill their role as messages by being translated into protein. Upon the first round of translation, the nuclear poly(A) binding protein (PABPN) is replaced with its cytoplasmic counterpart, PABP (Maquat et al., 2010; Sato and Maquat, 2009). Multiple cis-acting elements in mRNAs modulate translation. Primarily, the cap plays an essential role in engaging the translation initiation complex. One of these proteins, eIF4G, bridges the cap-bound translation factors to the poly(A) tail, by interacting simultaneously with eIF4E and PABP (Figure 1.1). This promotes circularization of mRNAs, which appears to facilitate ribosome recycling and improve translation efficiency (Costello et al., 2015; Hoshino, 2012; Marshall et al., 2014; Wells et al., 1998). In addition to the cap and poly(A) tail, other features in mRNAs can regulate translation. For instance, in the 5'UTR, presence or lack of strong RNA secondary structures can modulate ribosome scanning (Hinnebusch et al., 2016), and IRES (internal ribosome entry sites) can bypass cap requirements in viral infections and stress (Weingarten-Gabbay et al., 2016). In the 3'UTR, microRNA-targeting sites can inhibit translation and promote deadenylation and decay (See Chapter 3 and Jonas and Izaurralde (2015)) and

docking sites for other RNA-binding proteins can regulate translation and stability (Jia et al., 2013; Quenault et al., 2011). The coding sequence can also modulate translation and stability of mRNAs. In a particular message, the relative presence of translationally optimal and suboptimal codons affects the rates of translation elongation and consequently, protein output (Richter and Coller, 2015). In addition, a high frequency of suboptimal codons in a mRNA can greatly accelerate its decay (Presnyak, 2015).

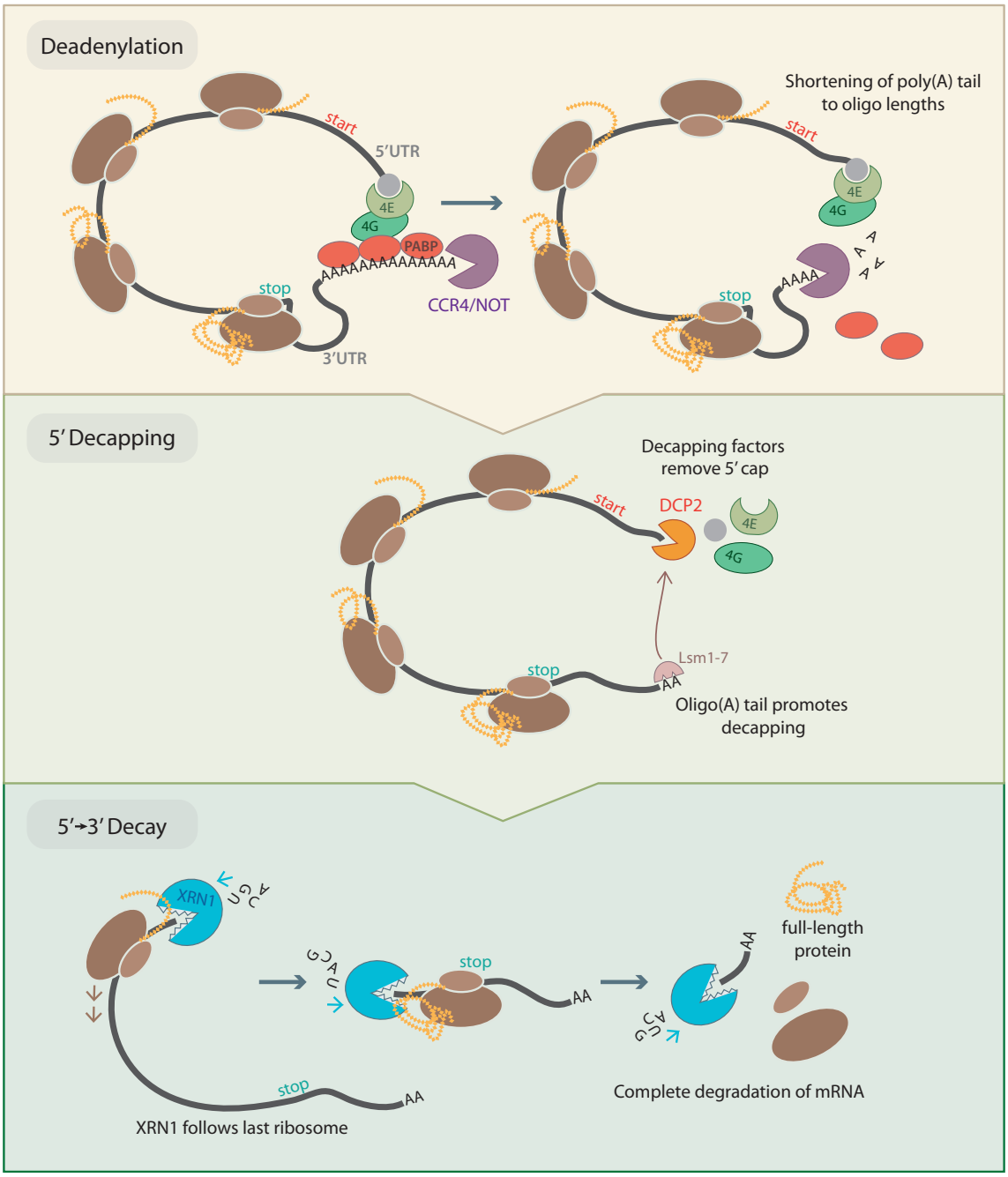
1.4 mRNA decay

For most normal mRNAs, decay occurs in a coordinated fashion that involves three consecutive steps: deadenylation, decapping, and 5'→3' processive degradation (Figure 1.2). First, mRNAs have their poly(A) tails shortened to oligo lengths (~8-12 adenosines), which exclude PABP binding to the tail (Coller and Parker, 2004; Decker and Parker, 1993; Garneau et al., 2007). In eukaryotes, CCR4/NOT is the major deadenylation complex, however other deadenylases such as PAN2/3 (poly(A) nuclease) and PARN (poly(A) specific ribonuclease) can also act on the poly(A) tail (Norbury, 2013; Nousch et al., 2013; Wahle and Winkler, 2013).

Once the oligo tail is unprotected, the Lsm complex recognizes the 3' end of the mRNA and promotes the recruitment of decapping factors such as DCP2 (mRNA-decapping enzyme 2) (Garneau et al., 2007). In yeast, decapping factors have also been found in direct association with the CCR4/NOT complex (Alhusaini and Coller, 2016). Once the mRNA is decapped, the exonuclease XRN1 quickly degrades the message in the 5'→3' direction (Garneau et al., 2007; Sun et al., 2013).

The direction of decay is important as there is increasing evidence that, in general, decay occurs co-translationally (Hu et al., 2009; Pelechano et al., 2015). XRN1 is blocked by and follows the most recently added ribosome, processively degrading the mRNA from the 5' to the 3' end. In this manner, the last protein product is fully synthesized prior to the complete decay of the message. 3'→5' decay of unprotected RNAs can also happen through the exosome, but its impact seems to be minor for most mRNAs under normal conditions (Parker, 2012; Pelechano et al., 2015).

Figure 1.2: The majority of normal mRNAs are degraded through deadenylation and decapping. In the first step of decay, deadenylases such as CCR4/NOT will shorten the poly(A) tail to oligo lengths (10-12As) that exclude PABP binding. The unprotected oligo tail is recognized by the Lsm1-7 complex, which promotes 5' decapping by recruiting DCP2. The decapped transcript is then quickly degraded by the exonuclease XRN1. As decay frequently occurs co-translationally, XRN1 is blocked by the last ribosome translating the message. This prevents the generation of truncated protein products during the decay process, as XRN1 follows the ribosome until translation of the last protein is completed.


Deadenylation
start
5′UTR
4E
4G
PABP
AAAAAAAAAAAAAA
CCR4/NOT
stop
3′UTR
Shortening of poly(A) tail
to oligo lengths
start
4E
4G
stop
AAAA
5′ Decapping
Decapping factors
remove 5′ cap
start
DCP2
4E
4G
Lsm1-7
stop
AA
Oligo(A) tail promotes
decapping
5′→3′ Decay
XRN1
stop
AA
XRN1 follows last ribosome
full-length
protein
Complete degradation of mRNA

1.5 PABP and mRNA stability

Having at least one PABP bound to the poly(A) tail seems to be an important requirement for mRNA stability (Chapter 3 and Park et al. (2016)). PABP is highly conserved in eukaryotes and the protein covers a footprint of 25-30 adenosine (A) residues in the tail (Baer and Kornberg, 1983; Smith et al., 1997; Wang et al., 1999). In yeast, the RNA binding domain of PABP requires 12 As, however that number is unknown in other organisms (Sachs et al., 1987). PABP binding can protect the tail from degradation (Coller and Parker, 2004; Coller et al., 1998; Decker and Parker, 1993). However, the role of PABP in regulating mRNA stability is complex. The PABP C-terminal domain (PABC) binds and activates the major deadenylase complexes PAN2/PAN3 and CCR4/NOT (Hoshino, 2012; Wahle and Winkler, 2013; Xie et al., 2014). PABC also interacts with GW182, a protein containing glycine-tryptophan (GW) repeats, to promote microRNA-mediated deadenylation (Jonas and Izaurralde, 2015; Wahle and Winkler, 2013; Xie et al., 2014). These contrasting roles of PABP suggest that the length of the poly(A) tail and, consequently, the number of PABPs bound could positively or negatively modulate RNA stability.

1.6 Poly(A) length: does it matter?

As mentioned in the previous section, having a tail long enough to accommodate at least one PABP seems to be a basic requirement for the stability (Chapter 3 and Park et al. (2016)) and translation mRNAs (Ivanov et al., 2016; Kahvejian et al., 2005). Beyond that point, much less is known about how important the size of the tail is. For a long time, evidence about the advantages of a shorter or long tail was

restricted to specific biological situations. For instance, during embryogenesis, a cytoplasmic poly(A) polymerase (CPEB) is recruited to select transcripts to extend their tails and activate their translation (Mendez and Richter, 2001). Similarly, in neurons, where several mRNAs are transported to the synapse before they are translated, CPEB promotes local tail extension and translation activation (Norbury, 2013). These findings have resulted in the belief that, in general, longer poly(A) tails were advantageous for translation and stability (Eckmann et al., 2011; Goldstrohm and Wickens, 2008; Wahle and Winkler, 2013). However, until recently, the investigation of these ideas was hindered due to the difficulty in measuring poly(A) tails in a transcriptome-wide manner.

To solve this problem, two high-throughput sequencing methods were recently developed to assay global poly(A) lengths (Chang et al., 2014; Subtelny et al., 2014). Surprisingly, these same studies concluded that poly(A) length was generally not associated with translational efficiency in non-embryonic cells. However, most of these observations were derived from cultured cells, and many questions remained about the relationship between poly(A) length and translation in entire organisms.

In this work, I have adapted the poly(A) sequencing method TAIL-seq (Chang et al., 2014) and developed aTAIL-seq (adapted TAIL-seq). Detailed protocols for aTAIL-seq and other methods of poly(A) analysis are found in Chapter 4. I used aTAIL-seq to examine the poly(A) profile of the model organism *Caenorhabditis elegans*. We were surprised to find that well-translated and abundant mRNAs contain relatively short tails. This discovery led us to uncover a striking inverse relationship between poly(A) length, mRNA stability and translation that is observed across eukaryotes. These themes are explored in detail in Chapters 3 and 5.

Chapter 2

Identification of miRNAs and their targets in *C. elegans*

2.1 Abstract

MicroRNAs (miRNAs) are small non-coding RNAs that direct post-transcriptional regulation of specific target genes. Since their discovery in *Caenorhabditis elegans*, they have been associated with the control of virtually all biological processes and are known to play major roles in development and cellular homeostasis. Yet, the biological roles of most miRNAs remain to be fully known. Furthermore, the precise rules by which miRNAs recognize their targets and mediate gene silencing are still unclear. Systematic identification of miRNAs and of the RNAs they regulate is essential to close these knowledge gaps. Studies in *C. elegans* have been instrumental not only in the discovery phase of miRNA biology but also in the elucidation of mechanisms regulating miRNA expression, target recognition and regulation. This chapter highlights some of the main challenges still

present in the field, while introducing the major studies and methods used to find miRNAs and their targets in the worm.

2.2 Introduction

MicroRNAs (miRNAs) are small, approximately 22 nucleotide (nt) RNAs that mediate post-transcriptional regulation of gene expression throughout the plant and animal kingdoms (Aalto and Pasquinelli, 2012). These small RNAs regulate a large portion of the transcriptome and are involved in virtually all biological processes (Friedman et al., 2009). A single miRNA has the potential to modulate the expression of hundreds or even thousands of targets, producing both specific and cumulative changes in gene expression that are enough to trigger developmental transitions and maintain the balance between homeostasis and disease (Esquela-Kerscher and Slack, 2006; Alvarez-Garcia and Miska, 2005; Kloosterman and Plasterk, 2006). A conservative, and likely underestimated, prediction suggests that at least 27% of *Caenorhabditis elegans* 3'UTRs are under selective pressure to preserve miRNA target sites (Jan et al., 2011).

The biogenesis of miRNAs is a complex and highly regulated process that has been the focus of extensive reviews (Finnegan and Pasquinelli, 2013; Krol et al., 2010) and is summarized in Figure 2.1. Briefly, miRNAs originate from stem-loops embedded in long primary transcripts (pri-miRNAs) or introns (mirtrons). The endonucleases Drosha and Dicer (DRSH-1 and DCR-1 in *C. elegans*) are responsible for processing most miRNAs into their mature forms. Drosha cleaves the miRNA stem-loop from the pri-miRNA, generating the precursor hairpin (pre-miRNA).

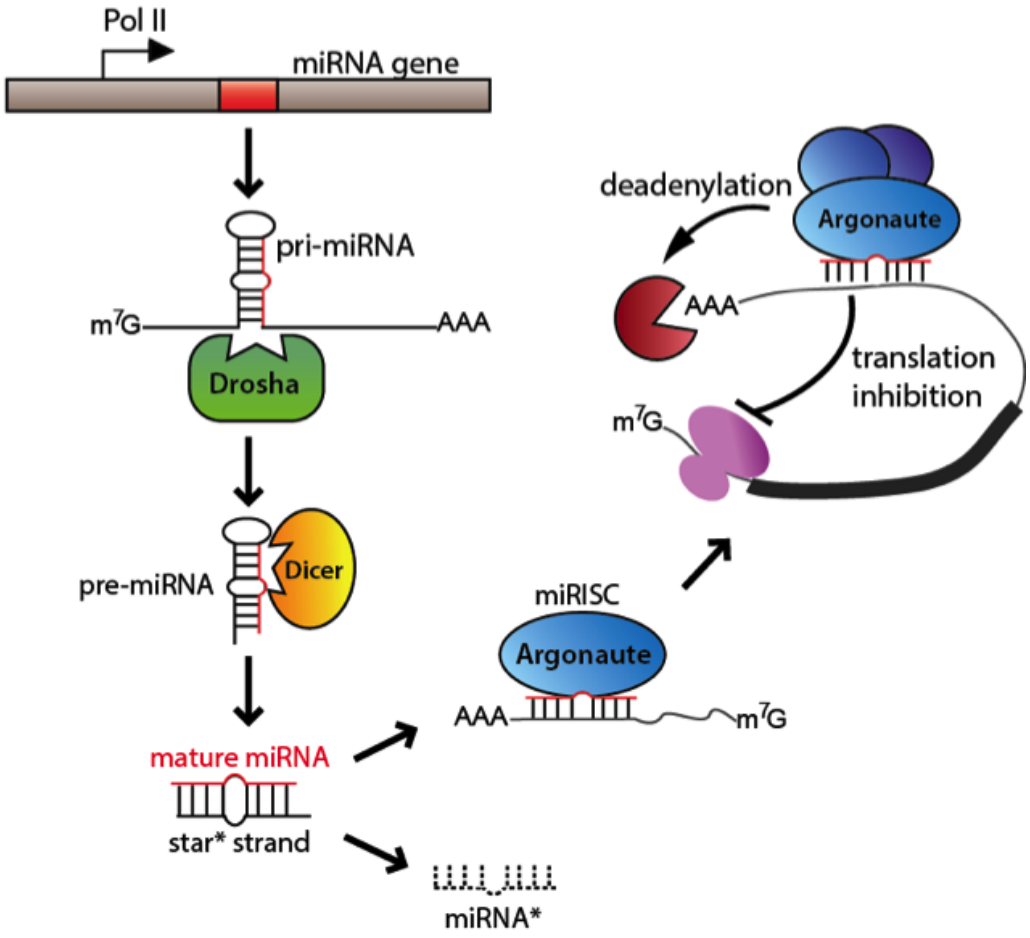


Figure 2.1: microRNA biogenesis. miRNAs are transcribed into long primary transcripts (pri-miRNAs) and processed by Drosha and Dicer (DRSH-1 and DCR-1 in C. elegans). Drosha cleaves the precursor hairpin (pre-miRNA) and Dicer generates a double stranded heteroduplex, formed by the mature miRNA and the miRNA* (star strand). The mature miRNA is loaded onto Argonaute proteins (ALG-1 and ALG-2 in C. elegans) forming miRNA-mediated silencing complexes (miRISCs) and the miRNA* is usually degraded. Targets are recognized by partial base pairing of the miRNA. The miRISC promotes silencing by recruiting factors that promote target degradation and/or translational inhibition.

Dicer then processes the pre-miRNA into a small double stranded heteroduplex, formed by the mature miRNA paired with the passenger (star) strand. The duplex is separated and the mature miRNA is loaded onto Argonaute proteins (ALG-1 and ALG-2 in *C. elegans*), forming miRNA-mediated silencing complexes (miRISCs). The miRNAs act as guides for miRISC by partially base pairing to specific target RNA sequences. Perfect pairing of nucleotides 2-7 of the miRNA, called the seed region, is involved in many but not all target interactions (Bartel, 2009). Once a target is recognized, silencing is achieved by recruiting factors that promote its degradation and/or translational inhibition (Fabian et al., 2010; Pasquinelli, 2012).

Ever since the discovery of miRNAs in C. elegans in 1993, many researchers have concentrated their efforts on identifying these small RNAs and their targets across species (Kozomara and Griffiths-Jones, 2011). C. elegans has many advantages for the study of miRNAs, such as an extensive library of strains, which includes mutants in many miRNA and target genes, (both validated and predicted) (Harris et al., 2010; Miska et al., 2007). There is also an abundance of data on spatial/temporal expression of small RNAs and protein-coding genes from several RNA-seq, microarray and reporter studies (Gerstein et al., 2010). Ultimately, the broad spectra of information on small RNA biology that is available in *C. elegans* makes it an ideal model organism for traditional and more sophisticated studies to elucidate general features of the miRNA pathway. This chapter aims to facilitate navigation through the major studies and methods involved in the identification of miRNAs and their targets in *C. elegans*.

2.3 miRNA identification and expression profiling in *C. elegans*

2.3.1 Finding miRNAs in the worm

Following the original discovery in 1993 of the first miRNA (*lin-4*) in *C. elegans* (Wightman et al., 1993; Lee et al., 1993), interest in small regulatory RNAs picked up steam 7 years later with the revelation that another miRNA, *let-7*, was in fact conserved across many organisms, including humans (Pasquinelli et al., 2000; Reinhart et al., 2000). Anticipating the existence of other genes like *lin-4* and *let-7*, several groups entered the race to systematically identify more members of this curious class in worms, mammalian cells, *Drosophila* and *Arabidopsis* (Lau et al., 2001; Lagos-Quintana et al., 2001; Lim et al., 2003; Reinhart et al., 2002; Lee and Ambros, 2001; Llave et al., 2002). Soon, thousands of novel small RNAs were described and annotated as miRNAs (Ambros et al., 2003; Griffiths-Jones et al., 2006; Kozomara and Griffiths-Jones, 2011).

Forward genetics enabled not only the early identification of miRNAs in *C. elegans* but also the first insights into their functions (Figure 2.3). Notably, *lin-4* (Wightman et al., 1993; Lee et al., 1993) and *let-7* (Reinhart et al., 2000) were characterized as essential regulators of developmental timing, whereas *lsy-6* was found to control neuronal patterning (Johnston and Hobert, 2003). However, the amount of time, labor and luck needed to identify miRNAs through classical genetics renders this approach impractical for executing large-scale searches for novel miRNAs. The understanding that miRNAs are derived from precursor stem-loops of ~65 nt that undergo Dicer processing to produce ~22 nt RNAs with 5' terminal

monophosphates and 3' terminal hydroxyl groups enabled more effective methods for discovering miRNAs (Hutvagner, 2001; Bernstein et al., 2001; Elbashir et al., 2001; Grishok et al., 2001). These features were used to search for new miRNAs by experimental and computational methods (Figure 2.2) by developing protocols for size exclusion and cloning of small RNAs as well as algorithms to analyze the genome for potential miRNA stem-loops (Lee and Ambros, 2001; Lau et al., 2001; Lagos-Quintana et al., 2001; Lim et al., 2003; Grad et al., 2003; Reinhart et al., 2002). Several of the candidate miRNAs (sequenced ~22 nt RNAs embedded in predicted hairpin structures) were confirmed by Northern blotting and, on occasion, by showing that they accumulated as precursors upon down-regulation of Dicer activity. By 2003, when almost 100 worm miRNAs had been identified and experimentally validated, estimates of the total number of *C. elegans* miRNAs varied from 120 (Lim et al., 2003) to over 300 (Grad et al., 2003). In 2006 the Bartel group performed high-throughput pyrosequencing of worm miRNA libraries, raising the number of experimentally detected worm miRNAs to 112 (Ruby et al., 2006). Although subsequent studies helped confirm some previously predicted miRNAs and identify a few new candidates (Gu et al., 2007; Zhang et al., 2007), the biggest jump in the discovery of new miRNAs came with advancements in sequencing technology. Using the Solexa deep sequencing platform, 66 new miRNAs were identified in *C. elegans* (Kato et al., 2009). This was also the first study to report high levels of star strand sequences, which are a good indicator of a hairpin having undergone *bona fide* miRNA processing. Previous efforts to globally sequence miRNAs in the worm had reported only 1% of star strand sequences among their libraries (Ruby et al., 2006), whereas in Kato et al. (2009) 35% of the newly found miRNA candidates

also had reads for the opposite strand. The latest version of miRBase, a database that compiles miRNA sequences and annotations across organisms, counts 223 precursors and 368 mature miRNAs and miRNAs* in *C. elegans* (Kozomara and Griffiths-Jones, 2011). Recently, there has been a push to improve the quality of miRBase miRNA annotations by utilizing even more sensitive RNA-seq technology. The greatly increased number of detected reads allows for more reliable identification of mature miRNAs and the alternative products generated during their processing (star and loop sequences) (Warf et al., 2011).

2.3.2 Profiling miRNA expression patterns in the worm

Determining where and when miRNAs are expressed provides important clues into their functions. From the first studies on miRNAs, there was evidence that these small regulators could exhibit distinct expression patterns during development (Reinhart et al., 2000; Wightman et al., 1993; Lee et al., 1993; Lee and Ambros, 2001; Lau et al., 2001; Lim et al., 2003; Lagos-Quintana et al., 2001; Grad et al., 2003). So far, the efforts to produce global profiles of miRNA expression in *C. elegans* have relied on deep sequencing (Kato et al., 2009), microarrays (Ibáñez-Ventoso et al., 2006; Gu et al., 2007) and promoter:reporter constructs (Martinez et al., 2008; Isik et al., 2010) (Figure 2.2). These approaches are complementary as the first two provide quantitative data on the timing of miRNA expression and the last can be used to assay tissue specificity.

Studies that followed miRNA expression from embryogenesis to young adulthood in *C. elegans* (Kato et al., 2009; Gu et al., 2007) revealed that temporal changes in abundance during development are a common feature of many worm miRNAs.

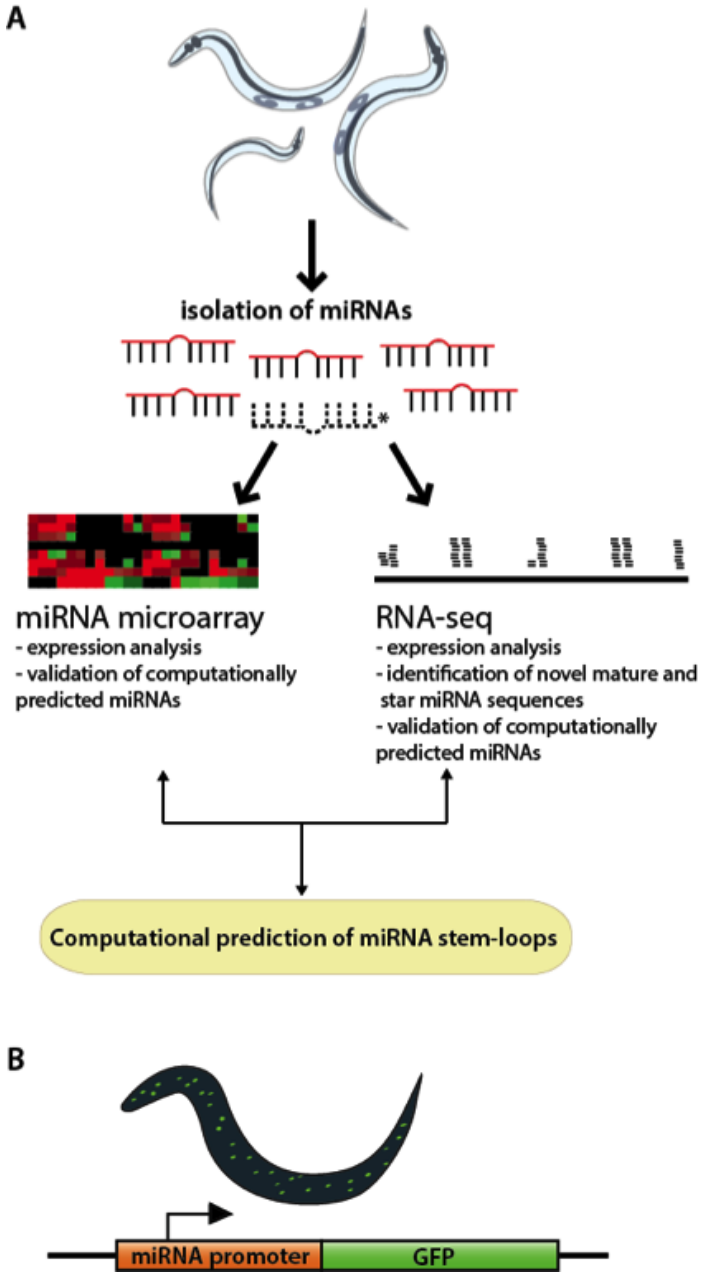


Figure 2.2: miRNA identification and expression profiling. (A) Identification and expression analysis of miRNAs through microarrays and RNA-seq. Global methods to detect miRNAs are used to analyze expression patterns and confirm predicted miRNAs, especially when the rare miRNA* strands are also identified. (B) Spatio-temporal analysis of miRNA expression through reporter constructs.

Microarray analysis on miRNAs isolated from miRISC demonstrated that the levels of over 60% of the detected miRNAs exhibit changes greater then four-fold across development (Gu et al., 2007) and small RNA deep sequencing indicated that the abundance of 16% of the miRNAs change by more than ten-fold (Kato et al., 2009). Interestingly, when miRNA expression was evaluated throughout the worm lifespan, there was a trend of global downregulation in miRNA levels after adulthood, with only a few miRNAs upregulated during aging (among them miR-34 and miR-231) (Ibáñez-Ventoso et al., 2006; Kato et al., 2011). However, a recent study shows that even if miRNAs are generally downregulated after development, they are still needed to maintain normal physiology and are important regulators of aging in *C. elegans* (Lehrbach et al., 2012).

Several studies report that members of the same miRNA family (comprised of miRNAs that share a seed sequence but not necessarily a genomic locus) tend to have similar expression patterns and, presumably, act on similar targets (Ibáñez-Ventoso et al., 2006; Kato et al., 2009; Gu et al., 2007; Martinez et al., 2008). miRNAs in a genomic cluster, which sometimes contain identical seed sequences, are also frequently expressed at the same time, in agreement with the assumption that most of them are processed from a single transcript (Ibáñez-Ventoso et al., 2006; Kato et al., 2009; Gu et al., 2007). However, a global assessment of worm miRNA promoter:GFP fusions across all developmental stages (Martinez et al., 2008), showed that some exceptions to overlapping profiles of related and/or clustered miRNAs do occur. Examples include the miR-75, miR-232 and miR-251 families, whose members exhibit distinct spatial expression patterns. Extreme differences in miRNA levels have also been observed for some co-expressed miRNAs. The

seven members of the *miR-35-41* cluster are expressed in the embryo and seem to derive from a common transcript (Lau et al., 2001). However, they vary greatly in expression levels: deep sequencing data (Kato et al., 2009) uncovered over 100,000 reads for miR-35 but only 40 for miR-41. These striking differences point to post-transcriptional regulation with modulation of miRNA processing or decay.

Analysis of transgenic worms expressing GFP driven by predicted miRNA promoters is a powerful approach for determining spatiotemporal expression *in vivo* (Martinez et al., 2008; Isik et al., 2010) (Figure 2.2B). However, this strategy is limited by the risk of excluding important elements in the cloned promoter regions and the inability to account for post-transcriptional events that might regulate production of the functional mature miRNA. Despite these caveats, expression data from Northern blots mostly correlated with the temporal manifestation of the GFP reporters (Martinez et al., 2008). Interestingly, it was observed that although functional miRNA promoters were found in almost all worm tissues, 50% of individual promoters seemed to be active only in 3 or fewer tissues. The promoters of *let-7*, *lin-4* and *miR-53* were among the few that were able to confer ubiquitous somatic tissue expression (Martinez et al., 2008). Additionally, both studies (Martinez et al., 2008; Isik et al., 2010) characterized several intronic miRNAs with expression patterns distinct from that of their host genes, indicating that these embedded miRNAs are transcribed independently.

2.4 Identifying miRNA targets

2.4.1 Finding targets by genetic suppression

The first miRNA targets were identified through phenotypic suppression (Figure 2.3A). This was possible because mutations in the *lin-4* and *let-7* miRNA genes cause obvious loss-of-function phenotypes. Both *lin-4* and *let-7* participate in the developmental timing pathway (Lee et al., 1993; Reinhart et al., 2000; Kaufman and Miska, 2010) and disruption of either miRNA gene results in failure to develop adult structures and inability to lay eggs (Ambros, 2008; Mondol and Pasquinelli, 2012). Additionally, *let-7* mutants often die through vulval bursting (Mondol and Pasquinelli, 2012). Since the best characterized miRNA-target interactions result in negative regulation, decreased expression of target genes can sometimes compensate for the absence of miRNA activity. Thus, suppression of the miRNA mutant phenotype by a secondary mutation or by RNAi depletion can be used to reveal potential targets.

Multiple let-7 miRNA targets have been identified through forward and reverse genetic screens in *C. elegans*. An advantage of *let-7* is that there are several mutant alleles providing a range of phenotypes that can be scored in different experimental designs. These phenotypes vary from non-conditional lethality, where mature let-7 is completely absent, to temperature sensitive lethality, where mature let-7 contains a point mutation and is expressed at very low levels, to more mild defects in seam cell and alae development, where ~2-fold less mature let-7 is produced (Reinhart et al., 2000; Mondol and Pasquinelli, 2012). The *lin-41* gene was uncovered as a potent suppressor of the bursting and lethal phenotypes associated with *let-7* (Reinhart et al., 2000; Slack et al., 2000). The recognition of two closely

spaced sequences in the *lin-41* 'UTRs with partial complementarity to let-7 miRNA helped establish the model of imperfect base-pairing between miRNAs and targets. Remarkably, regulation of *lin-41* by let-7 seems to be broadly conserved across species (Pasquinelli et al., 2000; O'Farrell et al., 2008; Schulman et al., 2005; Lancman et al., 2005). Several other let-7 targets in *C. elegans,* such as *hbl-1* (Abrahante et al., 2003; Lin et al., 2003), *daf-12* (Grosshans et al., 2005) and *let-60/RAS* (Johnson et al., 2005) have been identified through phenotypic suppression using mutagenesis or RNAi screens. Combinatorial approaches that consider mis-regulation of mRNA or protein levels in *let-7* mutants, predicted let-7 binding sites and suppression of *let-7* phenotypes have yielded an assortment of targets important for different aspects of the *let-7* pathway (Hunter et al., 2013; Grosshans et al., 2005; Lall et al., 2006; Jovanovic et al., 2010; Ding et al., 2008). It is perhaps surprising that reduced activity of a single target can suppress the loss of a miRNA, which has many targets. This phenomenon suggests that there are complex cross-regulatory and feedback loops connecting the pathways regulated by miRNAs.

Suppression genetics is a valuable approach for finding miRNA targets that are biologically relevant. It also has the advantage of not relying on pre-conceived notions of target recognition rules. However, it can be difficult to distinguish between direct and indirect targets and further authentication of the target is necessary. Also, miRNAs can regulate the expression of many different targets and it is likely that some phenotypes are very complex and do not depend on a single candidate gene. Furthermore, it seems that *C. elegans* miRNAs are highly redundant and that most miRNA mutants do not have detectable phenotypes under standard lab conditions (Miska et al., 2007). These limitations point to the importance of developing sensitive

biochemical methods and accurate prediction algorithms to facilitate the identification of miRNA targets.

Table 2.1: Comparison between algorithms for miRNA target site prediction in *C. elegans*. *Stringent (S), Moderate (M).

Target Prediction Tool	Seed-pairing*	Conservation	Site accessibility	Heteroduplex stability	Site number	Reference
TargetScan	S	✓			✓	Jan et al. 2011
PicTar	S	✓		✓	✓	Lall et al. 2006
Miranda	M	✓		✓	✓	Betel et al. 2008
mirWIP	M	✓	✓	✓	✓	Hammell et al. 2008
PITA	M		✓	✓	✓	Kertesz et al. 2007
rna22	M			✓		Miranda et al. 2006

2.4.2 Computational prediction of targets

miRNAs act as guides for selective gene regulation by tethering the effector proteins of the RISC complex to their targets. The majority of known functional miRNA/target interactions is mediated by base pairing of the miRNA seed sequence (nucleotides 2-7) with a complementary site located in the 'UTRs of the target mRNA (Bartel, 2009). The significance of seed pairing is also reinforced by the fact that, to date, many studies involving high-throughput identification of miRNA targets report

enrichment of seed sequences in their experimental datasets (Zisoulis et al., 2010; Zhang et al., 2007, 2009; Hunter et al., 2013; Chi et al., 2009; Hafner et al., 2010; Leung et al., 2011; Jovanovic et al., 2012; Lim et al., 2005). Although perfect seed pairing is clearly an important motif for recognizing miRNA targets, it is not the only one. Some of the best studied *let-7* targets in *C. elegans*, *lin-41* (Reinhart et al., 2000) and *hbl-1* (Abrahante et al., 2003; Lin et al., 2003), contain functional target sites with seed mismatches. Imperfections in seed pairing can sometimes be compensated by a high degree of complementarity in the 3' end of the miRNA with its target. Nevertheless, there are several examples of miRNA target sites that rely on more centered or distributed interactions and the precise set of rules governing target recognition is still largely unknown (Pasquinelli, 2012).

Currently, computational prediction of miRNA targets mostly relies on perfect or nearly perfect complementarity, allowing for occasional mismatches or G:U wobbles, of the seed region with its 'UTRs target site (Bartel, 2009). However, seed-pairing predictions in animals generate hundreds to thousands of possible targets for a given miRNA, including many false positives (Thomas et al., 2010). To increase the reliability of predictions, targets with sufficient seed complementarity are usually scored according to secondary features of their sites, such as conservation, structural accessibility of the site in the 3'UTR (measured by the amount of AU rich regions surrounding the site), overall binding energy of the heteroduplex and number of sites for a given miRNA (Thomas et al., 2010; Bartel, 2009; Pasquinelli, 2012).

Some of the more commonly used algorithms for target prediction in *C. elegans* are: TargetScan (Jan et al., 2011; Lewis et al., 2005), PicTar (Lall et al., 2006), Miranda (Betel et al., 2008), PITA (Kertesz et al., 2007), rna22 (Miranda et al.,

2006) and mirWIP (Hammell et al., 2008) (Table 2.1). TargetScan and PicTar both rely on stringent seed pairing. However, while TargetScan looks for perfect seed complementarity, PicTar allows for imperfect pairing as long as the binding energy is within the cutoff. Miranda, PITA, mirWIP and rna22 have more moderate seed pairing requirements, being more tolerant towards mismatches or wobbles. TargetScan, PicTar, Miranda and mirWIP consider conservation of the target sites. PITA relies heavily on the stability of the heteroduplex and on the structural accessibility of the site within the 3'UTR. The rna22 algorithm uses a distinct system that evaluates shared miRNA sequence patterns and heteroduplex stability. mirWIP is also unique in the fact that it was based and trained on an experimental dataset of targets in *C. elegans*. Table 2.1 provides a summarized comparison among these prediction algorithms.

The application of bioinformatics for prediction of miRNA targets has been very useful for enabling the identification and experimental analysis of many new target sites (Maragkakis et al., 2009). Yet, these prediction algorithms walk a fine line between sensitivity of target identification and reliability of their predictions (e.g. emphasizing conservation of target sites can increase their reliability, but it can also prevent the identification of many non-conserved functional targets). Furthermore, the use of different 3'UTR and miRNA databases along with substantial biases in prediction parameters yields an unsatisfying level of overlap among candidate targets identified by current algorithms (Hua et al., 2009; Hammell et al., 2008; Ritchie et al., 2009). Also, non-canonical miRNA pairing and the presence of regulatory sites outside of the 3'UTR still represent major challenges in target prediction, creating a high demand for experimental approaches that can help define broader and more

accurate target recognition rules.

2.4.3 Biochemical detection of miRNA targets

- RNA immunoprecipitation and microarray (RIP-chip) assays

The ability to capture miRNA targets through biochemical methods has offered many insights into the dynamics of target selection (Figure 2.3B-C). These approaches provide a snapshot of mRNAs bound by miRISC *in vivo* that does not rely on our limited knowledge of miRNA target recognition rules. Results from these types of experiments can be used to optimize target prediction algorithms and reveal new features of miRNA targeting. In *C. elegans*, the first attempt to isolate potential miRNA targets used DNA microarrays to analyze transcripts that co-immunoprecipitated with proteins related to GW182 (RIP-chip) (Figure 2.3B) (Zhang et al., 2007). The GW182 family proteins associate with Argonaute and recruit factors that mediate the deadenylation and translational repression of miRNA targets (Braun et al., 2013). In *C. elegans* the GW182 family members, AIN-1 and AIN-2, are thought to have largely overlapping roles in binding the miRISC Argonautes ALG-1 and ALG-2 and directing target mRNA repression (Zhang et al., 2007; Ding et al., 2005). Over 3500 mRNAs were found to be enriched in AIN-1/2 IPs, including most of the previously validated worm miRNA targets such as *lin-14, lin-28, hbl-1* and *let-60/ras* (Zhang et al., 2007). This strategy has also been used to detect mRNAs associated with miRISC at different worm stages, which supported a broad role for miRNAs in regulating development and signaling processes (Zhang et al., 2009). More recently, IPs of tagged AIN-2 expressed in specific tissues were used to identify targets in the intestine and muscle (Kudlow et al., 2012). This strategy

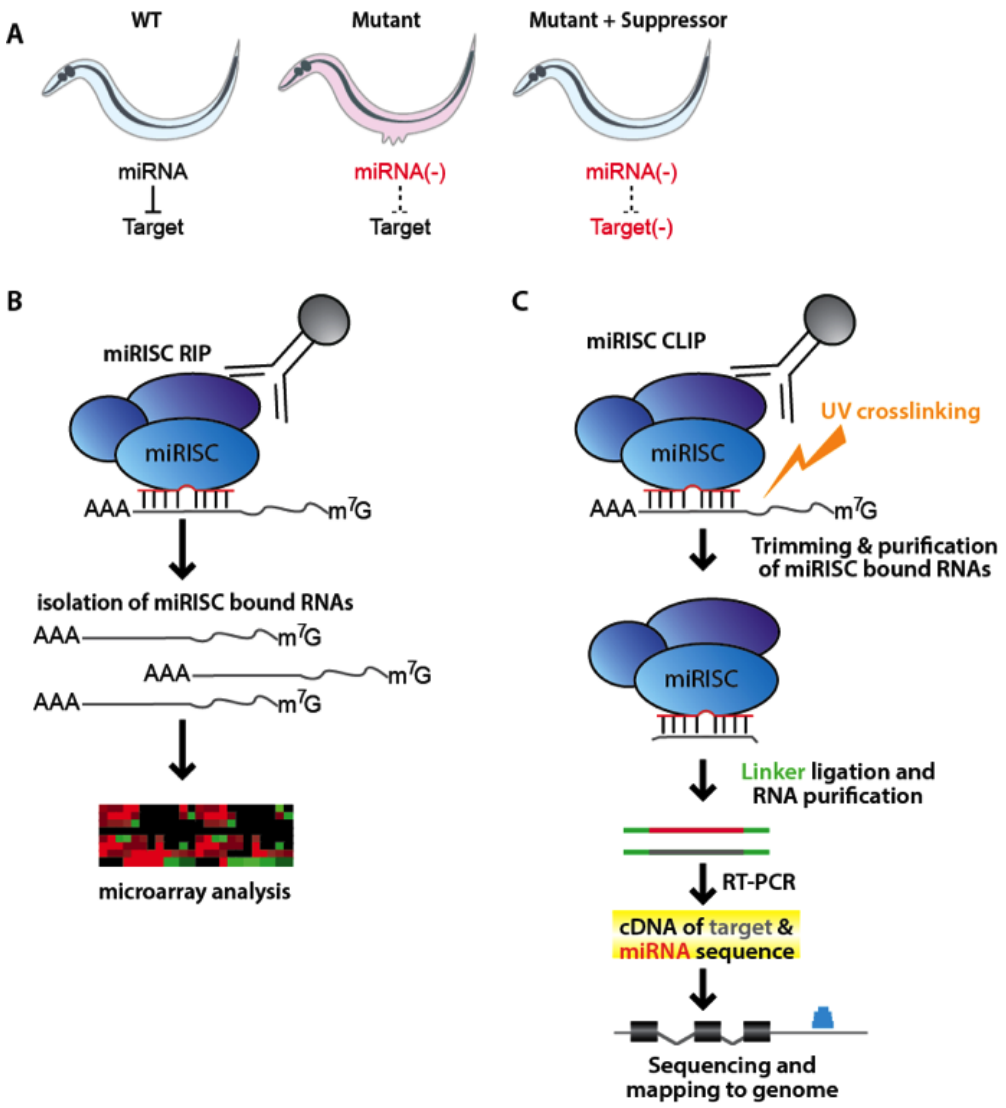


Figure 2.3: Experimental strategies for finding miRNA targets in *C. elegans.* (A) Suppressor genetics, (B) miRISC RIP-Chip, (C) miRISC CLIP-seq. See text for details.

revealed a strong bias for miRISC regulation of pathogen responsive genes in the gut and pointed to new roles for the miRNA pathway in host-pathogen interactions (Kudlow et al., 2012). One of the biggest caveats of RIP-chip approaches is that they are dependent on computer predictions to locate potential miRNA target sites within the isolated transcripts. Since these predictions can be biased towards particular pairing requirements and often only search for sites in 3'UTRs, critical features of endogenous miRNA target recognition can be missed. Also, the use of microarrays limits the pool of genes analyzed and prevents the identification of targets outside the scope of the chip, such as non-coding RNAs.

- Cross-linking immunoprecipitation of miRISC coupled with RNA sequencing (CLIP-seq)

Some of the sensitivity problems and biases in standard RIP approaches are solved when cross-linking is used to stabilize miRISC interactions for IP and deep sequencing is used to identify mRNA regions in direct contact with the complex (CLIP-seq) (Darnell, 2010) (Figure 2.3C). Cross-linking assures the fidelity of endogenous interactions and the use of RNA deep sequencing technology allows for a more global and unbiased identification of the interacting RNAs. The downside of this approach is that the ability to detect weaker interactions could lead to higher false-positive rates, as it is possible that Argonaute proteins "scan" sites that are not functional. The key innovation and advantage of CLIP-seq is that it reveals direct sites of interaction between miRISC and its targets. This simplifies prediction of the actual miRNA binding site and enables identification of non-canonical miRNA

binding patterns and locations. CLIP-seq of Argonaute proteins has uncovered thousands of targets sites in mouse brain (Chi et al., 2009), human cells (Hafner et al., 2010; Leung et al., 2011), and *C. elegans* (Zisoulis et al., 2010).

C. elegans offers several unique advantages for using CLIP-seq to identify endogenous miRNA target sites (Zisoulis et al., 2011). A strain with a mutation in *alg-1* that removes the epitope for the anti-ALG-1 antibody serves as a powerful negative control for distinguishing authentic interactions from background. A list of well-established miRNA targets provides a set of positive controls that should be identified in successful CLIP-seq experiments. Furthermore, a large collection of miRNA mutants offers a valuable resource for using CLIP-seq to determine which ALG-1 sites are dependent on specific miRNAs (Miska et al., 2007). By performing ALG-1 CLIP-seq in wild-type versus *alg-1* mutant worms at the L4 stage, almost 5,000 specific sites in over 3,000 genes were identified, including 10 out of 13 previously validated targets (Zisoulis et al., 2010). Compared to control, unbound sequences, ALG-1 interaction sites showed greater sequence conservation and accessibility and were associated with CU-rich motifs (Zisoulis et al., 2010). Moreover, this study provided the first global map of miRISC binding in a whole animal.

Recently, the more sensitive iCLIP (individual-nucleotide resolution CLIP) protocol has been optimized for *C. elegans* ALG-1 (Broughton and Pasquinelli, 2013). iCLIP follows the same basic principles of original CLIP-seq protocols with the addition of steps that improve the precision of mapping binding sites and greatly increase the recovery of target reads (König et al., 2010; Sugimoto et al., 2012). Efforts have also been made in developing techniques that will allow the recovery of the miRNAs paired with their target sites. In *C. elegans* one study attempted

to use miRNAs as reverse transcription primers in order to clone their target sites (Andachi, 2008). Although this approach identified a known target site, the broad applicability of this technique is yet to be established. In human cells, a method similar to CLIP-seq called CLASH (cross-linking, ligation and sequencing of hybrids) was modified to include a ligation step between the Argonaute bound miRNAs and their target RNA fragments (Helwak et al., 2013). Despite the inefficiency of this step (less then 2% of the reads were chimeras between miRNAs and targets), the method shows promise and provides further confirmation that many miRNA-target interactions are non-canonical and reside outside of 3'UTRs.

Currently, the IP-based methods for identifying miRNA targets provide no intrinsic information about the functional consequences of this interaction. Further validation by monitoring changes in RNA and protein levels is necessary to confirm an output of miRISC binding. Intriguingly, the results of ALG-1 CLIP-seq in *C. elegans* suggested that target recognition and regulation differ based on the position of the site. Consistent with CLIP in other species, the majority of the ALG-1 sites localized to 3'UTRs and coding exons. Target sites in 3'UTRs were highly enriched for the seed pairing motif and correlated with mRNA destabilization (Zisoulis et al., 2010). In contrast, neither of those features described targets with sites in coding exons. Thus, further investigations are needed to establish that those sites mediate regulation, perhaps at the translational level.

2.4.4 Support and validation for experimentally identified targets

The expression of biologically relevant targets is expected to change upon regulation by miRISC. Since target mRNA levels are often down-regulated in response to miRNA regulation (Pasquinelli, 2012), microarrays and RNA-seq are useful methods for globally analyzing changes in mRNA expression in the presence or absence of miRNA activity (Figure 2.4). A more quantitative assessment of select target mRNAs can be achieved through Northern blotting and qRT-PCR (Figure 2.4). Since loss of a specific miRNA is expected to produce direct and indirect effects, further validation of potential miRNA targets is required. Reporter genes fused to the 3'UTR of the predicted target should be assayed for the requirement of the complementary site as well as the miRNA for regulation. In these types of assays, care should be taken to test the reporter under physiological conditions. Transgenic over-expression of the reporter or miRNA or forced expression in non-native tissues can result in false conclusions about endogenous miRNA function.

There are two general methods, ribosome profiling and quantitative mass spectometry, for analyzing miRISC dependent changes in protein levels on a genome wide scale (Figure 2.4). Ribosome profiling (RP) offers an indirect approach to assess protein expression by monitoring the association of transcripts with translating ribosomes (Ingolia et al., 2009). Results from RP assays in mammalian cell culture led to the conclusion that mRNA degradation, as opposed to translational repression, is the predominant outcome of miRNA target regulation (Guo et al., 2010). This method was used to assess the contributions of mRNA destabilization and translational inhibition to the regulation of known miRNA targets important for devel-

opment in *C. elegans* (Stadler et al., 2012). Curiously, this high throughput approach led to conclusions about mRNA levels and ribosome association for some of these targets that seem to contradict previous more detailed studies (Bagga et al., 2005; Ding and Grosshans, 2009). Thus, multiple validation efforts should be utilized to assess the functional outcome of miRNA targeting.

The global analysis of miRNA dependent changes in protein levels through proteomics methods was first developed in mammalian cell culture (Figure 2.4) (Baek et al., 2008; Selbach et al., 2008). A more targeted proteomics approach has been used to detect miRNA targets in *C. elegans.* Selected reaction monitoring (SRM) was employed to quantify expression differences in proteins of interest from wild-type versus strains with mutations in specific miRNAs (Jovanovic et al., 2010, 2012). This method was combined with ALG-1 RIP assays to detect targets regulated by miR-58 at the mRNA and protein level (RIP-chip-SRM) (Jovanovic et al., 2012). Candidates dependent on miR-58 for association with ALG-1 were screened by quantitative proteomics and microarrays for differential expression in extracts from wild-type versus *miR-58* mutants. Based on these global assays, the authors concluded that miR-58 targets are largely resistant to destabilization and, instead, regulated primarily at the translational level (Jovanovic et al., 2012).

2.5 Integrating datasets and building functional networks

The biggest challenge faced by researchers sifting through the massive amount of data that has been generated in just over a decade of miRNA research

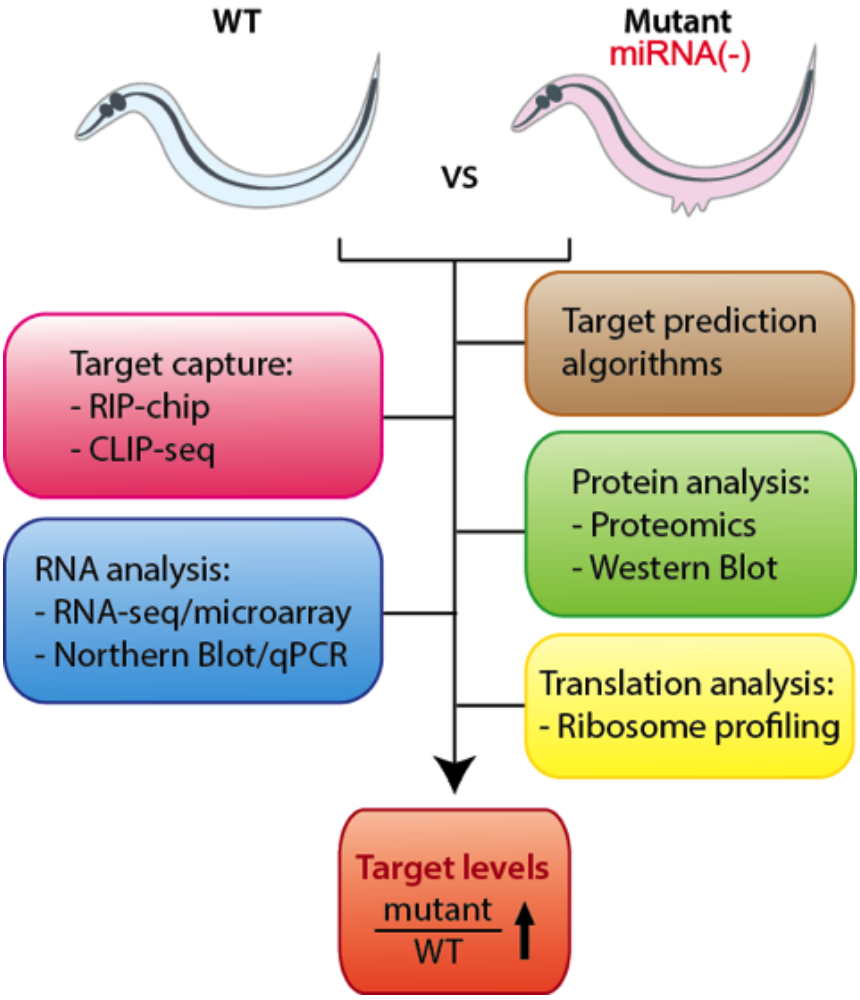


Figure 2.4: Combinatorial approaches for identifying and validating miRNA targets. Potential miRNA targets (predicted computationally or experimentally isolated) are tested for effects on protein or mRNA expression in wild-type versus mutant backgrounds to confirm regulation by a given miRNA.

is managing and integrating it. In order to facilitate this task, there have been efforts to validate and combine material from different sources. miRBase compiles a great amount of information on miRNAs, including the known miRNA sequences of several organisms as well as information on their expression, genomic location, stem-loop, family, cluster, and links to platforms that provide predicted or validated targets (Kozomara and Griffiths-Jones, 2011; Griffiths-Jones et al., 2006). One of these platforms is TarBase, a compilation of experimentally validated miRNA targets that summarizes the amount of evidence on the targets and provides links to the associated publications (Papadopoulos et al., 2009). Likewise, the 3'UTRome project compiles extensive data on *C. elegans* 3'UTRs, including the developmental stage(s) in which a certain 3'UTR was found and the method used to find it (Mangone et al., 2008, 2010). Another very useful approach for visualization of miRNA data is provided by starBase, which integrates several types of information in a genome browser format (Yang et al., 2011). This makes it possible for one to view numerous information tracks (e.g. CLIP-seq and Degradome-seq data, target predictions, 3'UTR annotations, etc.) in a combined and intuitive manner.

The modENCODE (model organism Encyclopedia of DNA Elements) project has aimed to systematically annotate functional genomic elements in flies and worms (Gerstein et al., 2010). As part of this project, Martinez et al. (2008) analyzed the interplay between miRNAs and transcription factors (TFs). Interestingly, by merging networks of TFs that bind miRNA promoters and TFs that could be regulated by miRNAs, the study proposed a very complex mesh of interactions and several types of regulatory motifs. Most striking among those were the 23 composite miRNA-TF feedback loops in which the TF that binds the promoter of a certain miRNA is

also a predicted target of that same miRNA (Martinez et al., 2008). Additionally, modENCODE has provided a wealth of data on the *C. elegans* transcriptome, including annotation of novel and known ncRNAs and assignment of transcription start sites (TSS). Also, several worm deep sequencing datasets have been generated to determine changes in expression of small RNAs upon development, aging and stress (modencode.org). Analyzing the modulation of miRNA levels and potential targets during those conditions could provide important clues to the functions of the many worm miRNAs whose biological roles are still unknown.

2.6 Conclusions

Our knowledge of miRNAs has evolved greatly considering the relatively short period of research since their discovery, and *C. elegans* has been instrumental in this progress. Numerous high quality studies have generated an enormous volume of data, calling for databases capable of integrating information from different experimental and computational sources. Despite the thousands of miRNAs now annotated in different organisms, the biological role of the majority of miRNAs is still largely unknown. Considering that the function of miRNAs is directly associated with their influence on gene expression, establishing reliable rules for target recognition remains a paramount challenge in the field. Complementary experimental and computational methods coupled with rigorous validation studies are essential for advancing the field and deciphering the complex miRNA-target interactions that shape most biological pathways.

2.7 Acknowledgements

We thank members of the Pasquinelli lab for critical review of this manuscript. This work was supported by funding from the NIH (GM071654), Keck, and Peter Gruber Foundations.

Chapter 2, in full, is a reprint of the material as it appears in Advances in Experimental Medicine and Biology by Sarah Azoubel Lima and Amy E. Pasquinelli, Springer, 2014. I was the primary author.

Chapter 3

Efficient translation promotes poly(A) pruning

3.1 Abstract

Poly(A) tails are important elements in mRNA translation and stability. However, recent genome-wide studies concluded that poly(A) length was generally not associated with translational efficiency in non-embryonic cells. To investigate if poly(A) tail size is coupled to gene expression in an intact organism, we used an adapted TAIL-seq protocol to measure poly(A) tails in *Caenorhabditis elegans*. Surprisingly, we found that well-expressed transcripts contain relatively short, well-defined tails. This attribute is dependent on translation efficiency, as transcripts enriched for optimal codons and ribosome association had the shortest tail sizes, while non-coding RNAs retained long tails. Across eukaryotes, short tails were a feature of abundant and well-translated mRNAs. Although this seems to contradict the dogma that deadenylation induces translational inhibition and mRNA decay,

it instead suggests that well-expressed mRNAs undergo pruning of their tails to accommodate a minimal number of poly(A) binding proteins, which may be ideal for protective and translational functions.

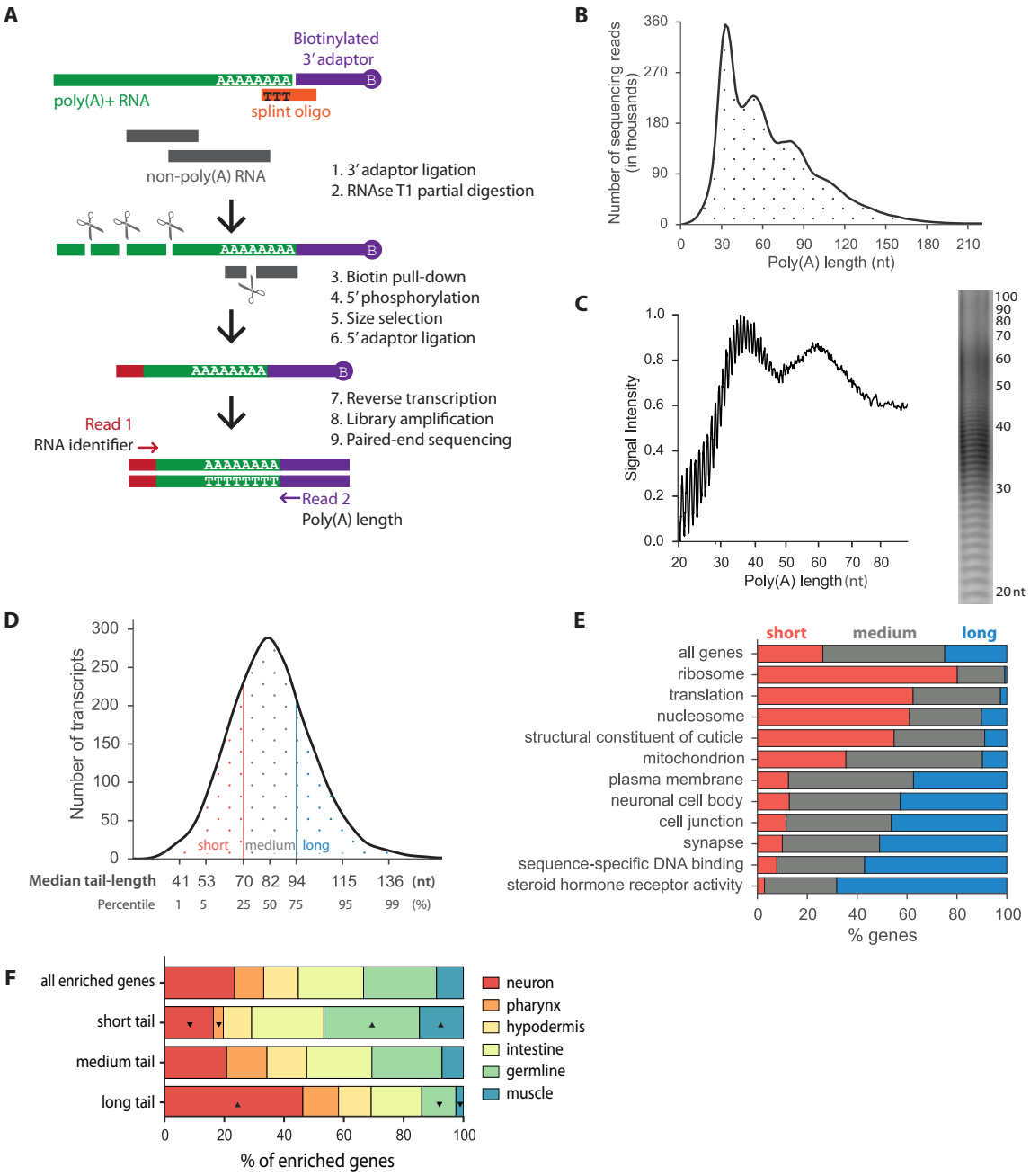
3.2 Main text

During transcriptional termination, the majority of eukaryotic mRNAs undergo polyadenylation, resulting in a 3' tail estimated to contain ~90 (yeast) or ~250 (animals) adenosines (Mangus et al., 2003). The poly(A) tail has been shown to be important for protection and translation of the mRNA (Goldstrohm and Wickens, 2008; Roy and Jacobson, 2013). These roles are largely mediated by poly(A) binding proteins (PABPs), which coat the tail. The direct interaction of PABP with the 5' cap binding complex factor eIF4G is thought to promote mRNA stability and translation by supporting formation of the closed-loop state (Goldstrohm and Wickens, 2008; Mangus et al., 2003; Roy and Jacobson, 2013). Conversely, PABP also binds deadenylation complexes (CCR4/NOT/Tob and PAN2/PAN3) and contributes to microRNA-mediated repression (Jonas and Izaurralde, 2015; Wahle and Winkler, 2013; Xie et al., 2014). These seemingly contradictory roles of PABP suggest that poly(A) tail length and, hence, the number of bound PABPs might determine mRNA fate. In early embryos and other cellular contexts, regulated cytoplasmic polyadenylation lengthens the tails of select mRNAs, resulting in their translational activation (Charlesworth et al., 2013; Jalkanen et al., 2014). Yet, recent studies that measured poly(A) tails of individual transcripts genome-wide did not identify a general association between tail size and translational efficiency in somatic cells

(Chang et al., 2014; Park et al., 2016; Subtelny et al., 2014). Since cellular context can regulate poly(A) size and function, we asked if tail size was associated with stability and translation of mRNAs in an intact animal.

Two distinct high-throughput sequencing methods have been developed to assay global poly(A) tail sizes: TAIL-seq (Chang et al., 2014) and PAL-seq (poly(A)-tail length profiling by sequencing) (Subtelny et al., 2014). We adapted the TAIL-seq protocol to analyze poly(A) tails in *Caenorhabditis elegans* because it utilizes a standard and direct sequencing platform. However, the TAIL-seq method relies on costly bead-based ribosomal RNA (rRNA) removal procedures that are ineffective or unavailable for many organisms, including *C. elegans*. Therefore, it was necessary to modify TAIL-seq to minimize contamination by rRNAs. Inspired by the PAL-seq method, we used a splint ligation approach, in which a DNA oligo bridges the last 9 adenosines of the poly(A) tail and the 3' adaptor, greatly favoring the ligation reaction of poly(A)+ RNAs over non-adenylated transcripts (Figure 3.1A). Our adapted TAIL-seq (aTAIL-seq) method produces reliable and reproducible libraries (Supplemental Figure 3.6, A to C), requires less starting material, and can be readily applied to measure poly(A) tails in any organism.

Figure 3.1: The *C. elegans* poly(A) profile. (A) Outline of aTAIL-seq procedure. A splint oligo is used to select for polyadenylated RNAs and exclude other RNA contaminants. (B and C) Global size distribution of *C. elegans* poly(A) tails measured by aTAIL-seq (B) and bulk poly(A) labeling (C). (D) Distribution of median poly(A) tail-length per gene. Genes with a median tail $\leq$ 70 nt were categorized as short-tailed, genes with a median tail $>$ 70 and $\leq$ 94 nt were categorized as medium-tailed and genes with a median tail $>$ 94 nt were categorized as long-tailed. (E) Functional annotations (Gene Ontology terms) significantly enriched for genes with short or long tails. The colored bars represent the percent of members in each tail-length category. (F) Tissue enrichment profiles for genes with short, medium or long tails. ▲ significant enrichment; ▼ significant depletion for a tissue category ($p<0.01$, Fisher test).


A
Biotinylated
3′ adaptor
AAAAAAAA
B
poly(A)+ RNA
TTT
splint oligo
non-poly(A) RNA
1. 3′ adaptor ligation
2. RNAse T1 partial digestion
3. Biotin pull-down
4. 5′ phosphorylation
5. Size selection
6. 5′ adaptor ligation
7. Reverse transcription
8. Library amplification
9. Paired-end sequencing
Read 1
RNA identifier
TTTTTTTT
Read 2
Poly(A) length
B
Number of sequencing reads (in thousands)
Poly(A) length (nt)
C
Signal Intensity
Poly(A) length (nt)
20 nt
D
Number of transcripts
short
medium
long
Median tail-length
Percentile
(nt)
(%)
E
short
medium
long
all genes
ribosome
translation
nucleosome
structural constituent of cuticle
mitochondrion
plasma membrane
neuronal cell body
cell junction
synapse
sequence-specific DNA binding
steroid hormone receptor activity
% genes
F
all enriched genes
short tail
medium tail
long tail
% of enriched genes
neuron
pharynx
hypodermis
intestine
germline
muscle

We used aTAIL-seq to investigate the poly(A) tail lengths of transcripts produced during the last larval stage of worm development (L4). We found that 90% of all individual mRNA molecules have tail lengths between 26 and 132 nucleotides (nt) and the median overall poly(A) length is 57 nt (Figure 3.1B). These sizes are comparable to the bulk tail lengths measured in mammalian (Chang et al., 2014; Subtelny et al., 2014) and *Drosophila* cells (Subtelny et al., 2014). Interestingly, the most abundant species of polyadenylated mRNAs were 33-34 nt (Figure 3.1B), which is close to the reported 25-30 nt footprint for a single PABP (Baer and Kornberg, 1983; Smith et al., 1997; Wang et al., 1999). Additionally, we observed a phasing pattern with peaks at the poly(A) sizes expected to occur with serial binding of PABP (Figure 3.1B), suggesting removal of unprotected 3' adenosines. Furthermore, the sharp drop in frequency of mRNAs with tail-lengths under 30 nt indicates that the minimal tail length required for stability corresponds to the size of one PABP footprint. We validated this phasing pattern with a ~34 nt peak by direct labeling and visualization of bulk poly(A) tails from total *C. elegans* RNA (Figure 3.1C), which was consistent with previous poly(A) profiling of nematode RNA by this method (Nousch et al., 2013).

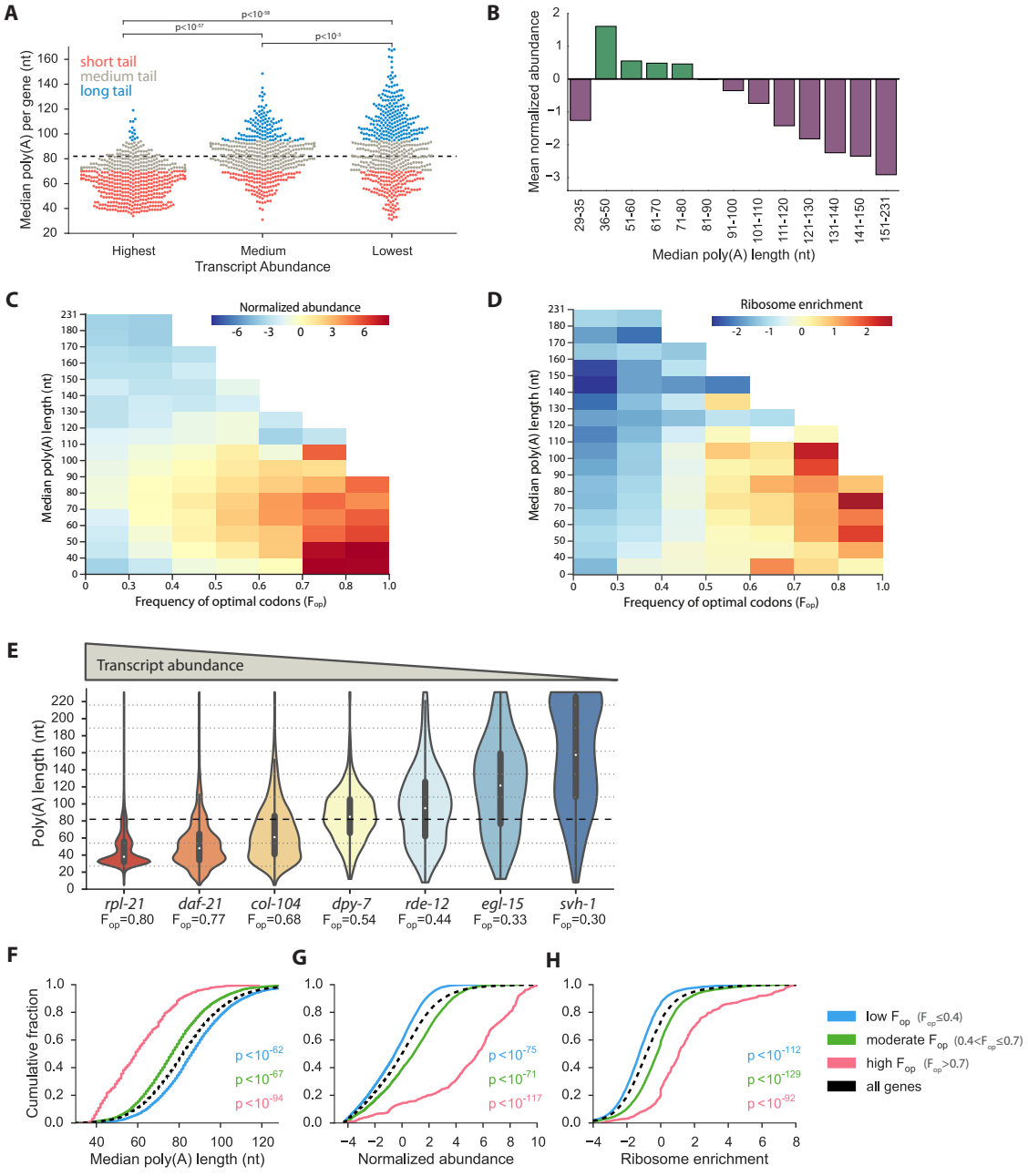
The aTAIL-seq method allowed us to analyze the tail distributions and median tail lengths of 13,601 protein coding gene transcripts with 10 or more poly(A) measurements. Within this comprehensive dataset, the most frequent median poly(A) length was 82 nt, with 90% of mRNAs having median tails ranging between 53 and 115 nt (Figure 3.1D). To investigate if there were functional classes of genes that tend to have longer or shorter poly(A) tails, genes were sorted according to their median tail lengths. We classified the quartiles of genes with the shortest (short: $\leq$

70 nt) and longest (long: $>$ 94 nt) median poly(A) tails (Figure 3.1D) and searched for enriched gene ontology (GO) terms within each category. Short-tailed transcripts were highly enriched for genes involved in translation, nucleosome components, and cuticular collagens (Figure 3.1E). Conversely, long-tailed transcripts were enriched for genes with regulatory functions, such as transcription factors, signal transduction proteins, mediators of neuronal activity, and hormone receptors. This difference in GO term enrichment for genes encoding transcripts with short and long median poly(A) tail lengths was also evident at the cellular level. Remarkably, many long tailed genes were predicted to be specific to neurons, whereas short-tailed transcripts were enriched for genes with germline and muscle expression (Figure 3.1F).

As shortening of the poly(A) tail is thought to promote mRNA instability (Goldstrohm and Wickens, 2008; Jalkanen et al., 2014; Roy and Jacobson, 2013), we were surprised to find that short-tailed transcripts were associated with highly expressed genes, such as those encoding ribosomal proteins (Figure 3.1E). However, this pattern would explain the disparity between the median tail of the global mRNA pool (57 nt) and the median poly(A) size per transcript (82 nt). Our finding indicates that the transcripts associated with short tails are very abundant, thus skewing the global poly(A) profile towards shorter poly(A) lengths (Figure 3.1, B and D). To compare steady state transcript levels to poly(A) size, we plotted the median tail lengths of mRNAs categorized by relative abundance (Figure 3.2A). This analysis revealed that the majority of highly expressed transcripts contained short tails, whereas the least abundant transcripts had longer tail distributions. When we binned genes according to median poly(A) tail lengths, we observed a striking

inverse correlation between poly(A) size and transcript abundance (Figure 3.2B and Supplemental Table 3.1). The mRNAs with shorter median poly(A) tail lengths were, on average, much more abundant than those with the longest tails. The rare exception was the small group of 33 transcripts with median tails in the 29-35 nt range, where many RNAs likely contain tails too short to accommodate a single PABP and are undergoing active degradation. This strong inverse relationship between tail length and transcript abundance was unexpected, as it is generally thought that longer tails are associated with stable and highly expressed RNAs (Goldstrohm and Wickens, 2008; Jalkanen et al., 2014).

Figure 3.2: Highly expressed mRNAs have short poly(A) tails. (A) Tail-length distribution is different in pools of genes with distinct expression levels. The transcript abundance categories represent the highest expressed genes, those closest to the median expression, and lowest expressed. All three distributions were significantly different (Mann-Whitney U test). (B) Global relationship between poly(A) length and abundance was measured by plotting the mean normalized abundance of bins of genes divided by median tail lengths. (C and D) Heat maps demonstrating the interplay of the frequency of optimal codons (F_{op}) and tail size with transcript abundance (C) and ribosome enrichment (Hendriks et al., 2014) (D). (E) Violin distribution plots with inlaid box-plots (white dot represents the median) of all tail-length measurements in genes with different frequencies of optimal codons (F_{op}) and abundance levels. (F to H) *C. elegans* genes were classified according to codon optimization, demonstrating a significant relationship between translation efficiency and the cumulative distribution of poly(A) length (F), transcript abundance (G) and ribosome enrichment (Hendriks et al., 2014) (H). Normalized abundance is calculated as the log2 of the fold-change of the number of tags in a transcript over the median transcript level. P-values were calculated using the Mann-Whitney U test between each codon optimization category and all genes sampled.


A
p<10-58
p<10-57
p<10-3
short tail
medium tail
long tail
Median poly(A) per gene (nt)
Highest
Medium
Lowest
Transcript Abundance
B
Mean normalized abundance
29-35
36-50
51-60
61-70
71-80
81-90
91-100
101-110
111-120
121-130
131-140
141-150
151-231
Median poly(A) length (nt)
C
Normalized abundance
Median poly(A) length (nt)
Frequency of optimal codons (Fop)
D
Ribosome enrichment
Median poly(A) length (nt)
Frequency of optimal codons (Fop)
E
Transcript abundance
Poly(A) length (nt)
rpl-21 Fop=0.80
daf-21 Fop=0.77
col-104 Fop=0.68
dpy-7 Fop=0.54
rde-12 Fop=0.44
egl-15 Fop=0.33
svh-1 Fop=0.30
F
Cumulative fraction
p<10-62
p<10-67
p<10-94
Median poly(A) length (nt)
G
p<10-75
p<10-71
p<10-117
Normalized abundance
H
p<10-112
p<10-129
p<10-92
Ribosome enrichment
low Fop (Fop≤0.4)
moderate Fop (0.4<Fop≤0.7)
high Fop (Fop>0.7)
all genes

We next asked if poly(A) tail size was associated with translation efficiency. In general, the ribosome occupancy and frequency of optimal codons in a given mRNA are indicators of its translation status (Ingolia, 2014; Presnyak et al., 2015; Quax et al., 2015). Additionally, it was recently shown that *Saccharomyces cerevisiae* transcripts with optimized codons have higher rates of translation elongation and are more stable than genes with suboptimal codons (Presnyak et al., 2015). Consistent with this report, we found that in *C. elegans* the most abundant transcripts were enriched for optimal codons (Figure 3.2C) and ribosome association (Figure 3.2D), using data from previously published ribosome profiling studies (Hendriks et al., 2014). Moreover, these favored translation substrates were strongly biased towards short poly(A) tails (Figure 3.2, C and D and Supplemental Table 3.1). However, for these genes, and almost all others, we were still able to detect transcripts with tail lengths consistent with the very long ($>$200 nt) poly(A) tails synthesized on nascent mRNAs (Eckmann et al., 2011). Specifically, we detected molecules with tail sizes $\geq$200 nt for 78% and $\geq$160 nt for 90% of all genes assayed (Supplemental Figure 3.7A). More variability was observed for the minimum and overall range of poly(A) tail sizes of mRNAs (Supplemental Figure 3.7, B and C). The finding that genes with the highest frequencies of optimal codons were represented by mRNAs that spanned the entire range of detectable tail sizes but were strongly biased for short tailed species (Figure 3.2E and Supplemental Figure 3.7C), suggests that poly(A) tail shortening to an optimal length, which we refer to as pruning, may be related to translational efficiency.

Examination of the distribution of poly(A) tail lengths for individual genes revealed distinct patterns based on transcript abundance and codon composition

(Figure 3.2E). For highly expressed and codon-optimized genes such as *rpl-21* (a ribosomal protein) and *daf-21* (HSP90 - a molecular chaperone), tail lengths ranged from 5-231 nt but concentrated prominently around lengths that would accommodate 1-2 PABPs (~30-60 nt). In contrast, less abundant mRNAs with poorly optimized codons, such as *egl-15* (fibroblast growth factor receptor) and *svh-1* (neuronal growth factor), tended to have much longer and more diffusely distributed poly(A) tail sizes. On a genome-wide scale, we observed significant differences in the distribution of median poly(A) lengths, abundance and ribosome enrichment for transcripts containing low, medium, and high levels of optimal codons (Figure 3.2, F to H).

To further investigate the relationship between translation and poly(A) tail pruning, we focused on a set of mRNAs undergoing translational activation or repression, using published RNA-seq and ribosome profiling time course data for *C. elegans* (Hendriks et al., 2014). During a two-hour window that spans the time point we used for aTAIL-seq, transcripts for 365 genes become at least 8-fold enriched while those for 341 genes become at least 8-fold depleted from ribosomes, after normalization to changes in mRNA abundance. Remarkably, the ribosome enriched transcripts, and presumably more actively translated group, had significantly shorter median poly(A) tail sizes compared to the transcripts undergoing translational repression (Figure 3.3A). Further evidence suggesting an inverse relationship between poly(A) tail size and translation surfaced from our aTAIL-seq analysis of annotated long non-coding RNAs (lncRNAs) (Nam and Bartel, 2012). In general, lncRNAs, including antisense RNAs, had long poly(A) tails and showed no evidence of the phasing seen for mRNAs (Figure 3.3B). Taken together, our

findings suggest that translation regulates pruning of poly(A) tails, either by actively promoting their shortening or by stabilizing the short-tailed mRNAs.

The association of abundant and highly translated mRNAs with short poly(A) tails seems to be conserved across eukaryotes. We analyzed published datasets for poly(A) tail lengths (Subtelny et al., 2014), ribosome enrichment (Subtelny et al., 2014), RNA stability (Pelechano et al., 2015; Presnyak et al., 2015; Schwanhäusser et al., 2011) , and translation (Schwanhäusser et al., 2011) for *S. cerevisiae, Drosophila* and mouse transcripts. We observed that highly translated mRNAs tended to have shorter tails (Figure 3.4, A and B, and Supplemental Figure 3.8A and Supplemental Table 3.1), higher steady state expression levels (Figure 3.4C and Supplemental Figure 3.8B and Supplemental Table 3.1), and longer half-lives (Figure 3.4D and Supplemental Table 3.1). Notably, the shorter relative median tail length of transcripts encoding ribosomal proteins was well conserved among the different organisms (Supplemental Figure 3.9A). Additionally, in the *C. elegans* dataset this class of mRNAs exhibited highly uniform median tail lengths of ~40 nt (Supplemental Figure 3.9A), with the largest fraction of tails sized to accommodate one, and to a lesser extent, two PABPs (Supplemental Figure 3.9B). Overall, these results suggest that pruned poly(A) tails are a feature of stable and efficiently translated mRNAs across species.

Initially, it was puzzling to find that the class of relatively unstable and poorly translated mRNAs had the longest median poly(A) tail sizes (Figures 3.2 and 3.4). One possibility is that this pool mainly consists of recently synthesized transcripts that have not yet been targeted for rapid decay. In yeast, unstable mRNAs have been shown to undergo rapid deadenylation to a ~10 nt oligo(A) tail length, followed by

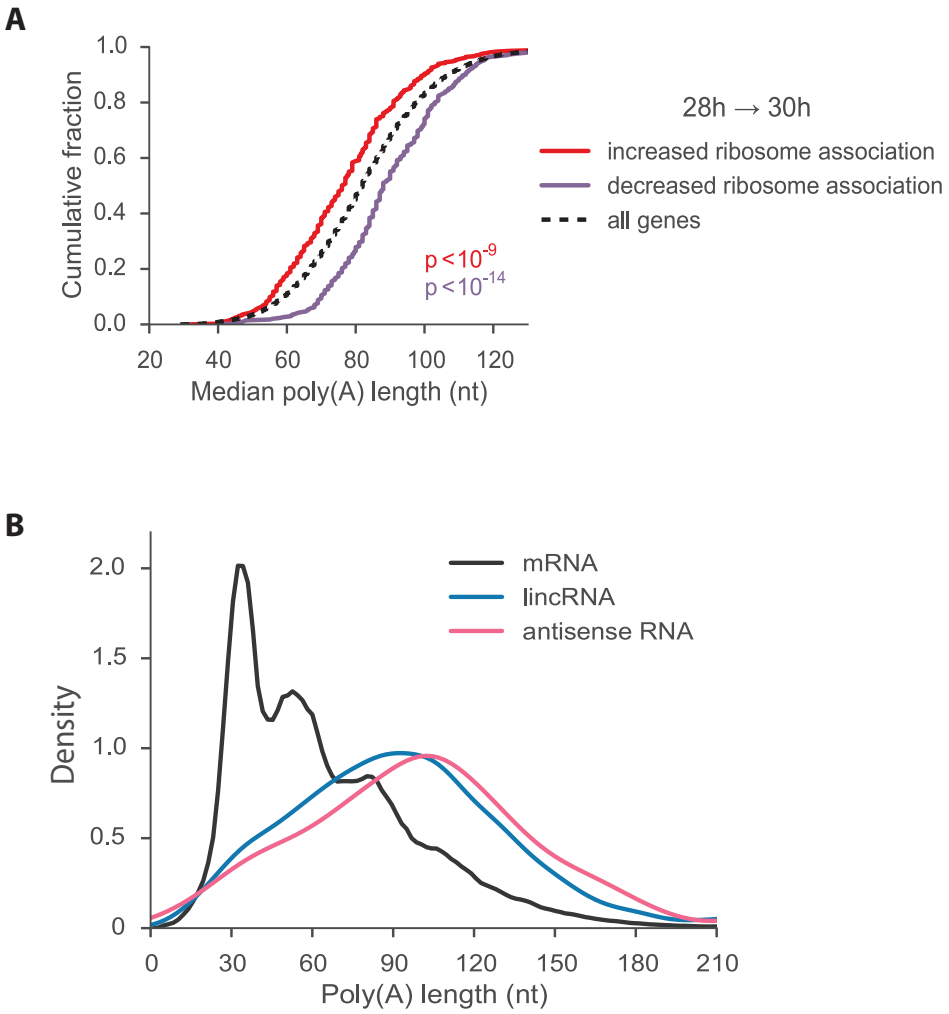
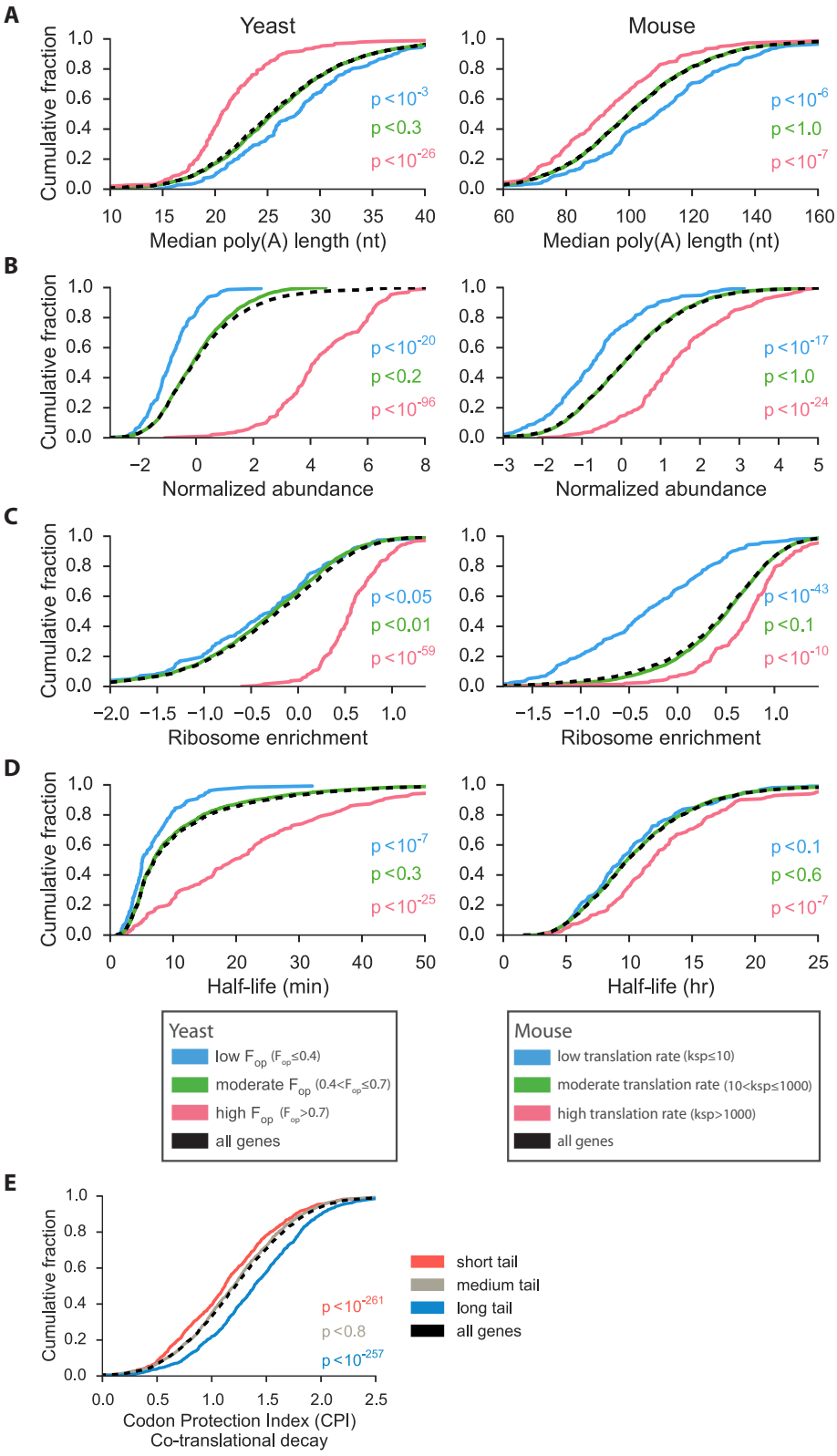


Figure 3.3: Translation promotes poly(A) tail pruning. (A) Cumulative median tail length distributions of genes that are enriched or depleted in the ribosomes (at least 8 fold) over a 2 hour period (Hendriks et al., 2014) that spans the time point used for aTAIL-seq (29 h). P-values were calculated using the Mann-Whitney U test between each category and all genes sampled. (B) Density plot comparing the bulk distribution of poly(A) tails between mRNAs and two classes of long non-coding RNAs: lincRNAs (long intervening non-coding RNAs) and antisense RNAs.

decapping and 5' to 3' exonucleolytic degradation (Coller and Parker, 2004; Decker and Parker, 1993). Although decay intermediates are rare in wild type cells (Hsu and Stevens, 1993; Hu et al., 2009), a recent study used deep sequencing methods (5PSeq) to identify decapped yeast mRNAs on a genome wide scale (Pelechano et al., 2015). Using published 5PSeq datasets for yeast mRNAs (Pelechano et al., 2015), we found that genes for transcripts with long median tails were represented by the highest levels of 5' decapped mRNAs (Figure 3.4E and Supplemental Table 3.1). The 5' decay intermediates only accounted for ~12% of cellular RNAs that could be captured by oligo(dT) isolation methods (Pelechano et al., 2015), which is consistent with the brief existence of decapped RNAs in wild type cells (Hsu and Stevens, 1993; Hu et al., 2009). Thus, many transcripts in the "long" poly(A) tail class may actually be detected in a transient state prior to rapid destabilization. Conversely, most "short" class transcripts seem to be those that accumulate with pruned poly(A) tails.

Figure 3.4: Poly(A) tail pruning is conserved in yeast and mouse. (A to D) Cumulative distribution plots showing the relationship between translation levels and poly(A) length (Subtelny et al., 2014) (A), ribosome enrichment (Subtelny et al., 2014) (B), transcript abundance (Subtelny et al., 2014) (C), and transcript half-lives (D) in *S. cerevisiae* (Presnyak et al., 2015) and mouse NIH3T3 (Schwanhäusser et al., 2011) cells. In mouse cells, translation rate constants (ksp) represent the number of proteins synthesized per mRNA per hour (Schwanhäusser et al., 2011). In yeast, translation rates are reflected in the codon optimization of the transcripts. (E) Relationship between poly(A) tail size and co-translational decay in yeast transcripts. Higher CPI (Codon Protection Index) values correspond to higher rates of co-translational 5' decapping (Pelechano et al., 2015). P-values were calculated using the Mann-Whitney U test between each tail size category (short = 1st quartile; medium = 2nd and 3rd quartiles; long = 4th quartile, based on median length) and all genes sampled.


A
Yeast
Mouse
Cumulative fraction
Median poly(A) length (nt)
p<10-3
p<0.3
p<10-26
p<10-6
p<1.0
p<10-7
B
Normalized abundance
p<10-20
p<0.2
p<10-96
p<10-17
p<1.0
p<10-24
C
Ribosome enrichment
p<0.05
p<0.01
p<10-59
p<10-43
p<0.1
p<10-10
D
Half-life (min)
Half-life (hr)
p<10-7
p<0.3
p<10-25
p<0.1
p<0.6
p<10-7
Yeast
low Fop (Fop≤0.4)
moderate Fop (0.4<Fop≤0.7)
high Fop (Fop>0.7)
all genes
Mouse
low translation rate (ksp≤10)
moderate translation rate (10<ksp≤1000)
high translation rate (ksp>1000)
all genes
E
Codon Protection Index (CPI)
Co-translational decay
p<10-261
p<0.8
p<10-257
short tail
medium tail
long tail
all genes

Previous poly(A) tail sequencing studies concluded that tail length was not associated with translational efficiency in non-embryonic cells (Chang et al., 2014; Park et al., 2016; Subtelny et al., 2014). However, the PAL-seq study reported that in yeast and mouse NIH3T3 cells tail sizes and measures of translation rates were negatively correlated (Rs = -0.12, $P<10\text{-}9$ (*S. cerevisiae)*; Rs = -0.20, $P<10\text{-}16$ (mouse)), findings confirmed in our analyses (Figure 3.4A and Supplemental Table 3.1). Additionally, the classes of transcripts found to have long or short tails by our study and PAL-seq (Subtelny et al., 2014) are largely in agreement, with short-tailed transcripts generally considered to be among the most abundant and well translated in the cell. Presently, it is unclear why other analyses of poly(A) tail size on individual genes in yeast or NIH3T3 cells found that ribosomal protein and other abundantly expressed transcripts had relatively long tails (Beilharz and Preiss, 2007; Chang et al., 2014). Those conclusions are at odds with single-gene Northern Blot or PCR based assays that have detected relatively short poly(A) tails on ribosomal protein mRNAs in yeast (Brown and Sachs, 1998; Subtelny et al., 2014; Tucker et al., 2001), mouse NIH3T3 cells (Wong et al., 2010), and worms (Supplemental Figure 3.9C). It is possible that, like with translation, gene specific control of poly(A) tail length is sensitive to differences in cellular contexts (Gowrishankar et al., 2006; Hilgers et al., 2006; Kleene et al., 2003).

Here we provide genome-wide evidence that short poly(A) tail sizes are a feature of abundant and efficiently translated mRNAs across eukaryotes. Although this challenges the longstanding idea that longer tails promote mRNA stability and translation (Eckmann et al., 2011; Goldstrohm and Wickens, 2008; Wahle and Winkler, 2013), it suggests that instead there might be an optimal tail size that results

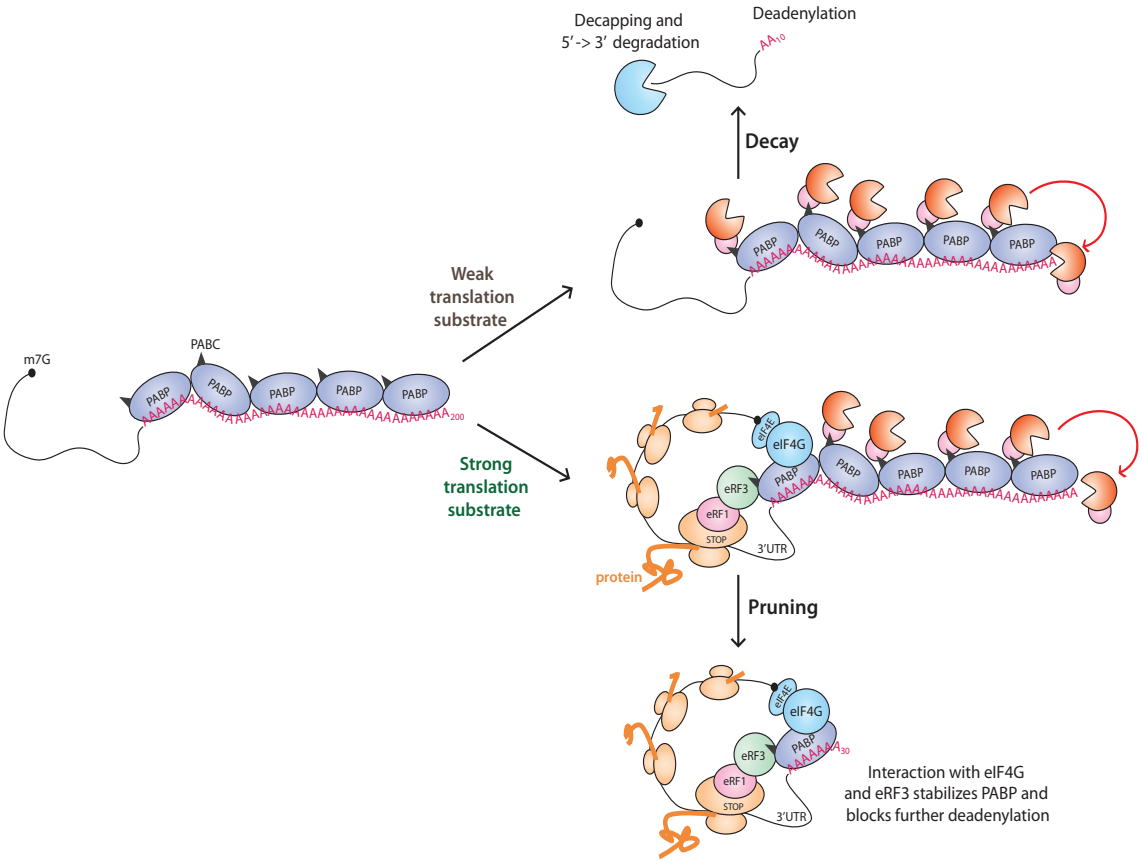


Figure 3.5: Poly(A) tail pruning model. Newly transcribed mRNAs receive long (>200 nt) tails, which are coated with PABP. The PABP C-terminal domain (PABC, black triangles) binds the CCR4/NOT/Tob and PAN2/3 deadenylation complexes (Wahle and Winkler, 2013; Xie et al., 2014). In strong translation substrates, interactions between a proximal PABP and translation initiation factor eIF4G promote a closed-loop structure and the translation termination factor eRF3 may compete with the deadenylases for binding the PABC domain (Hoshino, 2012; Xie et al., 2014). These interactions are predicted to stabilize the proximal PABP and prevent processive deadenylation of the transcript, allowing the tail to be pruned to an ideal length for mRNA stabilization and translation. Trimming of the poly(A) tail to limit the number of associated PABPs may be important for removing binding sites for factors that catalyze deadenylation and translational repression. For weak translation substrates, the deadenylases recruited to the PABC domain can act processively without the impediment of stabilizing interactions provided by translational activity, resulting in critically short tails that trigger decapping and 5'→3' degradation of the mRNA.

from a shortening process we refer to as pruning. Since poorly translated mRNAs and non-coding transcripts were found to contain long, less defined poly(A) tails, pruning seems to be associated with translational activity. We propose that the initially long poly(A) tails on newly synthesized transcripts become deadenylated to different extents depending on translational status: for well translated mRNAs, tail shortening ceases at lengths that accommodate a minimal number of PABPs, and for inefficiently translated mRNAs, deadenylation progresses to critically short lengths that trigger decapping and rapid mRNA decay (Figure 3.5). Processive deadenylation may result when the last PABP is dislodged from the poly(A) tail, and efficient translation may antagonize this event by stabilizing the PABP-poly(A) tail association, perhaps through direct interactions with initiation (eIF4G) and termination (eRF3) factors (Hoshino, 2012; Mangus et al., 2003). Numerous studies have pointed to dual, seemingly contradictory, roles for PABP in regulating mRNA stability. Whereas binding of PABP can protect the poly(A) tail from degradation (Coller and Parker, 2004; Coller et al., 1998; Decker and Parker, 1993), it also has been shown to recruit the major deadenylase complexes PAN2/PAN3 and (Hoshino, 2012; Wahle and Winkler, 2013; Xie et al., 2014). We predict that the multiple PABPs bound to initially long tailed transcripts engage deadenylation factors that either reduce the tails to lengths that exclude PABP binding, resulting in rapid decay, or that stall at short tail sizes bound by a minimal number of PABPs stably associated with actively translated mRNAs (Figure 3.5). While consistent with the well-established connection between translation and mRNA decay (Presnyak et al., 2015; Roy and Jacobson, 2013; Shoemaker and Green, 2012), this model implicates an optimal poly(A) tail length that is achieved through translational activity and, in turn, contributes to the stability

and efficient decoding of the mRNA.

3.3 Materials and methods

3.3.1 Nematode culture and RNA extraction

WT*Caenorhabditis elegans* (N2 Bristol) animals were cultured on OP50 bacteria at 25 °C, and collected at the last larval stage (mid-L4; 29 h time point). Standard worm synchronization methods were used (Porta-de-la Riva et al., 2012). RNA was extracted with Trizol and DNAse treated. RNA quality was measured by 260/280 ratio and confirmed by gel electrophoresis.

3.3.2 Bulk poly(A) labeling

1 µg total RNA (DNAse treated) was 3' labeled by performing a 3' ligation reaction containing 20 U T4 RNA Ligase (NEB) and 1 µM [32P]pCp (Perkin Elmer) overnight at 16 °C. Enzymes were inactivated at 68 °C for 5 min and unincorporated nucleotides were removed with MicroSpin G-50 columns (GE Healthcare). Labeled RNA was digested with 80 U RNAse T1 and 4 µg RNAse A (which cannot act on the poly(A) tail) for 2 h at 37 °C; 40 µg unlabeled yeast RNA was used as ballast. The reaction was stopped by Proteinase K digestion of the RNAses and the labeled poly(A)s were extracted with acid-phenol:chloroform:IAA and ethanol precipitated. Labeled poly(A) tails were resuspended in 20 µL, of which 5 µL were run on a long 15% Urea-PAGE sequencing gel along with labeled Decade RNA Marker (Ambion). The gel was dried onto whatman paper and scanned on a PhosphorImager.

3.3.3 Poly(A) analysis by northern blot

As detailed in Sallés et al. (1999), total RNA samples were digested with RNAse H (NEB) in the presence of a gene specific complementary oligonucleotide and, in the case of poly(A)- samples, also oligo$(dT)_{18}$. The samples were then resolved on a 6% Urea-PAGE minigel along with RNA Century marker (Ambion) for size determination of the fragments. Northern blotting was performed as described in Van Wynsberghe et al. (2011).

3.3.4 aTAIL-seq

aTAIL-seq was performed as in the original TAIL-seq (Chang et al., 2014), with the following modifications. *3' adaptor splint ligation*: A splint oligonucleotide was used to favor capture of poly(A)+ RNAs. We incubated 20 ug of total RNA in 5 uL with 1 uL 10 uM biotinylated 3' adaptor and 1 uL 10 uM splint oligonucleotide (5' -NNNGTCAGTTTTTTTTT-3') at room temperature for 5 min. Next, 1 uL 10X RNA ligase buffer (NEB), 0.5 uL of Superase-In (Ambion) and 1 uL T4 RNA ligase 2 truncated (NEB) were added and the ligation was performed overnight at 18 °C. *RNA Size selection*: After partially digesting the RNA from the ligation reaction with 2 U of RNAse T1 (1U /uL) for 5 min at 22 °C and performing the original protocol for biotin pull-down and on-bead 5' phosphorylation, we eluted the RNA and size-selected fragments of 250-1000 nt. This was done by gel extraction and purification from a 6% Urea-PAGE gel. Libraries were normalized, pooled and then sequenced in the Illumina MiSeq platform (51 x 251 bp paired end run) with PhiX control library and the spike-in controls mixture. The quantified fluorescent signals were saved and processed by tailseeker2.

3.3.5 aTAIL-seq data analysis

Base calling, trimming of adapter sequences, removal of duplicated reads and determination of poly(A) tail sizes were performed by tailseeker2. Reads were analyzed by mapping to the WS247 assembly of the *C. elegans* genome using RNA-STAR (Dobin et al., 2013). Poly(A) lengths were then assigned to individual coding genes by intersecting the mapped sequences with WormBase.org WS247 gene annotations using BEDTools (Quinlan and Hall, 2010). Assignment to WormBase annotated non-coding RNAs (Nam and Bartel, 2012) was determined after ruling out matches to other overlapping coding and non-coding transcripts. Sequenced tags without a poly(A) tail were discarded and represented less than 0.02 percent of the data. The minimal poly(A) length detected was 5 nt.

3.3.6 Frequency of optimal codons (F_{op}) and ribosome enrichment

Optimal codons have been identified for yeast (Presnyak et al., 2015), *C. elegans* (Stenico et al., 1994) and *D. melanogaster* (Akashi, 1994). F_{op} was calculated as in a previous study (Stenico et al., 1994), and represents the ratio of optimal codons relative to the total number of codons in a transcript, excluding codons for amino acids represented by a single codon (methionine and tryptophan) and stop codons. Values can range from 0 to 1, with a F_{op} of 1 meaning that every codon is optimal. Ribosome enrichment was determined by calculating the log2 fold change of RPKM values for each transcript in the ribosome fraction relative to total RNA using paired RNA-seq and ribosome profiling datasets (Hendriks et al., 2014;

Subtelny et al., 2014).

3.3.7 Gene Ontology (GO) and tissue enrichment analysis

GO terms associated with long and short-tailed gene pools were identified using DAVID (Huang et al., 2009). Analysis for tissue enrichment in long and short-tailed genes was performed by employing scores from a dataset of global predictions of tissue-specific gene expression in *C. elegans* (Chikina et al., 2009).

3.3.8 Datasets used in this study

C. elegans: ribosome profiling and RNA-seq (Hendriks et al., 2014); *S. cerevisiae*: poly(A) measurements (Subtelny et al., 2014), ribosome profiling and RNA-seq (Subtelny et al., 2014), RNA half-life (Presnyak et al., 2015), and co-translational 5' decapping (codon protection index of cycloheximide treated cells) (Pelechano et al., 2015). NIH3T3: poly(A) measurements (Subtelny et al., 2014), ribosome profiling and RNA-seq (Subtelny et al., 2014), RNA half-life (Schwanhäusser et al., 2011), and translation rates (Schwanhäusser et al., 2011). *Drosophila* S2: poly(A) measurements and RNA-seq (Subtelny et al., 2014). HeLa: poly(A) measurements (Subtelny et al., 2014).

3.4 Acknowledgments

We thank V.N. Kim, J. Lim, and H. Chang for providing a detailed TAIL-seq protocol, their algorithm (tailseeker2), and technical assistance; E. Van Nostrand and members of the Yeo lab for assistance with the Illumina MiSeq platform; J.

Chen and J. Broughton for programming support; G. Yeo, J. Lykke-Andersen, H. Cook-Andersen, M. Wilkinson and members of the Pasquinelli lab for suggestions and critical reading of the manuscript. This work was supported by grants from the NIH (GM071654) and UCSD Academic Senate to A.E.P.; S.A.L. received an international HHMI predoctoral fellow. A.E.P. and S.A.L. designed the project and wrote the paper. S.A.L. conducted the experiments and data analysis.

Chapter 3, in full, is a reprint of the submitted manuscript "Efficient translation promotes poly(A) pruning" by Sarah Azoubel Lima and Amy E. Pasquinelli. I was the primary author.

3.5 Supplementary materials

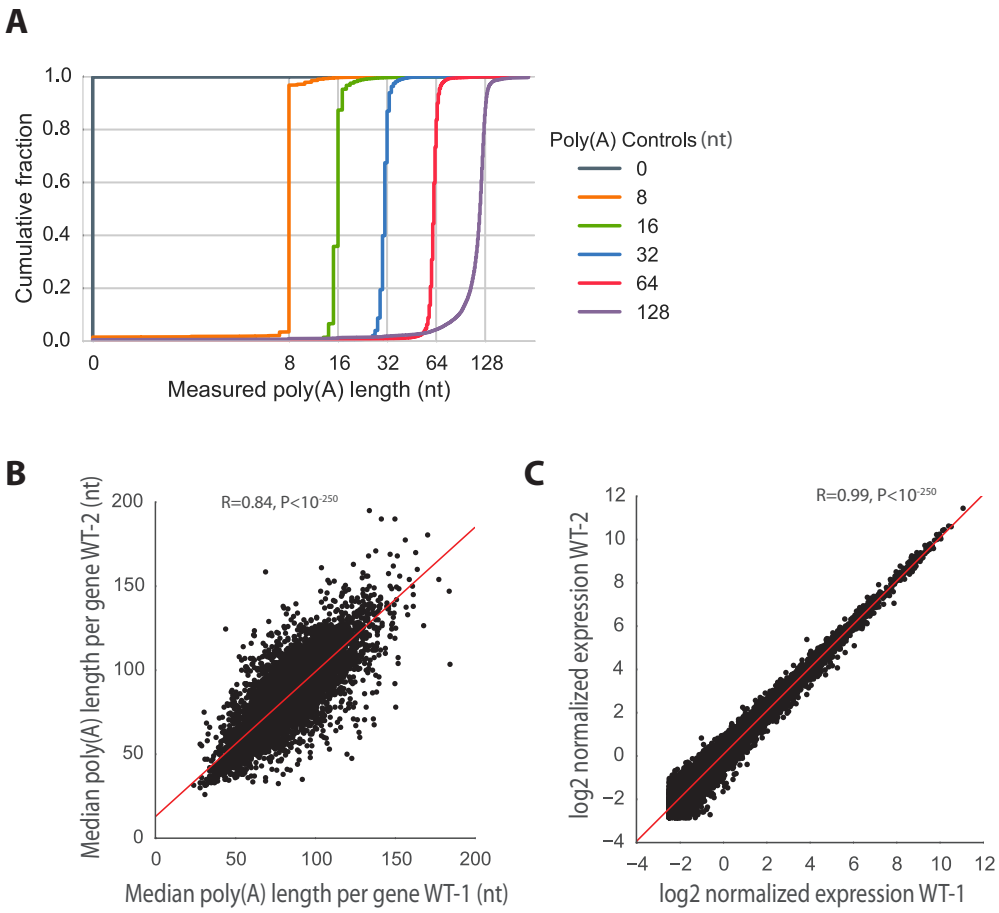


Figure 3.6: Validation of aTAIL-seq. (A) Cumulative frequency plot of poly(A) measurements of spike in controls of varying tail lengths. (B) Pearson correlation between median tail-length per gene in two biological replicates. (C) Pearson correlation of measured transcript abundance between two biological replicates.

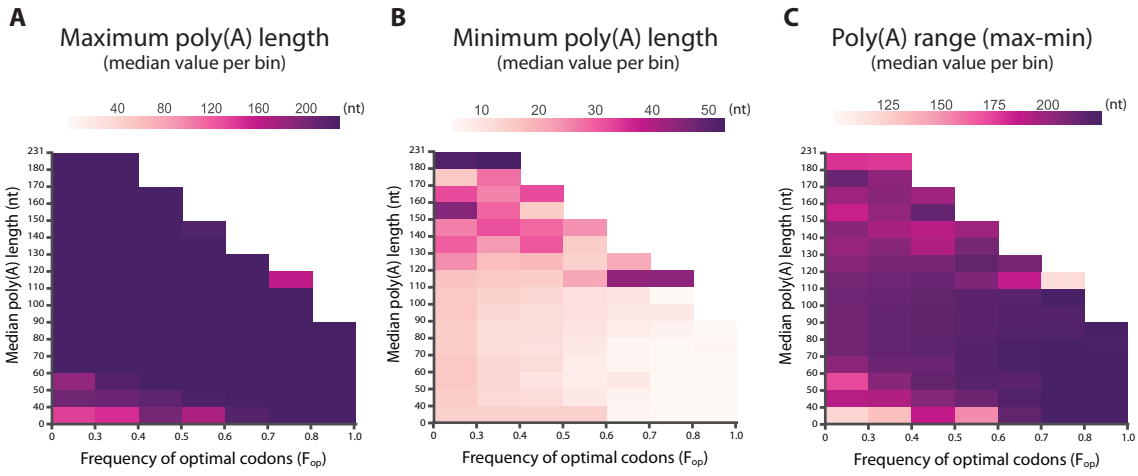


Figure 3.7: Range of poly(A) tails. (A to C) Heat maps show the median maximal tail length reported for each box (A), the median minimal tail length (B) and the tail range (C).

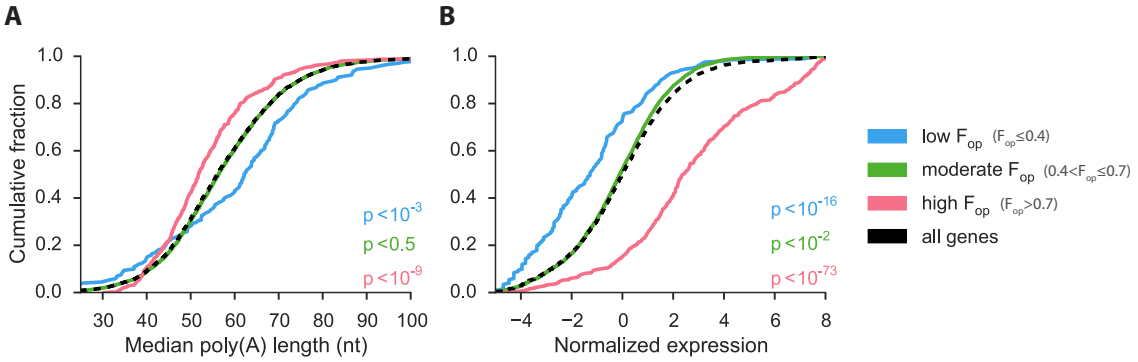
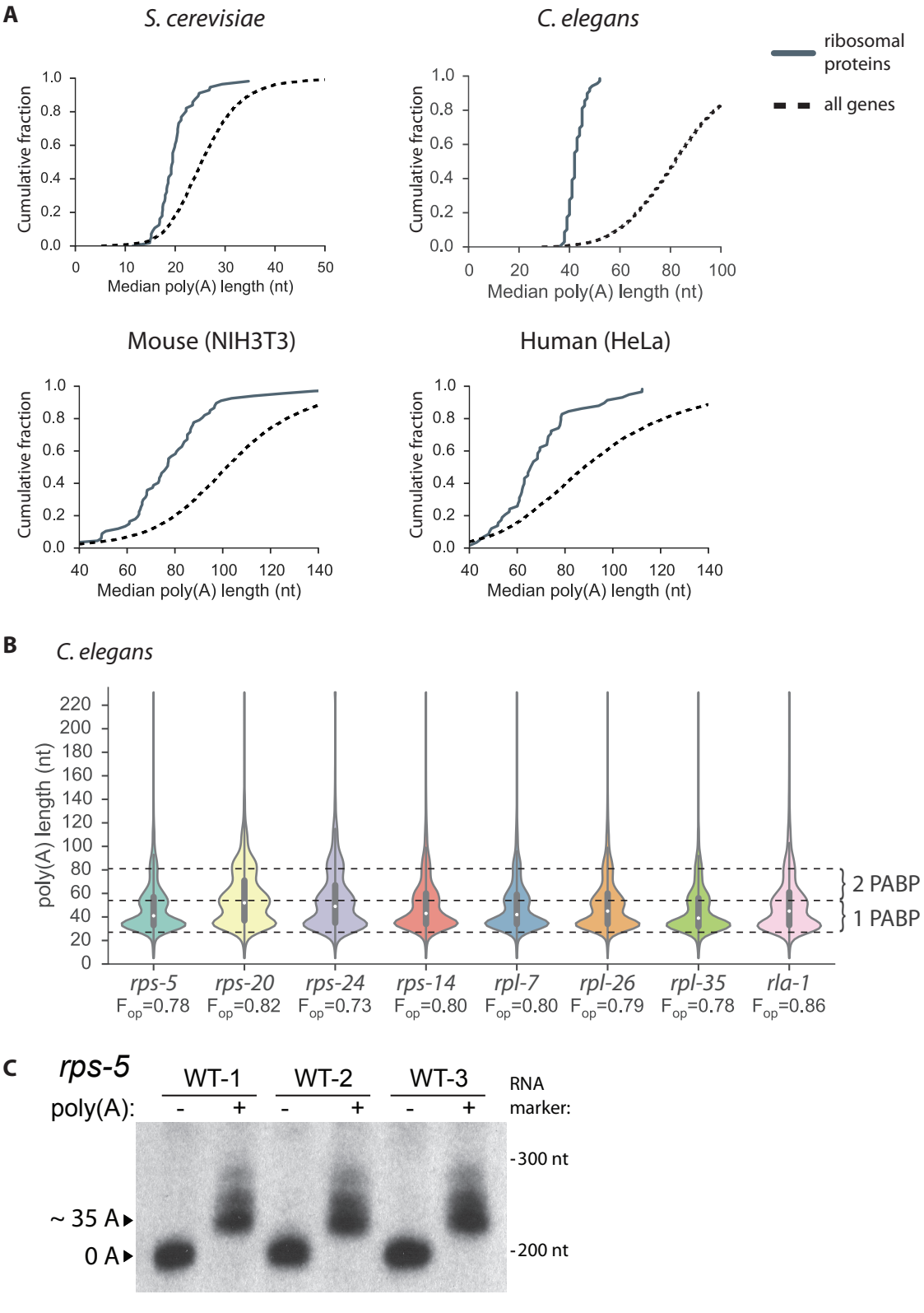


Figure 3.8: Evidence for poly(A) tail pruning in *Drosophila* S2 cells. (A and B) Cumulative distribution plots showing the relationship between codon optimization, poly(A) length (Subtelny et al., 2014) (A) and transcript abundance (Subtelny et al., 2014) (B).

Table 3.1: Spearman correlation values for poly(A) tail length, frequency of optimal codons (F_{op}) or NIH3T3 translation rates (TR), abundance, ribosome enrichment (RE), codon protection index (CPI), and other features in *C. elegans*, *S. cerevisiae*, *M. musculus* NIH3T3 and *Drosophila* S2 cells.

	C. elegans		***S. cerevisiae***		**Mouse NIH3T3**		***Drosophila* S2**	
Spearman correlation:	**R**	**P**	**R**	**P**	**R**	**P**	**R**	**P**
Poly(A) *vs* abundance	-0.24	$<10^{-174}$	-0.24	$<10^{-57}$	-0.27	$<10^{-53}$	-0.05	$<10^{-3}$
Poly(A) *vs* F_{op} (or TR)	-0.32	$<10^{-250}$	-0.17	$<10^{-25}$	-0.19	$<10^{-31}$	-0.09	$<10^{-10}$
Poly(A) *vs* RE	-0.21	$<10^{-133}$	-0.09	$<10^{-8}$	-0.08	$<10^{-6}$		
Poly(A) *vs* half-life			-0.11	$<10^{-9}$	-0.14	$<10^{-16}$		
Poly(A) *vs* CPI			0.2	$<10^{-29}$				
F_{op} (or TR) *vs* abundance	0.37	$<10^{-250}$	0.7	$<10^{-250}$	0.45	$<10^{-162}$	0.37	$<10^{-163}$
F_{op} (or TR) *vs* half-life			0.28	$<10^{-51}$	0.2	$<10^{-33}$		
F_{op} (or TR) *vs* RE	0.44	$<10^{-250}$	0.47	$<10^{-187}$	0.44	$<10^{-153}$		

Figure 3.9: Poly(A) tails of ribosomal protein transcripts. (A) Cumulative distribution plots of median poly(A) tail length for the mRNAs of ribosomal proteins in *S. cerevisiae* (Subtelny et al., 2014), *C. elegans,* NIH3T3 (Subtelny et al., 2014) and HeLa (Subtelny et al., 2014). (B) Violin distribution plots with inlaid box-plots (white dot represents the median) of all tail-length measurements for transcripts of 8 ribosomal proteins in *C. elegans*. (C) Northern blot for poly(A) length analysis of *rps-5* mRNA. To produce short 3' fragments that can be resolved by high-resolution Urea-PAGE gel, *rps-5* transcripts were digested using RNAse H with a gene-specific complementary oligonucleotide. Poly(A)- samples were simultaneously digested with oligo(dT) to remove the tail. The expected length of the poly(A)- fragment and a RNA marker were used to determine the poly(A) tail size in the three biological replicates.


A
S. cerevisiae
C. elegans
ribosomal proteins
all genes
Cumulative fraction
Median poly(A) length (nt)
Mouse (NIH3T3)
Human (HeLa)
B
C. elegans
poly(A) length (nt)
2 PABP
1 PABP
rps-5 Fop=0.78
rps-20 Fop=0.82
rps-24 Fop=0.73
rps-14 Fop=0.80
rpl-7 Fop=0.80
rpl-26 Fop=0.79
rpl-35 Fop=0.78
rla-1 Fop=0.86
C
rps-5
poly(A):
WT-1
WT-2
WT-3
RNA marker:
~ 35 A
0 A
-300 nt
-200 nt

Chapter 4

Adapted methods for investigating the poly(A) tail

4.1 Bulk poly(A) labeling

Bulk poly(A) labeling is useful to get a glimpse of the distribution of the global tail profile in a particular situation (Figure 4.2). It consists of radioactively labeling the 3' end of total RNA and digesting everything that is not a poly(A) tract with RNAse A (cuts affter C and U) and RNase T1 (cuts after G). The labeled poly(A)s are then run on a long sequencing gel for maximum resolution and imaged with a PhosphorImager (Figure 4.1). A radiolabeled marker is run along with the samples for estimation of poly(A) size.

4.1.1 Protocol

Radioactive 3' labeling of the RNA

1. In a screw cap tube, set up the following 30 µL reaction:
   (a) 3µL 10x NEB T4 RNA Ligase Buffer
   (b) 3µL 10mM ATP
   (c) 2 µg Total RNA
   (d) 3µL DMSO
   (e) 3µL PerkinElmer [32P] pCp
   (f) 1µL NEB T4 RNA Ligase 1 (20 units/µL)
   (g) Nuclease-free water (NFW) to complete 30µL
2. Incubate overnight at 16°C
3. Inactivate T4 RNA ligase at 68°C for 5min
4. Pulse down samples
5. Do the following to filter out unincorporated nucleotides using MicroSpin G-50 columns (GE)
   (a) Vortex column, loosen cap, snap off bottom, put column on collection tube
   (b) Spin 1min at 2000g, dispose of the liquid flow-through
   (c) Transfer column to a labeled screw cap sample tube
   (d) Add P32 labeled sample to resin

(e) Spin 2min at 2000g (unincorporated nucleotides are stuck in the resin, what passes through is your sample)

(f) Dispose of column in radioactivity waste

6. Label new screw cap tubes as "undigested" and save 1µL of each P32 labeled sample in these tubes to confirm 3' labeling; store at -80°C

Figure 4.1: Overview of the bulk poly(A) labeling method

Digestion of non-polyadenylated RNA

7. Make the following master mix to digest the labeled RNAs with RNAses A and T1 to remove non-poly(A) RNA

   per sample:

   (a) 0.4µL 10mg/ml RNAse A

   (b) 0.8µL 100u/µL RNAse T1

   (c) 16µL 5x RNAse digestion Buffer (recipe at the end)

   (d) 4µL 10mg/ml Yeast RNA (acts as ballast)

   (e) 29.8µL DEPC or Nuclease-free water

8. Add 51µL of the RNAse master mix to each sample

9. Incubate at 37°C for 2h
10. Stop the reaction by adding 20µL RNAse Stop Solution (recipe at the end) and incubating at 37°C for 30min
11. Add 100µL of Acid Phenol:Chloroform:Isoamylalcohol. Vortex well and spin at 16000g for 5min
12. Label new screw cap tubes and transfer aqueous phase to them (100µL)
13. Precipitate RNA by adding 2µL 10mg/ml Yeast RNA, 10µL NaOAC 3M, 2µL 20mg/ml glycogen and 300µL 100% Ethanol at -20°C overnight
14. Spin at top speed (~16000g) at 4°C for 15min; a nice pellet should form
15. Dispose of the radioactive supernatants appropriately
16. Wash pellets with 1ml 80% ethanol in DEPC H_2O by spinning at top speed for 10min in cold centrifuge; dispose of the radioactive wash appropriately
17. Dry pellets at room tempertature (RT) for 10min
18. Resuspend labeled poly(A) tails in 20µL DEPC H_2O (warming the H_2O to 68°C helps); if not using samples immediately, store at -80°C

Running the 3' labeled poly(A) tails on a long sequencing gel

Pouring the gel:

19. Wipe down sequencing gel plates and spacers with ddH_2O with lint-free tissue to assure they are perfectly clean (imperfections in the glass and dust cause bubbles during pouring)

20. Assemble sequencing gel plates with 0.4mm spacers and clips. One of the surfaces facing the gel should have the GelSlick on it and one should not (the gel will stick to the non-slick plate when disassembling). If needed, apply a few drops of GelSlick to one of the sides and rub evenly with a large kimwipe, wait a few minutes to dry. Check which side is slick by splashing with ddH_2O

21. Prop up the assembled plates on 2 Styrofoam racks so that it is high up and has the top part facing you

22. Make 50ml of 16% UreaPage solution with the UreaGel Sequagel system

    (a) 13ml diluent

    (b) 5ml buffer

    (c) 32ml concentrate

23. Prepare a beaker and a 60ml syringe (or a 50ml sterile pipette tip)

24. When ready to pour add 400µL 10% APS (ammonium persulfate) and 20µL Temed

25. After adding the APS and Temed quickly invert a few times and transfer solution to the clean beaker and pull it into the 60ml syringe

26. Tilt the plate so the top is up and so that one side is higher than the other. Slowly pour the PAGE solution with the syringe steadily from the lower side. Try to avoid bubbles

27. After all the space has been filled, insert the well comb

28. Allow gel to polymerize for ~1h

    Running the gel:

29. Make 2L 1X TBE

30. Carefully remove well comb and bottom spacer from the gel

31. Prop gel on rig and tighten it just so the upper buffer chamber does not leak (screwing too tight can lead to the glass cracking during the run)

32. Fill upper and lower chambers with 1X TBE (lower chamber can be filled half way)

33. Use clips to attach metal plate behind the gel, so that it helps distribute the heat

34. Stick a thermometer strip to the glass plate facing you

35. Use syringe to flush wells with 1X TBE

36. Start pre-run at ~32W, wait 30min to 1h so that the temperature of the gel is between 45°C and 60°C (50-60°C is ideal). Going over 60°C can cause the glass to break during the run

37. Prepare your RNA samples during the pre-run:

    (a) Mix 5µL labeled poly(A) RNA with 5µL 2X formamide RNA gel loading dye

    (b) Add 4µL DEPC H_2O and 5µL formamide dye to "uncut" samples

    (c) Get the P32 labeled Ambion Decade Marker (instructions on how to label the marker at the end)

(d) Boil RNA samples and marker for 5min immediately before loading; pulse down; put on ice

38. After pre-run turn power supply off and flush wells with 1X TBE a second time

39. Carefully load ~5µL of your radioactive samples with a 20µL filter tip p20; save the remaining 5µL in the -80°C

40. Run gel at 30-35W (found that 32W worked well) until blue dye is 8-10cm from the bottom of the gel (~2h)

Drying and exposing the gel:

41. Use a 50ml pipette to transfer buffer from the upper rig box to a beaker, until it sits under the gel level (buffer may be radioactive)

42. Lay down some diaper papers in the are the gel will be disassembled

43. Remove metal plate and loosen screws, carefully taking gel out of the rig

44. With a big spatula, gently prop apart glass plates; gel will be stuck on the non-slick glass surface

45. Hydrate gel by spraying lightly with ddH_2O; a dry gel will not stick to the whatman

46. Lay large 3MM whatman paper on top of gel; turn glass-gel-whatman sandwich upside down and gently peel away the whatman so that the gel sticks to it

47. Carefully lay a sheet of saran wrap on top of gel (minor grooves may form which will not affect imaging)

48. Put whatman-gel-saran wrap sandwich on top of the porous sheet of a gel dryer (whatman on the bottom), add the hard plastic sheet on top of the saran wrap and cover with the flexible rubbery sheet. Turn on vacuum and make sure the dryer is vacuum sealed

49. Vacuum dry the gel for 2h at 80°C and at least 30min with no heat; cooling down the gel before removing from the dryer helps in avoiding breaks (the no-heat step can be left overnight)

50. Expose gel to a phosphor screen for the desired time (usually at least 2 days) and image in a Typhoon scanner

Labeling the Decade RNA Marker (Ambion) with P32 ATP

1. Set up the following 10µL reaction in a screw cap tube:
   (a) 1µL Decade Marker RNA (stored at -80°C)
   (b) 1µL 10x PNK buffer
   (c) 6µL Nuclease-free H_2O
   (d) 1µL 32P ATP
   (e) 1µL T4 PNK
2. Incubate at 37°C for 1h
3. Add: 8µL Nuclease-free H_2O and 2µL cleavage reagent; Incubate at RT for exactly 5min
4. Clear labeled marker with a MicroSpin G-50 column (GE)
   (a) Vortex column, loosen cap, snap off bottom, put column on collection tube
   (b) Spin 1min at 2000g, dispose of the liquid flow-through
   (c) Transfer column to a labeled screw cap tube
   (d) Add P32 labeled marker to resin
   (e) Spin 2min at 2000g (unincorporated nucleotides are stuck in the resin, what passes through is the marker)
   (f) Dispose of column in radioactivity waste
5. Add 20µL 2X formamide RNA gel loading dye to labeled ladder
6. Boil for 5min, store at -80°C

Solutions (use RNAse free reagents)

RNAse Digestion Buffer:

- 5mM EDTA
- 300mM NaCl
- 10mM Tris-HCl (pH 7.5)

RNAse Stop Solution (make fresh):

- 2 mg/mL Proteinase K
- 130 mM EDTA
- 2.5% (w/v) SDS

4.1.2 Bulk poly(A) labeling reveals change in the poly(A) profile of aging worms

I have applied the bulk poly(A) labeling method to screen for differences in the profile of tails in different biological conditions (Figure 4.2). In this experiment, I found a marked difference in the poly(A) tails of aging worms, which are significantly shorter then the tails of young adult and late larval development animals. This result is discussed in Chapter 5.

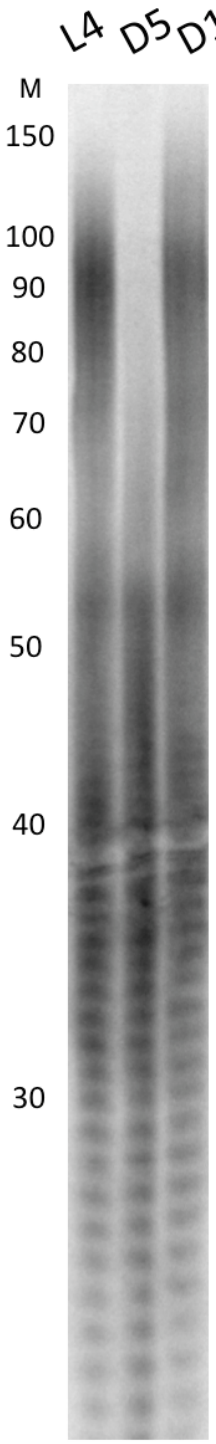


Figure 4.2: Aging causes shift in the worm bulk poly(A) profile. Bulk poly(A) labeling comparing late larval stage (L4) to young adults (day 1 - D1) and aged *C. elegans* (day 5 of adulthood - D5). Aged worms have their poly(A) profile significantly shifted towards shorter tails. M: Decade Marker (nt).

4.2 ePAT (extension Poly(A) Test)

ePAT provides a fast way to investigate gene-specific poly(A) lengths. The ePAT method was developed by Jänicke et al. (2012) as an improvement on the traditional Ligation-Mediated Poly(A) Test (LM-PAT) assay. ePAT uses Klenow DNA polymerase to add a 3' extension onto RNAs by using a DNA oligo template. This DNA oligo (ePAT anchor primer - see below) anneals to the end of the poly(A) tail and provides the 3' adapter sequence for Klenow to extend. As a control, a TVN (TVN anchor primer - see below) oligo is utilized to identify the polyadenylation site and allow estimation of the tail site. The TVN anneals with the first few adenosine residues in the tail and anchors at the polyadenylation site by the use of 2 variable nucleotides (VN).

Here I provide a detailed e-PAT protocol with some improvements in the quality of gel visualisation of the PCR products. An example of a sucessful ePAT gel is provided in Figure 4.3.

4.2.1 Protocol

ePAT cDNA

1. Mix in a 0.2ml PCR tube:
   (a) 0.5 to 1µg of RNA
   (b) 1µL 100µM anchor primer (A1696)
   (c) Complete to 8µL with Nuclease-Free H_2O
2. Use a thermocycler to incubate at 80°C for 5min

3. Allow samples to cool to room temperature

4. Add 12µL of the following master mix to each sample:

   (a) 4µL Nuclease-Free H_2O

   (b) 4µL 5X Superscript III Buffer

   (c) 1µL 0.1M DTT

   (d) 1µL 10mM dNTP

   (e) 1µL RNAsin plus

   (f) 1µL Klenow polymerase (Large Fragment; not exo-)

5. Gently mix (tap tube and invert) and pulse down

6. Run the following protocol in the thermocycler:

   (a) 25°C 1h

   (b) 80°C 10min (Klenow inactivation)

   (c) 55°C 1min

   (d) 55°C hold; keep tubes at 55°C in the block

   (e) add 1µL SuperScript III RT to each tube (It is important to do this with the tubes still in the PCR block; Mix by pipetting up and down gently)

   (f) 55°C 1h

   (g) 80°C 10min (SuperScript inactivation)

   (h) 37°C hold; allow samples to stabilize at 37°C

   (i) add 1µL RNAse H to each tube

(j) 37°C 20min

(k) 4°C hold

7. Store cDNAs in the -20°C

Anchored (TVN) cDNA

8. Start with 0.5 to 1µg of RNA in 10µL DEPC H_2O
9. Prepare Mix 1 for n+1 reactions:

   per reaction volume: 2µL

   (a) 1µL 10mM dNTPs

   (b) 1µL TVN anchor primer 100µM (A2952)

10. Add 2µL Mix 1 to 10µL of normalized RNA
11. Incubate at 65°C for 5min in PCR machine
12. Transfer to ice and incubate for 1min
13. While incubating reactions, prepare Mix 2 for n+1 reactions:

    per reaction volume: 7µL

    (a) 4µL 5X First Strand Buffer

    (b) 2µL 0.1M DTT

    (c) 1µL RNasin Plus

14. Pre-heat Mix 2 at 42°C and pre-heat a PCR machine at 42°C

15. After incubating samples on ice, add 7µL of pre-heated Mix 2 to each tube (19µL total) and immediately put tubes in the 42°C thermocycler
16. Add 1µL SuperScript III RT to each reaction (Mix by gently pipetting up and down while keeping samples in the 42°C block)
17. Run following protocol on thermocycler:
    (a) 1h at 42°C, 1h at 52°C, hold at 4°C
    (b) Add 1µL RNase H to each reaction.
    (c) Incubate at 37°C for 20min
18. Store TVN cDNAs at -20°C

ePAT PCR

19. Dilute ePAT cDNA (1/6) and the TVN cDNA (1/10)
20. Run PCR with a forward primer in the 3'UTR of target gene (shorter PCR products work better and make poly(A) changes easier to observe; ideal size is between 80-120bp)
21. The reverse primer for the PCR is the same as the RT primer (A1696)
22. The PCR can run for 18 to 40 cycles, depending on the gene. Annealing temperatures are usually in the range of 58-60°C
23. Use a Sodium Borate agarose gel (recipe at the end) for best resolution; Usually, a 1.7% gel works best
24. Run a TVN PCR along with the ePAT samples

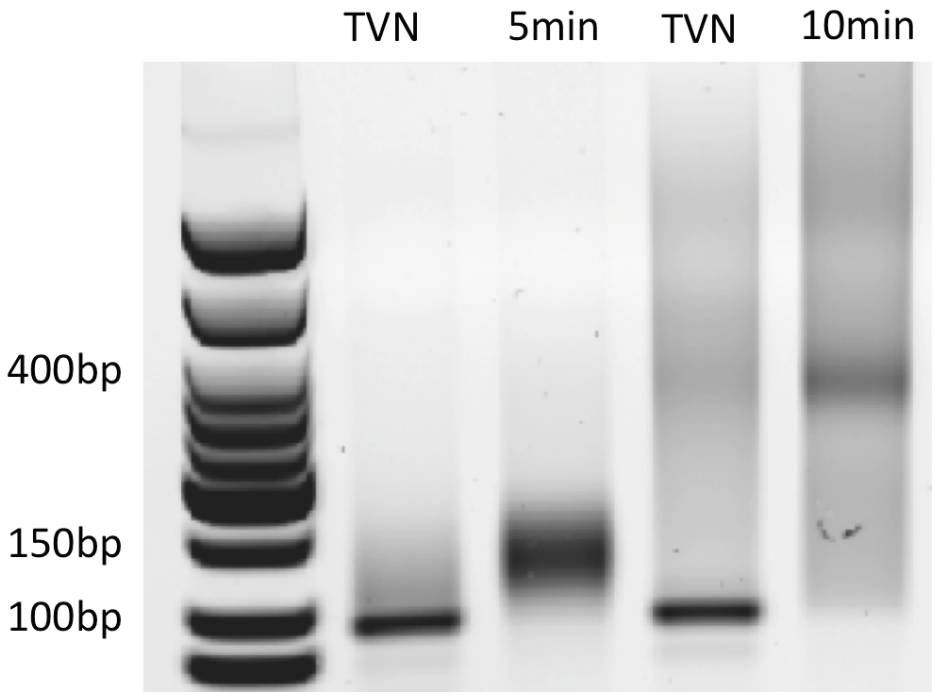


Figure 4.3: ePAT successfully detects differences in polyadenylation. An *in vitro* transcribed RNA was polyadenylated *in vitro* for 5 or 10min. The TVN control shows where the tail is anchored to the 3'UTR and allows the estimation of poly(A) lengths

ePAT oligos:

- ePAT anchor primer (A1696):

  GCGAGCTCCGCGGCCGCGTTTTTTTTTTTT

- TVN anchor primer (A2952):

  GCGAGCTCCGCGGCCGCGTTTTTTTTTTTTVN

Sodium Borate (SB) agarose gel buffer:

- To make a 1L 20x stock solution:

  1. Add 8g NaOH (use solid pellets) and 47g boric acid to 900ml ddH_2O
  2. Use stir bar to mix
  3. Adjust volume to 1L with ddH_2O
  4. Filter solution (0.2 micron) and test pH with a strip; pH should be ~8.2
  5. Use 1x SB buffer to make and run your gels
  6. These gels can run very fast without overheating. General guidelines are 200V for a 50ml gel and 250V for a 100ml gel

4.3 adapted TAIL-seq (aTAIL-seq)

aTAIL-seq was adapted from TAIL-seq (Chang et al., 2014) to remove the need for bead-based ribosomal RNA removal; which is not optimal for *C. elegans* and many other organisms. aTAIL-seq intead uses a splint oligo, which greatly favors the capture of polyadenylated RNA over non-polyadenylated RNAs. For a overview of the aTAIL-seq method, see Figure 3.1. For a more detailed discussion on the improvements and analysis of aTAIL-seq, see Chapter 3 and Chapter 5.

4.3.1 Protocol

Preparation of input RNA

- Extract RNA using Trizol (include DNAse treatment)
- Quantify RNA and make sure RNA is excellent quality by running gel and nanodrop
- Start with 20µg RNA per sample

3' adaptor ligation

1. Use 200 µL PCR tubes and perform the reaction in a thermocycler
2. Mix the following reagents:
   (a) 5.5 µL 20µg RNA
   (b) 1 µL 10µM biotinylated and adenylated 3' adaptor (A3220)
   (c) 1 µL 10µM splint oligo (A3462)

3. Incubate the mixture at room temperature (RT) for 5 min
4. Add the following reagents and briefly mix and pulse down
    (a) 1 µL 10X RNA ligase buffer (B0216S, NEB)
    (b) 0.5 µL Superase-In (Ambion)
    (c) 1 µL T4 RNA ligase 2 truncated (M0242S, NEB)
    (d) Total: 10 µL
5. Incubate at 18°C overnight (O/N)

RNase T1 treatment

6. To the 10µL ligation reactions add:
    (a) 10µL Nuclease-free Water (NFW)
    (b) 80µL 1x sequencing buffer
    (c) Total: 100µL
7. In a thermocycler:
    (a) Incubate at 50°C for 5 min
    (b) Hold samples at 22 °C
    (c) Add 2 µL RNase T1 (1 U/ µL) (AM2283, Ambion) and mix well
    (d) Incubate at 22 °C for 5 min
8. Transfer to 1.5ml tube containing 220 µL Precipitation/Inactivation buffer, mix well

9. Add 0.5µL GlycoBlue
10. Incubate at -80 °C for 30min to 1hr
11. Centrifuge at max speed for 30 min at 4 °C
12. Wash with 1mL 75% Ethanol twice (spin for 5 min max speed in cold centrifuge for each wash)
13. Resuspend the pellet with 50 µL NFW, put on ice

Biotin pull-down

- Use a thermomixer and 1.5ml tubes in this step)

14. Bead preparation (Dynabeads M-280 Streptavidin, 11206D)
    (a) Mix dynabeads to resuspend
    (b) Transfer 50 µL of dynabeads to 1.5mL tube
    (c) Place the tube on magnetic stand for at least 1 min
    (d) Remove supernatant and add an equal volume of solution A (*always use fresh solution A)
    (e) Incubate at RT for 2 min
    (f) Repeat solution A wash
    (g) Remove supernatant and add an equal volume of solution B
    (h) Remove supernatant and resuspend beads with an equal volume of 2X B&W buffer

15. Bead binding

    (a) Mix RNA and beads (final volume 100 µL)

    (b) Incubate at 25 °C for 15 min, shaking on thermomixer (4min no shake, 15sec 1200rpm)

16. Bead wash

    (a) Wash beads with 1X B&W buffer 100 µL twice

    (b) Wash beads with 1X PNK buffer 100 µL once

5' phosphorylation (on beads)

17. Remove PNK buffer wash add the following PNK reaction mastermix to the beads:

    (a) 5 µL 10X PNK buffer

    (b) 2 µL 10mM ATP

    (c) 1 µL Superase-In (Ambion)

    (d) 2 µL, T4 PNK (NEB)

    (e) 40 µL NFW

    (f) Total: 50 µL

18. Incubate at 37 °C for 30 min, shaking on thermomixer (4min no shake, 15sec 1200rpm)

19. Wash beads with 1X PNK buffer 100 µL twice

RNA Fragment size selection (Urea-PAGE gel)

20. Elution from beads:

    (a) To beads, add 13 µL of 2X RNA formamide loading dye and mix well. Be sure that beads were completely resuspended

    (b) Incubate at 95 °C for 3 min (thermomixer)

    (c) Immediately place the tube on magnetic stand and collect supernatant

    (d) Keep eluted RNA on ice

    (e) Repeat elution once more and combine volumes together (Total: 26 µL per sample)

21. Gel purification

    (a) Load both the eluted RNA and RNA Century Plus Marker (AM7145) on a 6% Urea-PAGE gel (1X TBE)

    **Beforing loading: pre-run for 30min at 150V, and flush the wells with 1X TBE)

    (b) Run the gel for ~70 min at 150 V (1X TBE)

    (c) Stain the gel with SYBR gold (S-11494, Invitrogen) for 5 min at RT

    (d) Cut the gel between 250 and 750 nt on blue light transilluminator and put piece in 1.5ml tube

    (e) Crush gel slice in 200µL 0.3M NaCl with sterile pestle until it is in very small fragments

    (f) Add another 600 µL 0.3M NaCl (for a total of 800µL) and rotate O/N at 4°C

22. (Next day) Transfer gel debris into Spin-X columns (Corning, 8162)
    **Divide each sample into 2 columns (400 µL each)
23. Centrifuge at maximum speed at RT for 5 min
24. Take flow through and add 0.5 µL GlycoBlue, 40 µL 3M NaOAc, and 1 ml pre-chilled 100% ethanol
25. Incubate at -80 °C for 30min to 1hr
26. Centrifuge at max speed for 30 min at 4 °C
27. Wash with 1mL 75% ethanol twice (5min spins)
28. Resuspend the first sample pellet with 4.2 µL NFW
29. Transfer the 4.2 µL to the second sample tube and resuspend second pellet to integrate sample

5' adaptor ligation

- Use 200 µL tubes and perform the ligation reaction in a thermocycler

30. Combine the following reagents in a PCR tube:
    (a) 4.2 µL Purified RNA
    (b) 1 µL 5 µM 5' adaptor (RNA oligo - A3219)
31. Incubate the mixture at 70 °C for 2 min and transfer immediately onto ice
32. Add the following reagents and mix well:

(a) 0.8 µL 10X RNA ligase buffer (B0216S, NEB)

(b) 0.4 µL Superase-In (Ambion)

(c) 0.8 µL 10mM ATP

(d) 0.8 µL T4 RNA ligase 1 (M0204S, NEB)

(e) Total: 8 µL

33. Incubate at 28°C for 1hr, 25°C for 6 hr, 22°C for 6 hr, and hold at 4°C O/N

cDNA preparation

- Use 200 µL tubes and perform reaction in a thermocycler

34. Mix the following reagents:

(a) 8 µL Ligated RNA

(b) 1 µL 4 µM RT primer (RTP) (A3221)

35. Incubate the mixture at 70 °C for 2 min and place the tube immediately on ice

36. Add the following:

(a) 4 µL Nuclease-free Water (NFW)

(b) 4 µL 5X first strand buffer (Invitrogen)

(c) 1 µL 10mM dNTP mixture

(d) 1 µL 100mM DTT (Invitrogen)

(e) 1 µL Superscript III Reverse-Transcriptase (Invitrogen)

(f) Total: 20 µL

37. Incubate at 50 °C for 1hr, heat inactivate at 70 °C for 15min, pause at 4°C

Small scale PCR

- Use 200 µL tubes and perform reaction in a thermocycler

38. Small scale (1/25) PCR to decide the number of PCR cycles

    (Normally test between 11-15 cycles for each sample)

39. Mix the following reagents:

    (a) 0.4 µL cDNA

    (b) 2 µL 2X Phusion HF MasterMix (NEB)

    (c) 0.08 µL RP1 Forward primer (25 µM) (A3222)

    (d) 0.08 µL RPIX Reverse primer (25 µM) (From Illumina small RNA-seq kit)

    (e) 1.44 µL NFW

    (f) Total: 4 µL

40. Run the following PCR program:

    (a) 98 °C; 30 sec

    (b) 98 °C; 10 sec

    (c) 60 °C; 30 sec

    (d) 72 °C; 45 sec

    (e) Cycle (b)-(d) 11 to 15X

    (f) 72 °C; 5 min

    (g) Hold at 4 °C

41. Add 1 µL DNA loading dye to each sample

42. Run in a 6% non-denaturing PAGE gel for ~40 min at 200 V (1X TBE)

43. Stain the gel with EtBr for 5 min at RT

44. Image gel, library smears should be visible and smooth

45. Establish ideal cycle number for each sample

    As few as possible (less duplicated reads when sequencing)

Large scale PCR to prepare aTAIL-seq library

- Use 200 µL tubes and perform reaction in a thermocycler

46. Mix the following reagents:

    (a) 10 µL cDNA

    (b) 25 µL 2X Phusion HF MasterMix (NEB)

    (c) 1 µL RP1 Forward primer (25 µM)

    (d) 1 µL RPIX Reverse primer (25 µM)

    (e) 13 µL NFW

    (f) Total: 50 µL

47. Perform PCR as in step 40, with number of cycles determined by step 45

    *As few as possible (try to keep number under 16 cycles)

Purification of DNA library by AMPure XP beads

- This step size-selects the library, removing the primers seen in Figure 4.4

48. Transfer the PCR mix to 1.5 ml tube and add 35 µL (0.7X) of AMPure XP beads
49. Mix thoroughly by pipetting, incubate at RT (25 °C) for 5 min
50. Place the tube on magnetic stand for 2 min
51. Remove supernatant
52. Add 200 µL of 70% Ethanol to each tube
    *Always make fresh 70% Ethanol
53. Incubate for at RT (25 °C) 30 sec: Do not disturb the separated magnetic beads during wash steps
54. Repeat for a total of two washes
55. Air dry the beads for 3 min: Be careful not to over dry (beads will appear cracked)
56. Add 100 µL of NFW, mix well, incubate at RT (25 °C) for 5 min
57. Place the tube on magnetic stand for 1 min
58. Collect supernatant: Be careful of magnetic beads contamination
59. Repeat elution once and combine them together

60. Add 100% Ethanol 1000 µL, 3M NaOAc 20 µL, GlycoBlue 0.5 µL
61. Incubate at -80 °C for 30min -1hr
62. Centrifuge at max speed for 30 min at 4 °C
63. Wash with 75% Ethanol twice (5min spins)
64. Resuspend the pellet with 6-10 µL NFW

Sequencing

- Use Tapestation and Qubit to analyze libraries for quality and concentration. The normalized and pooled libraries are sequenced in the Illumina MiSeq platform (51 x 251 bp paired end run) with 5-10% PhiX control library and the spike-in controls mixture. The quantified fluorescent signals are saved and processed by tailseeker2.

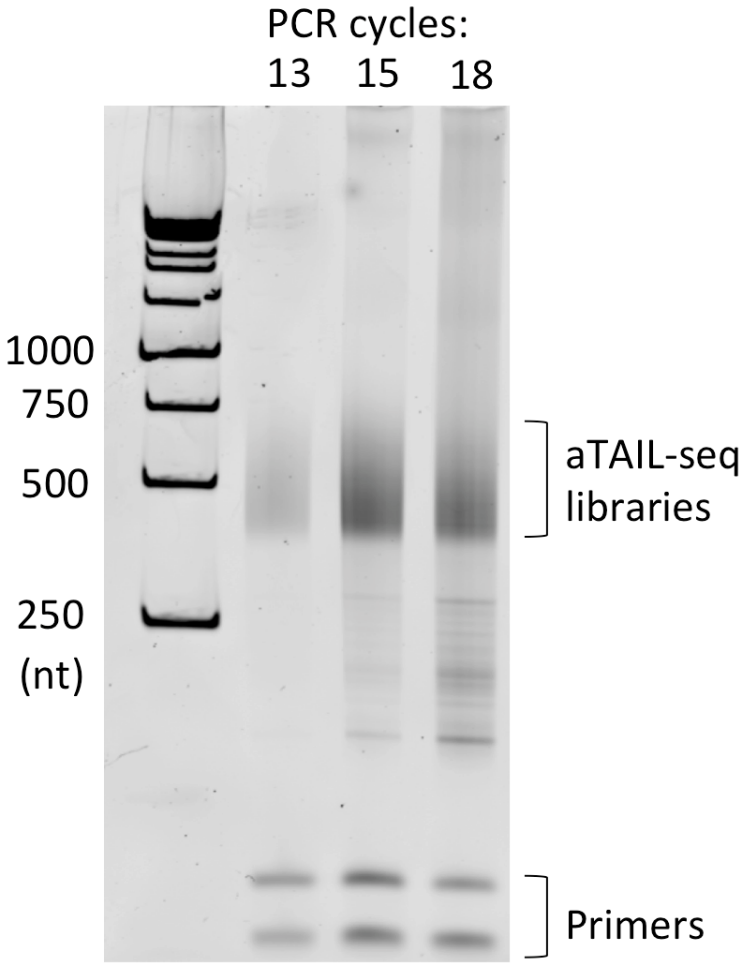


Figure 4.4: aTAIL-seq small scale PCR. Libraries were run on a 6% non-denaturing PAGE gel along with DNA size marker. For this library, 13 PCR cycles are ideal to minimize duplications and unspecific amplification products

4.3.2 Materials and solutions

Buffer recipes:

- 1X Sequencing buffer

  20 mM sodium citrate pH 5, 1 mM EDTA, 7 M urea

- 1X PNK buffer

  50mM Tris-Cl pH 7.4, 10mM MgCl2, 0.5% NP-40

- 1X solution A

  (**Always make fresh solution A and use recent 1M NaOH stock)

  0.1 M NaOH, DEPC-treated 0.05 M NaCl

- 1x solution B

  DEPC-treated 0.1 M NaCl

- 2X Binding & Wash buffer (B&W)

  10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl

Oligos:

- Splint oligo for 3' ligation (A3462):

  NNNGTCAGTTTTTTTTT

- Biotinylated adapter for 3' ligation (A3220):

  /5App/CTGACNNNNNNNNNNNNNNNTGGAATT

  CTCGGGTGCCAAGGC/iBiodT//iBiodT//3ddC/

- RNA adapter for 5' ligation (A3219):

  GUUCAGAGUUCUACAGUCCGACGAUC

- RNA RT primer (RTP) (A3221):

  GCCTTGGCACCCGAGAATTCCA

- RNA PCR Primer (RP1) (A3222):

  AATGATACGGCGACCACCGAGATCTACACGTTC

  AGAGTTCTACAGTCCGA

- RNA PCR Primer, Variable Index (RPIX)

  These are primers from the Illumina small RNA-seq kit;

  Several indexes exist; Example of primer with Index 1 sequence below:

  CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTG

  GAGTTCCTTGGCACCCGAGAATTCCA

Chapter 5

Conclusions

The poly(A) tail has long been recognized as an important regulatory element in the translation and stability of mRNAs. However, there is still a lot to learn about the dynamic control of poly(A) length and how it relates to translation and decay. Due to technical restraints, most past observations linking poly(A) length to translation have been limited to a few very specific biologic scenarios and particular transcripts. Now, with novel sequencing methods designed to measure global poly(A) profiles (Chang et al., 2014; Subtelny et al., 2014), we can investigate the poly(A) lengths of specific genes from various organisms, tissues and biological conditions.

In this work, we have built upon the work of Chang et al., (2014) and adapted the poly(A) sequencing method TAIL-seq. We then used aTAIL-seq (adapted TAIL-seq) to perform a in-depth analysis of the poly(A) profile of the model organism *C. elegans* in its late larval development. Now, in view of this and other recent works (Chang et al., 2014; Lim et al., 2016; Park et al., 2016; Subtelny et al., 2014), we understand that the relationship between poly(A) length, translation and stability is much more interconnected and nuanced than we originally thought.

5.1 How does aTAIL-seq improve upon TAIL-seq?

The major innovation of aTAIL-seq upon TAIL-seq is the use of a splint oligo, which greatly favors the capture of polyadenylated RNA over non-polyadenylated RNAs (Figure 3.1A). The decision to use the splint was driven by the fact that for *C. elegans* (and other organisms), we do not currently have a cost-effective and efficient method to remove ribosomal RNA contamination from a large amount of total RNA. My early attempts with the original TAIL-seq protocol resulted in libraries containing mostly ribosomal RNA contamination. By adding the splint in the 3' ligation reaction, libraries improved from 80-90% contamination to 80-90% usable poly(A) reads (Figure 4.4). Our splint is designed to work with the original 3' adaptor used by TAIL-seq (Chang et al., 2014) and accounts for the randomized sequences in that 3' adaptor by including 3 variable nucleotides in its 5' end (Chapter 4.3).

The addition of the splint also reduces the size range of RNA fragments needed for sequencing from 500-1000nt to 250-750nt (Figure 3.1A), as there is reduced peril of contamination with non-polyadenylated small RNAs. The smaller size range for the library allows for better formation of clusters on the Illumina sequencing chip, contributing to the improvement of the number of quality reads.

One caveat of adding a splint oligo to prevent ligation of non-poly(A) RNAs is the possible loss of information about 3' tail modification. If a transcript has other nucleotides added to its tail, we might not be able to capture that RNA, as the splint is designed to bridge the last 9 adenosines of the tail to the adaptor. Studies using the original TAIL-seq method have found 3' end tail modifications in mammalian cells that are associated with mRNA stability (Chang et al., 2014; Lim et al., 2014). However, these transcripts with modified tails seem to correspond to a small fraction

of the total pool of mRNAs. While a novel and important observation, the general significance of these modifications to global RNA regulation, and their conservation in other organisms, are still to be determined. In the future, there is potential for including alternative splints designed to capture multiple or specific 3' modifications in aTAIL-seq.

Another caveat is the requirement for a minimum number of adenosines in the tail. However, that does not seem to be a problem in our system, as we capture tails as short as 5 nt and we do not observe a large accumulation of deadenylated species whenever we assay poly(A) size through northern blots (Figure 3.9C). Overall our global poly(A) profile indicates that these short-tailed species are not abundant and likely very unstable(Figure 3.1B).

5.2 Poly(A) length and stability

We were surprised to find that, in *C. elegans*, the majority of mRNA transcripts in the organism have poly(A) tails that can accommodate only a single PABP (Figure 3.1B) (Baer and Kornberg, 1983; Smith et al., 1997; Wang et al., 1999). Also striking was the sharp drop in frequency of transcripts that have tails shorter than the PABP footprint. This indicates that having at least one PABP protein bound to the poly(A) tail is a general requirement for mRNA stability, which is consistent with observations in mammalian cells (Chapter 3 and Park et al. (2016)). Furthermore, we can draw a similar observation from the phasing pattern in Figure 3.1B. The frequency peaks correspond to the tail sizes that occur with serial binding of PABP to the tail. The drop in frequency between those lengths suggests that unprotected adenosines are

removed from the tail. The most likely explanation is that unprotected poly(A) tails are easily targeted by deadenylases or by the exosome.

As shortening of the poly(A) tail is generally thought to promote mRNA instability (Goldstrohm and Wickens, 2008; Jalkanen et al., 2014; Roy and Jacobson, 2013), we were surprised to find that the relationship between poly(A) length and abundance is inverted once the tail is long enough to contain one or more PABPs (Figure 3.2B). However, this raises the question if steady-state abundance necessarily corresponds to stability. The answer was reassuring: across eukaryotes, shorter poly(A) tails are associated with mRNAs that have longer half-lives (Figures 3.1D and Table 3.1) and mRNAs with longer average length poly(A) tails are more frequently degraded in yeast (Figure 3.4E).

Yet, *how do unstable mRNAs have long poly(A) tails if the first step of decay is deadenylation?* One possibility is that decay intermediates are very short lived. Thus, I assume we mostly detect a "younger" pool of mRNAs that is not yet targeted for degradation. The poly(A) lengths likely reflect the long tails attached to nascent transcripts (Eckmann et al., 2011). The lack of shorter tail lengths for many of these transcripts also indicates that, once started, deadenylation for decay purposes is likely very processive, leading to rapid decapping and degradation (Figure 1.2). This is consistent with past studies that show that decay intermediates are rare under normal circumstances (Hsu and Stevens, 1993; Hu et al., 2009).

However, *is a long average length poly(A) tail a consequence of mRNA instability or does it promote instability?* This is an important question, which is yet to be directly addressed. There are two possible answers. 1) It could be that a short mRNA half-life is defined by other factors, such as translation rate and 3'UTR

sequence elements. In this way, being detected with a steady state long poly(A) tail is just a consequence of most of the mRNA being rapidly targeted for decay. 2) Another possibility is that there is a combinatorial effect, and that poly(A) length (in addition to the factors) does have some impact on stability. This is an interesting idea, as the C-terminal domain of PABP (PABC) is known to interact and activate major deadenylases complexes such as CCR4/NOT/Tob (Hoshino, 2012; Wahle and Winkler, 2013; Xie et al., 2014). Furthermore, CCR4/NOT has been found to promote decay even when its deadenylation capacity is deficient (Alhusaini and Coller, 2016). PABP has also been found to stimulate miRNA target regulation (Moretti et al., 2012). PABC also interacts with the miRISC factor GW182 (Tritschler et al., 2010). GW182 is responsible for actively recruiting deadenylases (such as CCR4/NOT) and decapping enzymes to promote decay of miRNA-repressed mRNAs (Tritschler et al., 2010). Conversely, when recruited by GW182, CCR4/NOT can effectively displace PABP from the tail (Zekri et al., 2013). Together, this evidence suggests that, perhaps the activated CCR4/NOT (by an interaction with PABP or GW182) could also remove PABPs from the tail and completely deadenylate the mRNA, triggering its rapid decay. Therefore, it is also possible that carrying multiple PABPs (due to having a long tail) negatively affects the transcript's stability.

5.3 Translation and the poly(A) tail

In early development, cytoplasmic extension of the poly(A) tail is responsible for activating translation of select genes (Mendez and Richter, 2001). The same has been observed for specific neuronal genes, which are locally activated for translation

by tail extension at the synapse (Norbury, 2013). Largely, these findings have been expanded into the general idea that a longer poly(A) tail was ideal for translation and stability (Eckmann et al., 2011; Goldstrohm and Wickens, 2008; Wahle and Winkler, 2013).

However, we were surprised to find that well-translated and stable mRNAs were strikingly associated with shorter poly(A) lengths across eukaryotes (Figures 3.2 and 3.4). In *C. elegans*, highly translated transcripts (such as ribosomal proteins) tend to have tails that accommodate only 1-2 PABPs. This process that associates translation with the shortening of the tail to a minimal functional length was named "poly(A) pruning".

There are still many questions yet to be answered about the mechanisms of poly(A) pruning. For instance: *does efficient translation promote pruning or is pruning a consequence of the stabilization of mRNAs due to translation?* For most organisms, we measured translation efficiency by analyzing the transcriptome for its ribosome occupancy and the frequency of optimal codons. Having a high fraction of optimal codons accelerates translation elongation rates, as tRNAs are readily available to the ribosome (Richter and Coller, 2015). The frequency of optimal codons can also greatly affect the stability of transcripts, and mRNAs with a majority of optimal codons are very stable (Presnyak et al., 2015). In fact, it is becoming increasingly clear that translation and stability are highly interwoven processes in which the ribosome plays the central role. For abnormal or defective mRNAs, the ribosome is in charge of quality control; for normal mRNAs, efficient translation prevents degradation (Presnyak et al., 2015; Roy and Jacobson, 2013; Shoemaker and Green, 2012).

So while it is a possibility that the longer lifespans of well-translated mRNAs enables their tails to reach short lengths, there is also evidence that translation can actively promote poly(A) pruning. For instance, during spermiogenesis, protamines and other transcripts are transcribed and remain untranslated for days until translation is developmentally activated. For several mRNAs, poly(A) shortening accompanies translation activation. Remarkably, untranslated mRNAs accumulate with long tails of ~160 nt and the translated pool is shortened to ~30nt, which corresponds to a single PABP (Kleene, 1989, 1993; Kleene et al., 1984; Steger, 1999).

Still, if translation actively promotes pruning, *what is the connection between the translation machinery and the poly(A) tail?* One candidate for connecting the translation machinery and the poly(A) tail is the translation terminator eRF3. As illustrated in Figure 3.5, eRF3 can bind simultaneously to eRF1 (the termination factor that interacts with the ribosome to recognize the stop codon) and the PABC domain of PABP (Hoshino, 2012). PABP also participates in a well-established interaction with the initiation factor eIF4G (Mangus et al., 2003). As both these interactions utilize different PABP domains (Xie et al., 2014), it is possible that binding to both eRF3 and eIF4G strongly anchors a proximal PABP to the tail, protecting the RNA from complete deadenylation and instead resulting in a pruned tail.

This raises the question: *are there any benefits to having a pruned tail?* As discussed earlier in section 5.2, it is possible that carrying excess PABP can have a negative effect on the stability of mRNAs. This model suggests that a short poly(A) length, achieved through translational activity, could limit the number of PABPs and contribute to the formation of a stable, efficient translation unit (Figure 3.5). Studies using strategies of translation inhibition, reporters with different levels of codon

optimization, and depletion of eRF3 and different deadenylases will be crucial to further clarify this proposed mechanism.

5.4 Future directions

As transcriptome-wide studies of the poly(A) profile are a recent possibility, there is a general lack of information about poly(A) dynamics in different organisms, tissues, and biological situations. Here, we have shown that the *C. elegans* poly(A) profile undergoes a striking shift towards shorter tails in aging worms (Figure 4.2).

I propose two scenarios that could explain this change in global tail lengths in aging. 1) As the worms become post-reproductive and old, the transcriptome could be shifted towards the transcription of housekeeping genes (which have short tails) and many regulatory transcripts (containing long tails) that were once upregulated during development and embryogenesis are no longer expressed or reduced. 2) Another possibility is that the change is not due to differential expression of housekeeping *versus* regulatory genes, but reflects a shift in gene-specific poly(A) profiles. That is, many genes that during development had long tails, shift towards shorter tails upon aging. Interestingly, a preliminary analysis of RNA-seq data comparing young and aged *C. elegans* revealed that transcript levels for most deadenylases are upregulated upon aging, while PABP expression is reduced (Figure 5.1). This suggests that deadenylation activity may be intensified in the older worms, which is consistent with the shift towards shorter poly(A) tails. To test these hypotheses, an important future direction would be to perform aTAIL-seq to measure poly(A) tails during *C. elegans* development and aging. The use of

long and short-lived animals, as well as deadenylase mutants will also aid in the elucidation of the mechanism behind this shift.

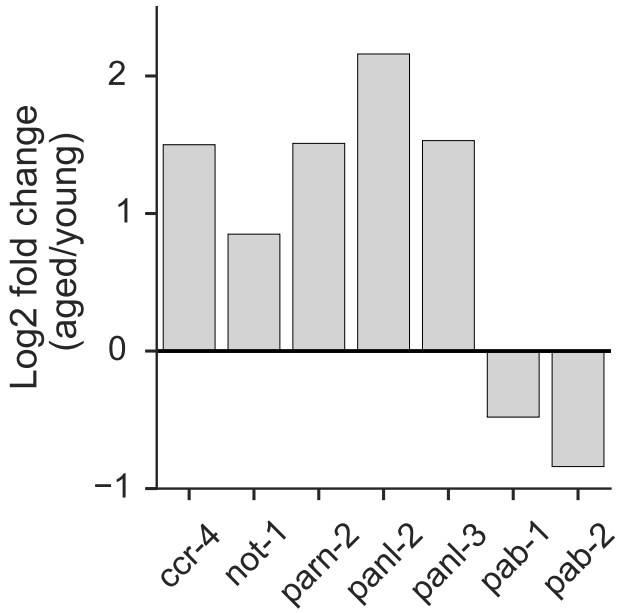


Figure 5.1: Preliminary analysis of RNA-seq data comparing last larval stage with day 5 adult worms. Several genes involved in poly(A) metabolism are differentially regulated upon aging. The deadenylases CC4/NOT (ccr-4 and not-1), PAN2/3 (panl-2 and panl-3), and PARN (parn-2) are upregulated at the mRNA level. On the other hand, the transcript levels of cytoplasmic PABPs (pab-1 and pab-2) are reduced in the older worms.

It is clear that many unanswered questions remain about the lengths and functions of poly(A) tails. Nonetheless, the recently developed methods to measure poly(A) tails on a global scale will likely speed discovery in the field. Below, I have outlined some of the most evident lines of enquiry that prompt further investigation:

- Does poly(A) tail length directly affect RNA stability?
- What is the mechanism behind poly(A) pruning?
- Does poly(A) pruning influence translation efficiency?
- What are the causes and consequences of the shift in the poly(A) profile during aging?
- How does the poly(A) profile respond to different stresses?
- Are there novel or conserved poly(A) tail modifications present in C. elegans and other organisms?

References

Aalto, A. P. and Pasquinelli, A. E. (2012). Small non-coding RNAs mount a silent revolution in gene expression. *Current Opinion in Cell Biology*, 24(3):333–340.

Abrahante, J. E., Daul, A. L., Li, M., Volk, M. L., Tennessen, J. M., Miller, E. A., and Rougvie, A. E. (2003). The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs.

Akashi, H. (1994). Synonymous codon usage in Drosophila melanogaster: Natural selection and translational accuracy. *Genetics*, 136(3):927–935.

Alhusaini, N. and Coller, J. (2016). The deadenylase components Not2p, Not3p, and Not5p promote mRNA decapping. *RNA (New York, N.Y.)*, pages 1–13.

Alvarez-Garcia, I. and Miska, E. A. (2005). MicroRNA functions in animal development and human disease. *Development*, 132(21):4653–4662.

Ambros, V. (2008). The evolution of our thinking about microRNAs. *Nature Medicine*, 14(10):1036–1040.

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., Dreyfuss, G., Eddy, S. R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., and Tuschl, T. (2003). A uniform system for microRNA annotation. *RNA (New York, N.Y.)*, 9(3):277–9.

Andachi, Y. (2008). A novel biochemical method to identify target genes of individual microRNAs: Identification of a new Caenorhabditis elegans let-7 target. *RNA*, 14(11):2440–2451.

Apponi, L. H., Leung, S. W., Williams, K. R., Valentini, S. R., Corbett, A. H., and Pavlath, G. K. (2010). Loss of nuclear poly(A)-binding protein 1 causes defects in myogenesis and mRNA biogenesis. *Human Molecular Genetics*, 19(6):1058–1065.

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, 455(7209):64–71.

Baer, B. W. and Kornberg, R. D. (1983). The protein responsible for the repeating structure of cytoplasmic poly(A)-ribonucleoprotein. *Journal of Cell Biology*, 96(March):717–721.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A. E. (2005). Regulation by let-7 and lin-4 miRNAs Results in Target mRNA Degradation. *Cell*, 122(4):553–563.

Bartel, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions.

Beilharz, T. H. and Preiss, T. (2007). Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA (New York, N.Y.)*, 13(7):982–97.

Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature reviews. Genetics*, 15(3):163–75.

Bernstein, E., Caudy, a. a., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409(6818):363–366.

Betel, D., Wilson, M., Gabow, A., Marks, D. S., and Sander, C. (2008). The microRNA.org resource: Targets and expression. *Nucleic Acids Research*, 36.

Braun, J. E., Huntzinger, E., and Izaurralde, E. (2013). The role of GW182 proteins in miRNA-mediated gene silencing. *Advances in Experimental Medicine and Biology*, 768:147–163.

Broughton, J. P. and Pasquinelli, A. E. (2013). Identifying argonaute binding sites in caenorhabditis elegans using iCLIP. *Methods*, 63(2):119–125.

Brown, C. E. and Sachs, A. B. (1998). Poly(A) Tail Length Control in Saccharomyces cerevisiae Occurs by Message-Specific Deadenylation. *Mol. Cell. Biol.*, 18(11):6548–6559.

Chang, H., Lim, J., Ha, M., and Kim, V. N. (2014). TAIL-seq: Genome-wide determination of poly(A) tail length and 3' end modifications. *Molecular Cell*, 53(6):1044–1052.

Charlesworth, A., Meijer, H. A., and de Moor, C. H. (2013). Specificity factors in cytoplasmic polyadenylation. *Wiley Interdisciplinary Reviews: RNA*, 4(4):437–461.

Chi, S. W., Zang, J. B., Mele, A., and Darnell, R. B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, 460(7254):479–86.

Chikina, M. D., Huttenhower, C., Murphy, C. T., and Troyanskaya, O. G. (2009). Global prediction of tissue-specific gene expression and context-dependent gene networks in Caenorhabditis elegans. *PLoS computational biology*, 5(6):e1000417.

Coller, J. and Parker, R. (2004). Eukaryotic mRNA decapping. *Annual review of biochemistry*, 73(1):861–90.

Coller, J. M., Gray, N. K., and Wickens, M. P. (1998). mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. *Genes & development*, 12:3226–3235.

Costello, J., Castelli, L. M., Rowe, W., Kershaw, C. J., Talavera, D., Mohammad-Qureshi, S. S., Sims, P. F. G., Grant, C. M., Pavitt, G. D., Hubbard, S. J., and Ashe, M. P. (2015). Global mRNA selection mechanisms for translation initiation. *Genome Biology*, 16(1):10.

Darnell, R. B. (2010). HITS-CLIP: Panoramic views of protein-RNA regulation in living cells. *Wiley Interdisciplinary Reviews: RNA*, 1(2):266–286.

Decker, C. J. and Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes & Development*, 7(8):1632–1643.

Ding, L., Spencer, A., Morita, K., and Han, M. (2005). The developmental timing regulator AIN-1 interacts with miRISCs and may target the argonaute protein ALG-1 to cytoplasmic P bodies in C. elegans. *Molecular Cell*, 19(4):437–447.

Ding, X. C. and Grosshans, H. (2009). Repression of C. elegans microRNA targets at the initiation level of translation requires GW182 proteins. *The EMBO journal*, 28(3):213–22.

Ding, X. C., Slack, F. J., and Großhans, H. (2008). The let-7 microRNA interfaces extensively with the translation machinery to regulate cell differentiation. *Cell Cycle*, 7(19):3083–3090.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21.

Eckmann, C. R., Rammelt, C., and Wahle, E. (2011). Control of poly(A) tail length.

Elbashir, S. M., Elbashir, S. M., Lendeckel, W., Lendeckel, W., Tuschl, T., and Tuschl, T. (2001). RNA interference is mediated 1- and 22-nucleotide RNAs. *Genes Dev.*, 15:188–200.

Esquela-Kerscher, A. and Slack, F. J. (2006). Oncomirs - microRNAs with a role in cancer. *Nature reviews. Cancer*, 6(4):259–69.

Fabian, M. R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annual review of biochemistry*, 79:351–379.

Finnegan, E. F. and Pasquinelli, A. E. (2013). MicroRNA biogenesis: regulating the regulators. *Critical reviews in biochemistry and molecular biology*, 48(1):51–68.

Friedman, R. C., Farh, K. K.-H., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome research*, 19(1):92–105.

Garneau, N. L., Wilusz, J., and Wilusz, C. J. (2007). The highways and byways of mRNA decay. *Nature reviews. Molecular cell biology*, 8(2):113–126.

Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., Yip, K. Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., Alves, P., Chateigner, A., Perry, M., Morris, M., Auerbach, R. K., Feng, X., Leng, J., Vielle, A., Niu, W., Rhrissorrakrai, K., Agarwal, A., Alexander, R. P., Barber, G., Brdlik, C. M., Brennan, J., Brouillet, J. J., Carr, A., Cheung, M.-S., Clawson, H., Contrino, S., Dannenberg, L. O., Dernburg, A. F., Desai, A., Dick, L., Dosé, A. C., Du, J., Egelhofer, T., Ercan, S., Euskirchen, G., Ewing, B., Feingold, E. A., Gassmann, R., Good, P. J., Green, P., Gullier, F., Gutwein, M., Guyer, M. S., Habegger, L., Han, T., Henikoff, J. G., Henz, S. R., Hinrichs, A., Holster, H., Hyman, T., Iniguez, A. L., Janette, J., Jensen, M., Kato, M., Kent, W. J., Kephart, E., Khivansara, V., Khurana, E., Kim, J. K., Kolasinska-Zwierz, P., Lai, E. C., Latorre, I., Leahey, A., Lewis, S., Lloyd, P., Lochovsky, L., Lowdon, R. F., Lubling, Y., Lyne, R., MacCoss, M., Mackowiak, S. D., Mangone, M., McKay, S., Mecenas, D., Merrihew, G., Miller, D. M., Muroyama, A., Murray, J. I., Ooi, S.-L., Pham, H., Phippen, T., Preston, E. A., Rajewsky, N., Rätsch, G., Rosenbaum, H., Rozowsky, J., Rutherford, K., Ruzanov, P., Sarov, M., Sasidharan, R., Sboner, A., Scheid, P., Segal, E., Shin, H., Shou, C., Slack, F. J., Slightam, C., Smith, R., Spencer, W. C., Stinson, E. O., Taing, S., Takasaki, T., Vafeados, D., Voronina, K., Wang, G., Washington, N. L., Whittle, C. M., Wu, B., Yan, K.-K., Zeller, G., Zha, Z., Zhong, M., Zhou, X., Ahringer, J., Strome, S., Gunsalus, K. C., Micklem, G., Liu, X. S., Reinke, V., Kim, S. K., Hillier, L. W., Henikoff, S., Piano, F., Snyder, M., Stein, L., Lieb, J. D., and Waterston, R. H. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science (New York, N.Y.)*, 330(6012):1775–87.

Goldstrohm, A. C. and Wickens, M. (2008). Multifunctional deadenylase complexes diversify mRNA control. *Nature reviews. Molecular cell biology*, 9(4):337–344.

Gowrishankar, G., Winzen, R., Dittrich-Breiholz, O., Redich, N., Kracht, M., and Holtmann, H. (2006). Inhibition of mRNA deadenylation and degradation by different types of cell stress. *Biological chemistry*, 387(3):323–327.

Grad, Y., Aach, J., Hayes, G. D., Reinhart, B. J., Church, G. M., Ruvkun, G., and Kim, J. (2003). Computational and experimental identification of C. elegans microRNAs. *Molecular Cell*, 11(5):1253–1263.

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research*, 34:D140–4.

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., and Mello, C. C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell*, 106(1):23–34.

Grosshans, H., Johnson, T., Reinert, K. L., Gerstein, M., and Slack, F. J. (2005). The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in C. elegans. *Developmental Cell*, 8(3):321–330.

Gu, S. G., Pak, J., Barberan-Soler, S., Ali, M., Fire, A., and Zahler, A. M. (2007). Distinct ribonucleoprotein reservoirs for microRNA and siRNA populations in C. elegans. *RNA*, 13(9):1492–1504.

Guo, H., Ingolia, N. T., Weissman, J. S., and Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308):835–840.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A. C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., and Tuschl, T. (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell*, 141(1):129–141.

Hammell, M., Long, D., Zhang, L., Lee, A., Carmack, C. S., Han, M., Ding, Y., and Ambros, V. (2008). mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nature methods*, 5(9):813–9.

Harris, T. W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W. J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., Fernandes, J., Han, M., Kishore, R., Lee, R., Muller, H.-M., Nakamura, C., Ozersky, P., Petcherski, A., Rangarajan, A., Rogers, A., Schindelman, G., Schwarz, E. M., Tuli, M. A., Van Auken, K., Wang, D., Wang, X., Williams, G., Yook, K., Durbin, R., Stein, L. D., Spieth, J., and Sternberg, P. W. (2010). WormBase: a comprehensive resource for nematode research. *Nucleic Acids Research*, 38:D463–D467.

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 153(3):654–665.

Hendriks, G.-J., Gaidatzis, D., Aeschimann, F., and Großhans, H. (2014). Extensive Oscillatory Gene Expression during C. elegans Larval Development. *Molecular Cell*, 53(3):380–392.

Hilgers, V., Teixeira, D., and Parker, R. (2006). Translation-independent inhibition of mRNA deadenylation during stress in Saccharomyces cerevisiae. *RNA (New York, N.Y.)*, 12(10):1835–45.

Hinnebusch, A. G., Ivanov, I. P., and Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science*, 352(6292):1413–1416.

Hoshino, S.-i. I. (2012). Mechanism of the initiation of mRNA decay: role of eRF3 family G proteins. *Wiley interdisciplinary reviews. RNA*, 3(6):743–57.

Hsu, C. L. and Stevens, A. (1993). Yeast cells lacking 5'–>3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Mol. Cell. Biol.*, 13(8):4826–4835.

Hu, W., Sweet, T. J., Chamnongpol, S., Baker, K. E., and Coller, J. (2009). Co-translational mRNA decay in Saccharomyces cerevisiae. *Nature*, 461(7261):225–9.

Hua, Y.-J., Tang, Z.-Y., Tu, K., Zhu, L., Li, Y.-X., Xie, L., and Xiao, H.-S. (2009). Identification and target prediction of miRNAs specifically expressed in rat neural tissue. *BMC genomics*, 10:214.

Hunter, S. E., Finnegan, E. F., Zisoulis, D. G., Lovci, M. T., Melnik-Martinez, K. V., Yeo, G. W., and Pasquinelli, A. E. (2013). Functional genomic analysis of the let-7 regulatory network in Caenorhabditis elegans. *PLoS Genet*, 9(3):e1003353.

Hutvagner, G. (2001). A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA. *Science*, 293(5531):834–838.

Ibáñez-Ventoso, C., Yang, M., Guo, S., Robins, H., Padgett, R. W., and Driscoll, M. (2006). Modulated microRNA expression during adult lifespan in Caenorhabditis elegans. *Aging Cell*, 5(3):235–246.

Ingolia, N. T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. *Nature reviews. Genetics*, 15(3):205–13.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., and Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (New York, N.Y.)*, 324(5924):218–23.

Isik, M., Korswagen, H. C., and Berezikov, E. (2010). Expression patterns of intronic microRNAs in Caenorhabditis elegans. *Silence*, 1(1):5.

Ivanov, A., Mikhailova, T., Eliseev, B., Yeramala, L., Sokolova, E., Susorov, D., Shuvalov, A., Schaffitzel, C., and Alkalaeva, E. (2016). PABP enhances release factor recruitment and stop codon recognition during translation termination. *Nucleic Acids Research*, page gkw635.

Jalkanen, A. L., Coleman, S. J., and Wilusz, J. (2014). Determinants and implications of mRNA poly(A) tail size - Does this protein make my tail look big? *Seminars in Cell & Developmental Biology*, 34:24–32.

Jan, C. H., Friedman, R. C., Ruby, J. G., and Bartel, D. P. (2011). Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. *Nature*, 469(7328):97–101.

Jänicke, A., Vancuylenberg, J., Boag, P. R., Traven, A., Beilharz, T. H., Janicke, A., Vancuylenberg, J., Boag, P. R., Traven, A., and Beilharz, T. H. (2012). ePAT: a simple method to tag adenylated RNA to measure poly(A)-tail length and other 3' RACE applications. *RNA (New York, N.Y.)*, 18(6):1289–95.

Jia, J., Yao, P., Arif, A., and Fox, P. L. (2013). Regulation and dysregulation of 3'UTR-mediated translational control. *Current opinion in genetics & development*, 23(1):29–34.

Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K. L., Brown, D., and Slack, F. J. (2005). RAS is regulated by the let-7 microRNA family. *Cell*, 120(5):635–647.

Johnston, R. J. and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. *Nature*, 426(6968):845–849.

Jonas, S. and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nature reviews. Genetics*, 16(7):421–433.

Jovanovic, M., Reiter, L., Clark, A., Weiss, M., Picotti, P., Rehrauer, H., Frei, A., Neukomm, L. J., Kaufman, E., Wollscheid, B., Simard, M. J., Miska, E. A., Aebersold, R., Gerber, A. P., and Hengartner, M. O. (2012). RIP-chip-SRM - A new combinatorial large-scale approach identifies a set of translationally regulated bantam/miR-58 targets in C. elegans. *Genome Research*, 22(7):1360–1371.

Jovanovic, M., Reiter, L., Picotti, P., Lange, V., Bogan, E., Hurschler, B. a., Blenkiron, C., Lehrbach, N. J., Ding, X. C., Weiss, M., Schrimpf, S. P., Miska, E. a., Grosshans, H., Aebersold, R., and Hengartner, M. O. (2010). A quantitative targeted proteomics approach to validate predicted microRNA targets in C. elegans. *Nature methods*, 7(10):837–42.

Kahvejian, A., Svitkin, Y. V., Sukarieh, R., M'Boutchou, M.-N., and Sonenberg, N. (2005). Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes & development*, 19(1):104–13.

Kato, M., Chen, X., Inukai, S., Zhao, H., and Slack, F. J. (2011). Age-associated changes in expression of small, noncoding RNAs, including microRNAs, in C. elegans. *RNA (New York, N.Y.)*, 17(10):1804–20.

Kato, M., De Lencastre, A., Pincus, Z., and Slack, F. J. (2009). Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during C. elegans development. *Genome Biology*, 10(5):R54.

Kaufman, E. J. and Miska, E. A. (2010). The microRNAs of Caenorhabditis elegans.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. TL - 39. *Nature genetics*, 39 VN - r(10):1278–1284.

Kleene, K. C. (1989). Poly(A) shortening accompanies the activation of translation of five mRNAs during spermiogenesis in the mouse. *Development (Cambridge, England)*, 106(2):367–373.

Kleene, K. C. (1993). Multiple controls over the efficiency of translation of the mRNAs encoding transition proteins, protamines, and the mitochondrial capsule selenoprotein in late spermatids in mice.

Kleene, K. C., Cataldo, L., Mastrangelo, M. A., and Tagne, J. B. (2003). Alternative patterns of transcription and translation of the ribosomal protein L32 mRNA in somatic and spermatogenic cells in mice. *Experimental Cell Research*, 291(1):101–110.

Kleene, K. C., Distel, R. J., and Hecht, N. B. (1984). Translational regulation and deadenylation of a protamine mRNA during spermiogenesis in the mouse. *Developmental Biology*, 105(1):71–79.

Kloosterman, W. P. and Plasterk, R. H. (2006). The diverse functions of microRNAs in animal development and disease. *Dev Cell*, 11(4):441–450.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D. J., Luscombe, N. M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Structural & Molecular Biology*, 17(7):909–915.

Kozomara, A. and Griffiths-Jones, S. (2011). MiRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, 39(SUPPL. 1).

Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nature reviews. Genetics*, 11(9):597–610.

Kudlow, B. A., Zhang, L., and Han, M. (2012). Systematic Analysis of Tissue-Restricted miRISCs Reveals a Broad Role for MicroRNAs in Suppressing Basal Activity of the C. elegans Pathogen Response. *Molecular Cell*, 46(4):530–541.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science (New York, N.Y.)*, 294(5543):853–8.

Lall, S., Grün, D., Krek, A., Chen, K., Wang, Y. L., Dewey, C. N., Sood, P., Colombo, T., Bray, N., MacMenamin, P., Kao, H. L., Gunsalus, K. C., Pachter, L., Piano, F., and Rajewsky, N. (2006). A genome-wide map of conserved MicroRNA targets in C. elegans. *Current Biology*, 16(5):460–471.

Lancman, J. J., Caruccio, N. C., Harfe, B. D., Pasquinelli, A. E., Schageman, J. J., Pertsemlidis, A., and Fallon, J. F. (2005). Analysis of the regulation of lin-41 during chick and mouse limb development. *Developmental Dynamics*, 234(4):948–960.

Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science (New York, N.Y.)*, 294(5543):858–62.

Lee, R. C. and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. *Science (New York, N.Y.)*, 294(5543):862–4.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854.

Lehrbach, N. J., Castro, C., Murfitt, K. J., Abreu-Goodger, C., Griffin, J. L., and Miska, E. a. (2012). Post-developmental microRNA expression is required for normal physiology, and regulates aging in parallel to insulin/IGF-1 signaling in C. elegans. *RNA*, 18(12):2220–2235.

Leung, A. K., Young, A. G., Bhutkar, A., Zheng, G. X., Bosson, A. D., Nielsen, C. B., and Sharp, P. A. (2011). Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol*, 18(2):237–244.

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1):15–20.

Lim, J., Ha, M., Chang, H., Kwon, S., Simanshu, D., Patel, D., and Kim, V. (2014). Uridylation by TUT4 and TUT7 Marks mRNA for Degradation. *Cell*, 159(6):1365–1376.

Lim, J., Lee, M., Son, A., Chang, H., and Kim, V. N. (2016). mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. *Genes & development*, 30(14):1671–82.

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S., and Johnson, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 433(7027):769–73.

Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B., and Bartel, D. P. (2003). The microRNAs of Caenorhabditis elegans. *Genes & development*, 17(8):991–1008.

Lin, S. Y., Johnson, S. M., Abraham, M., Vella, M. C., Pasquinelli, A., Gamberi, C., Gottlieb, E., and Slack, F. J. (2003). The C. elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable MicroRNA target.

Llave, C., Xie, Z., Kasschau, K. D., and Carrington, J. C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science (New York, N.Y.)*, 297(5589):2053–2056.

Macdonald, C. C. and Redondo, J. L. (2002). Reexamining the polyadenylation signal: Were we wrong about AAUAAA? *Molecular and Cellular Endocrinology*, 190(1-2):1–8.

Mangone, M., Macmenamin, P., Zegar, C., Piano, F., and Gunsalus, K. C. (2008). UTRome.org: A platform for 3'UTR biology in C. elegans. *Nucleic Acids Research*, 36(SUPPL. 1).

Mangone, M., Manoharan, A. P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S. D., Mis, E., Zegar, C., Gutwein, M. R., Khivansara, V., Attie, O., Chen, K., Salehi-Ashtiani, K., Vidal, M., Harkins, T. T., Bouffard, P., Suzuki, Y., Sugano, S., Kohara, Y., Rajewsky, N., Piano, F., Gunsalus, K. C., and Kim, J. K. (2010). The landscape of C. elegans 3'UTRs. *Science (New York, N.Y.)*, 329(5990):432–5.

Mangus, D. A., Evans, M. C., and Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome biology*, 4(7):223.

Maquat, L. E., Tarn, W.-Y. Y., and Isken, O. (2010). The pioneer round of translation: features and functions. *Cell*, 142(3):368–74.

Maragkakis, M., Reczko, M., Simossis, V. A., Alexiou, P., Papadopoulos, G. L., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., Vergoulis, T., Koziris, N., Sellis, T., Tsanakas, P., and Hatzigeorgiou, A. G. (2009). DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic acids research*, 37:W273–6.

Marshall, E., Stansfield, I., and Romano, M. C. (2014). Ribosome recycling induces optimal translation rate at low ribosomal availability. *Journal of the Royal Society, Interface / the Royal Society*, 11 VN - r(98):20140589.

Martinez, N. J., Ow, M. C., Reece-Hoyes, J. S., Barrasa, M. I., Ambros, V. R., and Walhout, A. J. M. (2008). Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. *Genome Research*, 18(12):2005–2015.

Mendez, R. and Richter, J. D. (2001). Translational control by CPEB: a means to the end. *Nature reviews. Molecular cell biology*, 2(7):521–529.

Miranda, K. C., Huynh, T., Tay, Y., Ang, Y. S., Tam, W. L., Thomson, A. M., Lim, B., and Rigoutsos, I. (2006). A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell*, 126(6):1203–1217.

Miska, E. A., Alvarez-Saavedra, E., Abbott, A. L., Lau, N. C., Hellman, A. B., McGonagle, S. M., Bartel, D. P., Ambros, V. R., and Horvitz, H. R. (2007). Most Caenorhabditis elegans microRNAs Are Individually Not Essential for Development or Viability. *PLoS Genetics*, 3(12):e215.

Mondol, V. and Pasquinelli, A. E. (2012). Let's Make It Happen. The Role of let-7 MicroRNA in Development. *Current Topics in Developmental Biology*, 99:1–30.

Moretti, F., Kaiser, C., Zdanowicz-Specht, A., and Hentze, M. W. (2012). PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding. *Nature Structural & Molecular Biology*, 19(6):603–608.

Nam, J.-w. and Bartel, D. P. (2012). Long noncoding RNAs in C. elegans. *Genome research*, 22(12):2529–40.

Norbury, C. J. (2013). Cytoplasmic RNA: a case of the tail wagging the dog. *Nat Rev Cancer*, 13(10):643–653.

Nousch, M., Techritz, N., Hampel, D., Millonigg, S., and Eckmann, C. R. (2013). The Ccr4-Not deadenylase complex constitutes the main poly(A) removal activity in C. elegans. *Journal of cell science*, 126(Pt 18):4274–85.

O'Farrell, F., Esfahani, S. S., Engström, Y., and Kylsten, P. (2008). Regulation of the Drosophila lin-41 homologue dappled by let-7 reveals conservation of a regulatory mechanism within the LIN-41 subclade. *Developmental Dynamics*, 237(1):196–208.

Papadopoulos, G. L., Reczko, M., Simossis, V. A., Sethupathy, P., and Hatzigeorgiou, A. G. (2009). The database of experimentally supported targets: A functional update of TarBase. *Nucleic Acids Research*, 37.

Park, J.-E., Yi, H., Kim, Y., Chang, H., and Kim, V. N. (2016). Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Molecular Cell*, 62(3):462–471.

Parker, R. (2012). RNA degradation in Saccharomyces cerevisae. *Genetics*, 191(3):671–702.

Pasquinelli, A. E. (2012). MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nature reviews. Genetics*, 13(4):271–82.

Pasquinelli, a. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Müller, P., Spring, J., Srinivasan, a., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808):86–89.

Pelechano, V., Wei, W., and Steinmetz, L. M. (2015). Widespread co-translational RNA decay reveals ribosome dynamics. *Cell*, 161(6):1400–1412.

Porta-de-la Riva, M., Fontrodona, L., Villanueva, A., and Cerón, J. (2012). Basic Caenorhabditis elegans methods: synchronization and observation. *Journal of visualized experiments : JoVE*, (64):e4019.

Presnyak, V. (2015). *Effects of Codon Usage on Mrna Translation and Decay*. PhD thesis.

Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K. E., Graveley, B. R., and Coller, J. (2015). Codon optimality is a major determinant of mRNA stability. *Cell*, 160(6):1111–1124.

Quax, T. E. F., Claassens, N. J., Söll, D., and van der Oost, J. (2015). Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular cell*, 59(2):149–61.

Quenault, T., Lithgow, T., and Traven, A. (2011). PUF proteins: repression, activation and mRNA localization. *Trends in Cell Biology*, 21(2):104–112.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, a. E., Bettinger, J. C., Rougvie, a. E., Horvitz, H. R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403(6772):901–906.

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., and Bartel, D. P. (2002). MicroRNAs in plants. *Genes and Development*, 16(13):1616–1626.

Richter, J. D. and Coller, J. (2015). Perspective Pausing on Polyribosomes : Make Way for Elongation in Translational Control. *Cell*, 163(2):292–300.

Ritchie, W., Flamant, S., and Rasko, J. E. J. (2009). Predicting microRNA targets and functions: traps for the unwary. *Nature Methods*, 6(6):397–398.

Roy, B. and Jacobson, A. (2013). The intimate relationships of mRNA decay and translation. *Trends in Genetics*, 29(12):691–699.

Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D. P. (2006). Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in C. elegans. *Cell*, 127(6):1193–1207.

Sachs, a. B., Davis, R. W., and Kornberg, R. D. (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):3268–3276.

Sallés, F. J., Richards, W. G., and Strickland, S. (1999). Assaying the polyadenylation state of mRNAs. *Methods (San Diego, Calif.)*, 17(1):38–45.

Sato, H. and Maquat, L. E. (2009). Remodeling of the pioneer translation initiation complex involves translation and the karyopherin importin. *Genes & Development*, 23(21):2537–2550.

Schulman, B. R. M., Esquela-Kerscher, A., and Slack, F. J. (2005). Reciprocal expression of lin-41 and the microRNAs let-7 and mir-125 during mouse embryogenesis. *Developmental Dynamics*, 234(4):1046–1054.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209):58–63.

Shoemaker, C. J. and Green, R. (2012). Translation drives mRNA quality control. *Nature structural & molecular biology*, 19(6):594–601.

Slack, F. J., Basson, M., Liu, Z., Ambros, V., Horvitz, H. R., and Ruvkun, G. (2000). The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Molecular cell*, 5(4):659–669.

Smith, B. L., Gallie, D. R., Le, H., and Hansma, P. K. (1997). Visualization of Poly(A)-Binding Protein Complex Formation with Poly(A) RNA Using Atomic Force Microscopy. *Journal of Structural Biology*, 119(2):109–117.

Stadler, M., Artiles, K., Pak, J., and Fire, A. (2012). Contributions of mRNA abundance, ribosome loading, and post- or peri-translational effects to temporal repression of C. elegans heterochronic miRNA targets. *Genome Research*, 22(12):2418–2426.

Steger, K. (1999). Transcriptional and translational regulation of gene expression in haploid spermatids. *Anatomy and embryology*, 199(6):471–487.

Stenico, M., Lloyd, A. T., and Sharp, P. M. (1994). Codon usage in Caenorhabditis elegans : delineation of translational selection and mutational biases. *Nucleic Acids Research*, 22(13):2437–2446.

Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H., and Bartel, D. P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature*, 508(7494):66–71.

Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome biology*, 13(8):R67.

Sun, M., Schwalb, B., Pirkl, N., Maier, K. C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Global analysis of Eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Molecular Cell*, 52(1):52–62.

Thomas, M., Lieberman, J., and Lal, A. (2010). Desperately seeking microRNA targets. *Nature structural & molecular biology*, 17(10):1169–1174.

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

Tritschler, F., Huntzinger, E., and Izaurralde, E. (2010). Role of GW182 proteins and PABPC1 in the miRNA pathway: a sense of déjà vu. *Nature reviews. Molecular cell biology*, 11(5):379–384.

Tucker, M., Valencia-Sanchez, M. A., Staples, R. R., Chen, J., Denis, C. L., and Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. *Cell*, 104(3):377–386.

Van Wynsberghe, P. M., Kai, Z. S., Massirer, K. B., Burton, V. H., Yeo, G. W., and Pasquinelli, A. E. (2011). LIN-28 co-transcriptionally binds primary let-7 to regulate miRNA maturation in Caenorhabditis elegans. *Nat Struct Mol Biol*, 18(3):302–308.

Wahle, E. and Winkler, G. S. (2013). RNA decay machines: Deadenylation by the Ccr4-Not and Pan2-Pan3 complexes. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(6-7):561–570.

Wang, Z., Day, N., Trifillis, P., and Kiledjian, M. (1999). An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA in vitro. *Molecular and cellular biology*, 19(7):4552–4560.

Warf, M. B., Johnson, W. E., and Bass, B. L. (2011). Improved annotation of C. elegans microRNAs by deep sequencing reveals structures associated with processing by Drosha and Dicer. *RNA (New York, N.Y.)*, 17(4):563–77.

Weingarten-Gabbay, S., Elias-Kirma, S., Nir, R., Gritsenko, A. A., Stern-Ginossar, N., Yakhini, Z., Weinberger, A., and Segal, E. (2016). Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science*, 351(6270):1–24.

Wells, S. E., Hillner, P. E., Vale, R. D., and Sachs, A. B. (1998). Circularization of mRNA by Eukaryotic Translation Initiation Factors. *Molecular Cell*, 2(1):135–140.

Wickramasinghe, V. O. and Laskey, R. A. (2015). Control of mammalian gene expression by selective mRNA export. *Nature reviews. Molecular cell biology*, 16(7):431–442.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, 75(5):855–862.

Wong, Y. Y., Moon, A., Duffin, R., Barthet-Barateig, A., Meijer, H. a., Clemens, M. J., and de Moor, C. H. (2010). Cordycepin inhibits protein synthesis and cell adhesion through effects on signal transduction. *The Journal of biological chemistry*, 285(4):2610–21.

Xie, J., Kozlov, G., and Gehring, K. (2014). The "tale" of poly(A) binding protein: The MLLE domain and PAM2-containing proteins. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(11):1062–1068.

Yang, J. H., Li, J. H., Shao, P., Zhou, H., Chen, Y. Q., and Qu, L. H. (2011). StarBase: A database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Research*, 39(SUPPL. 1).

Zekri, L., KuzuoÄ§lu-Öztürk, D., and Izaurralde, E. (2013). GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. *The EMBO journal*, 32(7):1052–65.

Zhang, L., Ding, L., Cheung, T. H., Dong, M.-Q., Chen, J., Sewell, A. K., Liu, X., Yates, J. R., and Han, M. (2007). Systematic identification of C. elegans miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Molecular cell*, 28(4):598–613.

Zhang, L., Hammell, M., Kudlow, B. a., Ambros, V., and Han, M. (2009). Systematic analysis of dynamic miRNA-target interactions during C. elegans development. *Development (Cambridge, England)*, 136(18):3043–3055.

Zisoulis, D. G., Lovci, M. T., Wilbert, M. L., Hutt, K. R., Liang, T. Y., Pasquinelli, A. E., and Yeo, G. W. (2010). Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. *Nature structural & molecular biology*, 17(2):173–9.

Zisoulis, D. G., Yeo, G. W., and Pasquinelli, A. E. (2011). Comprehensive identification of miRNA target sites in live animals. *Methods in molecular biology (Clifton, N.J.)*, 732:169–85.