

Regulatory Consequences of mRNA Poly(A)-Tail Length Changes

by

Stephen W. Eichhorn

B.A., Biochemistry and Molecular Biology (2010)
Reed College

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2017

© 2017 Massachusetts Institute of Technology
All rights reserved

Signature redacted

Signature of Author: _____

Stephen W. Eichhorn
Department of Biology
May 25, 2017

Signature redacted

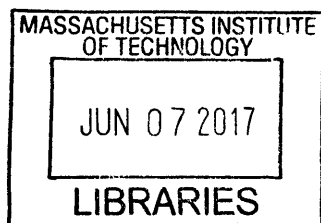
Certified by: _____

David P. Bartel
Professor of Biology
Thesis Supervisor

Signature redacted

Accepted by: _____

Amy E. Keating
Professor of Biology
Co-Chair, Biology Graduate Committee



ARCHIVES



77 Massachusetts Avenue
Cambridge, MA 02139
<http://libraries.mit.edu/ask>

DISCLAIMER NOTICE

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available.

Thank you.

Appendix section B contains text that runs off the bottom page margin.

Regulatory Consequences of mRNA Poly(A)-Tail Length Changes

by

Stephen W. Eichhorn

Submitted to the Department of Biology on May 25, 2017
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Abstract

Eukaryotic mRNAs have a cap structure at their 5' ends and a poly(A) tail at their 3' ends, and the proteins that bind these features increase the stability and translation of an mRNA. The influence of the poly(A) tail on translation was discovered decades ago, but primarily with regard to the idea that an mRNA with a poly(A) tail is better translated than one without. The influence of differences in tail length on translation had been assessed for just a few mRNAs, and in these cases long-tailed mRNAs were better translated than short-tailed mRNAs.

We measured the poly(A)-tail length and translational efficiency of mRNAs corresponding to thousands of different genes in 35 different cell types or contexts. Extending previous single-gene studies, we found a global relationship between tail length and translational efficiency in *Drosophila* oocytes, and *Drosophila*, *Xenopus*, and zebrafish embryos. Surprisingly, in all three species, the strong coupling between tail length and translational efficiency was lost once the embryos reached gastrulation, and there was no coupling in the post-embryonic contexts we examined. We thus demonstrated that poly(A)-tail length is a major determinant of translational efficiency during early animal development and discovered a broadly conserved developmental switch in translational control.

During the tail-length regulatory regime of the early embryo, a protein or microRNA might regulate translation by changing the poly(A)-tail length of an mRNA, interacting with the translation machinery, or both mechanisms. We characterized the mechanism used by two translational regulatory proteins in *Drosophila*, finding that they predominantly act by regulating tail length. Likewise, in early zebrafish embryos, microRNAs repress the translation of their hundreds of mRNA targets by shortening poly(A) tails. Our findings indicate that much of the translational regulation in early development is achieved by regulating poly(A)-tail lengths.

Outside of early embryonic contexts, microRNAs regulate gene expression by causing both translational repression and mRNA degradation. We greatly expanded the mammalian cell types and contexts in which the steady-state and pre-steady-state effects of a microRNA had been examined globally for endogenous mRNAs. In all post-embryonic contexts with substantial microRNA-mediated repression, the predominant mode of repression was mRNA degradation.

Thesis Supervisor: David P. Bartel
Title: Professor

Acknowledgements

I have had the good fortune of working with the most outstanding group of scientists during my time in graduate school. The members of my lab, the MIT biology community, and collaborators outside of MIT are responsible for my growth and development as a scientist, and have deeply affected how I think about science. My advisor David Bartel taught me how to rigorously and carefully go from an initial question to a meaningful finding. Our conversations and his advice have had the largest impact on how I approach scientific problems.

I thank my thesis committee members, Wendy Gilbert and Phil Sharp, for their advice and guidance throughout my graduate research. I am also thankful to Wendy for encouraging me to more readily share my thoughts with others. I thank Terry Orr-Weaver for her mentorship and interesting discussions.

I am indebted to the past and present member of the Bartel lab—your suggestions, creativity, and rigor have been a large part of what I enjoy so much about our lab. Thank you Vikram, Vincent, Namita, Danny, Grace, Anna, Tim, Wenwen, David G, Matt, Justin, Huili, Junjie, Katrin, Sue-Jean, Wendy, Elena, Ben, David K., Jamie, Kathy, Sean, Jeff, Jin-Wu, Michael, Olivia, Charlie, Lori, Lena, Noah, Asia, Igor, David W., Xuebing, and Coffee. I am also very grateful to Laura Resteghini, our administrative lab manager, for lots of help over the years. I have particularly relied on discussion with Jamie, Tim, Sean and Ben. Jamie helps me see results from a new perspective and tie together seemingly disparate findings. Tim brings his tireless work ethic and creativity to all of our conversations. Sean has helped me think more quantitatively about biological problems, and Ben always has a careful and considered perspective.

All of my research has been done in close collaboration with others. My collaboration with Alex was a highlight of my graduate research. I learned a lot from him, and I quickly found out that his rigor and attention to detail meant that when he suggested doing something a certain way, there was a really good reason for it. My collaboration with Iva opened up exciting new avenues of research and taught me a great deal about developmental biology. Her knowledge and foresight led to many of our most interesting observations. My research would not have been possible without the excellence of Sumeet and the entire Whitehead Genome Technology Core. Thanks to my many other collaborators for introducing new and exciting areas of biology to me.

I am also lucky to have many people that bring balance, excitement and fun to my life. Foremost, Jamie, your support means a great deal to me and I am always excited to go on adventures with you. Cory, Tim and Christi, thanks for your friendship, fun meals and weekends away. Thanks to my classmates (especially Matt, Kara, Stacie, and Ben), running friends, and the St. Louis contingent for the fun times we have shared.

Finally, I want to thank my family. Thanks to my parents for their enduring support, for instilling in me the value of education, and for all they have done and continue to do for me. I thank my sisters for their friendship, and my nieces for being so very fun. I thank my aunts, uncles, cousins and other relatives for their support and encouragement. I thank my grandmother for believing in me and prizing education so highly.

Table of Contents

Abstract	3
Acknowledgements	5
Chapter 1. Introduction	8
mRNA 5' and 3' ends	9
Cap-dependent translation initiation	13
MicroRNA-mediated repression	16
Translational regulation	21
Translational regulation during animal development	24
Figures.....	12
References.....	31
Chapter 2. Poly(A)-tail profiling reveals an embryonic switch in translational control	41
Chapter 3. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues	93
Chapter 4. mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in <i>Drosophila</i> oocytes and early embryos	148
Chapter 5. Future Directions	203
Appendix A. Widespread changes in the posttranscriptional landscape at the <i>Drosophila</i> oocyte-to-embryo transition	216
Appendix B. Altered translation of GATA1 in Diamond-Blackfan anemia	231
Appendix C. Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation	241
Appendix D. Genome-Scale Networks Link Neurodegenerative Disease Genes to α-Synuclein through Specific Molecular Pathways	255
<i>Curriculum vitae</i>	285

Chapter 1

Introduction

Eukaryotic and prokaryotic mRNAs are strikingly different at their termini. The 5' end of a eukaryotic mRNA has a 7-methylguanosine “cap”, whereas prokaryotic mRNAs have a 5' triphosphate. At the 3' end, eukaryotic mRNAs have a poly(A) tail (a long, non-templated stretch of adenosines), whereas prokaryotic mRNAs end with genome-templated sequence. These divergent 5' and 3' ends underlie many of the fundamental differences in how organisms in these domains of life translate and degrade their mRNAs.

Prokaryotic translation occurs through an internal initiation mechanism, wherein a ribosome binds within an mRNA and initiates translation at that site. Many prokaryotic mRNAs encode multiple proteins within an individual transcript, and the internal initiation mechanism enables translation of all the polypeptides encoded by these polycistronic mRNAs (Kozak, 1999). The observation that eukaryotic mRNAs were likely monocistronic and initiated translation at a single site provided an early indication that eukaryotic translation initiation operates through a different mechanism (Jacobson and Baltimore, 1968). Evaluating the known mRNA sequences and translation products revealed that translation began at the most upstream start codon of each transcript, and led to the prediction that eukaryotic translation initiation occurs through a scanning mechanism in which the 40S ribosome is recruited to the 5' end of the mRNA and then scans 5' to 3' for the start codon (Kozak, 1978). The 5' cap, ubiquitous among eukaryotic mRNAs (Shatkin, 1976), provided a plausible feature for the recruitment of ribosomes to the 5' end of mRNAs (Kozak, 1978), and this model of recruitment was strengthened by evidence that the 5' cap greatly increased both mRNA stability and translation (Shatkin, 1976). The discovery of cap-binding proteins that stimulate the translation of capped

RNAs (Sonenberg et al., 1978; Sonenberg et al., 1979), and the identification of an interaction between the cap-binding proteins and an initiation factor that associates with the 40S ribosome (Sonenberg et al., 1978; Sonenberg et al., 1979), filled out the molecular details of ribosomal recruitment to mRNA 5' ends. Subsequent research has borne out predictions made by the scanning model of translation initiation (Kozak, 1989).

With translation initiation happening entirely at the 5' end of an mRNA, it initially seemed odd that the poly(A) tail (Rosenthal et al., 1983; Huarte et al., 1987; Rosenthal and Ruderman, 1987; McGrew et al., 1989; Gallie, 1991) and proteins binding to the 3' untranslated region (UTR) of an mRNA (Sonenberg, 1994) were able to regulate translation initiation. Proteins bound at the 3' end of an mRNA have since been shown to interact, directly or indirectly, with translation initiation factors at the 5' end, and the regulatory potential of the 3' ends of mRNAs is now well appreciated (Sonenberg and Hinnebusch, 2009). In general these mechanistic studies evaluated only a handful of mRNAs or reporter RNAs, leaving open many fundamental questions about translational regulation. This dissertation describes global studies of translational regulation and poly(A)-tail length for endogenous mRNAs in a variety of systems. The findings from these studies extend and clarify previous ideas about translational regulation, and identify novel aspects of translational control.

mRNA 5' and 3' ends

Several nuclear processing events extensively modify eukaryotic mRNAs during and immediately following their transcription. Just after the onset of transcription, a 7-methylguanosine is added to the 5' end of the mRNA. As transcription continues, the intronic sequences that interrupt adjacent exons are spliced out, and once fully synthesized, the mRNA is

cleaved to form the 3' end of the transcript. The 3' end of nearly all mRNAs is then extended with a string of adenosines (Proudfoot et al., 2002). Whereas the correct splicing of an mRNA is essential for it to encode the intended protein product and contain the intended regulatory sequences, the 5' and 3' modifications that bookend the mRNA engender it with increased stability and translational capacity in the cytoplasm.

Degradation of most eukaryotic mRNAs occurs in an ordered fashion that begins at the 3' end of the mRNA (Chen and Shyu, 2011). In a typical mammalian cell, an mRNA emerges from the nucleus with a poly(A) tail that is approximately 200 nucleotides (nts) long (Proudfoot et al., 2002), and shortening the poly(A) tail is the first and rate-limiting step of degradation (Chen and Shyu, 2011). Deadenylases shorten the poly(A) tail one adenosine at a time to a final length of 10 to 60 nts (Chen and Shyu, 2011). Decapping activators preferentially bind to short-tailed mRNAs (Chowdhury et al., 2007) and recruit the decapping complex, which cleaves off the 5' cap and leaves behind an mRNA with an accessible 5' phosphate (Coller and Parker, 2004). XRN1, the predominant 5' to 3' cytoplasmic exoribonuclease, catalyzes the final step of degradation by recognizing the 5' phosphate and degrading the mRNA into individual nucleotides (Chen and Shyu, 2011). Although the absolute poly(A)-tail lengths of mRNAs differ in other eukaryotic systems, the general principle that deadenylation precedes decapping, and is rate limiting for the degradation of most mRNAs, is deeply conserved among eukaryotes (Coller and Parker, 2004; Chen and Shyu, 2011).

Proteins that bind the 5' cap and poly(A) tail confer greater stability on mRNAs. Poly(A) tails are generally thought to exist as a complex with cytoplasmic poly(A)-binding protein (PABP) (Mangus et al., 2003; Goldstrohm and Wickens, 2008), implying that in most contexts PABP is sufficiently abundant to bind all polyadenylated mRNAs. Indeed, in the cell culture

setting where PABP abundance has been measured quantitatively there was three-fold more PABP than necessary to bind all of the possible binding sites in poly(A) tails (Gorlach et al., 1994). PABP bound to a poly(A) tail decreases the rate of deadenylation by both physically protecting the poly(A) tail and specifically antagonizing some of the deadenylases complexes (Mangus et al., 2003; Goldstrohm and Wickens, 2008). PABP also seems to prevent premature recruitment of the decapping activators, which ensures that decapping does not bypass deadenylation (Caponigro and Parker, 1995). At the other end of the mRNA, the 5' cap is bound by the translation initiation factor eIF4E, which decreases the accessibility of the 5' cap to the decapping machinery (Schwartz and Parker, 2000). Even without eIF4E, the 5' cap increases the stability of an mRNA because the 5'-5' triphosphate linkage that connects the 7-methylguanosine to the adjacent nucleotide renders the mRNA inaccessible to XRN1 (Stevens, 1978). An mRNA without a poly(A) tail would be rapidly decapped and degraded in the cytoplasm, and would likely have a cellular half-life similar to that of an mRNA without a 5' cap.

Beyond their influence on mRNA stability, the proteins that bind the 5' cap and poly(A) tail increase how efficiently an mRNA is translated. eIF4E binds to the 5' cap as part of the eIF4F complex, which also contains eIF4G and eIF4A (Hinnebusch, 2011)(Figure 1). Collectively, this complex promotes the translation of the bound mRNA by recruiting the ribosome to the mRNA, and by resolving mRNA secondary structure that might otherwise impede ribosome binding (Hinnebusch, 2011). The mechanism by which PABP bound to the poly(A) tail influences translation, and the magnitude of its effect, is less clear. PABP is thought to promote translation by interacting with eIF4G, forming a molecular bridge between the 5' and 3' ends of the mRNA (Wells et al., 1998; Hinnebusch, 2011)(Figure 1). This closed-loop

conformation may promote translation by stabilizing the binding of eIF4F to the 5' cap (Kahvejian et al., 2005; Amrani et al., 2008; Hinnebusch, 2011). There has also been speculation

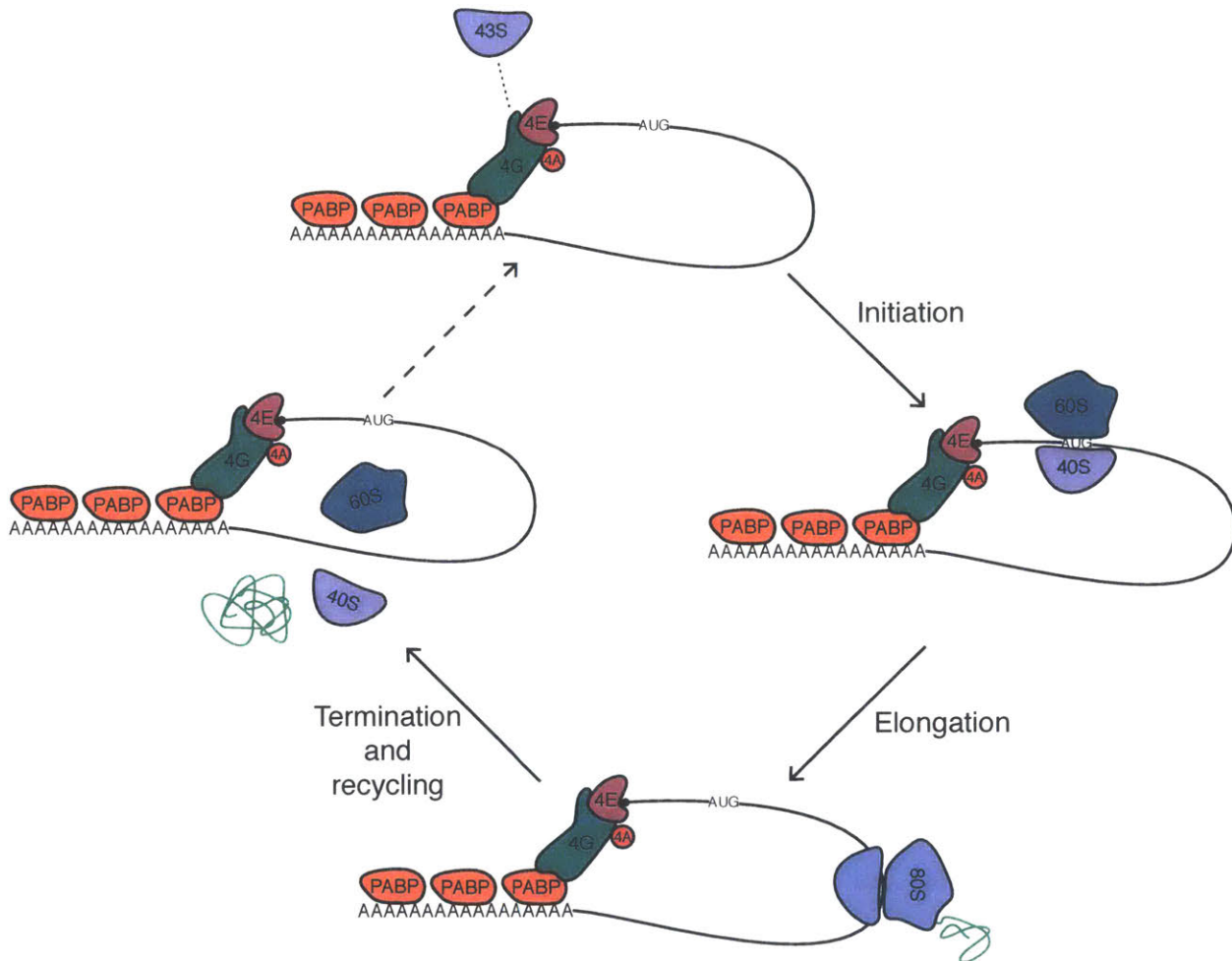


Figure 1. Overview of cap-dependent translation

eIF4E binds to the 5' cap of the mRNA, PABP binds to the poly(A) tail, and eIF4G bridges the 5' and 3' ends of the mRNA by binding both eIF4E and PABP. The 43S preinitiation complex (PIC), composed of the 40S ribosomal subunit and several initiation factors, is recruited to the 5' end of the mRNA by eIF4G (interaction shown as a dashed line). Once bound, the 43S PIC scans towards the 3' end of the mRNA until it reaches the start codon of the open reading frame (ORF). The initiation phase concludes when the 60S ribosomal subunit joins the 40S ribosomal subunit, forming an 80S ribosome ready to translate the ORF. During the elongation phase, the 80S ribosome translocates through the ORF producing a nascent polypeptide (shown in green). Once reaching the stop codon, termination causes the 80S ribosome to release the full-length polypeptide, and recycling splits the 80S into the 40S and 60S subunits. A dashed arrow depicts the mRNA beginning another cycle of translation. In practice many ORFs are translated by more than one ribosome at a time, and a new round of initiation can begin while other ribosomes are in the elongation, termination and recycling phases.

that the closed-loop conformation increases translation by making it easier for terminating ribosomes to be re-recruited to the same mRNA to perform additional rounds of translation (Philipps, 1965; Baglioni et al., 1969; Gingras et al., 1999). However, both wild-type PABP and a mutant PABP that cannot bind eIF4G stimulated translation to a similar degree when tethered to an mRNA (Gray et al., 2000). Additionally, in many cases it is hard to disentangle the influence of PABP on translation from its influence on stability (Kozak, 2004), as increasing either would yield the same result in many assays. Although many assays demonstrate that an mRNA with a poly(A) tail is better translated than one without (Munroe and Jacobson, 1990; Gallie, 1991; Iizuka et al., 1994), the magnitude of stimulation is often modest, and whether it acts synergistically with the 5' cap as predicted by the closed-loop model is less clear (Kozak, 2004).

Cap-dependent translation initiation

Each mRNA is composed of three functionally distinct sequence regions: the 5' UTR, the open reading frame (ORF), and the 3' UTR. The 5' and 3' UTRs contain regulatory information that influences the protein output, stability, and localization of an mRNA, whereas the sequence of the ORF encodes a protein product. The ORF can also contain regulatory information that influences the lifecycle of an mRNA. The fundamental process of translating the sequence in an ORF into a protein product is carried out by the ribosome.

Translation of an ORF by a ribosome can be divided into four phases: initiation, elongation, termination and recycling. During initiation, the 40S and 60S ribosomal subunits are recruited to the mRNA, ultimately resulting in a complete 80S ribosome sitting at the start codon of the ORF. The ribosome poised on the start codon then begins the elongation phase, decoding

the mRNA one codon at a time to produce the polypeptide encoded by the ORF. Once the ribosome reaches the stop codon, termination causes release of the polypeptide, and recycling splits apart the 80S ribosome back into the 40S and 60S subunits (Figure 1). The process of initiation is described in detail here.

In eukaryotic cells, the predominant mechanism of translation initiation is dependent on the binding of eIF4F to the 5' cap of the mRNA (Kozak, 1989). Cap-bound eIF4F performs the critical function of bringing the 43S preinitiation complex (PIC), containing the 40S ribosomal subunit, in proximity with the 5' UTR of the mRNA. In addition to the 40S ribosomal subunit, the 43S PIC is composed of the translation initiation factors eIF1, eIF1A, eIF3, eIF5, and the ternary complex of eIF2, GTP, and methionyl-initiator tRNA (Jackson et al., 2010; Hinnebusch, 2011). A direct interaction between eIF4G and eIF3 brings the 43S PIC to the 5' UTR (LeFebvre et al., 2006), at which point the 43S PIC can attach to the mRNA. The 43S PIC attaches to the mRNA near the start of the 5' UTR and then scans towards the ORF (Jackson et al., 2010; Hinnebusch, 2011). Pairing between the start codon and the anticodon loop of the initiator tRNA causes scanning to halt (Cigan et al., 1988), locking the 43S PIC at the start codon.

The initiation factors in the 43S PIC are critical for scanning through the 5' UTR and stopping at the start codon. eIF1 and eIF1A promote scanning and prevent nucleotide triplets with partial complementarity to the anticodon loop from halting scanning too early (Pestova et al., 1998; Pestova and Kolupaeva, 2002), while eIF5 prevents scanning from proceeding past the start codon (Jackson et al., 2010; Hinnebusch, 2011). Once pairing between the start codon and the anticodon loop is established, eIF2 hydrolyzes its bound GTP, and this reaction is dependent on eIF5 (Jackson et al., 2010; Hinnebusch, 2011). The irreversible hydrolysis of GTP triggers a rearrangement in the 43S PIC to produce a stable 48S PIC positioned at the start codon, and is

followed by dissociation of the initiation factors that were present during scanning (Jackson et al., 2010; Hinnebusch, 2011).

The final step of initiation is the joining of the 60S ribosomal subunit to the 48S PIC to create an elongation-competent ribosome. eIF5B catalyzes the joining of the 60S ribosomal subunit to the 40S ribosomal subunit (Pestova et al., 2000), which results in an 80S ribosome with the start codon positioned in the P site of the ribosome and paired to the anticodon loop of the initiator tRNA (Jackson et al., 2010; Hinnebusch, 2011). Translation initiation concludes with the 80S ribosome poised to begin translation elongation, which begins when a cognate tRNA is brought into the ribosome A site and the first polypeptide bond is formed (Dever and Green, 2012).

Many viruses encode proteins that shut down cap-dependent translation, and utilize cap-independent mechanism to continue synthesizing proteins from their own transcripts. This strategy enables the viral transcripts to outcompete cellular mRNAs for the translational capacity of the cell, and prevents cellular responses to the viral infection. The specific mechanism of cap-independent translation initiation varies among viral transcripts, but the general principle is that a sequence within the viral transcript acts to directly recruit the 40S ribosomal subunit, bypassing the need for recruitment by eIF4F at the 5' cap. Many of these internal ribosomal entry sites (IRESs) act by binding an initiation factor, such as eIF4G, which then recruits the ribosome directly to that site. Other IRESs recruit the ribosomal subunits, or the ribosome itself, to the mRNA, bypassing all of the initiation factors (Hellen and Sarnow, 2001).

Beyond viral transcripts, a handful of cellular mRNAs have also been shown to contain sequences within their 5' UTRs that can promote cap-independent translation (Hellen and Sarnow, 2001; Gilbert, 2010; Jackson, 2013). Additionally, a recent study indicates that a large

number of sequences found in cellular mRNAs can promote cap-independent translation, implying that cap-independent translation may occur on far more cellular mRNAs than previously appreciated (Weingarten-Gabbay et al., 2016). The ability to perform cap-independent translation of certain cellular mRNAs may be important in growth or stress conditions when cap-dependent translation is inhibited, analogous to the use of cap-independent translation in viruses. However, most cellular sequences with reported IRES activity have only been assessed in reporter assays, and often these sequences have a weak ability to promote translation when benchmarked to cap-dependent translation (Gilbert, 2010). Understanding the contribution of cap-independent translation to cellular protein synthesis will require studying these sequences in their native context.

MicroRNA-mediated repression

MicroRNAs (miRNAs) are a class of small, noncoding RNAs that posttranscriptionally regulate gene expression in plants and animals. In animals, miRNAs bind to their mRNA targets as part of the RNA-induced silencing complex (RISC), composed of a miRNA and an Argonaute (Ago) protein, with the set of mRNAs targeted based on partial complementarity between the miRNA and the mRNA. Effective miRNA binding sites are typically found within the 3' UTR of an mRNA, and have contiguous pairing to nucleotides 2–7 of the miRNA. Increased efficacy is often seen when a site also has an adenosine across from nucleotide 1, a pair to nucleotide 8, or both (Bartel, 2009). Over 600 unique miRNAs have been confidently identified in humans, and over half of the human protein-coding genes show evidence of conserved targeting by a miRNA (Bartel, 2009; Friedman et al., 2009). As expected from the huge number of genes directly

regulated by miRNAs, there is a growing list of biological processes impacted by miRNA-mediated repression.

The posttranscriptional regulation mediated by miRNAs occurs through translational repression and mRNA degradation (Huntzinger and Izaurralde, 2011)(Figure 2). Early reports of miRNA-mediated repression emphasized their role as translational repressors, and for a long time they were thought to act solely through this mode of repression (Wightman et al., 1993; Olsen and Ambros, 1999; Seggerson et al., 2002). In general, miRNA-mediated translational

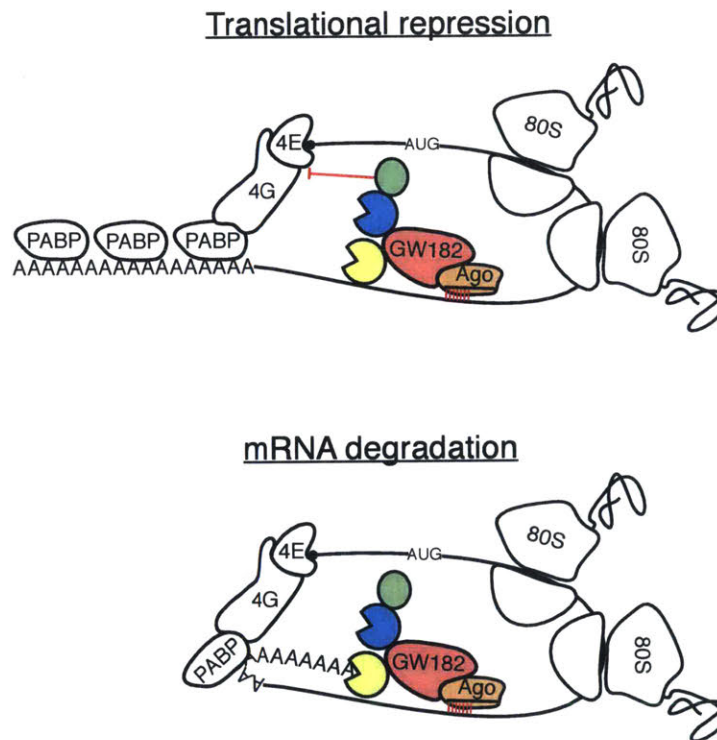


Figure 2. miRNA-mediated translational repression and mRNA degradation
 miRNA-mediated translational repression (top) and mRNA degradation (bottom) are mediated by the same protein complex. The RNA-induced silencing complex, composed of an Argonaute protein (Ago) and a microRNA (shown as a black line in Ago), binds to an mRNA based on partial complementarity between the 3' untranslated region of the mRNA and the 5' end of the microRNA (pairing shown in red). Ago recruits the large scaffold protein GW182, which in turn recruits the Pan2/Pan3 deadenylase complex (shown in yellow) and the CCR4/Not deadenylase complex (shown in blue). DDX6 (shown in green) binds to CCR4/Not, and is thought to mediate translational repression by inhibiting the eIF4E–eIF4G interaction (shown as a red bar). miRNA-mediated mRNA degradation results from the deadenylase complexes accelerating the rate of deadenylation of the bound mRNA.

repression occurs through regulation of translation initiation (Humphreys et al., 2005; Pillai et al., 2005; Guo et al., 2010), although several other mechanisms of translational repression have also been observed (Olsen and Ambros, 1999; Seggerson et al., 2002; Maroney et al., 2006; Nottrott et al., 2006; Petersen et al., 2006). The second mode of repression, mRNA degradation (Bagga et al., 2005; Krutzfeldt et al., 2005; Lim et al., 2005), was discovered in later studies and is caused by the miRNA accelerating the rate of deadenylation of its mRNA targets (Behm-Ansmant et al., 2006; Giraldez et al., 2006; Wu et al., 2006). The more rapid deadenylation causes degradation through the standard mRNA degradation pathway (Rehwinkel et al., 2005; Behm-Ansmant et al., 2006; Chen et al., 2009). Despite not being identified in the initial studies, mRNA degradation accounts for the most of the overall effect of a miRNA at steady state (Baek et al., 2008; Hendrickson et al., 2009; Guo et al., 2010).

Biochemical studies have delineated the molecular effectors of miRNA-mediated mRNA degradation. RISC binds to the large scaffold protein GW182 (Liu et al., 2005; Meister et al., 2005; Rehwinkel et al., 2005; Behm-Ansmant et al., 2006), which in turn binds both the CCR4/Not deadenylase complex and the Pan2/Pan3 deadenylase complex (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011)(Figure 2). The recruitment of these deadenylases to an mRNA target accelerates the rate of deadenylation, causing decapping and degradation to occur sooner than it otherwise would in the lifetime of the mRNA (Chen et al., 2009). The increased rate of degradation and unchanged rate of transcription leads to a reduction in mRNA target abundance.

miRNA-mediated translational repression is less well-characterized at a mechanistic level, and may be caused by more than one mechanism of repression (Jonas and Izaurralde, 2015). The CCR4/Not deadenylase complex can induce both mRNA degradation and

translational repression, which has led to the idea that the two modes of repression involve many of the same proteins (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011). At least a portion of miRNA-mediated translational repression seems to be mediated by DDX6, a decapping activator and translational repressor that binds to the CCR4/Not deadenylase complex (Chen et al., 2014b; Mathys et al., 2014)(Figure 2). Depletion of DDX6 partially derepresses miRNA targets (Chu and Rana, 2006), and the recruitment of DDX6 to a reporter RNA that does not undergo degradation accounts for most of the translational repression of that RNA (Mathys et al., 2014). However, interpreting these experiments from a mechanistic perspective is not entirely straightforward because DDX6 has a known function as a decapping activator (Presnyak and Collier, 2013). It is unclear whether any of the DDX6-mediated translational repression can be accounted for by decapping, and further study of this promising candidate is needed to elucidate the impact of DDX6 on endogenous mRNAs during miRNA-mediated repression. At least one other mechanism of translational repression can also occur, as some of the translational repression mediated by CCR4/Not is independent of DDX6 (Mathys et al., 2014). Developing an understanding of the relationship between the proteins recruited by RISC and the modes of miRNA-mediated repression on endogenous mRNAs remains of major interest to the field.

The striking overlap of proteins involved in miRNA-mediated mRNA degradation and translational repression may be indicative of a causal relationship between the two modes of repression (Jonas and Izaurralde, 2015). Interestingly, a causal relationship between mRNA degradation and translational repression in either direction is plausible. Translation prevents targeting of mRNAs to P-bodies (Parker and Sheth, 2007), sites of mRNA deadenylation and degradation, and in this way translational repression could promote mRNA degradation. There is also a more general precedent for translation causally influencing mRNA stability, such as

pathways where translation regulates the stability of aberrant transcripts (Parker, 2012) or mRNAs with low codon optimality (Presnyak et al., 2015; Bazzini et al., 2016; Mishima and Tomari, 2016). Conversely, mRNA deadenylation and decapping causes loss of PABP and eIF4E, and could thereby cause translational repression. It is also possible that there is no causal relationship between the two modes of repression, as many studies have demonstrated that miRNA-mediated translational repression can occur on non-polyadenylated or non-degradable RNAs (Pillai et al., 2005; Wu et al., 2006; Eulalio et al., 2009), and miRNA-mediated mRNA degradation can occur on non-translatable RNAs (Wu et al., 2006; Eulalio et al., 2009).

Regardless of a causal or non-causal relationship between miRNA-mediated translational repression and mRNA degradation, measurements of the contribution of each mode of repression to the overall effect of the miRNA inform on how miRNAs repress their mRNA targets. At steady state, most miRNA-mediated repression occurs through mRNA degradation (Baek et al., 2008; Hendrickson et al., 2009; Guo et al., 2010), and thus the overall effect of a miRNA is well approximated by simply analyzing changes in mRNA abundance. Whether this approximation is accurate prior to steady state was investigated globally for endogenous mRNAs in zebrafish embryos (Bazzini et al., 2012), and for reporter RNAs in cultured human (Bethune et al., 2012) and fly cells (Djuranovic et al., 2012). At the onset of miRNA-mediated repression, a phase of translational repression was observed prior to mRNA degradation in all three systems (Bazzini et al., 2012; Bethune et al., 2012; Djuranovic et al., 2012). The timing of deadenylation relative to the onset of translational repression was also analyzed for two mRNA targets in zebrafish and for the reporter RNAs in human and fly cells, and there was no evidence of deadenylation during the initial phase of translational repression (Bazzini et al., 2012; Bethune et al., 2012; Djuranovic et al., 2012). These studies demonstrate that the relative contributions of the two modes of

repression change during the approach to steady state (Bazzini et al., 2012; Bethune et al., 2012; Djuranovic et al., 2012), and that biologically relevant repression could be missed by not considering translational repression in pre-steady-state contexts.

The observation of translational repression prior to mRNA degradation does not support assignment of a causal relationship, as translational repression could simply occur more quickly than mRNA degradation. Whether substantial translational repression also occurs during the early phases of repression for endogenous mRNAs outside of early developmental contexts, and in mammalian systems, is critical for understanding how to integrate these results into studies of the biological effects of miRNAs.

Translational regulation

Regulatory mechanisms acting on translation contribute to a variety of cellular responses to stimuli and stresses, help restrict the expression of certain proteins to a specific location, and contribute to maintaining homeostatic levels of proteins during normal cell growth. In general, translational regulation offers a temporal advantage over transcriptional regulation because it can be executed more quickly and is reversed more rapidly. Although the initiation, elongation, and termination phases of translation can all be regulated, most of known regulatory mechanisms act on translation initiation. Compared to regulating elongation or termination, regulating translation initiation has the advantage of saving the energy that would have gone into producing a partial or complete protein product from the mRNA that is being repressed.

A cellular response to certain stresses, such as amino acid starvation, is a global reduction in translation initiation to decrease the metabolic burden placed on the cell (Sonenberg and Hinnebusch, 2009). Although the mechanism of translational regulation varies somewhat

between stress conditions, many of them modulate the activity of initiation factors through phosphorylation. The specific case of amino acid starvation causes the phosphorylation of eIF2, which leads to a reduction in the amount of ternary complex (eIF2–GTP–methionyl-initiator tRNA). This represses the translation of nearly all mRNAs, as ternary complex is essential for translation initiation (Sonenberg and Hinnebusch, 2009). In addition to phosphorylation of eIF2, cellular stress responses can globally regulate translation by inducing phosphorylation of eIF4E or eIF4G, and dephosphorylation of proteins that bind eIF4E (Sonenberg and Hinnebusch, 2009).

RNA-binding proteins (RBPs) or RISC can target subsets of mRNAs for translational repression. These proteins or ribonucleoproteins each use distinct recognition motifs to bind to specific sets of mRNAs and regulate their translation, and recognition motifs falling within the 5' or 3' UTR are typically the most effective. A typical RBP regulates mRNAs transcribed from tens to hundreds of genes, with the precise mechanism of translational repression varying between RBPs. For example, when the intracellular concentration of iron is low, an iron-regulatory protein binds near the start of the 5' UTR of an mRNA involved in iron storage and represses it by physically blocking the 40S ribosomal subunit from loading onto the mRNA (Muckenthaler et al., 1998; Meyron-Holtz et al., 2004). In contrast to this steric block, RBPs that bind within the 3' UTR are generally thought to repress translation primarily by recruiting proteins that bind to either the 5' cap and thus prevent eIF4E binding, or to eIF4E in a manner mutually exclusive with eIF4G (Sonenberg and Hinnebusch, 2009; Szostak and Gebauer, 2013). Regulation imparted RBPs that bind in 3' UTRs is prevalent during early animal development, and is covered in detail below.

The translation of individual mRNAs can be tuned by short ORFs within their 5' UTR. A substantial proportion of mRNAs have one or more short ORFs within their 5' UTRs (between

40 and 50% of human and rodent mRNAs depending on the criteria and transcript annotations) (Yamashita et al., 2003; Iacono et al., 2005; Calvo et al., 2009), and these upstream ORFs (uORFs) are thought to repress the translation of the main ORF (Calvo et al., 2009; Ingolia et al., 2009; Sonenberg and Hinnebusch, 2009; Ingolia et al., 2011; Johnstone et al., 2016). During scanning, a ribosome encounters the start codon of a uORF prior to the start codon of the main ORF, and some proportion of the time the ribosome initiates and translates the uORF, and then terminates and recycles. This siphons away a ribosome that would have otherwise translated the main ORF, decreasing the protein output of that mRNA (Sonenberg and Hinnebusch, 2009). Many uORFs have a suboptimal potential to initiate translation, explaining why the ribosome is often able to scan past the start codon of the uORF and initiate translation on the main ORF (Ingolia et al., 2011). In this way, uORFs tune the translation of mRNAs, and their inclusion can be regulated through the use of alternative transcription start sites (Ingolia et al., 2011; Rojas-Duran and Gilbert, 2012).

Translation and translational regulation can feed back through several known mechanisms to influence mRNA stability. mRNA surveillance pathways, such as non-sense mediated decay, require translation to sense mistakes in the coding sequence of an mRNA and target the mRNA for destruction (Parker, 2012). Aside from aberrant mRNAs, a global relationship has recently been uncovered between the elongation rate of ribosomes on an mRNA and the stability of that mRNA. This relationship is most apparent in yeast, and has also been reported in animal systems (Presnyak et al., 2015; Bazzini et al., 2016; Mishima and Tomari, 2016). More relevant for considering the impact of regulating translation initiation, a link between translation initiation and mRNA stability has been reported in bacteria (Deana and Belasco, 2005) and yeast (Schwartz and Parker, 1999; Sheth and Parker, 2003; Collier and

Parker, 2005; Teixeira et al., 2005). In bacteria, the ribosome physically protects the mRNA from endonucleolytic cleavage and so greater stability is conferred by increased translation initiation (Deana and Belasco, 2005). However, endonucleolytic cleavage is not a major pathway of mRNA decay in eukaryotic systems (Chen and Shyu, 2011), and thus the shielding model from bacterial systems cannot explain observations in yeast where perturbing translation initiation led to mRNA destabilization (Schwartz and Parker, 1999; Sheth and Parker, 2003; Coller and Parker, 2005; Teixeira et al., 2005). In many cases, perturbing translation initiation also increases the accessibility of the 5' cap or poly(A) tail to cytoplasmic enzymes (Schwartz and Parker, 1999; Coller and Parker, 2005), making it hard to determine whether the change in mRNA stability was the result of decreased translation initiation, or decreased shielding of the 5' cap and poly(A) tail from the decapping and deadenylase complexes. For this reason it remains ambiguous whether repressing translation initiation causes mRNA destabilization in eukaryotic systems.

Translational regulation during animal development

Early animal development is a particularly fascinating context for the study of translational regulation. In contrast to most other cell types, oocytes and early embryos from nearly all animal species are transcriptionally silent. During early oogenesis, maternal cells produce all the mRNAs needed for early development and deposit them in the oocyte. This process of maternal deposition stops by late oogenesis, after which point no new mRNAs are deposited in the oocyte, and the deposited mRNAs remain completely stable into the earliest stages of embryogenesis. The mRNA metabolism of the embryo radically changes during the maternal-to-zygotic transition (MZT), when a wave of degradation of maternal mRNAs

coincides with the onset of transcription from the zygotic genome (Figure 3). After the MZT, the mRNA metabolism of the embryo more closely resembles that of other non-embryonic contexts (Tadros and Lipshitz, 2009). The absence of transcription and mRNA degradation in oocytes and early embryos results in translational and post-translational regulation being entirely responsible for regulating gene expression in these contexts.

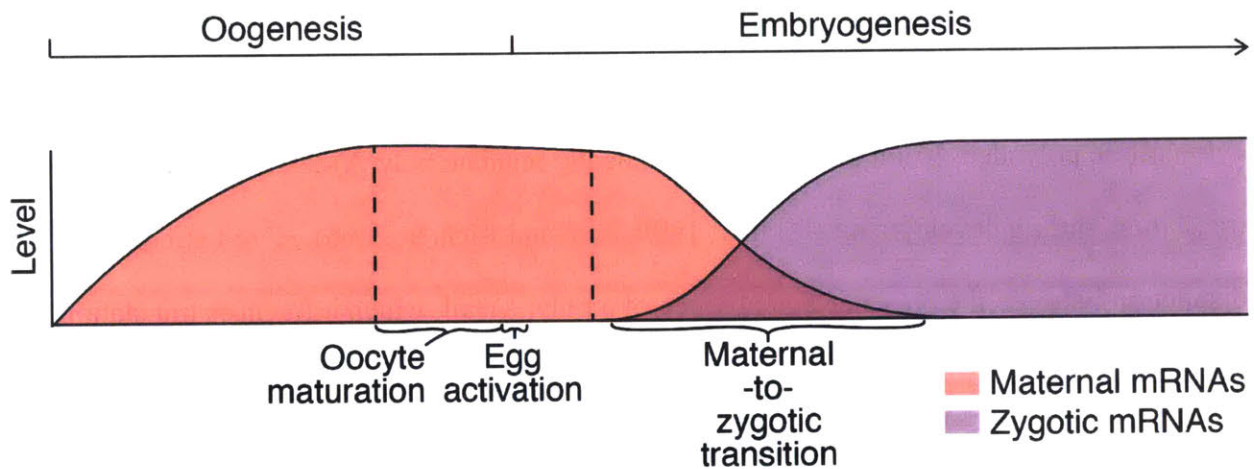


Figure 3. Schematic of mRNA metabolism during oogenesis and embryogenesis

Dynamics of maternal and zygotic mRNA abundance during oogenesis and embryogenesis, with the relative timing of major developmental events indicated (reproduced from Eichhorn et al., 2016). Dashed lines enclose the period of development that occurs without any transcription or mRNA destabilization.

Late oogenesis and early embryogenesis are incredibly dynamic periods of development. During late oogenesis, the oocyte proceeds through the meiotic cell-cycle and becomes competent for fertilization (Von Stetina and Orr-Weaver, 2011). Once the oocyte is mature, egg activation or fertilization induces a rewiring of cellular identity, converting a differentiated oocyte into a totipotent embryo (Horner and Wolfner, 2008). The fertilized embryo then enters the mitotic cell-cycle, and shortly thereafter initiates the changes necessary to activate the zygotic genome (Tadros and Lipshitz, 2009)(Figure 3). Translational regulation is central to producing the changes in gene expression that drive these major developmental transitions.

Additionally, the expression of many mRNAs involved in establishing embryonic polarity is restricted to the appropriate domain through translational regulation (Lasko, 2012).

Changes in poly(A)-tail length play a major role in regulating gene expression during early development (Richter, 1999). In most cell types, after a poly(A) tail is added to an mRNA it can only get shorter over time (Chen and Shyu, 2011). Early developmental systems have the unique ability to re-extend poly(A) tails in the cytoplasm through a mechanism known as cytoplasmic polyadenylation (Richter, 1999). The opposing processes of deadenylation and cytoplasmic polyadenylation are used to dynamically regulate poly(A)-tail length, and in turn translation, during development (Richter, 1999; Kim and Richter, 2006). Here I am drawing a distinction between the presence and absence of a poly(A) tail, which influences translation in all contexts (Hinnebusch, 2011), and the length of a poly(A) tail, which has only been demonstrated to influence translation in select developmental and neuronal contexts (Richter, 1999).

In *Xenopus* oocytes, many mRNAs have short poly(A) tails and are poorly translated (Sagata et al., 1980; Hyman and Wormington, 1988). This includes *c-mos* mRNAs (Sheets et al., 1994), which encode a kinase essential for oocyte maturation (Sagata et al., 1988). As the oocytes proceed through maturation, many of these short poly(A) tails, including those of *c-mos* mRNAs, are re-extended by cytoplasmic polyadenylation and these mRNAs are translationally upregulated (Dworkin and Dworkin-Rastl, 1985; Paris and Philippe, 1990; Paris and Richter, 1990; Sheets et al., 1994; Stebbins-Boaz and Richter, 1994; Sheets et al., 1995). Blocking the tail-length extension of a reporter RNA blocked its translational upregulation (McGrew et al., 1989), indicating that the concomitant tail lengthening and translational upregulation of many cellular mRNAs likely reflects cytoplasmic polyadenylation causing translational upregulation.

Subsequent studies in *Xenopus* and *Drosophila* have provided additional support for this hypothesis (Salles et al., 1994; Sheets et al., 1995; Barkoff et al., 1998).

The translational upregulation caused by cytoplasmic polyadenylation could have been due to either the act of cytoplasmic polyadenylation influencing translation, or the consequence of cytoplasmic polyadenylation (a longer poly(A)-tail) influencing translation. A phenotypic assay for *Xenopus* oocyte maturation was used to distinguish between these possibilities. The translational upregulation of *c-mos* is essential for oocyte maturation, and cleaving off the poly(A) tail of endogenous *c-mos* blocked its translational upregulation and oocyte maturation (Sheets et al., 1995; Barkoff et al., 1998). Appending the cleaved *c-mos* mRNA with a sequence that undergoes cytoplasmic polyadenylation (Sheets et al., 1995) or a 130 nt poly(A) tail (Barkoff et al., 1998) substantially rescued oocyte maturation, but appending a 30 nt poly(A) tail failed to rescue (Barkoff et al., 1998). Thus, *c-mos* translation is influenced by the length of the poly(A) tail attached to it and not just the presence or absence of a poly(A) tail. A similar result was obtained in fly embryos (Salles et al., 1994), and together these studies indicated that during early development, mRNAs with longer poly(A)-tails are better translated than their short-tailed counterpart. However, it is important to note that increasing the tail length of an mRNA is not always sufficient to upregulate its translation (Simon et al., 1992; Barkoff et al., 1998). For example, if *Xenopus* oocytes were not stimulated with progesterone, appending either a short or long poly(A)-tail to *c-mos* mRNAs resulted in the same amount of c-Mos protein (Barkoff et al., 1998). This may indicate that the act of cytoplasmic polyadenylation contributes to translational upregulation in certain contexts, although there are several other possible explanations.

As implied by the ability of increases in poly(A)-tail length to upregulate translation, shortening of poly(A) tails can repress translation (Richter, 1999). Smaug, a well-characterized

RBP in *Drosophila*, binds to a stem-loop motif within the 3' UTR of mRNAs and mediates their deadenylation through a direct interaction with the CCR4/Not deadenylase complex (Smibert et al., 1996; Semotok et al., 2005). *nanos* mRNA is targeted by Smaug, and the resulting deadenylation causes translational repression of *nanos* mRNA (Jeske et al., 2006; Zaessinger et al., 2006). In addition to *nanos*, Smaug is known to translationally repress hundreds of other mRNA targets (Chen et al., 2014a), but it is unknown if Smaug represses these targets through deadenylation or a different mechanism. As development proceeds and the degradation pathway becomes active, Smaug-mediated deadenylation serves as a major mechanism for degrading maternal mRNAs (Tadros et al., 2007; Chen et al., 2014a). Several other RBPs are known to translationally repress their mRNA targets through a similar mechanism of recruiting a deadenylase (Kadyrova et al., 2007; Temme et al., 2010; Newton et al., 2015).

Smaug is capable of mediating a second, deadenylation independent, mechanism of translational repression (Jeske et al., 2006; Jeske et al., 2011). In a cell-free lysate system Smaug can translationally repress a reporter RNA that cannot be deadenylated or that has no poly(A) tail, both of which must be through a deadenylation-independent mechanism (Jeske et al., 2006; Jeske et al., 2011). The deadenylation-independent translational repression mediated by Smaug involves CUP (Nelson et al., 2004) and potentially other proteins (Jeske et al., 2011; Pinder and Smibert, 2013), but the precise mechanism remains unclear. CUP is thought to repress translation initiation by binding to eIF4E and preventing eIF4G binding (Nelson et al., 2004; Jeske et al., 2011), although the mutually exclusive binding has only been shown with peptide fragments of CUP and eIF4G (Nelson et al., 2004). This model of how Smaug mediates deadenylation-independent translational repression does not explain how Smaug is able to repress cap-independent translation mediated by the cricket paralysis virus IRES (Jeske et al., 2011), an

IRES that bypasses the requirement for all initiation factors (Hellen and Sarnow, 2001). Lack of clarity on the mechanism of repression notwithstanding, Smaug can repress translation through both a deadenylation-dependent and deadenylation-independent mechanism (Jeske et al., 2006; Zaessinger et al., 2006; Jeske et al., 2011; Chen et al., 2014a). The detailed analyses of Smaug have not yet determined how much of the translational repression it causes is attributable to the different mechanisms of repression, and for many RBPs that regulate translation it is unknown whether or not deadenylation explains none, some, or all of their regulatory effects.

The activation of the mRNA degradation pathway during development restores the typical coupling between deadenylation and decapping, and results in the clearance of short-tailed mRNAs (Tadros and Lipshitz, 2009). mRNAs that are translationally repressed through deadenylation-dependent mechanisms are targeted for degradation at this time, whereas those repressed through deadenylation-independent mechanisms could persist and become re-expressed. This difference may help explain why one mechanism of regulation might be preferred over another for different mRNAs. For an RBP that causes both deadenylation-dependent and -independent translational repression, there is no understanding of whether a subset of targets are regulated by deadenylation and a subset by deadenylation-independent translational repression, or if they are all regulated by both mechanisms. Differences in recognition motifs, the presence of additional regulators on certain mRNA targets, or differences in mRNA localization could all give rise to unique regulatory outcomes for the targets of an individual RBP.

Is there a causal relationship between poly(A)-tail length and translational efficiency outside of early development and neurons? Cytoplasmic polyadenylation has been reported in cells entering mitosis (Novoa et al., 2010) and senescing cells (Burns and Richter, 2008).

Additionally, a global correlation between poly(A)-tail length and translational efficiency has been reported in fission and budding yeasts (Beilharz and Preiss, 2007; Lackner et al., 2007). These observations indicate that a causal relationship between tail length and translational efficiency may exist more generally in eukaryotic systems. Coupling tail length and translational efficiency in transcriptionally active cells would have the effect of increasing the expression of nascent transcripts, as in a typical cell a poly(A) tail is longest when it emerges from the nucleus. This would limit the utility of regulating expression through mRNA stability, suggesting that perhaps only in cases where transcription and mRNA degradation are absent (or distant, as in the case of neuronal synapses) does this mechanism of translational control exist. Further study of endogenous mRNAs in mammalian cells and classic developmental contexts will elucidate the overall impact of poly(A)-tail length on translational efficiency.

This dissertation describes the regulatory consequences of changes in poly(A)-tail length. Methods for making global measurements of poly(A)-tail length and translational efficiency are applied to a variety of cellular contexts, including early development in vertebrates and an invertebrate, mammalian cell lines, a mammalian tissue, and budding and fission yeasts. These analyses describe the relationship between poly(A)-tail length and translational efficiency globally for endogenous mRNAs. In the context of early development, the contribution of tail-length changes to translational regulation during major developmental transitions is analyzed, as is the mechanism of repression used by Smaug and other translational regulators. Additionally, the kinetics and relative contribution of microRNA-mediated translational repression and mRNA degradation are defined for endogenous mRNAs in mammalian cells.

References

- Amrani, N., Ghosh, S., Mangus, D.A., and Jacobson, A. (2008). Translation factors promote the formation of two states of the closed-loop mRNP. *Nature* *453*, 1276-1280.
- Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. *Nature* *455*, 64-71.
- 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*, 553-563.
- Baglioni, C., Vesco, C., and Jacobs-Lorena, M. (1969). The role of ribosomal subunits in mammalian cells. *Cold Spring Harb Symp Quant Biol* *34*, 555-565.
- Barkoff, A., Ballantyne, S., and Wickens, M. (1998). Meiotic maturation in *Xenopus* requires polyadenylation of multiple mRNAs. *Embo J* *17*, 3168-3175.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* *136*, 215-233.
- Bazzini, A.A., Del Viso, F., Moreno-Mateos, M.A., Johnstone, T.G., Vejnar, C.E., Qin, Y., Yao, J., Khokha, M.K., and Giraldez, A.J. (2016). Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *Embo J* *35*, 2087-2103.
- Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* *336*, 233-237.
- Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* *20*, 1885-1898.
- Beilharz, T.H., and Preiss, T. (2007). Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA* *13*, 982-997.
- Bethune, J., Artus-Revel, C.G., and Filipowicz, W. (2012). Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* *13*, 716-723.
- Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol Cell* *44*, 120-133.
- Burns, D.M., and Richter, J.D. (2008). CPEB regulation of human cellular senescence, energy metabolism, and p53 mRNA translation. *Genes Dev* *22*, 3449-3460.
- Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A* *106*, 7507-7512.

- Caponigro, G., and Parker, R. (1995). Multiple functions for the poly(A)-binding protein in mRNA decapping and deadenylation in yeast. *Genes Dev* 9, 2421-2432.
- Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat Struct Mol Biol* 18, 1218-1226.
- Chen, C.Y., and Shyu, A.B. (2011). Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* 2, 167-183.
- Chen, C.Y., Zheng, D., Xia, Z., and Shyu, A.B. (2009). Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat Struct Mol Biol* 16, 1160-1166.
- Chen, L., Dumelie, J.G., Li, X., Cheng, M.H., Yang, Z., Laver, J.D., Siddiqui, N.U., Westwood, J.T., Morris, Q., Lipshitz, H.D., *et al.* (2014a). Global regulation of mRNA translation and stability in the early *Drosophila* embryo by the Smaug RNA-binding protein. *Genome Biol* 15, R4.
- Chen, Y., Boland, A., Kuzuoglu-Ozturk, D., Bawankar, P., Loh, B., Chang, C.T., Weichenrieder, O., and Izaurralde, E. (2014b). A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. *Mol Cell* 54, 737-750.
- Chowdhury, A., Mukhopadhyay, J., and Tharun, S. (2007). The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA* 13, 998-1016.
- Chu, C.Y., and Rana, T.M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *PLoS Biol* 4, e210.
- Cigan, A.M., Feng, L., and Donahue, T.F. (1988). tRNAⁱ(met) functions in directing the scanning ribosome to the start site of translation. *Science* 242, 93-97.
- Coller, J., and Parker, R. (2004). Eukaryotic mRNA decapping. *Annu Rev Biochem* 73, 861-890.
- Coller, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. *Cell* 122, 875-886.
- Deana, A., and Belasco, J.G. (2005). Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev* 19, 2526-2533.
- Dever, T.E., and Green, R. (2012). The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb Perspect Biol* 4, a013706.
- Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* 336, 237-240.
- Dworkin, M.B., and Dworkin-Rastl, E. (1985). Changes in RNA titers and polyadenylation during oogenesis and oocyte maturation in *Xenopus laevis*. *Dev Biol* 112, 451-457.

- Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2009). Deadenylation is a widespread effect of miRNA regulation. *RNA* *15*, 21-32.
- Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F., *et al.* (2011). miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nat Struct Mol Biol* *18*, 1211-1217.
- Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* *19*, 92-105.
- Gallie, D.R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev* *5*, 2108-2116.
- Gilbert, W.V. (2010). Alternative ways to think about cellular internal ribosome entry. *J Biol Chem* *285*, 29033-29038.
- Gingras, A.C., Raught, B., and Sonenberg, N. (1999). eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev Biochem* *68*, 913-963.
- Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* *312*, 75-79.
- Goldstrohm, A.C., and Wickens, M. (2008). Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol* *9*, 337-344.
- Gorlach, M., Burd, C.G., and Dreyfuss, G. (1994). The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. *Exp Cell Res* *211*, 400-407.
- Gray, N.K., Collier, J.M., Dickson, K.S., and Wickens, M. (2000). Multiple portions of poly(A)-binding protein stimulate translation in vivo. *Embo J* *19*, 4723-4733.
- Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* *466*, 835-840.
- Hellen, C.U., and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev* *15*, 1593-1612.
- Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol* *7*, e1000238.
- Hinnebusch, A.G. (2011). Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiol Mol Biol Rev* *75*, 434-467, first page of table of contents.
- Horner, V.L., and Wolfner, M.F. (2008). Transitioning from egg to embryo: triggers and mechanisms of egg activation. *Dev Dyn* *237*, 527-544.

- Huarte, J., Belin, D., Vassalli, A., Strickland, S., and Vassalli, J.D. (1987). Meiotic maturation of mouse oocytes triggers the translation and polyadenylation of dormant tissue-type plasminogen activator mRNA. *Genes Dev* 1, 1201-1211.
- Humphreys, D.T., Westman, B.J., Martin, D.I., and Preiss, T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci U S A* 102, 16961-16966.
- Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* 12, 99-110.
- Hyman, L.E., and Wormington, W.M. (1988). Translational inactivation of ribosomal protein mRNAs during *Xenopus* oocyte maturation. *Genes Dev* 2, 598-605.
- Iacono, M., Mignone, F., and Pesole, G. (2005). uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene* 349, 97-105.
- Iizuka, N., Najita, L., Franzusoff, A., and Sarnow, P. (1994). Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from *Saccharomyces cerevisiae*. *Mol Cell Biol* 14, 7322-7330.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218-223.
- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789-802.
- Jackson, R.J. (2013). The current status of vertebrate cellular mRNA IRESs. *Cold Spring Harb Perspect Biol* 5.
- Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 11, 113-127.
- Jacobson, M.F., and Baltimore, D. (1968). Polypeptide cleavages in the formation of poliovirus proteins. *Proc Natl Acad Sci U S A* 61, 77-84.
- Jeske, M., Meyer, S., Temme, C., Freudenreich, D., and Wahle, E. (2006). Rapid ATP-dependent deadenylation of nanos mRNA in a cell-free system from *Drosophila* embryos. *J Biol Chem* 281, 25124-25133.
- Jeske, M., Moritz, B., Anders, A., and Wahle, E. (2011). Smaug assembles an ATP-dependent stable complex repressing nanos mRNA translation at multiple levels. *Embo J* 30, 90-103.
- Johnstone, T.G., Bazzini, A.A., and Giraldez, A.J. (2016). Upstream ORFs are prevalent translational repressors in vertebrates. *Embo J* 35, 706-723.

- Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* 16, 421-433.
- Kadyrova, L.Y., Habara, Y., Lee, T.H., and Wharton, R.P. (2007). Translational control of maternal Cyclin B mRNA by Nanos in the *Drosophila* germline. *Development* 134, 1519-1527.
- 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 Dev* 19, 104-113.
- Kim, J.H., and Richter, J.D. (2006). Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. *Mol Cell* 24, 173-183.
- Kozak, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* 15, 1109-1123.
- Kozak, M. (1989). The scanning model for translation: an update. *J Cell Biol* 108, 229-241.
- Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187-208.
- Kozak, M. (2004). How strong is the case for regulation of the initiation step of translation by elements at the 3' end of eukaryotic mRNAs? *Gene* 343, 41-54.
- Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., and Stoffel, M. (2005). Silencing of microRNAs in vivo with 'antagomirs'. *Nature* 438, 685-689.
- Lackner, D.H., Beilharz, T.H., Marguerat, S., Mata, J., Watt, S., Schubert, F., Preiss, T., and Bahler, J. (2007). A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol Cell* 26, 145-155.
- Lasko, P. (2012). mRNA localization and translational control in *Drosophila* oogenesis. *Cold Spring Harb Perspect Biol* 4.
- LeFebvre, A.K., Korneeva, N.L., Trutschl, M., Cvek, U., Duzan, R.D., Bradley, C.A., Hershey, J.W., and Rhoads, R.E. (2006). Translation initiation factor eIF4G-1 binds to eIF3 through the eIF3e subunit. *J Biol Chem* 281, 22917-22932.
- 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, 769-773.
- Liu, J., Rivas, F.V., Wohlschlegel, J., Yates, J.R., 3rd, Parker, R., and Hannon, G.J. (2005). A role for the P-body component GW182 in microRNA function. *Nat Cell Biol* 7, 1261-1266.
- 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 Biol* 4, 223.

- Maroney, P.A., Yu, Y., Fisher, J., and Nilsen, T.W. (2006). Evidence that microRNAs are associated with translating messenger RNAs in human cells. *Nat Struct Mol Biol* *13*, 1102-1107.
- Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Mol Cell* *54*, 751-765.
- McGrew, L.L., Dworkin-Rastl, E., Dworkin, M.B., and Richter, J.D. (1989). Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev* *3*, 803-815.
- Meister, G., Landthaler, M., Peters, L., Chen, P.Y., Urlaub, H., Luhrmann, R., and Tuschl, T. (2005). Identification of novel argonaute-associated proteins. *Curr Biol* *15*, 2149-2155.
- Meyron-Holtz, E.G., Ghosh, M.C., Iwai, K., LaVaute, T., Brazzolotto, X., Berger, U.V., Land, W., Ollivierre-Wilson, H., Grinberg, A., Love, P., *et al.* (2004). Genetic ablations of iron regulatory proteins 1 and 2 reveal why iron regulatory protein 2 dominates iron homeostasis. *Embo J* *23*, 386-395.
- Mishima, Y., and Tomari, Y. (2016). Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Mol Cell* *61*, 874-885.
- Muckenthaler, M., Gray, N.K., and Hentze, M.W. (1998). IRP-1 binding to ferritin mRNA prevents the recruitment of the small ribosomal subunit by the cap-binding complex eIF4F. *Mol Cell* *2*, 383-388.
- Munroe, D., and Jacobson, A. (1990). mRNA poly(A) tail, a 3' enhancer of translational initiation. *Mol Cell Biol* *10*, 3441-3455.
- Nelson, M.R., Leidal, A.M., and Smibert, C.A. (2004). *Drosophila* Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. *Embo J* *23*, 150-159.
- Newton, F.G., Harris, R.E., Sutcliffe, C., and Ashe, H.L. (2015). Coordinate post-transcriptional repression of Dpp-dependent transcription factors attenuates signal range during development. *Development* *142*, 3362-3373.
- Nottrott, S., Simard, M.J., and Richter, J.D. (2006). Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol* *13*, 1108-1114.
- Novoa, I., Gallego, J., Ferreira, P.G., and Mendez, R. (2010). Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. *Nat Cell Biol* *12*, 447-456.
- Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* *216*, 671-680.

- Paris, J., and Philippe, M. (1990). Poly(A) metabolism and polysomal recruitment of maternal mRNAs during early *Xenopus* development. *Dev Biol* *140*, 221-224.
- Paris, J., and Richter, J.D. (1990). Maturation-specific polyadenylation and translational control: diversity of cytoplasmic polyadenylation elements, influence of poly(A) tail size, and formation of stable polyadenylation complexes. *Mol Cell Biol* *10*, 5634-5645.
- Parker, R. (2012). RNA degradation in *Saccharomyces cerevisiae*. *Genetics* *191*, 671-702.
- Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. *Mol Cell* *25*, 635-646.
- Pestova, T.V., Borukhov, S.I., and Hellen, C.U. (1998). Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* *394*, 854-859.
- Pestova, T.V., and Kolupaeva, V.G. (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev* *16*, 2906-2922.
- Pestova, T.V., Lomakin, I.B., Lee, J.H., Choi, S.K., Dever, T.E., and Hellen, C.U. (2000). The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* *403*, 332-335.
- Petersen, C.P., Bordeleau, M.E., Pelletier, J., and Sharp, P.A. (2006). Short RNAs repress translation after initiation in mammalian cells. *Mol Cell* *21*, 533-542.
- Philipps, G.R. (1965). Haemoglobin Synthesis and Polysomes in Intact Reticulocytes. *Nature* *205*, 567-570.
- Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* *309*, 1573-1576.
- Pinder, B.D., and Smibert, C.A. (2013). microRNA-independent recruitment of Argonaute 1 to nanos mRNA through the Smaug RNA-binding protein. *EMBO Rep* *14*, 80-86.
- Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., *et al.* (2015). Codon optimality is a major determinant of mRNA stability. *Cell* *160*, 1111-1124.
- Presnyak, V., and Collier, J. (2013). The DHH1/RCKp54 family of helicases: an ancient family of proteins that promote translational silencing. *Biochim Biophys Acta* *1829*, 817-823.
- Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. *Cell* *108*, 501-512.
- Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* *11*, 1640-1647.

- Richter, J.D. (1999). Cytoplasmic polyadenylation in development and beyond. *Microbiol Mol Biol Rev* 63, 446-456.
- Rojas-Duran, M.F., and Gilbert, W.V. (2012). Alternative transcription start site selection leads to large differences in translation activity in yeast. *RNA* 18, 2299-2305.
- Rosenthal, E.T., and Ruderman, J.V. (1987). Widespread changes in the translation and adenylation of maternal messenger RNAs following fertilization of *Spisula* oocytes. *Dev Biol* 121, 237-246.
- Rosenthal, E.T., Tansey, T.R., and Ruderman, J.V. (1983). Sequence-specific adenylations and deadenylations accompany changes in the translation of maternal messenger RNA after fertilization of *Spisula* oocytes. *J Mol Biol* 166, 309-327.
- Sagata, N., Oskarsson, M., Copeland, T., Brumbaugh, J., and Vande Woude, G.F. (1988). Function of c-mos proto-oncogene product in meiotic maturation in *Xenopus* oocytes. *Nature* 335, 519-525.
- Sagata, N., Shiokawa, K., and Yamana, K. (1980). A study on the steady-state population of poly(A)+RNA during early development of *Xenopus laevis*. *Dev Biol* 77, 431-448.
- Salles, F.J., Lieberfarb, M.E., Wreden, C., Gergen, J.P., and Strickland, S. (1994). Coordinate initiation of *Drosophila* development by regulated polyadenylation of maternal messenger RNAs. *Science* 266, 1996-1999.
- Schwartz, D.C., and Parker, R. (1999). Mutations in translation initiation factors lead to increased rates of deadenylation and decapping of mRNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol* 19, 5247-5256.
- Schwartz, D.C., and Parker, R. (2000). mRNA decapping in yeast requires dissociation of the cap binding protein, eukaryotic translation initiation factor 4E. *Mol Cell Biol* 20, 7933-7942.
- Seggerson, K., Tang, L., and Moss, E.G. (2002). Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation. *Dev Biol* 243, 215-225.
- Semotok, J.L., Cooperstock, R.L., Pinder, B.D., Vari, H.K., Lipshitz, H.D., and Smibert, C.A. (2005). Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early *Drosophila* embryo. *Curr Biol* 15, 284-294.
- Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. *Cell* 9, 645-653.
- Sheets, M.D., Fox, C.A., Hunt, T., Vande Woude, G., and Wickens, M. (1994). The 3'-untranslated regions of c-mos and cyclin mRNAs stimulate translation by regulating cytoplasmic polyadenylation. *Genes Dev* 8, 926-938.
- Sheets, M.D., Wu, M., and Wickens, M. (1995). Polyadenylation of c-mos mRNA as a control point in *Xenopus* meiotic maturation. *Nature* 374, 511-516.

- Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* *300*, 805-808.
- Simon, R., Tassan, J.P., and Richter, J.D. (1992). Translational control by poly(A) elongation during *Xenopus* development: differential repression and enhancement by a novel cytoplasmic polyadenylation element. *Genes Dev* *6*, 2580-2591.
- Smibert, C.A., Wilson, J.E., Kerr, K., and Macdonald, P.M. (1996). smaug protein represses translation of unlocalized nanos mRNA in the *Drosophila* embryo. *Genes Dev* *10*, 2600-2609.
- Sonenberg, N. (1994). mRNA translation: influence of the 5' and 3' untranslated regions. *Curr Opin Genet Dev* *4*, 310-315.
- Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* *136*, 731-745.
- Sonenberg, N., Morgan, M.A., Merrick, W.C., and Shatkin, A.J. (1978). A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proc Natl Acad Sci U S A* *75*, 4843-4847.
- Sonenberg, N., Rupprecht, K.M., Hecht, S.M., and Shatkin, A.J. (1979). Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *Proc Natl Acad Sci U S A* *76*, 4345-4349.
- Stebbins-Boaz, B., and Richter, J.D. (1994). Multiple sequence elements and a maternal mRNA product control cdk2 RNA polyadenylation and translation during early *Xenopus* development. *Mol Cell Biol* *14*, 5870-5880.
- Stevens, A. (1978). An exoribonuclease from *Saccharomyces cerevisiae*: effect of modifications of 5' end groups on the hydrolysis of substrates to 5' mononucleotides. *Biochem Biophys Res Commun* *81*, 656-661.
- Szostak, E., and Gebauer, F. (2013). Translational control by 3'-UTR-binding proteins. *Brief Funct Genomics* *12*, 58-65.
- Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. *Dev Cell* *12*, 143-155.
- Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development* *136*, 3033-3042.
- Teixeira, D., Sheth, U., Valencia-Sanchez, M.A., Brengues, M., and Parker, R. (2005). Processing bodies require RNA for assembly and contain nontranslating mRNAs. *RNA* *11*, 371-382.

Temme, C., Zhang, L., Kremmer, E., Ihling, C., Chartier, A., Sinz, A., Simonelig, M., and Wahle, E. (2010). Subunits of the *Drosophila* CCR4-NOT complex and their roles in mRNA deadenylation. *RNA* 16, 1356-1370.

Von Stetina, J.R., and Orr-Weaver, T.L. (2011). Developmental control of oocyte maturation and egg activation in metazoan models. *Cold Spring Harb Perspect Biol* 3, a005553.

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

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell* 2, 135-140.

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, 855-862.

Wu, L., Fan, J., and Belasco, J.G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci U S A* 103, 4034-4039.

Yamashita, R., Suzuki, Y., Nakai, K., and Sugano, S. (2003). Small open reading frames in 5' untranslated regions of mRNAs. *C R Biol* 326, 987-991.

Zaessinger, S., Busseau, I., and Simonelig, M. (2006). Oskar allows nanos mRNA translation in *Drosophila* embryos by preventing its deadenylation by Smaug/CCR4. *Development* 133, 4573-4583.

Chapter 2

Poly(A)-tail profiling reveals an embryonic switch in translational control

Alexander O. Subtelny^{1,2,3,4*}, Stephen W. Eichhorn^{1,2,3*}, Grace R. Chen^{1,2,3}, Hazel Sive^{2,3},
David P. Bartel^{1,2,3}

¹Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

²Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

³Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

⁴Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139, USA

*These authors contributed equally to this work

A.O.S. developed PAL-seq, generated tail-length measurements, and performed associated analyses. S.W.E. performed ribosome profiling, RNA-seq and associated analyses. G.R.C. performed zebrafish injections and assisted with staging. D.P.B. supervised with help from H.S. All authors helped design the study and write the manuscript.

Published as:

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, 66-71.

Abstract

Poly(A) tails enhance the stability and translation of most eukaryotic mRNAs, but difficulties in globally measuring poly(A)-tail lengths have impeded greater understanding of poly(A)-tail function. Here, we describe poly(A)-tail length profiling by sequencing (PAL-seq) and apply it to measure tail lengths of millions of individual RNAs isolated from yeasts, cell lines, *Arabidopsis* leaves, mouse liver, and zebrafish and frog embryos. Poly(A)-tail lengths were conserved between orthologous mRNAs, with those encoding ribosomal proteins and other classes of “housekeeping” proteins tending to be shorter. As expected, tail lengths were coupled to translational efficiency in early zebrafish and frog embryos. However, this strong coupling diminished at gastrulation and was absent in our non-embryonic samples, which indicated a rapid developmental switch in the nature of translational control. This switch complements an earlier switch to zygotic transcriptional control and explains why the predominant effect of microRNA-mediated deadenylation concurrently shifts from translational repression to mRNA destabilization.

Introduction

Most eukaryotic mRNAs end with poly(A) tails, which are added by a nuclear poly(A) polymerase, following cleavage of the primary transcript during transcriptional termination (Moore and Proudfoot, 2009). These tails are then shortened by deadenylases (Goldstrohm and Wickens, 2008; Chen and Shyu, 2011), and in some contexts, such as in animal oocytes and early embryos or at neuronal synapses, they can be re-extended by cytoplasmic poly(A) polymerases (Richter, 1999). In the cytoplasm, the mRNA poly(A) tail promotes translation and inhibits decay (Goldstrohm and Wickens, 2008; Weill et al., 2012).

Although poly(A) tails must exceed a minimal length to promote translation, an influence of tail length beyond this minimum is largely unknown. The prevailing view is that longer tails generally lead to increased translation (Eckmann et al., 2011; Weill et al., 2012). This idea partly stems from the known importance of cytoplasmic polyadenylation in activating certain genes in specific contexts (i.e., oocytes, early embryos, and synapses)(Richter, 1999; Weill et al., 2012) and the increased translation observed in *Xenopus* oocytes and *Drosophila* embryos when appending synthetic tails of increasing length onto an mRNA (Salles et al., 1994; Barkoff et al., 1998). Support for a more general coupling of tail length and translation comes from studies of yeast extracts (Preiss et al., 1998) and cells(Beilharz and Preiss, 2007; Lackner et al., 2007). However, the general relationship between tail length and translational efficiency has not been reported outside of yeast, primarily because transcriptome-wide measurements have not been feasible for the longer-tailed mRNAs found in other organisms.

Results and Discussion

Poly(A)-tail length profiling by sequencing (PAL-seq)

We developed a high-throughput sequencing method that accurately measures individual poly(A) tails of any physiological length (Fig. 1a). After generating sequencing clusters and before sequencing, a primer hybridized immediately 3' of the poly(A) sequence is extended using a mixture of dTTP and biotin-conjugated dUTP as the only nucleoside triphosphates and conditions that had been optimized to yield full-length extension products without terminal mismatches (Extended Data Fig. 1a). This key step of the protocol quantitatively marks each cluster with biotin in proportion to the length of the poly(A) tail (Fig. 1a, step 11). After sequencing the 36 nucleotides immediately 5' of the poly(A) site, the flow cell is incubated with

fluorophore-tagged streptavidin, which binds the biotin incorporated during primer-extension to impart fluorescence intensity proportional to the poly(A)-tract length. To account for the density of each cluster, this raw intensity is normalized to that of the fluorescent bases added during sequencing by synthesis (Nutiu et al., 2011), thereby yielding a normalized fluorescence-intensity for the poly(A) tail of each transcript, paired with a sequencing read that identifies its poly(A) site and thus the gene from which it originated

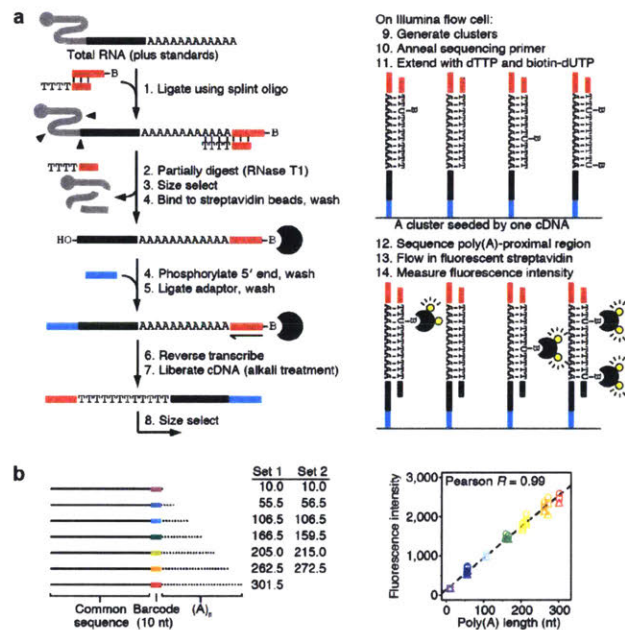


Figure 1. Global measurement of poly(A)-tail lengths.

a, Outline of PAL-seq. With the aid of a splint oligonucleotide, 3' termini ending with at least eight A residues are preferentially ligated to a biotinylated adaptor oligonucleotide. After partial digestion with ribonuclease T1 (which cuts after G residues), 3' fragments of ligation products are captured on streptavidin beads. Captured RNAs are phosphorylated, ligated to an adaptor oligonucleotide at their 5' termini, and then reverse transcribed. Complementary DNAs (cDNAs) are liberated from the beads and loaded directly onto a flow cell of an Illumina Genome Analyzer II sequencer. After standard cluster generation, a sequencing primer that hybridizes immediately 3' of the poly(A) sequence is annealed and extended using a mixture of dTTP and biotin-conjugated dUTP. The flow cell is then placed onto the sequencing machine for sequencing by synthesis, which for each cluster provides the sequence immediately 5' of the poly(A) site. After sequencing, the flow cell is incubated with fluorophore-tagged streptavidin, which binds the biotin to impart a fluorescence intensity proportional to the poly(A)-tract length.

b, Median streptavidin fluorescence intensities for two sets of mRNA-like molecules with listed poly(A)-tail lengths, which were added to the 3T3 (circle), HEK293T (triangle), and HeLa (square) samples and used for tail-length calibration.

A cocktail of mRNA-like standards with known tail lengths (Extended Data Fig. 1b) was spiked into each starting sample to produce a standard curve for converting normalized fluorescence intensities to poly(A)-tail lengths (Fig. 1b). We refer to each of these tail-length measurements paired with its identifying sequence as a poly(A) tag.

Although recovery of tags from the standards varied somewhat, it did not vary systematically with tail length, which indicated that length-related biases did not affect our measurements (Extended Data Fig. 1c). Additional analyses indicated that mRNA degradation was not a potential source of bias against longer poly(A) tails (Extended Data Fig. 2a).

Because alternative start sites or alternative splicing can generate different transcripts with the same poly(A) site, we considered our results with respect to unique gene models (abbreviated as “genes”) rather than to transcripts (even though polyadenylation occurs on RNA transcripts, not genes). Moreover, tags for alternative poly(A) sites of the same gene were pooled, unless stated otherwise. With this pipeline, analysis of RNA from NIH3T3 mouse fibroblasts (3T3 cells) yielded at least one tag from 10,094 unique protein-coding genes (including 97% of the 9,976 genes with at least one mRNA molecule per cell) and ≥ 100 tags from 2,873 genes, coverage typical of most samples (Supplementary Information Table 1).

Tail-length diversity within each species

Median tail lengths in various mammalian cells (range, 67–96 nt) exceeded those in *Arabidopsis* leaves and *Drosophila* S2 cells (51 and 50 nt, respectively), which exceeded those in budding and fission yeasts (27 and 28 nt, respectively) (Fig. 2a). Similar differences between mammalian, fly, plant, and yeast cells were observed when comparing tail-length averages for individual genes (Fig. 2b). For genes within each species, mean tail lengths varied, with the 10th and 90th

percentiles differing by 1.4–1.6 fold. Variation was also observed for different mRNA transcripts from the same gene (Fig. 2c). For most genes the distributions were unimodal, with the mode approaching the mean (Fig. 2d). Poly(A)-tail lengths increased when progressing through cleavage, blastula, and gastrula stages of zebrafish embryonic development [2, 4 and 6 hours post-fertilization (hpf), respectively] and analogous stages of frog (*Xenopus laevis*) development (Fig. 2a,b and d).

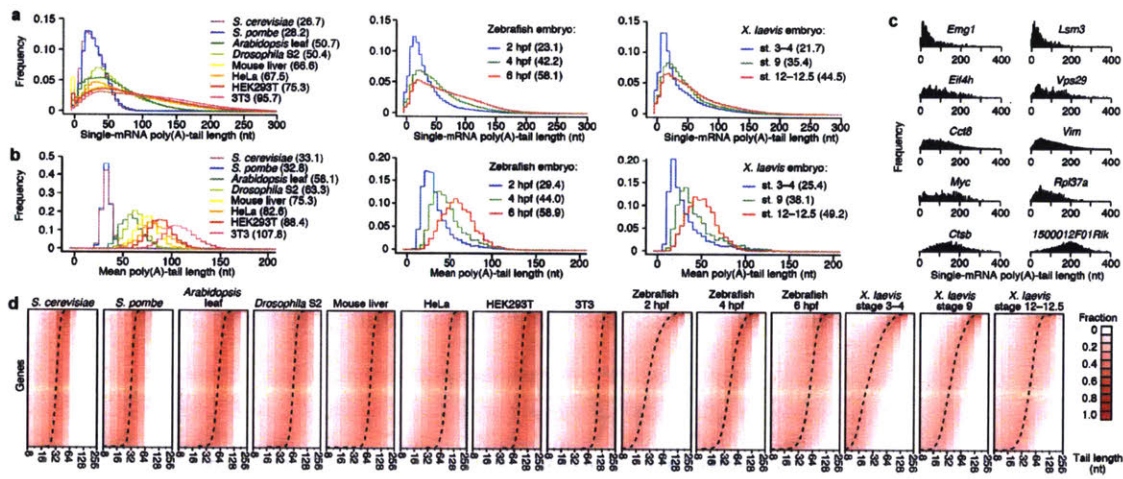


Figure 2. Poly(A)-tail lengths in yeast, plant, fly and vertebrate cells.

a, Global mRNA tail-length distributions. For each sample, histograms tally tail-length measurements for all poly(A) tags mapping to annotated 3' UTRs (bin size = 5 nt). Median tail lengths are in parentheses. The leftmost bin includes all measurements <0 nt. **b,** Intergenic tail-length distributions. For each sample, histograms tally average tail lengths for protein-coding genes with ≥ 50 tags (yeasts, zebrafish and *Xenopus*) or ≥ 100 tags (other samples). Median average tail lengths are in parentheses. **c,** Intragenic tail-length distributions for ten genes spanning the spectrum of average tail lengths in 3T3 cells. **d,** Intragenic tail-length distributions. Heatmaps show the frequency distribution of tail lengths for each gene tallied in panel b. The color intensity indicates the fraction of the total for the gene (key). Genes are ordered by average tail length (dashed line). Results from the first budding-yeast sample are reported in this figure.

Comparison of tail lengths for orthologous genes in human (HeLa and HEK293T) and mouse cells (3T3 and liver) revealed moderately strong correlations, indicating that tail lengths are conserved [Extended Data Table 1, Spearman R (R_s) as high as 0.46]. When analyzing gene classes that tended to have longer or shorter tails, the most striking and pervasive enrichment was

for ribosomal-protein genes and other ‘house-keeping’ genes among the short-tailed genes (Extended Data Table 2). This enrichment was strong in yeast, despite previous reports that ribosomal-protein genes tend to have long tails (Beilharz and Preiss, 2007; Lackner et al., 2007). To address this and other discrepancies with the previous yeast studies (Extended Data Fig. 3a and b), we used an independent method to measure the poly(A)-tail lengths of eight yeast genes, including four ribosomal-protein genes. The results were much more consistent with our measurements than with the previous measurements (Extended Data Fig. 3 and 4). Both previous reports used the polyadenylation state microarray (PASTA) method, which fractionates RNAs by stepwise thermal elution from poly(U)-Sepharose. Although many studies have successfully used poly(U)-Sepharose fractionation to detect tail-length changes for the same genes in different contexts (Rosenthal et al., 1983; Palatnik et al., 1984; Paynton et al., 1988), detecting differences between different genes in the same context is more challenging. Our results suggest that PASTA, as previously implemented in yeasts (Beilharz and Preiss, 2007; Lackner et al., 2007), is less suitable than PAL-seq for intergenic comparisons, although we cannot exclude the possibility that the discrepancies arose from different growth conditions.

The types of genes with shorter or longer tails differed between the embryonic samples and the other samples (Extended Data Table 1). Genes in the early embryo might not necessarily have the same tail lengths as their orthologs do in other contexts because prior to the maternal-to-zygotic transition (MZT), which occurs at ~3 hpf in zebrafish (Kane and Kimmel, 1993) and at approximately stage 8 in *X. laevis* (Newport and Kirschner, 1982), transcription is not yet active, and some maternal transcripts are masked for later use while others are subject to cytoplasmic polyadenylation (Weill et al., 2012). At 6 hpf in zebrafish, ribosomal protein mRNAs had switched from being enriched in shorter-tail genes to being enriched in longer-tail genes

(Extended Data Table 1), perhaps because these were mostly newly synthesized transcripts, which tended to have longer tails at this stage (Extended Data Fig. 5).

Because deadenylation is an important early step in eukaryotic mRNA decay (Decker and Parker, 1993; Goldstrohm and Wickens, 2008; Chen and Shyu, 2011), we examined the relationship between poly(A)-tail length and published mRNA-stability values (Extended Data Table 2). Tail-length and half-life were slightly negatively correlated in HeLa and 3T3 cells ($R_s = -0.048$ and -0.16 , respectively) and variably correlated in yeast, depending on the source of the half-life measurements ($R_s =$ ranging from -0.44 to 0.23). The weak relationships in HeLa and 3T3 cells would be expected if mRNAs with different half-lives have similar steady-state tail-length distributions with the less stable mRNAs transiting through the distributions more quickly.

No strong, easily interpretable correlations between tail length and mRNA features (3' UTR length, ORF length, total length, splice-site number and splice-site density) or expression (steady-state accumulation and nuclear-to-cytoplasmic ratio) were observed (Extended Data Table 1). Of these, the strongest correlations were between tail length and accumulation level (R_s ranging from -0.44 to 0.25), and between tail length and total mRNA length (R_s range, -0.12 to 0.36) or features related to mRNA length. Support for the latter relationship was also observed in intragenic comparisons, which revealed a weak positive relationship between tail length and the length of tandem 3'-UTR isoforms (Extended Data Fig. 6a). In early zebrafish embryos this relationship between 3'-UTR isoforms was even more pronounced when a predicted cytoplasmic polyadenylation element (CPE) (Richter, 1999; Aanes et al., 2011) was present in the unique region of the longer isoform (Extended Data Fig. 6b).

Transient coupling of tail length and translation in embryos

Most reports of increased translation of longer-tailed mRNAs are from studies in oocytes and early embryos (Richter, 1999; Weill et al., 2012). To examine whether this phenomenon reported in early embryos for a few genes applies transcriptome-wide, we performed ribosome footprint profiling and RNA-seq to measure translational efficiencies (TEs)(Ingolia et al., 2009) from the embryonic samples used to measure tail lengths. We found that in early embryos (cleavage and blastula stages) of both fish and frog, mean poly(A)-tail length correlated strongly with TE (Fig. 3a, R_s ranging from 0.62 to 0.77). Indeed, no other mRNA feature has been reported to correlate so well with TE in any system.

In these early embryonic stages a 2-fold increase in tail length corresponded to a large increase in TE—up to a 15-fold increase when doubling the tail from 16 to 32 nt in 4 hpf zebrafish (Fig. 3a). Although longer-tailed mRNAs were more likely to contain a CPE, the relationship between tail length and TE for CPE-containing mRNAs was no different from that of other mRNAs (Extended Data Fig. 7a). In theory, this coupling might not be causal, or it might be causal but strictly due to either translational inhibition causing tail shortening or translational activity preventing tail shortening. Alternatively, all or at least some of the coupling might result from longer tail lengths causing more efficient translation in the early embryo. We favor this last possibility because it agrees with the known importance of cytoplasmic polyadenylation for activating genes in maturing *Xenopus* oocytes (McGrew et al., 1989; Paris and Richter, 1990; Barkoff et al., 1998), early *Xenopus* embryos (Paris and Philippe, 1990; Simon et al., 1992) and certain other vertebrate contexts (Vassalli et al., 1989; Gebauer et al., 1994; Wu et al., 1998; Oh et al., 2000; Burns and Richter, 2008; Novoa et al., 2010). Even more

importantly, it agrees with the increased translation observed in *Xenopus* oocytes when appending prosthetic poly(A) tails of increasing length onto an mRNA (Barkoff et al., 1998).

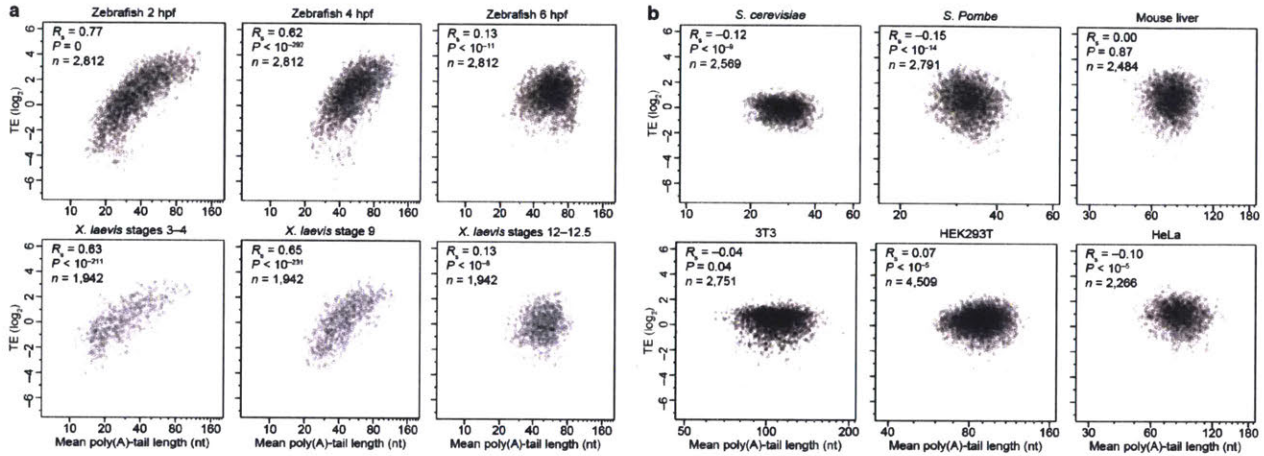


Figure 3. Transient coupling between poly(A)-tail length and TE.

a, The relationship between mean tail length and TE for genes with ≥ 50 poly(A) tags from embryonic samples at the indicated developmental stages. For each stage, tail lengths and TEs were obtained from the same sample. *MGC116473* fell outside the plot for *X. laevis*, stages 3–4, and *LOC100049085* fell outside for stages 12–12.5. **b**, Relationship between mean tail length and TE in the indicated cell types, considering genes with ≥ 50 tags for yeasts and ≥ 100 tags for others. With the exception of HeLa (Guo et al., 2010), tail lengths and TEs were obtained from the same samples. Budding-yeast *YBR196C*, *YLR355C* and *YDL080C*, fission-yeast *SPCC63.04.1*, mouse-liver *NM_007881* and *NM_145470*, HEK293T *NM_001007026*, *NM_021058*, and *NM_003537* and HeLa *NM_001007026* fell outside their respective plots.

The strong coupling observed in the blastula largely disappeared in gastrulating embryos (Fig. 3a, $R_s = 0.13$ and 0.17). Moreover, we observed no positive correlation of a meaningful magnitude between mean poly(A)-tail length and TE in HeLa cells, HEK293T cells, 3T3 cells, mouse liver, budding yeast, or fission yeast ($R_s = -0.10, 0.07, -0.04, 0.00, -0.12$, and -0.15 , respectively) (Fig. 3b). Our results in yeasts differed from those reported earlier (Beilharz and Preiss, 2007; Lackner et al., 2007), which we again attribute to limitations of methods used previously to rank tail lengths of different genes. In 3T3 cells, metabolic-labeling studies have been used to infer protein-synthesis rates (Schwanhäusser et al., 2011), which correlated with our TEs ($R_s = 0.44, P < 10^{-158}$) and did not correlate positively with tail length ($R_s = -0.20, P < 10^{-16}$).

Taken together, our results suggested that beginning at gastrulation, translational control undergoes a mechanistic change that uncouples TE from poly(A)-tail length.

Intragenic comparison of tail length and translation

The simplest interpretation of the weak or negative correlations we observed between tail length and TE in yeast and mammalian cells is that increasing tail length over the physiological range does not enhance translation in these contexts. However, our comparisons of average tail length and average TE between genes (Fig. 3b) might have missed a relationship that would be observed when looking at differentially translated mRNAs from the same gene. To address this possibility, we fractionated lysate from 3T3 cells to isolate mRNAs associated with different numbers of ribosomes and measured the tail lengths in each fraction (Fig. 4a). To learn how poly(A)-tail length related to ribosome density for individual genes, we plotted mean tail-length values as a function of the number of bound ribosomes and fit the data for each gene with a straight line (Fig. 4b). The slopes of these lines were generally quite small, and most were slightly negative (Fig. 4b), whereas positive slopes would have been expected if longer tails enhanced translation. Thus, the increase in median length observed between the lightest and heaviest fractions when considering bulk tail lengths (Fig. 4a; 66 and 82 nt, respectively) did not indicate a relationship between longer tails and enhanced translation but instead might have reflected the positive correlation between ORF length and tail length observed in 3T3 cells (Extended Data Table 1; $R_s = 0.36$) The trend of mostly negative slopes prevailed even when excluding data from mRNA not associated with any ribosomes (Extended Data Fig. 7b), or when examining subsets of genes with higher or lower translation efficiency, or with longer or shorter mean tail lengths (Extended Data Fig. 7c). This global intragenic analysis (Fig. 4) supports the

conclusion drawn from intergenic analyses (Fig. 3): In all yeast and mammalian contexts we examined, and presumably in most other cellular contexts, mRNAs with longer poly(A)-tails are not more efficiently translated.

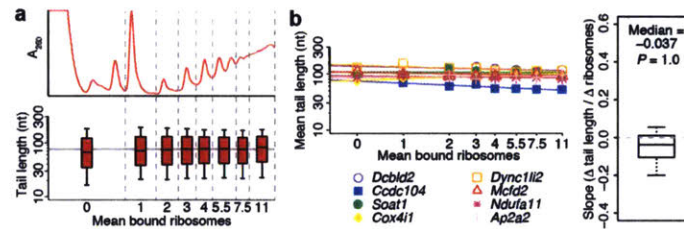


Figure 4. No detectable intragenic coupling between poly(A)-tail length and TE.

a, Global analysis of tail lengths across the polysome profile for 3T3 cells. UV absorbance indicates the mean number of ribosomes bound per mRNA for each fraction from the sucrose gradient (top, fractions demarcated with vertical dashed lines). Boxplots show distributions of tail lengths in each fraction for all tags mapping to annotated 3' UTRs (bottom). Boxplot percentiles are line, median; box; 25th and 75th percentiles; whiskers, 10th and 90th percentiles. The horizontal line indicates the median overall median of the median tail lengths. **b**, Relationship between tail lengths and ribosomes bound per mRNA when comparing mRNAs from the same gene. For each gene, the data from panel **a** were used to plot the mean tail length as a function of bound ribosomes. Log-log plots for eight randomly selected genes with ≥ 50 poly(A) tags in ≥ 6 fractions are shown (left), with lines indicating linear least-squared fits to the data (adding a pseudocount of 0.5 ribosomes to the fraction with zero ribosomes). The boxplot shows the distribution of slopes for all genes with ≥ 50 poly(A) tags in ≥ 4 fractions (right; one-sided, one-sample Wilcoxon test; boxplot percentiles as in panel **a**).

Explaining the shift in the ultimate effects of microRNAs

MicroRNAs (miRNAs) are 22-nt RNAs that pair to sites in mRNAs to target these messages for posttranscriptional repression (Bartel, 2009). Global measurements indicate that miRNA targeting causes mostly mRNA destabilization, with translational repression comprising a detectable but minor component of the overall repression (Baek et al., 2008; Hendrickson et al., 2009; Guo et al., 2010; Bazzini et al., 2012). The only known exception is the transient translational repression observed in early zebrafish embryos (Bazzini et al., 2012). At 4 hpf miR-430 targeting causes mostly translational repression with very little mRNA destabilization, whereas by 6 hpf the outcome shifts to mostly mRNA destabilization (Bazzini et al., 2012). Because miR-430 is

induced only ~1.5 h before the 4-hpf stage, these results are interpreted as revealing the dynamics of miRNA action, in which an early phase of translational repression gives way to a later phase in which destabilization dominates (Bazzini et al., 2012). When considering that miRNA targeting promotes poly(A)-tail shortening through the recruitment of deadenylase complexes (Braun et al., 2012), our results suggests an alternative mechanism for the shift in miRNA regulatory outcomes. In this mechanism, miRNAs mediate tail shortening at both 4 and 6 hpf, but because of the switch in the nature of translational control (as well as destabilization of short-tailed mRNAs at later stages), tail shortening has very different consequences in the two stages: At 4 hpf, tail shortening predominantly decreases TE, whereas at 6 hpf, it predominantly decreases mRNA stability.

To integrate miRNA-mediated repression with effects on tail length, we injected one-cell zebrafish embryos with miRNAs that are normally not present in the early embryo and examined the influence of these injected miRNAs on ribosome-protected fragments (RPFs), mRNA levels and poly(A)-tail lengths at 2, 4 and 6 hpf. Injecting miR-155 caused RPFs from many of its predicted targets to decrease relative to RPFs from no-site control mRNAs (Fig. 5a). Despite the decrease in RPFs, target mRNA levels did not change relative to the controls at 2 and 4 hpf, indicating that at these stages miR-155 targeting caused mostly translational repression. In contrast, RPF changes were accompanied by nearly commensurate mRNA reductions at 6 hpf, indicating that by this stage the outcome of repression had shifted to mostly mRNA destabilization (Fig. 5a). Thus, the shift in miRNA regulatory outcome that occurs between 4 and 6 hpf is not specific to miR-430 or its targets. With respect to mechanism, the observation of this shift between 4 and 6 hpf, even though the injected miR-155 was present and active much earlier than was miR-430, indicated that the shift reflected the unusual regulatory regime operating pre-

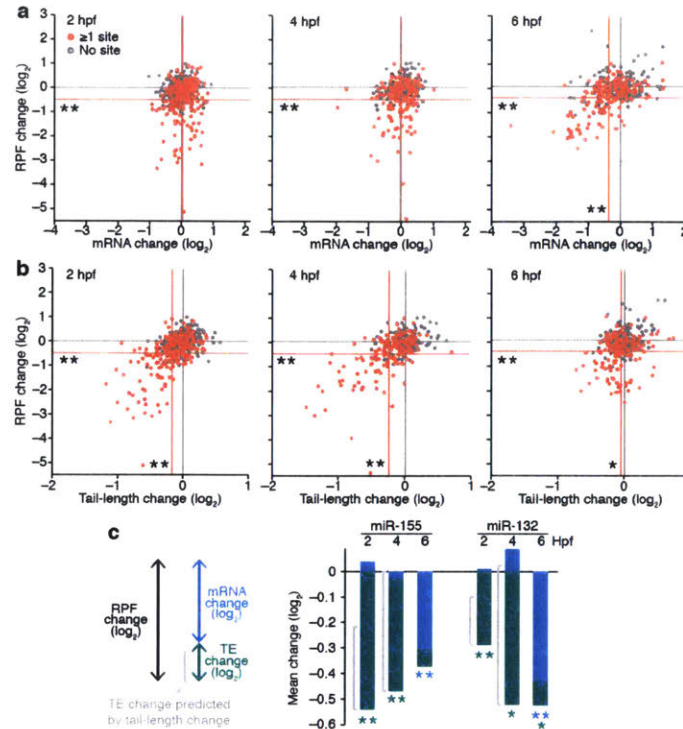


Figure 5. The influence of miR-155 on ribosomes, mRNA and tails in the early zebrafish embryo.

a, The relationship between changes in ribosome protected fragments (RPFs) and changes in mRNA levels after injecting miR-155. Changes observed between injected and mock-injected embryos are plotted at the indicated stages for predicted miR-155 target genes (red, genes with ≥ 1 miR-155 site in their 3' UTR) and control genes (gray, genes that have no miR-155 site yet resemble the predicted targets with respect to 3' UTR length). To ensure that differences observed between 4 and 6 hpf were not the result of examining different genes, only site-containing genes and no-site control genes present at both 4 and 6 hpf are shown for these samples. Lines indicate mean changes for the respective gene sets, with statistically significant differences between the sets indicated (*, $P \leq 0.05$; **, $P < 10^{-4}$, one-tailed Kolmogorov–Smirnov test). Because injected miRNAs partially inhibited miR-430-mediated repression, genes with miR-430 sites were not considered. All data were normalized to the median changes observed for the controls. **b,** The relationship between RPF changes and mean tail-length changes after injecting miR-155. Tail-lengths were determined using PAL-seq, otherwise as in **a**. **c,** A developmental switch in the dominant mode of miRNA-mediated repression. The schematic (left) depicts the components of the bar graphs, showing how the RPF changes were comprised of both mRNA changes and TE changes. The compound bar graphs show the fraction of repression attributed to mRNA degradation (blue) and TE (green) for the indicated stage, depicting the overall impact of miR-155 (center; plotting results from **a** and **b** for genes with sites) and miR-132 (right, plotting results from Extended Data Fig. 8b for genes with sites). Slight, albeit statistically insignificant, increases in mRNA for predicted targets resulted in blue bars extending above the axis. For samples from the stages in which tail length and TE are coupled, a bracket adjacent to the compound bar indicates the fraction of repression attributable to shortened tails. Significant changes for each component are indicated with asterisks of the corresponding color (*, $P \leq 0.05$; **, $P < 10^{-4}$, one-tailed Kolmogorov–Smirnov test).

gastrulation embryos (in which TE is sensitive to tail length) more than it reflected the dynamics of miRNA action.

The tail-length results further supported a mechanism involving shifting consequences of tail-length shortening. Predicted miR-155 targets had shortened tails at 2 and 4 hpf (Fig. 5b), which explained most of the miRNA-induced translational repression observed at these stages (Fig. 5c). By 6 hpf, the tail-length decreases observed at 4 hpf had mostly abated for predicted miR-155 targets (Fig. 5b), and these mRNAs were instead less abundant (Fig. 5a), in concordance with their extent of deadenylation at 4 hpf (Extended Data Fig. 8a). These observations agreed with the idea that tail shortening at later developmental stages destabilizes mRNAs, and suggested that the miRNA-mediated deadenylation occurring during the earlier developmental stages promotes decay later. With shorter tails no longer associated with reduced translation (Fig. 3) and instead associated with reduced mRNA levels, the ultimate consequence of miRNA-mediated repression shifted from translational repression to mRNA destabilization (Fig. 5c). Analogous results were obtained after injecting a different miRNA, miR-132 (Fig. 5c, Extended Data Fig. 8).

Because tail lengths were no longer strongly coupled to TE (Fig. 3), tail-length changes did not explain the decrease in mean TE observed at 6 hpf for miR-132 predicted targets (Fig. 5c). We conclude that in this and other systems for which poly(A)-tail length is uncoupled from TE, the translational repression often detected as a minor component of the overall repression (Baek et al., 2008; Hendrickson et al., 2009; Guo et al., 2010) arises from a mechanism different from the one that dominates pre-gastrulation.

Taken together, our results provide a compelling explanation for miRNA-mediated translational repression in the pre-gastrulation zebrafish embryo: MicroRNAs induce poly(A)

shortening, which decreases TE at this developmental period. Our results also explain why the pre-gastrulation zebrafish embryo is the only known context for which translational repression is the dominant outcome of miRNA-mediated regulation: In all other contexts examined, tail-length shortening causes mRNA destabilization with little or no effect on TE.

Two gene-regulatory regimes

Our results from yeast, mammalian cells in culture, and mouse liver refute the prevailing view that poly(A)-tail length broadly influences TE. In doing so, our results add to the known differences between the regulatory regime operating in these cells and that operating in early metazoan embryos, revealing that these cells lack the coupling between poly(A)-tail length and TE observed in the early embryo.

This newly recognized difference can be rationalized in light of the potential interplay among regulatory options available in the two regulatory regimes. The yeast, mammalian, and mid-gastrulation cells were all transcriptionally active, which offers ample opportunities for nuclear control of gene expression. Moreover, active transcription enables unstable mRNAs to be replaced should the need arise, thereby expanding the contexts in which differential mRNA stability can be exploited for gene control. Thus, an additional layer of control in which TE depends on poly(A)-tail length would be dispensable. More importantly, because this type of coupling would lower output from older mRNA molecules, which in the absence of cytoplasmic polyadenylation would often have shorter poly(A) tails, the utility of gene regulation through mRNA stability would be compromised. In this conventional regulatory regime, long-lived mRNAs would have less value if they were translated less efficiently because of their shorter tails.

For embryos at the cleavage stage, the regulatory regime is very different. These embryos are not transcriptionally active, which not only precludes the use of transcriptional and other nuclear processes to alter gene expression programs but also limits the use of differential mRNA stability, because degraded mRNAs cannot be replaced until zygotic transcription begins (Kane and Kimmel, 1993). Perhaps as a consequence, many mRNAs with short tails were observed (Fig. 2a), consistent with the known stability of short-tailed mRNAs in early embryos (Audic et al., 1997; Aanes et al., 2011). With no prospect of using differential transcription and limited potential to use differential mRNA stability, cells of these embryos apparently harness differential tail length for global gene control. This result expands on the known behavior of individual genes in *Xenopus* embryos (Paris and Philippe, 1990; Simon et al., 1992) and the observation that early embryonic cells have robust cytoplasmic polyadenylation, which increases the utility of a tail-length regulatory mechanism (Richter, 1999). Compared to metazoan cells subject to the standard regulatory regime (including 6-hpf zebrafish embryos and the mammalian cells examined), the cleavage-stage embryos had more uniform intragenic tail lengths and more variable intergenic lengths (Fig. 2d), as required for efficient harnessing of the tail-length regulatory regime. With their tail-length distribution also shifted towards shorter tails (Fig. 2b), the cleavage-stage embryos can most efficiently exploit the tail-length differences with the greatest impact (Fig. 3a).

The transition between these two very different gene-regulatory regimes was rapid but not immediate. Despite their zygotic transcription, late-blastula embryos still coupled tail length with translation. Indeed, to the extent that newly transcribed zygotic mRNA tended to have longer tails than did the maternally inherited mRNAs (Extended Data Fig. 5), the continued

coupling observed in this hybrid regime would act to increase the relative output from these newly minted mRNAs, thereby sharpening the MZT.

We suspect that the tail-length regulatory regime observed in early embryos also operates in other biological systems in which transcription is repressed (or distant) and cytoplasmic polyadenylation is active, which include early embryos of other metazoan species, maturing oocytes, and neuronal synapses (Weill et al., 2012). The ability to measure poly(A)-tail lengths at single-mRNA resolution should provide important insights in these systems.

Acknowledgements

We thank D. Weinberg, V. Auyeung, I. Ulitsky, C. Jan, J.-W. Nam, A. Shkumatava, S.-J. Hong, Y. Erlich and the Whitehead Genome Technology Core (V. Dhanapal, L. Francis, S. Gupta, J. Love and T. Volkert) for helpful discussions, J.-W. Nam, I. Ulitsky, and D. Weinberg for assistance with transcript annotation, C. Bresilla, X. Guo, S.-J. Hong, and A. Rothman for experimental assistance, and D. Weinberg for comments on the manuscript. Supported by NIH grant GM067031 (D.B.) and an NIH Medical Scientist Training Program fellowship (A.S.). D.B. is an investigator of the Howard Hughes Medical Institute.

Author Contributions A.S. developed PAL-seq, generated tail-length measurements, and performed associated analyses. S.E. performed ribosome profiling, RNA-seq and associated analyses. G.C. performed zebrafish injections and assisted with staging. All authors helped design the study and write the manuscript.

Author Information Sequencing data and the processed data for each gene are available at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE52809. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.B. (dbartel@wi.mit.edu).

Methods summary

Cytoplasmically enriched lysates were prepared from HEK293T, 3T3, mouse liver, *X. laevis*, fission yeast and zebrafish samples, as well as the second budding-yeast sample, were harvested and divided into three portions, one each for PAL-seq, RNA-seq, and ribosome profiling. For other contexts, if RNA-seq and ribosome profiling were performed, samples for these were harvested independently of those used for PAL-seq. PAL-seq was performed as outlined in Fig. 1a. RNA-seq and ribosome profiling were performed essentially as described previously (Guo et al., 2010). Poly(A) tags were mapped to a reference genome (or transcriptome) of the species, carrying forward those that mapped uniquely to the genome (or transcriptome) and also overlapped with the 3' UTR of a transcript model chosen to represent a gene. RPF and RNA-seq tags were mapped to the ORFs, as described previously (Guo et al., 2010), except tags mapping fully within the first 50 nt of an ORF were discarded (to exclude signal from ribosomes that might have initiated after cycloheximide was added). Each mRNA with a 3' UTR that had at least one 7-nt site matching the miRNA seed region (Bartel, 2009) was predicted to be a target of that miRNA. Genes that had no 6-nt miRNA seed match anywhere within their transcript were classified as no-site genes, from which a set of no-site control genes was selected such that its 3'-UTR length distribution matched that of the predicted targets.

References

- Aanes, H., Winata, C.L., Lin, C.H., Chen, J.P., Srinivasan, K.G., Lee, S.G., Lim, A.Y., Hajan, H.S., Collas, P., Bourque, G., *et al.* (2011). Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res* 21, 1328-1338.
- Audic, Y., Omilli, F., and Osborne, H.B. (1997). Postfertilization deadenylation of mRNAs in *Xenopus laevis* embryos is sufficient to cause their degradation at the blastula stage. *Mol Cell Biol* 17, 209-218.
- Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. *Nature* 455, 64-71.
- Barkoff, A., Ballantyne, S., and Wickens, M. (1998). Meiotic maturation in *Xenopus* requires polyadenylation of multiple mRNAs. *The EMBO journal* 17, 3168-3175.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215-233.
- Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* 336, 233-237.
- Beilharz, T.H., and Preiss, T. (2007). Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA* 13, 982-997.
- Braun, J.E., Huntzinger, E., and Izaurralde, E. (2012). A Molecular Link between miRISCs and Deadenylases Provides New Insight into the Mechanism of Gene Silencing by MicroRNAs. *Csh Perspect Biol* 4.
- Burns, D.M., and Richter, J.D. (2008). CPEB regulation of human cellular senescence, energy metabolism, and p53 mRNA translation. *Genes & development* 22, 3449-3460.
- Chen, C.Y., and Shyu, A.B. (2011). Mechanisms of deadenylation-dependent decay. *Wiley interdisciplinary reviews RNA* 2, 167-183.
- 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, 1632-1643.
- Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., *et al.* (2012). Landscape of transcription in human cells. *Nature* 489, 101-108.
- Eckmann, C.R., Rammelt, C., and Wahle, E. (2011). Control of poly(A) tail length. *Wiley interdisciplinary reviews RNA* 2, 348-361.
- Gebauer, F., Xu, W., Cooper, G.M., and Richter, J.D. (1994). Translational control by cytoplasmic polyadenylation of *c-mos* mRNA is necessary for oocyte maturation in the mouse. *The EMBO journal* 13, 5712-5720.

- Goldstrohm, A.C., and Wickens, M. (2008). Multifunctional deadenylase complexes diversify mRNA control. *Nature reviews Molecular cell biology* 9, 337-344.
- Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M.D., and Hughes, T.R. (2004). Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol* 24, 5534-5547.
- Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835-840.
- Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millan-Zambrano, G., Barkai, O., Chavez, S., Perez-Ortin, J.E., Darzacq, X., and Choder, M. (2013). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* 153, 1000-1011.
- Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS biology* 7, e1000238.
- Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717-728.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218-223.
- Jan, C.H., Friedman, R.C., Ruby, J.G., and Bartel, D.P. (2011). Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature* 469, 97-101.
- Kane, D.A., and Kimmel, C.B. (1993). The zebrafish midblastula transition. *Development* 119, 447-456.
- Lackner, D.H., Beilharz, T.H., Marguerat, S., Mata, J., Watt, S., Schubert, F., Preiss, T., and Bahler, J. (2007). A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol Cell* 26, 145-155.
- Larsson, E., Sander, C., and Marks, D. (2010). mRNA turnover rate limits siRNA and microRNA efficacy. *Mol Syst Biol* 6, 433.
- 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* 294, 858-862.
- McGrew, L.L., Dworkin-Rastl, E., Dworkin, M.B., and Richter, J.D. (1989). Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes & development* 3, 803-815.

- Meijer, H.A., Bushell, M., Hill, K., Gant, T.W., Willis, A.E., Jones, P., and de Moor, C.H. (2007). A novel method for poly(A) fractionation reveals a large population of mRNAs with a short poly(A) tail in mammalian cells. *Nucleic acids research* 35, e132.
- Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136, 688-700.
- Munchel, S.E., Shultzaberger, R.K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. *Mol Biol Cell* 22, 2787-2795.
- Newport, J., and Kirschner, M. (1982). A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell* 30, 687-696.
- Novoa, I., Gallego, J., Ferreira, P.G., and Mendez, R. (2010). Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. *Nature cell biology* 12, 447-456.
- Nutiu, R., Friedman, R.C., Luo, S., Khrebtukova, I., Silva, D., Li, R., Zhang, L., Schroth, G.P., and Burge, C.B. (2011). Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nature biotechnology* 29, 659-664.
- Oh, B., Hwang, S., McLaughlin, J., Solter, D., and Knowles, B.B. (2000). Timely translation during the mouse oocyte-to-embryo transition. *Development* 127, 3795-3803.
- Palatnik, C.M., Wilkins, C., and Jacobson, A. (1984). Translational control during early *Dictyostelium* development: possible involvement of poly(A) sequences. *Cell* 36, 1017-1025.
- Pall, G.S., Codony-Servat, C., Byrne, J., Ritchie, L., and Hamilton, A. (2007). Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic acids research* 35, e60.
- Paris, J., and Philippe, M. (1990). Poly(A) metabolism and polysomal recruitment of maternal mRNAs during early *Xenopus* development. *Dev Biol* 140, 221-224.
- Paris, J., and Richter, J.D. (1990). Maturation-specific polyadenylation and translational control: diversity of cytoplasmic polyadenylation elements, influence of poly(A) tail size, and formation of stable polyadenylation complexes. *Mol Cell Biol* 10, 5634-5645.
- Paynton, B.V., Rempel, R., and Bachvarova, R. (1988). Changes in state of adenylation and time course of degradation of maternal mRNAs during oocyte maturation and early embryonic development in the mouse. *Dev Biol* 129, 304-314.
- Preiss, T., Muckenthaler, M., and Hentze, M.W. (1998). Poly(A)-tail-promoted translation in yeast: implications for translational control. *RNA* 4, 1321-1331.
- Richter, J.D. (1999). Cytoplasmic polyadenylation in development and beyond. *Microbiology and molecular biology reviews* : MMBR 63, 446-456.

- Rosenthal, E.T., Tansey, T.R., and Ruderman, J.V. (1983). Sequence-specific adenylations and deadenylations accompany changes in the translation of maternal messenger RNA after fertilization of *Spisula* oocytes. *J Mol Biol* 166, 309-327.
- Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature* 448, 83-86.
- Salles, F.J., Lieberfarb, M.E., Wreden, C., Gergen, J.P., and Strickland, S. (1994). Coordinate initiation of *Drosophila* development by regulated polyadenylation of maternal messenger RNAs. *Science* 266, 1996-1999.
- Salles, F.J., Richards, W.G., and Strickland, S. (1999). Assaying the polyadenylation state of mRNAs. *Methods* 17, 38-45.
- Schurer, H., Lang, K., Schuster, J., and Morl, M. (2002). A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic acids research* 30, e56.
- 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, 337-342.
- Shalem, O., Dahan, O., Levo, M., Martinez, M.R., Furman, I., Segal, E., and Pilpel, Y. (2008). Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol Syst Biol* 4, 223.
- Simon, R., Tassan, J.P., and Richter, J.D. (1992). Translational control by poly(A) elongation during *Xenopus* development: differential repression and enhancement by a novel cytoplasmic polyadenylation element. *Genes & development* 6, 2580-2591.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545-15550.
- 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. *Mol Cell* 52, 52-62.
- Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Lariviere, L., Maier, K.C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res* 22, 1350-1359.
- Ulitsky, I., Shkumatava, A., Jan, C.H., Subtelny, A.O., Koppstein, D., Bell, G.W., Sive, H., and Bartel, D.P. (2012). Extensive alternative polyadenylation during zebrafish development. *Genome Res* 22, 2054-2066.

Vassalli, J.D., Huarte, J., Belin, D., Gubler, P., Vassalli, A., O'Connell, M.L., Parton, L.A., Rickles, R.J., and Strickland, S. (1989). Regulated polyadenylation controls mRNA translation during meiotic maturation of mouse oocytes. *Genes & development* 3, 2163-2171.

Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. (2002). Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A* 99, 5860-5865.

Weill, L., Belloc, E., Bava, F.A., and Mendez, R. (2012). Translational control by changes in poly(A) tail length: recycling mRNAs. *Nature structural & molecular biology* 19, 577-585.

Wu, L., Wells, D., Tay, J., Mendis, D., Abbott, M.A., Barnitt, A., Quinlan, E., Heynen, A., Fallon, J.R., and Richter, J.D. (1998). CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of α -CaMKII mRNA at synapses. *Neuron* 21, 1129-1139.

Methods

PAL-seq. Total RNA or RNA from cytoplasmically enriched lysate (~1–50 µg) was supplemented with two mixes of tail-length standards and trace marker RNA containing an internal ³²P-label (*) (ugagguaguagguuguau-agu*caauccuaaucauuccaauccuaaucauucaaaaaaaaaa, IDT). Polyadenylated ends in the mixture of cellular RNA and standards were ligated to a 3'-biotinylated adapter DNA oligonucleotide (p-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGACACATAC-biotin, IDT) in the presence of a splint oligonucleotide (TTCCGATCTTTTTTTTTT, IDT) using T4 Rnl2 (NEB) in an overnight reaction at 18°C. The RNA was partially digested with RNase T1 (Ambion) as described (Jan et al., 2011), extracted with phenol-chloroform, ethanol precipitated and then purified on a denaturing polyacrylamide gel (selecting 104–750 nt fragments), which removed residual unreacted 3' adapter. The extent of digestion by RNase T1 was estimated using the internally labeled marker RNA, by comparing the relative abundances of the full-length marker and cleavage products. The marker RNA was also used to quantify the efficiency of the preceding adapter ligation. Splinted-ligation products were captured on streptavidin M-280 Dynabeads (Invitrogen) and, while still bound to the beads, 5' phosphorylated with 3'-phosphatase-deficient T4 polynucleotide kinase (NEB) and ligated to a 5' adapter oligonucleotide (C3.spacer-CAAGCAGAAGACGGCATAACGAGTTCAGAGTTCTAcaguccgacgauc, IDT; uppercase, DNA; lowercase, RNA) using T4 Rnl1 (NEB) in an overnight reaction at 22°C. Following reverse transcription using SuperScript II (Invitrogen) and a primer oligonucleotide (AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACG, IDT), complementary DNA (cDNA) was liberated from the beads by base hydrolysis and purified on a denaturing

polyacrylamide gel (selecting 166–790 nt DNA). Purified cDNA was denatured at room temperature in 5–100 mM NaOH, neutralized with addition of HT1 hybridization buffer (Illumina) and applied to an Illumina flow cell (at a typical concentration of 1.0–1.2 pM). Standard cluster generation, linearization, 3'-end blocking and primer hybridization were performed on a cBot cluster generation system (Illumina).

After transferring the flow cell to a Cluster Station designed for an Illumina Genome Analyzer, the sequencing primer was extended using a reaction mix containing 100 units/ml Klenow polymerase (NEB), 200 nM dTTP and either 10 nM (yeast samples) or 4 nM (other samples) biotin-16-dUTP (Roche). Extension was for 30 minutes at 37°C, flowing a fresh aliquot ($\geq 50 \mu\text{l}$) of reaction mix every two minutes to replenish dNTPs. Following primer extension, the flow cell was placed on a Genome Analyzer II sequencer (Illumina) for 36 cycles of standard sequencing-by-synthesis. After three additional cycles of cleavage to remove any residual sequencing fluorophores, the flow cell was washed with buffer [40.25 mM phosphate buffered saline (PBS), pH 7.4, 0.1% Tween], blocked with streptavidin-binding buffer [300 $\mu\text{g/ml}$ bovine serum albumin (NEB), 40.25 mM PBS, pH 7.4, 0.1% Tween], washed with buffer again, and then imaged, as done previously (Nutiu et al., 2011). This cycle of wash, block, wash, image was then repeated with a binding step inserted after the blocking step, in which the flow cell was incubated with 30 nM Alexa Fluor 532 Streptavidin (Invitrogen) in streptavidin-binding buffer for 10 minutes at 20°C. This expanded cycle was then repeated two more times, but with 100 nM streptavidin included at the binding step. Fluorescence was captured in the T and G channels because the wavelength of the excitation laser for these channels (532 nm) was identical to the fluorophore excitation wavelength. The sequential imaging confirmed that the second 100 nM

streptavidin incubation did not increase mean cluster intensity (monitored in real time as part of the “first base report”), which indicated saturation of available biotin.

Calculation of poly(A)-tail lengths. Raw images taken during sequencing-by-synthesis and after binding of fluorescent streptavidin were processed with Firecrest image-analysis and Bustard base-calling software (Illumina, version 1.9.0, using default parameters) to generate a read (FASTQ) file and another file containing the position of each cluster, the read sequence, the quality score for each base, and the base intensities in all four channels for every cycle of sequencing and streptavidin binding. Reads (from all tiles) were aligned to a reference genome (hg18 for human, mm9 for mouse, dm3 for fly, danRer7 for fish, tair10 for *Arabidopsis*, Spombe1 for fission yeast, and sacCer3 for budding yeast) or a reference transcriptome (Unigene mRNA sequences for *X. laevis*) using the Bowtie program for short-read mapping and the parameters ‘-l 25 -n 2 -m 1 -3 z’, where z was the number of streptavidin-binding cycles plus one. Reads containing ambiguous base calls (as indicated by characters “N” or “.”) at any position in the first 36 nt were discarded, as were reads mapping to multiple genomic loci. Reads that did not map to the genome were aligned to Bowtie indexes corresponding to the tail-length standards. For the remaining reads, mapping to the genome and standards was repeated, accounting for the possibility that the read failed to map because the sequence extended past the poly(A)-proximal fragment of the transcript and into the 5' adaptor. This mapping was reiterated for 16 rounds (to capture tags of ≥ 20 nt), with each round considering previously unmapped reads in which the 5' adaptor sequence started a nucleotide closer to the beginning of the read (requiring a perfect match to only the final 6 nt of the adaptor after the fifth round). Before each round of mapping, the adapter sequence was stripped by adjusting the Bowtie “-3” parameter.

For the *Arabidopsis* sample, the first (3'-most) base was of low quality, and raw images from the first cycle of sequencing were excluded when performing image analysis and base calling. Consequently, sequence reads were 35 nt long, and only 15 rounds of iterative adapter trimming/sequence mapping were done.

For each read carried forward as mapping to a single locus (of either the genome or the length standards), the cluster fluorescence intensity in the T channel after the first 100 nM streptavidin flow-in was recorded as the raw streptavidin fluorescence intensity. From this raw intensity, the intensity after the 0 nM flow-in was subtracted as background, and the resulting background-subtracted intensity was divided by the relative cluster intensity observed during sequencing-by-synthesis, which normalized for the density of molecules within the cluster. The relative cluster intensity was calculated by first dividing the fluorescence intensity of every sequenced base in the read by the median intensity for that base among all clusters with the same base at the same position, and then taking the average of the resulting values over the length of the read. Normalized streptavidin intensities were transformed to poly(A)-tail lengths using linear regression parameters derived from the median intensities of the standards and their mode poly(A)-tail lengths. For yeast, *Arabidopsis*, *Drosophila*, *Xenopus* and zebrafish samples, only the standards with tails of 10, 50 and 100 nt were used in the linear regression. For the other samples, all of the standards were used except for one with a 300 nt tail (barcode sequence = CUCACUAUAC), which was typically not sufficiently abundant for accurate measurement of its tail length. Each tail length was then paired with the genomic (or standard) coordinates to yield a poly(A) tag.

Assigning poly(A) tags to genes. Reference transcript annotations were downloaded (in refFlat format) from the UCSC Genome browser or another database (Ensembl for zebrafish, Unigene for *X. laevis*, TAIR for *Arabidopsis* and PomBase for *S. pombe*). For human, mouse, zebrafish and fly, transcript 3' ends were re-annotated using poly(A) sites identified by 3P-seq (Jan et al., 2011) and the workflow described previously (Ulitsky et al., 2012). For each gene, a representative transcript model was chosen as the one that had the longest ORF and the longest 3' UTR corresponding to that ORF. These reference transcript databases are available for anonymous download at <http://web.wi.mit.edu/bartel/pub/publication.html>. *S. cerevisiae* representative transcript models were from the GSE53268. Poly(A) tags that overlapped the 3' UTR of the representative transcript model by at least one nucleotide were assigned to that gene. Poly(A) tags with length measurements <-50 and >1000 (which included $<0.0009\%$ of the tags in any sample) were excluded from all analyses. Mean tail-length measurements <1 nt (which included measurements from 11 analyzed genes) were replaced with a value of 1.0 nt in the intragenic analysis across the polysome gradient (Fig. 4b). When considering the depth of a representative PAL-seq dataset from 3T3 cells, we considered 1.0 RPKM as the RNA-seq level indicating an average of one mRNA molecule per cell. This estimate was conservative, in that a comparison to published mRNA abundances in 3T3 cells (Schwanhäusser et al., 2011) indicated that 1.0 RPKM from our experiment corresponded to about 0.2 mRNA molecules per 3T3 cell.

RNA preparation for PAL-seq. For libraries made from fission yeast, HEK293T, 3T3, mouse-liver, *X. laevis*, and mock-, miR-132- and miR-155-injected zebrafish samples, as well as the second budding-yeast sample, RNA was extracted from a portion of the lysates that were also used for ribosome profiling and RNA-seq. These cleared lysates were enriched in cytoplasm. For

libraries made from HeLa and polysome-gradient samples, RNA was extracted from similar cytoplasmically enriched lysates. For the polysome gradient fractionation (Fig. 4a), lysate preparation and centrifugation were performed as for ribosome profiling, but without nuclease digestion prior to fractionation. For other libraries, total RNA was used. The correlation observed when comparing PAL-seq results from cytoplasmically enriched and total RNA from HeLa cells (Extended Data Fig. 2b, $R_s = 0.84$) resembled that observed between comparable 3T3 biological replicates (Extended Data Fig. 2b, $R_s = 0.83$). The measured lengths in both types of RNA preparation were similar, despite the possibility that total RNA might have included more long-tailed mRNAs due to a population of nascent mRNAs that had full-length tails and were awaiting export. However, not all nuclear mRNAs had full-length tails (as some were still in the process of being polyadenylated at the time of harvesting), and the nuclear population of mRNAs awaiting export presumably comprised a small fraction of the cellular mRNAs.

Tail-length standards. The common 5' region of each standard and the unique 3' region, consisting of the standard-specific barcode and poly(A)-tail (Extended Data Fig. 1b), were synthesized separately and then ligated together to make full-length standards. To generate each 3' region, a 5'-phosphate-bearing RNA oligonucleotide (IDT) consisting of the barcode segment followed by a 10 nt poly(A) segment was extended with *E. coli* poly(A)-polymerase (NEB), with ATP concentration and reaction time adjusted to yield of tails of the desired length. To narrow the tail-length distribution, extension products were sequentially purified on two denaturing polyacrylamide gels, excising products with tails of the desired length range and reducing the variability of tailed RNA to be mostly within ~5–25 nt, depending on the length of tail added (Extended Data Fig. 1b). The 5' region of the standards was synthesized by in vitro transcription

of a template containing *Renilla* luciferase sequence followed by that of a modified HDV ribozyme (Schurer et al., 2002). After gel-purification of the 5' HDV self-cleavage product, the 2', 3'-cyclic phosphate at its 3' end was removed with T4 polynucleotide kinase (NEB; 3000 μ l reaction containing 30,000 U enzyme and 100 mM MES-NaOH, pH 5.5, 10 mM MgCl₂, 10 mM β -mercaptoethanol, 300 mM NaCl, 37°C, 6 h). After another gel purification, the dephosphorylated product was joined to the poly(A)-tailed barcode oligonucleotide by splinted ligation using T4 Rnl2 (NEB) and a bridge oligonucleotide with 10 nt of complementarity to each side of the ligation junction. Ligation products were gel-purified and mixed in desired ratios before being added to RNA samples for PAL-seq.

Ribosome footprint profiling. Immediately prior to harvesting, cultured mammalian cells were incubated with media containing 100 μ g/ml cycloheximide for 10 minutes at 37°C to stop translation elongation. Cells were washed twice with ice-cold 9.5 mM PBS, pH 7.3 containing 100 μ g/ml cycloheximide, and lysed by adding lysis buffer [10 mM Tris-HCl, pH 7.4, 5 mM MgCl₂, 100 mM KCl, 2 mM dithiothreitol, 100 μ g/ml cycloheximide, 1% Triton X-100, 500 units/ml RNasin Plus, and protease inhibitor (1X complete, EDTA-free, Roche)] and triturating four times with a 26-gauge needle. After centrifuging the crude lysate at 1300g for 10 minutes at 4°C, the supernatant was removed and flash-frozen in liquid nitrogen. Cultured *S. pombe* cells were grown to mid-log phase and then harvested by filtering off the media and flash freezing the remaining paste, which was then manually ground into a fine powder with a mortar and pestle while being bathed in liquid nitrogen. The powder was thawed on ice, resuspended in lysis buffer and processed as described for the other lysates. Zebrafish embryos were enzymatically dechorionated and then incubated in 100 μ g/ml cycloheximide in E3 buffer (5 mM NaCl, 0.17

mM KCl, 0.33 mM CaCl₂, 0.33 mM MgSO₄) for 5 minutes at room temperature. The embryos were then transferred into lysis buffer and flash-frozen. *X. laevis* embryos were chemically dejellied after fertilization and flash-frozen in lysis buffer without cycloheximide pre-treatment. Once thawed, these samples were clarified as above and then processed in the same manner as other lysates. Prior to dissecting liver, a 6-week-old, male C57BL/6 mouse was sacrificed by cervical dislocation. The liver was excised, flash-frozen, and manually ground and processed as described for *S. pombe*. Ribosome profiling and RNA-seq were performed on the cleared lysate as described (Guo et al., 2010), using RiboMinus-treated RNA for the *S. pombe* RNA-seq sample, and poly(A)-selected RNA for all others, with a detailed protocol available at <http://bartellab.wi.mit.edu/protocols>. *S. cerevisiae* RPF and RNA-seq data were from GSE53268 and were derived from the same sample as the budding-yeast PAL-seq sample 2.

RPF and RNA-seq tags were mapped to the ORFs, as described previously (Guo et al., 2010) (using the assemblies and transcript models used for PAL-seq), except reads with 5' ends mapping within the first 50 nt of each ORF were disregarded. This was done to minimize a bias from ribosomes accumulating at or shortly after the start codon, which results from translation initiation events continuing in the face of cycloheximide-inhibited elongation (Ingolia et al., 2009). Because of this bias, genes with shorter ORFs have artifactually higher TEs if all the bound ribosomes are considered (as in conventional polysome gradient analysis). This cycloheximide effect might have distorted the TE measurements in studies that calculated ribosome densities using polysome gradient fractionation followed by microarray analysis, including those reporting a positive correlation between ribosome density and poly(A)-tail length (Beilharz and Preiss, 2007; Lackner et al., 2007). However, this effect could not have influenced the conclusions of our polysome-gradient experiment, because our analysis focused on intragenic

comparisons (Fig. 4b). TEs were considered only for genes exceeding a cutoff of 10 RPM (reads per million uniquely mapped reads) in the RNA-seq library. For the analysis of miRNA effects, only genes exceeding a cutoff of 10 RPM in the mock-injected RNA-seq and RPF libraries, and ≥ 50 PAL-seq tags in the mock-injected and miRNA-injected samples were considered.

Statistics, reagents and animal models. All statistical tests were two-sided unless indicated otherwise. No power testing was done to anticipate the sample size needed for adequate statistical power. No randomization or blinding was used for miRNA injection experiments. Features of mRNAs (e.g. poly(A)-tail length, mRNA length, expression level, etc.) were not normally distributed, nor were changes in expression due to miRNA-mediated repression. Therefore, non-parametric measures or tests were used when making comparisons involving such quantities, and these tests do not make assumptions about equal variance between groups. Mammalian cell lines were obtained from ATCC, and S2 cells were the same as in (Ruby et al., 2007) (i.e. adapted to growth in serum-free media). The BY4741 strain was used for *S. cerevisiae*, 972 for *S. pombe*, Columbia for *Arabidopsis*, and AB for zebrafish. All animal experiments were performed in accordance with a protocol approved by the MIT Committee on Animal Care.

Zebrafish injections. Zebrafish embryos were injected at the one-cell stage with 1 nl of 10 μ M miRNA duplex (miR-155 and miR-132) or buffer alone using a PLI-100 Plus Pico-Injector. Duplexes were made by combining RNAs corresponding to either miR-132 (uaacagucuacagccauggucg) and miR-132* (accguggcauuagauuguuacu) or miR-155 (uuaaugcuaaucgugauaggggu) and miR-155* (accuaugcuguuagcauuauuc) in annealing buffer (30

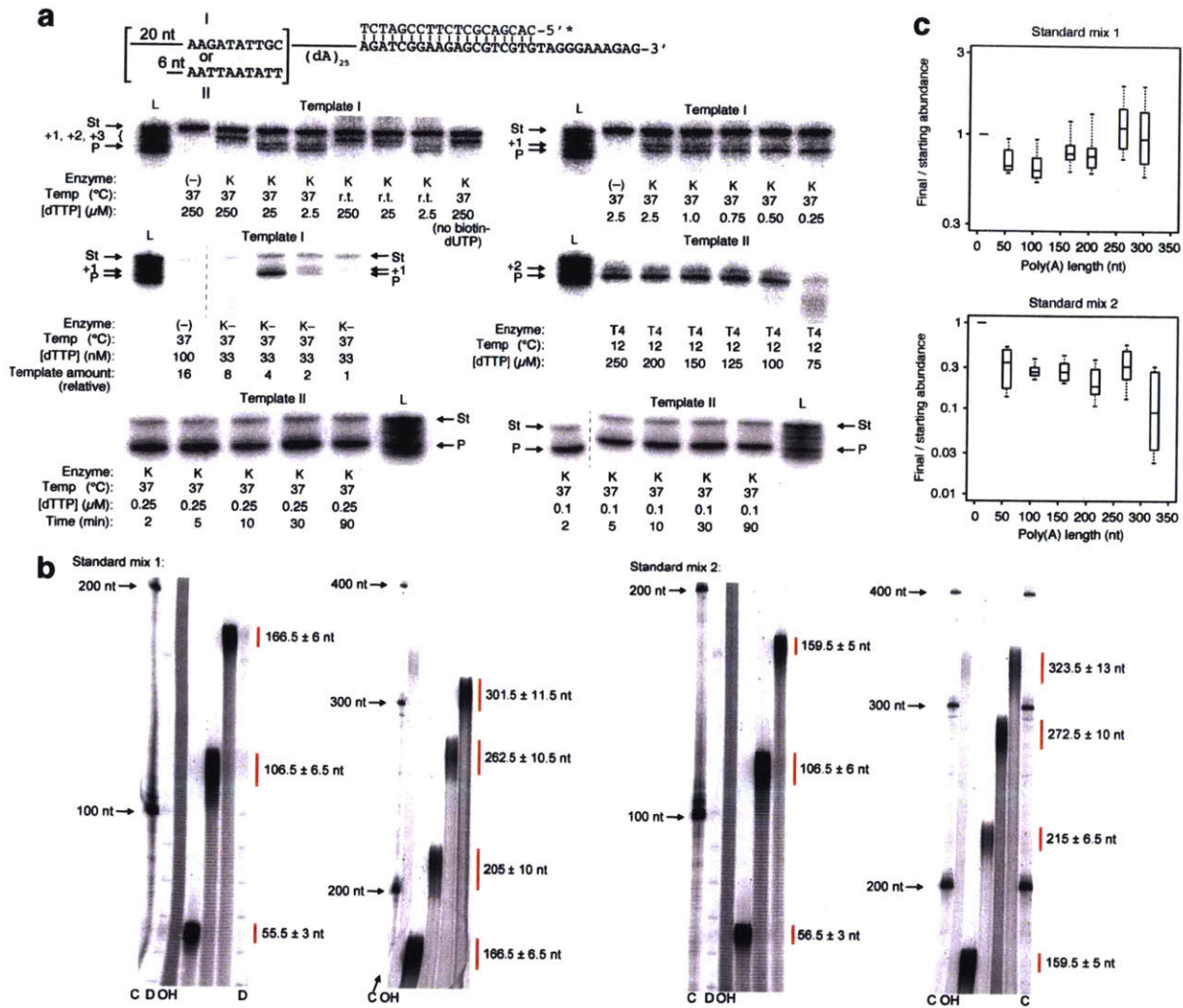
mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.1 mM EDTA), heating to 90°C for 1 minute, and slow cooling to room temperature over several hours. Injected embryos were incubated in E3 buffer at 28°C until time of harvesting.

Predicted miRNA targets. Each mRNA with a 3' UTR that had at least one 7-nt site matching the miRNA seed region (Bartel, 2009) was predicted to be a target of that miRNA. Genes that had no 6-nt miRNA seed match anywhere within their transcript were classified as no-site genes, from which a set of no-site control genes was selected such that its 3'-UTR length distribution matched that of the predicted targets.

Calculation of the relationship between poly(A)-tail length and TE. For experiments in which zebrafish embryos were mock-injected or injected with miR-132 or miR-155, least-squares second-order polynomial regression was performed to determine the change in \log_2 TE for each change in \log_2 poly(A)-tail length. To prevent microRNA effects on TE and/or tail length from influencing any relationship, the regression analyses were performed after excluding genes for which the mRNAs contained a perfect match to either the seed (nucleotides 2–7 of the miRNA) of miR-430 (the predominant endogenous miRNA at 4 and 6 hpf) or the seed of the injected miRNA. These regression results were used to estimate the TE change attributable to tail-length change for each gene.

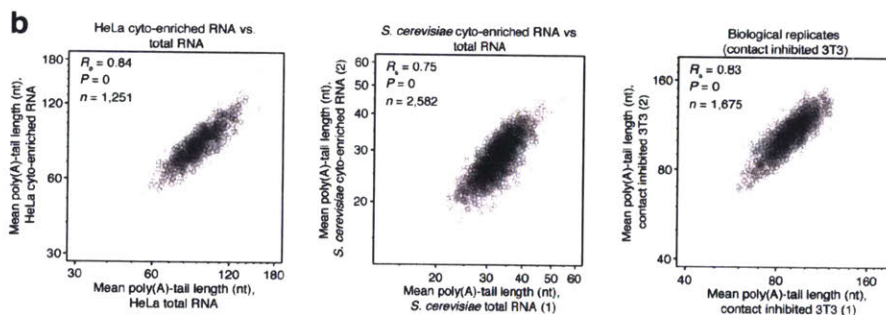
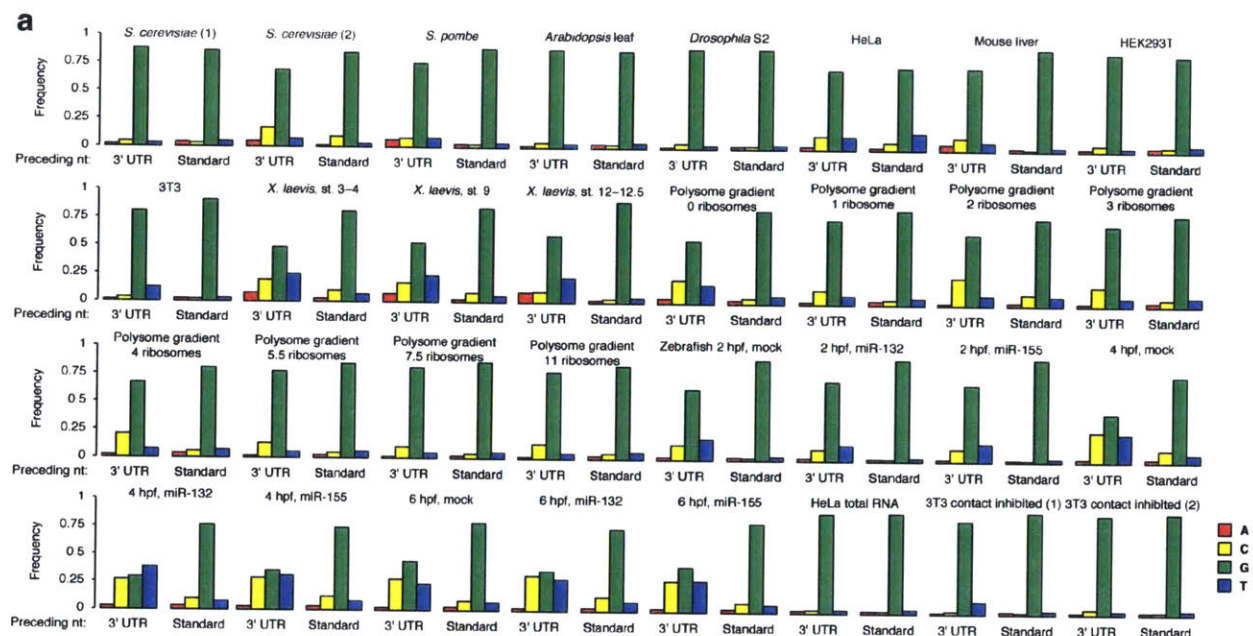
Poly(A)-tail measurements on RNA blots. Single-gene poly(A)-tail lengths were measured on RNA blots after directed RNase H cleavage of the interrogated mRNA. Standard methods (Salles et al., 1999) were modified to enable higher resolution for shorter tails (<50 nt), such as those

found on yeast mRNAs. Total RNA (3–20 μg) was heat-denatured for 5 minutes at 65°C in the presence or absence of 33 pmol/ μg total RNA of (dT)₁₈ (IDT), and in the presence of 25 pmol of a DNA oligonucleotide (or gapmer oligonucleotide, which had 16 DNA nucleotides flanked on each side by five 2'-*O*-methyl RNA nucleotides) that was complementary to the 3'-terminal region of the interrogated mRNA. After snap-cooling on ice, the RNA was treated with RNase H (Invitrogen) for 30 minutes at 37°C in a 20 μl reaction according to the manufacturer's instructions. The reaction was stopped by addition of gel loading buffer (95% formamide, 18 mM EDTA, 0.025% SDS, dyes) and then analyzed on RNA blots resembling those used for small-RNA detection (Lau et al., 2001) (Detailed RNA blot protocol available at <http://bartellab.wi.mit.edu/protocols>). Briefly, after separation of the RNA on a denaturing polyacrylamide gel and transfer onto a Hybond-NX membrane (GE Healthcare), the blot was treated with EDC (*N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide; Sigma-Aldrich), which crosslinked the 5' phosphate of the 3'-terminal RNase H cleavage product to the membrane (Pall et al., 2007). The blot was then hybridized to a probe designed to pair to the region spanning the RNase H cleavage site and the poly(A) site. Comparison of these 3'-terminal fragments with and without poly(A) tails revealed the length of the tails.



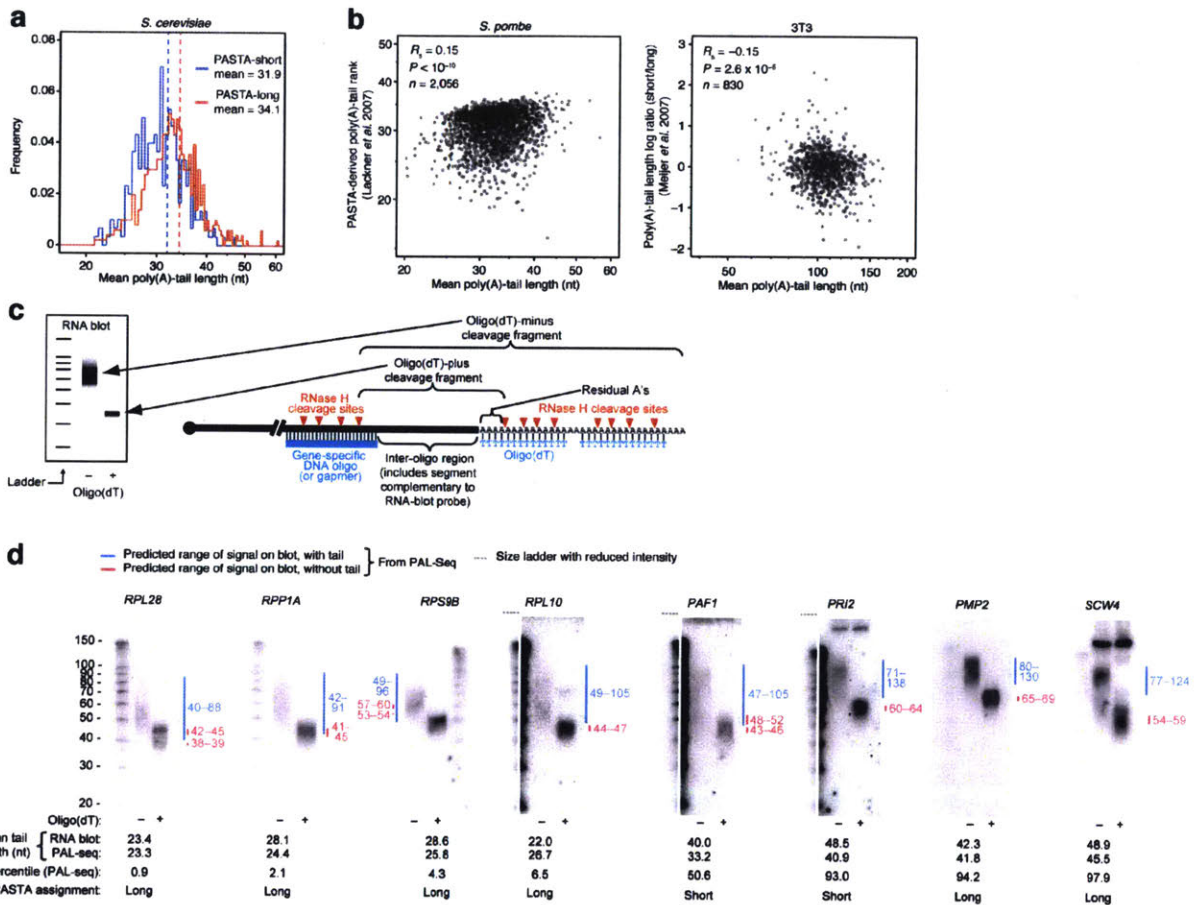
Extended Data Figure 1. Development and characterization of the PAL-seq method. **a**, Optimization of the primer-extension reaction. A 5'-radiolabeled primer was annealed to a single-stranded DNA template containing a (dA)₂₅ tract immediately upstream of the primer-binding site (top schematic). Two templates (I and II), which differed at the segment immediately 5' of the poly(dA) tract, were used in the experiments shown. Primer extension was performed with either Klenow fragment [K, New England Biolabs (NEB)], Klenow fragment lacking 3'-to-5' exonuclease activity (K-, NEB), or T4 DNA polymerase (T4, NEB). Reactions contained the recommended buffer and enzyme concentrations and a 50:1 mixture of dTTP:biotin-16-dUTP (unless indicated otherwise) at the dTTP concentrations indicated. In one experiment (center left), the dTTP concentration was kept constant, and the concentration of the primer-template duplex was varied instead. Reactions were incubated for five minutes, unless stated otherwise (bottom two panels), at the indicated temperature (Temp), then stopped and in most cases supplemented with gel-mobility standard (St), which was a ³²P-labeled synthetic oligonucleotide that had four extra dT residues appended to the intended the full-length primer-extension product (P). Products were resolved on denaturing polyacrylamide gels, alongside a size ladder, which was a mixture of ³²P-labeled oligonucleotides that differed from the full-length primer-extension product by -1, 0, +1, +2, +3, and +4 dT's (three of these are indicated as

+1, +2, +3), and visualized using a phosphorimager. Klenow fragment was capable of yielding the full-length primer-extension product without abundant side-products containing one or more additional untemplated nucleotides. The other two enzymes gave a mixture of products in all conditions tested. Incorporation of untemplated nucleotides by Klenow fragment was favored at lower temperature (upper left) and was largely eliminated by using very low dTTP concentrations at 37°C (upper right panel and bottom two panels). Under these favorable conditions the distribution of products did not change when the reaction times were extended beyond five minutes (bottom). **b**, Poly(A)-tail lengths of the synthetic standards added to each biological sample. Each standard was generated by splinted ligation of a common 5' RNA to a standard-specific 3' RNA consisting of a unique 10-nt barcode sequence followed by a poly(A) tail of either 10, ~50, ~100, ~150, ~200, ~250, or ~300 nt. Because the poly(A) tails >10 nt were generated enzymatically, their lengths were approximate, and after gel-purification the barcode-poly(A) RNAs retained some microheterogeneity in their lengths. To determine the actual poly(A) lengths of the gel-purified barcode-poly(A) RNAs following synthesis, each RNA was ³³P-labeled at its 5' terminus and analyzed on denaturing polyacrylamide gels under conditions that enable single-nucleotide resolution for long polynucleotides. The values to the right of each panel indicate the modes and approximate ranges of the poly(A) tail lengths (after accounting for the 10-nt barcodes). Also shown are marker lanes with ³³P-labeled Century Plus ladder (C+, Ambion), ³³P-labeled Decade ladder (D, Ambion) and a partial base-hydrolysis ladder of the labeled barcode-poly(A) RNA used to make the ~300 nt standard of mix 2 (OH). **c**, The relative yield of each poly(A)-length standard. For each standard in the indicated mix, the yield of poly(A) tags relative to that of the A₁₀ standard was plotted, after normalizing to the starting ratio determined from analysis of 5'-labeled mix on a denaturing polyacrylamide gel (panel **b**). The boxplots show the distribution of yields for 32 PAL-seq libraries (line, median; box, 25th and 75th percentiles; whiskers, 10th and 90th percentiles).



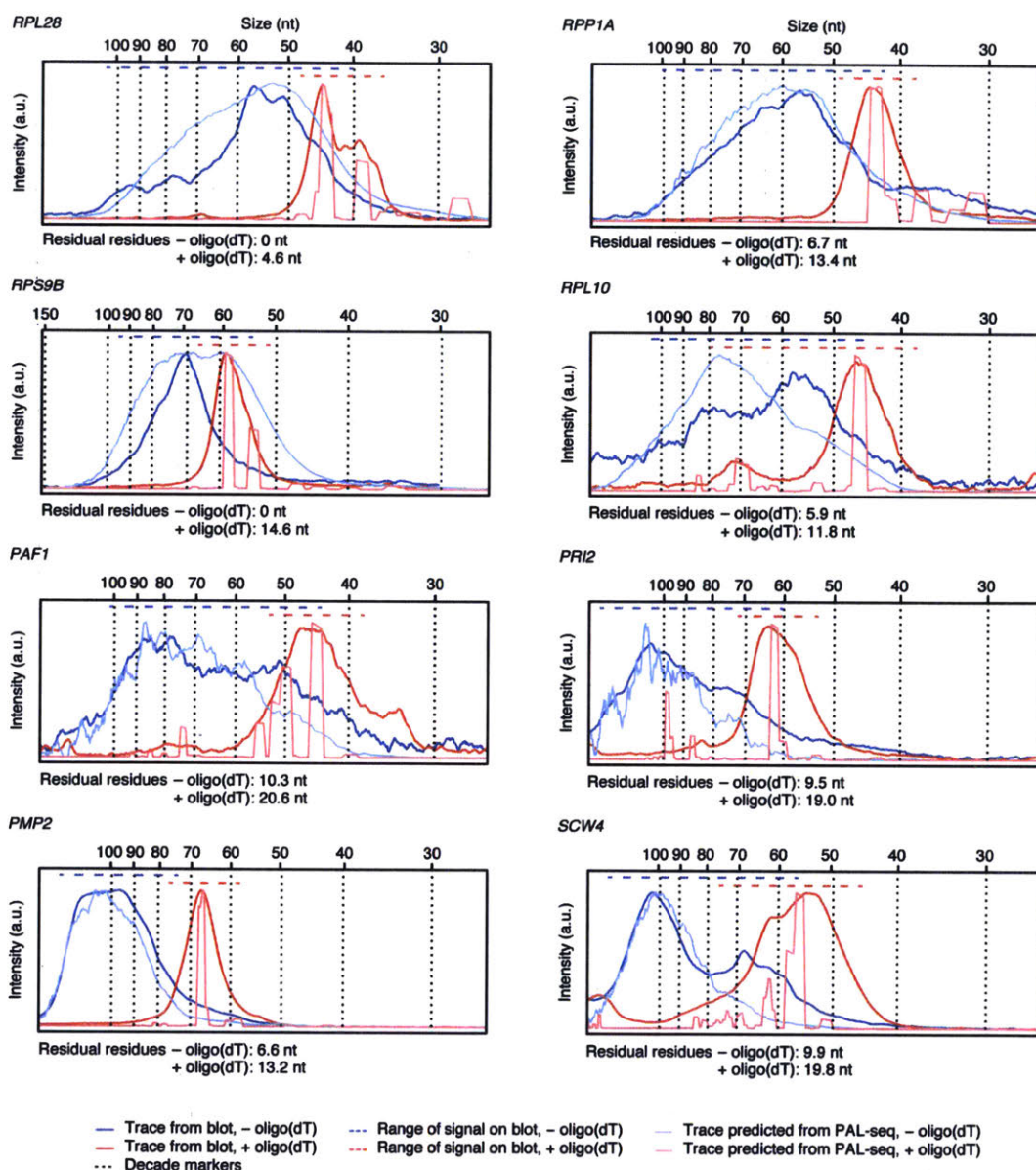
Extended Data Figure 2. Validation of PAL-seq performance. **a**, Evidence against non-specific RNA degradation. Plotted are nucleotide identities at the positions immediately upstream of poly(A) tags that both mapped uniquely to the genome and ranged from 21–30 nt in length (a range chosen to be long enough to enable mostly unique mapping to the genome yet short enough to include enough 5' adapter nucleotides in a 36-nt read to clearly identify the 5' end of the tag). Frequencies were normalized to the aggregate nucleotide composition of positions 23–31 in either uniquely genome- or standard-mapping tags that extended the full length of the reads (36 nt). Because RNase T1 cuts after G's, the nucleotide preceding each 22–30-nt tag was expected to be G, unless the mRNA had been cut for some other reason. The high frequency of G indicated that most mRNA fragments had not been cut for other reasons, which also implied that for these samples the poly(A) tails had also remained intact. We are unable to explain the high signal for an upstream U or C in some samples. Nonetheless, the frequency of an upstream A was low, which indicated that there had been little cleavage after A's, again implying that the poly(A) tails had remained intact. In the case of *Arabidopsis* leaf, for which the raw sequence reads had the first base removed, estimation of RNA integrity was performed as described above but with the indicated length ranges shortened by one nucleotide (e.g. informative poly(A) tags were 21–29 nt long). **b**, Consistent results from samples that function as biological replicates. Plotted are the relationships between average poly(A)-tail lengths generated using total HeLa RNA or RNA from a cytoplasmically enriched HeLa lysate (left), between average poly(A)-tail

lengths generated using budding-yeast total RNA or RNA from a cytoplasmically enriched yeast lysate (sample 1 and 2, respectively; middle), and between average poly(A)-tail lengths generated using cytoplasmically enriched lysates from two different 3T3 cell lines (right). Although the 3T3 lines were each engineered to enable inducible expression of a miRNA (either miR-1 or miR-155), the miRNA was not induced in the cells used for this comparison. (The 3T3 cells of this experiment happened to be growth-arrested, whereas the 3T3 cells in the experiments reported in the main text were actively dividing, but because growth arrest had little influence on tail length, these samples were suitable for illustrating the reproducibility of the method.) *NM_001007026* fell outside the plot for HeLa, and *YDL080C* fell outside the plot for yeast.

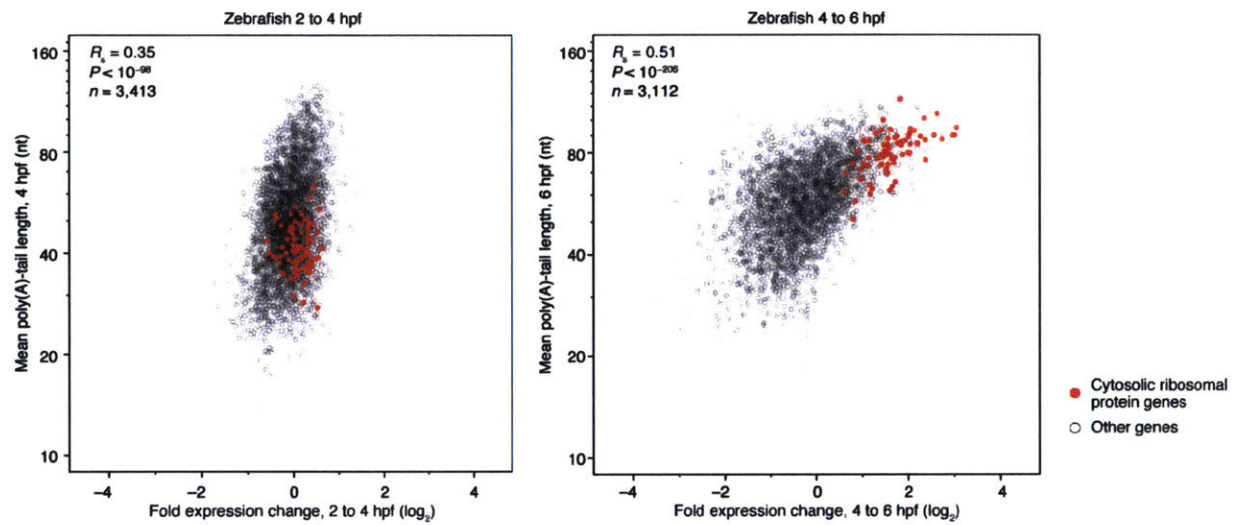


Extended Data Figure 3. Discrepancies between the results of PAL-seq and those of previous methods and experimental evaluation of PAL-seq results in *S. cerevisiae*. **a**, Comparison of *S. cerevisiae* poly(A)-tail lengths measured by performing PAL-seq on total RNA to the previous results from PASTA analysis (Beilharz and Preiss, 2007). Plotted are mean poly(A)-tail lengths measured by PAL-seq for genes previously classified as having short and long tails (PASTA-short and PASTA-long, respectively) (Beilharz and Preiss, 2007). The vertical dashed lines indicate the mean for each group as measured by PAL-seq. **b**, Comparison between PAL-seq measurements and either PASTA-derived poly(A)-tail ranks in fission yeast (Lackner et al., 2007) (left), or results of a related method reporting log ratios of short- and long-tail fractions in actively dividing 3T3 cells (Meijer et al., 2007) (right). **c**, Schematic of tail-length measurements using RNA blots. A DNA oligonucleotide or a gapmer (chimeric oligonucleotide with DNA flanked by 2'-*O*-methyl-RNA) was designed to pair near the 3' end of the mRNA. This oligonucleotide directed RNase H cleavage, thereby generating 3'-terminal mRNA fragments with lengths suitable for high-resolution analysis on RNA blots. Some of each sample was also incubated with oligo(dT), which directed RNase H removal of most of the poly(A) tail. Cleavage fragments were resolved on RNA blots resembling those normally used for microRNA analysis and detected by probing for the inter-oligo region of the mRNA. The average poly(A)-tail length was calculated as the difference in the average sizes of the oligo(dT)-minus and oligo(dT)-plus fragments, plus the average number of residual adenosine residues that remained because of incomplete digestion of the poly(A) tail (residual A's). For each reaction guided by a gene-

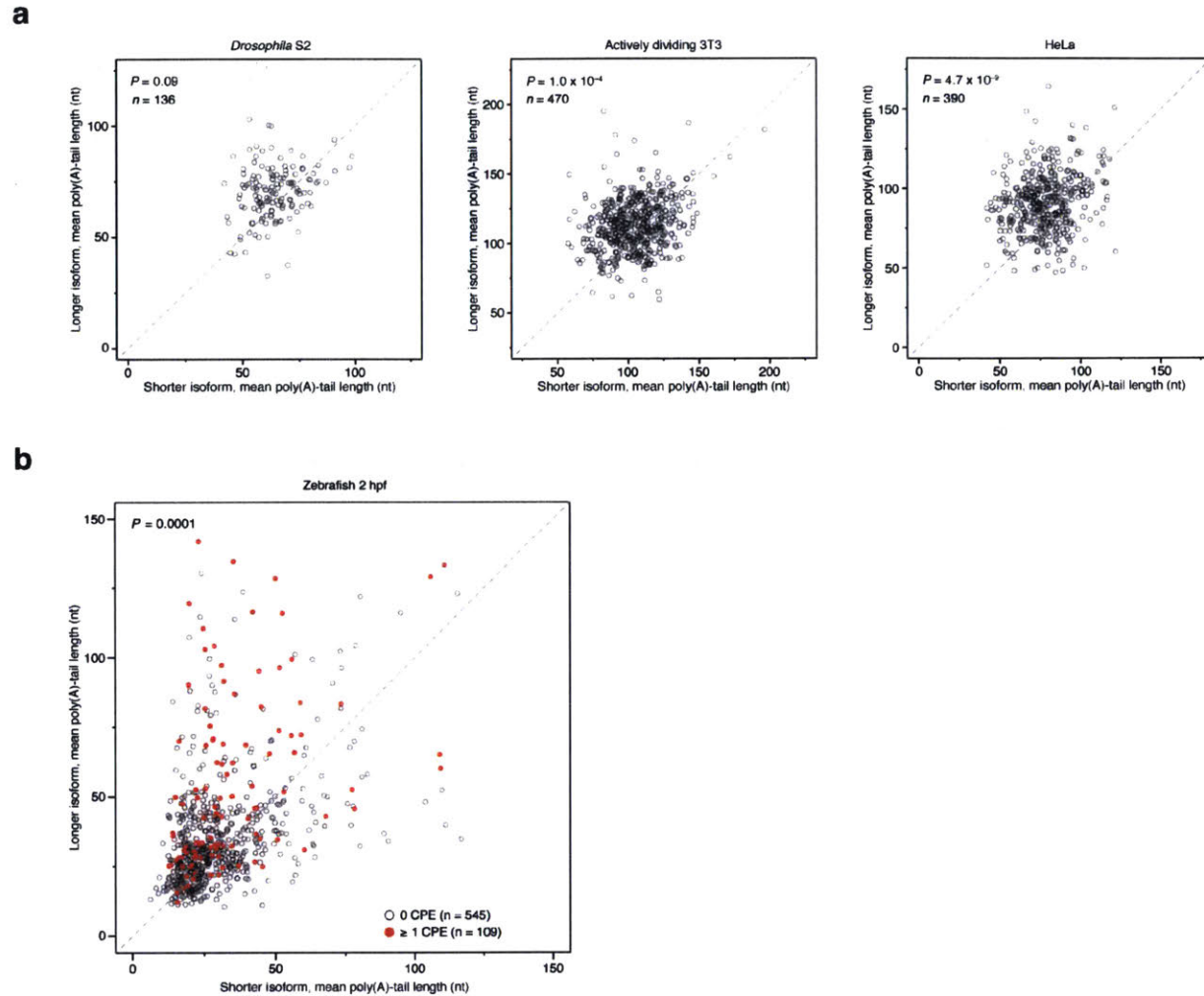
specific DNA oligo, the average number of residual adenosines was estimated as half the difference between the known length of the inter-oligo region and the observed length of the oligo(dT)-plus fragment. For the two reactions guided by a gene-specific gapmer (*RPL28* and *RPS9B*), the inter-oligo region extended through the residues pairing to one of the 2'-*O*-methyl-RNA segments, and cleavage was assumed to occur across from the most poly(A)-proximal DNA residue. Thus, the average number of residual adenosines was estimated as the difference between the inter-oligo region and the observed length of the oligo(dT)-plus fragment. **d**, RNA blots used to measure poly(A)-tail lengths, as described in panel **c**, with the length information determined by PAL-seq (on total RNA) and PASTA (Beilharz and Preiss, 2007) indicated below each blot for comparison. For each lane, the range of high signal predicted based on PAL-seq results (Extended Data Fig. 4) is shown as a line next to the blot [with and without oligo(dT), red and blue, respectively]. These predicted sizes accounted for the residual nucleotides flanking the inter-oligo region, using the migration of the oligo(dT)-plus fragment to estimate the residual nucleotides on one or both ends as described in panel **c**. Genes chosen for analysis were required to be adequately expressed and to have a relatively homogeneous cleavage and poly(A) site, as determined by 3P-seq (data not shown). [Nonetheless, some genes, such as *RPL28* had frequent poly(A) events falling into noncontiguous segments, as reflected by the two ranges marked in red.] A preference was also given to ribosomal protein genes and those with discordant poly(A)-tail lengths when comparing the results of PAL-seq and PASTA.



Extended Data Figure 4. The signal distributions for the RNA blots (Extended Data Fig. 3d) compared with those predicted using PAL-Seq. Predicted traces from PAL-Seq accounted for the estimated number of residual nucleotides flanking the inter-oligo region after RNase H cleavage, as described (Extended Data Fig. 3c). The offsets added to account for these residual nucleotides are indicated below each plot. The horizontal dashed lines above each plot indicate the range of the signal determined by visual inspection of the RNA blots in Extended Data Fig. 3d [oligo(dT)-plus and minus, blue and red, respectively]. Vertical dashed lines indicate the migration of Decade markers (Ambion). The vertical axes are in arbitrary units (a.u.). The range of the high signal predicted based on PAL-Seq data (signal exceeding 33% of the maximum) was determined using these plots and shown on Extended Data Fig. 3d as vertical lines next to the RNA blots. [For some genes, poly(A)-site heterogeneity caused the signal exceeding 33% to map to noncontiguous segments.]

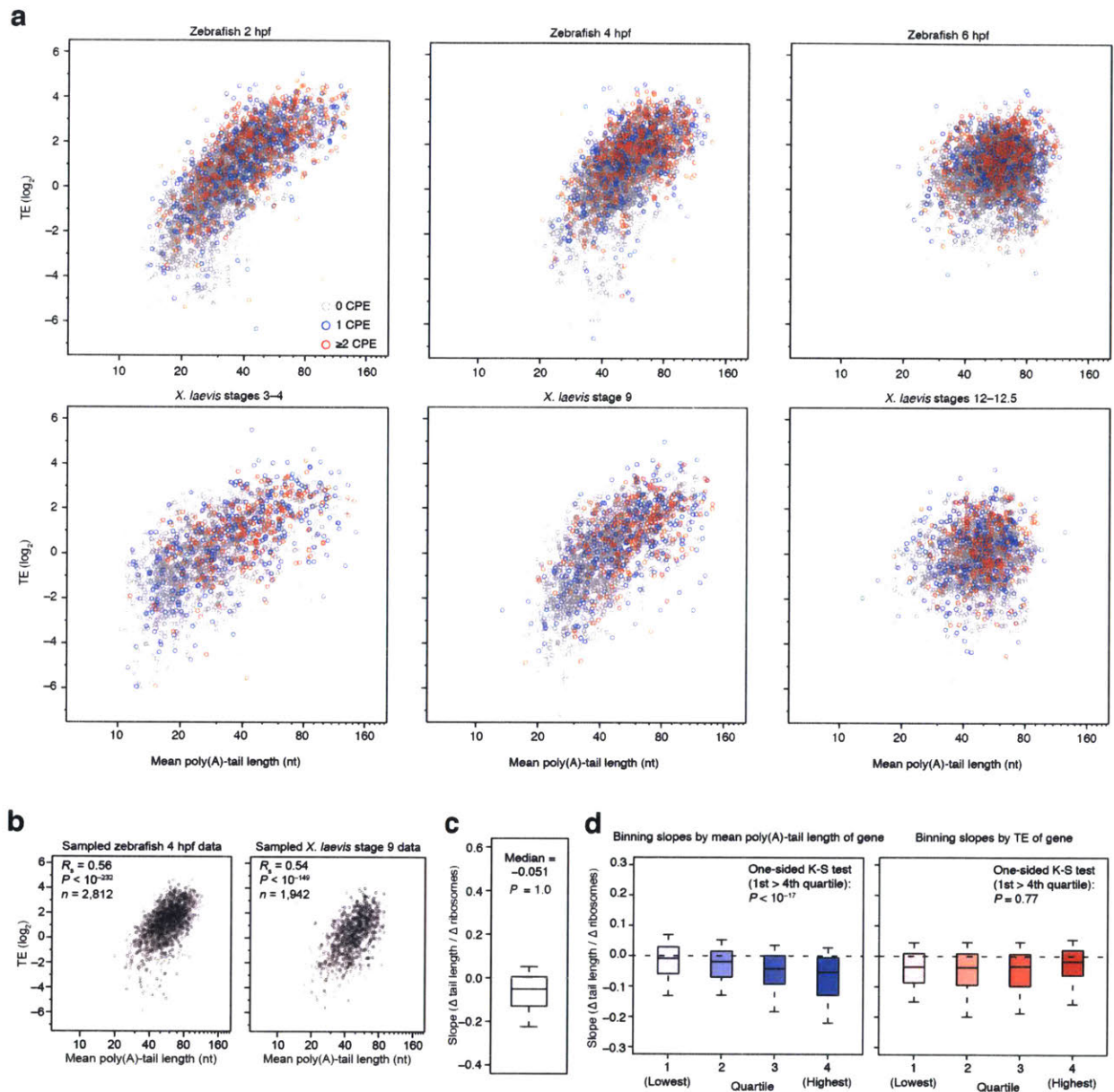


Extended Data Figure 5. Relationship between poly(A)-tail length and changes in gene expression during zebrafish embryogenesis. Changes in gene expression between the indicated embryonic stages, as measured by mRNA-seq, are plotted as a function of the mean poly(A)-tail length at the latter stage.

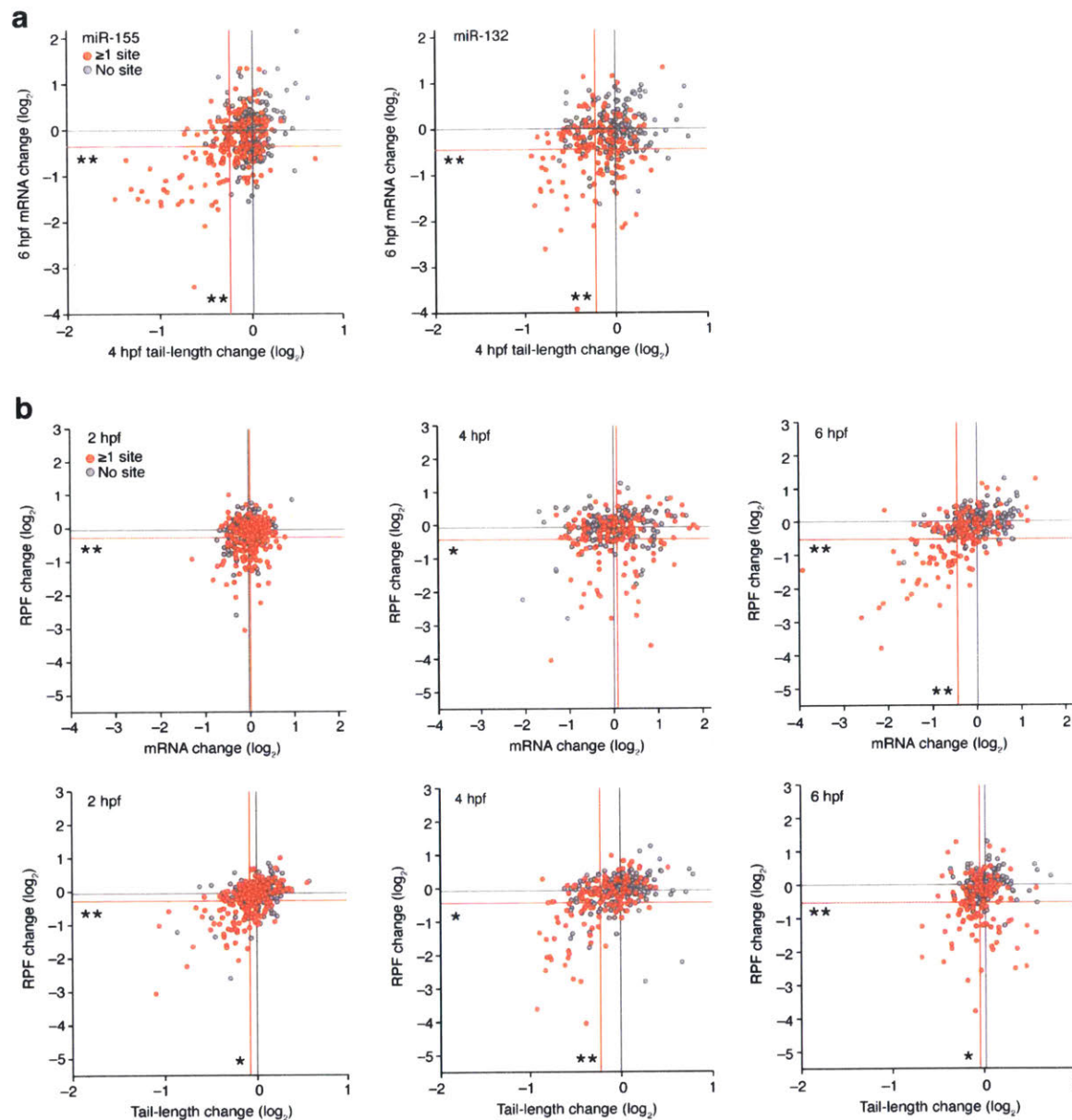


Extended Data Figure 6. Poly(A)-tail lengths of tandem alternative 3' UTR isoforms. **a**, Comparison of average poly(A)-tail lengths for proximal (short) and distal (long) isoforms in the indicated cell lines. Results are plotted for isoforms that were each represented by ≥ 25 poly(A) tags and had alternative poly(A) sites ≥ 500 nt apart. For genes with more than one isoform pair meeting these criteria, the pair with poly(A) sites furthest apart was selected. In the scatterplot for HeLa, points for NM_001007026 and NM_003913 fell outside the boundaries of the plot. P values, χ^2 test evaluating whether the relationship between isoform length and tail length differs from that expected by chance. **b**, Average poly(A)-tail lengths for proximal (short) and distal (long) 3' UTR isoforms in 2 hpf zebrafish embryos, comparing results for genes that either contain (red), or do not contain (blue), a cytoplasmic polyadenylation element (CPE) anywhere within the region unique to the distal isoform. A CPE was defined as U_{12} with a single non-U anywhere within the 12 nucleotides. For a CPE found in the unique region to be counted as present, a canonical poly(A) signal (AAUAAA) also had to exist in the last 30 nt of the distal isoform (Simon et al., 1992; Aanes et al., 2011). For each gene with a CPE within the region unique to the distal isoforms, five genes with unique distal regions of comparable length ($\pm 10\%$) but lacking a CPE are also shown. Poly(A) tags from three zebrafish 2 hpf PAL-seq libraries

(mock-, miR-132-, and miR-155-injected) were combined prior to calculating average tail length for each isoform. Tandem isoform pairs with a target site for miR-132 or miR-155 in the region unique to the distal isoform were not considered. Only genes for which both tandem isoforms had ≥ 25 poly(A) tags, and for which the alternative poly(A) sites were 50–500 nt apart, are plotted. For genes for which isoform choice affected inclusion of a CPE, the isoform pair representing that gene was chosen as the two isoforms with the most 5'-proximal poly(A) sites that flanked a CPE and satisfied the above criteria. For the pool of genes from which controls were chosen, two adjacent isoforms were picked randomly. *P*-value, Fisher's exact test, comparing genes with a CPE in the unique region to the controls.



Extended Data Figure 7. Relationship between poly(A)-tail length and translational efficiency, classifying genes based on CPE content, tail length, or translational efficiency. **a**, The same data as in Fig. 3a, except genes were classified based on whether their 3' UTR contained no CPE (gray), one CPE (blue), or two or more non-overlapping CPE's (red). **b**, Box plot as in Fig. 4b for the same set of genes, with slopes calculated using tail lengths from all but the zero-ribosome bound fraction. **c**, Box plots as in Fig. 4b, creating four equal bins of genes based on either overall mean poly(A)-tail length (left) or TE (right). The same slopes were used as in Fig. 4b, but only considering genes containing a TE value and ≥ 100 poly(A) tags in the actively dividing 3T3 sample.



Extended Data Figure 8. The influence of miRNAs on ribosomes, mRNA and tails in the early zebrafish embryo. **a**, The relationship between changes in tail length at 4 hpf (as determined by PAL-seq) and changes in mRNA abundance at 6 hpf (as determined by RNA-seq), after injecting miR-155 (left) or miR-132 (right). Changes observed between injected and mock-injected embryos are plotted for predicted miRNA target genes (red, genes with ≥ 1 cognate miRNA site in their 3' UTR) and control genes (gray, genes that have no cognate miRNA site yet resemble the targets with respect to 3' UTR length). Lines indicate mean changes for the respective gene sets; statistically significant differences between the gene sets for each of the two parameters are indicated (*, $P \leq 0.05$; **, $P < 10^{-4}$, one-tailed Kolmogorov-Smirnov test). Because injected miRNAs partially inhibited miR-430-mediated repression, genes with a site complementary to nucleotides 2–7 of miR-430 were not considered. All data were normalized to the median

changes observed for the controls. **b**, The relationship between changes in ribosome protected fragments (RPFs) and changes in mRNA levels (top), and between RPFs and changes in tail length (bottom) after injecting miR-132. At 2, 4 and 6 hpf, embryos were analyzed using ribosome profiling (which detects RPFs), mRNA-seq, and PAL-seq. Plots are as in Fig. 5a.

Tail length conservation				mRNA length			
Samples	<i>n</i>	<i>R_s</i>	<i>P</i> value	Sample	<i>n</i>	<i>R_s</i>	<i>P</i> value
HeLa to HEK293T	1620	0.74	< 10 ⁻²⁷⁶	<i>S. cerevisiae</i>	3265	0.039	0.025
HeLa to 3T3	1259	0.46	< 10 ⁻⁶⁷	<i>S. pombe</i>	2911	0.12	< 10 ⁻¹⁰
HeLa to mouse liver	1095	0.21	< 10 ⁻¹¹	<i>Arabidopsis</i> leaf	1425	0.30	< 10 ⁻⁶⁹
HeLa to S2	1087	0.16	< 10 ⁻⁶	<i>Drosophila</i> S2	3488	0.31	< 10 ⁻⁷⁷
HeLa to <i>S. cerevisiae</i>	671	0.056	0.14	HeLa	2362	0.32	< 10 ⁻⁹⁸
HEK293T to 3T3	1907	0.40	< 10 ⁻⁷⁴	Mouse liver	3415	-0.12	< 10 ⁻¹¹
HEK293T to mouse liver	1815	0.22	< 10 ⁻²⁰	HEK293T	4773	0.31	< 10 ⁻¹⁰⁶
HEK293T to S2	1877	0.16	< 10 ⁻¹¹	3T3	2873	0.36	< 10 ⁻⁶⁷
HEK293T to <i>S. cerevisiae</i>	1221	0.078	0.0063	Zebrafish 2 hpf	6749	0.19	< 10 ⁻⁹²
3T3 to mouse liver	1548	0.37	< 10 ⁻⁵¹	Zebrafish 4 hpf	5692	0.068	< 10 ⁻⁶
3T3 to S2	1194	0.20	< 10 ⁻¹¹	Zebrafish 6 hpf	4594	0.013	0.37
3T3 to <i>S. cerevisiae</i>	737	0.11	0.0034	Splice-site number			
mouse liver to S2	1238	-0.068	0.016	Sample	<i>n</i>	<i>R_s</i>	<i>P</i> value
mouse liver to <i>S. cerevisiae</i>	784	0.0028	0.94	<i>S. cerevisiae</i>	3265	-0.093	< 10 ⁻⁴
S2 to <i>S. cerevisiae</i>	959	0.094	0.0038	<i>S. pombe</i>	2911	0.021	0.25
<i>S. pombe</i> to <i>S. cerevisiae</i>	1379	0.22	< 10 ⁻¹⁵	<i>Arabidopsis</i> leaf	1425	0.39	< 10 ⁻⁵¹
mRNA expression level				<i>Drosophila</i> S2	3488	0.30	< 10 ⁻⁷³
Sample	<i>n</i>	<i>R_s</i>	<i>P</i> value	HeLa	2362	0.22	< 10 ⁻²⁷
<i>S. cerevisiae</i>	3199	-0.44	< 10 ⁻¹⁴⁸	Mouse liver	3415	0.0097	0.57
<i>S. pombe</i>	2791	-0.31	< 10 ⁻⁶¹	HEK293T	4773	0.23	< 10 ⁻⁹⁵
HeLa ³⁵	2266	-0.053	0.012	3T3	2873	0.31	< 10 ⁻⁸⁵
Mouse liver	2484	0.085	2.2 x 10 ⁻⁵	Zebrafish 2 hpf	6749	0.12	< 10 ⁻²⁹
HEK293T	4509	-0.23	< 10 ⁻⁵³	Zebrafish 4 hpf	5692	0.066	< 10 ⁻⁶
3T3	2751	-0.23	< 10 ⁻³⁶	Zebrafish 6 hpf	4594	0.077	< 10 ⁻⁶
Zebrafish 2 hpf	3693	-0.13	< 10 ⁻¹⁵	Splice-site density			
Zebrafish 4 hpf	3413	-0.011	0.53	Sample	<i>n</i>	<i>R_s</i>	<i>P</i> value
Zebrafish 6 hpf	3112	0.25	< 10 ⁻⁴³	<i>S. cerevisiae</i>	3265	-0.041	0.018
3'-UTR length				<i>S. pombe</i>	2911	-0.074	6.0 x 10 ⁻⁵
Sample	<i>n</i>	<i>R_s</i>	<i>P</i> value	<i>Arabidopsis</i> leaf	1425	0.24	< 10 ⁻¹⁸
<i>S. cerevisiae</i>	3265	0.063	0.00033	<i>Drosophila</i> S2	3488	-0.0021	0.90
<i>S. pombe</i>	2911	0.24	< 10 ⁻³⁷	HeLa	2362	-0.098	1.8 x 10 ⁻⁶
<i>Arabidopsis</i> leaf	1425	0.035	0.18	Mouse liver	3415	0.13	< 10 ⁻¹³
<i>Drosophila</i> S2	3488	0.28	< 10 ⁻⁶¹	HEK293T	4773	-0.079	< 10 ⁻⁷
HeLa	2362	0.23	< 10 ⁻²⁸	3T3	2873	-0.055	0.0030
Mouse liver	3415	-0.15	< 10 ⁻¹⁸	Zebrafish 2 hpf	6749	-0.062	< 10 ⁻⁶
HEK293T	4773	0.24	< 10 ⁻⁶²	Zebrafish 4 hpf	5692	0.0070	0.60
3T3	2873	0.24	< 10 ⁻³⁶	Zebrafish 6 hpf	4594	0.086	< 10 ⁻⁸
Zebrafish 2 hpf	6749	0.17	< 10 ⁻⁴¹	mRNA nuclear-to-cytoplasmic ratio			
Zebrafish 4 hpf	5692	0.084	< 10 ⁻⁹	Sample	<i>n</i>	<i>R_s</i>	<i>P</i> value
Zebrafish 6 hpf	4594	-0.035	0.016	HeLa ⁴⁷	2340	0.28	< 10 ⁻⁴³
ORF length				mRNA half-life			
Sample	<i>n</i>	<i>R_s</i>	<i>P</i> value	Sample	<i>n</i>	<i>R_s</i>	<i>P</i> value
<i>S. cerevisiae</i>	3265	0.022	0.21	<i>S. cerevisiae</i> ⁴⁸	3027	0.048	0.0079
<i>S. pombe</i>	2911	0.031	0.093	<i>S. cerevisiae</i> ⁴⁹	2592	-0.094	1.5 x 10 ⁻⁶
<i>Arabidopsis</i> leaf	1425	0.28	< 10 ⁻²⁷	<i>S. cerevisiae</i> ⁵⁰	1802	0.044	0.063
<i>Drosophila</i> S2	3488	0.23	< 10 ⁻⁴²	<i>S. cerevisiae</i> ⁵¹	2320	-0.11	< 10 ⁻⁷
HeLa	2362	0.29	< 10 ⁻⁴⁵	<i>S. cerevisiae</i> ⁵²	3168	0.045	0.011
Mouse liver	3415	-0.016	0.34	<i>S. cerevisiae</i> ⁵³	3256	-0.44	< 10 ⁻¹⁴⁹
HEK293T	4773	0.25	< 10 ⁻⁴⁸	<i>S. cerevisiae</i> ⁵⁴	1096	0.23	< 10 ⁻¹³
3T3	2873	0.36	< 10 ⁻⁹⁰	<i>S. cerevisiae</i> ⁵⁵	2824	-0.35	< 10 ⁻⁸¹
Zebrafish 2 hpf	6749	0.14	< 10 ⁻³⁸	HeLa ⁵⁶	642	-0.048	0.23
Zebrafish 4 hpf	5692	0.021	0.11	3T3 ⁵¹	1780	-0.16	< 10 ⁻¹¹
Zebrafish 6 hpf	4594	0.029	0.046				

Extended Data Table 1. Relationships between poly(A)-tail lengths of orthologous genes in samples from different species (or the same gene, when the samples are from the same species), and poly(A)-tail length and other mRNA features, including 3' UTR length, ORF length, mRNA length, splice-site number, splice-site density, mRNA abundance, nuclear-to-cytoplasmic abundance ratio, and half-life. When calculating splice-site density, a pseudocount of one was added to the number of splice sites in an mRNA. For the comparisons between poly(A)-tail length and expression level, mRNA abundances were measured by mRNA-Seq; data from HeLa

were from (Guo et al., 2010). For the relationship between poly(A)-tail length and mRNA nuclear-to-cytoplasmic abundance ratio, measurements of nuclear and cytoplasmic mRNA abundance in HeLa cells were from (Djebali et al., 2012). mRNA half-lives for budding yeast, HeLa and 3T3 mRNAs were from (Holstege et al., 1998; Wang et al., 2002; Grigull et al., 2004; Shalem et al., 2008; Munchel et al., 2011; Sun et al., 2012; Haimovich et al., 2013; Sun et al., 2013), (Larsson et al., 2010) and (Schwanhäusser et al., 2011), respectively.

<i>S. cerevisiae</i>	HEK293T
cytosolic small ribosomal subunit (-4.358, 0)	structural constituent of ribosome (-3.26, 0)
cytosolic large ribosomal subunit (-4.1807, 0)	cytosolic large ribosomal subunit (-2.4247, 0.0118)
oxidoreductase activity (-3.1264, 0)	proteasome complex (-2.2871, 0.0263)
actin binding (-2.6333, 0.0013)	epidermal growth factor receptor signaling pathway (2.3589, 0.0417)
hydrogen ion transmembrane transporter activity (-2.6072, 0.0014)	clathrin coat assembly (2.3693, 0.0475)
protein refolding (-2.5312, 0.0028)	branched chain family amino acid catabolic process (2.3775, 0.0499)
glycolysis (-2.4683, 0.0047)	cell-cell junction (2.7562, 0.0065)
microsome (-2.347, 0.0106)	
proteasome complex (-2.3188, 0.0126)	3T3
aminoacyl-tRNA ligase activity (-2.1318, 0.0379)	cytosolic large ribosomal subunit (-3.7571, 0)
mating projection tip (-2.1125, 0.0427)	proteasome complex (-2.9927, 0.0001)
ATP biosynthetic process (-2.1002, 0.0461)	electron transport chain (-2.6717, 0.0016)
kinetochore (2.28, 0.0369)	translation initiation factor activity (-2.4636, 0.0052)
RNA Pol II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription (2.6171, 0.0021)	cytosolic small ribosomal subunit (-2.3661, 0.0098)
transmembrane transport (2.7216, 0.0007)	transmembrane receptor protein tyrosine kinase signaling pathway (2.1827, 0.0391)
sequence-specific DNA binding (3.2069, 0)	transcription, DNA-dependent (2.2119, 0.0352)
	protein serine/threonine kinase activity (2.2641, 0.0296)
	proteinaceous extracellular matrix (2.3521, 0.0178)
	collagen (2.6353, 0.0048)
<i>S. pombe</i>	Liver
cytosolic large ribosomal subunit (-4.8841, 0)	cellular response to oxidative stress (2.557, 0.0119)
cytosolic small ribosomal subunit (-3.9765, 0)	proton transport (2.5639, 0.0163)
cytoplasmic translational elongation (-3.5414, 0)	
glycolysis (-2.5283, 0.0027)	zebrafish 2 hpf
translation initiation factor activity (-2.3317, 0.0121)	ribosome (-3.3771, 0)
RNA Pol II core promoter proximal region sequence-specific DNA binding (2.8398, 0)	ATP synthesis coupled proton transport (-3.1102, 0)
	ATP hydrolysis coupled proton transport (-3.0592, 0)
	glycolysis (-2.611, 0.0012)
	proteasome core complex (-2.3519, 0.0097)
	cytochrome-c oxidase activity (-2.2381, 0.0186)
	lipid binding (-2.0649, 0.0492)
	convergent extension involved in gastrulation (2.1808, 0.0396)
	ATP-dependent helicase activity (2.2225, 0.0344)
	double-stranded RNA binding (2.2407, 0.0352)
	RNA splicing (2.4446, 0.0101)
	protein kinase activity (2.5718, 0.0049)
	zebrafish 4 hpf
	ribosome (-3.2965, 0)
	ATP hydrolysis coupled proton transport (-2.5236, 0.0044)
	spindle pole (-2.3207, 0.0171)
	ATP synthesis coupled proton transport (-2.2413, 0.0297)
	glycolysis (-2.1842, 0.0394)
	fin regeneration (2.1141, 0.0497)
	convergent extension involved in axis elongation (2.1437, 0.0425)
	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (2.1683, 0.0383)
	cell-cell adherens junction (2.1891, 0.0351)
	double-stranded RNA binding (2.2892, 0.0186)
	ATP-dependent helicase activity (2.4816, 0.0042)
	nuclear pore (2.6064, 0.0025)
	RNA splicing (2.6176, 0.0025)
	zebrafish 6 hpf
	cell redox homeostasis (-2.2444, 0.0393)
	transferase activity, transferring glycosyl groups (-2.1834, 0.0434)
	vesicle-mediated transport (-2.132, 0.0489)
	ATP-dependent helicase activity (2.2093, 0.0187)
	DNA-dependent ATPase activity (2.3181, 0.0094)
	cell adhesion (2.3641, 0.0065)
	regulation of cell cycle (2.4049, 0.0057)
	proteasome complex (2.4547, 0.0042)
	nucleosomal DNA binding (2.4853, 0.0042)
	structural constituent of ribosome (4.8594, 0)
<i>Arabidopsis leaf</i>	
structural constituent of ribosome (-2.4958, 0.0188)	
lipid binding (-2.4453, 0.0143)	
cysteine biosynthetic process (-2.2569, 0.0341)	
rRNA modification (-2.2477, 0.0253)	
response to absence of light (2.2135, 0.0279)	
response to auxin stimulus (2.2338, 0.0263)	
abscisic acid mediated signaling pathway (2.2619, 0.0249)	
endoplasmic reticulum unfolded protein response (2.2861, 0.0246)	
negative regulation of programmed cell death (2.601, 0.0015)	
response to chitin (2.6803, 0.0006)	
response to mechanical stimulus (2.9409, 0)	
<i>Drosophila S2</i>	
mitochondrial large ribosomal subunit (-4.7635, 0)	
cytosolic small ribosomal subunit (-4.0894, 0)	
cytosolic large ribosomal subunit (-3.7405, 0)	
mitochondrial small ribosomal subunit (-3.1553, 0)	
mitochondrial respiratory chain complex I (-2.9286, 0.0001)	
ATP synthesis coupled proton transport (-2.5931, 0.001)	
pyruvate metabolic process (-2.2117, 0.0024)	
ubiquinol-cytochrome-c reductase activity (-2.1951, 0.026)	
tRNA binding (-2.1493, 0.0346)	
small nuclear ribonucleoprotein complex (-2.1231, 0.038)	
proteasome regulatory particle, lid subcomplex (-2.0899, 0.0435)	
proteinaceous extracellular matrix (2.5315, 0.0175)	
HeLa	
cytosolic large ribosomal subunit (-2.7068, 0.0004)	
mitochondrial ribosome (-2.8985, 0.0006)	
chaperonin-containing T-complex (-2.2595, 0.0143)	
proteasome complex (-2.2316, 0.0166)	
exosome (RNase complex) (-2.2024, 0.0199)	
nitric oxide metabolic process (-2.0896, 0.0399)	
cell-cell junction (2.1558, 0.0471)	
positive regulation of transcription from RNA Pol II promoter (2.205, 0.0377)	
extracellular matrix (2.242, 0.0351)	
glycosphingolipid metabolic process (2.2986, 0.025)	
melanosome (2.5133, 0.0045)	
antigen processing and presentation of exogenous peptide antigen via MHC class II (2.5749, 0.0039)	
lysosome (2.9548, 0)	

Extended Data Table 2. Gene ontology (GO) categories enriched in shorter- or longer-tail genes, as determined by gene set enrichment analysis (GSEA) (Subramanian et al., 2005). For each sample, GSEA was performed on genes ranked based on their mean poly(A)-tail length. The normalized enrichment score (NES) and false-discovery rate (Q value) are indicated in

parentheses next to each enriched GO category. A negative NES indicates a category enriched in the shorter-tail genes, whereas a positive value indicates a category enriched in the longer-tail genes. Enriched GO categories were manually curated to eliminate redundant or uninformative categories. For mouse liver no GO categories were significantly enriched ($Q < 0.05$) in either the shorter- or longer-tail genes.

Chapter 3

mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues

Stephen W. Eichhorn^{1,2,3*}, Huili Guo^{1,2,3,4,5,6*}, Sean E. McGeary^{1,2,3}, Ricard A. Rodriguez-Mias⁷, Chanseok Shin^{1,2,8}, Daehyun Baek^{1,2,9,10,11}, Shu-hao Hsu¹², Kalkpana Ghoshal¹², Judit Villén⁷, David P. Bartel^{1,2,3}

¹Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

³Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁴Institute of Molecular and Cell Biology, Singapore 138673, Singapore

⁵Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore

⁶Lee Kong Chian School of Medicine, Nanyang Technological University-Imperial College, Singapore 639798, Singapore

⁷Department of Genome Sciences, University of Washington, Seattle, WA 98105, USA

⁸Department of Agricultural Biotechnology, Plant Genomics and Breeding Institute, Seoul National University, Seoul 151-921, Republic of Korea

⁹Center for RNA Research, Institute for Basic Science, Seoul 151-747, Republic of Korea

¹⁰School of Biological Sciences, Seoul National University, Seoul 151-747, Republic of Korea

¹¹Bioinformatics Institute, Seoul National University, Seoul 151-742, Republic of Korea

¹²Department of Pathology, Ohio State University, Columbus, OH 43210, USA

*These authors contributed equally to this work

S.W.E., H.G., and D.P.B. designed the study. S.W.E. and H.G. performed ribosome profiling and RNA-seq, and S.W.E. did the associated analyses. S.E.M. did the mathematical modeling. R.A.R.-M. and J.V. performed the proteomics. C.S. cultured primary cells. D.B. analyzed microarray and proteomics data. S.-h.H. and K.G. harvested liver tissue. S.W.E., H.G., S.E.M., and D.P.B., wrote the paper, with input from the other authors.

Published as:

Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S-h., Ghoshal, K., Villén, J., Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell* 56, 104-115.

Abstract

MicroRNAs (miRNAs) regulate target mRNAs through a combination of translational repression and mRNA destabilization, with mRNA destabilization dominating at steady state in the few contexts examined globally. Here, we extend the global steady-state measurements to additional mammalian contexts and find that regardless of the miRNA, cell type, growth condition or translational state, mRNA destabilization explains most (66% to >90%) miRNA-mediated repression. We also determine the relative dynamics of translational repression and mRNA destabilization for endogenous mRNAs as a miRNA is induced. Although translational repression occurs rapidly, its effect is relatively weak, such that by the time consequential repression ensues, the effect of mRNA destabilization dominates. These results imply that consequential miRNA-mediated repression is largely irreversible and provide other insights into the nature of miRNA-mediated regulation. They also simplify future studies, dramatically extending the known contexts and time points for which monitoring mRNA changes captures most of the direct miRNA effects.

Introduction

MicroRNAs are small, non-coding RNAs that post-transcriptionally regulate the expression of most mammalian genes (Bartel, 2009; Friedman et al., 2009). Acting as the specificity components of ribonucleoprotein silencing complexes, miRNAs pair with target mRNAs at sites complementary to the miRNA 5' region. Most effective sites map to 3' untranslated regions (3' UTRs) and pair perfectly with the miRNA seed (nucleotides 2–7), with an additional pair at nucleotide 8 and/or an A across from nucleotide 1 (Bartel, 2009).

Although early reports of gene regulation by miRNAs emphasized their role as translational repressors (Wightman et al., 1993; Olsen and Ambros, 1999; Seggerson et al., 2002), subsequent studies revealed that miRNAs can also induce mRNA degradation (Bagga et al., 2005; Krutzfeldt et al., 2005; Lim et al., 2005). This degradation is a consequence of miRNA-mediated deadenylation of target mRNAs (Behm-Ansmant et al., 2006; Giraldez et al., 2006; Wu et al., 2006), which causes these mRNAs to undergo decapping and then 5'–3' decay (Rehwinkel et al., 2005; Behm-Ansmant et al., 2006; Chen et al., 2009). The discovery of this second mode of repression raised the question as to the relative contributions of translational repression and mRNA degradation to reducing the protein abundance of regulated genes.

Large-scale analyses comparing protein and mRNA changes of predicted miRNA targets after introducing or deleting individual mammalian miRNAs found that protein changes generally correspond to changes in polyadenylated mRNA abundance (Baek et al., 2008). More precise measurements comparing changes in translational efficiency (TE) to changes in mRNA again found that mRNA degradation explains the majority of miRNA-mediated repression, with translational repression contributing roughly 10–25% of the overall repression (Hendrickson et al., 2009; Guo et al., 2010). These global measurements of TE and mRNA (or protein and mRNA) were made at relatively late time points (12–32 hours after introducing the miRNA or long after induction of an endogenous miRNA), and thus are thought to reflect the steady-state effects of the miRNA (Baek et al., 2008; Hendrickson et al., 2009; Guo et al., 2010). When miRNAs are expressed at constant levels, steady-state measurements are ideal for quantifying the relative contributions of translational repression and mRNA degradation because they integrate effects occurring throughout the life cycle of each targeted transcript.

If generalizable to other cell types and conditions, these high-throughput steady-state measurements, which indicate that mRNA changes closely approximate the overall effects of a miRNA on target gene expression, would be welcome news for those placing mammalian miRNAs into gene-regulatory networks and quantifying their impact on gene expression, since measuring changes in mRNA levels is much easier than measuring changes in protein levels or TE. However, protein/TE and mRNA effects have been globally compared in only two cell lines, HeLa cells (Baek et al., 2008; Selbach et al., 2008; Guo et al., 2010) and HEK293T cells (Hendrickson et al., 2009), and a single primary cell type, mouse neutrophils (Baek et al., 2008; Guo et al., 2010), which leaves open the possibility that translational repression might dominate in most other mammalian contexts.

The observation that mRNA destabilization can account for most repression at steady state has prompted a search for time points in which translational repression might explain a larger proportion of the repression. Two studies examined the dynamics of miRNA-mediated repression on inducible reporter genes as these genes begin to be expressed in fly and human cells (Bethune et al., 2012; Djuranovic et al., 2012), and another examined the effects of miR-430 on its endogenous targets in the zebrafish embryo (Bazzini et al., 2012). In blastula-stage zebrafish embryos (4 hours postfertilization [hpf]), miR-430 substantially reduces the TE of its targets with little effect on their stability, whereas by gastrulation (6 hpf), the relative contributions of TE and mRNA destabilization closely resemble those observed previously at steady state in mammalian systems (Baek et al., 2008; Hendrickson et al., 2009; Guo et al., 2010; Bazzini et al., 2012). Because miR-430 is strongly induced shortly before the blastula stage, the large amount of translational repression observed in the blastula stage, followed by the mRNA

destabilization observed later in the gastrula stage was proposed to reflect the fundamental dynamics of miRNA-mediated repression (Bazzini et al., 2012).

The idea that miRNA-mediated translational repression precedes mRNA degradation cannot be disputed—an mRNA molecule can undergo translational repression only before it has been degraded and thus its translational regulation must precede regulation at the level of its stability in the same way that transcriptional regulation must precede translational regulation. However, subsequent insight into the shift in regulatory regime occurring as zebrafish embryos progress from pre- to post-gastrulation has overturned the idea that the miR-430 observations reflect the dynamics of miRNA-mediated repression (Subtelny et al., 2014). Prior to gastrulation, mRNA poly(A)-tail length and TE are coupled, and short-tailed mRNAs are stable. These two unique conditions enable miRNA-mediated deadenylation to cause translational repression without mRNA destabilization (Subtelny et al., 2014). The transition to mostly mRNA decay is due to a change in these conditions at gastrulation, such that coupling between tail length and TE is lost and short-tailed mRNAs become less stable, which causes the consequence of miRNA-mediated deadenylation to shift from translational repression to mRNA destabilization (Subtelny et al., 2014). When considering this shifting regulatory regime, the miR-430 results do not provide insight into the dynamics of the two modes of miRNA-mediated repression for endogenous mRNAs, nor do they demonstrate that miRNA-mediated translational repression occurring through a deadenylation-independent mechanism ever mediates meaningful changes in the expression of endogenous mRNAs. This being said, the miR-430 study is notable as the first (and only) to identify an endogenous setting in which the effects of a miRNA cannot be approximated by changes in mRNA levels (Bazzini et al., 2012). Because of the regulatory regime operating in the pre-gastrulation zebrafish embryo (and presumably in other early

embryos or other unusual settings, such as neuronal synapses), measuring mRNA changes misses essentially all of the effects of miRNAs in this setting (Subtelny et al., 2014).

The two studies that monitor reporter genes rather than endogenous transcripts to examine miRNA repression dynamics both report that a phase of substantial translational repression occurs prior to detectable mRNA deadenylation or decay (Bethune et al., 2012; Djuranovic et al., 2012). However, the updated understanding of the miR-430 results reopens the question of whether such a phase also occurs for endogenous mRNAs. Although reporters can faithfully represent endogenous genes, several observations led us to suspect that when measuring the effects of miRNAs there might be a difference between reporters and endogenous genes. First, even at very early time points in zebrafish embryonic development, most repression of endogenous mRNAs is attributable to miRNA-mediated deadenylation rather than direct translational repression (Subtelny et al., 2014). Second, at steady state, the fractional repression attributed to translational repression of the reporters (Bethune et al., 2012; Djuranovic et al., 2012) exceeds that typically observed for endogenous mRNAs in mammalian cells (Baek et al., 2008; Hendrickson et al., 2009; Guo et al., 2010). Similarly, the magnitude of repression observed for reporters vastly exceeds that typically observed for endogenous mRNAs in mammalian cells.

Here we substantially expand the contexts and conditions for which the repressive effects on endogenous mRNAs are examined. We measured the consequence of deleting specific miRNAs on the mRNA and translation (or protein) of predicted targets in mouse liver, primary macrophages, and activated and non-activated primary B cells, thereby adding four additional biological settings to the previous two settings (mouse neutrophils and zebrafish embryos) in which translational effects on endogenous targets have been broadly measured. We also

measured the translational effects on endogenous mRNAs after adding specific miRNAs in two additional cell lines (U2OS cells and NIH3T3 cells) and two additional conditions (growth-arrested cells and translationally inhibited cells). In all cases, mammalian miRNAs predominantly acted to decrease target mRNA levels, with relatively small contributions from translational repression. We then examined the repression dynamics of endogenous mRNAs and did not observe an early phase in which dominant translation effects imparted substantial repression. We conclude that although translational repression is rapid, its effect is relatively weak, and thus by the time consequential repression ensues, the effect of mRNA destabilization dominates.

Results

Negligible contribution of nuclear or deadenylated RNA to TE values

The adaptation of ribosome profiling to mammalian cells has provided new opportunities to quantitatively assess the influence of miRNAs on TE (Guo et al., 2010). Ribosome profiling uses high-throughput sequencing of ribosome-protected fragments (RPFs) to determine the positions of millions of ribosomes on mRNAs (Ingolia et al., 2009). To assess the TE of a gene, RPFs mapping to its open reading frame are normalized to its mRNA abundance, as determined by RNA-seq.

When comparing samples with and without a particular miRNA, the change in RPFs for a target of that miRNA reflects the aggregate effects of mRNA degradation and translational repression, while the change in mRNA reflects only the component attributable to degradation. After accounting for the change in RPFs attributed to mRNA degradation, the residual change in

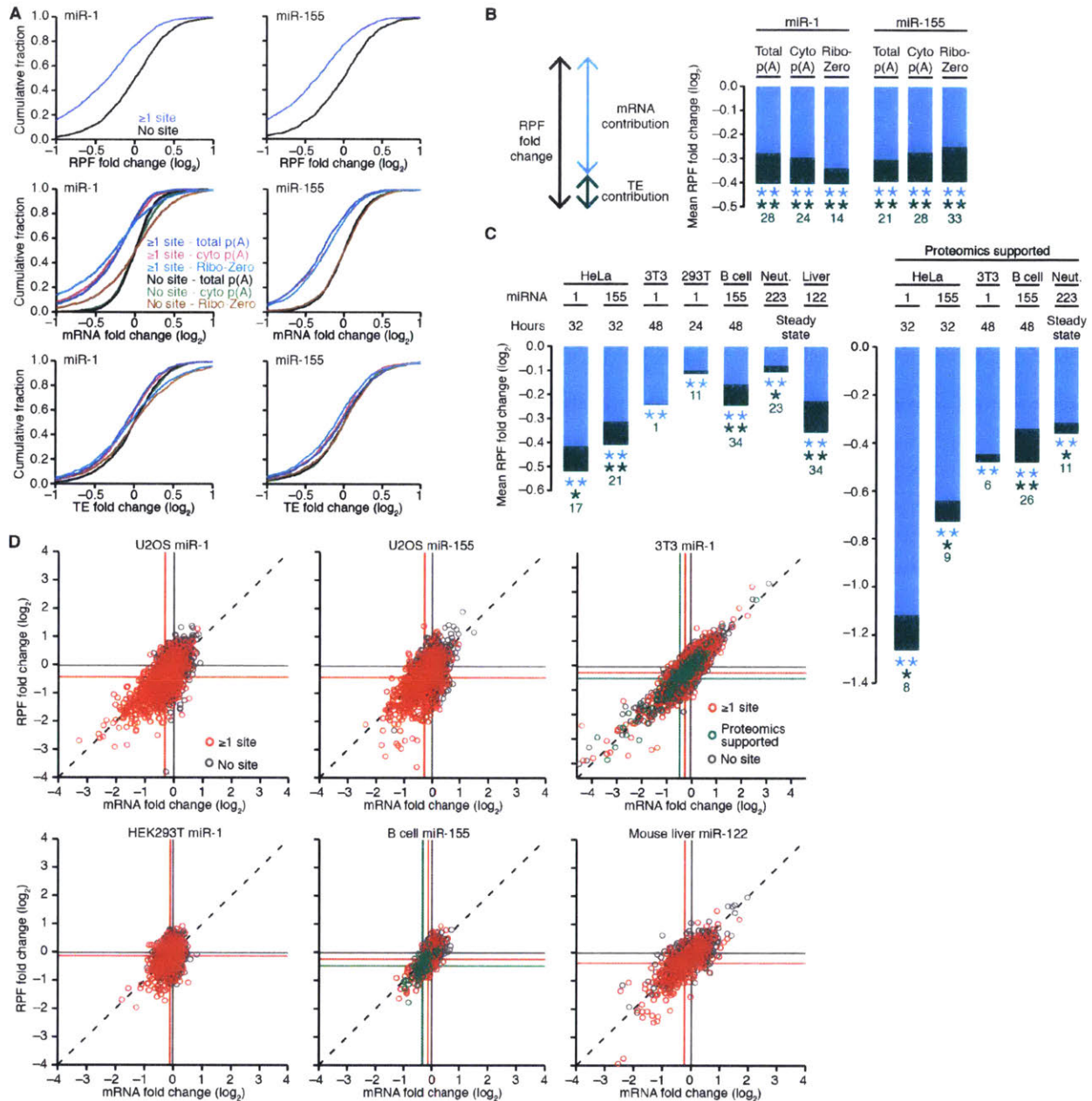


Figure 1. Steady-State Changes in Gene Expression Due to miRNAs

(A) The influence of using different types of mRNA enrichment when measuring the effects of miRNAs on mRNA levels and TE. Plots show cumulative distributions of changes in RPFs (top), mRNA (middle), and TE (bottom) after transfection of either miR-1 (left) or miR-155 (right) into U2OS cells. The impact of the miRNA on genes with at least one site to the cognate miRNA in their 3' UTR (≥ 1 site; $n = 1321$ and 1075 for miR-1 and miR-155, respectively) is compared to that of control genes (no site; $n = 1205$ and 1056 , respectively), which were chosen from the genes with no site to the cognate miRNA throughout their entire transcript to match the 3'-UTR-length distributions of site-containing genes. The three types of mRNA enrichment were poly(A)-selected total RNA, poly(A)-selected cytoplasmic RNA, and tRNA/rRNA-depleted total RNA [total p(A), cyto p(A), and Ribo-zero, respectively]. RNA-seq analyses of these preparations were used to calculate mRNA and TE changes, with results plotted as indicated in

the key. Data were normalized to the median changes observed for the controls. See also, Figure S1.

(B) A simplified representation of the results in panel A showing for each experiment the mean RPF fold change (\log_2) attributable to changes in mRNA (blue) and TE (green), after subtracting the mean RPF change of the no-site control genes. The schematic (left) depicts the components of the compound bar graphs (right). Significant changes for each component are indicated with asterisks of the corresponding color [$*$, $P \leq 0.05$; $**$, $P \leq 0.0001$, one-tailed Kolmogorov–Smirnov test (K–S test)], with the percent change attributed to lower TE also shown in green below each bar. See also, Figure S2.

(C) The steady-state effects of miRNAs in a variety of cell types, shown using compound bar graphs like those of panel B. For comparison with our current results, previously published results from HeLa and neutrophils (neut.) (Guo et al., 2010) are also plotted after reanalysis using the current methods (including the method for choosing no-site control cohorts). When available, proteomic-supported targets were also analyzed (right). For HeLa and neutrophil these were the ones selected previously (Guo et al., 2010), and for the other samples these were selected from our proteomics data as the subset of site-containing genes with fold changes (\log_2) ≤ -0.3 in the presence of the miRNA. Experiments with cell lines compared cells with and without the miRNA introduced by either transfection (HeLa and 293T) or induction from a transgene (3T3). Experiments with B cells, neutrophils and liver compared cells/tissues isolated from wild-type and miRNA-knockout mice. The hours indicate the time following transfection (HeLa and 293T), induction (3T3), or activation (B cells). See also, Figure S3, Table S1 and Table S2.

(D) Comparison of mRNA and RPF changes for individual genes analyzed in panels (A)–(C). For U2OS cells, the results for the poly(A)-selected cytoplasmic RNA are shown. The dashed line is for $y = x$; the vertical and horizontal lines indicate the mean fold changes for the correspondingly colored groups of genes. Red, genes with ≥ 1 3'-UTR site to the cognate miRNA; grey, no site to the miRNA selected as in panel A; green, proteomics-supported targets (Table S1 and Table S2). Data were normalized to the median changes observed for the controls. A comparable analysis of the HeLa and neutrophil data has been published (Guo et al., 2010).

RPFs reflects a change in TE, which is interpreted as the miRNA-mediated translational repression acting on the message at the moment the ribosomes were arrested.

Previously, we observed little miRNA-mediated translational repression in mammalian cells, with the concern that these modest TE changes might actually be overestimates (Guo et al., 2010). An overestimation would occur if some polyadenylated mRNA were sequestered away from the compartment containing both miRNAs and ribosomes, as would be the case for mRNAs awaiting export from the nucleus. In this case, miRNA-mediated degradation of mRNAs only in the cytoplasm would lead to a larger relative loss of RPFs (which are only from the cytoplasm)

than mRNA fragments (which are from both the nucleus and cytoplasm), thereby inflating the apparent translational repression. To address this concern, we performed ribosome profiling on miRNA- and mock-transfected U2OS cells, and in parallel performed RNA-seq on poly(A)-selected RNA from both whole-cell lysates and cytoplasmic fractions. The efficacy of fractionation was demonstrated by the depletion of pre-ribosomal RNAs (pre-rRNAs) in the cytoplasmic fraction (Figure S1A). Following transfection of miR-1, a miRNA not normally expressed in U2OS cells, repression was observed, with significant degradation of mRNAs with at least one miR-1 3'-UTR site (Figure 1A). The amount of degradation was indistinguishable in the RNA-seq libraries made with either whole-cell or cytoplasmic mRNA, and thus the amount of translational repression was similarly indistinguishable (Figure 1A). The same was observed with miR-155, another miRNA not normally expressed in U2OS cells, demonstrating that a nuclear mRNA-sequestration artifact does not detectably elevate the signal for miRNA-mediated translational repression in mammalian cells.

A second concern involved the measurement of poly(A)-selected RNA. Monitoring changes in poly(A)-selected RNA leaves unanswered the question of whether repressed mRNAs are degraded or merely deadenylated, and under-recovery of partially deadenylated messages during poly(A)-selection might overestimate the amount of mRNA degradation that has occurred. To address this concern, we generated a third set of RNA-seq libraries from the aforementioned U2OS cells, starting with whole-cell RNA preparations that were not poly(A)-selected and instead were depleted of both tRNAs and rRNAs. Greatly increased RNA-seq coverage of replication-dependent histone mRNAs, which lack poly(A) tails, illustrated our ability to detect RNAs regardless of poly(A)-tail length (Figure S1B). Results for miRNA-dependent changes in tRNA/rRNA-depleted RNA were similar to those of poly(A)-selected

RNA (Figure 1A), which indicated that changes in accumulation of mRNA refractory to poly(A) selection were negligible. These results imply that the absolute amount of deadenylated mRNAs and other intermediates underrepresented in poly(A)-selected RNA is small, even for repressed mRNAs, presumably because these decay intermediates are rapidly decapped and degraded. Thus, concerns that translational repression measurements might have been either under- or over-estimates appear to be unfounded; comparing TEs calculated by simply normalizing RPF changes to those of poly(A)-selected RNA accurately measures translational repression in mammalian cells.

To aid comparisons, the results in Figure 1A can be summarized in compound bar graphs (Figure 1B). For each experiment, the mean RPF fold change (distance that the compound bar extends below zero) indicates the overall repression. The mRNA contribution (blue component of the compound bar) indicates the extent to which mRNA degradation explains this repression, and any residual RPF change is the TE contribution (green component), which reflects the translational repression of the remaining mRNA. Based on the RPF reductions attributable to these two repression modes, their relative contributions to repression are then calculated (Figure S2). Of the two modes, mRNA degradation dominates in U2OS cells (Figure 1B), despite the presence of a P-body subtype reported to impart increased translational repression (Castilla-Llorente et al., 2012).

Dominant mRNA destabilization in many contexts

We expanded our analysis to examine the steady-state effects of gaining or losing a miRNA in additional cell lines and biological contexts. These experiments included studies comparing RPF and mRNA measurements in liver from wild-type mice, which expresses miR-122, to those in

liver from mice lacking the *mir-122* gene. Similarly, the effects of miR-155 in activated primary murine B cells were measured comparing cells from wild-type mice to those lacking the *mir-155* gene. These loss-of-function experiments enabled analysis of endogenous targets in their endogenous settings. The effects on predicted targets of endogenous miR-122 in mouse liver, endogenous miR-155 in primary mouse B cells, induced miR-1 (expressed from a transgene) in 3T3 cells, and transfected miR-1 in HEK293T cells, all resembled the published effects of endogenous miR-223 in neutrophils and transfected miRNAs in HeLa cells (Figure 1C). In all settings, reduced mRNA levels explained most of the steady-state RPF reduction observed in the presence of the miRNA, implying that miRNAs predominantly act to reduce target mRNA levels. Nonetheless, mean RPF reduction attributable to translational repression was observed, ranging from 1–34% of the total, depending on the experiment.

Because a 7–8-nt site to a miRNA is not always sufficient to mediate miRNA targeting, high-throughput proteomic measurements can be used to identify high-confidence targets by identifying site-containing genes with less protein in the presence of the miRNA (Guo et al., 2010). With this in mind, we performed a quantitative proteomics experiment using SILAC (stable isotope labeling with amino acids in culture) to identify a set of genes with reduced protein after inducing miR-1 in 3T3 cells (Table S1) and pulsed SILAC (Selbach et al., 2008) to identify those responding to miR-155 in activated B cells (Table S2). These proteomics-supported targets showed greater mean repression than did the complete set of genes with ≥ 1 site, as expected if they were enriched in direct targets of the miRNA (Figure 2C). For new and published experiments with proteomics-supported targets, the fractional repression attributed to translational repression ranged from 6–26%, somewhat narrower than the range observed when

considering all mRNAs with sites, perhaps because a focus on the more confidently identified targets decreased experimental variability.

Although the amount of repression attributed to translational repression did not always reach statistical significance, our results are consistent with the idea that a small amount of translational repression occurs for each direct target in each context. As was found previously (Baek et al., 2008; Hendrickson et al., 2009; Guo et al., 2010), a gene-by-gene analysis of results from each of the newly examined settings revealed no compelling evidence for a subset of genes repressed at only the translational level (Figure 1D), although the possibility of a few such genes cannot be ruled out.

Matching mRNA and proteomic results for less proficient miRNAs

In pilot experiments aimed at extending our studies to other endogenous contexts, we used wild-type and miRNA-deleted mice to acquire mRNA microarray data for macrophages and neutrophils with and without miR-21, and B cells with and without miR-150. Although these miRNAs were each among the most frequently sequenced miRNAs in the respective wild-type cells (Figure S3A), we observed weak miRNA effects when comparing sets of genes with and without 3'-UTR sites to the cognate miRNA (Figure S3B).

A potential explanation for the weak signals observed by mRNA profiling was that most of the repression was occurring through translational repression rather than mRNA degradation. However, when we used quantitative proteomics to test this possibility, the proteomics results mirrored those of the mRNA arrays, providing no evidence for substantial translational repression (Figure S3B). Thus the weak repression signals observed at the mRNA level for endogenously expressed miR-21 and miR-150 were not due to a discrepancy between mRNA

changes and the overall effects of miRNA-mediated repression. These results add to the growing list of endogenous settings for which mRNA changes accurately represent the effects of miRNA-mediated repression. This list now includes miR-223 in neutrophils (Baek et al., 2008; Guo et al., 2010), miR-21 in macrophages and neutrophils (Figure S3B), miR-122 in liver (Figure 1C), miR-150 in primary B cells (Figure S3B), and miR-155 in activated B cells (Figure 1C).

Dynamics of endogenous mRNA repression by inducible miRNAs

The shifting regulatory regime in the early zebrafish embryo, which changes the consequences of miRNA-mediated poly(A)-tail shortening, confounded the previous attempt to determine the dynamics of the two modes of repression for endogenous messages (Bazzini et al., 2012; Subtelny et al., 2014). Therefore, we set out to characterize the regulatory dynamics of miRNA-mediated repression of endogenous mRNAs and determine if there might be an endogenous setting in which these dynamics could give rise to a phase of substantial translation-dominated repression, as previously observed in reporter experiments (Bethune et al., 2012; Djuranovic et al., 2012).

Perhaps the most dynamic mammalian miRNA is miR-155, which is rapidly and strongly induced in B and T cells upon activation (Thai et al., 2007). In primary murine B cells, we observed a nearly 10-fold increase 4 h after activation with lipopolysaccharide, IL-4, and anti-CD40 (Figure 2A). Although presumably not as strong as for miR-430 in zebrafish embryos [which is expressed from as many as 93 loci (Giraldez et al., 2005)], miR-155 induction was nonetheless stronger than that of other mammalian miRNAs, in that no other mammalian miRNA has been reported to increase so rapidly to a high level of expression.

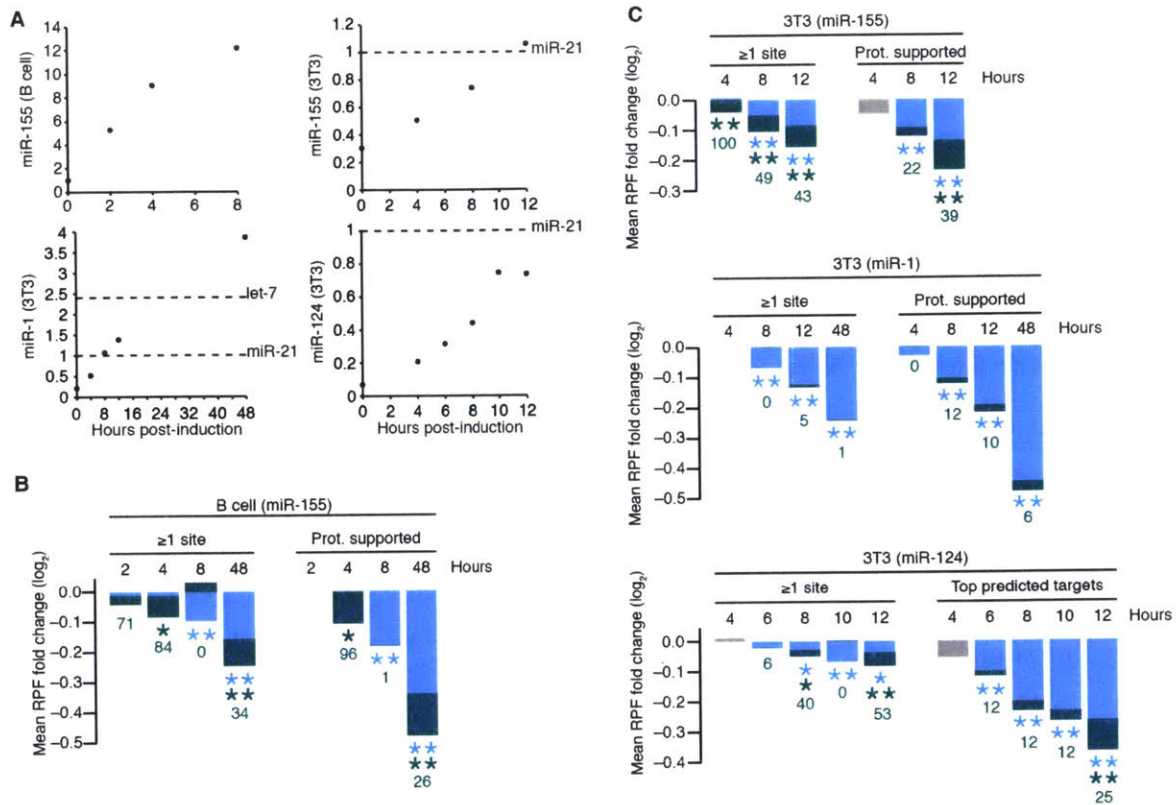


Figure 2. Minor Impact of Translational Repression at all Times in Mammalian Cells

(A) Induction of miRNAs in activated murine B cells and in contact-inhibited NIH3T3 cells engineered to inducibly express miR-1, miR-124, or miR-155. Induction was monitored using RNA blots, probing for the induced miRNA. For samples from B cells, the membrane was reprobed for endogenous U6 snRNA, which served as a loading control for normalization, and expression is plotted relative to that of the non-activated cells. For samples from 3T3 cells, synthetic standards for the induced miRNAs and endogenous miR-21 were included on the blot, and used for absolute quantification. Expression is plotted relative to that of miR-21, with relative expression of the let-7 family (inferred from small-RNA sequencing data) also shown.

(B) The contributions of mRNA decay and translational repression following miR-155 induction in primary murine B cells. The same sets of site-containing and control genes are analyzed in all time points. Otherwise, as in Figure 1C. The 48 h time point is replotted from Figure 1C and was from a preparation of cultured B cells independent from that used for the earlier time points. See also, Table S2.

(C) The contributions of mRNA decay and translational repression following induction of miR-155 (top), miR-1 (middle), or miR-124 (bottom) in the corresponding contact-inhibited 3T3 cell lines. In the absence of proteomics data for miR-124, the top 100 site-containing genes, as ranked by total context+ score (Garcia et al., 2011) regardless of site conservation, were analyzed to focus on a subset of site-containing genes likely to be regulated by miR-124. Otherwise, as in panel B. The miR-1 48 h time point is replotted from Figure 1C and is from the same experiment as the earlier time points. See also, Table S1 and Table S3.

To assess the dynamics of translational repression and mRNA decay during miR-155-mediated repression, we isolated B cells from wild-type and miR-155 knockout mice, activated these cells, and then performed ribosome profiling and RNA-seq to monitor miRNA-dependent TE and mRNA changes occurring soon after induction. At 2 h post-activation, repression of genes with ≥ 1 site was detectable, but neither the mRNA nor TE component was significantly decreased on its own. At 4 h, the small amount of repression was predominantly attributable to reduced TE (Figure 2B). By 8 h, the proportion attributed to translational repression abated, and at this time point the mRNA degradation so closely approached overall repression that the mean mRNA change for genes with ≥ 1 site slightly exceeded the mean RPF change (Figure 2B). Because this slight excess was not observed for the proteomics-supported targets or in similar experiments with other miRNAs, we attribute it to experimental variability rather than evidence of translational activation. After 48 h, mRNA degradation continued to dominate (Figure 2B; as already shown in the steady-state analyses of Figure 1C), which indicated that B cells resemble other cells with respect to steady-state repression.

Although we found some evidence for translational repression dominating early in miR-155 induction, the amount of repression observed during this brief period was much weaker than that observed during the analogous phase of reporter experiments. Thus, we cannot claim to have found a mammalian setting with an early phase of substantial translational repression of endogenous messages, i.e., a time at which substantial repression would be missed if only mRNA changes were monitored. To further explore repression dynamics in mammalian cells, we created stable, miRNA-inducible 3T3 cell lines in which doxycycline treatment rapidly induced the expression of a miRNA not normally expressed in 3T3 cells (either miR-1, miR-124, or miR-155) to levels comparable to those of miR-21 and the let-7 miRNA family (Figure 2A), which

are the miRNA and miRNA family most frequently sequenced for these cells (Rissland et al., 2011). The major advantages of such cell lines for studying the dynamics of translational repression and decay on endogenous messages is that, in contrast to B cells, miRNA induction does not accompany significant developmental changes, allowing the miRNA effects to be more easily isolated. With these lines we performed ribosome profiling and RNA-seq soon after miRNA induction, comparing translational efficiencies and mRNA expression levels with those of un-induced cells.

To account for the 2–3 h lag prior to the appearance of increased mature miRNA, the first time-point examined was 4 h post-induction. At 4 h, the miR-155–expressing line showed significant repression of genes with ≥ 1 site, all of which was attributed to translational repression (Figure 2C). At later time points, mRNA degradation dominated, as observed in B cells. For the miR-1–expressing line, 4 h was too early to observe significant repression for genes with ≥ 1 site, and by 8 h, mRNA degradation already dominated (Figure 2C), suggesting that we had missed any potential translation-dominant phase. For miR-124, a translation-dominant phase also was not observed (Figure 2C), presumably because induction was too gradual to achieve significant repression at early time points. (As we did not acquire murine proteomics data for miR-124, the top predicted targets were used instead of proteomics-supported targets.) Because miRNA induction in vivo is rarely more rapid than that achieved for miR-124 in our inducible line, we suggest that the miR-124 results are representative of most endogenous settings.

Minimal influence of translational stress and state

Having investigated eight different cell types and six different miRNAs, and having considered both pre-steady-state and later time points without identifying a setting with substantial overall

repression in which translational effects dominated, we turned to the potential influence of cellular state. Studies of *lin-4*-mediated repression in *C. elegans* suggest that starvation might tip the balance towards more translational inhibition with less mRNA degradation (Holtz and Pasquinelli, 2009), presumably because starvation influences global translational activity.

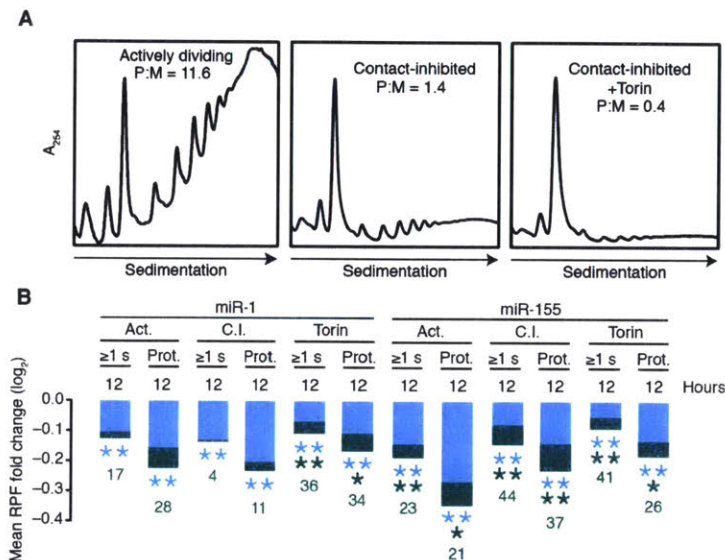


Figure 3. Negligible Influence Translational Stress or State on the Repression Mode
(A) Polysome profiles showing the translational activity of actively dividing (left), contact-inhibited (middle), and Torin1-treated contact-inhibited (right) miR-1 inducible 3T3 cells. Profiles are normalized to the monosome peak, with the polysome-to-monomosome ratio (P:M) indicated.

(B) The contributions of mRNA decay and translational repression following miR-1 or miR-155 induction in the corresponding 3T3 cell lines in the indicated states. Otherwise, as in Figure 2C. Results for contact-inhibited 3T3 cells expressing miR-1 and miR-155 were recalculated so as to only consider site-containing and no-site genes present in all samples. Abbreviations: Act., actively dividing; C.I., contact-inhibited; Torin, contact-inhibited and Torin1-treated; ≥1 s, genes with at least one site to the cognate miRNA in their 3' UTR; Prot., proteomics-supported targets. See also, Figure S4.

Therefore, we compared the relative contributions of TE and mRNA degradation for 3T3 cells in three translational states: 1) dividing cells, which have very active translation [polysome to monosome ratio (P:M) = 11.6], 2) contact-inhibited cells (P:M = 1.4), and 3) contact-inhibited cells under Torin1-induced mTOR inhibition (P:M = 0.4) (Figure 3A). We found no pervasive difference in the relative contribution of translational repression to miR-1- and miR-155-

mediated repression between these states (Figure 3B), despite the ~30-fold range in translational activity. Thus, translational stress, and more generally the translational state, does not have a perceptible global impact on the mode of miRNA-mediated regulation in these mammalian cells.

Because translating ribosomes displace miRNA-directed silencing complexes, which renders miRNA sites in the path of the ribosome much less effective than those ≥ 15 nt downstream of the stop codon (Grimson et al., 2007), we reasoned that the efficacy of sites within open reading frames (ORFs) might increase in conditions of reduced translational activity. Indeed, relative to the efficacy of 3'-UTR sites, the efficacy of ORF sites did appear to increase when translation was repressed with Torin1 (Figure S4A), which supported the model in which displacement of bound miRNAs by translating ribosomes is the predominant reason that ORF sites are ineffective.

Discussion

The principles of repression dynamics

Our results in 3T3 and B cells, considered in light of the fundamental differences between the nature of translational repression and mRNA destabilization, lead to the following principles regarding the miRNA-mediated repression of endogenous mRNAs in mammalian cells:

Compared to translational repression, detectable mRNA destabilization occurs after more of a lag, presumably because mRNA decay takes longer than inhibiting translation initiation. Because of this relative lag, after unusually robust miRNA induction, we can detect a short phase resembling that observed in reporter experiments, in which most of the repression is from decreased TE. However, the lag in destabilization does not last long, and destabilization soon dominates. To illustrate these principles, we simulated the repression time course of a rapidly

induced miRNA for which 80% of the steady-state repression is through mRNA destabilization and 20% is through translational repression (Figure 4). In our simulation, translational repression begins immediately upon miRNA–mRNA association, and mRNA degradation occurs through an increased degradation rate for the miRNA-bound mRNA. This approach yields an early phase in which translational repression dominates, consistent with that observed in our experimental time courses (Figure 4B). The transition from mostly translational repression to mostly mRNA destabilization takes place at 5.7 h (Figure 4C), when relatively little overall repression (9.7% RPF decrease, compared to a 50% decrease at steady state) is occurring (Figure 4B). Our example simulates very rapid miRNA induction; within 6 h the induced miRNA reaches levels that would make it the highest expressed miRNA in 3T3 cells (Figure 4A), similar to or faster than the induction observed in our 3T3 cell lines (Figure 2A). Slowing the induction rate by about half would result in this transition occurring at a point of even less repression (6.6% RPF decrease), and thus in most mammalian contexts, miRNA induction would be too slow to yield detectable repression during the phase in which TE changes dominate. For an early phase of substantial repression mediated primarily through TE changes, miRNA induction would have to be stronger than that ever reported, which is consistent with our inability to find a mammalian context with substantial translation-based repression.

Decreases in mRNA and TE lead to decreased protein from the targeted messages, and this change in protein is what matters to the cell. Despite the ultimate importance of the protein changes, measuring these changes over time is less informative for analyzing miRNA repression dynamics than is measuring RPF and mRNA changes, which more directly captures the molecular effects of the miRNA in inhibiting translation and destabilizing mRNA. RPF and mRNA measurements are also more suitable for quantitative comparisons for two reasons: 1)

they enabled accurate comparisons of more miRNA targets, and 2) they were each acquired using analogous methods that measured differences at one moment in time without the complications that arise from pre-steady-state measurements of protein changes. With regard to these complications, protein differences detected using direct labeling or standard metabolic labeling (e.g., SILAC) cannot distinguish between protein synthesized before or after induction of the miRNA and thus are unsuitable for pre-steady-state measurements because they would underestimate the impact on newly synthesized protein. Pulsed SILAC differentiates between pre-existing and newly synthesized protein but as currently implemented still entails an extended period (≥ 6 h for global measurements) of metabolic labeling (Schwanhausser et al., 2009; Huo et al., 2012), which compromises its utility for observing the results of the first few hours of repression.

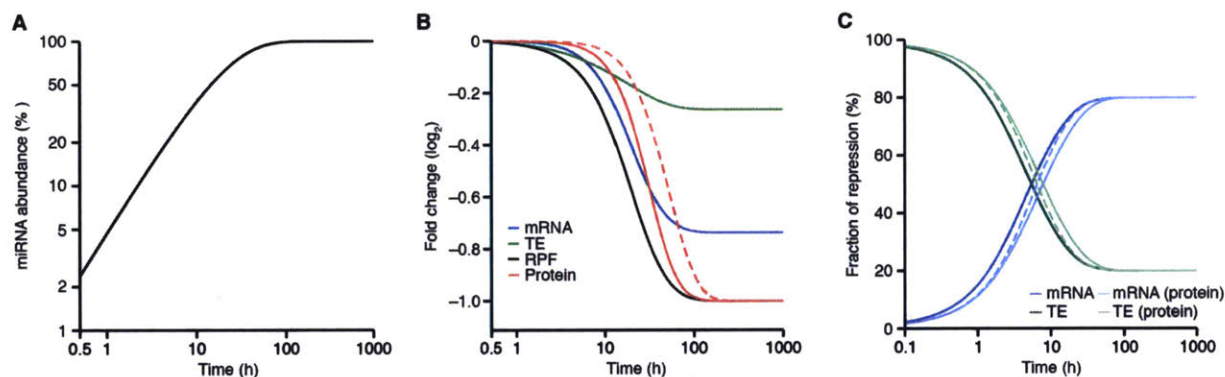


Figure 4. Simulated Dynamics of miRNA-Mediated Repression

(A) Simulation of rapid miRNA induction that begins with no miRNA and rises to a concentration exceeding that of the highest expressed endogenous miRNA in 3T3 cells within 6 h.

(B) Changes in target mRNA (blue), TE (green), RPF (black), and protein (red; solid line, 10 h protein half-life; dashed line, 100 h protein half-life) levels resulting from the miRNA induction in panel A.

(C) The relative contributions of mRNA decay and translational repression to the overall repression in panel A, when measured at either the RPF level (dark blue and dark green, respectively) or the protein level (light blue and light green; solid lines, 10 h protein half-life; dashed lines, 1 h protein half-life).

Despite the advantages of measuring RPF and mRNA changes, we note that during pre-steady-state conditions, the relative TE and mRNA effects can underestimate the relative contribution of translational repression to miRNA-mediated repression at the protein level. For example, at a given time point reduced mRNA might explain 80% of the RPF effect, leaving only 20% of the reduced protein synthesis at that moment to be explained by translational repression, but when considering the reduced protein levels (not current protein synthesis) more repression might have been due to translational repression. This is because the reduced protein levels are a function of the miRNA effects integrated up to the current time point, which includes earlier periods in which translational repression might have represented a greater share of the decreased protein synthesis.

The extent to which the relative contribution of translational repression would be underestimated depends on three factors: 1) the extent to which translational repression represents a greater share of the overall repression at the earlier time periods, 2) the relative strength of the overall repression during earlier periods, and 3) the stability of the protein. Our results indicate that with respect to the second factor, the relative strength of the overall repression during earlier periods is low in mammalian contexts, which implies that any underestimate of the contribution of translational repression to the reduction in protein levels would be minimal. In our simulation the greatest underestimate was observed at 5.7 h, when TE changes explained 49% of the reduction in protein synthesis at that moment and 58% of the reduction in protein accumulation, assuming intermediate protein stability (10 h protein half-life; Figure 4C). A shorter protein half-life further diminished the small differential between protein synthesis and protein accumulation (Figure 4C), whereas a longer half-life delays the onset of any consequential miRNA effect on protein abundance to a period well beyond the onset of

substantial mRNA decay (Figure 4B). In sum, monitoring protein levels rather than TE would not increase the prospects for finding a mammalian setting in which substantial translational repression dominates.

Comparison of fish embryos and mammalian contexts

Attempts to characterize the dynamics of the two modes of miRNA-mediated repression in zebrafish embryos were confounded by two unique features of fish and frog embryos prior to gastrulation: 1) a strong coupling between poly(A)-tail length and translational efficiency and 2) an unusual mRNA metabolism wherein mRNAs with short poly(A)-tails are stable. These features do not necessarily preclude analysis of dynamics, but in these contexts changes in TE due to miRNA-mediated deadenylation must be accounted for independently of changes in TE due to direct miRNA-mediated translational repression. Indeed, when the repression due to mRNA decay is thought of as including deadenylation-dependent translational repression, mRNA decay is the predominant mode of miRNA-mediated repression at all time points analyzed in zebrafish (Subtelny et al., 2014) just as it is at all but the earliest time points in mammalian cells. An important difference between most mammalian systems and early developmental systems (and presumably neuronal synapses or other systems with the aforementioned features) is that, in the latter, effects on translation must be measured to accurately capture the impact of the miRNA on gene expression, and effects on deadenylation must be measured to understand how repression is achieved. However, neither system seems to have a phase in which deadenylation-independent translational repression performs substantial repression without even stronger repression detectable by mRNA changes.

Mechanistic interpretations

Although translational repression and mRNA decay both lead to reduced protein synthesis, the mechanism used for repression has important biological implications. To the extent that repression occurs through translational repression, rapid recovery would be possible without requiring new transcription. This would, for example, be the case in early zebrafish embryos, where the repression of miRNA targets could be rapidly reversed through cytoplasmic polyadenylation. In most settings, however, reversal of miRNA-mediated repression requires new transcription, as mRNA decay constitutes the major mode of repression.

When miRNA-mediated mRNA decay was first reported, it was proposed to occur either through active recruitment of mRNA-degradation machinery or as a secondary effect of inhibiting translation (Lim et al., 2005). Although we observe translational repression prior to the decay of endogenous mRNAs in some experiments, this temporal relationship does not imply that mRNA decay is a consequence of translational repression, because it is also consistent with mRNA decay simply being a slower process. Indeed, several observations favor the model that the decay occurs through active recruitment of mRNA-degradation machinery rather than as a secondary effect of inhibiting translation. First, miRNA targeting can destabilize reporter transcripts that cannot be translated, which indicates that mRNA destabilization is not merely a secondary effect of reducing the number of ribosomes translating an mRNA (Mishima et al., 2006; Wu et al., 2006; Eulalio et al., 2007; Wakiyama et al., 2007; Eulalio et al., 2009; Fabian et al., 2009), although it does not rule out models in which only translationally repressed mRNAs can be destabilized. Second, direct biochemical interactions link miRNAs to Argonaute, Argonaute to TNRC6, and TNRC6 to the deadenylase complexes (the PAN2-PAN3 complex and the CCR4-NOT complex) that shorten the poly(A) tail (Braun et al., 2012), thereby showing how

the mRNA-degradation machinery can be actively recruited independent of either the act or consequence of translational repression. Finally, our work greatly expands the number of mammalian systems examined and shows that in each of these systems mRNA destabilization explains a large majority (from 70% to >90%) of the miRNA-mediated repression observed at steady state.

The idea that the mRNA destabilization might be a secondary consequence of inhibiting translation would be more plausible if a larger fraction of the steady-state repression was through translational repression; otherwise, the mRNA destabilization is out of proportion to the translational repression. We are not aware of any mammalian examples in which translationally repressed messages are so destabilized as a secondary consequence of this repression that the amount of steady-state destabilization exceeds the amount of steady-state translational repression. Indeed, the idea that mammalian messages might be destabilized solely as a secondary consequence of reduced ribosome occupancy or density appears to be largely an extrapolation from observations made in bacteria and yeast, but not mammalian cells (Muhlrad et al., 1995; Schwartz and Parker, 1999; Deana and Belasco, 2005). When examining mammalian mRNAs in general (irrespective of miRNA targeting), we find only a very weak correlation between TE and mRNA half-life in mammalian cells (Figure S4B, $R^2 = 0.004$ and 0.001 for 3T3 and HeLa, respectively), and others have shown that repression of translational initiation through the iron response element (a textbook example of mammalian translational repression) does not impart detectable destabilization of either its endogenous host mRNAs (Coccia et al., 1992; Melefors et al., 1993; Kim et al., 1996) or a reporter transcript (Hentze et al., 1987). Thus, when considered together, the available evidence strongly supports a model in which miRNAs actively recruit the deadenylation machinery and the ensuing deadenylation, decapping and decay

comprises the major mode of miRNA-mediated repression of endogenous targets in mammalian cells.

Some translational repression accompanies mRNA destabilization as a minor component of endogenous target repression in mammalian cells. Like mRNA destabilization, this translational repression also appears to depend on recruitment of CCR4-NOT, but three observations indicate that this repression is not simply a consequence of shortened poly(A)-tails. First, mRNAs without poly(A) tails can be translationally repressed (Wu et al., 2006; Eulalio et al., 2008; Eulalio et al., 2009; Braun et al., 2011; Chekulaeva et al., 2011; Zekri et al., 2013). Second, mutant complexes lacking deadenylase activity can nonetheless promote translational repression (Cooke et al., 2010). Third, tail length and TE are not correlated in most mammalian settings (Subtelny et al., 2014). Thus, the two modes of miRNA-mediated repression seem to represent two independent ramifications of recruiting the deadenylation complexes.

Reconciling results with single-gene studies of mRNA and protein changes

The conclusion that mRNA destabilization is the major mode of miRNA-mediated repression agrees with many previous observations monitoring protein and mRNA changes of single target genes after perturbing a miRNA. Among the >30,000 research studies of mammalian miRNAs, there are also counter examples, in which single-gene measurements seem to suggest a greater role for translational repression (Poy et al., 2004; O'Donnell et al., 2005; Zhao et al., 2005; Chen et al., 2006). An advantage of our approach is that we simultaneously examine thousands of genes, comparing the changes of both mRNA level and TE for hundreds of genes that have at least one miRNA site to those of hundreds of genes that lack a site and thus serve as internal controls. The aggregate result of this global approach should reflect the overall contributions of

mRNA destabilization and translational repression, whereas a single-gene study might choose a non-representative example and reach a conclusion that does not apply more generally to the targets of the miRNA.

This raises the question as to what might explain a single-gene result in which a miRNA-dependent change is observed in protein (i.e., with an immunoblot) but not mRNA (e.g., with quantitative reverse-transcription-PCR), which would appear as an outlier in our analyses. Might such outliers represent targets that are repressed at the level of translation without being destabilized? Although this possibility cannot be excluded, changes observed among our control genes that lack miRNA sites raise doubts about its validity. In most experiments (the possible exception being U2OS cells transfected with miR-155), a similar number of these control genes also change at the level of translation without being destabilized (Figure 1D). The observation that this behavior usually does not depend on the presence of a site to the miRNA suggests that either indirect effects of the miRNA or experimental variability explain the presence of most outliers that appear to be changing only at the level of translation.

Other single-gene examples for which translational repression is reported to be the major mode of miRNA-mediated regulation examine reporter mRNAs rather than endogenous mRNAs (Doench and Sharp, 2004; Kiriakidou et al., 2004; Nelson et al., 2004; Yekta et al., 2004; Pillai et al., 2005). Interestingly, the fractional component of regulation attributable to translational repression generally seems to be higher for reporters than for endogenous genes. We have begun experiments that aim to understand this difference between reporter and endogenous genes. Once this difference is understood, reporters could be developed that better recapitulate the regulation of endogenous genes, which would provide more relevant tools for studying the mechanism and dynamics of miRNA-mediated repression.

Experimental Procedures

RNA Isolation

For RNA-seq, total RNA was extracted from B cells and U2OS cells using TRI reagent.

Cytoplasmic RNA was extracted from U2OS cells using TRI reagent from cytoplasmic fractions that were separated from nuclear fractions by differential centrifugation. Briefly, whole-cell lysate prepared as described (Guo et al., 2010) was centrifuged at 1,300g for 10 min, and the resulting supernatant was collected as the cytoplasmic fraction while the pellet obtained was collected as the nuclear fraction. To prepare rRNA/tRNA-depleted U2OS total RNA, total RNA was first treated with the Ribo-Zero rRNA removal kit (Epicentre BioTechnologies) according to manufacturer's instructions. The resulting rRNA-depleted RNA sample was then spin-filtered using Ultra-4 centrifugal filters with Ultracel-100 membranes (Amicon) by centrifuging at 5,000g for 10 min at 4°C. The filtrate was enriched in tRNAs and was discarded, and the retentate was collected as the rRNA/tRNA-depleted RNA sample. RNA for all other samples was prepared by extracting RNA from ribosome profiling lysates with TRI reagent as described (Subtelny et al., 2014). Except in the case of the tRNA/rRNA-depleted U2OS RNA sample, the extracted RNA was poly(A)-selected as described (Subtelny et al., 2014).

Ribosome Footprint Profiling and RNA-seq

For B cell and U2OS samples, ribosome profiling and RNA-seq were performed essentially as described (Guo et al., 2010), with the only difference being how the RNA was isolated or enriched the cases of U2OS cytoplasmically enriched RNA and tRNA/rRNA-depleted total RNA. All other samples were prepared as described (Subtelny et al., 2014). Detailed protocols

are available at <http://bartellab.wi.mit.edu/protocols.html>. Reference transcript annotations were downloaded (in refFlat format) from the UCSC Genome browser, and for each gene, the longest transcript was chosen as a representative transcript model. RPF and RNA-seq reads were mapped to ORFs as described, which excluded the first 50 nt of each ORF so as to eliminate signal from ribosomes that initiated after adding cycloheximide (Subtelny et al., 2014).

Accession numbers

The NCBI GEO accession number for all ribosome profiling, RNA-seq and sRNA sequencing data is GSE60426.

Author Contributions

S.W.E., H.G., and D.P.B designed the study. S.W.E and H.G. performed ribosome profiling and RNA-seq, and S.W.E. did the associated analyses. S.E.M did the mathematical modeling. R.A.R-M and J.V. performed the proteomics. C.S. cultured primary cells. D.B. analyzed microarray and proteomics data. S.H. and K.G. harvested liver tissue. S.W.E., H.G., S.E.M., and D.P.B. wrote the paper, with input from the other authors.

Acknowledgements

We thank V. Agarwal and V. Auyeung for helpful discussions, D. Patrick, E. van Rooji and E. Olson for miR-21 knockout mice and wild-type controls, and the Whitehead Genome Technology Core for sequencing. Supported by NIH grants R01GM067031 (D.P.B) and R01CA193244 (K.G.). H.G. was supported by the Agency for Science, Technology and Research, Singapore. D.P.B is an investigator of the Howard Hughes Medical Institute.

References

- Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. *Nature* *455*, 64-71.
- 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*, 553-563.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* *136*, 215-233.
- Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* *336*, 233-237.
- Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* *20*, 1885-1898.
- Bethune, J., Artus-Revel, C.G., and Filipowicz, W. (2012). Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* *13*, 716-723.
- Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol Cell* *44*, 120-133.
- Braun, J.E., Huntzinger, E., and Izaurralde, E. (2012). A molecular link between miRISCs and deadenylases provides new insight into the mechanism of gene silencing by microRNAs. *Cold Spring Harb Perspect Biol* *4*.
- Castilla-Llorente, V., Spraggon, L., Okamura, M., Naseeruddin, S., Adamow, M., Qamar, S., and Liu, J. (2012). Mammalian GW220/TNGW1 is essential for the formation of GW/P bodies containing miRISC. *J Cell Biol* *198*, 529-544.
- Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat Struct Mol Biol* *18*, 1218-1226.
- Chen, C.Y., Zheng, D., Xia, Z., and Shyu, A.B. (2009). Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat Struct Mol Biol* *16*, 1160-1166.
- Chen, J.F., Mandel, E.M., Thomson, J.M., Wu, Q., Callis, T.E., Hammond, S.M., Conlon, F.L., and Wang, D.Z. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* *38*, 228-233.
- Coccia, E.M., Profita, V., Fiorucci, G., Romeo, G., Affabris, E., Testa, U., Hentze, M.W., and Battistini, A. (1992). Modulation of ferritin H-chain expression in Friend erythroleukemia cells: transcriptional and translational regulation by hemin. *Mol Cell Biol* *12*, 3015-3022.
- Cooke, A., Prigge, A., and Wickens, M. (2010). Translational repression by deadenylases. *J Biol Chem* *285*, 28506-28513.

Deana, A., and Belasco, J.G. (2005). Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev* *19*, 2526-2533.

Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* *336*, 237-240.

Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. *Genes Dev* *18*, 504-511.

Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008). GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol* *15*, 346-353.

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2009). Deadenylation is a widespread effect of miRNA regulation. *RNA* *15*, 21-32.

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.F., Doerks, T., Dorner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes Dev* *21*, 2558-2570.

Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A., *et al.* (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Mol Cell* *35*, 868-880.

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

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lcy-6* and other microRNAs. *Nat Struct Mol Biol* *18*, 1139-1146.

Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P., and Schier, A.F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. *Science* *308*, 833-838.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* *312*, 75-79.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engle, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* *27*, 91-105.

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

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol* 7, e1000238.

Hentze, M.W., Rouault, T.A., Caughman, S.W., Dancis, A., Harford, J.B., and Klausner, R.D. (1987). A cis-acting element is necessary and sufficient for translational regulation of human ferritin expression in response to iron. *Proc Natl Acad Sci U S A* 84, 6730-6734.

Holtz, J., and Pasquinelli, A.E. (2009). Uncoupling of *lin-14* mRNA and protein repression by nutrient deprivation in *Caenorhabditis elegans*. *RNA* 15, 400-405.

Huo, Y., Iadevaia, V., Yao, Z., Kelly, I., Cosulich, S., Guichard, S., Foster, L.J., and Proud, C.G. (2012). Stable isotope-labelling analysis of the impact of inhibition of the mammalian target of rapamycin on protein synthesis. *Biochem J* 444, 141-151.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218-223.

Kim, H.Y., LaVaute, T., Iwai, K., Klausner, R.D., and Rouault, T.A. (1996). Identification of a conserved and functional iron-responsive element in the 5'-untranslated region of mammalian mitochondrial aconitase. *J Biol Chem* 271, 24226-24230.

Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A. (2004). A combined computational-experimental approach predicts human microRNA targets. *Genes Dev* 18, 1165-1178.

Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., and Stoffel, M. (2005). Silencing of microRNAs in vivo with 'antagomirs'. *Nature* 438, 685-689.

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, 769-773.

Melefors, O., Goossen, B., Johansson, H.E., Stripecke, R., Gray, N.K., and Hentze, M.W. (1993). Translational control of 5-aminolevulinate synthase mRNA by iron-responsive elements in erythroid cells. *J Biol Chem* 268, 5974-5978.

Mishima, Y., Giraldez, A.J., Takeda, Y., Fujiwara, T., Sakamoto, H., Schier, A.F., and Inoue, K. (2006). Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. *Curr Biol* 16, 2135-2142.

Muhrad, D., Decker, C.J., and Parker, R. (1995). Turnover mechanisms of the stable yeast PGK1 mRNA. *Mol Cell Biol* 15, 2145-2156.

Nelson, P.T., Hatzigeorgiou, A.G., and Mourelatos, Z. (2004). miRNP:mRNA association in polyribosomes in a human neuronal cell line. *RNA* 10, 387-394.

- O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V., and Mendell, J.T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435, 839-843.
- Olsen, P.H., and Ambros, V. (1999). The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* 216, 671-680.
- Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005). Inhibition of translational initiation by *Let-7* MicroRNA in human cells. *Science* 309, 1573-1576.
- Poy, M.N., Eliasson, L., Krutzfeldt, J., Kuwajima, S., Ma, X., Macdonald, P.E., Pfeffer, S., Tuschl, T., Rajewsky, N., Rorsman, P., *et al.* (2004). A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* 432, 226-230.
- Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* 11, 1640-1647.
- Rissland, O.S., Hong, S.J., and Bartel, D.P. (2011). MicroRNA destabilization enables dynamic regulation of the miR-16 family in response to cell-cycle changes. *Mol Cell* 43, 993-1004.
- Schwanhauser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009). Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* 9, 205-209.
- Schwartz, D.C., and Parker, R. (1999). Mutations in translation initiation factors lead to increased rates of deadenylation and decapping of mRNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol* 19, 5247-5256.
- Seggerson, K., Tang, L., and Moss, E.G. (2002). Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation. *Dev Biol* 243, 215-225.
- Selbach, M., Schwanhauser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58-63.
- 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, 66-71.
- Thai, T.H., Calado, D.P., Casola, S., Ansel, K.M., Xiao, C., Xue, Y., Murphy, A., Frenthewey, D., Valenzuela, D., Kutok, J.L., *et al.* (2007). Regulation of the germinal center response by microRNA-155. *Science* 316, 604-608.
- Wakiyama, M., Takimoto, K., Ohara, O., and Yokoyama, S. (2007). Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes Dev* 21, 1857-1862.
- 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, 855-862.

Wu, L., Fan, J., and Belasco, J.G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci U S A* *103*, 4034-4039.

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. *Science* *304*, 594-596.

Zekri, L., Kuzuoglu-Ozturk, D., and Izaurralde, E. (2013). GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. *Embo J* *32*, 1052-1065.

Zhao, Y., Samal, E., and Srivastava, D. (2005). Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* *436*, 214-220.

Supplemental Experimental Procedures

Cell Lines and Culture Conditions

U2OS and HEK293T cells were cultured in DMEM supplemented with 10% fetal bovine serum at 37°C with 5% CO₂, and split every second or third day at ~90% confluency. miRNA-inducible NIH3T3 cells were cultured in DMEM supplemented with 10% donor calf serum and 2 µg/ml puromycin at 37°C with 5% CO₂, and split every second or third day at ~75% confluency. For experiments involving actively dividing 3T3 cells, 6.6 million cells were plated on a 500-cm² plate, and after 48 h, these cells were split and plated at the same density. After 24 h cells, cells were mock treated or treated with 2 µg/ml doxycycline (Clontech), and then harvested at the appropriate time point. Contact-inhibited 3T3 cells were grown under the same conditions, except after plating they were grown for 5 days before inducing with doxycycline. When applicable, contact-inhibited cells were treated with 250 nM Torin1 at the same time as induction.

For SILAC experiments, miRNA-inducible 3T3 cells were cultured in light or heavy media by supplementing DMEM lacking lysine (Invitrogen) with 10% dialyzed donor calf serum (Corning; dialysis was performed using a 10,000 mw cutoff against PBS) and either light or heavy (¹³C₆) lysine (Invitrogen). Cells were maintained in the appropriate SILAC media for at least ten doublings. Light- and heavy-labeled cells were grown for 5 days without splitting to achieve contact inhibition, induced with doxycycline or mock-induced, respectively, and then harvested after 48 h. Proteins were extracted, reduced with 5 mM dithiothreitol, alkylated with 15 mM iodoacetamide, and digested with lysyl endopeptidase (LysC) at a 1:100 (enzyme/substrate) ratio. Peptides were desalted by solid-phase extraction and offline fractionated by basic-pH (pH = 9) reverse-phase chromatography into 12 fractions. Fractions

were analyzed by LC-MS/MS on a hybrid linear ion trap-Orbitrap (Orbitrap Velos) mass spectrometer, with data analysis essentially as described (Baek et al., 2008).

Primary Cells and Tissue

All animal experiments were performed in accordance with protocols approved by the MIT and Ohio State University Committees on Animal Care. All B cell, neutrophil, and macrophage experiments were performed using tissue from 3-month-old male mice. For B cells, spleens were obtained from miR-155 knockout, miR-150 knockout and wild-type mice (Jackson Laboratory, stock numbers 007745, 007750 and 000664, respectively). Spleens were mechanically disrupted, red blood cells were lysed using Ammonium Chloride Solution (StemCell Technologies), and then B cells were isolated using anti-mouse CD45R/B220 magnetic particles (BD Biosciences) according to manufacturer's instructions and cultured in RPMI media. Quantitative proteomics measurements involving miR-150 knockout B cells and the corresponding wild-type cells used a reductive dimethylation labeling approach (Hsu et al., 2003). Proteins were digested as described above for 3T3 samples. Peptides were desalted by solid-phase extraction and dissolved in 1M MES pH 6. Labeling reaction proceeded for 10 min with normal (light) or deuterated versions (heavy) of 4% formaldehyde and 600 mM sodium cyanoborohydride, and was quenched with 5% trifluoroacetic acid. Peptides were fractionated and analyzed by mass spectrometry as described above for 3T3 samples, with data analysis and microarrays essentially as described (Baek et al., 2008).

Proteomics experiments on miR-155 used a pulsed-SILAC (pSILAC) approach, and relative quantification was performed on protein synthesized after activation. Wild-type and miR-155 knockout B cells were grown in regular RPMI media, containing light lysine.

Activation was performed by treatment with LPS (100 $\mu\text{g/ml}$), IL-4 (10 ng/ml), and anti-CD-40 antibody (10 $\mu\text{g/ml}$). For the 48 h time point, B cells from wild-type mice were activated in RPMI media supplemented with 4,4,5,5-D₄-L-Lysine (medium lysine, Cambridge Isotope Laboratories, Inc.), while B cells from miR-155 knockout mice were activated in RPMI media supplemented with U-¹³C₆, U-¹⁵N₂-Lysine (heavy lysine, Cambridge Isotope Laboratories, Inc.). At 2, 4, 8, and 48 hours post-activation, 100 $\mu\text{g/ml}$ cycloheximide was added to arrest translation. After incubating for 10 min at 37°C, cells were harvested in ice-cold PBS supplemented with cycloheximide and split into three portions for RNA-seq, ribosome profiling and proteomics. Equal numbers of cells from each genotype harvested at the 48 h time point were mixed for pSILAC. Protein digestion, peptide separation and mass spectrometric analysis were performed as described above for 3T3 samples. Data was analyzed as described (Baek et al., 2008), in this case focusing on the relative quantification between the heavy and medium proteins. For neutrophils and macrophages, bone marrow hematopoietic progenitors from miR-21 knockout mice and wild-type mice were isolated as described (Baek et al., 2008). For neutrophil experiments, hematopoietic progenitors were cultured as described (Baek et al., 2008). For macrophage experiments, hematopoietic progenitors were cultured for 6 days in RPMI 1640 supplemented with M-CSF and SCF (20 ng/ml and 50 ng/mL , respectively; PeproTech Inc., NJ) and then for 4 days with media supplemented with M-CSF (20 ng/ml) (Davies and Gordon, 2005). Microarray analyses involving miR-21 knockout macrophages and neutrophils and the corresponding wild-type cells were as described (Baek et al., 2008), and proteomic analyses were as described above for 3T3 samples.

For the liver experiment, 6-week-old male miR-122 knockout mice and wild-type littermates were anesthetized with Ketamine (90 mg/kg) and Zylazine (10 mg/kg), and the portal

vein was cannulated and perfused with ice-cold PBS with 100 µg/ml cycloheximide under aseptic conditions. Distal portions of the liver lobes were dissected, flash frozen in liquid nitrogen, and then used to prepare a lysate as described (Subtelny et al., 2014).

Transfection

U2OS cells were mock-transfected or transfected with 100 nM miRNA duplex using Lipofectamine 2000 at ~50% confluence according to manufacturer's instructions. At 24 h post-transfection, cells were incubated with cycloheximide (100 µg/ml) for 8 min at 37°C to arrest translation, and harvested as described (Guo et al., 2010). HEK293T cells were treated the same, except they were transfected at ~90% confluence.

Inducible Cell Lines

The sequence of *pre-miR-1a-1*, *-124a-3*, or *-155*, (primer pairs:

CGCg gatccATGTGTGAGAGAGACTGAGACACA and

CCGgaattcTCGGCCTCCCGAGGCCCTGCCGGT,

CGCg gatccGCTGGAGCATTCGCGCCCCTCAGG and

CCGgaattcGGCCACCGGGGGCCGGGGCTCGCC, and

CGCg gatccTTTCTCTTTGCAGGTGCTGCAAAC and

CCGgaattcGTCTGACATCTACGTTTCATCCAGC, respectively; relevant restriction sites in

lower case) along with 100 bp of upstream and 100 bp of downstream flanking genomic

sequence was inserted downstream of the tet-on promoter in the pRetroX-tight-pur plasmid

(Clontech). HEK293T cells were transfected with each plasmid vector and pCL-Eco using

Lipofectamine 2000, media was changed at 6 h and 24 h post-transfection, and then retroviral

particles were harvested after a further 24 h. 3T3 cells were co-infected with retroviral particles with one of the miRNA vectors and an rtTA vector in culture media supplemented with 4 µg/ml polybrene. After 48 h, cells were plated at very low density to allow individual colony formation and grown in culture media with 2 µg/ml puromycin. Once colonies were visible, they were isolated, tested for miRNA induction, and those showing the best induction were selected for use.

Analysis of miRNA Effects

The effects of a miRNA on target gene expression were determined by comparing the fold changes in RNA and RPF measurements observed for genes containing a site to the cognate miRNA in their 3' UTR relative to those observed for a cohort of control genes, which was randomly selected (without replacement) from the genes with no site to the cognate miRNA throughout their entire transcript to match the 3'-UTR-length distribution of site-containing genes. Unless stated otherwise, to account for the variability in changes observed among different possible cohorts, 1001 different cohorts of control genes were randomly chosen from the set of no-site genes and used to determine the percent contribution of RNA degradation for the set site-containing genes. The control cohort yielding the median percent contribution was chosen and used for the analysis of the set of site-containing genes. For samples within a time course, the cohort of no-site genes was chosen by performing this procedure on the last time point, and for comparisons across multiple samples, such as that done for U2OS cells, a no-site cohort was chosen that yielded a value for the RNA-degradation contribution that fell within 1% of the median value obtained for each sample when choosing 1001 cohorts for that sample. In all

cases, the proteomics-supported targets were compared to the no-site cohort generated for the genes with ≥ 1 site.

RNA Blots

For small-RNA blots, total RNA was isolated using TRI reagent from induced miR-1, miR-124, and miR-155 3T3 cells and activated B cells. RNA blots were performed as described (Pall et al., 2007), with a detailed protocol available at <http://bartellab.wi.mit.edu/protocols.html>. RNA blots from 3T3 samples included absolute quantification standards for the induced miRNA and miR-21. Endogenous U6 snRNA or miR-21 was used as a loading control for B cell and 3T3 samples, respectively.

Small-RNA Sequencing

Small-RNA (sRNA) sequencing was performed as described (Chiang et al., 2010), and sRNA reads were aligned to the complete set of mature mouse miRNAs (miRBase version 21) (Griffiths-Jones et al., 2008). Alignments were made based on the first 20 nt of each mature sequence, and reads from all miRNAs with an identical seed region (nucleotides 2–8) were grouped into a family and reported as the lowest-numbered miRNA within that family.

Mathematical Modeling of miRNA-Mediated Repression

A mathematical model of the consequences of miRNA-mediated repression at the mRNA and protein level was constructed, in which each molecular species is represented by a continuous dynamical variable. The system includes three RNA species: an expressed miRNA, a target mRNA that binds the miRNA with high affinity and is translated to produce a corresponding

protein, and a background mRNA pool that also binds the miRNA but with lower affinity than the target. The free miRNA i can reversibly bind either free target mRNA m or free background mRNA b to form complexes c_m and c_b , respectively, which are subject to increased degradation, and in the case of complex c_m , decreased translation of protein p . The kinetics of this system are given by six coupled first-order differential equations:

$$\frac{di}{dt} = \alpha_i + (\beta_m + \beta_c + k_m^-)c_m + (\beta_b + \beta_c + k_b^-)c_b - (\beta_i + \beta_{dil} + k_m^+m + k_b^+b)i \quad [1a]$$

$$\frac{dm}{dt} = \alpha_m + (\beta_i + k_m^-)c_m - (\beta_m + \beta_{dil} + k_m^+)m \quad [1b]$$

$$\frac{db}{dt} = \alpha_b + (\beta_i + k_b^-)c_b - (\beta_b + \beta_{dil} + k_b^+)b \quad [1c]$$

$$\frac{dc_m}{dt} = k_m^+im - (\beta_i + \beta_m + \beta_c + \beta_{dil} + k_m^-)c_m \quad [1d]$$

$$\frac{dc_b}{dt} = k_b^+ib - (\beta_i + \beta_b + \beta_c + \beta_{dil} + k_b^-)c_b \quad [1e]$$

$$\frac{dp}{dt} = \tau_m m + \tau_c c_m - (\beta_p + \beta_{dil})p \quad [1f]$$

Each free RNA species n (where $n = i, m, \text{ or } b$) is continuously generated through the process of transcription α_n , and lost due to cellular decay processes β_n , and the diluting effect of cell growth β_{dil} . Because both dilution and decay cause the cellular concentration of a species to decrease exponentially in the absence of transcription, the two processes are treated identically. The miRNA-bound forms of the mRNA c_n (where $n = m \text{ or } b$) are reversibly formed with association rate constant k_n^+ and dissociation rate constant k_n^- . Both the miRNA and mRNA within each of the two complexes are still subject to the same complex-independent decay processes described above, with the miRNA-bound mRNA experiencing an additional degradation rate, β_c , due specifically to its association with the miRNA. Lastly, the protein product p is continuously generated by translation of both the free and miRNA-bound form of the target mRNA (m and c_m)

given by the corresponding translation rate constant τ_n ($n = m$ or c), and is lost as a result of decay, β_p , and dilution, β_{dil} . miRNA-mediated translational repression is specified by assigning the miRNA-bound mRNA translation rate constant τ_c a value less than that of the free target mRNA, τ_m . Similarly, miRNA-mediated mRNA destabilization is specified by a non-zero value for the miRNA-dependent mRNA decay rate constant, β_c .

Parameter values were chosen from the literature or based on the results of our experimental time courses. The rate of dilution by cell growth reflects the 24 h doubling time observed for our 3T3 cell lines. The miRNA transcription and decay rate constants (α_i and β_i) used in Figure 4 produce a steady-state concentration of 100,000 copies per cell (cpc), with 90% induction by 48 h, resulting in levels similar to those observed for the miR-1 induction time course in 3T3 cells (Figure 2A). The transcription and decay rate constants for target mRNA (α_m and β_m) were chosen to reflect the median values for mRNA copy number and half-life [17 cpc and 9 h, respectively (Schwanhausser et al., 2011)]. The background mRNA pool b was assigned a transcription rate constant $\alpha_b = 30,000 \times \alpha_m$ to generate a background pool of miRNA target sites in excess of total miRNA (Denzler et al., 2014) by at least 4-fold, with the same decay constant (β_b) as for target mRNA. Translation and protein-decay rate constants (τ_m and β_p), were chosen such that the steady-state level of protein reached 50,000 cpc (Schwanhausser et al., 2011), with protein half-life varying from 1 to 10 to 100 h (Figure 4B and C), which bracketed the median protein half-life in 3T3 cells, reported to be 46 h (Schwanhausser et al., 2011). Values for target-miRNA association and dissociation rate constants k_m^+ and k_m^- were from *in vitro* data (Wee et al., 2012). For the complex formed with background mRNA, we used the same diffusion-limited association constant as for the target mRNA ($k_b^+ = k_m^+$), and a dissociation rate constant $k_b^- = 5 \times k_m^-$, to model a scenario in which target mRNA exhibits

greater relative occupancy of, and therefore greater repression by, the miRNA than does the background mRNA pool. As τ_c and β_c specify the amount of miRNA-mediated translational repression and mRNA destabilization in the model, their values were chosen empirically such that by 70 h, the overall miRNA-mediated repression reaches an approximate fold-change (\log_2) of -1 , with 80% of the steady-state repression conferred by mRNA destabilization (on the low end of the relative contribution of mRNA destabilization typically seen at steady state in mammalian cells; Figure 1C).

Each simulation was performed by solving the above differential system using the core R environment (R Core Team, 2014) with the package deSolve (Soetaert et al., 2010), with mRNA and protein levels beginning at non-repressed steady-state values, and miRNA levels beginning at zero:

$$i_0, c_{m0}, c_{b0} = 0 \quad [2a]$$

$$m_0 = \frac{\alpha_m}{\beta_m + \beta_{dil}} \quad [2b]$$

$$b_0 = \frac{\alpha_b}{\beta_b + \beta_{dil}} \quad [2c]$$

$$p_0 = \frac{\tau_m m_0}{\beta_p + \beta_{dil}} \quad [2d]$$

Because translational repression is assessed experimentally with the fundamental assumption that the instantaneous measurement of RPFs normalized to transcript length will scale linearly with the translation rate, we consider the term $\tau_m m + \tau_c c_m$ as representative of RPF measurement within this model, coming from both the free and miRNA-bound target mRNA. This allows for calculation of the fold change (\log_2) in RPFs θ_R , target mRNA levels θ_M , TE of the target mRNA θ_T , and protein levels over time θ_p . The curves in Figure 4B are generated using the following equations:

$$\theta_R = \log_2 \frac{\tau_m m + \tau_c c_m}{\tau_m m_0} \quad [3a]$$

$$\theta_M = \log_2 \frac{m + c_m}{m_0} \quad [3b]$$

$$\theta_T = \log_2 \frac{\tau_m m + \tau_c c_m}{\tau_m (m + c_m)} \quad [3c]$$

$$\theta_p = \log_2 \frac{p}{p_0} \quad [3d]$$

Note that the relationships between equations [3a–c] are such that: $\theta_M + \theta_T = \theta_R$, which is consistent with the experimental practice of subtracting fold change in mRNA levels from the fold change in RPFs to calculate the fold change in TE.

In Figure 4C, the contribution of mRNA destabilization and translational repression to the overall effect of the miRNA is assessed with respect to both the instantaneous percent reduction in RPFs as well as the percent reduction in accumulated protein. Because upon degradation of an mRNA, no RPFs can be ascribed to that molecule, irrespective of the extent of translational repression acting on that mRNA prior to its degradation, the absolute amount of instantaneous RPFs lost due to miRNA-mediated mRNA degradation is equal to the average number of RPFs per mRNA in the absence of repression (represented by τ_m), multiplied by the absolute difference in mRNA caused by miRNA-mediated mRNA degradation:

$$\Delta RPFs_M = \tau_m (m_0 - (m + c_m)) \quad [4]$$

The difference in absolute numbers of RPFs due to translational repression, by comparison, only comes from mRNAs that are extant (i.e., have not yet been degraded by miRNA-mediated destabilization), and is given by the difference in the average number of RPFs per mRNA between free and miRNA-bound target mRNA ($\tau_m - \tau_c$), multiplied by the level of miRNA-bound target mRNA:

$$\Delta RPF_{S_T} = (\tau_m - \tau_c) c_m \quad [5]$$

To calculate the relative contribution of mRNA destabilization and translational repression to the percent reduction in accumulated protein, we expanded the model with two additional dynamical variables p_M and p_T , which serve as a cumulative record of percent reduction in protein due to each mode of repression. Although these variables do not represent molecular species present in the system, their behavior is determined by translation and decay rates as if they represented protein present in the system. We use the corresponding expression for instantaneous loss of RPFs (given by equation [4] or [5]) as the rate of protein production for p_M and p_T , respectively, and the standard protein decay and dilution rate constants, yielding the differential equations:

$$\frac{dp_M}{dt} = \tau_m (m_0 - (m + c_m)) - (\beta_p + \beta_{dil}) p_M \quad [6a]$$

$$\frac{dp_T}{dt} = (\tau_m - \tau_c) c_m - (\beta_p + \beta_{dil}) p_T \quad [6b]$$

These equations were added to the model described by equations [1a–f] to simulate the dynamics of p_M and p_T as the mRNA undergoes miRNA-mediated repression. The curves for the percent reduction due to mRNA and TE (Figure 4C, dark blue and dark green, respectively) are calculated by dividing each of the instantaneous RPF values from equations [4] and [5] by the sum of both. The curves for the percent reduction due to mRNA and TE at the protein level (light blue and light green) are calculated identically, but using p_M , p_T and the sum of both.

At standard conditions, parameter values were as follows: $\alpha_i = 2.86 \times 10^{-9} \text{ M}^{-1} \text{ h}^{-1}$, $\beta_i = 1.91 \times 10^{-2} \text{ h}^{-1}$, $\alpha_m = 1.08 \times 10^{-12} \text{ M}^{-1} \text{ h}^{-1}$, $\alpha_b = 3.23 \times 10^{-8} \text{ M}^{-1} \text{ h}^{-1}$, $\beta_m = \beta_b = 7.70 \times 10^{-2} \text{ h}^{-1}$, $\beta_c = 1.17$

$\times 10^{-1} \text{ h}^{-1}$, $\beta_{dil} = 2.89 \times 10^{-2} \text{ h}^{-1}$, $k_m^+ = k_b^+ = 7.20 \times 10^{10} \text{ M}^{-1} \text{ h}^{-1}$, $k_m^- = 1.84 \text{ h}^{-1}$, $k_b^- = 9.18 \text{ h}^{-1}$,
 $\tau_m = 289 \text{ h}^{-1}$, $\tau_c = 209 \text{ h}^{-1}$, $\beta_p = 6.93 \times 10^{-2} \text{ h}^{-1}$. To reduce the protein half-life ten-fold (Figure
4B; dashed line), three values were changed: $\tau_m = 2.12 \times 10^3 \text{ h}^{-1}$, $\tau_c = 1.53 \times 10^3 \text{ h}^{-1}$, $\beta_p = 6.93 \times$
 10^{-1} h^{-1} , whereas to increase protein half-life ten-fold (Figure 4C; dashed lines), these values
were changed again: $\tau_m = 105 \text{ h}^{-1}$, $\tau_c = 76.1 \text{ h}^{-1}$, $\beta_p = 6.93 \times 10^{-3} \text{ h}^{-1}$.

Literature values for molecular abundances reported in cpc were converted to molar
concentrations (n_M) using the following relation, in which Avogadro's constant $N_A = 6.022 \times$
 10^{23} , 3T3 cell volume $v_{3T3} = 3.63 \times 10^{-12} \text{ L}$ (Swanson et al., 1991), and the proportion of the 3T3
cytoplasmic cell volume $\theta_{cyto} = 76.6\%$ (Swanson et al., 1991):

$$n_M = \frac{n_{cpc}}{N_A v_{3T3} \theta_{cyto}} \quad [7]$$

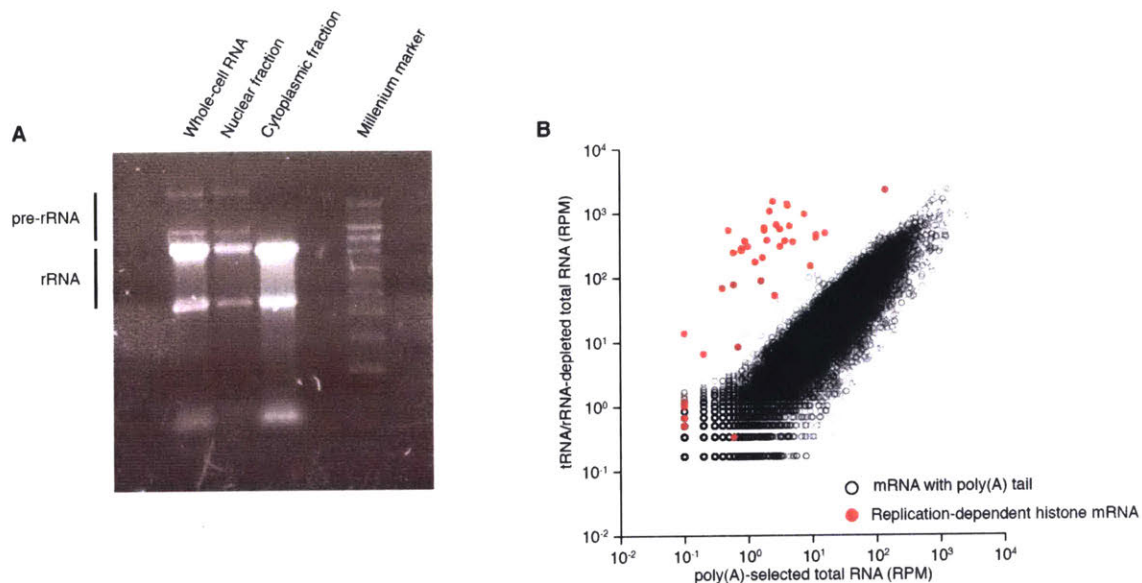


Figure S1, Related to Figure 1A. Enrichment of Different RNA Populations

(A) Analysis of RNA fractions on an agarose gel, showing the expected depletion of pre-rRNAs in the cytoplasmic fraction.

(B) Comparison of RNA-seq results for poly(A)-selected RNA and tRNA/rRNA-depleted total RNA, showing the expected difference in recovery for mRNAs of replication-dependent histone genes (red), which predominantly lack poly(A) tails, and much smaller differences for mRNAs of other genes (black open circles), which have poly(A) tails.

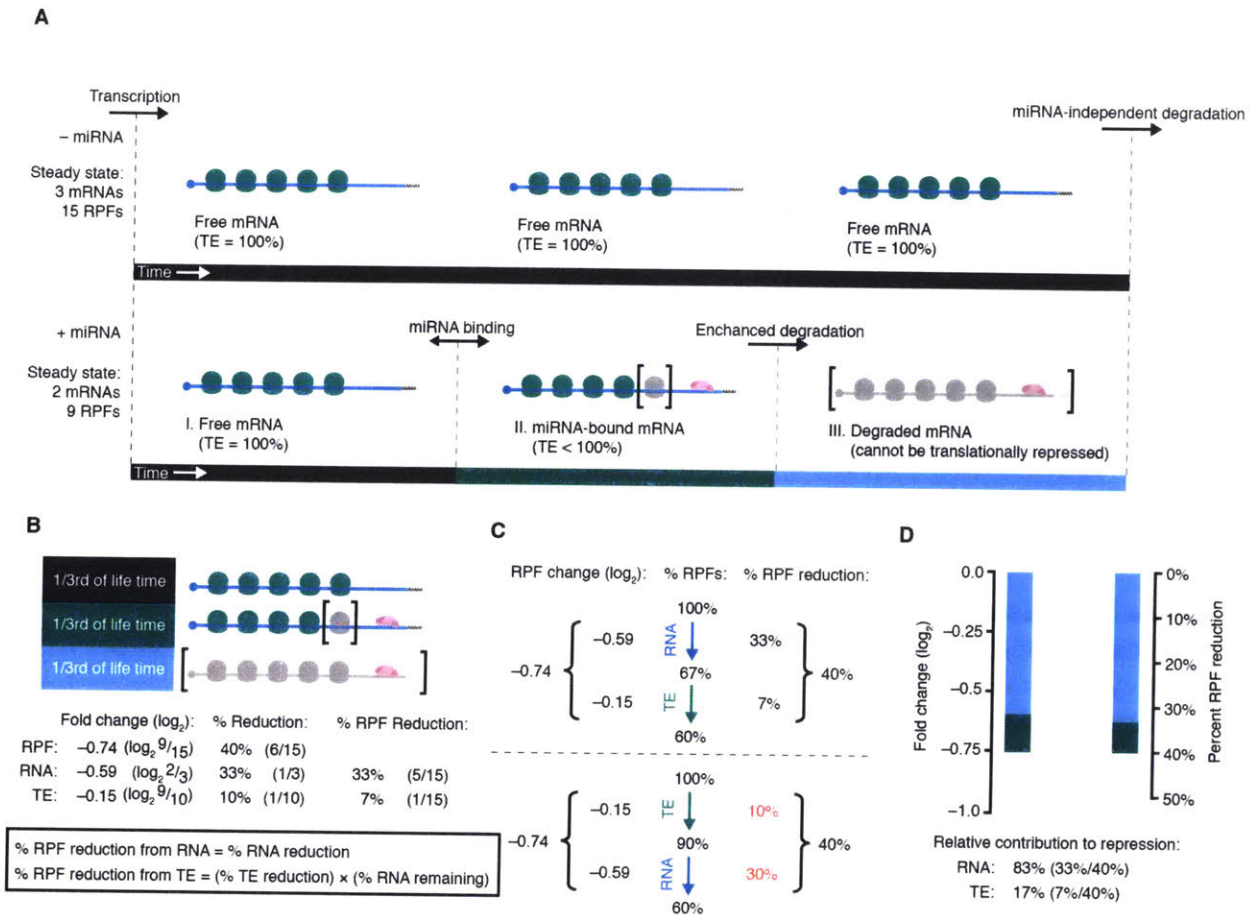


Figure S2, Related to Figure 1B. Calculating the Contribution of Translational Repression to Overall Repression

(A) Schematic of the temporal trajectory of an mRNA target in either the absence (upper) or presence (lower) of the miRNA. In the absence of the miRNA, the mRNA target is present at three steady-state copies, each with 5 RPFs. The trajectory of the target undergoing miRNA-mediated repression can be divided into three stages: stage I, in which the mRNA has been transcribed, is not bound by a miRNA silencing complex, and is being translated at its maximum efficiency; stage II, in which the mRNA has been bound by a miRNA silencing complex (purple), and is translated with sub-maximal TE; and stage III, in which the mRNA has been irreversibly degraded. Stage III begins at the point that degradation prevents any additional translation and lasts for the remainder of mean lifetime of the unrepressed mRNA. During stage III, there is no translatable mRNA and thus no TE to consider. The RNA loss is the fractional duration of stage III compared to the total duration of all three stages, and the translational repression is a function of the relative durations of stages I and II, as well as the decrease in translation during stage II. Ribosomes and mRNAs that would have been present in the absence of the miRNA are drawn in gray and enclosed in brackets.

(B) Example of the repressed mRNA from panel A at steady state, for which the three temporal stages of repression each have an equal duration. Because the steady-state mRNA population consists of molecules uniformly distributed along the temporal trajectory, the relative occupancy of each stage is proportional to its fractional duration. Below the diagram are the fold changes

(\log_2) in steady state RPFs, RNA molecules, and TE when comparing the repressed to the unrepressed condition, and the corresponding percent reduction for each of these three values. Additionally, the percent of the total reduction in RPFs contributed by reduction in RNA and TE is provided in the right-most column. Equations relating the percent reduction in RPFs given by percent reduction in RNA and TE are shown in the box below the calculated values.

(C) Importance of accounting for enhanced mRNA degradation before calculating the percent reduction due to reduced TE. The percent reduction in RPFs is calculated using the fold change (\log_2) values from panel B, considering either enhanced RNA degradation (top) or reduced TE first (bottom). Although both approaches give the same overall reduction in RPFs (40%, in agreement with the 6 grey ribosomes out of 15 total in panel B), accounting for the change in RNA before the change in TE gives the correct fraction of RPFs lost due to enhanced degradation (33%, 5/15 of the ribosomes of panel B) and the correct fraction of RPFs lost due to reduced TE (7%, 1/15 of the ribosomes of panel B), whereas the opposite order does not (red values).

(D) Compound bars showing the component of the RPF fold change (\log_2 , left) or the percent RPF reduction (right) attributable to mRNA degradation (blue) and reduced TE (green) for the example in panel B. The bars correspond to the values from the top portion of panel C. Use of the percent reductions to calculate relative contributions to overall repression is shown below the graph.

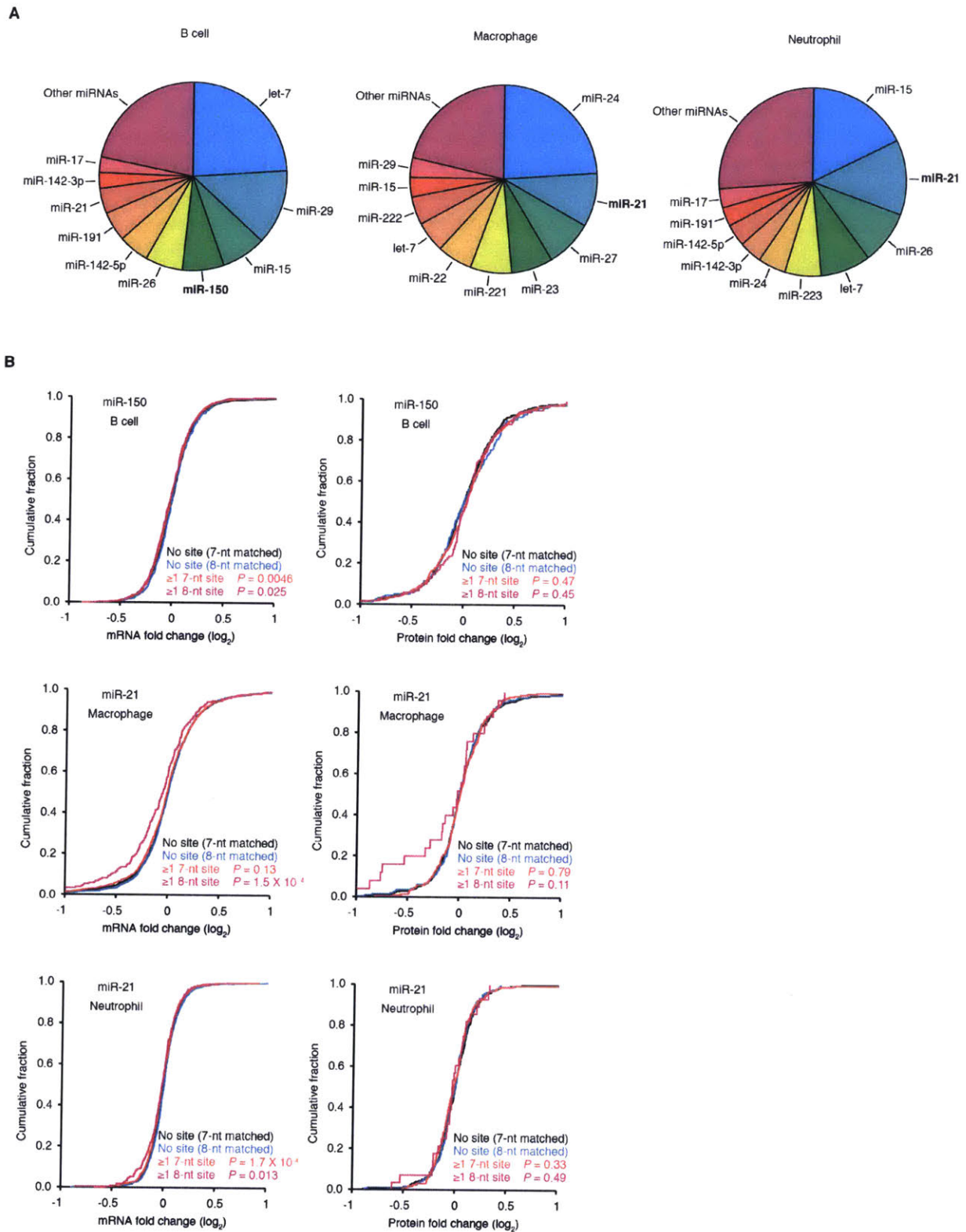


Figure S3, Related to Figure 1C. Correspondence Between mRNA and Protein Changes After Knocking Out Genes for Either miR-150 or miR-21

(A) Pie charts showing the proportion of the ten most highly expressed miRNAs and the aggregate of all other expressed miRNAs in B cells, macrophages, and neutrophils.

(B) Plots show cumulative distributions of mRNA or protein changes (left and right, respectively) at steady state in either murine B cells after knocking out *mir-150* (top), murine macrophages after knocking out *mir-21* (middle), or murine neutrophils after knocking out *mir-21* (bottom). Results are reported for two sets of site-containing genes, one comprising all genes with at least one 7-nt 3'-UTR site to the cognate miRNA (but no 8-nt sites), and another restricted to those with at least one 8-nt 3'-UTR site to the cognate miRNA. To maximize the number of genes analyzed, all genes exceeding the expression cutoff in the array data were analyzed regardless of whether they also exceeded the cutoff in the proteomics data, and vice versa. A set of control genes was generated for each set of site-containing genes by randomly choosing genes with no site to the cognate miRNA throughout their entire transcript to match the 3'-UTR-length distribution of site-containing genes, but without any repeated sampling. Otherwise, as in Figure 1A. *P* values are for the difference in the distribution for each set of genes with sites and that of the corresponding set of genes without sites (one-tailed K-S test).

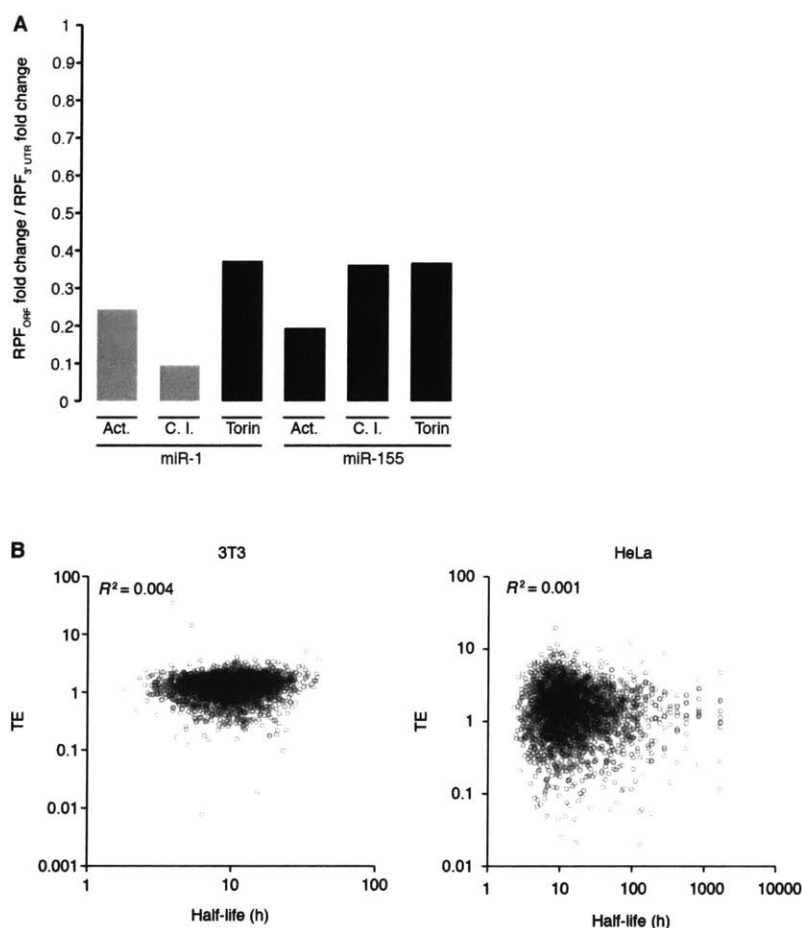


Figure S4, Related to Figure 3. The Impact of Global Translational Activity on Repression Mediated Through ORF Sites and the Relationship Between TE and mRNA Half-life.

(A) The relative efficacy of the ORF sites in different translational states, calculated as the RPF_{ORF} change divided by the $RPF_{3' UTR}$ change. The RPF_{ORF} change was the mean RPF change following 12 h of miRNA induction for genes containing at least one 7–8-nt site to the cognate miRNA in the ORF and no sites elsewhere in the transcript, normalized to the median change observed for a set of control genes that matched the ORF-length distribution of the site-containing genes but had no site throughout their transcripts. The $RPF_{3' UTR}$ change was the mean RPF change for genes containing at least one 7–8-nt site to the cognate miRNA in the 3' UTR and no sites elsewhere in the transcript, normalized to the median change for the set of control genes that matched the 3'-UTR-length distribution of the site-containing genes but had no site throughout their transcripts. In contrast to how the cohort of no-site genes was usually selected, the cohort yielding the median RPF-fold change for each set of site-containing genes was selected. Black bars indicate that repression of ORF-site genes was significant ($P \leq 0.05$, one-tailed K–S test), and in all cases repression of UTR-site genes was significant ($P \leq 0.05$, one-tailed K–S test).

(B) The relationship between TE and mRNA half-life. TEs for actively dividing 3T3 and HeLa cells, as measured by ribosome profiling and RNA-seq, are plotted in relation to previously published mRNA half-life measurements from 3T3 (Schwanhausser et al., 2011) and HeLa cells (Larsson et al., 2010) grown under conditions comparable to those of this study. All negative

half-life values were excluded from the HeLa data. 3T3 data are from the uninduced actively-dividing miR-1 cells, and HeLa data are from the mock-transfected cells. A 10 RPM cutoff was applied to the RNA-seq data, but not the RPF data, so as to not exclude any poorly translated genes.

Supplemental References

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

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E., *et al.* (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev* 24, 992-1009.

Davies, J.Q., and Gordon, S. (2005). Isolation and culture of murine macrophages. *Methods Mol Biol* 290, 91-103.

Denzler, R., Agarwal, V., Stefano, J., Bartel, D.P., and Stoffel, M. (2014). Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol Cell* 54, 766-776.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36, D154-158.

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

Hsu, J.L., Huang, S.Y., Chow, N.H., and Chen, S.H. (2003). Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* 75, 6843-6852.

Larsson, E., Sander, C., and Marks, D. (2010). mRNA turnover rate limits siRNA and microRNA efficacy. *Mol Syst Biol* 6, 433.

Pall, G.S., Codony-Servat, C., Byrne, J., Ritchie, L., and Hamilton, A. (2007). Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res* 35, e60.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing (Vienna, Austria, R Foundation for Statistical Computing).

Schwanhausser, 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, 337-342.

Soetaert, K., Petzoldt, T., and Setzer, R.W. (2010). Solving Differential Equations in R: Package deSolve. *J Stat Softw* 33.

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, 66-71.

Swanson, J.A., Lee, M., and Knapp, P.E. (1991). Cellular dimensions affecting the nucleocytoplasmic volume ratio. *J Cell Biol* 115, 941-948.

Wee, L.M., Flores-Jasso, C.F., Salomon, W.E., and Zamore, P.D. (2012). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* 151, 1055-1067.

Chapter 4

mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila* oocytes and early embryos

Stephen W. Eichhorn^{1,2}, Alexander O. Subtelny^{1,2,3}, Iva Kronja^{2,4}, Jamie C. Kwasnieski^{1,2}, Terry L. Orr-Weaver^{2,4*}, David P. Bartel^{1,2*}

¹Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, Cambridge, USA

²Department of Biology, Massachusetts Institute of Technology, Cambridge, USA

³Harvard-MIT Division of Health Sciences and Technology, Cambridge, USA

⁴Whitehead Institute for Biomedical Research, Cambridge, USA

S.W.E., A.O.S., and I.K.: Conception and design, acquisition of data, analysis and interpretation of data, drafting or revising the article. J.C.K.: Analysis and interpretation of data, drafting or revising the article. T.L.O.-W. and D.P.B.: Conception and design, analysis and interpretation of data, drafting or revising the article

Published as:

Eichhorn, S.W., Subtelny, A.O., Kronja, I., Kwasnieski, J.C., Orr-Weaver, T.L., and Bartel, D.P. (2016). mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila* oocytes and early embryos. *eLife* 5:e16955.

Abstract

Because maturing oocytes and early embryos lack appreciable transcription, posttranscriptional regulatory processes control their development. To better understand this control, we profiled translational efficiencies and poly(A)-tail lengths throughout *Drosophila* oocyte maturation and early embryonic development. The correspondence between translational-efficiency changes and tail-length changes indicated that tail-length changes broadly regulate translation until gastrulation, when this coupling disappears. During egg activation, relative changes in poly(A)-tail length, and thus translational efficiency, were largely retained in the absence of cytoplasmic polyadenylation, which indicated that selective poly(A)-tail shortening primarily specifies these changes. Many translational changes depended on PAN GU and Smaug, and both acted primarily through tail-length changes. Our results also revealed the presence of tail-length-independent mechanisms that maintained translation despite tail-length shortening during oocyte maturation, and prevented essentially all translation of *bicoid* and several other mRNAs before egg activation. In addition to these fundamental insights, our results provide valuable resources for future studies.

Introduction

The generation of a zygote through fertilization marks the onset of development for the offspring, and requires the proper completion of sperm and oocyte development in the respective parents. Spermatogenesis produces haploid sperm capable of penetrating the oocyte, whereas oogenesis produces differentiated oocytes that are stockpiled with maternal nutrients, proteins, and mRNAs, and have outer layers that protect the embryo and enable fertilization. In most animals the oocyte is arrested in meiosis, and fertilization leads to initiation of mitosis as the

oocyte nucleus completes meiosis and fuses with the haploid sperm nucleus. Fertilization also leads to changes in mRNA translation and protein stability, which support a period of development driven off of maternal stockpiles. After this initial stage of maternal control, which lasts for 1–2 mitotic divisions in mammals and 13 mitotic divisions in *Drosophila*, widespread transcription begins from the zygotic nuclei (Tadros and Lipshitz, 2009).

The production of viable offspring requires three key developmental events: oocyte maturation, the oocyte-to-embryo transition (OET), and the maternal-to-zygotic transition (MZT) (Figure 1A). Oocyte maturation involves the release of the primary meiotic arrest at prophase I and progression of the oocyte nucleus into meiotic divisions to produce a mature oocyte (egg) capable of being fertilized (Von Stetina and Orr-Weaver, 2011). The mature oocyte is arrested in metaphase II in most vertebrates and metaphase I in insects. The OET requires egg activation, which is coupled to fertilization in most animals. Egg activation releases the secondary meiotic arrest, allowing completion of meiosis (Horner and Wolfner, 2008). The MZT marks the transfer of control of development from the mother to the zygote as maternal mRNAs are degraded, transcription from the zygotic genome begins, and embryonic development becomes dependent on zygotic gene products (Tadros and Lipshitz, 2009). In *Xenopus*, zebrafish, and *Drosophila* the major activation of zygotic transcription occurs as the cell cycle lengthens and gastrulation begins, a developmental period referred to as the midblastula transition.

In nearly all animals, oocyte maturation and the OET occur in the absence of transcription, and thus any changes in protein levels must occur through posttranscriptional regulatory mechanisms. Indeed, widespread changes in protein levels without corresponding changes in mRNA levels are observed at these developmental transitions (Kronja et al., 2014a; Kronja et al., 2014b). Posttranscriptional regulation of gene expression during oocyte maturation, the OET,

and early embryogenesis faces unique challenges, including stably storing maternal mRNAs through the prolonged periods of oogenesis and maintaining some of these mRNAs in a translationally inactive state (referred to as masking) while selectively translating others at the proper times in development. During mouse oocyte maturation ~1300 mRNAs are translationally upregulated and ~1500 are translationally downregulated (Chen et al., 2011). Similarly, during *Drosophila* egg activation ~1000 mRNAs are translationally upregulated, and ~500 are translationally repressed (Kronja et al., 2014b).

One long-appreciated mechanism of translational regulation during development acts through control of mRNA poly(A)-tail length (Richter, 2007; Weill et al., 2012). The observation of short poly(A) tails in oocytes from amphibians and marine invertebrates led to the proposal that short poly(A) tails help mask maternal mRNAs, preventing translation without triggering the destabilization that typically results from deadenylation. Specific mRNAs are then targeted for cytoplasmic polyadenylation, and the lengthened poly(A) tails in turn cause translational upregulation of these mRNAs (McGrew et al., 1989), as demonstrated for the mRNAs for *mos* and several *cyclin B* genes during oocyte maturation in *Xenopus* (Sheets et al., 1994; Barkoff et al., 1998). The Cytoplasmic Polyadenylation Element Binding Protein (CPEB) plays a critical role in controlling translation of maternal mRNAs, both in recruiting Maskin to repress translation and later in promoting polyadenylation by binding cytoplasmic poly(A) polymerase together with Cleavage and Polyadenylation Specificity Factor (CPSF) (Richter, 2007; Weill et al., 2012; Ivshina et al., 2014). In *Drosophila*, the cytoplasmic poly(A) polymerase encoded by the *wispy* gene promotes poly(A)-tail lengthening during both oocyte maturation and egg activation (Benoit et al., 2008; Cui et al., 2008; Cui et al., 2013), and increased poly(A)-tail lengths of mRNAs for *cyclin B*, *cortex*, *bicoid*, *torso*, and *Toll* coincide with the translational

upregulation of these mRNAs (Salles et al., 1994; Lieberfarb et al., 1996; Vardy and Orr-Weaver, 2007a; Benoit et al., 2008). Moreover, appending synthetic poly(A) tails of increasing lengths to the *bicoid* mRNA promotes the translation of this mRNA, which demonstrates a causal relationship between increased poly(A)-tail length and increased translation in early *Drosophila* embryos (Salles et al., 1994).

Translational repressors, some of which have been identified through genetic analysis of pattern formation in the *Drosophila* embryo, also act during oogenesis and early embryogenesis (Vardy and Orr-Weaver, 2007b; Besse and Ephrussi, 2008; Barckmann and Simonelig, 2013). These repressors can impact either poly(A)-tail length (Temme et al., 2014), the translational machinery, or both. Brain Tumor and Pumilio (PUM) bind to target mRNAs and recruit the CCR4-NOT deadenylase complex, which shortens poly(A) tails (Kadyrova et al., 2007; Temme et al., 2010; Newton et al., 2015). Bruno binds to target mRNAs during oogenesis and recruits the eIF4E-binding protein Cup, which represses translation by both blocking the interaction between eIF4E and eIF4G (Nakamura et al., 2004) and promoting deadenylation while repressing decapping, which traps mRNAs in a translationally repressed state (Igreja and Izaurralde, 2011). Similarly, the RNA-binding protein Smaug (SMG) is reported to recruit either Cup or the CCR4-NOT deadenylase complex, with the consequent repression of translation (Nelson et al., 2004; Semotok et al., 2005). *smg* mRNA is translationally repressed in the mature oocyte by PUM, and at egg activation this repression is relieved by the PAN GU (PNG) protein kinase complex (Tadros et al., 2007). PNG is also thought to regulate other translational regulators.

mRNA expression arrays, RNA-seq and ribosome-footprint profiling have enabled global analyses of mRNA levels and relative translational efficiencies (TEs) in *Drosophila* oocytes and

embryos (Qin et al., 2007; Dunn et al., 2013; Chen et al., 2014; Kronja et al., 2014b). Measuring poly(A)-tail lengths has been more challenging due to the homopolymeric nature of the tail, and this technical hurdle has restricted studies to measuring tail lengths of a small number of *Drosophila* mRNAs (Salles et al., 1994; Lieberfarb et al., 1996) or to estimating relative differences in tail lengths more broadly using hybridization-based methods (Cui et al., 2013). However, methods are now available to measure poly(A)-tails of essentially any length for millions of individual mRNA molecules (Chang et al., 2014; Subtelny et al., 2014). In zebrafish and *Xenopus*, high-throughput measurements demonstrate that poly(A)-tail length and TE are globally coupled during early embryonic development (Subtelny et al., 2014), as expected from single-gene studies (McGrew et al., 1989; Sheets et al., 1994; Lieberfarb et al., 1996; Barkoff et al., 1998; Pesin and Orr-Weaver, 2007; Benoit et al., 2008). Surprisingly, however, this coupling disappears at gastrulation and is not observed in any of the post-embryonic samples examined (Subtelny et al., 2014). Thus, soon after a developmental shift in transcriptional control (the MZT), these vertebrate species undergo a developmental shift in translational control, in which TE abruptly changes from being strongly coupled to tail length to being uncoupled.

Drosophila melanogaster affords the opportunity to define changes in poly(A)-tail length and TE during oocyte maturation, the OET, and the MZT; facilitated by the high quality of both the genome assembly and the coding-sequence (CDS) annotations, the wealth of knowledge on regulatory proteins active during early development, the availability of genetic mutants in these known regulatory proteins, and the ability to isolate sufficient quantities of oocytes and embryos at distinct developmental stages for high-throughput sequencing techniques. Here we systematically analyze changes in poly(A)-tail length and TE during oocyte maturation, the OET and the MZT, and delineate the roles of Wispy, PNG, and SMG in regulating translation during

these crucial periods. These studies reveal fundamental insights into the regulatory mechanisms and regimes operating during these key developmental transitions.

Results

A conserved switch in the nature of translational control

We performed poly(A)-tail length profiling by sequencing (PAL-seq) (Subtelny et al., 2014) to measure poly(A)-tail lengths globally at eight developmental stages in *Drosophila melanogaster*, beginning with oocytes at the primary meiotic arrest point and ending with embryos late in gastrulation, at germ-band retraction (Figure 1A). PAL-seq determines the poly(A)-tail length of millions of mRNA molecules, providing for each molecule both a fluorescent measurement that reports on the poly(A)-tail length and a sequencing read that reports the cleavage and poly(A) site. Because alternative start sites or alternative splicing can generate different transcripts with the same poly(A) site, we considered our results with respect to a curated set of unique gene models. Moreover, because alternative cleavage and polyadenylation can generate different poly(A) sites for mRNAs from the same gene, we used our PAL-seq data to identify for each gene the alternative isoform with the longest 3' UTR and then combined its tail-length measurements with those of any shorter tandem 3'-UTR isoforms. Thus, the “mRNAs” of our analyses each corresponded to a unique gene model representing a different protein-coding gene and the aggregated tail-length results from many (if not all) of the alternative transcripts from that gene. In parallel, we performed ribosome-footprint profiling and RNA-seq to determine TEs (Ingolia et al., 2009), which also report aggregate results for alternative isoforms, allowing comparison of mean poly(A)-tail length and TE for thousands of mRNAs that passed our quantification cutoffs at each stage.

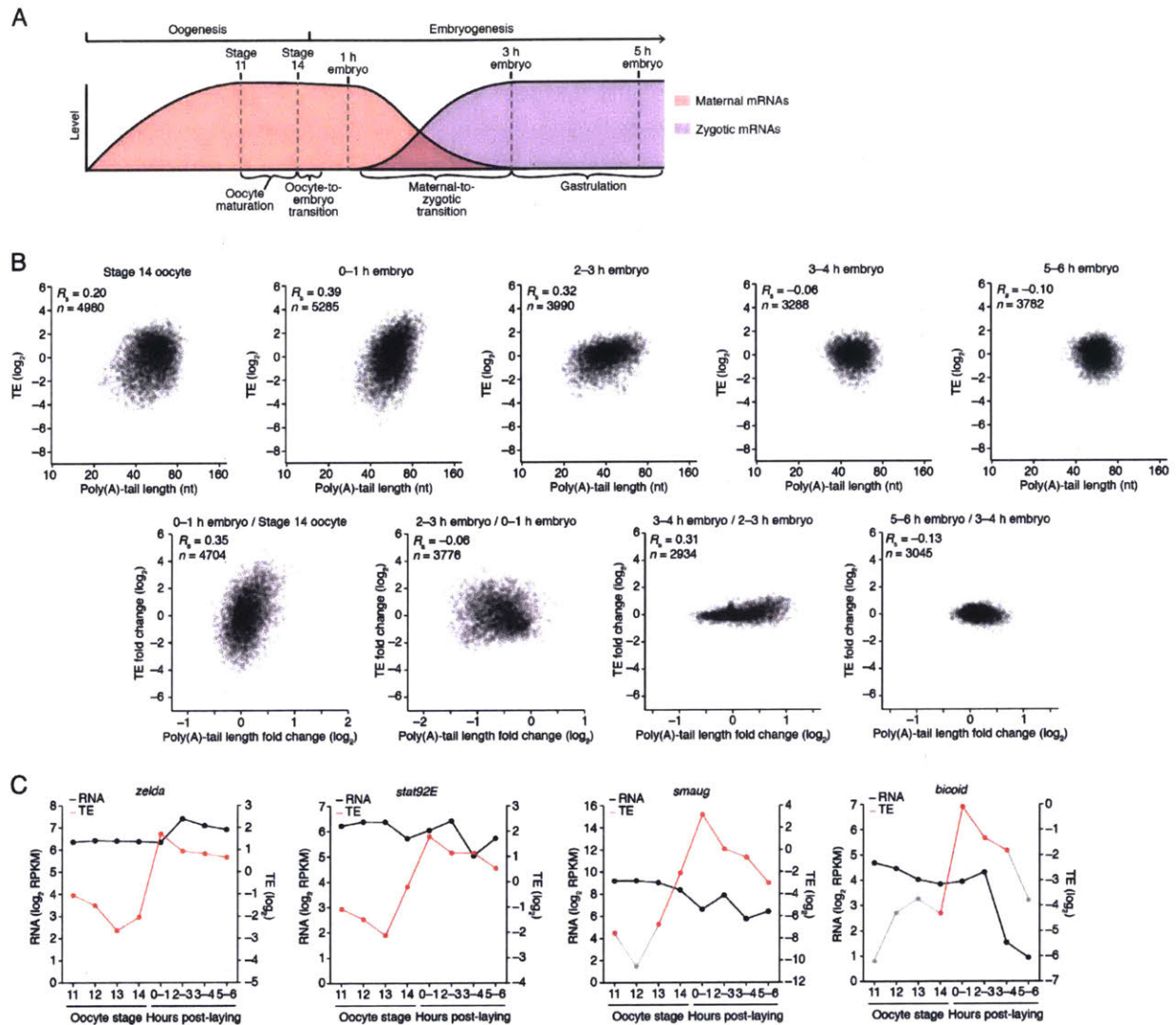


Figure 1. Transient coupling between poly(A)-tail length and translational efficiency in embryos.

(A) Time line of the developmental transitions probed in this study, illustrating the presence of maternal and zygotic mRNAs (pink and purple, respectively). (B) Relationship between mean poly(A)-tail length and relative translational efficiency (TE) at the indicated developmental stage (*top*), and the relationship between tail-length and TE changes observed between adjacent stages (*bottom*). For each stage, tail-length and TE measurements were from the same sample. In the top panels, results are plotted for all mRNAs that had ≥ 100 poly(A)-tail length measurements (tags), ≥ 10.0 reads per million mapped reads (RPM) in the RNA-seq data, and >0 RPM in the ribosome-profiling data. In each sample, TE values (\log_2) were median centered by subtracting the median TE value of that sample from each value (median values in stage 14 oocytes, 0–1 h embryos, 2–3 h embryos, 3–4 h embryos, and 5–6 h embryos were -0.2523 , -0.1329 , 0.4799 , 0.6469 , and 0.6447 , respectively). In the bottom panels, results are plotted for all mRNAs that had ≥ 100 poly(A) tags and ≥ 10.0 RPM in the RNA-seq data of both samples, and ≥ 10.0 RPM in the ribosome-profiling data of one of the two samples and >0 RPM in the other. TE fold-change values (\log_2) were median centered for each comparison (median values, 0.2027 , -0.0047 ,

0.0701, and -0.003 from left to right). The Spearman correlation coefficient (R_s) and number of mRNAs (n) are shown in each plot. The TE data for stage 14 were from Kronja et al. (2014b). (C) Abundance and TE of *zelda*, *stat92E*, *smaug*, and *bicoid* mRNA at the indicated developmental stages (RPKM is reads per kilobase per million mapped reads). TE values (\log_2) for each stage were median centered as described above, using the median values (\log_2) reported in the legends of Figure 1B and Figure 2. The data for stage 14 were from Kronja et al. (2014b). If no RPF reads were observed in a sample, TEs were calculated using a pseudocount of 1 read. Points for which TE is based on either a single read or pseudocount are in grey, as are the lines connecting to them. Other TEs were based on ≥ 25 RPF reads, except the TEs for *bicoid* in stage 14 oocytes and 3–4 h embryos (3 and 9, respectively) and the TE for *smaug* in stage 13 oocytes (11 reads).

In *Drosophila*, the first 13 division cycles occur in a nuclear syncytium. Cellularization (at 2–3 hours of embryogenesis) marks completion of the MZT and is immediately followed by the morphogenetic movements of gastrulation. Thus, early embryogenesis is controlled by maternal mRNAs, and based on the studies of mRNAs from a few genes (Salles et al., 1994; Lieberfarb et al., 1996), those with longer poly(A)-tails would be predicted to have higher TEs, as observed in vertebrate systems (Subtelny et al., 2014). Indeed, poly(A)-tail length and TE were correlated in cleavage-stage embryos (0–1 h embryos), and the slope of this relationship was steep, such that small differences in poly(A)-tail length corresponded to large differences in TE (Figure 1B; e.g., the median TE of mRNAs with a poly(A)-tail length between 70–80 nt was >5 fold greater than that of mRNAs with a poly(A)-tail length between 30–40 nt). Likewise, when progressing through the OET, the changes in TE correlated with the changes in tail length (Figure 1B, 0–1 h embryo / Stage 14 oocyte).

Although a positive correlation between poly(A)-tail length and TE persisted in the 2–3 h embryos, the slope of this relationship strongly diminished such that the median TE of mRNAs with a poly(A)-tail length between 70–80 nt was <2 fold greater than that of mRNAs with a poly(A)-tail length between 30–40 nt, as expected if poly(A)-tail length changes were no longer used to influence translation in these cellularized and gastrulating embryos. Indeed, the changes

in poly(A)-tail length that occurred between cleavage-stage and cellularized embryos were not positively correlated with the changes in TE (Figure 1B, 2–3 h embryo / 0–1 h embryo). The positive correlation between poly(A)-tail length and TE disappeared in subsequent stages of gastrulation (3–4 h embryos and 5–6 h embryos), and although tail lengths changed between these stages, the TE changes spanned a very narrow range (Figure 1B, 3–4 h embryo / 2–3 h embryo and 5–6 h embryo / 3–4 h embryo). The tail-length changes observed between these stages presumably resulted from transcription and mRNA decay as well as deadenylation, whereas those in earlier stages resulted primarily from cytoplasmic polyadenylation and deadenylation.

The narrower range of TE values and TE changes following the onset of gastrulation (Figure 1B) was consistent with a diminished role of translational regulation in specifying protein output after the embryo begins to modulate gene expression by altering mRNA abundance. Most importantly, these results showed that the developmental shift in the nature of translational control observed in vertebrates also occurs in flies; in *Drosophila*, just as in zebrafish and *Xenopus* (Subtelny et al., 2014), the coupling between poly(A)-tail length and TE observed in the early embryo disappeared around gastrulation.

An analysis of the temporal patterns of translational regulation occurring from the OET through the onset of gastrulation identified several significantly enriched patterns, although all possible patterns were observed (Figure 1—figure supplement 1). Among the five enriched profiles were those for mRNAs that were translationally upregulated at the OET and then stayed elevated during early embryonic development, those for mRNAs that were upregulated at the OET and then downregulated during early embryonic development, and those for mRNAs that were downregulated before the OET and remained repressed (Figure 1—figure supplement 1).

Because ribosome-profiling data reflect relative rather than absolute levels of translation of an mRNA across developmental stages, these results should be interpreted as reflecting changes in how well an mRNA was translated relative to all other mRNAs in the same stage. Thus, an mRNA translated at the same absolute efficiency in two stages might be perceived as downregulated if translation of many other mRNAs increased. Nonetheless, these dynamic and substantial changes in TE observed during the OET and early embryogenesis revealed a large-scale and rapid change in the translational capacity of mRNAs relative to each other during early development.

Pairwise comparisons within the mRNA, translation, and tail-length datasets (from RNA-seq, ribosome-footprint profiling, and PAL-seq, respectively) for the different developmental stages further supported these overall conclusions (Supplementary File 1). As expected, mRNA abundances were highly correlated between all stages prior to the onset of transcription (Supplementary File 1). Over these same stages, poly(A)-tail lengths for many different mRNAs shortened or lengthened, but these changes seemed to have minimal impact on mRNA abundance measurements, consistent with stability of shorter-tail mRNAs during this time (and also indicating that the poly(A)-selection that was performed as part of the RNA-seq protocol did not bias our RNA abundance, and thus TE, measurements). In contrast to the highly correlated mRNA abundances, the number of ribosome-protected fragments (RPFs) differed substantially between stages prior to the onset of transcription (Supplementary File 1), consistent with dynamic and robust translational regulation during this time. Once transcription began, widespread changes in both mRNA levels and RPFs were observed between stages (Supplementary File 1). Poly(A)-tail lengths varied both prior to and after the onset of transcription, and generally mirrored the pattern of correlations observed for RPFs up until the

MZT (Supplementary File 1). Little-to-no positive correlation between poly(A)-tail lengths was observed when comparing samples with no active transcription to those after the onset of robust transcription (Supplementary File 1), presumably because of the influence of nascent transcripts on mean poly(A)-tail lengths.

The striking increases in translation observed soon after fertilization presumably generated factors needed for later developmental transitions. For example, translation of mRNAs for *zelda* and *stat92E*, two transcription factors required for the onset of zygotic transcription at the MZT (Liang et al., 2008; Harrison and Eisen, 2015), increased markedly during the OET (Figure 1C), as expected if translational activation of these mRNAs produces the proteins necessary to initiate the MZT, and accounting for the previously noted increase in Zelda protein in early embryos (Harrison and Eisen, 2015). Likewise, translation of *smaug* and *bicoid* mRNA increased markedly during the OET (Figure 1C), yielding factors essential for embryonic development and patterning, respectively. Although translation of *stat92E* and *smaug* began to increase at stage 14, translation of both was still relatively low until after the OET (Figure 1C), consistent with the roles their translation products during early embryonic development.

Translational regulation during oocyte maturation

Having established that *Drosophila* embryos resemble those of fish and frog with respect to the coupling between TE and poly(A)-tail length, we sought to extend these findings to oocytes. The ability to isolate individual oocytes at distinct developmental stages in *Drosophila* enabled us to measure poly(A)-tail lengths and TEs and examine their relationship at key stages of oogenesis. Stage 11 oocytes are arrested in prophase I. Nuclear envelope breakdown, a hallmark of the release of this arrest, occurs between stages 12 and 13 (Von Stetina et al., 2008; Hughes et al.,

2009), and the oocytes arrest in metaphase I during stage 14, the final stage of oogenesis. Thus, comparison of oocytes from stages 11 through 14 captures the events of meiotic progression and oocyte differentiation that occur over the course of oocyte maturation.

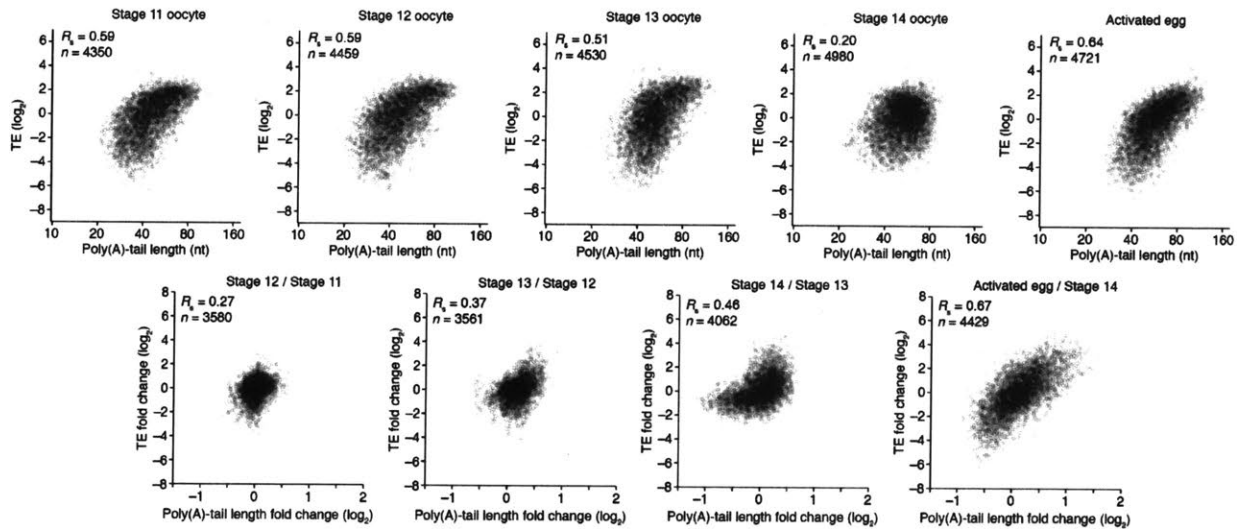


Figure 2. Coupling between poly(A)-tail length and translational efficiency in oocytes. Relationship between mean poly(A)-tail length and TE in either oocytes at the indicated stage or activated eggs (*top*), and the relationship between tail-length and TE changes observed between adjacent stages (*bottom*). TE values (\log_2) were median centered (median values in stage 11, 12, 13, and 14 oocytes and activated eggs were -0.6641 , -0.9006 , -1.38 , -0.2523 , and -0.0158 , respectively). TE fold-change values (\log_2) were median centered for each comparison (median values, -0.1961 , -0.3969 , 1.0886 , and 0.2123 from left to right). The TE data for stage 14 oocytes and activated eggs were from Kronja et al. (2014b), and the plot of mean poly(A)-tail length and TE for the stage 14 oocyte is modified from Figure 1B. Otherwise, these panels are as in Figure 1B.

Poly(A)-tail lengths and TEs were well correlated at stages 11, 12, and 13 ($R_s \geq 0.51$), but poorly correlated at stage 14 ($R_s = 0.20$) (Figure 2). Previously, the relationship between poly(A)-tail length and translation had been examined for the *c-mos* and *cyclin B* mRNAs during *Xenopus* oocyte maturation (Sheets et al., 1994; Barkoff et al., 1998; Pique et al., 2008). Our results extended these analyses to an invertebrate and to a global scale, showing that, as for the *c-mos* and *cyclin B* mRNAs in the *Xenopus* oocyte, tail-length increases correspond to translational activation in the maturing *Drosophila* oocyte. Thus, the global coupling between poly(A)-tail

length and TE, which had previously been examined only after fertilization (Subtelny et al., 2014), extends well before fertilization to potentially influence key developmental transitions in the oocyte.

Of the mRNAs exceeding our read cutoffs, nearly half (46%) had net TE changes of ≥ 2 fold over the course of oocyte maturation, with mRNAs for 1264 and 1037 genes up- or downregulated, respectively (Figure 2—figure supplement 1A). In contrast, far fewer mRNAs (1.5%) changed in abundance ≥ 2 fold (25 mRNAs increased, and 52 mRNAs decreased), indicating the posttranscriptional nature of this regulation. More detailed comparison of TEs for each mRNA from stages 11 through 14 revealed unexpectedly dynamic and complex temporal patterns of translational control (Figure 3A). The relative TEs for some mRNAs increased progressively at each stage of oogenesis (Figure 3A cluster 1), and this set of mRNAs included *cyclin B* (Figure 3B), whose translational control via polyadenylation is reported to be crucial for oocyte maturation (Benoit et al., 2005). Some of the downregulated mRNAs similarly showed progressive decreases in TE through the stages of oocyte maturation (Figure 3A cluster 26). However, more complex patterns of developmental changes in TE were more commonly observed (Figure 3A). Moreover, changes in TE correlated poorly with previously measured changes in protein abundance (Kronja et al., 2014a), potentially due to a large role of posttranslational regulation during this time (Figure 2—figure supplement 1B, $R_s = 0.15$).

The net changes in TE and poly(A)-tail length over the course of oocyte maturation were largely concordant for both translationally up- and downregulated mRNAs, indicating strong coupling between TE and poly(A)-tail length during oocyte maturation (Figure 4A). Nonetheless, a potential deviation from the typical coupled behavior occurred between oocyte stages 13 and 14, when a set of mRNAs that underwent substantial tail-length reductions did not

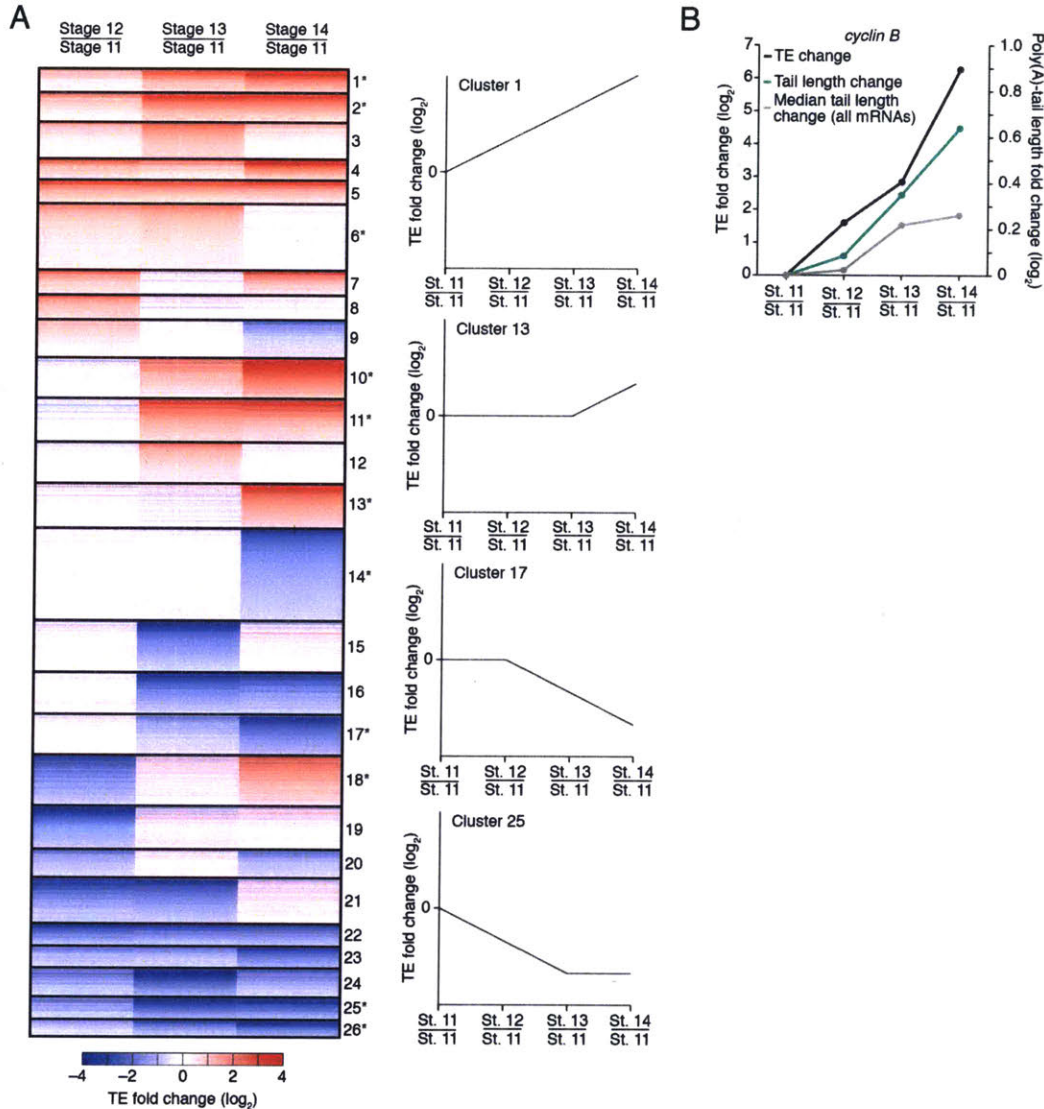


Figure 3. Dynamics of translational regulation during oocyte maturation.

(A) Distinct patterns of TE changes during oocyte maturation. TE values (\log_2) for mRNAs in oocytes from stages 11 through 14 that had ≥ 10.0 RPM in the RNA-seq data for all samples, and ≥ 10.0 RPM in the ribosome-profiling data for at least one sample and > 0.0 RPM in the others were median centered as in Figure 2. The median-centered TEs for stage 12, 13, and 14 oocytes were normalized to those of stage 11 oocytes and then clustered into defined patterns. The heatmap shows clustered TE changes for all mRNAs with ≥ 0.5 or ≤ -0.5 \log_2 fold change between any two samples. Clusters are identified by number, and stylized graphs illustrate the TE dynamics of sample clusters. Clusters that are significantly overrepresented are marked with an asterisk (p value < 0.05 , following Bonferonni correction). (B) Changes in TE (black) and mean poly(A)-tail length (green) of *cyclin B* mRNA during oocyte maturation. TEs (\log_2) were normalized as in Figure 3A, and the median of the changes in mean poly(A)-tail length (\log_2) for all mRNAs over these stages is shown as a grey line. Fold changes (\log_2) calculated relative to stage 11 oocytes are indicated on the respective y axes.

undergo concordant TE reductions (Figure 2, Stage 14 / Stage 13, and Figure 2—figure supplement 2A). These mRNAs tended to have both long poly(A)-tails and efficient translation in stage 13 oocytes, and then remained well translated compared to other mRNAs in stage 14 oocytes, despite undergoing substantial tail shortening (Figure 2—figure supplement 2B). Notable among this set were mRNAs for 22 of the 30 genes encoding subunits of the proteasome (Figure 2—figure supplement 2A and B).

To examine the relationship between tail-length changes and TE changes for mRNAs of key factors in oocyte maturation, we focused on those of three of the most highly upregulated cell-cycle regulators involved in meiotic progression, Cyclins B and B3 and the Cdc20 ortholog Fizzy (Jacobs et al., 1998; Swan and Schupbach, 2007). Between oocyte stages 11 through 14, the TEs of these mRNAs increased 103, 248, and 319 fold (compared to a median TE change of only 1.3 fold), respectively, whereas poly(A)-tail lengths increased 1.6, 2.4, and 2.0 fold (compared to the median of 1.2 fold) (Figure 4A). In addition to these examples, mRNAs from 121 other genes were very poorly translated in stage 11 oocytes ($TE \leq 0.1$) and became ≥ 10 -fold better translated during oocyte maturation (Figure 2—figure supplement 1C). As with *cyclin B*, *cyclin B3* and *fizzy* mRNAs, these also had modest increases in tail length (Figure 2—figure supplement 1C; median tail-length fold change = 1.7 for mRNAs with $TE \leq 0.1$ in stage 11 oocytes and a ≥ 10 -fold increase in TE in stage 14 oocytes), raising the question of the degree to which modest tail-length increases might explain the large TE increases observed during oocyte maturation.

One possibility is that a subset of the mRNAs for a gene underwent extensive polyadenylation, and this change in a subset of mRNAs did not cause much change in the mean tail-length value despite causing a large change in TE. However, for *cyclin B*, *cyclin B3* and *fizzy*

mRNAs, the distribution of poly(A)-tail lengths remained unimodal as overall lengths increased, which indicated that the focus on the mean tail-length change in our analyses did not miss specific regulation of a subset of mRNAs for a gene (Figure 4B). Also helpful for considering the effects of increasing tail lengths during oocyte maturation is the steep relationship between poly(A)-tail length and TE during this developmental period (Figure 2). The steepness of this relationship in stage 11, 12, and 13 oocytes implied a 6- to 7-fold increase in the median TE for mRNAs with a poly(A)-tail length between 70–80 nt relative to those with a poly(A)-tail length between 30–40 nt. In stage 14 oocytes, the poor correlation between tail length and TE resulted in only a 1.8-fold increase in median TE when comparing mRNAs with such tail lengths. Thus during oocyte maturation, the modest increases observed in mean poly(A)-tail lengths presumably explain part, but not all, of the robust translational activation of *cyclin B*, *cyclin B3*, *fizzy*, and mRNAs from many other genes. Moreover, a broad range of TE values was observed for mRNAs with the same mean poly(A)-tail length, presumably as a consequence of additional regulatory processes that are independent of poly(A)-tail length. To the extent that these additional processes act concordantly with tail-length–dependent processes, they would increase the apparent effect of tail lengthening.

Translational regulation during the oocyte-to-embryo transition

Although egg activation occurs rapidly (within ~20 minutes) (Horner and Wolfner, 2008), this developmental transition induced TE changes for mRNAs from many genes, some of which exceeded 30 fold (Figure 2, lower right) (Kronja et al., 2014b). As in the preceding period of oocyte maturation, tail-length changes corresponded to TE changes (Figure 4A and C). Indeed, during egg activation the correlation between tail-length change and TE change exceeded that

observed for any other transition (Figure 1B and Figure 2, lower panels; Figure 4A). Thus during egg activation, changes in poly(A)-tail length can account for a large proportion of the translational changes.

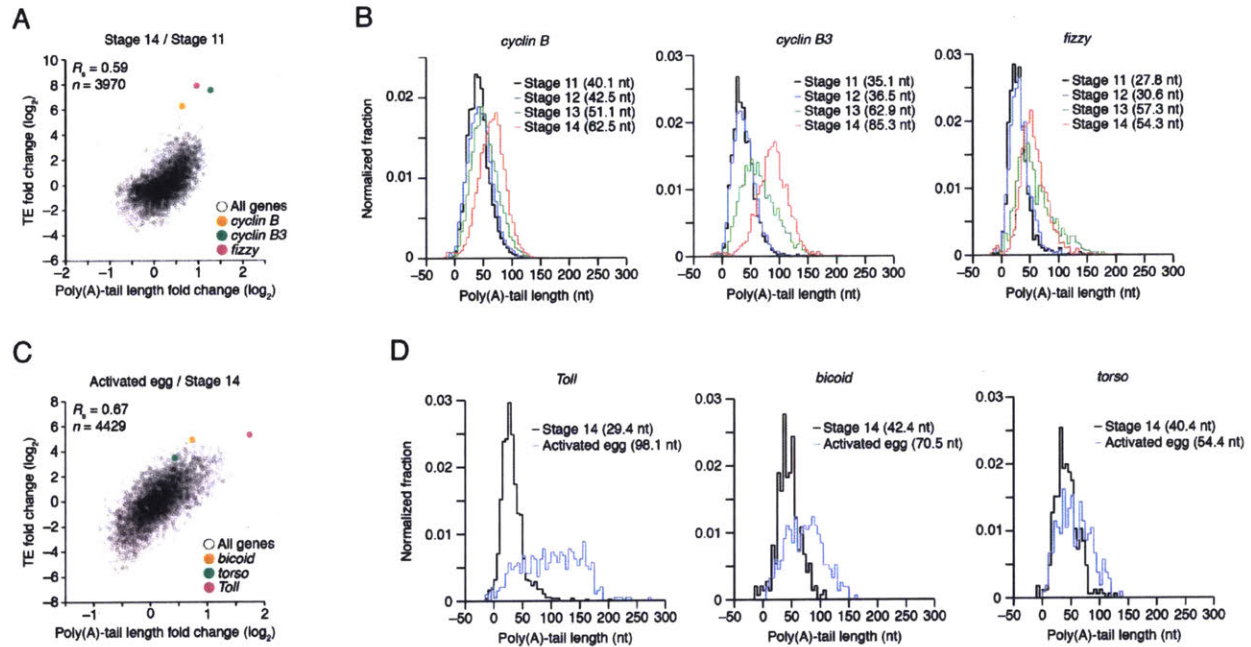


Figure 4. TEs and Poly(A) tail lengths of selected cell-cycle regulators during oocyte maturation.

(A) Relationship between net tail-length and TE changes observed after oocyte maturation, highlighting the behavior of *cyclin B*, *cyclin B3*, and *fizzy*. The TE data for stage 14 oocytes were from Kronja et al. (2014b). Values from stage 14 oocytes are compared to those of stage 11 oocytes (median value used for median centering, 0.389); otherwise, as in Figure 1B. (B) Distributions of poly(A)-tail lengths of mRNAs from *cyclin B* (left), *cyclin B3* (middle), and *fizzy* (right) at the indicated developmental stages. Mean poly(A)-tail lengths are in parentheses. (C) Relationship between tail-length and TE changes observed after egg activation, plotting the same points as in Figure 2 (bottom right), but highlighting the behavior of *Toll*, *bicoid*, and *torso*. (D) Distributions of poly(A)-tail lengths of mRNAs from *Toll* (left), *bicoid* (middle), and *torso* (right) at the indicated developmental stages; otherwise, as in Figure 4B.

mRNAs for the embryonic patterning genes *bicoid*, *torso*, and *Toll* are translationally upregulated during egg activation in a manner reported to depend on changes in poly(A)-tail length (Salles et al., 1994). In our analyses, translation of *bicoid* mRNA was undetectable or detected with very few reads in oocytes from stages 11 through 14, but it became much more

highly translated following egg activation, increasing by 36 fold in the activated egg relative to the stage 14 oocyte (median TE change = 1.2 fold). mRNAs for *torso* and *Toll* were somewhat better translated in stage 14 oocytes, but nonetheless their TEs increased 13 and 46 fold, respectively (Figure 4C). During this time, the *Toll* poly(A)-tail length increased 3.3 fold (median tail length change = 1.1 fold), which corresponded well to the trend observed globally for tail-length and TE changes (Figure 4C), suggesting that its increased tail-length might fully account for its increased TE. However, concordantly acting tail-length-independent processes that increase the apparent effect of tail-length changes might also be influencing *Toll* translation. Indeed, our subsequent experiments with *wispy*-mutant animals showed that in the absence of cytoplasmic polyadenylation, *Toll* underwent substantial translational upregulation during egg activation and remained one of the most efficiently translated mRNAs in the *wispy*-mutant laid egg, despite having a mean tail length of only 9 nt. The poly(A)-tail lengths of *bicoid* and *torso* increased by only 1.7 and 1.3 fold, respectively, putting them outside the global trend (Figure 4C), and the unimodal tail-length distributions showed again that our focus on the mean change did not miss a unique behavior of a subpopulation of transcripts (Figure 4D). Thus, the translational regulation of these mRNAs seems to also involve tail-length-independent mechanisms.

The strong translational upregulation of *bicoid* mRNA, from essentially untranslated in oocytes from stages 11 through 14 to relatively well translated in the activated egg, implied the influence of either a translational repressor acting on *bicoid* mRNA in the oocyte that was relieved at egg activation, or a translational activator acting on it during egg activation. Because *bicoid* mRNA had an average tail length of 42 nt in stage 14 oocytes, which was similar to that of many well translated mRNAs at this stage, we conclude that there was likely a translational

repressor acting on it in oocytes to prevent its translation prior to egg activation. The same is presumably true for other mRNAs that had little to no detectable translation in oocytes but then became efficiently translated following egg activation. For mRNAs such as *torso*, which had somewhat greater translation in stage 14 oocytes but subsequently experienced a TE increase that was out of proportion to its tail-length increase, either possibility might account for its translational upregulation at egg activation. Likewise, the 15-fold and 4-fold translational activation observed between stage 14 oocytes and 0–1 h embryos for *zelda* and *stat92E*, respectively (Figure 1C), were accompanied by only 1.2-fold changes in poly(A)-tail length for both mRNAs (median TE change = 1.2 fold and median tail length change = 1.1 fold), again implying that a component of their activation may have occurred through a tail-length independent mechanism (note that the poly(A)-tail length of *zelda* was quantified by only 31 and 72 tags in stage 14 oocytes and 0–1 h embryos, respectively).

Widespread impact of cytoplasmic polyadenylation on poly(A)-tail lengths but not TE

During oocyte maturation and egg activation, changes in poly(A)-tail length occur through the opposing activities of deadenylases and cytoplasmic poly(A)-polymerases, which shorten and lengthen poly(A) tails, respectively (Weill et al., 2012). Wispy is the cytoplasmic poly(A)-polymerase responsible for poly(A)-tail lengthening at the end of oocyte maturation and continuing through egg activation (Benoit et al., 2008; Cui et al., 2008). Indeed, within this developmental window, most maternal transcripts bind to oligo(dT) with less affinity in *wispy* mutants, indicating that many poly(A) tails shorten in the absence of Wispy (Cui et al., 2013), but the degree to which tail-lengths shorten and the impact on TE had not been previously measured.

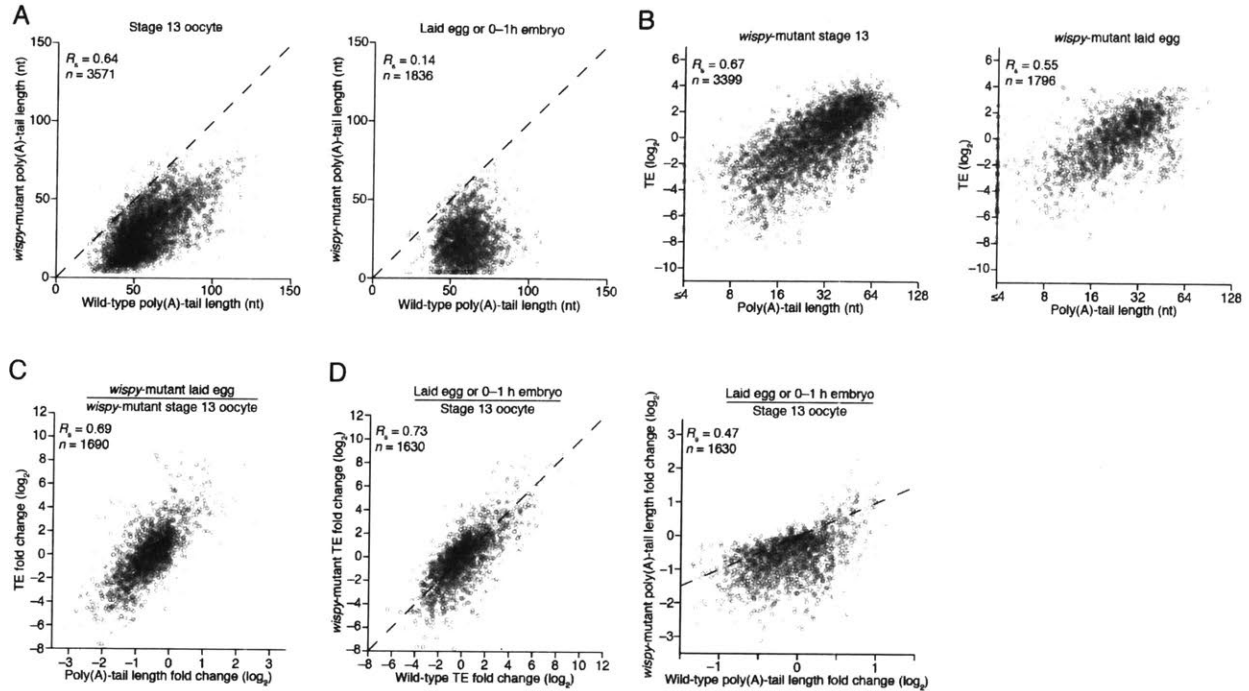


Figure 5. Widespread impact of Wispy on poly(A)-tail lengths but not TE.

(A) Comparison of mean poly(A)-tail lengths in wild-type and *wispy*-mutant stage 13 oocytes (*left*) and in wild-type cleavage-stage embryos and *wispy*-mutant laid eggs (*right*). Plotted are mean poly(A)-tail lengths of mRNAs with ≥ 100 poly(A) tags in both the wild-type and *wispy*-mutant sample at the indicated developmental stage. mRNAs that had mean tail-length values ≤ 4 nt are reported as 4 nt. The dashed line is for $y = x$. (B) Relationship between mean poly(A)-tail length and TE at the indicated developmental stage in *wispy*-mutant oocytes (*left*) and laid eggs (*right*). The TEs were calculated by dividing *wispy*-mutant stage 13 oocyte or laid egg RFP data by wild-type stage 13 oocyte or 0–1 h embryo RNA-seq data, respectively. TE values (\log_2) were median centered (median values in *wispy*-mutant stage 13 oocytes and laid eggs, -1.0551 and 0.3302 , respectively). mRNAs that had mean tail-length values ≤ 4 nt are reported as ≤ 4 nt; otherwise, as in Figure 1B. (C) Relationship between tail-length and TE changes observed between *wispy*-mutant laid eggs and stage 13 oocytes. TE fold-change values (\log_2) were median centered (median value, 0.8132); otherwise, as in Figure 1B. The mRNAs that seemed to have increased poly(A)-tail length over this time tended to have very short poly(A)-tails in stage 13 oocytes, suggesting that their positive fold-change values reflected difficulties in accurately measuring poly(A)-tails < 8 nt using PAL-seq rather than genuine increases in poly(A)-tail length. (D) Comparison of TE changes for wild-type and *wispy*-mutant samples during the OET (*left*) and comparison of tail-length changes for wild-type and *wispy*-mutant samples during the OET (*right*). TE fold-change values (\log_2) were median centered (median values for the wild-type and *wispy*-mutant samples, 0.7207 and 0.8367 , respectively). Dashed line is for $y = x$. Otherwise, this panel is as in Figure 1B.

Our PAL-seq measurements of poly(A)-tail lengths in *wispy*-mutant stage 13 oocytes and laid eggs confirmed the widespread shortening of poly(A) tails in the absence of Wispy (Figure

5A). Of the mRNAs from 3571 genes measured in stage 13 oocytes, 1801 had tails that were at least 50% shorter in *wispy*-mutant oocytes relative to wild-type, and for all but a few mRNAs the poly(A) tail was at least somewhat shorter in the absence of Wispy (Figure 5A). Although *wispy*-mutant eggs do not complete meiosis and thus do not begin nuclear divisions, most of the molecular changes characteristic of the OET are independent of these divisions, motivating a comparison of *wispy*-mutant laid eggs with wild-type early embryos. In *wispy*-mutant laid eggs, 1219 of the 1836 measured mRNAs had poly(A) tails at least 50% shorter than those in the analogous wild-type sample (0–1 h embryos), and again for all but a few mRNAs the poly(A) tail was shorter in the absence of Wispy (Figure 5A). Examination of the mRNAs with tails that were least affected by the absence of Wispy revealed a remarkable enrichment for mRNAs of ribosomal proteins, both in oocytes and laid eggs (Figure 5—figure supplement 1A; $p < 10^{-41}$ and $p < 10^{-44}$, respectively, one-tailed Wilcoxon rank sum test).

Strikingly, the global changes in poly(A)-tail length that occurred in the absence of Wispy did not affect the coupling between poly(A)-tail length and TE. In *wispy*-mutant stage 13 oocytes and laid eggs, poly(A)-tail length and TE were correlated to a similar extent as in wild-type oocytes and embryos (Figure 1B, 2, and Figure 5—figure supplement 1B). These correlations were observed even though the RNA-seq measurements used to calculate TEs were made using poly(A)-selected RNA, which in the *wispy*-mutant samples likely resulted in inflated TE estimates for mRNAs with very short poly(A)-tails, as these would be poorly recovered during the poly(A) selection. Indeed, RNA-seq measurements correlated with poly(A)-tail length in the *wispy*-mutant samples but not the wild-type samples (Figure 5—figure supplement 1C), indicating a poly(A)-selection bias in the *wispy*-mutant RNA-seq data but not in the wild-type data. Because neither developmental stage being considered had appreciable transcription

or RNA degradation, and because of the poly(A)-selection bias in the *wispy*-mutant RNA-seq data, we reasoned that the RNA-seq measurements from corresponding wild-type samples should most accurately reflect mRNA abundances in the *wispy*-mutant samples. Indeed, results obtained from determining TEs for the *wispy*-mutant samples using the *wispy*-mutant ribosome profiling and the wild-type RNA-seq (Figure 5B) were similar to those obtained when only *wispy*-mutant measurements were used (Figure 5—figure supplement 1B), except they had somewhat stronger correlations between poly(A)-tail length and TE, which supported the idea that these modified TEs more accurately reflected the true TEs at these stages.

Importantly, the strong correlation between changes in poly(A)-tail length and changes in TE that occurred at egg activation in wild-type animals was retained in *wispy*-mutant flies (Figures 1B and 5C, respectively), indicating that changes in the length of the poly(A)-tail—not the act of cytoplasmic polyadenylation per se—is what influences TE. Moreover, the changes in TE that occurred during egg activation in *wispy*-mutant samples strongly correlated with those in the wild-type samples, as did the relative changes in poly(A)-tail length (Figure 5D). Thus, although Wispy lengthens the poly(A)-tails of most mRNAs during egg activation, our results indicated that this lengthening was largely dispensable for reshaping TEs during egg activation, as the rank order of tail-length and TE changes at this time were largely preserved in the *wispy*-mutant samples.

The observation that cytoplasmic polyadenylation was not needed to generate the rank order of tail-length and TE changes for most mRNAs warranted further investigation, particularly when considering both the critical influence of cytoplasmic polyadenylation for generating these differences during vertebrate development and the importance of Wispy for relieving meiotic arrest during the OET. Perhaps some mRNAs most affected by the loss of Wispy dropped out of

our tail-length analysis because they had very short poly(A)-tails in the *wispy* mutant and thus were poorly recovered in our PAL-seq libraries. Although the tail lengths of such mRNAs might have been difficult to measure by PAL-seq, the RPF measurements of mRNAs for all genes can be analyzed without consideration of the corresponding tail-length or mRNA abundance measurements, either of which might selectively lose the set of mRNAs most dependent on Wispy for their regulation. The RPFs in wild-type and *wispy*-mutant stage 13 oocytes were well correlated, and the same was true for 0–1 h wild-type embryos compared with *wispy*-mutant laid eggs ($R_s = 0.76$ for both, Figure 5—figure supplement 1D). Although globally the RPFs were largely preserved in the *wispy*-mutant samples, some mRNAs had many RPFs in the wild-type samples but none in the *wispy*-mutant samples (Figure 5—figure supplement 1D). Included among these was *cortex*, a known target of Wispy, the upregulation of which is essential for the OET (Chu et al., 2001; Benoit et al., 2008).

We conclude that during egg activation in *Drosophila*, cytoplasmic polyadenylation extends the tails of nearly all mRNAs but does not play the major role in generating the relative differences in tail lengths and TEs observed between mRNAs, which implies that poly(A)-tail shortening instead plays the major role in generating these differences. Nonetheless, the mRNAs responsible for the *wispy*-mutant phenotype might be among those that are most strongly dependent on Wispy for their translational state.

PNG kinase complex up- and downregulates translation through changes in poly(A)-tail length

The PNG kinase, which is required for nuclear divisions in the early embryo, influences the translational up- or downregulation of hundreds of different mRNAs at egg activation (Figure

6A) (Kronja et al., 2014b), presumably through its regulation of translational regulators, such as SMG (Tadros et al., 2007). To characterize the extent to which these downstream regulators might use tail-length changes to influence TE, we profiled poly(A)-tail lengths and TEs in *png*-mutant stage 14 oocytes and activated eggs. Poly(A)-tail lengths and TEs were highly correlated between the *png*-mutant and wild-type stage 14 oocytes, whereas the correlation was substantially lower in *png*-mutant and wild-type activated eggs (Figure 6—figure supplement 1A). Additionally, the poly(A)-tail lengths and TEs in the *png*-mutant activated eggs were quite similar to those in the *png*-mutant stage 14 oocytes, whereas this was not the case for the same comparisons among wild-type samples (Figure 6—figure supplement 1B). The requirement of PNG for achieving the gene-expression profile of the wild-type activated egg but not the stage 14 oocyte was consistent with PNG kinase becoming active during egg activation to serve as a master regulator of the OET (Shamanski and Orr-Weaver, 1991; Kronja et al., 2014b).

The set of mRNAs dependent on PNG for their translational downregulation was predominantly those that underwent tail-length shortening during egg activation in wild-type samples (median TE log₂ fold change of -3.03 after median centering and median log₂ tail-length fold change of -0.35) (Figure 6B). In *png*-mutant embryos, these mRNAs failed to be deadenylated, as expected if tail-length shortening mediated their translational repression (median TE log₂ fold change of 0.10 after median centering and median log₂ tail-length fold change of 0.32, which closely resembled the median values for all mRNAs) (Figure 6B). Correspondingly, the set of mRNAs dependent on PNG for their translational upregulation was predominantly those that underwent tail-length extension during wild-type egg activation (median TE log₂ fold change of 2.98 after median centering and median log₂ tail-length fold change of 0.62), and these mRNAs failed to undergo tail lengthening in *png* mutants, again as

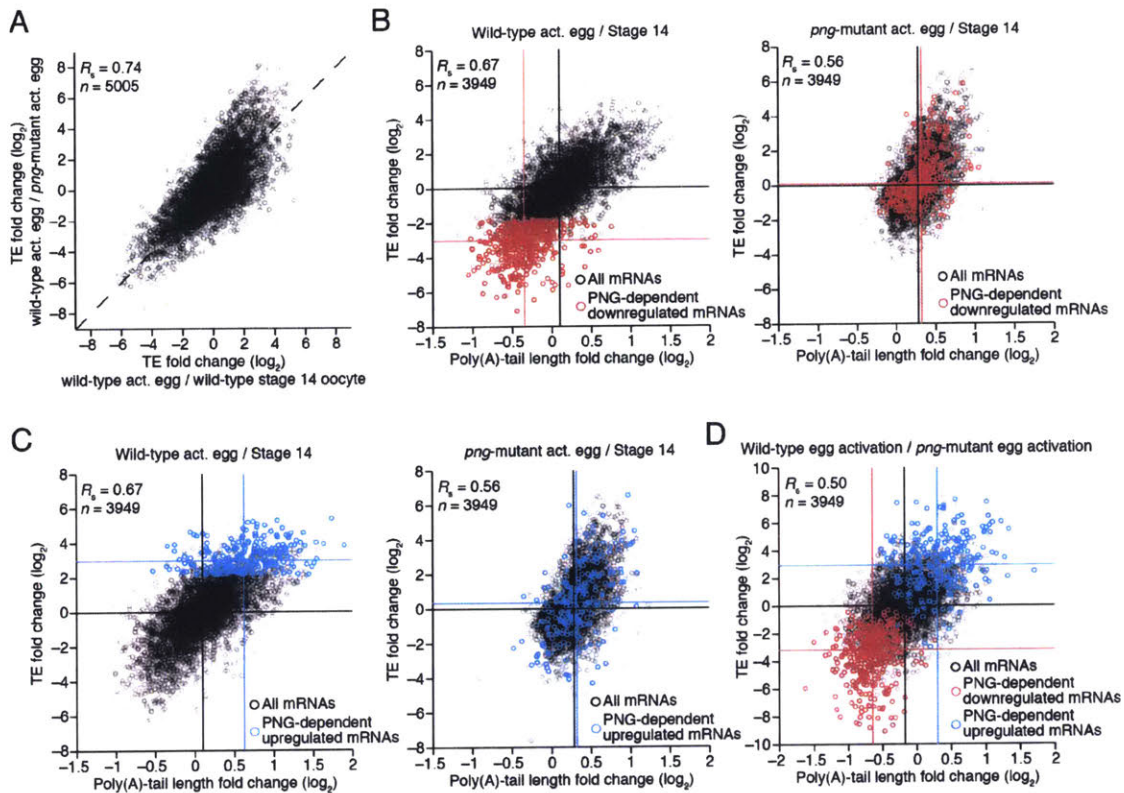


Figure 6. Widespread translational regulation by PAN GU mediated primarily by changes in poly(A)-tail length.

(A) Relationship between the TE changes in wild-type activated eggs relative to wild-type stage 14 oocytes and in wild-type activated eggs relative to *png*-mutant activated eggs. TE fold-change values (\log_2) were median centered (median values for wild-type activated egg relative to stage 14 oocyte and for wild-type activated egg relative to *png*-mutant activated egg, 0.3065 and 0.0881, respectively). Results are plotted for all mRNAs that had ≥ 10.0 RPM in the RNA-seq data of all samples, and ≥ 10.0 RPM in the ribosome profiling data of at least one of the two samples being compared and > 0.0 RPM in the other. mRNAs ≥ 4 -fold up- or downregulated in both comparisons were defined as the PNG-dependent up- or downregulated mRNAs, respectively. The TE data for wild-type stage 14 oocytes and activated eggs and *png*-mutant activated eggs were from Kronja et al. (2014b). Dashed line is for $y = x$. (B) Relationship between mean tail-length changes and TE changes after egg activation for wild-type activated eggs (*left*; modified from Figure 2) and *png*-mutant activated eggs (*right*). The mRNAs with PNG-dependent downregulation are highlighted in red (Supplementary File 3), analyzing and highlighting the same mRNAs in both plots. TE fold-change values (\log_2) were median centered (median for wild-type and *png*-mutant samples, 0.1366 and 0.6448, respectively). Black lines indicate median values for all mRNAs, and red lines indicate median values for the PNG-dependent downregulated mRNAs. The TE data for wild-type stage 14 oocytes and activated eggs and *png*-mutant activated eggs are from Kronja et al. (2014b). Otherwise, this panel is as in Figure 1B. (C) The plots of panel B, highlighting the PNG-dependent upregulated mRNAs in blue (Supplementary File 3). (D) Relationship between tail-length and TE changes during egg activation in wild-type relative to the *png*-mutant samples, analyzing and highlighting the same mRNAs as in panels B and C. TE fold-change values (\log_2) were median centered (median value, -0.727). Otherwise, this panel is as in Figure 6B.

expected if tail-lengthening mediated their translational activation (median TE log₂ fold change of 0.36 after median centering and median log₂ tail-length fold change of 0.32, which closely resembled the median values for all mRNAs) (Figure 6C). mRNAs that were translationally up- or downregulated by PNG also typically increased or decreased in poly(A)-tail length, respectively (Figure 6D), and the tail-length changes for either the up- or downregulated sets of mRNAs were largely dependent on PNG (Figure 6B and C), which showed that PNG plays a central role in regulating polyadenylation and deadenylation for the hundreds of mRNAs present at egg activation.

SMG-mediated repression occurs primarily through tail shortening and explains some but not all of the PNG-mediated translational repression

Because much of the PNG-mediated translational repression is attributed to its derepression of *smg* translation (Tadros et al., 2007), we examined the impact of SMG on the set of mRNAs that depended on PNG for tail-length shortening and translational repression. SMG is reported to downregulate translation of its target mRNAs through either direct recruitment of the deadenylation machinery or recruitment of CUP, which directly represses translation initiation and can also promote deadenylation (Nelson et al., 2004; Semotok et al., 2005; Igreja and Izaurralde, 2011). Our results showed that in the absence of SMG, the poly(A)-tail lengths for the set of PNG-dependent downregulated mRNAs were less shortened during egg activation (median log₂ tail-length fold changes of -0.22 and -0.04 for the wild-type and *smg* mutant, respectively, relative to the median of all mRNAs, $p < 10^{-27}$, one-tailed Wilcoxon rank sum test), and this coincided with their partial translational derepression (median log₂ TE fold change of -2.2 and -1.2, respectively, after median centering, $p < 10^{-32}$, one-tailed Wilcoxon rank sum test)

(Figure 7A). In contrast, no significant difference was observed for either poly(A)-tail length or TE changes for the PNG-dependent upregulated mRNAs (median \log_2 tail-length fold changes of 0.27 for both the wild-type and *smg* mutant relative to the median of all mRNAs, and median \log_2 TE fold changes of 2.8 and 2.7, respectively, after median centering for the wild-type and *smg* mutant, $p = 0.89$ and 0.37 for tail-length or TE changes, respectively, two-tailed Wilcoxon rank sum test), consistent with SMG functioning only as a repressor (Figure 7B). In general, mRNAs that were translationally repressed by SMG also decreased in poly(A)-tail length, which indicated that recruitment of deadenylases was the dominant mechanism of SMG-dependent translational repression occurring during egg activation and early embryogenesis (Figure 7C). Additionally, in the absence of SMG, a set of mRNAs reported to be bound by SMG in the early embryo (Chen et al., 2014) had longer poly(A) tails and were better translated following the OET (median \log_2 tail-length fold changes of -0.21 and -0.04 for the wild-type and *smg* mutant, respectively, relative to the median of all mRNAs, and median \log_2 TE fold changes of -1.0 and -0.15 , respectively, after median centering for the wild-type and *smg* mutant, $p < 10^{-16}$ and 10^{-11} for tail length or TE changes, respectively, one-tailed Wilcoxon rank sum test) (Figure 7—figure supplement 1).

Taken together, our results indicate that SMG mediates much of the PNG-dependent translational repression occurring during egg activation and early embryogenesis, with most of this repression occurring through poly(A)-tail shortening. Nevertheless, the residual repression observed in the absence of SMG (Figure 7A) indicates that one or more additional PNG-activated translational repressors also operate at this time.

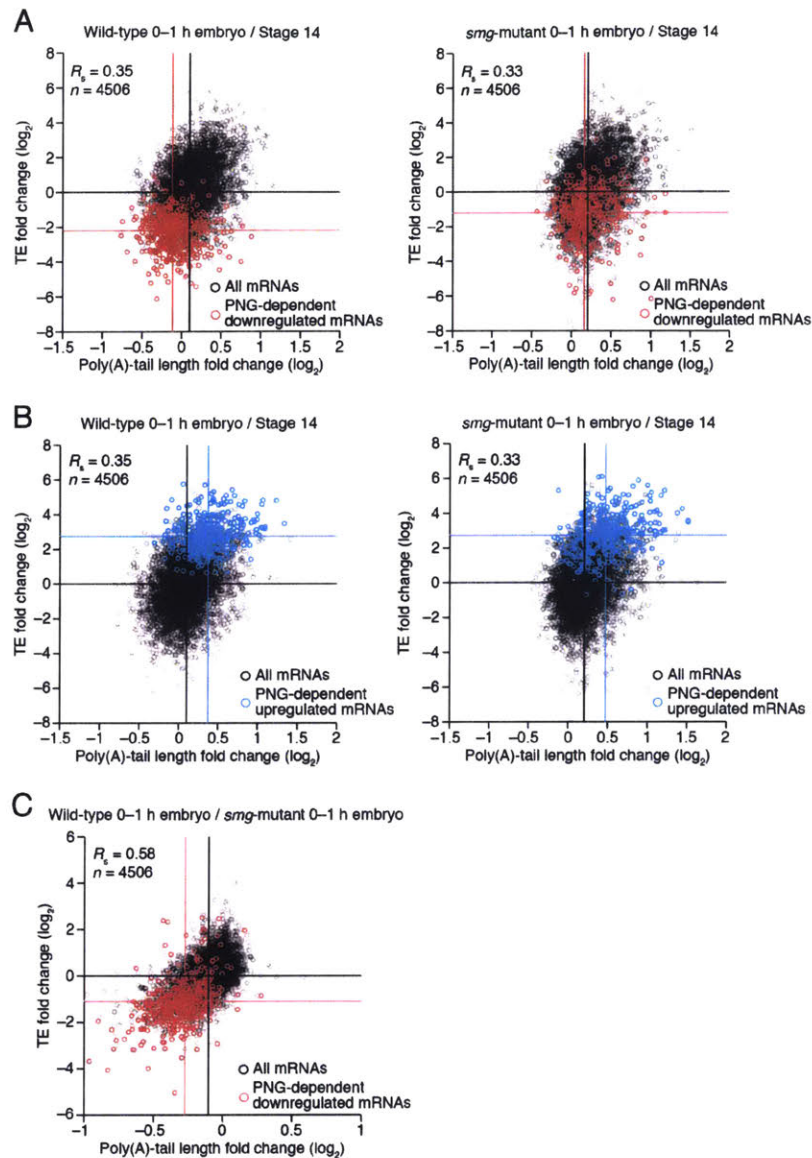


Figure 7. Translational regulation by Smaug mediated primarily by changes in poly(A)-tail length.

(A) Relationship between mean tail-length changes and TE changes during the OET for wild-type 0–1 h embryos (*left*) and *smg*-mutant 0–1 h embryos (*right*), comparing wild-type or *smg*-mutant 0–1 h embryos to wild-type stage 14 oocytes. The wild-type plot is redrawn from Figure 1B, but includes only the mRNAs that also passed the cutoffs for the *smg*-mutant comparison. TE fold-change values (\log_2) were median centered (median for the wild-type and *smg*-mutant samples, 0.175 and -0.0221 , respectively). The PNG-dependent downregulated mRNAs are in red (Supplementary File 3), analyzing and highlighting the same mRNAs in both plots. Otherwise, this panel is as in Figure 6B. (B) The plots of panel A, highlighting the PNG-dependent upregulated mRNAs in blue (Supplementary File 3). (C) Relationship between tail-length and TE changes for wild-type 0–1 h embryos relative to *smg*-mutant 0–1 h embryos, analyzing and highlighting the same mRNAs as in panel A. TE fold-change values (\log_2) were median centered (median value, 0.2634). Otherwise, this panel is as in Figure 6D.

Discussion

Our results from *Drosophila* oocytes and embryos increase the number of cellular environments known to possess global coupling between poly(A)-tail length and TE, which now include several stages of *Drosophila* oocytes as well as *Drosophila*, zebrafish, and *Xenopus* embryos prior to the onset of gastrulation. At least as early as the primary meiotic arrest point, changes in poly(A)-tail length largely explain changes in TE, and this regulatory strategy continues through early embryonic development. A switch around the onset of gastrulation uncouples poly(A)-tail length and TE in *Drosophila*, mirroring the developmental timing of the same uncoupling in zebrafish and *Xenopus* embryos. The uncoupling of TE from poly(A)-tail length is thus a broadly conserved developmental switch in the nature of translational control, occurring around the onset of gastrulation in both vertebrate and invertebrate embryos.

Changing the length of a poly(A) tail would influence both the number of poly(A)-binding proteins (PABPs) that could bind the tail and the probability that at least one PABP would be bound, with the former being more relevant in conditions of saturating PABP and the latter being more relevant in conditions in which tails are competing for PABP. If the number of PABPs that bind the tail were most relevant, then TE changes might have correlated better with absolute tail-length changes than with relative tail-length changes. However, repeating our analyses using absolute tail-length changes rather than \log_2 tail-length fold changes yielded the same results for the respective comparisons, with nearly identical R_s values (Figure 7—figure supplement 2), suggesting that PABP is limiting in the developmental stages in which tail length and TE are coupled, such that mRNAs are competing for PABP.

We presume that the correlation between poly(A)-tail length and TE that we observed in oocytes and pre-gastrulation embryos reflects a causal relationship, i.e., that changes in poly(A)-

tail length at least partially cause the corresponding changes in TE. The other possibilities are that changes in TE cause changes in tail length or that these two measurements merely correlate without having a causal relationship. Because previous single-gene studies clearly demonstrate a causal role for cytoplasmic polyadenylation, and poly(A)-tail length itself, in regulating the translation of mRNAs during these developmental times (McGrew et al., 1989; Salles et al., 1994; Barkoff et al., 1998), we interpret our observation of a global correlation between poly(A)-tail length and TE as reflecting a causal relationship between the two wherein tail-length changes cause TE changes. Given this interpretation, our results indicate that widespread changes in poly(A)-tail length broadly reshape the translational profiles of oocytes and early embryos.

Despite this strong qualitative conclusion, we cannot provide quantitative estimates of the proportion of each TE change that is attributable to each tail-length change, as additional translational regulatory mechanisms presumably act to reinforce or oppose the effects of tail-length–dependent translational regulation. Indeed, we suspect that tail-length–independent processes often act concordantly with tail-length–dependent processes and thereby amplify the apparent effects of tail-length changes. Such amplification explains why some mRNAs, including *cyclin B*, *cyclin B3*, *fizzy*, *zelda*, *stat92E*, *bicoid*, and *torso*, undergo translational activation that seems out of proportion to their tail-length changes.

A primary motivation for using *Drosophila* in our study was the opportunity to examine the impact of Wispy, PNG, and SMG on poly(A)-tail length and TE. Of particular interest was Wispy, the cytoplasmic poly(A) polymerase that acts during late oogenesis and early embryogenesis, and is essential for the OET (Benoit et al., 2008; Cui et al., 2008). Analysis of *wispy* mutants revealed key insights into both the mechanism of translational activation and the importance of tail shortening (as opposed to tail lengthening) in specifying translational control

in *Drosophila* maturing oocytes and early embryos. Although poly(A)-tail lengths were typically shorter in *wispy*-mutant stage 13 oocytes and laid eggs compared to the corresponding wild-type samples, tail length and TE were coupled in both of these *wispy*-mutant contexts, and the relative changes in poly(A)-tail lengths during the OET in wild-type samples were largely preserved in the *wispy*-mutant samples, as were the changes in TE. Thus, the act of cytoplasmic polyadenylation is not necessary for the coupling of poly(A)-tail length to translation. Moreover, cytoplasmic polyadenylation is largely dispensable for specifying which mRNAs are to be translationally activated or repressed, implying that in *Drosophila*, selective deadenylation is more important for imparting this specificity. This greater importance for deadenylation does not preclude a role for selective cytoplasmic polyadenylation in translational control in *Drosophila*—differences were observed when comparing TE changes in wild-type and *wispy*-mutant samples (Figure 5D), and when comparing RPF measurements between wild-type and *wispy*-mutant samples (Figure 5—figure supplement 1D), including for *cortex*, a critical OET regulator. However, the greater importance of selective deadenylation for specifying translational control in *Drosophila* differs from the prevailing paradigm of vertebrates, which centers on selective cytoplasmic polyadenylation of mRNAs with cytoplasmic polyadenylation elements (Pique et al., 2008; Weill et al., 2012). This suggests either a fundamental difference between invertebrates and vertebrates or that an underappreciated component of translational control in vertebrates occurs through more global cytoplasmic polyadenylation with selective deadenylation.

Despite the surprisingly modest perturbation of translational regulation that we observed in *wispy*-mutant laid eggs during egg activation, these laid eggs are blocked at the OET because of defective progression through meiosis and failure of pronuclei to fuse (Benoit et al., 2008; Cui et

al., 2008). Why does this developmental arrest occur if translational regulation is largely preserved during egg activation in *wispy* mutants? One possibility is that the absolute level of translation in *wispy*-mutant eggs is reduced such that they make an inadequate amount of one or more essential protein. The other possibility rests on the observation that although TE changes and RPFs were remarkably concordant in the *wispy*-mutant and wild-type samples, they were not identical, and thus an aberrant change in expression for one or more key factor, such as *cortex*, might be sufficient to trigger the developmental arrest of *wispy*-mutant laid eggs.

Activity of the PNG kinase during egg activation affects most mRNAs that undergo translational activity changes at this developmental transition, both those that are up- and downregulated (Kronja et al., 2014b). Although the direct targets of PNG kinase remain to be identified, PNG is required to relieve the repressive effects of PUM (Vardy and Orr-Weaver, 2007a) and thereby derepresses translation of *smg* and other mRNAs (Tadros et al., 2007). In wild-type activated eggs the mRNAs dependent on PNG for translational up- or downregulation exhibited corresponding increases or decreases in poly(A)-tail length, and these tail-length changes did not occur in *png* mutants. Thus, for both translationally activated and translationally repressed mRNAs, the downstream effectors of PNG appear to regulate translation primarily through changes in poly(A)-tail length.

Translational repressors in addition to SMG have been proposed to contribute to PNG-dependent regulation (Tadros et al., 2007). Our results are consistent with the idea that PNG activates at least one additional translational repressor, as the set of mRNAs dependent on PNG for their translational downregulation were still somewhat translationally repressed and deadenylated relative to all other mRNAs in *smg*-mutant embryos, although less than in wild-type embryos. Additionally, although SMG is reported to act partly through a mechanism that

does not involve deadenylation (Nelson et al., 2004; Semotok et al., 2005), essentially all the derepression observed in the *smg*-mutant embryos was accompanied by tail-length increases. Perhaps at other times SMG causes direct translational repression, but in 0–1 h embryos it seems to act predominantly through deadenylation.

Within the overall framework in which tail-length changes cause the up- or downregulation of mRNAs for a vast repertoire of different genes, we discovered unanticipated complexities and exceptions. One unanticipated complexity was the evidence for tail-length-independent processes that appear to amplify the tail-length-dependent effects, as already discussed. Additional unanticipated complexity was observed during oocyte maturation, in which surprisingly diverse patterns of regulatory dynamics produce the ultimate changes in TE. Single-gene studies in *Xenopus*, which show that *c-mos* mRNA becomes polyadenylated and translated prior to the mRNAs for several *cyclin B* genes (Richter, 2007; Pique et al., 2008), hint that the complex landscape of translational control observed in *Drosophila* extends to vertebrate oocyte maturation. The diverse temporal behaviors of translational regulation during oocyte maturation might reflect the evolving demands on the maturing oocyte, or perhaps simply differences in the timing of expression of RNA-binding proteins that interact with these different sets of mRNAs. Regardless, the observation of different temporal patterns of regulation indicates that rather than a being a sudden switch, oocyte maturation in *Drosophila* is a progressive process, whereby mRNAs are translationally activated or repressed at different times throughout this developmental transition.

Another surprise was the identification of a set of mRNAs for which tail-length changes did not correspond to TE changes during oocyte maturation. This exception to the general pattern occurred for the mRNAs that had undergone deadenylation between oocyte stages 13 and 14 yet

not only remained stable but also continued to be translated with relatively high efficiency (Figure 2—figure supplement 2). To escape the general regulatory environment of the cytoplasm, these mRNAs might exploit unique interactions with the translational machinery analogous to the translational activation of histone mRNAs through stem-loop binding protein (Cakmakci et al., 2008), or they might localize to a subcellular compartment that lacks coupling. The deadenylation of these mRNAs during oogenesis might prime them for rapid degradation during embryogenesis. Most mRNAs from genes encoding proteasome subunits fell into this class. Previous comparisons of the proteome and translation changes suggest that extensive proteolysis occurs at egg activation (Kronja et al., 2014b). Upregulation of the proteasome in late oogenesis might necessitate a mechanism to downregulate it rapidly during embryogenesis, perhaps facilitated by pre-deadenylating the corresponding mRNAs before egg activation.

The molecular mechanism that couples TE with poly(A)-tail length is unknown, as is the mechanism of the uncoupling that occurs at gastrulation. Nevertheless, our results, which show that widespread changes in poly(A)-tail length broadly reshape translational activity during oocyte and embryo development, demonstrate the importance of this coupling mechanism and provide a resource that documents the thousands of affected genes with information on the timing, magnitude, and inferred consequences of the tail-length changes for mRNAs from each of these genes. Moreover, the identification of unanticipated complexities and exceptions reveals the rich posttranscriptional regulatory landscape of oocytes and early embryos and points to additional mechanisms that operate for some mRNAs in specific settings, which can now be targeted for investigation.

Acknowledgments

We thank Mariana Wolfner for providing stocks, the Whitehead Genome Sequencing Core for sequencing, and Anthony Mahowald for comments on the manuscript. This work was supported by NIH grants GM39341 and GM118098 (T.L.O.-W.) and GM067031 (D.P.B.). A.O.S. was supported by NIH Medical Scientist Training Program fellowship T32GM007753, and I.K. was supported by the Feodor Lynen Postdoctoral Fellowship from the Alexander von Humboldt Foundation. T.L.O.-W. is an American Cancer Society Research Professor, and D.P.B. is an investigator of the Howard Hughes Medical Institute.

Materials and methods

Drosophila Stocks

All flies were kept at 18, 22 or 25 °C according to standard procedure (Greenspan, 1997).

Oregon R flies were used as a wild-type control. The null *png*¹⁰⁵⁸ allele was previously described (Shamanski and Orr-Weaver, 1991; Fenger et al., 2000). The *wispy* hemizygote mutants were derived from a cross between *wisp*⁴¹/*FM6* (Cui et al., 2008) (kindly provided by Mariana Wolfner, Cornell) and *Df(1)RA47/FM7c* flies (BL961, obtained from the Bloomington Drosophila Stock Center). The *smaug* hemizygote mutants resulted from a cross between *smg*¹/*TM3, Sb[1]* (BL5930, obtained from the Bloomington Drosophila Stock Center) and *Df(3L)Scf-R6, Diap1[1] st[1] cu[1] sr[1] e[s] ca[1]/TM3, Sb[1]* flies (BL4500, obtained from the Bloomington Drosophila Stock Center). Unfertilized eggs were collected from crosses to *twine*^{HB5} mutant males, which fail to make sperm (Courtot et al., 1992).

Oocyte and Embryo collection

Egg chambers were hand-dissected in Grace's Unsupplemented Insect Media (Gibco) from three-day old flies that had been fattened for two days with wet yeast at 22 °C. Approximately 300 egg chambers were collected per oogenesis stage. Oocyte stages 11 through 14 were distinguished using morphological criteria (Spradling, 1993). Activated eggs were collected as previously described (Kronja et al., 2014b). Samples were transferred to RPF lysis buffer (10 mM Tris-HCl, pH 7.4, 5 mM MgCl₂, 100 mM KCl, 2 mM DTT, 100 µg ml⁻¹ cycloheximide, 1% Triton X-100, 500 U ml⁻¹ RNasin Plus, and protease inhibitor (1x complete, EDTA-free, Roche); wild-type stage 14 oocyte and activated egg samples were transferred to RPF lysis buffer lacking

cycloheximide), centrifuged at 10000 rpm for 5 min, and the supernatants were collected and flash frozen in liquid nitrogen to be stored at -80°C . Ultimately, the yield was approximately 40-50 μg of total RNA in 70-120 μl lysate.

Embryos were collected at 25°C by discarding two initial one-hour collections to avoid collecting embryos held within females for prolonged time. Only the subsequent 0–1 h collection was then processed for sequencing analysis. The same embryo collection strategy was applied to *smg* or *wispy* mutants, though for the *wispy* mutant the collection was extended to 0–2 h as one hour was too short to gather a sufficient number of laid eggs for sequencing. To obtain *OrR* 2–3 h, 3–4 h and 5–6 h embryo samples, 0–1 h embryos (collected after two previous one-hour collections were discarded) were incubated in a “humid chamber” at 25°C for additional 2, 3 or 5 h, respectively.

Embryo Staging

Drosophila females can hold developing embryos or lay unfertilized eggs, and so to ensure that the vast majority of the embryos processed for downstream analyses were in the expected developmental stage, a fraction of each embryo collection (approximately 10-20%) was stained with DAPI. This approach was also used to confirm that the embryos laid by mutant mothers (*png*, *smg*, or *wispy*) were developmentally arrested as previously described. Briefly, collected embryos were dechorionated in 50% bleach for 3 min at room temperature, extensively washed with water, dried and then transferred into a scintillation vial that contained methanol. To remove the vitelline membrane from the embryos, an equal volume of heptane was added and the vial was vigorously shaken by hand for 2 min. Next, embryos were fixed overnight in methanol at 4°C . Embryos were then sequentially rehydrated in a mix of methanol:PBS (9:1, 3:1, 1:1,

1:3) for at least 5 min per step. After a wash in PBS, embryos were incubated for 15 min in a $1 \mu\text{g ml}^{-1}$ solution of DAPI in PBS, followed by 30 min wash in PBS containing 0.1% Triton X-100. Finally, after removing all the washing solution, samples were mounted in Vectashield.

Transcript Models

Reference transcript annotations were downloaded from the UCSC Genome browser (dm6 in refFlat format), and for each gene the longest transcript was chosen as the initial representative transcript model. Non-coding genes and overlapping genes on the same strand were removed from this analysis. PAL-seq tags from all wild-type samples were combined, filtered to remove tags with a poly(A) tail <20 nt, and then aligned with STAR (Dobin et al., 2013). 3' UTRs were annotated by identifying positions where ≥ 2 PAL-seq tags aligned at the same position between the stop codon of one gene and the transcription start site of the neighboring gene on the same strand. Isoforms were annotated as described previously (Jan et al., 2011), using a 60 nt window and a maximum 3'-UTR length of 4100 nt. If a new 3' terminus was annotated that was distal to the initial model, the 3' end of the model was extended to the distal isoform. If not, the initial representative transcript model was used. The resulting set of representative transcript models (referred to as “mRNAs”) used for our analyses is available at <http://www.ncbi.nlm.nih.gov/geo> under accession number GSE83616. Models for which the selected 3'-end isoform (either the initial refFlat 3' end or an extension based on our re-annotation) overlapped with a snoRNA or snRNA were included in the analysis of sequencing data but excluded from figures.

Ribosome Footprint Profiling, RNA-seq, and PAL-seq

Samples were split and prepared for ribosome-footprint profiling and RNA-seq as described (Kronja et al., 2014b), as well as for PAL-seq as described (Subtelny et al., 2014). Detailed protocols are available at <http://bartellab.wi.mit.edu/protocols.html>. The wild-type stage 14 oocyte and activated egg samples were prepared using RPF lysis buffer and sucrose gradients that did not contain cycloheximide, all other samples were prepared using RPF lysis buffer and sucrose gradients that contained cycloheximide. Because TE changes between wild-type stage 14 oocytes and activated eggs for samples prepared with and without cycloheximide correlated very well (Kronja et al., 2014b), the omission of cycloheximide did not affect our observations pertaining to these stages. RPF and RNA-seq reads were mapped to ORFs of representative transcript models as described, which excluded the first 50 nt of each ORF to eliminate signal from ribosomes that initiated after adding cycloheximide (Subtelny et al., 2014). PAL-seq tags were mapped to 3' UTRs of representative transcript models, and PAL-seq fluorescence intensities were converted to tail lengths as described (Subtelny et al., 2014).

Clustering TE dynamics

TE values (\log_2) for stage 11, 12, 13, and 14 oocytes, or for stage 14 oocytes, 0–1 h embryos, and 2–3 h embryos were normalized by the median TE value for the mRNAs in the corresponding sample from Figure 2 or Figure 1B, respectively, and were then entered into the Short Time-series Expression Miner (STEM) (Ernst and Bar-Joseph, 2006), which normalized the data relative to the first sample and then used the STEM clustering method to identify significantly enriched profiles (p value < 0.05 , following Bonferonni correction). The default settings were used, except a maximum unit change of one in model profiles was allowed between time points, and only mRNAs with ≥ 0.5 or ≤ -0.5 \log_2 fold change between any two samples

were placed in a cluster. A heatmap of all expression patterns was generated in Matlab using the normalized data.

Sequencing data

The RNA-seq and ribosome-footprint profiling data for wild-type stage 14 oocytes and activated eggs, and *png*-mutant activated eggs were previously published (Kronja et al., 2014b) and were reanalyzed for this study. The raw data for these previously published datasets are available at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE52799, and the processed data files used for this study are under accession number GSE83616. The PAL-seq data for those samples had not been published and are available along with all of the other RNA-seq, ribosome profiling, and PAL-seq data and processed data files under accession number GSE83616. All of the processed data are also available in Supplementary File 2.

Statistical analysis

Unless noted otherwise, all correlations report the Spearman correlation coefficient, and all statistical tests were two-sided. Comparing the TE changes observed between biological replicates for wild-type stage 14 oocytes and activated eggs correlated very well ($R_s = 0.94$) (Kronja et al., 2014b), even though cycloheximide was omitted from one of the replicates. Likewise, TE of activated egg highly correlated with that of 0–1 h embryo ($R_s = 0.88$). Although we did not analyze replicates of PAL-seq measurements in the current study, previous analysis of biological replicates demonstrate the reproducibility mean tail lengths measured using this technique ($R_s = 0.83$). Similar correlations were observed in the current study when comparing

stages for which little if any differences in tail lengths were expected, including activated egg and 0–1 h embryo ($R_s = 0.86$), stage 11 and stage 12 oocyte ($R_s = 0.94$), stage 12 and stage 13 oocyte ($R_s = 0.88$), as well as *png* and wild-type oocytes at stage 14 ($R_s = 0.86$), which was prior to PNG activation (Supplementary File 1 and Figure 6—figure supplement 1A).

References

- Barckmann, B., and Simonelig, M. (2013). Control of maternal mRNA stability in germ cells and early embryos. *Biochim Biophys Acta* 1829, 714-724.
- Barkoff, A., Ballantyne, S., and Wickens, M. (1998). Meiotic maturation in *Xenopus* requires polyadenylation of multiple mRNAs. *Embo J* 17, 3168-3175.
- Benoit, B., Mitou, G., Chartier, A., Temme, C., Zaessinger, S., Wahle, E., Busseau, I., and Simonelig, M. (2005). An essential cytoplasmic function for the nuclear poly(A) binding protein, PABP2, in poly(A) tail length control and early development in *Drosophila*. *Dev Cell* 9, 511-522.
- Benoit, P., Papin, C., Kwak, J.E., Wickens, M., and Simonelig, M. (2008). PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in *Drosophila*. *Development* 135, 1969-1979.
- Besse, F., and Ephrussi, A. (2008). Translational control of localized mRNAs: restricting protein synthesis in space and time. *Nat Rev Mol Cell Biol* 9, 971-980.
- Cakmakci, N.G., Lerner, R.S., Wagner, E.J., Zheng, L., and Marzluff, W.F. (2008). SLIP1, a factor required for activation of histone mRNA translation by the stem-loop binding protein. *Mol Cell Biol* 28, 1182-1194.
- 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. *Mol Cell* 53, 1044-1052.
- Chen, J., Melton, C., Suh, N., Oh, J.S., Horner, K., Xie, F., Sette, C., Blelloch, R., and Conti, M. (2011). Genome-wide analysis of translation reveals a critical role for deleted in azoospermia-like (*Dazl*) at the oocyte-to-zygote transition. *Genes Dev* 25, 755-766.
- Chen, L., Dumelie, J.G., Li, X., Cheng, M.H., Yang, Z., Laver, J.D., Siddiqui, N.U., Westwood, J.T., Morris, Q., Lipshitz, H.D., *et al.* (2014). Global regulation of mRNA translation and stability in the early *Drosophila* embryo by the Smaug RNA-binding protein. *Genome Biol* 15, R4.
- Chu, T., Henrion, G., Haegeli, V., and Strickland, S. (2001). Cortex, a *Drosophila* gene required to complete oocyte meiosis, is a member of the Cdc20/fizzy protein family. *Genesis* 29, 141-152.
- Courtot, C., Fankhauser, C., Simanis, V., and Lehner, C.F. (1992). The *Drosophila cdc25* homolog twine is required for meiosis. *Development* 116, 405-416.
- Cui, J., Sackton, K.L., Horner, V.L., Kumar, K.E., and Wolfner, M.F. (2008). Wispy, the *Drosophila* homolog of GLD-2, is required during oogenesis and egg activation. *Genetics* 178, 2017-2029.

- Cui, J., Sartain, C.V., Pleiss, J.A., and Wolfner, M.F. (2013). Cytoplasmic polyadenylation is a major mRNA regulator during oogenesis and egg activation in *Drosophila*. *Dev Biol* 383, 121-131.
- 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, 15-21.
- Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R., and Weissman, J.S. (2013). Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *Elife* 2, e01179.
- Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 7, 191.
- Fenger, D.D., Carminati, J.L., Burney-Sigman, D.L., Kashevsky, H., Dines, J.L., Elfring, L.K., and Orr-Weaver, T.L. (2000). PAN GU: a protein kinase that inhibits S phase and promotes mitosis in early *Drosophila* development. *Development* 127, 4763-4774.
- Greenspan, R.J. (1997). Fly pushing : the theory and practice of *Drosophila* genetics (Plainview, N.Y., Cold Spring Harbor Laboratory Press).
- Harrison, M.M., and Eisen, M.B. (2015). Transcriptional Activation of the Zygotic Genome in *Drosophila*. *Curr Top Dev Biol* 113, 85-112.
- Horner, V.L., and Wolfner, M.F. (2008). Transitioning from egg to embryo: triggers and mechanisms of egg activation. *Dev Dyn* 237, 527-544.
- Hughes, S.E., Gilliland, W.D., Cotitta, J.L., Takeo, S., Collins, K.A., and Hawley, R.S. (2009). Heterochromatic threads connect oscillating chromosomes during prometaphase I in *Drosophila* oocytes. *PLoS Genet* 5, e1000348.
- Igreja, C., and Izaurralde, E. (2011). CUP promotes deadenylation and inhibits decapping of mRNA targets. *Genes Dev* 25, 1955-1967.
- Ingolia, N.T., Ghaemmighami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218-223.
- Ivshina, M., Lasko, P., and Richter, J.D. (2014). Cytoplasmic polyadenylation element binding proteins in development, health, and disease. *Annu Rev Cell Dev Biol* 30, 393-415.
- Jacobs, H.W., Knoblich, J.A., and Lehner, C.F. (1998). *Drosophila* Cyclin B3 is required for female fertility and is dispensable for mitosis like Cyclin B. *Genes Dev* 12, 3741-3751.
- Jan, C.H., Friedman, R.C., Ruby, J.G., and Bartel, D.P. (2011). Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature* 469, 97-101.

- Kadyrova, L.Y., Habara, Y., Lee, T.H., and Wharton, R.P. (2007). Translational control of maternal Cyclin B mRNA by Nanos in the *Drosophila* germline. *Development* 134, 1519-1527.
- Kronja, I., Whitfield, Z.J., Yuan, B., Dzek, K., Kirkpatrick, J., Krijgsveld, J., and Orr-Weaver, T.L. (2014a). Quantitative proteomics reveals the dynamics of protein changes during *Drosophila* oocyte maturation and the oocyte-to-embryo transition. *Proc Natl Acad Sci U S A* 111, 16023-16028.
- Kronja, I., Yuan, B., Eichhorn, S.W., Dzek, K., Krijgsveld, J., Bartel, D.P., and Orr-Weaver, T.L. (2014b). Widespread changes in the posttranscriptional landscape at the *Drosophila* oocyte-to-embryo transition. *Cell Rep* 7, 1495-1508.
- Liang, H.L., Nien, C.Y., Liu, H.Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* 456, 400-403.
- Lieberfarb, M.E., Chu, T., Wreden, C., Theurkauf, W., Gergen, J.P., and Strickland, S. (1996). Mutations that perturb poly(A)-dependent maternal mRNA activation block the initiation of development. *Development* 122, 579-588.
- McGrew, L.L., Dworkin-Rastl, E., Dworkin, M.B., and Richter, J.D. (1989). Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev* 3, 803-815.
- Nakamura, A., Sato, K., and Hanyu-Nakamura, K. (2004). *Drosophila* cup is an eIF4E binding protein that associates with Bruno and regulates *oskar* mRNA translation in oogenesis. *Dev Cell* 6, 69-78.
- Nelson, M.R., Leidal, A.M., and Smibert, C.A. (2004). *Drosophila* Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. *Embo J* 23, 150-159.
- Newton, F.G., Harris, R.E., Sutcliffe, C., and Ashe, H.L. (2015). Coordinate post-transcriptional repression of Dpp-dependent transcription factors attenuates signal range during development. *Development* 142, 3362-3373.
- Pesin, J.A., and Orr-Weaver, T.L. (2007). Developmental role and regulation of cortex, a meiosis-specific anaphase-promoting complex/cyclosome activator. *PLoS Genet* 3, e202.
- Pique, M., Lopez, J.M., Foissac, S., Guigo, R., and Mendez, R. (2008). A combinatorial code for CPE-mediated translational control. *Cell* 132, 434-448.
- Qin, X., Ahn, S., Speed, T.P., and Rubin, G.M. (2007). Global analyses of mRNA translational control during early *Drosophila* embryogenesis. *Genome Biol* 8, R63.
- Richter, J.D. (2007). CPEB: a life in translation. *Trends Biochem Sci* 32, 279-285.

- Salles, F.J., Lieberfarb, M.E., Wreden, C., Gergen, J.P., and Strickland, S. (1994). Coordinate initiation of *Drosophila* development by regulated polyadenylation of maternal messenger RNAs. *Science* 266, 1996-1999.
- Semotok, J.L., Cooperstock, R.L., Pinder, B.D., Vari, H.K., Lipshitz, H.D., and Smibert, C.A. (2005). Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early *Drosophila* embryo. *Curr Biol* 15, 284-294.
- Shamanski, F.L., and Orr-Weaver, T.L. (1991). The *Drosophila* plutonium and pan gu genes regulate entry into S phase at fertilization. *Cell* 66, 1289-1300.
- Sheets, M.D., Fox, C.A., Hunt, T., Vande Woude, G., and Wickens, M. (1994). The 3'-untranslated regions of c-mos and cyclin mRNAs stimulate translation by regulating cytoplasmic polyadenylation. *Genes Dev* 8, 926-938.
- Spradling, A.C. (1993). Developmental Genetics of Oogenesis. In *The Development of Drosophila Melanogaster*, M.A. Bate, A., ed. (Woodbury, Cold Spring Harbor Laboratory Press), pp. 1-70.
- 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, 66-71.
- Swan, A., and Schupbach, T. (2007). The Cdc20 (Fzy)/Cdh1-related protein, Cort, cooperates with Fzy in cyclin destruction and anaphase progression in meiosis I and II in *Drosophila*. *Development* 134, 891-899.
- Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. *Dev Cell* 12, 143-155.
- Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development* 136, 3033-3042.
- Temme, C., Simonelig, M., and Wahle, E. (2014). Deadenylation of mRNA by the CCR4-NOT complex in *Drosophila*: molecular and developmental aspects. *Front Genet* 5, 143.
- Temme, C., Zhang, L., Kremmer, E., Ihling, C., Chartier, A., Sinz, A., Simonelig, M., and Wahle, E. (2010). Subunits of the *Drosophila* CCR4-NOT complex and their roles in mRNA deadenylation. *RNA* 16, 1356-1370.
- Vardy, L., and Orr-Weaver, T.L. (2007a). The *Drosophila* PNG kinase complex regulates the translation of *cyclin B*. *Dev Cell* 12, 157-166.
- Vardy, L., and Orr-Weaver, T.L. (2007b). Regulating translation of maternal messages: multiple repression mechanisms. *Trends Cell Biol* 17, 547-554.

Von Stetina, J.R., and Orr-Weaver, T.L. (2011). Developmental control of oocyte maturation and egg activation in metazoan models. *Cold Spring Harb Perspect Biol* 3, a005553.

Von Stetina, J.R., Tranguch, S., Dey, S.K., Lee, L.A., Cha, B., and Drummond-Barbosa, D. (2008). alpha-Endosulfine is a conserved protein required for oocyte meiotic maturation in *Drosophila*. *Development* 135, 3697-3706.

Weill, L., Belloc, E., Bava, F.A., and Mendez, R. (2012). Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat Struct Mol Biol* 19, 577-585.

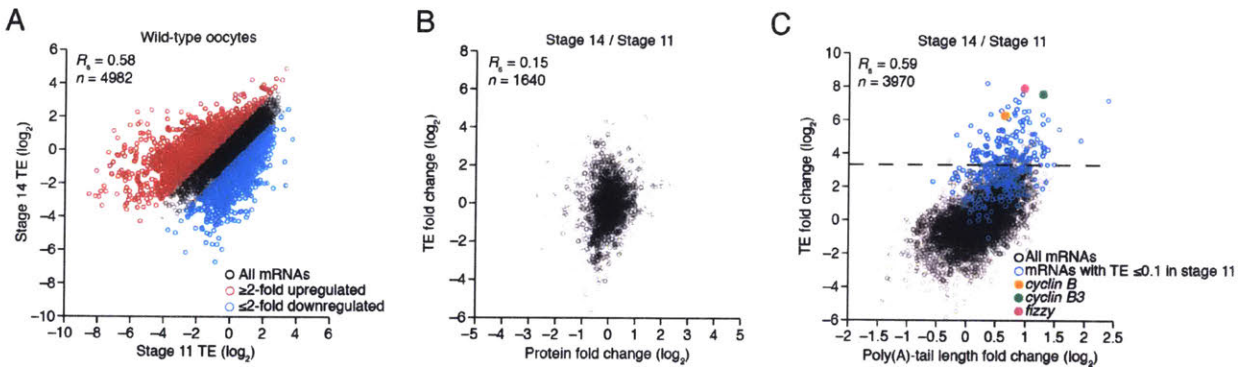


Figure 2—figure supplement 1. Widespread translational regulation during oocyte maturation.

(A) Relationship between TEs in stage 11 and stage 14 oocytes. TE values (log₂) were median centered (median values in stage 11 and stage 14 oocytes, -0.5153 and -0.1318 , respectively). Results are plotted for all mRNAs that had ≥ 10.0 RPM in the RNA-seq data of both samples and ≥ 10.0 RPM in the ribosome profiling data of at least one of the two samples and >0.0 RPM in the other. The median-centered TE fold-change values (log₂) of stage 14 relative to stage 11 (median, 0.0147) were used to identify mRNAs that were ≥ 2 -fold up- or downregulated in stage 14 oocytes, which are highlighted in red and blue, respectively; otherwise as in Figure 1B. (B) Relationship between the change in protein level (Kronja et al., 2014a) and change in TE observed after oocyte maturation. TE and protein fold-change (log₂) values were median centered (median values of 0.1236 and 0.0474 , respectively). Results are plotted for all mRNAs that had ≥ 10.0 RPM in the RNA-seq data, ≥ 10.0 RPM in the ribosome profiling of one of the two samples and >0.0 RPM in the other, and proteins with ≥ 2 unique peptides in each of three independent mass spectrometry experiments. (C) Relationship between tail-length and TE changes observed after oocyte maturation, redrawn from Figure 4A, but highlighting points for mRNAs with a TE ≤ 0.1 in stage 11 oocytes in blue (Supplementary File 3) and those for *cyclin B*, *cyclin B3*, and *fizzy* in yellow, green, and purple, respectively. The dashed line indicates a TE increase of 10 fold.

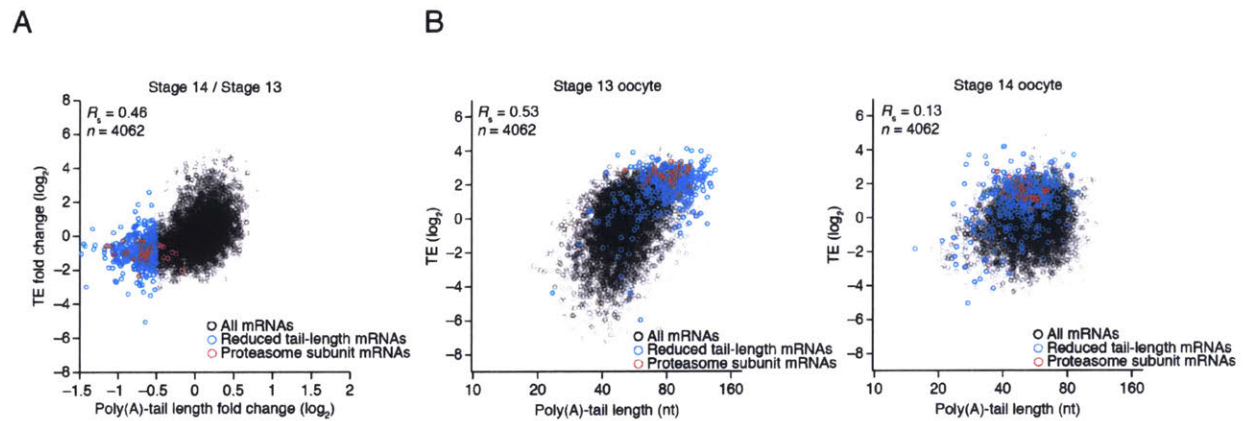


Figure 2—figure supplement 2. Uncoupled behavior observed for a subset of mRNAs between the last stages of oocyte maturation.

(A) The relationship between the tail-length and TE changes observed between stage 14 oocytes relative to stage 13 oocytes, redrawn from Figure 2, highlighting in blue mRNAs that had tail-length decreases of at least 29% (\log_2 fold change of -0.5), and in red mRNAs encoding proteasome subunits (Supplementary File 3). (B) The relationship between tail length and TE observed in stage 13 and 14 oocytes (*left* and *right*, respectively) analyzed as in Figure 2 but plotting results for only the mRNAs examined in panel A and highlighting the same sets of mRNAs as in panel A. TE values (\log_2) were median centered (median values for stage 13 and 14 oocytes, -1.2409 and -0.0779 , respectively).

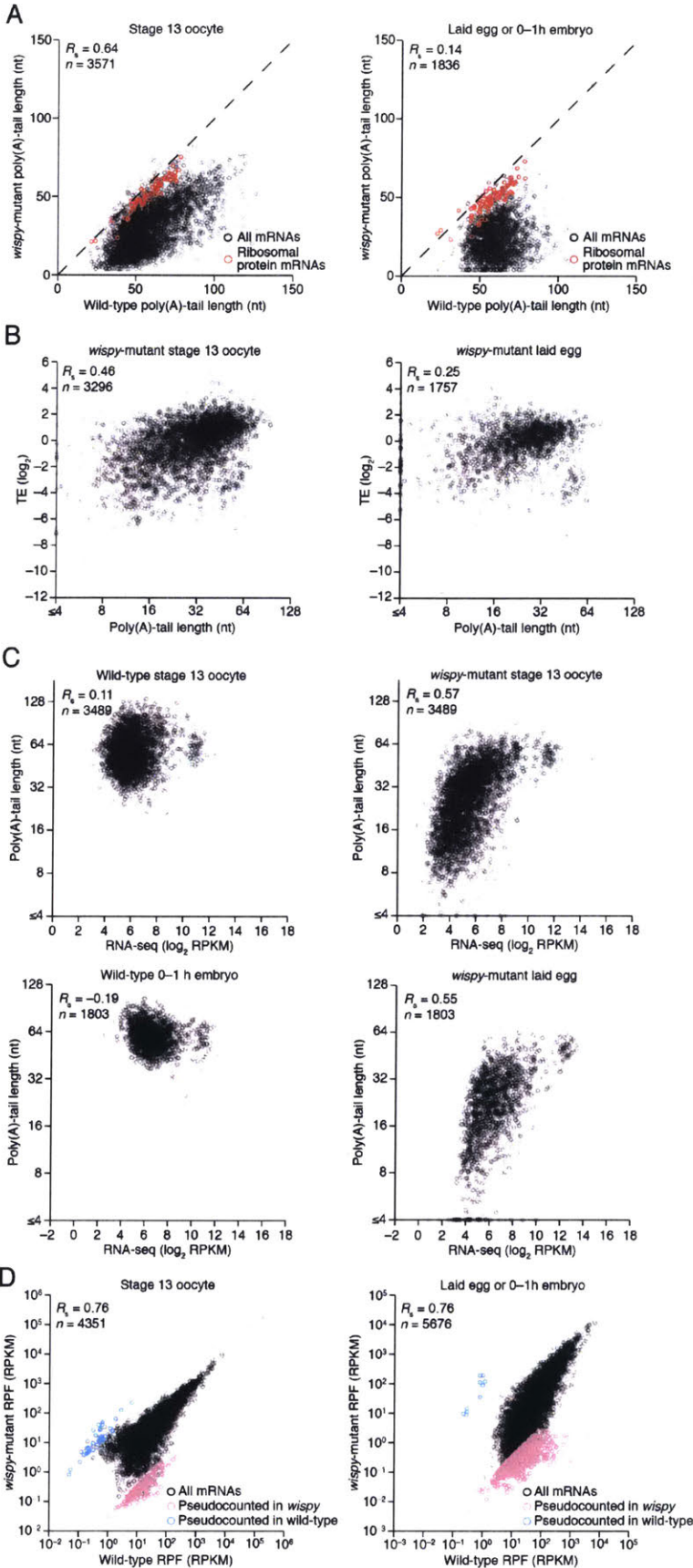


Figure 5—figure supplement 1. Impact of Wispy on the poly(A)-tail length of most mRNAs, which biased mRNA recovery during poly(A) selection.

(A) The plots from Figure 5A, highlighting mRNAs encoding ribosomal proteins in red (Supplementary File 3). (B) Relationship between mean poly(A)-tail length and TE at the indicated developmental stage in *wispy*-mutant samples. The TEs were calculated using *wispy*-mutant RNA-seq and ribosome profiling data, and are median centered (median values in stage 13 oocytes and laid eggs, -0.5729 and 0.5761 , respectively); otherwise, as in Figure 5B. (C) Relationship between measured RNA abundance and poly(A)-tail length in wild-type stage 13 oocytes and 0–1 h embryos, and in *wispy*-mutant stage 13 oocytes and laid eggs. Results are plotted for mRNAs that had ≥ 100 poly(A) tags in both the wild-type and *wispy*-mutant samples for the corresponding stage and ≥ 10.0 RPM in the RNA-seq data in the wild-type sample. The same mRNAs are plotted in corresponding wild-type and *wispy*-mutant samples. mRNAs that had mean tail-length values ≤ 4 nt are reported as ≤ 4 nt. (D) Comparison of RPF measurements for wild-type and *wispy*-mutant stage 13 oocytes (*left*) and cleavage-stage embryos (*right*). Results are plotted for all mRNAs with ≥ 10.0 RPM in the ribosome-profiling data of either the wild-type or *wispy*-mutant sample, and any mRNA with 0 reads in either sample was given a pseudocount of 1 read and highlighted in blue or purple (Supplementary File 3).

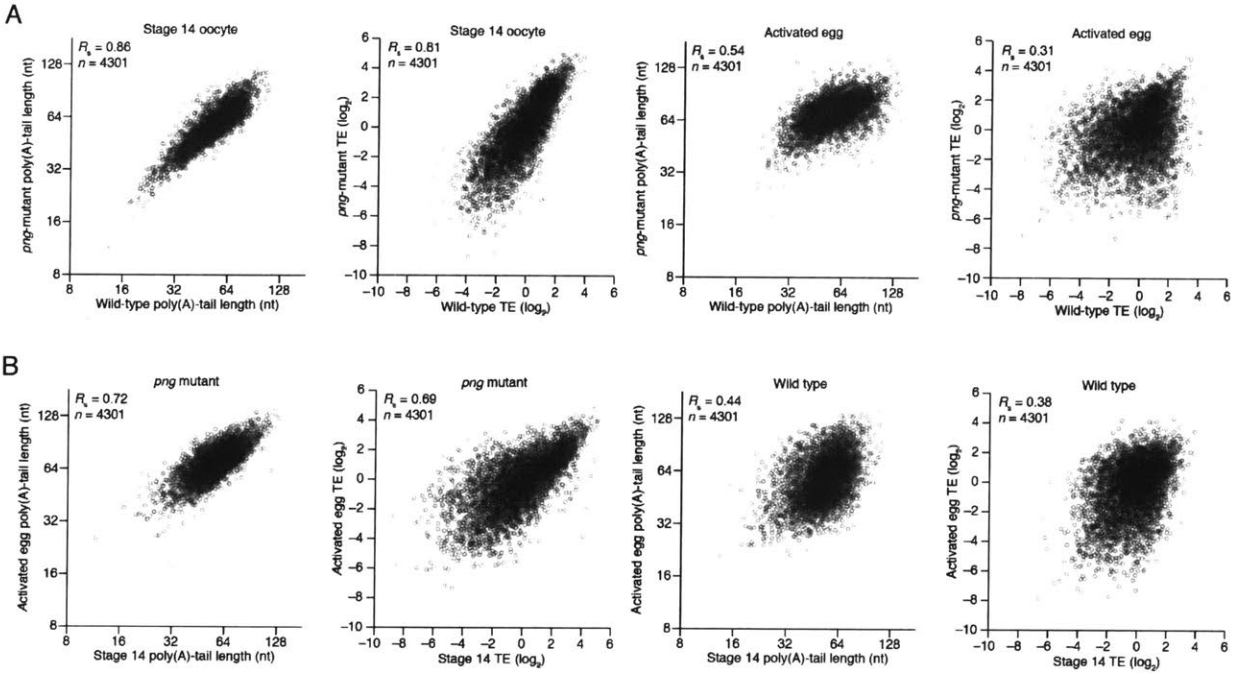


Figure 6—figure supplement 1. Perturbation of tail-length and TE during the oocyte-to-embryo transition in *png*-mutant samples.

(A) Comparison of mean poly(A)-tail lengths and TEs between wild-type and *png*-mutant stage 14 oocytes, and between wild-type and *png*-mutant activated eggs. TE values (\log_2) were median centered (median values in wild-type and *png*-mutant stage 14 oocytes, -0.2182 and -1.182 , respectively, and wild-type and *png*-mutant activated eggs 0.0724 and -0.2394 , respectively). Results are plotted for mRNAs that had ≥ 100 poly(A) tags, ≥ 10.0 RPM in the RNA-seq data, and >0 RPM in the ribosome profiling data for each sample. (B) Comparison of mean poly(A)-tail lengths and TEs between *png*-mutant stage 14 oocytes and activated eggs, and between wild-type stage 14 oocytes and activated eggs; otherwise, as in panel A.

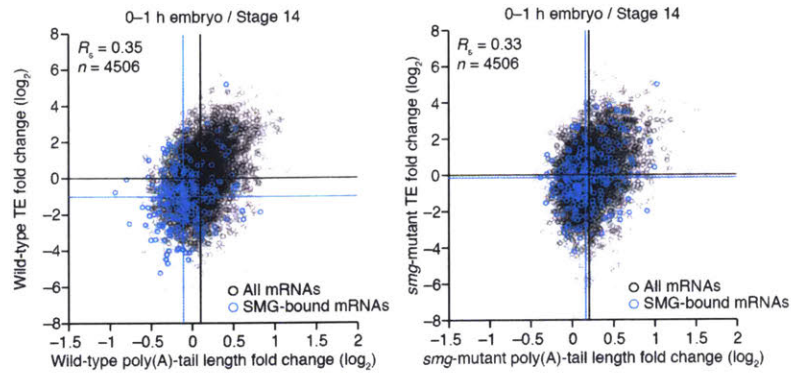


Figure 7—figure supplement 1. Smaug-dependent translational repression of Smaug binding targets, mediated primarily by changes in poly(A)-tail length.

Shown is the relationship between mean tail-length changes and TE changes during the OET for wild-type and *smg*-mutant 0–1 h embryos (*left* and *right*, respectively), highlighting mRNAs previously reported to be bound by SMG in the early embryo (Chen et al., 2014); otherwise as in Figure 6B.

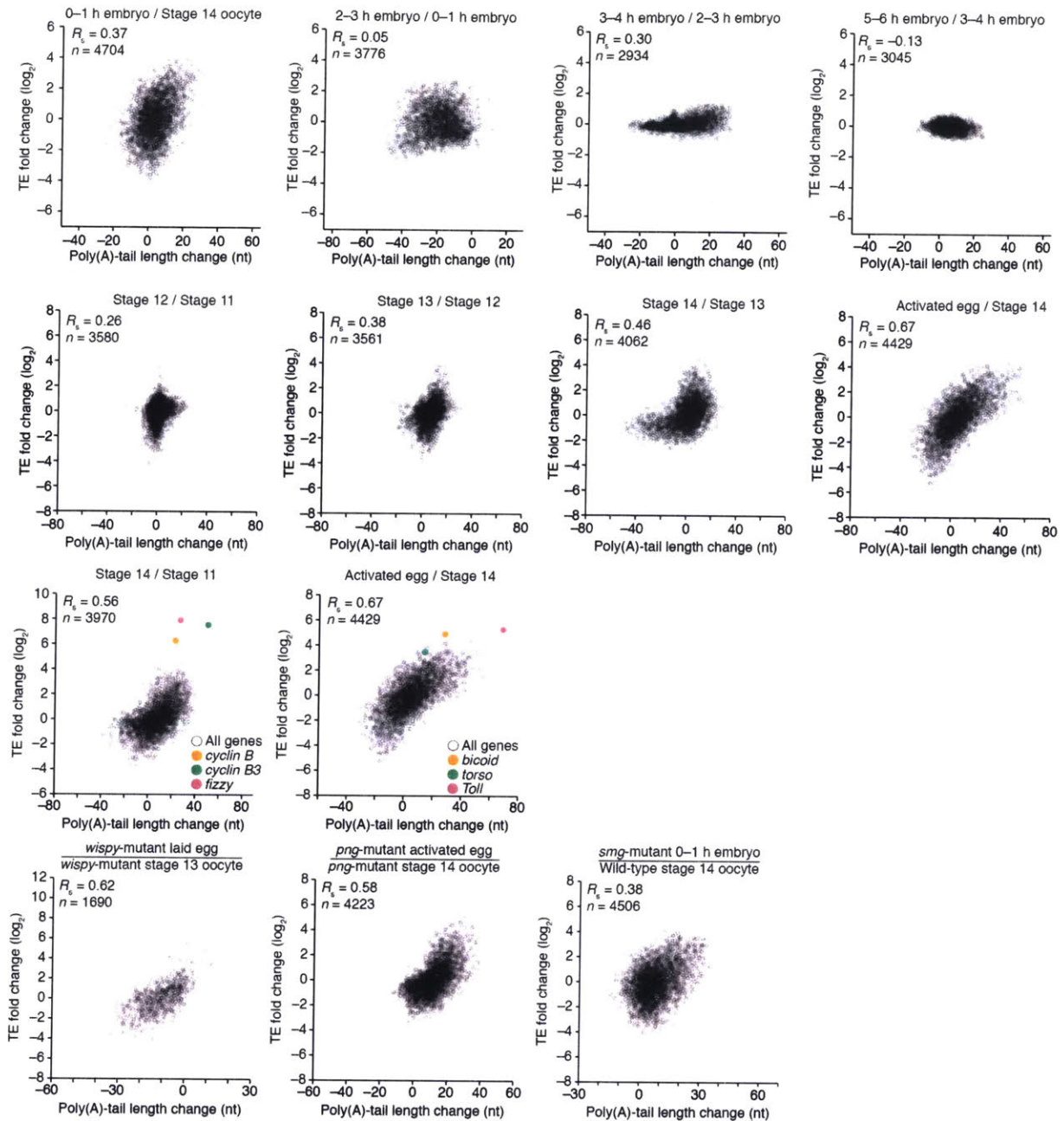


Figure 7—figure supplement 2. Relationship between poly(A)-tail length changes and TE changes throughout development, plotting absolute rather than relative tail-length changes. Shown are the relationships between the absolute differences in mean tail-length and TE changes observed between stages; otherwise, as in the corresponding panels from Figures 1, 2, 4, 5, 6, and 7.

Chapter 5

Future Directions

Translational regulation

Our studies have contributed to the fundamental understanding of how poly(A)-tail length influences translation in oocyte, embryos, and post-embryonic contexts (Subtelny et al., 2014; Eichhorn et al., 2016), as well as how mammalian microRNAs regulate their endogenous mRNA targets (Eichhorn et al., 2014). By globally studying post-transcriptional and translational regulation, we were able to learn both general regulatory principles, as well as specific instances of unusual regulation, and most importantly distinguish between the two. This encapsulates how next generation sequencing techniques have transformed the study of transcriptional, post-transcriptional, and translational regulation.

For over 40 years, it has been clear that many cellular mRNAs differ in how efficiently they are translated (Lodish, 1976). The application of next generation sequencing to measurements of translation expanded our understanding of translational efficiencies from a few well-studied mRNAs to all cellular mRNAs. In particular, ribosome profiling (Ingolia et al., 2009) has seen wide adoption and continues to advance our understanding of translation and translational regulation. Our ribosome profiling measurements in *Saccharomyces cerevisiae* indicate that the translational efficiencies of the mRNAs for nearly all expressed genes span an approximately 10-fold range (Subtelny et al., 2014; Weinberg et al., 2016). We measured a similarly narrow range in *Schizosaccharomyces pombe* and several mammalian cell types (Subtelny et al., 2014). However, this modest range of translational efficiencies should not dampen excitement about the importance of translational regulation in these cellular contexts—

many RNAi and CRISPRi experiments reduce the level of a protein by less than 10-fold, but nonetheless observe biologically relevant effects. Identifying and understanding the translational regulation of endogenous mRNAs is now within reach, and over time will provide a far more quantitative understanding of how sequence features and RNA-binding proteins determine the translational efficiency of an mRNA. Although the vast majority of genes span a narrow range of translational efficiencies, there are some poorly translated genes that are clear outliers. These include both well-understood and previously unknown cases of robust translational repression, and it will be interesting to understand how and why these mRNAs are regulated so strongly at the level of translation.

In many cell types and cellular states gene regulation is unusually reliant on translational regulation. This includes systems where transcription is absent (or physically distant), such as in oocytes, embryos, enucleate blood cells, and neuronal synapses, as well as systems with limited capacity for specific transcriptional regulation, such as in trypanosomes. Exploring these systems offers the ability to gain insight into translational regulation that is performing critical biological functions. Initial high-throughput studies in some of the aforementioned contexts have characterized an unusual mechanism of translational control (Bazzini et al., 2012; Subtelny et al., 2014) and made observations relevant to human disease (Jensen et al., 2014; Mills et al., 2016). In other contexts, one or more aspects of translation is dysregulated. This includes certain cancers (Ruggero, 2013), viral infections (Bushell and Sarnow, 2002), and stress conditions (Hinnebusch, 2011), along with diseases arising from mutation or deletion of certain ribosomal protein genes (Narla and Ebert, 2010). Characterizing how the changes in the translation apparatus in these contexts affect cellular mRNAs may reveal basic principles of translation in addition to insights specific to the disease or stress state.

Poly(A)-tail length and translational efficiency

Our research extended prior single-gene studies by demonstrating that there is a global relationship between poly(A)-tail length and translational efficiency in oocytes and early embryos. Furthermore, we discovered that this coupling between poly(A)-tail length and translational efficiency is lost around the time of gastrulation in *Drosophila*, *Xenopus*, and zebrafish embryos, and is absent in all post-embryonic contexts we examined. How do oocytes and early embryos measure poly(A)-tail lengths and relay that information to the translation machinery, and why is this mechanism lost as development proceeds? This is perhaps the most important remaining question from our work, and there are other interesting questions about when coupling is first established and whether it is present in mammals and other divergent animal species.

An obvious candidate for mechanistically coupling poly(A)-tail length and translational efficiency is PABP. It directly binds the poly(A) tail and is known to stimulate translation, making it an ideal molecule to regulate translation based on tail length. One simple model is that PABP is initially at a limiting concentration during early development and increases in abundance over time. In early development, longer poly(A)-tails would be more likely to have at least one PABP bound, conferring a translational advantage relative to short-tailed mRNAs. In later development, there would be sufficient PABP to bind all poly(A) tails and so longer poly(A)-tails would no longer confer a translational advantage.

Although appealing, prior studies cast doubt on the model of limiting PABP abundance. First, in sea urchin eggs, PABP levels are in 50-fold excess relative to poly(A)-tail binding sites (Drawbridge et al., 1990). We have not demonstrated coupling in sea urchin eggs, but I

anticipate that it would be present based on developmental similarity to systems we have studied. Second, although the predominant poly(A)-binding protein in post-embryonic contexts, PABPC1, is expressed at low levels in *Xenopus* oocytes and embryos (Stambuk and Moon, 1992), at least one other poly(A)-binding protein (known as ePABP) is also expressed in these contexts (Voeltz et al., 2001; Cosson et al., 2002). ePABP binds poly(A) RNA (Cosson et al., 2002), can interact with eIF4G (Cosson et al., 2002), and stimulates translation when tethered to a reporter RNA (Wilkie et al., 2005). Based on these properties, ePABP and PABPC1 seem to be very similar, and quantitative measurements of ePABP are necessary to determine if overall poly(A)-binding protein levels are at a limiting concentration in early *Xenopus* embryos. The limiting poly(A)-binding protein model could be directly tested by perturbing the levels of ePABP or PABPC1 during early development and measuring the relationship between tail length and translational efficiency.

The limiting PABP model as described above focuses on absolute protein levels, but it is also possible that PABP is regulated at a functional instead of absolute level through post-translational modifications. Several sites on PABPC1 can be phosphorylated, and altering the phosphorylation state of PABPC1 impacts its ability to bind RNA and translation initiation factors *in vitro* (Le et al., 2000). Thus, the excess abundance of PABP in sea urchin eggs (Drawbridge et al., 1990) might not actually reflect an excess in poly(A)-binding capacity if the amount of functional PABP is substantially smaller than the total amount of PABP. This possibility could be investigated by characterizing the modification state of PABP throughout early development.

It is formally possible that a unique property of ePABP is responsible for coupling tail length and translational efficiency, and this could be tested by replacing ePABP expression with

PABPC1 expression in early development. A mass-spectrometry approach to identify PABP-interacting proteins or proteins that bind long poly(A)-tails during early development may be useful in generating additional hypotheses about the proteins that are involved in measuring poly(A) tails.

Another hypothesis is that there is no mechanism to read out tail lengths in early development, but instead the unusual mRNA metabolism of this context is responsible for coupling tail length and translational efficiency. During early development deadenylation does not lead to decapping, such that mRNAs are able to have very short poly(A) tails at these times. As development proceeds through the MZT, short-tailed mRNAs are destabilized and transcription begins, resulting in an overall lengthening of poly(A) tails and a narrower distribution of tail lengths (Tadros and Lipshitz, 2009; Subtelny et al., 2014). Having determined the mean tail length for thousands of genes in early- and late-stage embryos, we computationally tested the hypothesis that a shift in the tail-length distribution caused a loss of coupling. In this analysis we sampled genes from early-stage embryos to match the distribution of mean tail-lengths in late-stage embryos. The correlation between tail length and translation efficiency in this matched dataset was still strong in the early-stage embryos (Subtelny et al., 2014), demonstrating that the change in the tail-length distribution does not account for the loss of coupling. It remains possible that a more subtle aspect of tail length that is not reflected in the mean for a given gene, such as the proportion of tails that are very short, is important for coupling tail length and translational efficiency.

Cytoplasmic polyadenylation, another part of the unusual mRNA metabolism of early development, is not required for coupling tail length and translational efficiency. In *Drosophila*, the cytoplasmic poly(A)-polymerase Wispy is active in early development, but the correlation

between tail length and translational efficiency was roughly the same in wild-type and Wispy-mutant oocytes and embryos (Eichhorn et al., 2016). Thus, the unusual mRNA metabolism of early development expands the potential to regulate tail length, and there is not yet evidence that it is important for mechanistically coupling tail length and translational efficiency. In all likelihood there will not be an individual protein or property of early developmental systems that is sufficient to couple tail length and translational efficiency, and characterization of the presence or absence of coupling in other species may enable hypotheses about mechanism to be generated through comparative analysis.

Characterizing tail length and translational efficiency in wild-type and mutant contexts led to several observations that merit further consideration. Our initial measurements of Toll mRNA seemed to reflect the classic paradigm of cytoplasmic polyadenylation regulating translation: Toll went from having a mean poly(A)-tail length of 29 nts to 98 nts during egg activation, and its translation was 50-fold upregulated. However, in the wispy-mutant samples the poly(A) tail of Toll mRNA remained short after egg activation, but its translation was upregulated to a similar degree as in the wild-type samples (Eichhorn et al., 2016). Without absolute measurements of translational efficiency it is not possible to understand how the translational upregulation observed in these two conditions compare to one another, but at a minimum it reflects a robust mechanism of tail-length independent translational regulation. Furthermore, in coupled systems there is a large range of translational efficiencies associated with poly(A) tails of a single length (Subtelny et al., 2014; Eichhorn et al., 2016). Tail-length independent regulatory mechanisms seem to be a large component of regulating gene expression during development, and understanding these mechanisms will likely be pertinent to gene regulation in both developmental and non-developmental contexts.

Smaug and PAN GU were initially thought to regulate mRNA targets through both tail-length dependent and independent mechanisms (Nelson et al., 2004; Semotok et al., 2005; Tadros et al., 2007; Vardy and Orr-Weaver, 2007). Our data indicate that they act primarily through tail-length dependent mechanisms, and that there is not a subset of their targets that is only regulated by a tail-length independent mechanism (Eichhorn et al., 2016). It would be valuable to use global analyses to characterize the mechanisms of regulation used by the many other RBPs that regulate translation during development. As is clear from our analyses, the broader picture offered by global analyses will refine the ideas that have come from studies of a handful of mRNA targets and ectopic substrates.

MicroRNA mechanism

Dissecting the mechanistic relationship between miRNA-mediated translational repression and mRNA degradation has been extremely challenging because a single protein complex mediates both modes of repression, and every protein in this complex is involved in both modes of repression (Jonas and Izaurralde, 2015). As a first step, we and others have observed that miRNAs cause a decrease in translation prior to a decrease in mRNA abundance. This was seen for reporter RNAs (Bethune et al., 2012; Djuranovic et al., 2012), as well as endogenous mRNAs (Bazzini et al., 2012; Eichhorn et al., 2014). Now that the order of events is known, the field can focus on dissecting the mechanistic relationship underlying these kinetic differences.

Three mechanistic relationships between miRNA-mediated translational repression and mRNA degradation could explain the observation that decreases in translation precede decreases in mRNA abundance. One possible relationship is that the changes induced by the mRNA

degradation pathway cause translational repression. Deadenylated and decapped mRNAs, as well as capped mRNAs that are lacking eIF4E, are intermediates that are necessarily produced during mRNA degradation and are poorly translated owing to their lack of eIF4E. If these intermediates persist for a meaningful amount of time prior to complete degradation, then a decrease in translation would be observed prior to a decrease in mRNA abundance. A better understanding of the kinetics of decapping after eIF4E removal, and XRN1 recruitment after decapping, would help clarify the plausibility of this model. It would also be informative to measure the timing of decapping relative to translational repression during miRNA-mediated repression. Another possible relationship is that miRNA-mediated translational repression causes mRNA destabilization. In this case, either decreases in the number of translating ribosomes or changes in the bound initiation factors cause reduced mRNA stability. Although there is a long precedent for changes in translation initiation affecting mRNA stability in yeast (Schwartz and Parker, 1999; Sheth and Parker, 2003; Collier and Parker, 2005; Teixeira et al., 2005), far less is known about such a connection in systems relevant to miRNA-mediated repression. The final possibility is that translational repression has no causal relationship with mRNA degradation, and simply occurs more quickly. As further experiments build support for one of these relationships, it will be interesting to consider the implications for other regulators of mRNA stability or translation.

Single-molecule sequencing approaches

Next generation sequencing has dramatically advanced the study of endogenous RNAs and their regulation. In spite of this, there are many applications and analyses that are limited by the current methods. For example, it is particularly challenging to characterize differential regulation of transcript isoforms, and in most analyses transcript isoforms are compressed into a

single “representative” isoform for a given gene despite the diversity that is known to exist. Another challenging application is determining the distribution of RNA modifications on mRNAs. Current approaches primarily yield population averages for all the mRNAs corresponding to a given gene, and offer limited ability to directly determine if distant modification sites are modified on the same mRNA molecule or different mRNA molecules. New technologies for high-throughput, single-molecule sequencing theoretically offer the ability to sequence long mRNA molecules in their entirety, and in a high-throughput manner. Such information will be valuable for evaluating the influence of transcript isoforms, or even single-nucleotide polymorphisms, on all manner of post-transcriptional regulation and will provide a more detailed view of many areas of RNA biology.

Conclusion

Nearly 50 years ago, a stretch of adenosine residues was found within eukaryotic mRNAs, and research over the subsequent decades demonstrated that this sequence increases the translation and stability of mRNAs in the cytoplasm. The ubiquitous presence of a poly(A) tail on mRNAs belies the fact that poly(A) tails are not all the same length, yet the consequence of these length differences had been unclear because the relationship between poly(A)-tail length and translation had been scarcely examined. Our research demonstrated that there is a global relationship between tail length and translational efficiency in oocytes and embryos, and that miRNAs and RNA-binding proteins exploit this coupled state by regulating translation through tail-length changes. As was true for our research, considering all endogenous genes often reveals principles of regulation that were missed or incompletely understood from earlier single-gene approaches. The application of global approaches to studying post-transcriptional and

translational regulation will continue to reveal previously underappreciated mechanisms of control and the breadth of their action.

References

- Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* 336, 233-237.
- Bethune, J., Artus-Revel, C.G., and Filipowicz, W. (2012). Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* 13, 716-723.
- Bushell, M., and Sarnow, P. (2002). Hijacking the translation apparatus by RNA viruses. *J Cell Biol* 158, 395-399.
- Coller, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. *Cell* 122, 875-886.
- Cosson, B., Couturier, A., Le Guellec, R., Moreau, J., Chabelskaya, S., Zhouravleva, G., and Philippe, M. (2002). Characterization of the poly(A) binding proteins expressed during oogenesis and early development of *Xenopus laevis*. *Biol Cell* 94, 217-231.
- Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* 336, 237-240.
- Drawbridge, J., Grainger, J.L., and Winkler, M.M. (1990). Identification and characterization of the poly(A)-binding proteins from the sea urchin: a quantitative analysis. *Mol Cell Biol* 10, 3994-4006.
- Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S.H., Ghoshal, K., Villen, J., and Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell* 56, 104-115.
- Eichhorn, S.W., Subtelny, A.O., Kronja, I., Kwasnieski, J.C., Orr-Weaver, T.L., and Bartel, D.P. (2016). mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila* oocytes and early embryos. *Elife* 5.
- Gruber, A.R., Martin, G., Muller, P., Schmidt, A., Gruber, A.J., Gumienny, R., Mittal, N., Jayachandran, R., Pieters, J., Keller, W., *et al.* (2014). Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat Commun* 5, 5465.
- Hinnebusch, A.G. (2011). Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiol Mol Biol Rev* 75, 434-467, first page of table of contents.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218-223.
- Jensen, B.C., Ramasamy, G., Vasconcelos, E.J., Ingolia, N.T., Myler, P.J., and Parsons, M. (2014). Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*. *BMC Genomics* 15, 911.

- Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* 16, 421-433.
- Le, H., Browning, K.S., Gallie, D.R. (2000). The phosphorylation state of poly(A)-binding protein specifies its binding to poly(A) RNA and its interaction with eukaryotic initiation factors (eIF) 4F, eIFiso4F, and eIF4B. *J Biol Chem* 275, 17452-17462.
- Lodish, H.F. (1976). Translational control of protein synthesis. *Annu Rev Biochem* 45, 39-72.
- Mayr, C. (2016). Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol* 26, 227-237.
- Mills, E.W., Wangen, J., Green, R., and Ingolia, N.T. (2016). Dynamic Regulation of a Ribosome Rescue Pathway in Erythroid Cells and Platelets. *Cell Rep* 17, 1-10.
- Narla, A., and Ebert, B.L. (2010). Ribosomopathies: human disorders of ribosome dysfunction. *Blood* 115, 3196-3205.
- Nelson, M.R., Leidal, A.M., and Smibert, C.A. (2004). Drosophila Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. *Embo J* 23, 150-159.
- Ruggero, D. (2013). Translational control in cancer etiology. *Cold Spring Harb Perspect Biol* 5.
- Schwartz, D.C., and Parker, R. (1999). Mutations in translation initiation factors lead to increased rates of deadenylation and decapping of mRNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol* 19, 5247-5256.
- Semotok, J.L., Cooperstock, R.L., Pinder, B.D., Vari, H.K., Lipshitz, H.D., and Smibert, C.A. (2005). Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early *Drosophila* embryo. *Curr Biol* 15, 284-294.
- Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* 300, 805-808.
- Spies, N., Burge, C.B., and Bartel, D.P. (2013). 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res* 23, 2078-2090.
- Stambuk, R.A., and Moon, R.T. (1992). Purification and characterization of recombinant *Xenopus* poly(A)(+)-binding protein expressed in a baculovirus system. *Biochem J* 287 (Pt 3), 761-766.
- 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, 66-71.
- Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG is a major regulator of

maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. *Dev Cell* *12*, 143-155.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development* *136*, 3033-3042.

Teixeira, D., Sheth, U., Valencia-Sanchez, M.A., Brengues, M., and Parker, R. (2005). Processing bodies require RNA for assembly and contain nontranslating mRNAs. *RNA* *11*, 371-382.

Vardy, L., and Orr-Weaver, T.L. (2007). The *Drosophila* PNG kinase complex regulates the translation of cyclin B. *Dev Cell* *12*, 157-166.

Voeltz, G.K., Ongkasuwan, J., Standart, N., and Steitz, J.A. (2001). A novel embryonic poly(A) binding protein, ePAB, regulates mRNA deadenylation in *Xenopus* egg extracts. *Genes Dev* *15*, 774-788.

Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J.A., Plotkin, J.B., and Bartel, D.P. (2016). Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell Rep* *14*, 1787-1799.

Wilkie, G.S., Gautier, P., Lawson, D., and Gray, N.K. (2005). Embryonic poly(A)-binding protein stimulates translation in germ cells. *Mol Cell Biol* *25*, 2060-2071.

Appendix A

Widespread changes in the posttranscriptional landscape at the *Drosophila* oocyte-to-embryo transition

Iva Kronja¹, Bingbing Yuan¹, Stephen W. Eichhorn^{1,2,3}, Kristina Dzek⁴, Jeroen Krijgsveld⁴, David P. Bartel^{1,2,3}, Terry L. Orr-Weaver^{1,2}

¹Whitehead Institute, Cambridge, MA 02142, USA

²Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

³Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

⁴European Molecular Biology Laboratory, 69117 Heidelberg, Germany

Published as:

Kronja, I., Yuan, B., Eichhorn, S.W., Dzek, K., Krijgsveld, J., Bartel, D.P., and Orr-Weaver, T.L. (2014). Widespread changes in the posttranscriptional landscape at the *Drosophila* oocyte-to-embryo transition. *Cell Rep* 7, 1495-1508.

Widespread Changes in the Posttranscriptional Landscape at the *Drosophila* Oocyte-to-Embryo Transition

Iva Kronja,¹ Bingbing Yuan,¹ Stephen W. Eichhorn,^{1,2,3} Kristina Dzeky,⁴ Jeroen Krijgsveld,⁴ David P. Bartel,^{1,2,3} and Terry L. Orr-Weaver^{1,2,*}

¹Whitehead Institute, Cambridge, MA 02142, USA

²Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

³Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

⁴European Molecular Biology Laboratory, 69117 Heidelberg, Germany

*Correspondence: weaver@wi.mit.edu

<http://dx.doi.org/10.1016/j.celrep.2014.05.002>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

SUMMARY

The oocyte-to-embryo transition marks the onset of development. The initial phase of this profound change from the differentiated oocyte to the totipotent embryo occurs in the absence of both transcription and mRNA degradation. Here we combine global polysome profiling, ribosome-footprint profiling, and quantitative mass spectrometry in a comprehensive approach to delineate the translational and proteomic changes that occur during this important transition in *Drosophila*. Our results show that PNG kinase is a critical regulator of the extensive changes in the translome, acting uniquely at this developmental window. Analysis of the proteome in *png* mutants provided insights into the contributions of translation to changes in protein levels, revealing a compensatory dynamic between translation and protein turnover during proteome remodeling at the return to totipotency. The proteome changes additionally suggested regulators of meiosis and early embryogenesis, including the conserved H3K4 demethylase LID, which we demonstrated is required during this period despite transcriptional inactivity.

INTRODUCTION

The oocyte-to-embryo transition marks the onset of development for any multicellular organism. This dramatic change involves the completion of meiosis in the oocyte, sperm entry, fusion of the male and female pronucleus, and the start of mitotic divisions (Horner and Wolfner, 2008b). These events accompany the profound developmental change from a differentiated oocyte into a totipotent embryo.

Studies on the restoration of cell potency have focused on the regulation of transcription (Young, 2011), but the oocyte-to-embryo transition necessitates a fundamentally different control mechanism. Following the primary arrest in prophase I, oocytes

are transcriptionally silent, and in all animals at least the first embryonic division occurs prior to the initiation of zygotic transcription (Tadros and Lipshitz, 2009). Organisms such as insects, fish, and amphibians rely on stockpiled maternal mRNAs. These organisms proceed through several hours of embryonic development (i.e., 12–13 division cycles in *Xenopus* and *Drosophila*) prior to the onset of zygotic transcription, which also triggers turnover of maternal mRNAs (Anderson and Lengyel, 1979; Zalokar, 1976). In *Drosophila*, a pathway to degrade maternal mRNAs is not active until 2 hr after egg laying (Tadros et al., 2003, 2007). Thus, the oocyte-to-embryo transition and early embryogenesis occur with constant mRNA levels (Tadros and Lipshitz, 2009).

Although many aspects of translational regulation in the oocyte-to-embryo transition remain to be elucidated, the effect of translation on meiotic progression has been extensively analyzed in two other developmental contexts: (1) in yeast meiosis, which proceeds without developmentally programmed arrests and in the presence of transcriptional control, extensive translational regulation nevertheless occurs (Brar et al., 2012; Carlile and Amon, 2008; Chu and Herskowitz, 1998), and (2) in metazoans, at maturation the oocyte exits the primary arrest in prophase I and progresses into meiotic divisions. Experiments in amphibians, mice, and marine invertebrates demonstrated a role for cytoplasmic polyadenylation in activating translation at oocyte maturation and showed that precisely timed translation of several mRNAs is required for progression through the meiotic divisions (Charlesworth et al., 2013; Chen et al., 2011; Gebauer et al., 1994; Tay et al., 2000).

Here, we define the regulatory steps of gene expression at the oocyte-to-embryo transition of *Drosophila*. In *Drosophila*, as in most animals, the mature oocyte is arrested at a secondary point in meiosis (Sagata, 1996). In insects, this arrest point is the metaphase of meiosis I. The trigger for the oocyte-to-embryo transition is egg activation. Egg activation and exit from meiosis in *Drosophila* take place as the oocyte passes into the uterus, regardless of whether it is fertilized. Instead of sperm entry, mechanical pressure as well as osmotic and Ca²⁺ changes are thought to initiate egg activation in *Drosophila* (Horner and Wolfner, 2008a).

Egg activation is presumably accompanied by changes in translation, as poly(A) tails on mRNAs are lengthened and proteins involved in developmental patterning and cell-cycle control are synthesized (Horner and Wolfner, 2008b). In addition, proteins are actively subjected to degradation during egg activation. Release of the metaphase I arrest and completion of meiosis requires the Anaphase Promoting Complex/Cyclosome (APC/C) to target Cyclin B for degradation (Pesin and Orr-Weaver, 2007; Swan and Schüpbach, 2007). We recently showed that a meiosis-specific form of the APC/C also contributes to the change from meiosis in the oocyte to mitosis in the embryo by mediating degradation of the meiotic protein Matrimony (Mtrm) (Whitfield et al., 2013).

The *Drosophila* PNG kinase is required for the onset of mitotic divisions in the embryo (Fenger et al., 2000; Shamanski and Orr-Weaver, 1991). This Ser/Thr kinase is a complex of a catalytic subunit, encoded by the *png* gene, and two activating subunits, the proteins GNU and PLU (Lee et al., 2003). The PNG kinase complex is present and acts solely at the oocyte-to-embryo transition. By promoting the translation of *cyclin B* after egg activation, the PNG kinase complex leads to Cyclin B/Cdk1 re-activation and entry into the first embryonic mitosis (Fenger et al., 2000; Vardy and Orr-Weaver, 2007). PNG has an indirect role later in embryogenesis in promoting degradation of maternal mRNAs. This degradation requires Smaug (SMG), whose translation also is dependent on PNG at egg activation (Tadros et al., 2007).

We now quantitate changes in translation and the proteome that occur during the oocyte-to-embryo transition, a combined approach not previously used in any organism for this time frame, and identify regulators of this key developmental change. Our comprehensive approach allowed us to determine the extent to which translational changes are reflected in the proteome. These studies reveal that both extensive translational and posttranslational regulatory mechanisms sculpt the proteome at the developmental change from oocyte to embryo, and they emphasize the value of examining both the translome and the proteome.

RESULTS

Translational Changes at Egg Activation

To assess globally the translational changes that occur at egg activation, we performed polysome profiling and ribosome footprinting of mature oocytes and activated eggs (Figure 1A). We used unfertilized *Drosophila* eggs instead of embryos because they undergo all the events of activation, including the completion of meiosis and the onset of expression of proteins that control embryonic patterning (Horner and Wolfner, 2008b). Unfertilized eggs offer an advantage over embryos: they allow monitoring of egg activation while avoiding the potentially confounding changes in protein synthesis and degradation that occur during the subsequent embryonic divisions of fertilized embryos.

In polysome profiling, mRNA-protein complexes are separated by fractionation through linear sucrose gradients (Beilharz and Preiss, 2004). The position of an mRNA within the sucrose gradient reflects its translational status: cosedimentation of mRNAs with ribonucleoproteins (RNPs) or ribosomal subunits suggests a lack of translation, whereas cosedimentation with

polysomes indicates active translation. An abundant monosome peak present in the polysome profiles of mature oocytes was lost at egg activation, accompanied by a >5-fold gain in the relative abundance of polysomes (Figure 1B). To define the translational changes for each mRNA, we isolated mRNAs from six regions within the gradient (Figure 1B) and quantified their abundance by mRNA sequencing (mRNA-seq).

We identified 5,088 mRNAs (~35% of the annotated genes) in sucrose gradient fractions in two independent polysome-profiling experiments (Table S1). As expected, mRNA abundance in unfractionated lysates of mature oocytes and activated eggs, which are devoid of mRNA transcription and degradation, was highly correlated ($R^2 = 0.89$; Figure S1A). This result also indicated that although cytoplasmic polyadenylation has been reported to accompany egg activation, poly(A) selection is a reliable method for isolating mRNA in mature oocytes and activated eggs (Horner and Wolfner, 2008b; Sallés et al., 1994). Importantly, the mRNA abundance within the same fraction from two independent polysome-profiling experiments was highly correlated ($R^2 = 0.72$ – 0.99 ; Figures S1B and S1C), which allowed us to calculate the average percentage distribution of each mRNA across the six fractions.

To confirm that the mRNAs that cosedimented with the polysomal peaks (defined here as ≥ 5 ribosomes) were actively engaged in translation, we treated lysates of mature oocytes and activated eggs with puromycin *ex vivo*, which causes premature termination of elongating ribosomes (Blobel and Sabatini, 1971). The lysates were fractionated and mRNA-seq was performed. Puromycin treatment of mature oocytes and activated eggs was efficient, as it led to complete disassembly of polysomes as measured by OD₂₅₄ (Figures S2A and S2B). Fractionation of puromycin-treated lysates of mature oocytes, but not of activated eggs, revealed that many mRNAs migrated within sucrose gradients in a translation-independent manner (Figures S2C and S2D). Following puromycin treatment of mature oocyte lysates, the median association of an mRNA with fractions corresponding to polysomes equaled a surprisingly high percentage of the total abundance (25%; Figure S2C). In contrast, in activated eggs after puromycin treatment, on average only 9% of the total mRNA cosedimented with the gradient regions corresponding to polysomes (Figure S2D). The large fraction of mRNA from mature oocytes that cosedimented, but was not actually associated with actively translating polysomes, highlights the necessity of the puromycin control in genome-wide polysome-profiling experiments. After correcting for translation-independent migration in the region corresponding to polysomes (see Supplemental Experimental Procedures), we classified 802 mRNAs as recruited onto polysomes, 729 mRNAs as released off the polysomes, and 3,557 mRNAs as unchanged during egg activation (Figures 1C and S2E; Table S4).

The discovery of a large number of translationally regulated mRNAs prompted us to employ an alternative genome-wide method to analyze the change in translational status of mRNAs at egg activation. Ribosome footprint profiling measures the number of ribosome-protected fragments (RPFs) derived from the mRNAs of each gene, which when normalized to mRNA abundance results in a single value for the translation efficiency (TE) of each gene (Ingolia et al., 2009). Moreover, ribosome

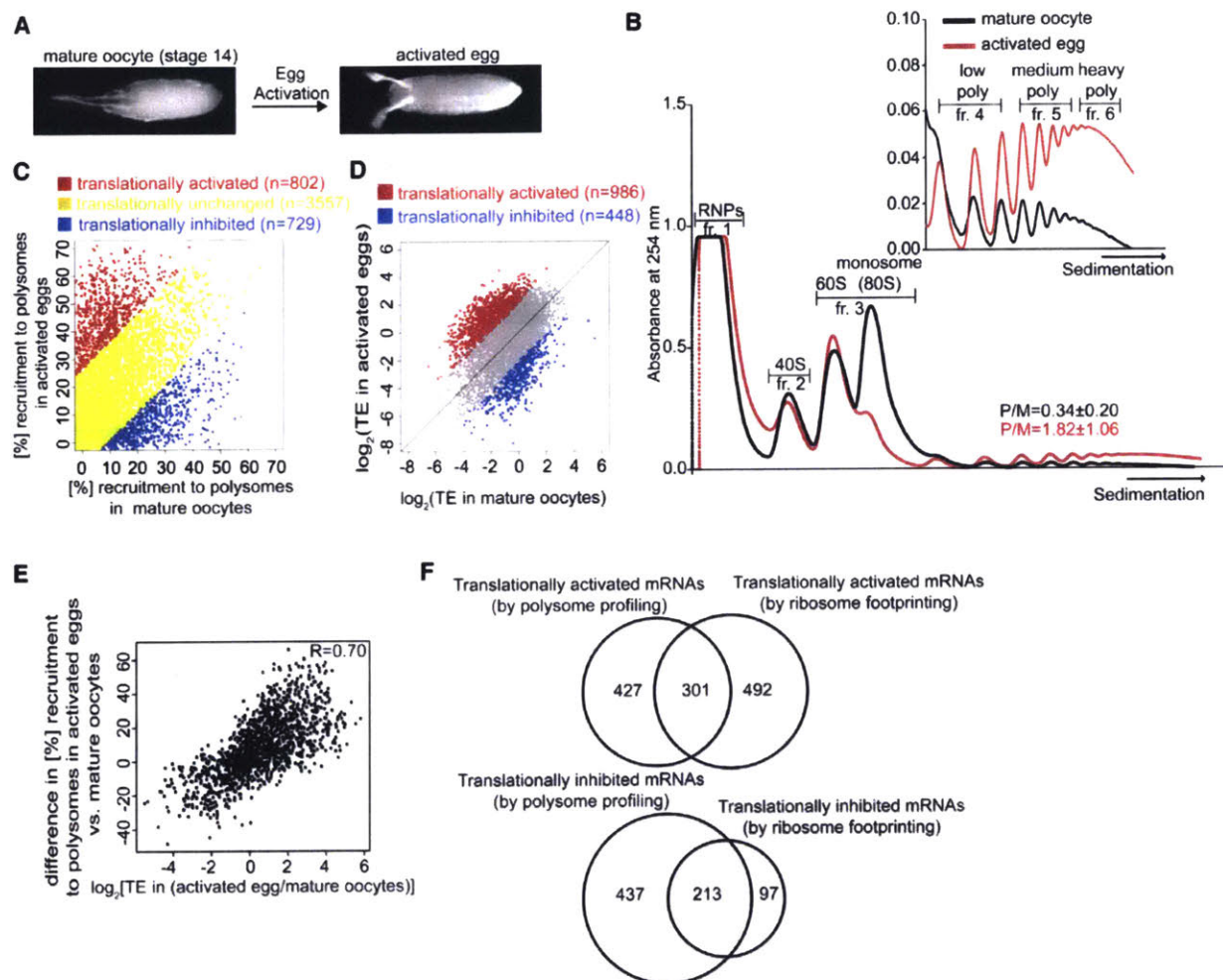


Figure 1. Translational Changes at Egg Activation

(A) Micrograph of a mature, stage 14 oocyte (preactivation state) and an activated egg (postactivation state). Anterior is on the left and dorsal is up.

(B) Representative profile of 254 nm absorbance for wild-type mature oocytes (black) and wild-type activated eggs (red). The inset is an enlargement of the polysomal section of the profile (starting from disomes). The six gradient fractions that were sequenced are labeled. Low polysomes correspond to 2–4 ribosomes, medium polysomes correspond to 5–9 ribosomes, and heavy polysomes correspond to ≥ 10 ribosomes. Polysome/monosome (P/M) ratios averaged from three biological replicate experiments are represented as mean \pm SD.

(C) Comparison of mRNAs associated with polysomes (≥ 5 ribosomes) in mature oocytes and activated eggs. Data were corrected for the presence of mRNAs in the same regions of the gradient after fractionation of puromycin-treated samples. mRNAs were categorized as translationally inhibited (blue) if they had at least 9.1% higher polysomal recruitment in mature oocytes than in activated eggs. Translationally activated (red) mRNAs had at least 26.4% higher polysomal recruitment in activated eggs than in mature oocytes. The cutoffs were chosen because they are 1 SD from the mean difference for all the identified mRNAs. The remaining, translationally unchanged, mRNAs are shown in yellow; 5,088 mRNAs are represented as the mean of two biological replicates.

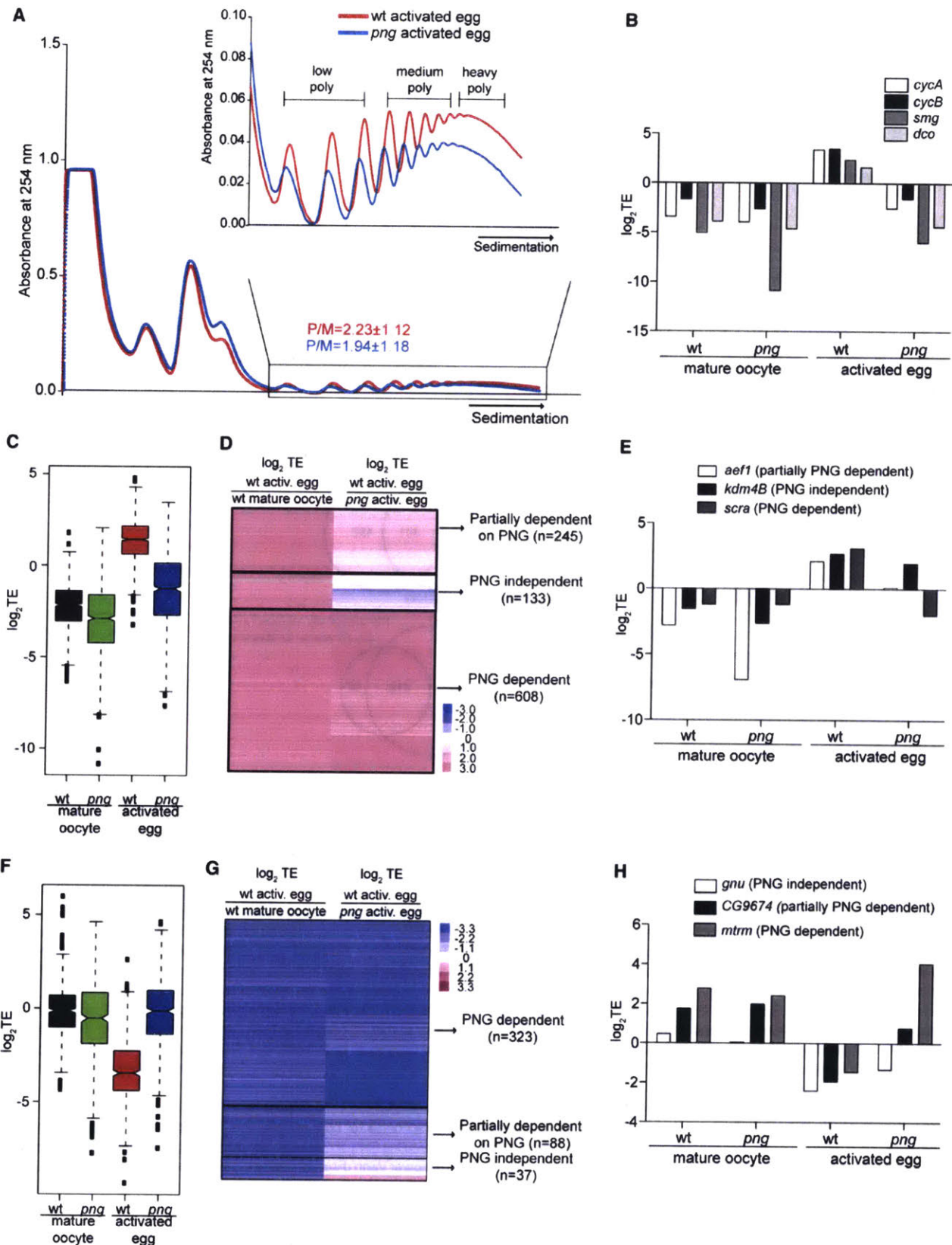
(D) Translation efficiencies (TEs, where $TE = \text{rpkm of RPFs}/\text{rpkm for mRNA abundance}$) in mature oocytes and activated eggs for 5,842 mRNAs. A total of 986 translationally activated mRNAs (red) have ~ 4.1 -fold higher TE in activated eggs than in mature oocytes in both replicates, whereas for 448 translationally inhibited mRNAs (blue) the TE ratio is ~ 4.1 -fold lower in activated eggs than in mature oocytes in both replicates. The cut-off of a ~ 4.1 -fold change in TEs was chosen because it represents 1 SD of the ratios of TEs between activated eggs and mature oocytes in the samples prepared in the absence of cycloheximide. The mean of two biological replicates is shown.

(E) Correspondence between the two complementary methods to measure the translational status of mRNAs in activated eggs versus mature oocytes; 4,580 mRNAs, detected by both approaches, are shown as the average of two biological replicates for both experiments. The Spearman R value is indicated.

(F) The upper Venn diagram compares the number of mRNAs identified by polysome profiling (left) or ribosome footprinting (right) as translationally upregulated at egg activation, applying the criteria described in Figures 1C and 1D, respectively. The lower Venn diagram presents translationally inhibited mRNAs. Here and for all of the other Venn diagrams, only factors identified by both approaches (or in all compared samples) are represented. See also Figures S1 and S2, and Tables S1 and S2.

footprinting reports the position of ribosomes on mRNAs. We observed an unusual accumulation of ribosomes on start codons specifically in mature oocytes, even though no cycloheximide

was used (Figure S2F). This feature, however, seems to be unrelated to translational regulation for two reasons: (1) even in mature oocytes, RPFs at the start codon comprised only a small



(legend on next page)

fraction of the RPFs for each gene (Figure S2F), and (2) initiation codon occupancy was not predictive of translational status in the activated egg (Figures S2G and S2H).

Two independent ribosome footprinting experiments yielded TE measurements for 5,842 mRNAs in mature oocytes and activated eggs (Figure S2I) and revealed drastic translational changes at egg activation ($R = 0.37$) (Figures 1D and S2J). *Mtrm*, the regulator of timing of oocyte maturation, and 447 other mRNAs were translationally repressed at egg activation, with ~ 4.1 -fold lower TE in activated eggs than in mature oocytes. A total of 986 mRNAs, including positive controls such as the embryonic patterning mRNAs *bcd*, *hb*, *cad*, *tor*, *Tl*, and *smg*, as well as mRNAs encoding cell-cycle regulators (*cycA*, *cycB*, and *stg*), had ~ 4.1 -fold higher TE in activated eggs than in mature oocytes and were translationally activated at egg activation (Tables S2 and S4). We found that 968 of the 986 mRNAs categorized as translationally activated at egg activation, and 390 of the 448 translationally repressed mRNAs were significantly translationally regulated with a p value of < 0.05 between the two replicate experiments. If only a p value cutoff of < 0.05 is applied, 2,435 mRNAs would be considered translationally activated and 989 would be considered translationally repressed. Because the mRNAs identified as translationally regulated at egg activation using a ~ 4.1 -fold-change cutoff were largely contained in the set selected by p value analysis, we chose the former approach for subsequent analyses, as it is more stringent.

The translational changes observed by ribosome footprinting and polysome profiling were in broad agreement (Spearman $R = 0.70$; Figure 1E), especially considering that these methods measure slightly different translational properties. Polysome profiling reveals the translationally active fraction of an mRNA. Ribosome footprinting provides information about the average number of ribosomes on each mRNA, with no insights as to whether only a small fraction of that mRNA is efficiently translated and the bulk is repressed.

Both methods identified 301 and 213 mRNAs as translationally upregulated or downregulated, respectively, at egg activation (Figure 1F). Our data set of 301 mRNAs translationally upregulated at egg activation was significantly enriched (false-discovery rate [FDR] < 0.05) for Gene Ontology (GO)-term categories such as "RNA processing," "mRNA binding," and "chromatin organization" (Figure S2K). Thus, translational acti-

vation might help prepare the egg for the maternal-to-zygotic transition, or these factors might serve transcription-independent roles during the first 2 hr of embryogenesis. Although hundreds of mRNAs were reported as translationally inhibited at egg activation in mice (Potireddy et al., 2006), none had been described in *Drosophila*. The 213 mRNAs we identified as translationally inhibited at egg activation showed enrichment in several GO-term categories, including "cell cycle process," consistent with egg activation involving completion of meiosis. Other enriched GO categories, such as "pyruvate metabolic process," "tRNA aminoacylation," and "nucleotide (ATP) binding," suggest that changes in metabolism or translation could be occurring at the exit of meiosis and entry into embryonic cycles (Figure S2L).

The PNG Kinase Is a Major Translational Regulator at Egg Activation

Having identified hundreds of mRNAs as translationally regulated at egg activation, our next goal was to discover key regulatory factors. One candidate was PNG, a Ser/Thr kinase that was previously demonstrated to promote translation of *cycB* and *smg* at egg activation (Tadros et al., 2007; Vardy and Orr-Weaver, 2007).

A slightly lower polysome/monosome ratio (P/M) in the *png* than wild-type activated eggs suggested a more systemic translational activation failure in *png* eggs (Figure 2A). These differences between wild-type and *png* activated eggs could not be attributed to mRNA expression levels, as mRNA abundance was highly correlated between these two samples (Pearson $R = 0.99$; Figure S3A). To identify translational targets of PNG, we performed ribosome footprinting of *png* mutant mature oocytes and activated eggs. The TEs of the positive controls, *cycB*, *cycA*, and *smg* were lower in *png* activated eggs than in wild-type (Figure 2B; Table S2). In *png* eggs, the TEs of all 986 mRNAs normally translationally upregulated at egg activation were closer to those of either wild-type or *png* mature oocytes than to those of wild-type activated eggs (Figure 2C).

A heatmap depicting the differences in TE in *png* activated eggs compared with the wild-type for each of the 986 translationally upregulated mRNAs shows that many, but not all, mRNAs require PNG function for activation (Figure 2D). A total of 608 mRNAs were dependent on PNG for their translational activation (e.g., *scra*; Figure 2E; Tables S2 and S4). In contrast,

Figure 2. PNG Kinase Regulates the Translational Status of the Majority of mRNAs at Egg Activation

(A) Representative profile of 254 nm absorbance for wild-type (WT, red) and *png* activated eggs (blue) as in Figure 1B, except that the P/M ratios are from two independent experiments.

(B) Graph showing the TEs of four mRNAs translationally upregulated at egg activation in WT, but not in *png* mutants (*png*).

(C) Box plot showing TEs in WT mature oocytes (black), *png* mature oocytes (green), WT (red), and *png* activated eggs (blue) for 986 mRNAs translationally upregulated at WT egg activation. The black lines within each box indicate the median, the edges of the boxes show the first and third quartiles of the values, and whiskers extend to the minimum and maximum values.

(D) A heatmap for 986 translationally activated mRNAs compares the TE ratios of WT activated eggs versus mature oocytes with the ratios of WT versus *png* activated eggs. Three classes defining dependence on PNG for translational activation at egg activation emerged. n, number of mRNAs in each category; TE, mean of two biological replicates.

(E) TEs in WT and *png* mature oocytes, as well as activated eggs, of three mRNAs representative of the three groups described in (D).

(F and G) Same as in (C) and (D), except that 448 translationally inhibited mRNAs are shown.

(H) Same as in (E), except that three mRNAs representative of the three groups classifying dependence on PNG for translational repression at egg activation are shown.

See also Figure S3 and Table S2.

we identified 133 mRNAs as PNG independent (e.g., *kdm4B*) and 245 as partially PNG dependent (e.g., *aef-1*; Figures 2D and 2E). We found that 38% of translationally activated mRNAs were at least partially independent of PNG, suggesting that there are one or more pathways acting in parallel to PNG to govern translation at egg activation, possibly serving divergent goals (Table S4). The PNG-independent pathway(s) may control translational activation of mRNAs encoding proteins involved in chromatin modifications (Figure S3B), whereas the PNG-dependent pathway seems to regulate mRNAs encoding protein kinases and factors involved in cell or tissue morphogenesis as well as embryonic development (Figure S3C).

We also found that many mRNAs translationally repressed at egg activation were dependent on PNG (Figure 2F). Based on differences in TEs in *png* versus wild-type activated eggs, we classified 323 mRNAs translationally inhibited at egg activation as dependent on PNG (e.g., *mtm*), 37 as PNG independent (e.g., *gnu*), and 88 as partially PNG dependent (e.g., CG9674; Figures 2G and 2H; Tables S2 and S4). PNG-independent mRNAs were not enriched in any particular GO-term category (FDR > 0.05).

To examine the developmental timing when PNG is active, we tested whether PNG impacts the translational status of its target mRNAs prior to egg activation. The difference in TE for both translationally activated and repressed mRNAs was more pronounced between wild-type and *png* activated eggs than between mutant and wild-type mature oocytes ($p < 0.0001$; Figures 2C and 2F). Combining polysome profiling and quantitative RT-PCR (qRT-PCR), we found a similar distribution of four different candidate mRNAs within the sucrose gradients of wild-type and *png* mature oocytes (Figure S3D). These results indicate that mRNAs are translationally regulated by PNG at, but not prior, to egg activation.

Translational Upregulation Can Strongly Contribute to Increased Protein Levels at Egg Activation

Translational activity is well correlated with protein levels in *S. cerevisiae* growing in nutrient-rich media (Ingolia et al., 2009). However, the extent of coupling between the transcriptome and proteome has not been investigated in a developmental context. To assess proteome changes accompanying *Drosophila* egg activation, we compared unfertilized eggs with mature oocytes using a quantitative proteomics approach. Between 3,648 and 4,187 proteins were quantified in three biological replicate experiments (Figure S4A). At egg activation, comparing the change in protein levels with either a change in polysomal recruitment (Figures S4B and S4C) or a change in TE (Figures 3A and S4D) of the encoding mRNAs revealed a relatively poor correlation ($R = 0.17$ or 0.2 , respectively). There were many examples of robust changes in TE corresponding to undetectable changes in protein levels. The correlation was not improved by considering changes in RPFs instead of changes in TE (Figure S4E). This is a consequence of egg activation occurring in the presence of constant mRNA levels (Figure S1A), thus resulting in a high correlation between changes in TE and changes in RPF (Spearman $R = 0.99$; Figure S4F).

We found that 291 proteins significantly increased after egg activation (FDR < 0.05; Figures 3A and 3B; Table S4). Among

them were well-characterized regulators of early embryogenesis that served as our positive controls, such as the Cdc25 phosphatase String (Stg), as well as the embryonic patterning factors Toll and SMG (Edgar et al., 1994; Sallés et al., 1994; Tadros et al., 2007; Tables S3 and S4). The list of proteins that increased at egg activation was enriched in some of the same GO-term categories as mRNAs translationally activated at egg activation, i.e., chromatin organization and modification, as well as protein amino acid phosphorylation (Figure 3C). Additionally, the proteins upregulated at egg activation participated in a variety of other processes, supporting the complexity of this developmental transition (Table S4).

Approximately 40% of proteins whose levels increased at egg activation were encoded by mRNAs with a significant increase (>4.1-fold) in TE (category I, 89 proteins; Figure 3D; Table S4). Increased levels of several category I proteins were confirmed by immunoblots (Figures 3E and S4G). Overall, translational upregulation strongly contributed to increases in protein levels at egg activation, as the \log_2 median fold-change in TE for mRNAs encoding these proteins was 1.32 as compared with 0.14 for the entire quantified proteome. However, mechanisms other than translational regulation also play a role. For example, category II contains mRNAs that are not significantly translationally upregulated yet encode proteins whose levels are increased (Figures 3D and S4H; Table S4). In these cases, protein accumulation might be caused by posttranslational or other changes that stabilize these proteins and increase their half-life even in the absence of new synthesis. Alternatively, these proteins might be inherently more stable.

Surprisingly, only 25% of mRNAs identified as translationally activated by ribosome footprinting (or 20% in the case of polysome profiling) encoded proteins whose levels increased at egg activation (Figures 3A, 3D, S4B–S4D, and S4H; Table S4). The rest of the translationally activated mRNAs, category III, encoded proteins that were either unchanged (e.g., Dlg1; Figure 3E) or actually decreased (e.g., Dup; Figure S4G). For example, the level of Dlg1 remained unchanged at egg activation despite a 23.6-fold increase in the TE of its mRNA (Figure 3E). The level of Dup decreased 1.4-fold at egg activation, although the TE of its mRNA increased 2.6-fold (Figure S4G; Tables S2 and S3).

One possible explanation for the existence of category III (unchanged) proteins is that increased protein degradation offsets higher synthesis rates in activated eggs. An alternative explanation is that preexisting stores of these proteins greatly exceed the newly synthesized protein, such that increases from translation make up an undetectably small fraction of the total protein pool. We note that if preexisting protein stores generally obscured our ability to observe newly synthesized protein, this effect was independent of the starting amounts of protein, as we observed no significant difference in the abundance of category III compared with category I proteins in mature oocytes (two-sample Wilcoxon test, $p = 0.09$; Figures S4I–S4K). We also did not detect a significant difference in mRNA abundance between these two categories (two-sample Wilcoxon test, $p = 0.17$; Figure S4L).

Regardless of the mechanism involved, the imperfect relationship between changes in protein levels and changes in TE

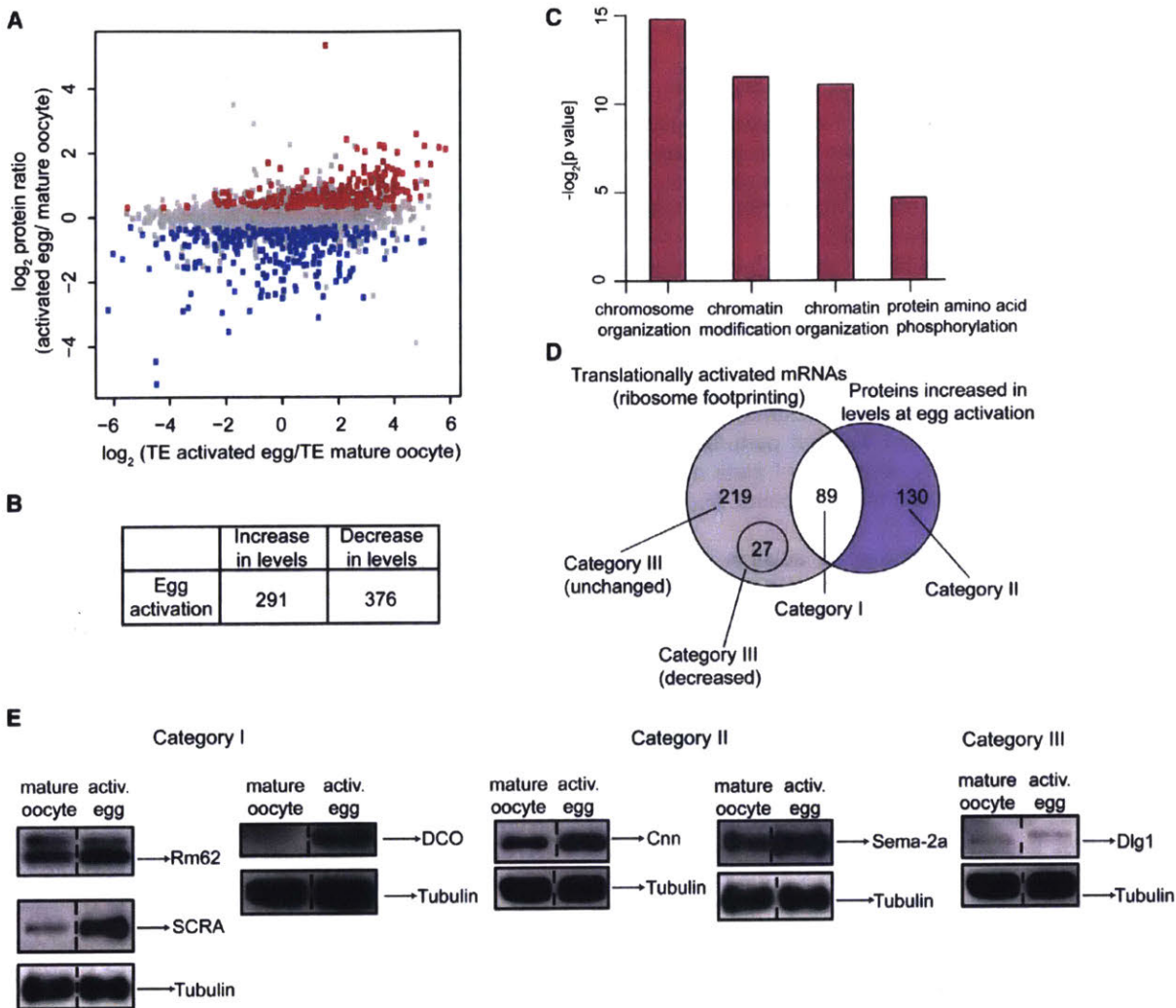


Figure 3. Protein Remodeling during the Oocyte-to-Embryo Transition

(A) Changes in protein levels and TE at egg activation. A scatterplot of the ratio of protein levels in activated eggs versus mature oocytes (mean of three biological replicates) and TE (mean of two biological replicates) is shown. The proteins that were scored as upregulated are shown in red and those scored as down-regulated are in blue; 2,934 data points are presented (identified by both quantitative MS and ribosome footprinting).

(B) Table summarizing the number of proteins that had increased or decreased levels during egg activation.

(C) GO-term categories (FDR, $p < 0.05$) for proteins that were more abundant in activated eggs than in mature oocytes.

(D) Venn diagram comparing the number of translationally activated mRNAs (Figure 1D) that encode proteins upregulated at egg activation (category I). Proteins whose levels increased at egg activation according to quantitative MS without a corresponding increase in TE are shown in category II (violet). mRNAs identified by ribosome footprinting as being translationally upregulated at egg activation but encoding proteins with unchanged or decreased levels at egg activation are in category III (gray). Only factors identified by both approaches are represented.

(E) Western blot validation of candidates belonging to the categories described in (D). Rm62, SCRA (the same membrane reprobated), and DCO are validation examples for category I; Cnn and Sema-2a (reprobated on Rm62/SCRA membrane) are representative of category II; and Dlg1 (reprobated on Cnn membrane) is representative of category III. In this and all subsequent western blots, tubulin was used as a loading control. The dashed line marks that one lane from the original blot is not shown.

See also Figure S4 and Tables S3 and S4.

highlights the value of our integrated approach. When we considered translational changes that were most relevant for egg activation, the acquisition of quantitative proteomics data enabled us to focus on those that had the greatest impact on protein levels (category I). Moreover, proteomics measurements revealed some changes that were not detectable when we examined translation changes (category II), whereas translation measurements uncovered translationally activated mRNAs that

were not detectable when we examined only protein changes (category III).

Lid Activity Is Required in Early Embryogenesis

The limited number of proteins that increase at activation may represent factors that are critical for early embryonic development. It was interesting that the levels of three histone demethylases significantly increased at egg activation, given that early

embryonic development in *Drosophila* is transcriptionally silent. The levels of Little Imaginal Discs (*Lid*) were increased 2.7-fold at egg activation, whereas those of *Kdm4A* and *Kdm4B* increased 3.6- and 2.3-fold, respectively (Table S3). *Lid* is a conserved and essential histone H4-me3-specific demethylase that has been shown to regulate gene expression, particularly of *hox* genes (Gildea et al., 2000; Lloret-Llinares et al., 2008), but it also may be involved in growth control through its association with *dMyc* (Secombe et al., 2007).

To study the role of *Lid* in the oocyte-to-embryo transition, we depleted *Lid* in the female germline by expressing RNAi against *lid* using the UAS-GAL4 system. qRT-PCR analysis of *lid* levels in mature oocytes of two different *lid* RNAi-expressing lines revealed that they were about 99% lower than in the control, demonstrating efficient knockdown (Figure S5A). Importantly, meiotic progression was not perturbed upon *lid* knockdown, as evidenced by metaphase I plate morphology that was comparable to that of the control (Figures S5B and S5C).

To examine the consequence of *lid* ablation on the early embryonic mitotic divisions, we measured the progression through early embryonic cycles by counting the percentage of laid embryos that (1) completed meiosis and thus contained polar bodies arranged into rosettes, but did not initiate embryonic divisions; (2) completed one to four embryonic cycles; (3) completed five to nine embryonic cycles; or (4) reached the syncytial blastoderm, corresponding to division cycles 10–13 (Figures 4A and S5D). The embryos from *lid* RNAi-ablated mothers showed slower progression through embryogenesis than the control. In particular, increased numbers of embryos were in the rosette stage or going through the first-to-fourth embryonic cycles. Consequently, fewer embryos reached later cycles (Figure 4A). There were subsequent embryonic defects, including a reduction in the number of gastrulating embryos and the appearance of embryos with pycnotic or aberrantly aggregated nuclei (Figures 4B and 4C). These data indicate that *Lid* has a role unrelated to transcription that is required during early embryogenesis for proper further embryonic development.

PNG-Mediated Translational Upregulation Provides Insights into the Balance of Translation and Protein Degradation for a Subset of Proteins

It was striking that at egg activation the majority of translationally activated mRNAs encoded proteins whose levels remained equal or even decreased (Figures 3D and S4B–S4D). To test the importance of translational upregulation for those mRNAs, we compared the proteomes of wild-type activated eggs with eggs laid by mothers mutant for the translational regulator PNG (Figure S6A; Table S3). The majority of proteins that increased at egg activation and were encoded by translationally upregulated mRNAs had lower levels in *png* eggs than in wild-type eggs (46 out of 75; Table S4; Figure 5A). Consistently, ~80% of these 46 mRNAs completely depended on PNG for translational upregulation at egg activation (Figure 5B). The other 29 mRNAs encoded proteins with unchanged levels in *png* activated eggs as compared with wild-type (Figures 5A and S6B; Table S4), most likely because a large fraction of them (~59%)

were at least partially independent of PNG for their translational activation at egg activation (Figure 5B).

Interestingly, there were 63 translationally upregulated mRNAs that encoded proteins with stable levels at egg activation but decreased abundance in *png* as compared with wild-type activated eggs (orange segment in Figures 5A and S6B; Table S4). We found that 86% of these 63 mRNAs were dependent on PNG for their translational upregulation at egg activation (Figure 5B). This group of proteins, represented by *Dlg1*, suggested that translational upregulation at egg activation might even be influencing proteins whose abundance does not change (Figure 5C). Another possibility was that inefficient maternal protein deposition in *png* mature oocytes led to the lower levels of proteins in *png* activated eggs versus wild-type (orange segment of Figure 5A), although this was not the case for *Dlg1* (Figure 5C). In addition, the majority of these 63 proteins were present at lower levels in *png* activated eggs as compared with *png* mature oocytes (Figures S6B and S6C; Table S3). Collectively, these results are consistent with the possibility that, for a subset of proteins, translational upregulation compensates for protein turnover occurring at egg activation to maintain rather than increase their levels at this developmental transition.

Downregulation of Protein Levels at Egg Activation Is Largely Posttranslationally Driven

A quantitative comparison of the proteomes of mature oocytes and activated eggs revealed that 376 proteins became significantly less abundant at egg activation ($p < 0.05$; Figures 3A and 3B; Table S4). This group was not enriched in any particular GO category (FDR > 0.05). Downregulation of proteins can be at least partially attributed to the activation of the APC/C at the completion of meiosis, resulting in the proteasomal degradation of its substrates, such as cell-cycle regulators *CycB3* and *Mtrm* (Pesin and Orr-Weaver, 2007; Whitfield et al., 2013). In addition, other degradation pathways are likely to be active.

For 265 of 376 significantly downregulated proteins, we obtained TE information about their mRNAs and found that 43 were translationally downregulated (category IV; Figure 6A; Table S4). In agreement with the proteomics data, immunoblots showed that the levels of Giant Nuclei (GNU, category IV) decreased after egg activation (Figure 6B). The number of downregulated proteins encoded by translationally inhibited mRNAs is significant, as only six of the translationally repressed mRNAs encode proteins whose levels increased during egg activation ($p < 10^{-7}$; Figure 6A). The levels of 226 proteins encoded by translationally inhibited mRNAs were not decreased at egg activation, presumably because the maternal stockpiles of these proteins are stable. We obtained comparable results when we measured the translation status at egg activation as a change in polysomal recruitment (Figure S6D).

Category V proteins showed decreased levels at egg activation even though their encoding mRNAs retained their translational status (Table S4). For example, *CycB3* protein was reduced 5.6-fold although the TE of its encoding mRNA increased 7.6-fold (Tables S2 and S3). Most of the proteins that were downregulated at egg activation have not been studied for their role in meiosis, and many are completely

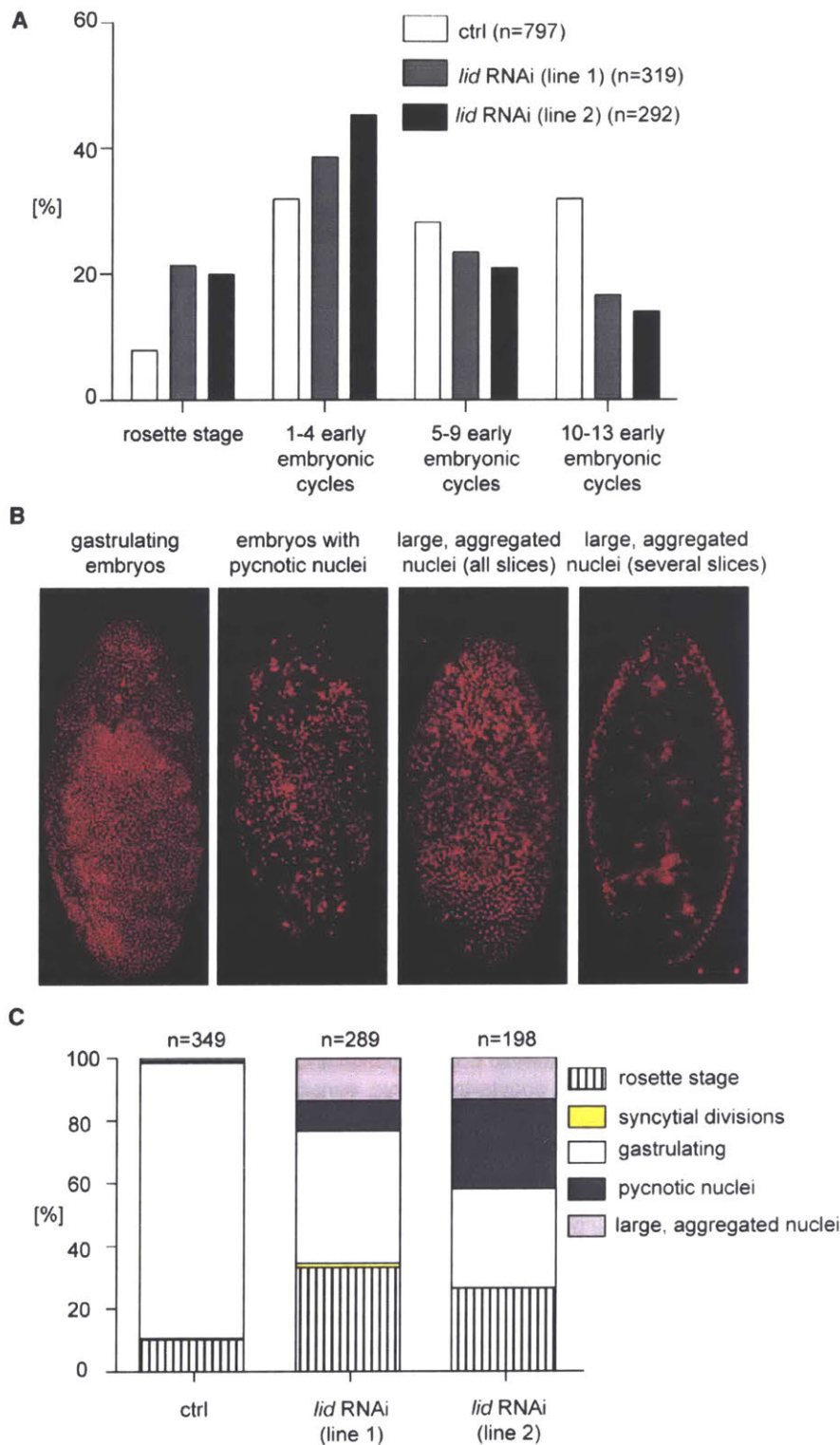


Figure 4. The Histone H3K4 Demethylase *lid* Is Required for Proper Early Embryonic Development

(A) Quantification of the percentage of 0–2 hr embryos that were present in different cycles of embryogenesis after completion of meiosis (as shown in Figure S5D). Embryos were laid either by mothers with only the maternal *a-tubulin-Gal4* driver (control) or by mothers in which a *lid* RNAi line was expressed using the *meta-tubulin-Gal4* driver. *lid* RNAi line 1 is BL35706; line 2 is BL36652. One representative experiment is shown. n, number of embryos scored.

(B) Representative images of 0–2 hr embryos that were aged for an additional 3 hr and stained with a DNA stain (propidium iodide, red). For gastrulating embryos (control) and embryos with pycnotic or aggregated nuclei laid by mothers expressing *lid* RNAi, maximal intensity projections of z stacks are shown. For embryos with large, aggregated nuclei, the figure shows the maximal intensity projections of the entire embryo as well as the optical sections in which aggregated nuclei are particularly visible. In all panels, the dorsal side of the embryo is shown with anterior at the top. Scale bar, 50 μ m. (C) Percentage of properly developed (gastrulating) and aberrantly developed embryos that were collected for 2 hr and aged for an additional 3 hr. Aberrantly developed embryos were classified as still at the rosette stage, in syncytial divisions, gastrulating, or displaying pycnotic or aggregated nuclei (as shown in B). The same genotypes as in (A) were examined. n, number of embryos scored. See also Figure S4.

regulated proteins decreased without translational inhibition of their mRNAs.

Failure in Translational Inhibition for a Subset of mRNAs Interferes with Successful Removal of the Proteins They Encode at the Oocyte-to-Embryo Transition

The modest contribution of translational inhibition to decreased protein levels at egg activation suggested that many of the proteins that were reduced at egg activation in the wild-type would also be reduced in the *png* mutant background. Indeed, 289 proteins were significantly reduced in *png* activated eggs as compared with *png* mature oocytes, and 193 of them were among the 283 proteins that were significantly reduced in wild-type eggs at activation (Figure 6C).

uncharacterized. In addition to revealing candidate regulators of the oocyte-to-embryo transition, these results indicate that for most of the proteins that were downregulated at egg activation, posttranslational rather than translational control governed the decrease in protein levels at egg activation. Most of these down-

Nonetheless, 42 proteins that were reduced at egg activation in the wild-type had higher levels in *png* versus wild-type activated eggs (Figure 6D). Approximately 31% of these proteins, including Mtrm, were encoded by mRNAs that were dependent on PNG for their translational shutoff. Consequently,

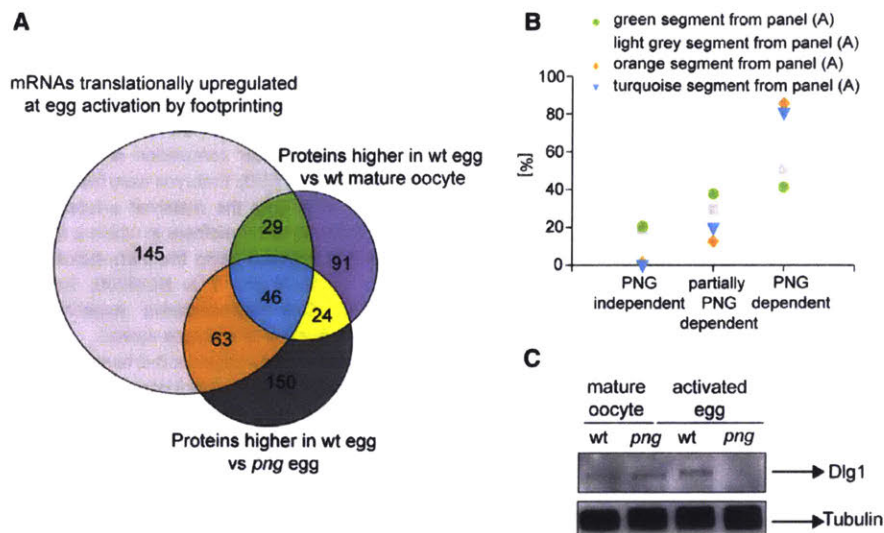


Figure 5. The Translational Regulator PNG Reveals the Importance of Translational Regulation for Homeostasis of Protein Levels for a Subset of Proteins at Egg Activation

(A) Same as Figure 3D, except that the comparison included the proteins whose levels are higher in WT than *png* activated eggs according to MS. Only factors identified in all compared samples are represented. The translationally upregulated mRNAs are from the WT.

(B) Percentage of mRNAs encoding proteins belonging to the light gray, orange, green, or turquoise segments in (A) that are independent, partially dependent, or dependent on PNG for translational upregulation at egg activation.

(C) Western blot validation of Dlg1, a candidate from the orange segment of the Venn diagram in (A). Dlg1 levels do not change at egg activation in the WT, although the protein shows altered mobility. In contrast, protein levels are decreased following activation of *png* mutants (reprobed on the same membrane as in Figure 6F). See also Figure S6 and Tables S3 and S4.

their TEs were overall significantly higher in *png* than in wild-type eggs (Mann-Whitney test, $p < 0.0001$; Figure 6E; Table S4).

The absence of a translational shutoff in *png* eggs allowed us to investigate the role of translational inhibition in decreased protein levels at egg activation, using Mtrm as an example. In parallel, we employed the *mr* mutant, which carries a mutation in the APC2 subunit, to weigh the contributions of APC/C-mediated proteasomal degradation versus translational inhibition to Mtrm protein levels (Kashevsky et al., 2002). Western blotting confirmed the proteome data showing that although there was a substantial decrease in Mtrm levels in both wild-type and *png* activated eggs as compared with mature oocytes, there was persistent Mtrm protein in *png* eggs (Figure 6F). Because the Mtrm levels that remained in *png* activated eggs were lower than those in *mr* activated eggs, we conclude that Mtrm levels are primarily regulated by targeting via the APC/C for proteasomal degradation. Thus, optimal protein decrease at egg activation in addition to protein degradation may require a PNG-dependent shutoff of translation.

DISCUSSION

Relationship between the Translatome and the Proteome

This study provides a demarcation of the contribution of translational and posttranslational regulation to proteome remodeling during a key developmental transition from oocyte to embryo. This transition requires a change from meiosis to mitosis as well as resetting of the oocyte to restore totipotency in the embryo. The absence of transcription and mRNA degradation in the oocyte and early embryo leads to exclusively translational and posttranslational control of gene expression. In this study, the oocyte-to-embryo transition was assessed through quantitative proteomics analysis combined with complementary translatome measurements. We found that translational upregulation strongly contributes to increased protein levels at egg activation.

The effect of translational shutdown on decreased protein levels is more modest, with decreased levels appearing to be largely driven by protein degradation.

The striking translational changes that occurred at egg activation, with as many as 802–986 translationally upregulated and 448–729 inhibited mRNAs, are comparable to those observed in other developmental studies. Polysome profiling at oocyte maturation and egg activation in mouse or in different stages of *Drosophila* embryogenesis also showed that hundreds of mRNAs were both released from and recruited to the polysomes (Chen et al., 2011, 2014; Potireddy et al., 2006; Qin et al., 2007). Thus, this window of development is highly dynamic across species and it relies on translational control.

Mechanism of Translational Control at Egg Activation

We observed both activation and inhibition of mRNA translation at the oocyte-to-embryo transition. Translational inhibition at egg activation may not be an actively regulated process. Rather, the onset of increased translational activity observed at egg activation may limit the available ribosomes, resulting in reduced access to ribosomes and consequently translational repression for the subset of mRNAs.

The Ser/Thr kinase PNG was previously demonstrated to activate the translation of *cycA*, *cycB*, and *smg*, but given the specificity of the mutant phenotypes and results from rescue experiments with overexpression of *cycB*, it was assumed to influence a circumscribed set of targets (Lee et al., 2001). In contrast, we find that PNG is a global regulator of the translational status of mRNAs exclusively at egg activation. In addition, although PNG has been thought to activate translation, our analysis reveals that it acts both positively and negatively, controlling the translational activation of at least 60% of regulated mRNAs and inhibition of 70%. Despite the critical role of PNG in controlling the translational status of the majority of mRNAs at egg activation, it is unlikely that the primary function of PNG kinase is to mediate an activation event upstream in the

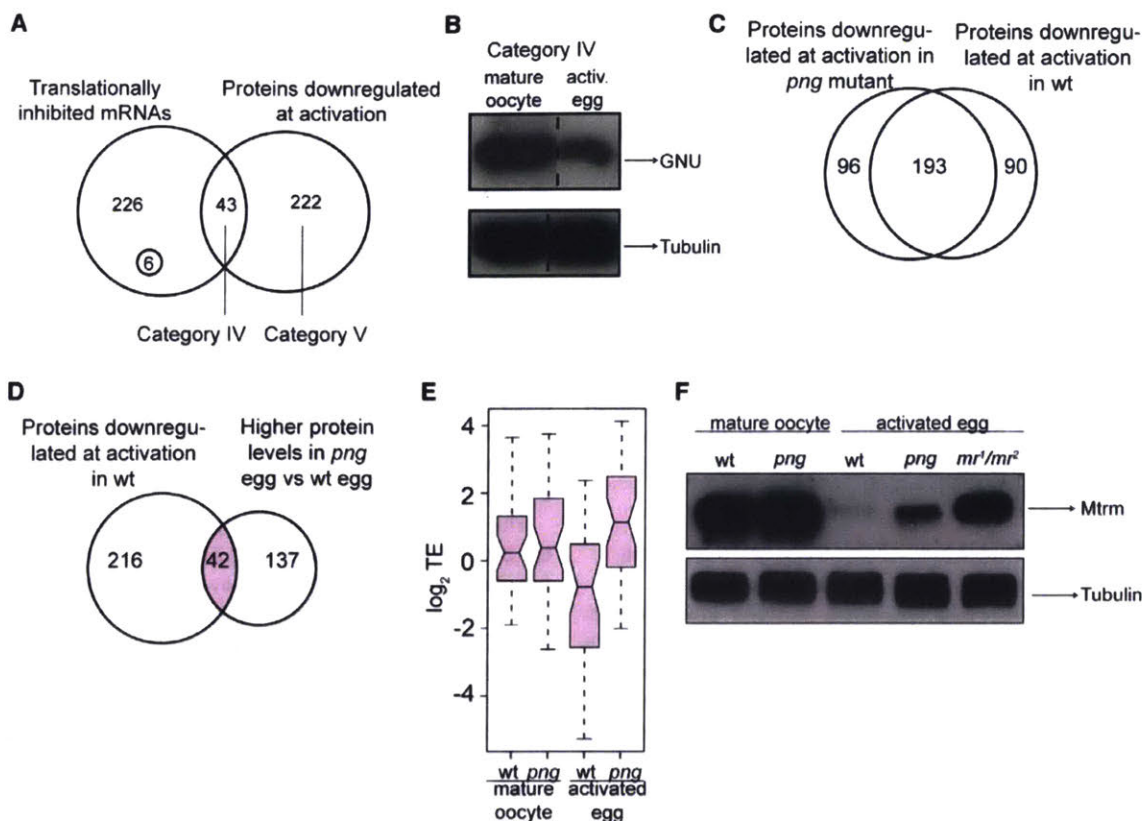


Figure 6. The Absence of a Translational Shutdown at Egg Activation May Interfere with Efficient Downregulation of a Subset of Proteins
 (A) Venn diagram showing the number of translationally inhibited mRNAs (Figure 1D) that encode proteins downregulated at egg activation (category IV). Other proteins whose levels decrease at egg activation according to quantitative MS are in category V. Among translationally inhibited mRNAs there are two groups: the majority of mRNAs encode proteins with unchanged levels at egg activation, whereas only six encode proteins whose levels increase. Only factors identified in both samples are represented.
 (B) Western blot validation of GNU, a protein from category IV.
 (C and D) Venn diagrams comparing the number of proteins that according to MS are downregulated at egg activation in WT or *png* activated eggs (C) or downregulated at activation in WT but more abundant in *png* activated eggs than in WT (D). Only factors identified in both compared samples are represented.
 (E) Box plot showing TEs of mRNAs encoding for 36 proteins in the overlap zone (shown in light pink) of Venn diagram in (D). Only 36 out of 42 mRNAs are presented because the other six were not identified by ribosome footprinting.
 (F) Western blot comparing Mtrm levels at egg activation in WT and *png* mutant background as well as in the activated eggs laid by mothers transheterozygous for female-sterile alleles of the APC2 subunit of APC/C, *morula* (*mr¹/mr²*). In *png* mutants in which translational inhibition of *mtrm* does not occur, Mtrm protein levels are elevated, although to a lesser extent than when the APC/C is mutated.
 See also Figure S6 and Tables S3 and S4.

egg activation pathway. Many activation events, such as eggshell hardening (Tadros et al., 2003) and downregulation of protein levels, occur in *png* mutants.

We postulate that many of the PNG-dependent translational effects reflect a direct role of the kinase. It will be interesting to investigate PNG substrates for candidates whose phosphorylation affects the TEs of mRNAs. It is likely, however, that translation of some mRNAs may be affected as a secondary or indirect effect of PNG activity.

One possible direct target of PNG that affects translation is the translational repressor Pumilio (PUM). The *png* embryonic cell-cycle defect is suppressed by *pum* mutations, which restore Cyclin B protein levels (Vardy and Orr-Weaver, 2007). PNG-PUM antagonism does not appear to be an exclusive mechanism through which PNG mediates translational activation, as there is

no enrichment of PUM-binding sites in the 3' UTRs of PNG-dependent translationally activated mRNAs compared with PNG-independent mRNAs (see Supplemental Discussion for additional 5' and 3' UTR analyses). This suggests that PNG controls translation via multiple targets. Some of its targets may also be translational regulators, as CG17514, a putative activator of translation, is translationally downregulated, whereas the translational repressor Cup is translationally upregulated, in *png* activated eggs.

The large number of puromycin-insensitive mRNAs we observed in mature oocytes was surprising and suggests that many mRNAs are localized to cytoplasmic granules or heavy RNPs at this developmental stage. The inefficient translation of these mRNAs in mature oocytes was corroborated by ribosome footprinting analysis, as they were found to associate with significantly fewer ribosomes compared with puromycin-sensitive

mRNAs (data not shown). Since puromycin only displaces actively translocating ribosomes, a fraction of puromycin-insensitive mRNAs may localize to cytosolic compartments where they are associated with stalled and nonelongating polysomes.

There is precedent for the notion that specific mRNAs are localized to cytoplasmic granules in oocytes. In *Drosophila*, the maternally deposited mRNAs for three patterning genes, *osk*, *grk*, and *bcd*, localize to large RNP complexes (Chekulaeva et al., 2006; Weil et al., 2012). It has been suggested that these cytoplasmic RNPs control both activation and repression of translation: *grk* mRNA is on the periphery of the complex and translated, whereas the internally localized *bcd* mRNA is repressed (Weil et al., 2012). Cytosolic granule association also may serve as a timer for translational activation. In zebrafish and mouse oocytes, association of *cycB* mRNA with cytosolic granules is not necessary for translational repression, but release out of the granules leads to premature CycB synthesis (Kotani et al., 2013).

Insights into Key Regulators of the Oocyte-to-Embryo Transition

By identifying translationally regulated mRNAs and proteins whose levels change at egg activation, this work also highlights potential key processes in the oocyte-to-embryo transition and reveals previously unrecognized regulators. As a proof of principle, we demonstrated that one of the 291 proteins that are upregulated at egg activation, the histone H3K4 demethylase LID, is required for timely progression through early embryonic cycles and proper further embryonic development. Other upregulated proteins may be required immediately at the onset of embryogenesis.

Downregulation of specific proteins at the oocyte-to-embryo transition also is important. This has been demonstrated for the Polo kinase inhibitor Mtrm, whose degradation is required for proper embryogenesis (Whitfield et al., 2013). The microtubule-severing enzyme katanin subunit Kat80 is controlled similarly (Tables S1–S4). In *C. elegans*, katanin (MEI-1) is required for assembly of the meiotic spindle, but failure to decrease its levels before embryonic mitosis leads to spindle defects (Quintin et al., 2003; Stitzel et al., 2006). The translational inhibition applied to these mRNAs may combine with proteasomal degradation for faster and more complete removal of meiotic proteins.

“Resetting” the Proteome at the Oocyte-to-Embryo Transition

Our transcriptome and proteome survey of *png* eggs provided an unexpected insight into the possible coupling of translational regulation and protein degradation at egg activation. There is a class of 63 mRNAs that are translationally upregulated without a perceptible increase in their protein levels in wild-type eggs. These proteins seem to be properly maternally deposited in *png* mutant oocytes, but their mRNAs require PNG for translational upregulation at egg activation. Lower levels of these proteins in *png* versus wild-type activated eggs showed that a failure of translational activation resulted in easily detected decreases in protein levels. Hence, the apparently constant levels of these proteins in the wild-type background were not merely a consequence of high maternal stores overwhelming the translational activation. Moreover, the proteome changes in

png mutants are unlikely to result from an additional independent effect of PNG on protein stability, as quantitative mass spectrometry (MS) data show that downregulation of protein levels occurs comparably in wild-type and *png* mutant eggs.

Thus, the simplest explanation for these results is that translational activation compensates for protein degradation, leading to no net change in protein abundance in the wild-type background. Although the *png* mutant allowed us to demonstrate this for 63 proteins, it is likely that additional proteins are reset. An additional possibility is that for some of these proteins, there is a spatial control in the embryo, with synthesis at one site and degradation at another. The increased protein degradation that normally occurs at egg activation probably is supported by mechanisms in addition to the APC/C, as we found at least 11 E2 Ubiquitin-conjugating enzymes, E3 Ubiquitin ligases, or their putative regulators among the proteins that were upregulated at egg activation.

The conclusions from our analysis of developmental control of the proteome at *Drosophila* egg activation contrast with a previous study in mouse fibroblasts in which the contribution of protein degradation to protein abundance is minor (Schwanhäusser et al., 2011). Unlike *Drosophila* activated eggs, mouse fibroblasts are transcriptionally active, a steady-state system that lacks maternally loaded proteins, potentially resulting in different requirements for translational regulation in controlling protein levels.

Why would a subset of proteins in activated eggs undergo an energetically costly process of increased synthesis and degradation only to maintain the same levels as in mature oocytes? We suggest that the developmental coordination between translation and protein stability restores the proteome, changing these proteins from their “oogenesis” form to their “embryogenesis” form. As some posttranslational modifications might interfere with the embryonic functions of a given protein, an expedient mechanism to remove these modifications might be to degrade and then resynthesize the proteins. Such a “resetting” of this subset of the proteome at egg activation resembles a phenomenon that occurs later in embryogenesis at the mid-blastula stage, in which the transcriptome is reset to allow zygotic control of development (Tadros and Lipshitz, 2009).

EXPERIMENTAL PROCEDURES

Quantitative Mass Spectrometry

Mature oocytes were hand dissected in Grace's Unsupplemented Insect Media (Gibco) from 4-day-old flies that had been fattened for 3 days with wet yeast at 22°C. Activated eggs collected for 0–2 hr (laid by wild-type *Oregon R* females mated with spermless *twine*^{HBS} males) were dechorionated, lysed, and sonicated, and the supernatants were frozen. Digestion of the proteins and stable isotope labeling of the peptides (peptide dimethylation) were performed as previously described (Wiñniewski et al., 2009; Boersema et al., 2009). The labeled peptides were fractionated, desalted, and separated using the nanoAcquity UPLC system (Waters), from which they were directed to an LTQ Orbitrap Velos (Thermo Fisher Scientific) using a Proxeon nanospray source. The MS raw data were processed using MaxQuant (version 1.1.1.25) (Cox and Mann, 2008) and MS/MS spectra were searched using the Andromeda search engine (Cox et al., 2011) against a Uniprot *Drosophila melanogaster* database. Statistical analysis of MS data was performed using the Limma package in R/Bioconductor (Gentleman et al., 2004). Further details are provided in Supplemental Experimental Procedures.

Western Blots

Samples were lysed and western blots performed as previously described (Whitfield et al., 2013). The antibodies used are described in Supplemental Experimental Procedures.

Polysome Analysis, Ribosome Footprint Profiling, RNA Isolation, and mRNA-Seq

Samples were lysed as described previously (Mermoud and Crippa, 1978) and flash frozen. For puromycin treatment, samples were prepared as described previously (Clark et al., 2000). The samples were run on 10%–50% linear sucrose gradients with 0.5 mg/ml cycloheximide. Cycloheximide was excluded for the gradients in the puromycin-treatment experiments. Following centrifugation and fractionation, prior to RNA extraction, 5 ng of *in vitro* transcribed Firefly luciferase (5 ng; Promega) and 5 ng of *S. cerevisiae* mRNA were added to each pooled fraction to allow for normalization between the fractions.

RNA was isolated from whole lysates of mature oocytes or activated eggs by homogenizing them in TRIzol (Invitrogen) according to the manufacturer's instructions. To extract RNA from sucrose gradient fractions, 0.5% SDS and 200 μ g/ml of Proteinase K (Sigma-Aldrich) were added for 30 min at 50°C, followed by RNA isolation using the hot acid phenol method.

For ribosome footprint profiling, samples were thawed on ice and then triturated four times with a 26-gauge needle. After centrifuging to clear the lysate, ribosome profiling and mRNA-seq were performed as described previously (Subtelny et al., 2014) using a detailed protocol available at <http://bartellab.wi.mit.edu/protocols.html>. A replicate of each sample was prepared with cycloheximide excluded from all solutions.

To sequence mRNAs, 1 μ g of total RNA was poly(A) selected using Sera-Mag magnetic oligo(dT) magnetic particles (Thermo Scientific). Barcoded mRNA-seq libraries were made according to the manufacturer's instructions (Illumina), with the exception that mRNA was fragmented using a RNA fragmentation kit (Ambion). The mRNA-seq analysis is described in Supplemental Experimental Procedures.

The average TE from two independent ribosome footprinting experiments obtained by our method of isolating and preparing lysates from 0–2 hr unfertilized eggs correlated well (Spearman $R = 0.89$) with TEs recently published for fertilized 0–2 hr embryos cryolysed in the absence of dechorionation (Dunn et al., 2013).

Immunofluorescence and Imaging

Embryos were collected at 25°C for 2 hr (and, if indicated, aged for an additional 3 hr at 25°C), dechorionated, fixed, and stained as described previously (Pesin and Orr-Weaver, 2007). Images were acquired on an LSM 700 microscope (Carl Zeiss) and processed using ImageJ software.

ACCESSION NUMBERS

The polysome profiling and ribosome footprinting sequencing data reported in this work have been deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GSE52799.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Discussion, Supplemental Experimental Procedures, six figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2014.05.002>.

ACKNOWLEDGMENTS

We thank A. Amon, W. Gilbert, J. Richter, M. Hara, B. Petrova, O. Rissland, and I. Ulitsky for helpful comments on the manuscript. We are grateful to J. Kirkpatrick from the EMBL Proteomics Core Facility for help in processing the samples for quantitative MS analysis. We thank G. Bell for help with bioinformatics analysis and T. DiCesare for help with the graphical abstract. We acknowledge A. Imhof, J. Brill, J. Price, T. Kaufman, E. Wahle, and M. Asano for the kind gifts of antibodies. The Sema-2a and Dlg1 antibodies developed by C. Goodman

were obtained from the Developmental Studies Hybridoma Bank. We thank the TRiP at Harvard Medical School (NIH/NIGMS R01-GM084947) and the Bloomington Stock Center for providing the transgenic RNAi fly stocks used in this study. I.K. was supported by a Feodor Lynen Postdoctoral Fellowship from the Alexander von Humboldt Foundation. This research was funded by NIH grant GM39341 (to T.L.O.-W.). D.P.B. is an HHMI investigator and T.L.O.-W. is an American Cancer Society Research Professor.

Received: November 4, 2013

Revised: March 10, 2014

Accepted: May 1, 2014

Published: May 29, 2014

REFERENCES

- Anderson, K.V., and Lengyel, J.A. (1979). Rates of synthesis of major classes of RNA in *Drosophila* embryos. *Dev. Biol.* 70, 217–231.
- Beilharz, T.H., and Preiss, T. (2004). Translational profiling: the genome-wide measure of the nascent proteome. *Brief. Funct. Genomics Proteomics* 3, 103–111.
- Blobel, G., and Sabatini, D. (1971). Dissociation of mammalian polyribosomes into subunits by puromycin. *Proc. Natl. Acad. Sci. USA* 68, 390–394.
- Boersema, P.J., Rajmakers, R., Lemeer, S., Mohammed, S., and Heck, A.J. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* 4, 484–494.
- Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335, 552–557.
- Carlile, T.M., and Amon, A. (2008). Meiosis I is established through division-specific translational control of a cyclin. *Cell* 133, 280–291.
- Charlesworth, A., Meijer, H.A., and de Moor, C.H. (2013). Specificity factors in cytoplasmic polyadenylation. *Wiley Interdiscip. Rev. RNA* 4, 437–461.
- Chekulaeva, M., Hentze, M.W., and Ephrussi, A. (2006). Bruno acts as a dual repressor of oskar translation, promoting mRNA oligomerization and formation of silencing particles. *Cell* 124, 521–533.
- Chen, J., Melton, C., Suh, N., Oh, J.S., Horner, K., Xie, F., Sette, C., Billech, R., and Conti, M. (2011). Genome-wide analysis of translation reveals a critical role for deleted in azoospermia-like (Dazl) at the oocyte-to-zygote transition. *Genes Dev.* 25, 755–766.
- Chen, L., Dumelie, J.G., Li, X., Cheng, M.H., Yang, Z., Laver, J.D., Siddiqui, N.U., Westwood, J.T., Morris, Q., Lipshitz, H.D., and Smibert, C.A. (2014). Global regulation of mRNA translation and stability in the early *Drosophila* embryo by the Smaug RNA-binding protein. *Genome Biol.* 15, R4.
- Chu, S., and Herskowitz, I. (1998). Gametogenesis in yeast is regulated by a transcriptional cascade dependent on Ndt80. *Mol. Cell* 1, 685–696.
- Clark, I.E., Wyckoff, D., and Gavis, E.R. (2000). Synthesis of the posterior determinant Nanos is spatially restricted by a novel cotranslational regulatory mechanism. *Curr. Biol.* 10, 1311–1314.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.
- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805.
- Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R., and Weissman, J.S. (2013). Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *eLife* 2, e01179.
- Edgar, B.A., Sprenger, F., Duronio, R.J., Leopold, P., and O'Farrell, P.H. (1994). Distinct molecular mechanisms regulate cell cycle timing at successive stages of *Drosophila* embryogenesis. *Genes Dev.* 8, 440–452.
- Fenger, D.D., Carminati, J.L., Burney-Sigman, D.L., Kashevsky, H., Dines, J.L., Elfving, L.K., and Orr-Weaver, T.L. (2000). PAN GU: a protein kinase that

- inhibits S phase and promotes mitosis in early *Drosophila* development. *Development* 127, 4763–4774.
- Gebauer, F., Xu, W., Cooper, G.M., and Richter, J.D. (1994). Translational control by cytoplasmic polyadenylation of *c-mos* mRNA is necessary for oocyte maturation in the mouse. *EMBO J.* 13, 5712–5720.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5, R80.
- Gildea, J.J., Lopez, R., and Shearn, A. (2000). A screen for new trithorax group genes identified little imaginal discs, the *Drosophila melanogaster* homologue of human retinoblastoma binding protein 2. *Genetics* 156, 645–663.
- Horner, V.L., and Wolfner, M.F. (2008a). Mechanical stimulation by osmotic and hydrostatic pressure activates *Drosophila* oocytes in vitro in a calcium-dependent manner. *Dev. Biol.* 316, 100–109.
- Horner, V.L., and Wolfner, M.F. (2008b). Transitioning from egg to embryo: triggers and mechanisms of egg activation. *Dev. Dyn.* 237, 527–544.
- Ingolia, N.T., Ghaemmhami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223.
- Kashevsky, H., Wallace, J.A., Reed, B.H., Lai, C., Hayashi-Hagihara, A., and Orr-Weaver, T.L. (2002). The anaphase promoting complex/cyclosome is required during development for modified cell cycles. *Proc. Natl. Acad. Sci. USA* 99, 11217–11222.
- Kotani, T., Yasuda, K., Ota, R., and Yamashita, M. (2013). Cyclin B1 mRNA translation is temporally controlled through formation and disassembly of RNA granules. *J. Cell Biol.* 202, 1041–1055.
- Lee, L.A., Elfring, L.K., Bosco, G., and Orr-Weaver, T.L. (2001). A genetic screen for suppressors and enhancers of the *Drosophila* PAN GU cell cycle kinase identifies cyclin B as a target. *Genetics* 158, 1545–1556.
- Lee, L.A., Van Hoewyk, D., and Orr-Weaver, T.L. (2003). The *Drosophila* cell cycle kinase PAN GU forms an active complex with PLUTONIUM and GNU to regulate embryonic divisions. *Genes Dev.* 17, 2979–2991.
- Lloret-Llinares, M., Carré, C., Vaquero, A., de Olano, N., and Azorín, F. (2008). Characterization of *Drosophila melanogaster* JmjC+N histone demethylases. *Nucleic Acids Res.* 36, 2852–2863.
- Mermod, J.J., and Crippa, M. (1978). Variations in the amount of polysomes in mature oocytes of *Drosophila melanogaster*. *Dev. Biol.* 66, 586–592.
- Pesin, J.A., and Orr-Weaver, T.L. (2007). Developmental role and regulation of cortex, a meiosis-specific anaphase-promoting complex/cyclosome activator. *PLoS Genet.* 3, e202.
- Potreddy, S., Vassena, R., Patel, B.G., and Latham, K.E. (2006). Analysis of polysomal mRNA populations of mouse oocytes and zygotes: dynamic changes in maternal mRNA utilization and function. *Dev. Biol.* 298, 155–166.
- Qin, X., Ahn, S., Speed, T.P., and Rubin, G.M. (2007). Global analyses of mRNA translational control during early *Drosophila* embryogenesis. *Genome Biol.* 8, R63.
- Quintin, S., Mains, P.E., Zinke, A., and Hyman, A.A. (2003). The mbk-2 kinase is required for inactivation of MEI-1/katanin in the one-cell *Caenorhabditis elegans* embryo. *EMBO Rep.* 4, 1175–1181.
- Sagata, N. (1996). Meiotic metaphase arrest in animal oocytes: its mechanisms and biological significance. *Trends Cell Biol.* 6, 22–28.
- Sallés, F.J., Lieberfarb, M.E., Wreden, C., Gergen, J.P., and Strickland, S. (1994). Coordinate initiation of *Drosophila* development by regulated polyadenylation of maternal messenger RNAs. *Science* 266, 1996–1999.
- 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, 337–342.
- Secombe, J., Li, L., Carlos, L., and Eisenman, R.N. (2007). The Trithorax group protein Lid is a trimethyl histone H3K4 demethylase required for dMyc-induced cell growth. *Genes Dev.* 21, 537–551.
- Shamanski, F.L., and Orr-Weaver, T.L. (1991). The *Drosophila plutonium* and *pan gu* genes regulate entry into S phase at fertilization. *Cell* 66, 1289–1300.
- Stitzel, M.L., Pellettieri, J., and Seydoux, G. (2006). The *C. elegans* DYRK kinase MBK-2 marks oocyte proteins for degradation in response to meiotic maturation. *Curr. Biol.* 16, 56–62.
- 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, 66–71.
- Swan, A., and Schüpbach, T. (2007). The Cdc20 (Fzy)/Cdh1-related protein, Cort, cooperates with Fzy in cyclin destruction and anaphase progression in meiosis I and II in *Drosophila*. *Development* 134, 891–899.
- Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development* 136, 3033–3042.
- Tadros, W., Houston, S.A., Bashirullah, A., Cooperstock, R.L., Semotok, J.L., Reed, B.H., and Lipshitz, H.D. (2003). Regulation of maternal transcript destabilization during egg activation in *Drosophila*. *Genetics* 164, 989–1001.
- Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. *Dev. Cell* 12, 143–155.
- Tay, J., Hodgman, R., and Richter, J.D. (2000). The control of cyclin B1 mRNA translation during mouse oocyte maturation. *Dev. Biol.* 221, 1–9.
- Vardy, L., and Orr-Weaver, T.L. (2007). The *Drosophila* PNG kinase complex regulates the translation of cyclin B. *Dev. Cell* 12, 157–166.
- Weil, T.T., Parton, R.M., Herpers, B., Soetaert, J., Veenendaal, T., Xanthakis, D., Dobbie, I.M., Halstead, J.M., Hayashi, R., Rabouille, C., and Davis, I. (2012). *Drosophila* patterning is established by differential association of mRNAs with P bodies. *Nat. Cell Biol.* 14, 1305–1313.
- Whitfield, Z.J., Chisholm, J., Hawley, R.S., and Orr-Weaver, T.L. (2013). A meiosis-specific form of the APC/C promotes the oocyte-to-embryo transition by decreasing levels of the Polo kinase inhibitor matrimony. *PLoS Biol.* 11, e1001648.
- Wiñiewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362.
- Young, R.A. (2011). Control of the embryonic stem cell state. *Cell* 144, 940–954.
- Zalokar, M. (1976). Autoradiographic study of protein and RNA formation during early development of *Drosophila* eggs. *Dev. Biol.* 49, 425–437.

Appendix B

Altered translation of GATA1 in Diamond-Blackfan anemia

Leif S. Ludwig^{1,2,3,10,11}, Hanna T. Gazda^{3,4,5}, Jennifer C. Eng^{2,3}, Stephen W. Eichhorn^{2,7}, Prathapan Thiru², Roxanne Ghazvinian⁴, Tracy I. George⁸, Jason R. Gotlib⁹, Alan H. Beggs^{4,5}, Colin A. Sieff^{1,5}, Harvey F. Lodish^{2,3,7}, Eric S. Lander^{3,6,7}, Vijay G. Sankaran^{1,2,3,5}

¹Division of Hematology/Oncology, The Manton Center for Orphan Disease Research, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA

²Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA

³Broad Institute, Cambridge, Massachusetts 02142, USA

⁴Division of Genetics and Program in Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, Massachusetts 02115, USA

⁵Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA

⁶Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

⁷Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

⁸Departments of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA

⁹Division of Hematology, Stanford University School of Medicine, Stanford, California 94305, USA

¹⁰Institute for Chemistry and Biochemistry, Freie Universität Berlin, Berlin 14195, Germany

¹¹Charité-Universitätsmedizin Berlin, Berlin 10117, Germany

L.S.L. and V.G.S. conceived the project; L.S.L., H.T.G., J.C.E., S.W.E., R.G., A.H.B., C.A.S. and V.G.S. performed the research; L.S.L., H.T.G., P.T., H.F.L., E.S.L. and V.G.S. analyzed data; T.I.G. and J.R.G. provided clinical assessments; and L.S.L., H.F.L., E.S.L. and V.G.S. wrote the paper with input from all authors.

Published as:

Ludwig, L.S., Gazda, H.T., Eng, J.C., Eichhorn, S.W., Thiru, P., Ghazvinian, R., George, T.I., Gotlib, J.R., Beggs, A.H., Sieff, C.A., Lodish, H.F., Lander, E.S., Sankaran, V.G. (2014). Altered translation of GATA1 in Diamond-Blackfan anemia. *Nat Med* 20, 748-753.

Altered translation of GATA1 in Diamond-Blackfan anemia

Leif S Ludwig¹⁻⁶, Hanna T Gazda^{4,7,8}, Jennifer C Eng^{3,4}, Stephen W Eichhorn^{3,9}, Prathapan Thiru³, Roxanne Ghazvinian⁷, Tracy I George¹⁰, Jason R Gotlib¹¹, Alan H Beggs^{7,8}, Colin A Sieff^{1,2,8}, Harvey F Lodish^{3,4,9}, Eric S Lander^{4,9,12} & Vijay G Sankaran^{1-4,8}

Ribosomal protein haploinsufficiency occurs in diverse human diseases including Diamond-Blackfan anemia (DBA)^{1,2}, congenital asplenia³ and T cell leukemia⁴. Yet, how mutations in genes encoding ubiquitously expressed proteins such as these result in cell-type- and tissue-specific defects remains unknown⁵. Here, we identify mutations in *GATA1*, encoding the critical hematopoietic transcription factor GATA-binding protein-1, that reduce levels of full-length GATA1 protein and cause DBA in rare instances. We show that ribosomal protein haploinsufficiency, the more common cause of DBA, can lead to decreased *GATA1* mRNA translation, possibly resulting from a higher threshold for initiation of translation of this mRNA in comparison with other mRNAs. In primary hematopoietic cells from patients with mutations in *RPS19*, encoding ribosomal protein S19, the amplitude of a transcriptional signature of *GATA1* target genes was globally and specifically reduced, indicating that the activity, but not the mRNA level, of GATA1 is decreased in patients with DBA associated with mutations affecting ribosomal proteins. Moreover, the defective hematopoiesis observed in patients with DBA associated with ribosomal protein haploinsufficiency could be partially overcome by increasing GATA1 protein levels. Our results provide a paradigm by which selective defects in translation due to mutations affecting ubiquitous ribosomal proteins can result in human disease.

Diamond-Blackfan anemia (DBA, OMIM 105650) is characterized by a specific reduction in the production of red blood (erythroid) cells and their precursors without defects in other hematopoietic lineages^{2,6}. In more than 50% of cases, DBA is caused by heterozygous loss-of-function mutations (haploinsufficiency) in 1 of 11 genes encoding ribosomal proteins¹. Moreover, recent studies have shown that haploinsufficiency of ribosomal proteins can contribute to other cell-type-specific diseases in humans, including congenital asplenia and T cell lymphocytic leukemia^{3,4}. How mutations that halve the quantity of ubiquitously expressed ribosomal proteins result in such

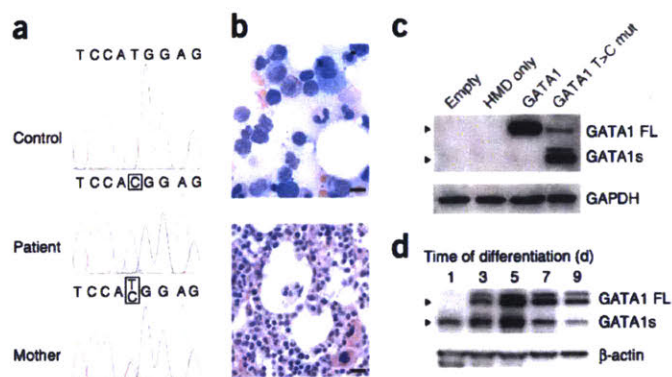
specific human disorders remains unknown. Numerous theories have been proposed for the pathogenesis of these diseases⁷. However, these models do not explain the cell-type specificity of DBA and other ribosomal disorders. The experimental evidence to support such pathogenic models for DBA is often contradictory⁸. Animal models of DBA do not faithfully mimic the disease, and the involvement of different molecular pathways in the hematopoietic defects observed is variable^{9,10}.

We reasoned that identifying genetic causes for the remaining 50% of DBA cases might provide insight into the pathogenesis of this disorder. We recently identified mutations in the *GATA1* gene using whole-exome sequencing¹¹, the first nonribosomal gene to our knowledge to be identified in DBA. *GATA1* encodes a key hematopoietic transcription factor essential for the specification of erythroid cells, as well as megakaryocytes and eosinophils, from early hematopoietic stem and progenitor cells^{11,12}. In humans, *GATA1* mRNA is alternatively spliced to produce two forms of the protein: a long (or full-length) form derived from inclusion of the second exon and a short form without this exon, which therefore lacks the N-terminal 83 amino acids^{11,13}. The *GATA1* mutations that we previously identified occur in the splice donor site of exon 2 and affect splicing by impairing the production of the mRNA encoding the full-length form.

It remained unclear whether the pathogenic mechanisms of these *GATA1* mutations are similar to those of ribosomal protein haploinsufficiency or whether *GATA1* mutations represent a distinct subset of DBA⁷. In an attempt to address this question, we undertook a systematic screening for new *GATA1* mutations in over 200 additional patients with DBA. We identified a highly informative mutation in a male patient who had received a clinical diagnosis of DBA (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1). This mutation changed the first translation initiation codon in *GATA1*, ATG, to an ACG codon (Fig. 1a). Consistent with X-linked inheritance, the patient's asymptomatic mother was a carrier for this mutation (Fig. 1a). When expressed in human 293T cells, wild-type *GATA1* cDNA predominantly produced the full-length form of the protein

¹Division of Hematology and Oncology, Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, Massachusetts, USA. ²Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. ³Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. ⁴Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ⁵Institute for Chemistry and Biochemistry, Freie Universität Berlin, Berlin, Germany. ⁶Charité-Universitätsmedizin Berlin, Berlin, Germany. ⁷Division of Genetics and Genomics, Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, Massachusetts, USA. ⁸Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA. ⁹Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. ¹⁰Department of Pathology, Stanford University School of Medicine, Stanford, California, USA. ¹¹Division of Hematology, Stanford University School of Medicine, Stanford, California, USA. ¹²Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. Correspondence should be addressed to V.G.S. (sankaran@broadinstitute.org).

Figure 1 A *GATA1* mutation in DBA impairs full-length *GATA1* protein production. (a) Sanger sequencing in the region of the first initiator codon from exon 2 of *GATA1* in a male patient with DBA, the patient's mother and a healthy control. (b) Bone marrow aspirate (top, 100× objective magnification, scale bar 10 μm) stained with Wright-Giemsa and a section from a bone marrow biopsy (bottom, 50× objective magnification, scale bar 30 μm) stained with H&E from the patient with DBA. (c) Western blot for detection of full-length (*GATA1* FL) and short (*GATA1s*) forms (arrowheads) of *GATA1* in extracts of 293T cells transfected with no vector (Empty), HMD lentiviral vector (HMD only), HMD-*GATA1* wild-type vector (*GATA1*) or the HMD-*GATA1* mutant vector (*GATA1* T>C mut). GAPDH was used as a loading control, (d) Western blots for detection of *GATA1* FL and *GATA1s* in extracts from primary differentiating erythroid cells in culture. β-actin was used as a loading control.



(Fig. 1c). In contrast, the ACG mutant cDNA predominantly produced the short form of *GATA1* lacking the first 83 amino acids. In addition, this mutant cDNA produced a low level of full-length *GATA1* (Fig. 1c and Supplementary Fig. 2), consistent with the observation that mammalian ribosomes are capable of initiating translation at specific non-AUG codons, including ACG¹⁴. We ruled out the possibility that the residual full-length *GATA1* production resulted from contamination through the use of independent clones and by sequencing (Supplementary Fig. 3). These expression analyses indicate that DBA can be caused by distinct mutations that reduce, but do not entirely abolish, production of full-length *GATA1*. To gain further insight into the physiological relevance of full-length *GATA1* activity, we examined expression of *GATA1* during human erythropoiesis and observed that expression of the full-length protein appears to be specifically upregulated in the course of erythroid differentiation (Fig. 1d).

These results, obtained from study of the *GATA1* initiator codon mutation and from our studies of human erythropoiesis, suggest that the level of *GATA1* full-length protein expression is critical for the promotion of normal erythropoiesis. We reasoned that the more commonly observed mutations in ribosomal protein genes in DBA might result in anemia by reducing production of the full-length *GATA1* protein, thus connecting these two seemingly disparate sets of molecular lesions. To test this hypothesis, we focused initially on the ribosomal protein gene *RPS19*, which is mutated in approximately 25% of DBA cases^{1,2}. We used a primary human erythroid culture system to study haploinsufficiency of *RPS19* using shRNA-mediated knockdown (Fig. 2a)^{15–17}. Reduced levels of *RPS19* protein expression were associated with reduced *GATA1* protein expression (of both the long and short forms) (Fig. 2a). Although this primary cell system contains mostly erythroid lineage cells, there is the possibility that cells can differentiate into non-erythroid cell types, or that differentiation within the erythroid lineage is impaired, which could confound this analysis¹⁶. We therefore turned to the clonal human erythroid K562 cell line to examine whether we could observe similar phenomena in a more homogenous cell population. Upon knockdown of *RPS19*, we observed reduced *GATA1* protein levels (Fig. 2b) without significant effects on the levels of major cellular proteins, as assessed by Coomassie blue staining of cell lysates (Fig. 2c). The levels of other erythroid-important proteins, including transferrin receptor, erythropoietin receptor (EPOR), Janus kinase-2, signal transducer and activator of transcription-5A (STAT5A) and T cell acute lymphocytic leukemia-1 (TAL1), were also unchanged by *RPS19* knockdown (Supplementary Fig. 4a). We observed reduced protein levels of both *RPS19* and *GATA1* within 4 d of infection with shRNA-encoding virus, before major effects on cell growth occurred (Supplementary Fig. 4). Notably, *GATA1* mRNA levels were not

affected (Fig. 2d), consistent with the notion that an effect on protein translation accounts for the decreased levels of *GATA1* protein.

To directly assess whether the decrease in *GATA1* protein levels is due to an effect on translation, we performed polysome profiling and fractionation (Fig. 2e). This approach allows for assessment of mRNA abundance in actively translating ribosomes that are present in multiple copies on a single mRNA species, thereby forming polysomes indicative of productive translation initiation^{18,19}. Whereas the level of *GATA1* mRNA associated with monosomes was similar or higher in *RPS19*-knockdown cells as compared with controls (Supplementary Fig. 5), there was a two- to threefold reduction in *GATA1* mRNA abundance in polysomes (Fig. 2f). In contrast, all other erythroid-important mRNAs tested showed a different pattern, with similar or increased abundance in larger polysomes following *RPS19* knockdown (Fig. 2g). As an alternative approach to demonstrate a selective reduction in *GATA1* translation, we labeled cells with the methionine analog *L*-azidohomoalanine²⁰. When we performed chemical detection of *L*-azidohomoalanine on immunoprecipitated *GATA1*, we found a reduced level of newly translated *GATA1* in cells with reduced *RPS19* levels (Supplementary Fig. 6). Global translation was decreased to ~40% of that in controls (Supplementary Fig. 7), consistent with the slowing of cell growth observed upon *RPS19* knockdown and the known global reduction of translation (48–75%) in patients with DBA^{21,22}.

To gain insight into the mechanism by which ribosomal protein haploinsufficiency results in the observed selective defect in translation, we examined the effect of reducing the expression of other DBA-associated ribosomal proteins. When we used shRNA constructs to reduce expression of RPL11, RPL5 or RPS24, which are collectively mutated in ~15% of DBA cases¹, we consistently noted decreased protein levels of *GATA1* (Fig. 3a–c). The degree of reduction in *GATA1* protein levels corresponded well with the extent of knockdown observed for the various ribosomal proteins (Fig. 3a–c and Supplementary Fig. 8). Highlighting the selectivity of this effect, the protein levels of other regulators of erythropoiesis such as EPOR, TAL1, transferrin receptor (CD71) and STAT5A showed no change after ribosomal protein knockdown (Fig. 3a–c), consistent with the results seen with *RPS19* knockdown.

The similar defect in *GATA1* translation resulting from reduction of ribosomal proteins from either the 40S or 60S subunit of the ribosome suggests that an impairment of translation initiation may underlie this observation, which is consistent with the concept that the availability of free ribosomes is the rate-limiting step in translation initiation²³. Eukaryotic translation initiation factors (eIFs) play a critical role in translation initiation. This process begins as eIF4E initially interacts with the 5' terminal cap of mRNAs and then binds eIF4G, which in

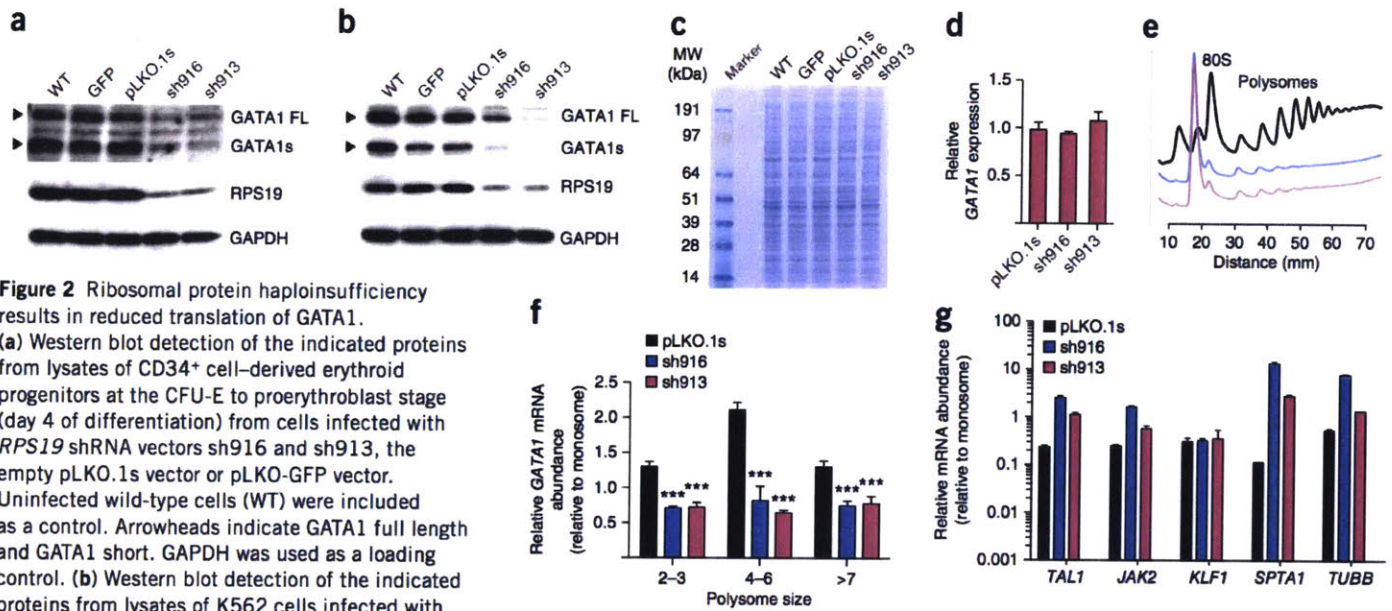


Figure 2 Ribosomal protein haploinsufficiency results in reduced translation of GATA1.

(a) Western blot detection of the indicated proteins from lysates of CD34⁺ cell-derived erythroid progenitors at the CFU-E to proerythroblast stage (day 4 of differentiation) from cells infected with *RPS19* shRNA vectors sh916 and sh913, the empty pLKO.1s vector or pLKO-GFP vector. Arrowheads indicate GATA1 full length and GATA1 short. GAPDH was used as a loading control. (b) Western blot detection of the indicated proteins from lysates of K562 cells infected with

RPS19 shRNA vectors sh916 and sh913. Arrowheads indicate GATA1 full length and GATA1 short. GAPDH was used as a loading control. (c) Coomassie staining of total protein lysates from control-infected or *RPS19* shRNA-infected cells. Molecular weight (MW) markers are shown on the left side of the gel. (d) *GATA1* mRNA levels by quantitative RT-PCR (normalized to β -actin) in *RPS19* shRNA-infected cells. Shown are results for exons 5–6; similar results were obtained for other exons (data not shown); $n = 3$ per group. The error bars shown represent the mean \pm s.d. (e) Polysome profiles from control- or *RPS19* shRNA-treated K562 cells 4 d following infection. The 80S ribosome and polysomes are labeled. The traces are shown offset from one another on the arbitrary y axis (relative absorbance at 254 nm) for ease of visualizing the data; the x axis shows distance along the sucrose gradient. (f) Relative abundance of *GATA1* mRNA in polysome fractions in K562 cells by quantitative RT-PCR (normalized to β -actin; *** $P < 0.0001$ using the unpaired two-tailed Student's *t*-test). Data are shown as the mean \pm s.d. ($n = 3$ per group). (g) Relative abundance of the indicated transcripts as measured by quantitative RT-PCR in larger polysomes (>4 ribosomes) compared to monosomes from K562 cells. Data are normalized to β -actin. Data are plotted on a log₁₀ scale for ease of viewing the range of relative abundance of various erythroid-important mRNAs. Data are shown as the mean \pm s.d. ($n = 3$ per group).

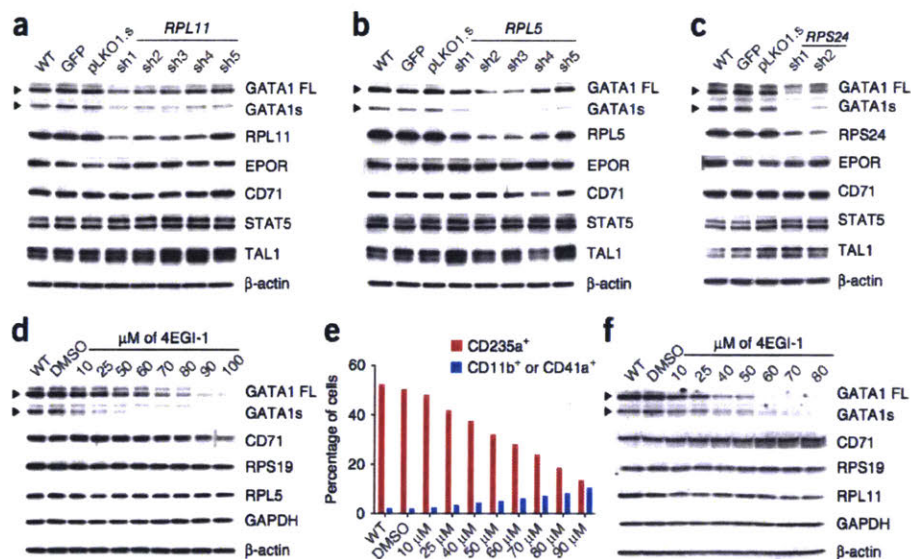
(c) Coomassie staining of total protein lysates from control-infected or *RPS19* shRNA-infected cells. Molecular weight (MW) markers are shown on the left side of the gel. (d) *GATA1* mRNA levels by quantitative RT-PCR (normalized to β -actin) in *RPS19* shRNA-infected cells. Shown are results for exons 5–6; similar results were obtained for other exons (data not shown); $n = 3$ per group. The error bars shown represent the mean \pm s.d. (e) Polysome profiles from control- or *RPS19* shRNA-treated K562 cells 4 d following infection. The 80S ribosome and polysomes are labeled. The traces are shown offset from one another on the arbitrary y axis (relative absorbance at 254 nm) for ease of visualizing the data; the x axis shows distance along the sucrose gradient. (f) Relative abundance of *GATA1* mRNA in polysome fractions in K562 cells by quantitative RT-PCR (normalized to β -actin; *** $P < 0.0001$ using the unpaired two-tailed Student's *t*-test). Data are shown as the mean \pm s.d. ($n = 3$ per group). (g) Relative abundance of the indicated transcripts as measured by quantitative RT-PCR in larger polysomes (>4 ribosomes) compared to monosomes from K562 cells. Data are normalized to β -actin. Data are plotted on a log₁₀ scale for ease of viewing the range of relative abundance of various erythroid-important mRNAs. Data are shown as the mean \pm s.d. ($n = 3$ per group).

turn recruits multiple factors, including eIF4A and eIF3, and allows the ribosome to overcome restrictive motifs in the 5' untranslated region (5' UTR) to scan for the appropriate translation initiation AUG codon²⁴. We therefore used a selective inhibitor of the eIF4E-eIF4G interaction (termed 4EGI-1) to examine whether inhibition of this interaction would also selectively affect *GATA1* translation^{25–27}. Notably, treatment of K562 cells with 4EGI-1 for 48 h led to a marked decrease in *GATA1* protein, whereas other proteins tested were largely unaffected (Fig. 3d). This result shows that *GATA1* mRNA has a more stringent requirement for eIF-dependent translation initiation compared to a number of other transcripts. Knockdown of the expression of *RPS19*, *RPL11*, *RPL5* or *GATA1* in primary CD34⁺ cell cultures resulted in a decreased ratio of erythroid to nonerythroid cells (as assessed using the erythroid marker CD235a, also known as glycoporphin A) (Supplementary Figs. 9 and 10), a phenotype characteristic of DBA^{15,16}. Treatment of the primary cell cultures with 4EGI-1 caused a similar and dose-dependent decrease in the relative amount of erythroid cells (Fig. 3e and Supplementary Fig. 11). Notably, this decrease was associated with a reduction in *GATA1* protein levels but not with changes in the levels of other tested proteins (Fig. 3f). The decreased erythropoiesis observed with 4EGI-1 treatment could largely be rescued by overexpression of *GATA1* (Supplementary Fig. 12), suggesting that *GATA1* is a critical downstream target when the eIF4E-eIF4G interaction is inhibited. Collectively, these results show that the reduction in *GATA1* translation observed with reduced ribosomal protein levels probably reflects a broad sensitivity of *GATA1* mRNA to impairment of translation initiation.

Cells can physiologically regulate gene expression through variation of mRNA translation initiation potential, as different mRNAs

can have variable barriers to translation^{23,24}. A number of studies have demonstrated that transcripts with highly structured 5' UTRs are more poorly translated and have a need for increased initiation potential by the cell^{23,24,28,29}. As the 5' end of human *GATA1* mRNA has never been fully characterized, we performed rapid amplification of the 5' cDNA ends (5' RACE) of human *GATA1* transcripts from both primary erythroid and K562 cells. This analysis revealed a previously unreported 5' end of *GATA1* mRNA that we found on the majority of clones encoding either the short or the long mRNA isoform (Supplementary Fig. 13a,b). The 5' end of human *GATA1* mRNA was predicted to be highly structured (Supplementary Fig. 13c), and thus translation may be more readily impaired in settings where the translation initiation potential is reduced, such as with ribosomal protein haploinsufficiency^{23,29,30}. In support of this notion, we found using a reporter assay that the *GATA1* 5' UTR restricted translation to a greater extent than did other 5' UTRs of similar length (Supplementary Fig. 13d). Additionally, other transcripts with highly structured 5' UTRs showed reduced association with larger polysomes (Supplementary Fig. 14), whereas the association of transcripts with unstructured 5' UTRs with larger polysomes was not affected or was upregulated (Fig. 2g). Therefore, downregulation of translation of *GATA1* mRNA—one of a select group of transcripts with a highly structured 5' UTR—may act as an Achilles' heel during hematopoietic development to impair erythropoiesis specifically in the context of DBA, consistent with our finding that rare mutations in the *GATA1* gene itself are capable of causing DBA. It is possible that other undefined regulatory or structural motifs in the 5' UTR of *GATA1* mRNA could explain the observed effect on translation in the setting of reduced ribosomal protein levels, but we have been

Figure 3 Ribosomal protein deficiency or translation factor inhibition impairs GATA1 translation. (a–c) Western blot detection of the indicated proteins in lysates of primary erythroid cells at day 5 following infection with shRNAs targeting *RPL11* (a), *RPL5* (b) or *RPS24* (c). Arrowheads indicate GATA1 full length and GATA1 short. (d) Western blot detection of the indicated proteins in lysates of K562 erythroid cells at 48 h after treatment with increasing concentrations of the eIF4E–eIF4G interaction inhibitor 4EGI-1. Arrowheads indicate GATA1 full length and GATA1 short. (e) Erythropoiesis (as measured by the percentage of CD235a⁺ cells) and myeloid cell production (as measured by the percentage of CD11b⁺ or CD41a⁺ cells) at 48 h after treatment of primary hematopoietic cells with 4EGI-1. (f) Western blot detection of the indicated proteins in lysates of CD34⁺ cell-derived primary erythroid cells with increasing concentrations of 4EGI-1. Arrowheads indicate GATA1 full length and GATA1 short.

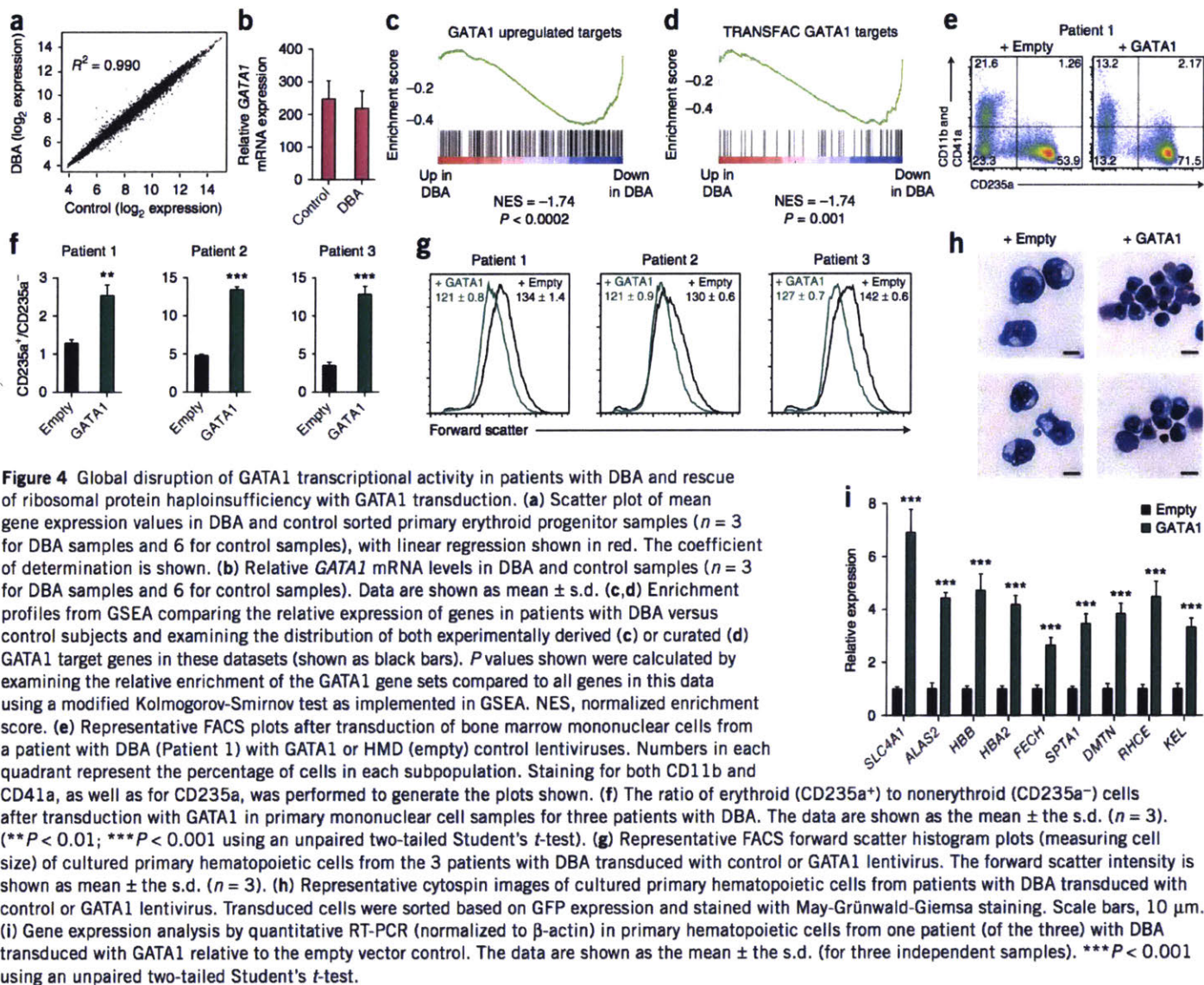


unable to identify any characteristic motifs in this sequence. Mouse *Gata1* mRNA has a much shorter and less structured 5' UTR than that of humans³¹, which may explain the failure of ribosomal protein haploinsufficiency to downregulate translation of this gene and cause major impairments of erythropoiesis in mouse models⁸.

To test whether our findings are of relevance in patients with DBA with mutations in genes encoding ribosomal proteins, we sought to assess GATA1 in erythroid cells from these patients. Given the difficulty of obtaining a sufficient number of stage-matched erythroid cells from patients with DBA to enable examination of GATA1 protein levels, we used RNA expression data to determine whether there was an alteration in the expression of GATA1 target genes. We performed global gene expression profiling on sorted erythroid progenitors (CD34⁺CD71^{high}CD45RA⁻ bone marrow mononuclear cells) from three patients with DBA with known *RPS19* mutations (Online Methods) and six control subjects. This cell surface phenotype is known to selectively enrich for early erythroid progenitors (erythroid burst-forming and erythroid colony-forming units, or BFU-Es and CFU-Es, respectively); similar colony numbers are obtained from healthy individuals and patients with DBA^{32,33}. After normalization of the gene expression data, we examined both global gene expression and expression of a curated set of GATA1 target genes (Supplementary Fig. 15). The global gene expression profiles showed a high degree of correlation between DBA and control erythroid progenitors ($R^2 = 0.990$; Fig. 4a), indicating that we obtained comparable populations of cells from the patients and control subjects. We then applied the gene set enrichment analysis (GSEA) algorithm to assess the expression of GATA1 target genes³⁴. Consistent with our hypothesis, we noted significant global downregulation of GATA1 target genes in erythroid progenitors from patients with DBA compared with controls (Fig. 4c,d, Supplementary Fig. 15 and Supplementary Tables 2–4). Of over 600 target gene sets for other transcription factors (based on the presence of transcription factor binding sites; Online Methods), none demonstrated such a marked and global downregulation, highlighting the selectivity of the effect on GATA1 activity. Notably, the expression of target genes for several other erythroid-important transcription factors and for the tumor suppressor p53 showed no significant changes (Supplementary Fig. 16 and Supplementary Table 5). *GATA1* mRNA levels were comparable between DBA and control samples (Fig. 4b), providing

support for the idea that altered GATA1 activity in DBA cells is mediated at the translational rather than transcriptional level. These results demonstrate that GATA1 activity is decreased in primary samples from patients with DBA with *RPS19* mutations, consistent with the idea that GATA1 dysfunction underlies the erythroid differentiation defect in DBA.

We next examined whether reduced GATA1 levels are sufficient to impair erythropoiesis. When either ribosomal proteins or GATA1 were targeted with shRNAs, we consistently observed increased apoptosis, decreased cell numbers and decreased erythroid differentiation of primary hematopoietic cells (Supplementary Figs. 4, 9 and 10). These findings are consistent with observations made in patients with DBA with mutations in the genes encoding these ribosomal proteins or GATA1¹¹. We modified the *GATA1* cDNA to contain an optimized 5' UTR lacking known structural barriers to translation, such that it should be effectively translated²⁴. Infection with lentivirus harboring this modified *GATA1* cDNA rescued the increased apoptosis of erythroid cells with *RPS19* knockdown, whereas the mutant *GATA1* cDNA that predominantly produces the short form of the protein showed only partial rescue (Supplementary Fig. 17a,b). To assess whether GATA1 could rescue the defect in erythroid differentiation observed in primary hematopoietic cells with reduced ribosomal protein levels, we concomitantly transduced primary cultured CD34⁺ cells with shRNAs targeting *RPL11* or *RPL5* and with the modified *GATA1* cDNA or appropriate controls. Introduction of *GATA1* into these cells improved erythropoiesis (Supplementary Fig. 17c) to an extent comparable to that observed with introduction of the deficient ribosomal protein gene itself³⁵, whereas the mutant version of *GATA1* had a lower level of rescue activity. Notably, introduction of *GATA1* lentivirus into primary bone marrow mononuclear cells from patients with DBA resulted in a two- to fourfold increase in the ratio of CD235a⁺ erythroid to CD235a⁻ nonerythroid cells (Fig. 4e,f). Not only was the frequency of erythroid cells increased, but there was also an improvement in erythroid differentiation, as indicated by smaller cell size (a reduced cell size characterizes more mature erythroid cells), mature erythroid cell morphology and increased expression of genes critical for terminal erythroid maturation (Fig. 4g–i and Supplementary Fig. 18)³⁶. These findings show that GATA1, although likely not the only target of dysregulated protein translation in DBA, is a key factor in mediating the erythroid-specific defect observed in this condition.



Given the limitations in comparing erythroid differentiation in samples from patients with DBA, which show delayed and impaired maturation, with erythroid differentiation in cells from normal individuals³⁷, the extent of this rescue is difficult to quantitatively assess and is likely to be partial. Nonetheless, these results suggest that modulation of GATA1 protein levels in the context of ribosomal protein haploinsufficiency can improve erythropoiesis, with implications for potential therapeutic approaches to ameliorate DBA and related forms of anemia, such as that observed in 5q-myelodysplastic syndrome³⁵.

Our finding of impaired translation of *GATA1* mRNAs by ribosomes in DBA adds to an increasing appreciation for the role of translational control in mediating cell-type-specific gene expression⁵. We showed that mutations in the *GATA1* gene itself can cause DBA by reducing full-length protein levels, and that the more common mutations in genes encoding ribosomal proteins result in impaired erythropoiesis through, at least in part, the same mechanism of decreased GATA1 protein production (Supplementary Fig. 19). Although prior work has implicated p53 in mediating some of the defects observed in DBA^{7,15} and has shown that GATA1 regulates the p53 pathway³⁸, the defects we observed in K562 cells upon GATA1 or ribosomal protein knockdown occurred in the absence of p53 (ref. 39). Thus, the cell-type

specific nature of the DBA phenotype is probably mediated by multiple pathways, including the p53 pathway, that lie downstream of GATA1. Indeed, we observed that p53 expression is upregulated upon reduction of RPS19 or GATA1 levels in primary hematopoietic cells, suggesting that these factors act through a common pathway (Supplementary Fig. 20).

More generally, our findings suggest that mRNAs that are inefficiently translated by the ribosome experience a further reduction in protein translation under conditions of limited ribosome abundance (or with other impairments in translation initiation), such that synthesis of their encoded proteins would be selectively impaired^{23,40}. Alterations in protein translation have been implicated in diverse human diseases including autism, cancer and other blood disorders^{26,27}. Our finding of selectively impaired protein translation resulting from ribosomal protein haploinsufficiency provides a paradigm for understanding the cell-type-specific defects observed in DBA and conditions such as congenital asplenia^{2,3} and childhood leukemia⁴. In addition, the deeper understanding of DBA pathogenesis that is presented here suggests potential therapeutic avenues that would involve targeting of GATA1 protein production through either gene therapy or small-molecule approaches for amelioration of the anemia observed in this disease.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. The microarray data can be found in the Gene Expression Omnibus with accession number [GSE41817](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful to the patients with DBA and their families for their inspiration and encouragement. We thank A. Wakabayashi and J. Ulirsch for valuable assistance and D.G. Nathan, S.H. Orkin, D.A. Williams, S.T. Chou, G.W. Bell, C.R. Walkley, J. Flygare and D.P. Bartel for valuable comments and advice. We are grateful to L. Solomon, L. Gaffney and T. DiCesare for assistance with illustrations. This work was supported by the German National Academic Foundation (to L.S.L.), US National Institutes of Health (NIH) grants R01 HL107558 and K02 HL111156 and a grant from the DBA Foundation (to H.T.G.), NIH grant P01 HL32262 (to H.F.L.), NIH grant U54 HG003067-09 (to E.S.L.) and NIH grant R21 HL120791-01 and a March of Dimes Basil O'Connor Scholar Award (to V.G.S.).

AUTHOR CONTRIBUTIONS

L.S.L. and V.G.S. conceived the project; L.S.L., H.T.G., J.C.E., S.W.E., R.G., A.H.B., C.A.S. and V.G.S. performed the research; L.S.L., H.T.G., P.T., H.F.L., E.S.L. and V.G.S. analyzed data; T.I.G. and J.R.G. provided clinical assessments; and L.S.L., H.F.L., E.S.L. and V.G.S. wrote the paper with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Gazda, H.T. *et al.* Frameshift mutation in p53 regulator RPL26 is associated with multiple physical abnormalities and a specific pre-ribosomal RNA processing defect in diamond-blackfan anemia. *Hum. Mutat.* **33**, 1037–1044 (2012).
- Lipton, J.M. & Ellis, S.R. Diamond-Blackfan anemia: diagnosis, treatment, and molecular pathogenesis. *Hematol. Oncol. Clin. North Am.* **23**, 261–282 (2009).
- Bolze, A. *et al.* Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. *Science* **340**, 976–978 (2013).
- De Keersmaecker, K. *et al.* Exome sequencing identifies mutation in *CNOT3* and ribosomal genes *RPL5* and *RPL10* in T-cell acute lymphoblastic leukemia. *Nat. Genet.* **45**, 186–190 (2013).
- Xue, S. & Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* **13**, 355–369 (2012).
- Nathan, D.G., Clarke, B.J., Hillman, D.G., Alter, B.P. & Housman, D.E. Erythroid precursors in congenital hypoplastic (Diamond-Blackfan) anemia. *J. Clin. Invest.* **61**, 489–498 (1978).
- Weiss, M.J., Mason, P.J. & Bessler, M. What's in a name? *J. Clin. Invest.* **122**, 2346–2349 (2012).
- Horos, R. *et al.* Ribosomal deficiencies in Diamond-Blackfan anemia impair translation of transcripts essential for differentiation of murine and human erythroblasts. *Blood* **119**, 262–272 (2012).
- Torihara, H. *et al.* Erythropoiesis failure due to RPS19 deficiency is independent of an activated Tp53 response in a zebrafish model of Diamond-Blackfan anaemia. *Br. J. Haematol.* **152**, 648–654 (2011).
- Jaako, P. *et al.* Mice with ribosomal protein S19 deficiency develop bone marrow failure and symptoms like patients with Diamond-Blackfan anemia. *Blood* **118**, 6087–6096 (2011).
- Sankaran, V.G. *et al.* Exome sequencing identifies *GATA1* mutations resulting in Diamond-Blackfan anemia. *J. Clin. Invest.* **122**, 2439–2443 (2012).
- Orkin, S.H. & Zon, L.I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
- Wechsler, J. *et al.* Acquired mutations in *GATA1* in the megakaryoblastic leukemia of Down syndrome. *Nat. Genet.* **32**, 148–152 (2002).
- Peabody, D.S. Translation initiation at an ACG triplet in mammalian cells. *J. Biol. Chem.* **262**, 11847–11851 (1987).
- Dutt, S. *et al.* Haploinsufficiency for ribosomal protein genes causes selective activation of p53 in human erythroid progenitor cells. *Blood* **117**, 2567–2576 (2011).
- Flygare, J. *et al.* Deficiency of ribosomal protein S19 in CD34⁺ cells generated by siRNA blocks erythroid development and mimics defects seen in Diamond-Blackfan anemia. *Blood* **105**, 4627–4634 (2005).
- Sankaran, V.G. *et al.* MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. *Proc. Natl. Acad. Sci. USA* **108**, 1519–1524 (2011).
- Rousseau, D., Kaspar, R., Rosenwald, I., Gehrke, L. & Sonenberg, N. Translation initiation of ornithine decarboxylase and nucleocytoplasmic transport of cyclin D1 mRNA are increased in cells overexpressing eukaryotic initiation factor 4E. *Proc. Natl. Acad. Sci. USA* **93**, 1065–1070 (1996).
- Kondrashov, N. *et al.* Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell* **145**, 383–397 (2011).
- Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A. & Schuman, E.M. Selective identification of newly synthesized proteins in mammalian cells using biorthogonal noncanonical amino acid tagging (BONCAT). *Proc. Natl. Acad. Sci. USA* **103**, 9482–9487 (2006).
- Cmejlova, J. *et al.* Translational efficiency in patients with Diamond-Blackfan anemia. *Haematologica* **91**, 1456–1464 (2006).
- Garçon, L. *et al.* Ribosomal and hematopoietic defects in induced pluripotent stem cells derived from Diamond Blackfan anemia patients. *Blood* **122**, 912–921 (2013).
- Shah, P., Ding, Y., Niemczyk, M., Kudla, G. & Plotkin, J.B. Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589–1601 (2013).
- Sonenberg, N., Hershey, J.W.B. & Mathews, M. *Translational Control of Gene Expression* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 2000).
- Moerke, N.J. *et al.* Small-molecule inhibition of the interaction between the translation initiation factors eIF4E and eIF4G. *Cell* **128**, 257–267 (2007).
- Santini, E. *et al.* Exaggerated translation causes synaptic and behavioural aberrations associated with autism. *Nature* **493**, 411–415 (2013).
- Gkogkas, C.G. *et al.* Autism-related deficits via dysregulated eIF4E-dependent translational control. *Nature* **493**, 371–377 (2013).
- Gebauer, F. & Hentze, M.W. Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell Biol.* **5**, 827–835 (2004).
- Babendure, J.R., Babendure, J.L., Ding, J.H. & Tsien, R.Y. Control of mammalian translation by mRNA structure near caps. *RNA* **12**, 851–861 (2006).
- Kozak, M. An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* **115**, 887–903 (1991).
- Kobayashi, E., Shimizu, R., Kikuchi, Y., Takahashi, S. & Yamamoto, M. Loss of the *Gata1* gene IE exon leads to variant transcript expression and the production of a *GATA1* protein lacking the N-terminal domain. *J. Biol. Chem.* **285**, 773–783 (2010).
- Gazda, H.T. *et al.* Defective ribosomal protein gene expression alters transcription, translation, apoptosis, and oncogenic pathways in Diamond-Blackfan anemia. *Stem Cells* **24**, 2034–2044 (2006).
- Lansdorp, P.M. & Dragowska, W. Long-term erythropoiesis from constant numbers of CD34⁺ cells in serum-free cultures initiated with highly purified progenitor cells from human bone marrow. *J. Exp. Med.* **175**, 1501–1509 (1992).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
- Ebert, B.L. *et al.* Identification of *RPS14* as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335–339 (2008).
- Hu, J. *et al.* Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis *in vivo*. *Blood* **121**, 3246–3253 (2013).
- Moniz, H. *et al.* Primary hematopoietic cells from DBA patients with mutations in *RPL11* and *RPS19* genes exhibit distinct erythroid phenotype *in vitro*. *Cell Death Dis.* **3**, e356 (2012).
- Trainor, C.D., Mas, C., Archambault, P., Di Lello, P. & Omichinski, J.G. *GATA-1* associates with and inhibits p53. *Blood* **114**, 165–173 (2009).
- Chylicki, K. *et al.* p53-mediated differentiation of the erythroleukemia cell line K562. *Cell Growth Differ.* **11**, 315–324 (2000).
- Lodish, H.F. Model for the regulation of mRNA translation applied to haemoglobin synthesis. *Nature* **251**, 385–388 (1974).

ONLINE METHODS

Cell culture. 293T cells (ATCC) were maintained in DMEM with 10% FBS (FBS), 2 mM L-glutamine and 1% penicillin-streptomycin (P/S). K562 cells (ATCC) were maintained at a density between 0.1×10^6 and 1×10^6 cells per milliliter in RPMI 1640 medium supplemented with 10% FBS, 2 mM L-glutamine and 1% P/S. Culture of primary human cells is described below. Cells were incubated at 37 °C with 5% CO₂. Where indicated, cultures were supplemented with 10–100 μM 4EGI-1 (sc-202597, Santa Cruz Biotechnology) or DMSO.

Culture of human adult peripheral blood–mobilized CD34⁺ progenitors was performed using a two-stage culture method, as described previously^{17,41}. CD34⁺ cells were obtained from magnetically sorted mononuclear samples of G-CSF–mobilized peripheral blood from donors and were frozen after isolation. Cells were obtained from the Fred Hutchinson Cancer Research Center, Seattle, USA. Cells were thawed and washed into PBS with 1% FBS, pelleted and then seeded in StemSpan SFEM medium (StemCell Technologies, Inc.) with 1 × CC100 cytokine mix (Stem Cell Technologies, Inc.) and 1% P/S. Cells were maintained in this expansion medium at a density between 0.1×10^6 and 1×10^6 cells per milliliter, with medium changes every other day as necessary. Cells were kept in expansion medium for a total of 5 d. After this expansion phase, the cells were reseeded into StemSpan SFEM medium with 1% P/S, 20 ng/mL SCF (PeproTech, Inc.), 1 U/mL Epo (Amgen), 5 ng/mL IL-3 (PeproTech, Inc.), 2 μM dexamethasone (Sigma-Aldrich) and 1 mM β-estradiol (Sigma-Aldrich). Cells were maintained in differentiation medium, with medium changes every other or every third day as needed. Cells were maintained at a density between 0.1×10^6 and 1×10^6 cells per milliliter.

Bone marrow mononuclear cells from patients with DBA and healthy controls were isolated from human bone marrow aspirates using Ficoll-Paque Plus (17-1440-02, GE Healthcare) density gradient centrifugation. Mononuclear cells were cultured in IMDM with 3% human type AB plasma, 2% human AB serum, 1% P/S, 3 U/ml heparin, 200 μg/ml Holo-Transferrin (Sigma-Aldrich), 10 μg/ml insulin, 10 ng/ml SCF (PeproTech, Inc.), 1 ng/ml IL-3 (PeproTech, Inc.), 2 μM dexamethasone (Sigma-Aldrich) and 3 U/ml Epo (Amgen), similarly to recently described protocols for culturing human erythroid cells³⁶. Cells were cultured for 1–2 d before infection as described below and analyzed 4–5 d after infection.

Bone marrow mononuclear cells from patients with DBA or healthy donors were collected after appropriate informed consent was obtained. Patient 1 was male and was 6 months old at the time of bone marrow collection. The patient had been noted to have fussiness and pallor at 4 months of age, and a complete blood count performed at that time showed a hemoglobin level of 4.3 g/dL. The patient remained on monthly transfusions to avoid symptomatic anemia. The patient had no physical anomalies. The erythrocyte adenosine deaminase level was elevated at 1.83 IU per gram of hemoglobin. All 11 known ribosomal protein genes, as well as *GATA1*, were sequenced in this patient, and no definitive pathogenic mutations were found. Patients 2 and 3 were both male and were 34 and 27 years old, respectively, at the time of bone marrow collection. These patients both had a macrocytic anemia that was diagnosed in infancy. Both individuals had required intermittent transfusions, but details of their medical history were limited, as only select records were available for review. Both patient 2 and patient 3 were dependent on steroid therapy to prevent the need for transfusions. Neither individual had any physical anomalies, and erythrocyte adenosine deaminase levels were elevated at 0.99 and 1.92 IU per gram of hemoglobin (normal range 0.33–0.96 IU per gram of hemoglobin), respectively. Patients 2 and 3 had the *RPS19* mutations c.3G>A Met1Ile and c.185G>A Arg62Gln, respectively.

Study approval. All patients or their families provided written informed consent to participate in this study. The use of human cells was approved by the Institutional Biosafety Committee and Institutional Review Board of Boston Children's Hospital and the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology.

Lentiviral vectors and infection. The shRNA constructs targeting human *RPS19* (sh913 and sh916, RefSeq ID NM_001022), human *GATA1* (sh19-23, RefSeq ID NM_002049), human *RPS24* (sh1-2, RefSeq ID NM_001026), human *RPL5* (sh1-5 RefSeq ID NM_000969) and human *RPL11* (sh1-5 RefSeq ID NM_000975) were obtained from the Mission shRNA

collection (Sigma-Aldrich). The constructs were in the pLKO.1-puro lentiviral vector. The sequences of the shRNAs used in this study are listed in **Supplementary Table 6**.

As controls, the lentiviral vectors pLKO-GFP and pLKO.1s (the empty pLKO.1 vector with a 1.2-kb stuffer element) were used (The RNAi Consortium of the Broad Institute of MIT and Harvard). For rescue experiments, erythroid cells were cotransduced with shRNAs targeting ribosomal proteins with either the HMD control, HMD-GATA1 or HMD-GATA1 mutant, which contain the respective cDNAs. The HMD-GATA1 construct was made by subcloning the *GATA1* cDNA into the HMD lentiviral vector using the EcoRI and XhoI restriction sites in this vector. The HMD-GATA1 T>C mutant construct was made by performing site-directed mutagenesis of the original HMD-GATA1 construct using the QuikChange II site-directed mutagenesis kit (Agilent). Double-transduced cells were identified by puromycin selection and GFP expression driven by an IRES-GFP in the HMD vector.

For production of lentiviruses, 293T cells were transfected with the appropriate viral packaging and genomic vectors (pVSV-G and pDelta8.9) using FuGene 6 reagent (Promega) according to the manufacturer's protocol. The medium was changed the day after transfection to RPMI 1640 or StemSpan SFEM medium. After 24 h, this medium was collected and filtered using an 0.45-μm filter immediately before infection of primary hematopoietic or K562 cells. The cells were mixed with viral supernatant in the presence of 8 μg/ml polybrene (Millipore) in a 6-well plate at a density of 250,000–500,000 cells per well. The cells were spun at 2,000 r.p.m. for 90 min at 22 °C and left in viral supernatant overnight. The medium was replaced the morning after infection. Selection of infected cells was started 24 h after infection with 1 μg/ml or 2 μg/ml puromycin for primary hematopoietic cells and K562 cells, respectively. The infection efficiency for pLKO-GFP-infected cells was assessed by measuring the frequency of GFP⁺ cells by flow cytometry 48 h post infection. Typically, the frequency of GFP⁺ cells was between 30–60% and >95% for primary hematopoietic cells and K562 cells, respectively.

Quantitative RT-PCR. Isolation of RNA was performed using the miRNeasy Mini Kit (Qiagen). An on-column DNase (Qiagen) digestion was performed according to the manufacturer's instructions. RNA was quantified by a NanoDrop spectrophotometer (Thermo Scientific). Reverse transcription was carried out using the iScript cDNA synthesis kit (Bio-Rad). Real-time PCR was performed using the ABI 7900 Machine Real-Time PCR system and SYBR green PCR Master Mix (Applied Biosystems). Quantification was performed using the ΔΔC_T method. Normalization was performed using β-actin mRNA as a standard, unless otherwise indicated. The primers used for quantitative RT-PCR are listed in **Supplementary Table 7**.

5' RACE. 5' RACE was conducted using total RNA from K562 cells and *in vitro* cultured primary erythroid cells using the FirstChoice RLM-RACE Kit (AM1700, Life Technologies) according to the manufacturer's instructions. RACE PCR products were subcloned using the TOPO TA cloning kit (450641, Invitrogen) followed by Sanger sequencing. The primers used for 5' RACE are listed in **Supplementary Table 7**.

Western blotting. Cells were harvested at indicated time points, washed twice in PBS, resuspended in RIPA lysis buffer (50 mM Tris-HCl at pH 7.4, 150 mM NaCl, 0.1% SDS, 1% NP-40, 0.25% sodium deoxycholate, 1 mM DTT) supplemented with 1 × Complete Protease Inhibitor Cocktail (Roche) and incubated for 30 min on ice. After centrifugation at 14,000 r.p.m. for 10 min at 4 °C to remove cellular debris, the remaining supernatant was transferred to a new tube, supplemented with sample buffer and incubated for 10 min at 70 °C. Equal amounts of proteins were separated by SDS gel electrophoresis using the NuPAGE Bis-Tris gel system (Invitrogen) and MOPS running buffer. Subsequently, proteins were transferred onto a PVDF membrane using NuPAGE transfer buffer (Invitrogen). Membranes were blocked with 3% BSA-PBST for 1 h and probed with *GATA1* goat polyclonal antibody (M-20, sc-1234, Santa Cruz Biotechnology) at a 1:500 dilution, *RPS19* mouse monoclonal antibody (WW-4, sc-100836, Santa Cruz Biotechnology) at a 1:500 dilution, *RPL5* goat polyclonal (D-20, sc-103865, Santa Cruz Biotechnology) at a 1:500 dilution, *RPL11* goat polyclonal (N-17, sc-25931, Santa Cruz

Biotechnology) at a 1:500 dilution, RPS20 goat polyclonal (G-15, sc-55035, Santa Cruz Biotechnology) at a 1:500 dilution, RPS24 rabbit polyclonal (ab102986, Abcam) at a 1:1,000 dilution, EPOR rabbit polyclonal (M-20, sc-697, Santa Cruz Biotechnology) at a 1:500 dilution, CD71 rabbit polyclonal (H-300, sc-9099, Santa Cruz Biotechnology) at a 1:500 dilution, JAK2 rabbit polyclonal (Clone 06-255, Millipore) at a 1:750 dilution, STAT5A rabbit polyclonal (C-17, sc-835, Santa Cruz Biotechnology) at a 1:500 dilution, TAL1 goat polyclonal (C-21; sc-12984, Santa Cruz Biotechnology) at a 1:500 dilution, p53 rabbit monoclonal (7F5; 2527, Cell Signaling Technology) at a 1:1,000 dilution, β -actin mouse monoclonal (AC-15, Sigma) at a 1:2,500 dilution or GAPDH mouse monoclonal antibody (6C5; sc-32233, Santa Cruz Biotechnology) at a 1:1,000 dilution in 3% BSA-PBST for 1 h at room temperature or overnight at 4 °C. Membranes were washed four times with PBST, incubated with donkey anti-mouse, anti-goat or anti-rabbit peroxidase-coupled secondary antibodies (715-035-150, 705-035-147 or 711-035-152, respectively; Jackson ImmunoResearch) at a 1:5,000 to 1:10,000 dilution in 3% BSA-PBST for 1 h at room temperature, washed three times with PBST and incubated for 1 min with Western Lightning Plus-ECL substrate (PerkinElmer). Proteins were visualized by exposure to scientific imaging film (Kodak).

Flow cytometry analysis and apoptosis. For flow cytometry analysis, *in vitro* cultured erythroid cells were washed in PBS and stained with propidium iodide (PI), 1:60 APC-conjugated CD235a (glycophorin A, clone HIR2, 17-9987-42, eBioscience), 1:60 FITC-conjugated CD71 (OKT9, 11-0719-42, eBioscience), 1:60 PE-conjugated CD41a (HIP8, 12-0419-42, eBioscience) and 1:60 PE-conjugated CD11b (ICRF44, 12-0118-42, eBioscience). For apoptosis analysis, the Annexin V-APC staining kit was used according to the manufacturer's instructions (550474, BD Pharmingen). FACS analysis was conducted on a BD Bioscience LSR II flow cytometer. Data were analyzed using FlowJo 8.6.9 (TreeStar).

Flow cytometry-activated cell sorting of human bone marrow populations. Mononuclear cells from human bone marrow aspirates were isolated by Ficoll-Paque PLUS (17-1440-02, GE Healthcare) density gradient centrifugation. Cells were preincubated with human Fc receptor binding inhibitor (14-9161, eBioscience) and stained with propidium iodide (PI) and antibodies against CD235a, CD41a and CD71 as described above. Cell sorting was conducted on a BD Bioscience Aria I.

Untranslated region reporter constructs and luciferase reporter assay. For generation of the luciferase reporter constructs, the 5' UTR of the luciferase gene in the pGL3-Promoter Vector (Promega) was replaced by the 5' UTRs of *GAPDH*, *ACTB* or *GATA1*. The 5' UTRs were inserted 14 bp downstream of the major transcription initiation site of the SV40 promoter. 293T cells were seeded in a 24-well plate at a density of 50,000 cells per well. For transfection of 293T cells, 350 ng of each of the luciferase reporter vectors were cotransfected with 35 ng per well of the pRL-SV40 vector (Promega) using the FuGene 6 reagent according to the manufacturer's protocol (Promega). Cells were incubated at 37 °C with 5% CO₂ until analysis at 48 h. For measuring luciferase reporter activity, the Dual-Glo Luciferase assay system (E2920, Promega) was used according to the manufacturer's protocol. Briefly, cells were resuspended in Dual-Glo Luciferase assay reagent and incubated at room temperature for 10–30 min, followed by measurement of firefly luminescence on a Safire 2 microplate reader (Tecan). Subsequently, Dual-Glo Stop and Glo reagent was added to the suspension and incubated for 10–30 min at room temperature, followed by measurement of *Renilla* luciferase activity. For each sample, the ratio of firefly/*Renilla* luminescence was calculated and normalized to the signal of the GAPDH-5' UTR construct.

Polysome profiling. K562 cells were incubated with 100 μ g/ml of cycloheximide for 10 min at 37 °C, washed twice with ice-cold PBS containing 100 μ g/ml of cycloheximide and lysed in 10 mM Tris-HCl (pH 7.4), 5 mM MgCl₂, 100 mM KCl, 2 mM DTT, 100 μ g/ml cycloheximide, 500 U/ml RNasin (Promega) and 1 \times Complete Protease Inhibitor, EDTA-free (Roche) by passing the lysate through a 26-gauge needle 4 times. Polysomes were separated on a 10–50% linear sucrose gradient containing 20 mM HEPES-KOH (pH 7.4), 5 mM MgCl₂, 100 mM KCl, 2 mM DTT and 100 μ g/ml cycloheximide

and centrifuged at 36,000 r.p.m. for 2 h in a SW41 rotor in an XE-90 ultracentrifuge (Beckman Coulter). Gradients were fractionated using a Bicom Gradient Station fractionator. Absorbance at 254 nm was used to visualize the gradients using an Econo UV monitor (Bio-Rad). Fractions were collected and phenol-chloroform extraction was performed to isolate RNA. β -actin primers were used for normalization of abundance of the mRNA of interest in monosome and polysome gradient fractions, similarly to what has been previously described¹⁹.

Protein labeling and detection. Click-iT chemistry was conducted for metabolic labeling of proteins. Briefly, K562 cells were washed with warm PBS and incubated in methionine-free RPMI medium (R7513, Sigma-Aldrich) with 10% FBS and 2 mM L-glutamine for 1 h at 37 °C, 5% CO₂ to deplete methionine reserves. For labeling, Click-iT AHA (L-azidohomoalanine, C10102, Life Technologies) was added at a final concentration of 50 μ M for 4 h at 37 °C, 5% CO₂. Cells were harvested, washed twice in PBS, resuspended in RIPA lysis buffer (50 mM Tris-HCl at pH 7.4, 150 mM NaCl, 0.1% SDS, 1% NP-40, 0.25% sodium deoxycholate) supplemented with 1 \times Complete Protease Inhibitor Cocktail (Roche) and incubated for 30 min on ice. After centrifugation at 14,000 r.p.m. for 10 min at 4 °C to remove cellular debris, the remaining supernatant was transferred to a new tube. 50 μ g of protein lysate was used for the Click reaction with tetramethylrhodamine alkyne (TAMRA, T10183, Life Technologies) using the Click-iT Protein Reaction Buffer Kit (C10276, Invitrogen) in a total volume of 200 μ l according to the manufacturer's instructions. Total cell proteins were separated by SDS-gel electrophoresis, and TAMRA signal was detected on a Typhoon 9200 imager (Amersham Biosciences) using 532 nm excitation and 580 nm long-pass emission. After TAMRA imaging, gels were post-stained with SYPRO Ruby Protein Gel Stain (S12001, Invitrogen) and imaged on a Typhoon 9200 imager using 473 nm excitation and 610 nm long-pass emission.

Immunoprecipitation coupled to Click-iT on beads labeling. Whole cell lysate was used for immunoprecipitation conducted with a goat polyclonal antibody against GATA1 (M-20, sc-1234, Santa Cruz Biotechnology) bound to Dynabeads Protein G (10003D, Life Technologies) for 3 h with rotation at 4 °C. The Dynabeads-antibody-antigen complex was washed three times with RIPA buffer and resuspended in 50 μ l RIPA buffer, and the Click-iT reaction was performed using TAMRA alkyne and the Click-iT Protein Reaction Buffer Kit (C10276, Invitrogen) for 1 h at 4 °C in a total reaction volume of 200 μ l. The immunoprecipitate was then washed once in RIPA buffer and then bound proteins were eluted in 40 μ l 2 \times LDS buffer incubated at 70 °C for 10 min. Proteins were separated by SDS-gel electrophoresis followed by TAMRA detection as described above or western blot analysis using antibodies against GATA1.

Sanger sequencing and analysis. Sanger sequencing of patient samples was performed using standard PCR-based methods and analyzed as has been described¹¹.

Microarray gene expression analysis. Bone marrow mononuclear cell samples were obtained from three patients with DBA with *RPS19* mutations (a frameshift mutation at codon 84, Arg62Gln, and Leu131Arg), as well as six control samples from healthy individuals. CD71^{high}CD34⁺CD45⁻ cell populations from mononuclear cells were sorted as described³³, and RNA was isolated from these samples. The RNA was hybridized to Affymetrix HG-133A microarrays, and the arrays were scanned for further analysis. Subsequently, the array data were normalized using the RMA normalization method as implemented in the Bioconductor package in R (<http://www.bioconductor.org/>). These data are deposited under accession number GSE41817 in the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>).

GATA1 target genes were derived from the GSE628 data set by comparing either the 21 h vs. 0 h time points (the majority of known GATA1 upregulated targets have been shown to be upregulated at the 21 h timepoint) or the 30 h vs. 0 h time points (ref. 42). Genes with a log₂ change of >1.0 and with a *P* value of <0.05 at the two sets of time points were used to derive gene sets¹⁷, which are lists of genes that were used for the enrichment analysis discussed below. In addition, GATA1 targets were derived from the TRANSFAC database (<http://www.gene-regulation.com/pub/databases.html>)⁴³. These gene sets were

used to run the gene set enrichment analysis (GSEA) algorithm, as described³⁴. As controls, we used gene sets that contain genes that share a transcription factor binding site defined in the TRANSFAC using the C3 transcription factor target set from the MSigDB database (<http://www.broadinstitute.org/gsea/msigdb/index.jsp>).

Statistical analyses. All pairwise comparisons were assessed using an unpaired two-tailed Student's *t*-test, unless otherwise indicated

in the main text. Results were considered significant if the *P* value was <0.05.

41. Sankaran, V.G. *et al.* Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322**, 1839–1842 (2008).
42. Welch, J.J. *et al.* Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood* **104**, 3136–3147 (2004).
43. Matys, V. *et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–D110 (2006).

Appendix C

Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation

David E. Weinberg^{1,8}, Premal Shah^{2,8,9}, Stephen W. Eichhorn^{3,4,5}, Jeffrey A. Hussmann^{6,7}, Joshua B. Plotkin², David P. Bartel^{3,4,5}

¹Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA

²Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

³Howard Hughes Medical Institute

⁴Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

⁵Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁶Institute for Computational Engineering and Sciences, University of Texas, Austin, TX 78712, USA

⁷Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712, USA

D.E.W., P.S., S.W.E., J.B.P., and D.P.B. designed the study. D.E.W., P.S., J.B.P., and D.P.B. wrote the manuscript, with help from the other authors. D.E.W. and S.W.E. prepared RNA-seq and ribosome-profiling libraries, respectively, under the supervision of D.P.B. D.E.W. performed initial data analysis. P.S. performed detailed data analyses and simulations with help from J.B.P. J.A.H. contributed new analytical tools.

Published as:

Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J., Plotkin, J., and Bartel, D.P. (2016). Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep* 14, 1787-1799.

Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation

David E. Weinberg,^{1,8,*} Premal Shah,^{2,8,9} Stephen W. Eichhorn,^{3,4,5} Jeffrey A. Hussmann,^{6,7} Joshua B. Plotkin,² and David P. Bartel^{3,4,5}

¹Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA

²Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

³Howard Hughes Medical Institute

⁴Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

⁵Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁶Institute for Computational Engineering and Sciences, University of Texas, Austin, TX 78712, USA

⁷Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712, USA

⁸Co-first author

⁹Present address: Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA

*Correspondence: david.weinberg@ucsf.edu

<http://dx.doi.org/10.1016/j.celrep.2016.01.043>

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

SUMMARY

Ribosome-footprint profiling provides genome-wide snapshots of translation, but technical challenges can confound its analysis. Here, we use improved methods to obtain ribosome-footprint profiles and mRNA abundances that more faithfully reflect gene expression in *Saccharomyces cerevisiae*. Our results support proposals that both the beginning of coding regions and codons matching rare tRNAs are more slowly translated. They also indicate that emergent polypeptides with as few as three basic residues within a ten-residue window tend to slow translation. With the improved mRNA measurements, the variation attributable to translational control in exponentially growing yeast was less than previously reported, and most of this variation could be predicted with a simple model that considered mRNA abundance, upstream open reading frames, cap-proximal structure and nucleotide composition, and lengths of the coding and 5' UTRs. Collectively, our results provide a framework for executing and interpreting ribosome-profiling studies and reveal key features of translational control in yeast.

INTRODUCTION

Although most cellular mRNAs use the same translation machinery, the dynamics of translation can vary between mRNAs and within mRNAs, often with functional consequences. For example, strong secondary structure within the 5' UTR of an mRNA can impede the scanning ribosome, thereby reducing the rate of protein synthesis (Kozak, 1986; Andersson and Kur-

land, 1990; Bulmer, 1991; Kudla et al., 2009; Tuller et al., 2010, 2011; Plotkin and Kudla, 2011; Ding et al., 2012; Bentele et al., 2013). The accessibility of the 5' cap (Godefroy-Colburn et al., 1985; Richter and Sonenberg, 2005) and the presence of small open reading frames (ORFs) within 5' UTRs referred to as upstream ORFs (uORFs) (Kozak, 1986; Ingolia et al., 2009; Brar et al., 2012; Zur and Tuller, 2013) can also modulate the rate of translation initiation (Sonenberg and Hinnebusch, 2009). Likewise, codon choice, mRNA structure, and the identity of the nascent polypeptide can influence elongation rates (Varenne et al., 1984; Brandman et al., 2012). In addition, differences in elongation rates can influence co-translational protein folding, localization of the mRNA or protein, and in extreme cases the rate of protein production (Kimchi-Sarfaty et al., 2007; Xu et al., 2013; Zhou et al., 2013). Finally, stop-codon readthrough can introduce alternative C-terminal regions that affect protein stability, localization, or activity (Dunn et al., 2013). Despite known examples of regulation at each of these stages of translation, translation is largely controlled at initiation, which is rate limiting for most mRNAs (Andersson and Kurland, 1990; Bulmer, 1991; Chu and von der Haar, 2012; Shah et al., 2013).

Variation in protein abundances observed in yeast cells largely reflects variation in mRNA abundances, indicating that much of gene regulation occurs at the level of mRNA synthesis and decay (Greenbaum et al., 2003; Csárdi et al., 2015). However, differences in translation rates also contribute. Studies using microarrays for global polysome profiling indicate that ribosome densities for different mRNAs vary over a 100-fold range (from 0.03 to 3.3 ribosomes per 100 nucleotides), indicating extensive translation control in *Saccharomyces cerevisiae* (Arava et al., 2003). More recently, the use of ribosome-footprint profiling has enabled transcriptome-wide analyses of translation using high-throughput sequencing, which again suggested a nearly 100-fold range of translational efficiencies (TEs) in log-phase yeast (Ingolia et al., 2009).

The ribosome-profiling method has itself undergone refinements over the last few years. Here, we build upon these advances and present improved ribosome-profiling and mRNA sequencing (mRNA-seq) datasets for log-phase yeast. Comparisons to many previous datasets reveal protocol-specific biases that can influence interpretation of ribosome-profiling experiments. With these insights, we then address several classical questions and ongoing debates in protein translation, such as the influence of tRNA abundances and nascent-peptide sequence on elongation rates. Our improved datasets also constrict the differences in TEs observed in log-phase yeast, such that the gene-to-gene variability that does remain can be largely predicted using a simple statistical model that considers only six features of the mRNAs.

RESULTS

Less Perturbed Ribosome Footprints

Protocols for analyzing polysome profiles or capturing ribosome footprints (referred to as ribosome-protected fragments, or RPFs) typically involve treating cells with the elongation inhibitor cycloheximide (CHX) to arrest the ribosomes prior to harvest (Ingolia et al., 2009; Gerashchenko et al., 2012; Zinshteyn and Gilbert, 2013; Artieri and Fraser, 2014; McManus et al., 2014). An advantage of CHX pre-treatment is that it prevents the run-off of ribosomes that can otherwise occur during harvesting (Ingolia et al., 2009). However, this treatment can also have some undesirable effects. Because CHX does not inhibit translation initiation or termination, pre-treatment of cultures leads to ribosome accumulation at start codons and depletion at stop codons (Ingolia et al., 2011; Guydosh and Green, 2014; Pelechano et al., 2015). In addition, because CHX binding to the 80S ribosome is both non-instantaneous and reversible, the kinetics of CHX binding and dissociation presumably allow newly initiated ribosomes to translocate beyond the start codon. Another possible effect of CHX treatment is that ribosomes might preferentially arrest at specific codons that do not necessarily correspond to codons that are more abundantly occupied by ribosomes in untreated cells. Although effects of CHX pre-treatment have minimal consequence for analyses performed at the gene level, i.e., comparisons of the same gene in different conditions, or comparisons between different genes after discarding reads in the 5' regions of ORFs, CHX pre-treatment may have severe consequences for analyses that require single-codon resolution.

The potential effects of CHX pre-treatment near the start codon have been discussed since the introduction of ribosome profiling, where an alternative protocol with flash-freezing and no CHX pre-treatment is also presented (Ingolia et al., 2009). Indeed, many recent ribosome-profiling experiments avoid CHX pre-treatment (Gardin et al., 2014; Gerashchenko and Gladyshev, 2014; Guydosh and Green, 2014; Jan et al., 2014; Lareau et al., 2014; Pop et al., 2014; Williams et al., 2014; Nedialkova and Leidel, 2015). However, consensus on the ideal protocol has not yet been reached, in part because the influence of alternative protocols on the interpretation of translation dynamics has not been systematically analyzed.

Here, we implemented a filtration and flash-freezing protocol to rapidly harvest yeast cultures. Importantly, this protocol minimized the time the cells experience starvation, which leads to rapid ribosome run-off (Ingolia et al., 2009; Gardin et al., 2014; Guydosh and Green, 2014). The protocol did include CHX in the lysis buffer to inhibit elongation that might occur during RNase digestion, although we doubt this precaution was necessary.

The original ribosome-profiling protocol also used cDNA circularization (Ingolia et al., 2009), while some subsequent protocols instead ligate to a second RNA adaptor prior to cDNA synthesis (Guo et al., 2010). Both approaches can introduce sequence-specific biases at the 5' ends of reads, which are not expected to influence results of analyses performed at the level of whole mRNAs but might influence results of codon-resolution analyses. Borrowing from methods developed for small-RNA sequencing (Jayaprakash et al., 2011; Sorefan et al., 2012), we minimized these biases by ligating a library of adaptor molecules that included all possible sequences at the eight nucleotides nearest to the ligation junction. Using this ligation protocol with a rapidly harvested, flash-frozen sample, we generated 74.3 million RPFs for log-phase yeast.

The 5' Ramp of Ribosomes

Using the 5' ends of RPFs, we inferred the codon at the A site of each footprint (Ingolia et al., 2009). Analysis of all mapped reads revealed the expected three-nucleotide periodicity along the ORFs, as well as ribosome accumulation at the start and stop codons (Figures 1A and 1B).

To examine the global landscape of 80S ribosomes, we averaged the position-specific RPF densities of individual genes into a composite metagene, in which each gene was first normalized for its overall density of RPFs (i.e., RPKM of RPFs) and then weighted equally in the average (Equation S10). A small excess of ribosome density was observed in the first ~200 codons compared to the remainder of the ORF (Figure 1C). The trend toward decreasing ribosome density with codon position was also evident on a gene-by-gene basis: 82% of genes exhibited declining raw RPF reads along their entire gene-length, based on linear-regression of RPF reads with codon position (binomial test, $p < 10^{-15}$), with the 5'-to-3' decrease in ribosome densities for a gene of average length (~500 codons) averaging ~43%.

Much larger 5' ramps are observed in other studies (Ingolia et al., 2009; Gerashchenko et al., 2012; Zinshteyn and Gilbert, 2013; Artieri and Fraser, 2014; Guydosh and Green, 2014; McManus et al., 2014), which is attributed to their use of CHX pre-treatment (Ingolia et al., 2009; Gerashchenko and Gladyshev, 2014) (Figure S1). However, CHX pre-treatment cannot explain the more modest ramp observed in our dataset, since our protocol did not involve such treatment.

The 5' ramp of ribosomes has previously been attributed to slower elongation due to preferential use of codons corresponding to low-abundance cognate tRNAs at the 5' ends of genes (Tuller et al., 2010). To determine the contribution of codon usage, we tested whether differences in RPF densities between the 5' and 3' ends of genes depended on codon choice. Surprisingly, for each of the 61 sense codons, the average density of RPFs was 33% greater on average when the codon fell within

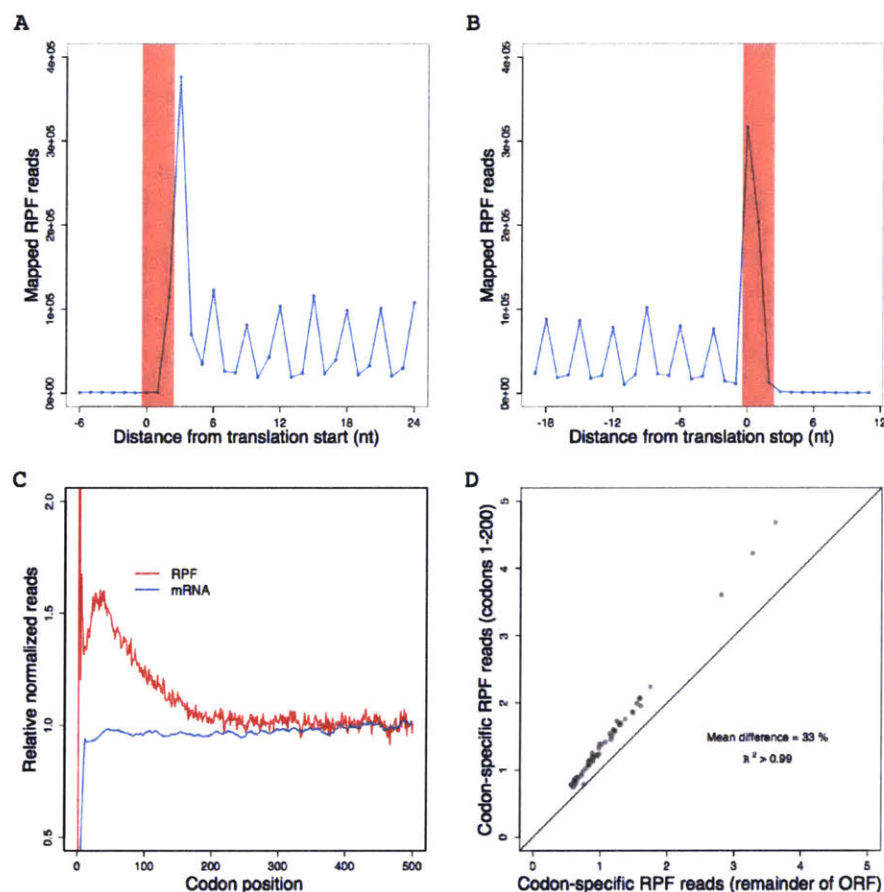


Figure 1. Less Perturbed RPFs Reveal a Codon-Independent 5' Ramp

(A and B) Metagenome analyses of RPFs. Coding sequences were aligned by their start (A) or stop (B) codons (red shading). Plotted are the numbers of 28–30-nt RPF reads with the inferred ribosomal A site mapping to the indicated position along the ORF.

(C) Metagenome analyses of RPFs and RNA-seq reads (mRNA). ORFs with at least 128 total mapped reads between ribosome-profiling (red) and RNA-seq (blue) samples were individually normalized by the mean reads within the ORF, and then averaged with equal weight for each codon position across all ORFs (e'_j in Equation S10 and h'_j in Equation S14).

(D) Comparison of codon-specific RPFs as a function of the 5' ramp. For each of the codons, densities of RPFs with ribosomal A sites mapping to that codon were calculated using either only the ramp region of each ORF (codons 1–200) or the remainder of each ORF (v^5_k in Equation S16 and v^3_k in Equation S17, respectively). The diagonal line indicates the result expected for no difference between the two regions.

See also Figure S1.

At least three considerations help explain why CHX pre-treatment would disrupt the correlation between tRNA abundances and measured ribosome densities at the A site. The first is that CHX, once bound to a ribosome, allows

the first 200 codons of an ORF (Figures 1D and S1), which showed that differential codon usage alone cannot explain the 5' ramp. Consistent with these experimental results, simulation of protein translation in a yeast cell, using a whole-cell stochastic model of yeast translation (Shah et al., 2013), indicated that codon ordering could account for at most a 20% ramp (Figure S1). Thus, codon ordering might explain some of the ~60% ramp observed in our dataset, but the majority of this ramp is likely caused by other mechanisms (see Discussion).

Codon-Specific Elongation Dwell Times Are Inversely Correlated with tRNA Abundances

The 61 sense codons varied in their average RPF densities by more than 6-fold (Figure 1D), indicating that different codons are decoded at different rates. Molecular biologists have long assumed that such differences in elongation rates are caused by corresponding differences in the cellular abundances of cognate tRNAs (Andersson and Kurland, 1990; Bulmer, 1991). Several early experiments provide empirical support for this view (Varenne et al., 1984; Sørensen and Pedersen, 1991), but early analyses of ribosome-profiling results do not find any relationship between ribosome density and cognate tRNA abundance expected from this model (Ingolia et al., 2011; Li et al., 2012; Qian et al., 2012; Charneski and Hurst, 2013; Zinsh-teyn and Gilbert, 2013). However, the datasets analyzed in these studies were all from experiments that used CHX pre-treatment.

for an additional round of elongation before halting ribosomes (Schneider-Poetsch et al., 2010; Gardin et al., 2014; Lareau et al., 2014), which alone would remove any correlation at the A site. Second, CHX binding is reversible, and at concentrations typically used in ribosome-profiling protocols, additional rounds of elongation might occur between CHX-binding events. Third, CHX prevents translocation of the ribosome by binding to the E site, with space for a deacylated tRNA (Schneider-Poetsch et al., 2010), and thus CHX binding affinity presumably varies with features of the E site and the tRNA in it. Thus, in the presence of CHX pre-treatment, the ribosome density at a site is likely more a function of the on and off rates of CHX binding than a function of differential isoaccepting tRNA availability. Indeed, recent analyses of profiling results obtained without CHX pre-treatment have observed modest correlations between tRNA abundances and ribosome-densities at the A site (Gardin et al., 2014; Lareau et al., 2014).

When examining earlier ribosome-profiling datasets, we found that whenever CHX pre-treatment was employed, the relationship between ribosome occupancy and tRNA abundance was both insignificant ($p > 0.05$) and in the opposite direction than expected (Figures S2C–S2E). Moreover, the concordance between these CHX pre-treatment datasets indicated a systematic bias (Figure S2), suggesting that an orthogonal set of mRNA sequence biases influence CHX binding. In contrast, for every dataset without CHX pre-treatment, we found that ribosome

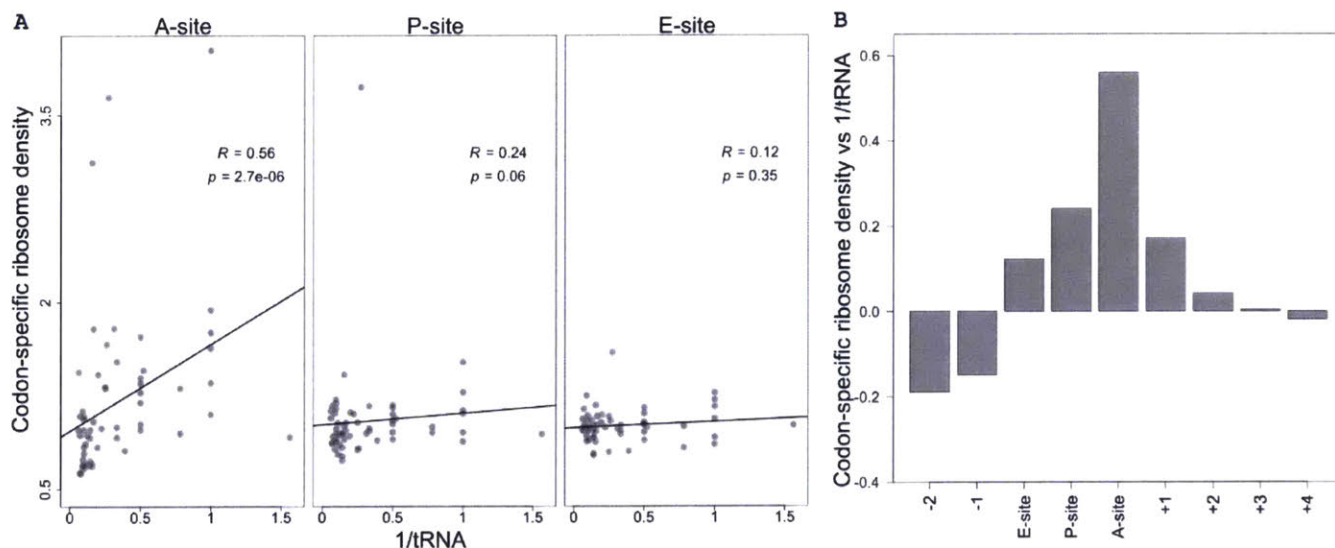


Figure 2. Codons Corresponding to Lower-Abundance tRNAs Are Decoded More Slowly

(A) Correlation between codon-specific excess ribosome densities and cognate tRNA abundances. Codons within RPFs were assigned to the A-, P-, and E-site positions based on the distance from the 5' ends of fragments, and codon-specific excess ribosome densities were calculated (v_k in Equation S19). Cognate tRNA abundances for each codon were estimated using the genomic copy numbers of iso-accepting tRNAs and wobble parameters (Table S2). Spearman R values are shown, with their significance (p values).

(B) The correlations of codon-tRNA abundance at different positions relative to the A site. Analysis was as in (A) using varying offsets from the A-site position within RPFs (x axis) to calculate Spearman correlations (y axis).

See also Figures S2 and S3 and Tables S1 and S2.

densities were inversely correlated with tRNA abundances (Figures S2C–S2E).

In our dataset, we found that codon-specific excess ribosome densities (v_k in Equation S19) were strongly anti-correlated with cognate tRNA abundances, as estimated by copy numbers of tRNA genes and wobble parameters (Figures 2A and 2B). This strong anti-correlation was also observed with direct estimates of tRNA abundances obtained from our RNA-seq measurements (Figure S2A; Table S1). As expected, the correlation was specific to the codon within the A site, with residual correlations at the P and E sites, which were potentially caused by some 5' heterogeneity of RPFs.

Taken together, these results strongly support the idea that differential cognate tRNA abundances influence differential elongation times of codons in the absence of CHX. Without CHX pretreatment, we also observed widespread pausing after polybasic tracts (Figure S3) but not at P-site proline codons (Figure S2), which has been the subject of some debate (Supplemental Information).

Slower Elongation at Regions Encoding Inter-domain Linkers

The modulation of elongation rates by either tRNA abundances (Figure 2A) or polybasic stretches (Figure S3) might influence the kinetics of co-translational folding. Indeed, slower elongation rates within inter-domain linkers relative to the adjacent domains is reported to coordinate co-translational folding of nascent polypeptides (Thanaraj and Argos, 1996; Kimchi-Sarfaty et al., 2007; Pechmann and Frydman, 2013). However, systematic

experimental evidence for such differences in elongation rates has been lacking.

To examine whether our ribosome-profiling data reveal such differences, we first used InterProScan classifications (Jones et al., 2014) based on the Superfamily database (Wilson et al., 2009) to partition coding sequences into domain and linker regions. We then calculated the mean normalized RPF densities (z_j in Equation S7) for codons within the domain- and linker-encoding regions and found significantly lower densities in regions of genes that fell within domains compared those that fell outside of domains (Figure 3; mean difference 0.094, paired t test, $p < 10^{-26}$). To eliminate any influence of the 5' ramp, we repeated the analysis excluding the first 200 codons. Although the size of the effect diminished (mean diff = 0.029), the difference in mean ribosome densities remained significant ($p = 0.0002$), indicating that the 5' ramp was not solely responsible for lower ribosome densities within domains (Figure S4A).

The trend toward relatively lower ribosome densities in domain regions held even when restricted to each individual amino acid, with the exceptions of cysteine residues and the single-codon-encoded methionine and tryptophan residues (Figure S4). Thus, differences in amino acid content between domains and linkers could not account for the observed differences in bound ribosome densities. Moreover, for 54 out of 61 sense codons, we found significantly lower ribosome densities in domains compared to linkers (one-sided t test, $p < 0.05$). For 26 out of 61 codons, we found significantly lower ribosome densities in domains even after excluding the first 200 codons (one-sided t test, $p < 0.05$). This result implied that differences in synonymous codon usage between domain and linker regions cannot alone

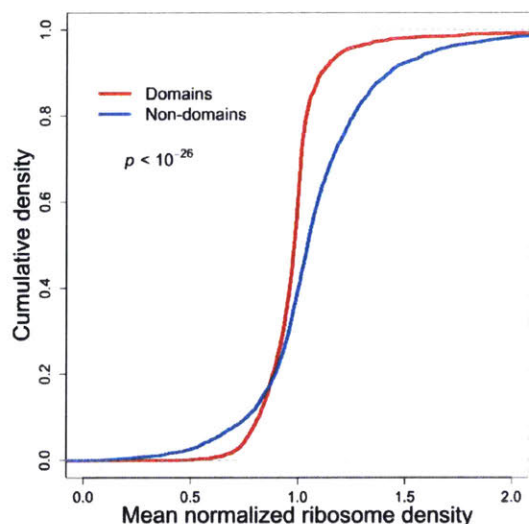


Figure 3. Elongation Dynamics Correlate Domain Architecture

Cumulative distributions of normalized ribosome densities within and outside of protein-folding domains. Mean normalized RPF densities (\bar{z}_p in Equation S7) for codons within the domain-encoding and non-domain-encoding regions were individually calculated for each ORF. Domain assignments were based on InterProScan classifications (Jones et al., 2014) obtained from the Superfamily database (Wilson et al., 2009). Statistical significance was evaluated using paired t test ($p < 10^{-26}$).

See also Figure S4.

account for the differences in ribosome densities. One possible mechanism for differential ribosome occupancy, independent of codon usage, is differential recruitment of chaperones and their associated effects on co-translational folding (Ingolia, 2014).

Similar results for densities in domain and linker regions were obtained when using InterProScan classifications (Bateman et al., 2002) instead of the Superfamily database (Figure S4B). Finally, consistent with other computational analyses (Pechmann and Frydman, 2013), differences in elongation rate were found at the level of protein secondary structures as well: regions corresponding to helices and sheets exhibited significantly lower RPF densities than regions corresponding to loops (Figure S4C). Taken together, these results provided systematic empirical support for the claim that co-translational folding requirements influence elongation rates. Nonetheless, the magnitude of this signal was very small, suggesting that slower inter-domain elongation either has very little impact or impacts very few genes.

Estimates of Protein-Synthesis Rates

Our results thus far indicated that the ribosome density at a given codon position is influenced by the abundance of cognate tRNAs and whether the codon is immediately downstream of a polybasic stretch, falls within a protein domain, or lies in the 5' region of the ORF. The non-uniform ribosome densities along individual ORFs imply that the overall RPF density on each gene (i.e., RPKM of RPFs) does not directly reflect the rate of protein synthesis (Li et al., 2014). For example, the RPF densities of genes enriched in more slowly elongated codons would tend to overestimate their protein-synthesis rates, and the same would be true for shorter ORFs. To more accurately quantify the protein-syn-

thesis rates of individual genes from RPF densities, we used empirically derived correction factors to account for the position- and codon-specific effects we observed (f_j in Equation S23). With these correction factors, the ~74.3 million sequenced RPFs enabled reliable estimates of protein-synthesis rates for 4,839 genes (Equation S28).

Accurate Measurement of Yeast mRNA Abundances

In addition to improving measurements of ribosome densities, we sought to improve measurements of mRNA abundances, which is also critical for accurately quantifying translational control. Prior experiments have typically measured yeast mRNA abundances by performing RNA-seq on poly(A)-selected RNA (Ingolia et al., 2009; Gerashchenko et al., 2012; Zinshteyn and Gilbert, 2013; Artieri and Fraser, 2014; Guydosh and Green, 2014; McManus et al., 2014). However, poly(A) selection might bias mRNA-abundance measurements. For example, mRNAs that lack a poly(A) tail of sufficient length to stably hybridize to oligo(dT) might not be as efficiently recovered. Although *S. cerevisiae* is not known to contain translated mRNAs that altogether lack a poly(A) tail, the lengths of poly(A) tails found on *S. cerevisiae* mRNAs are relatively short, with a median length of 27 nt (Subtelny et al., 2014). Another source of potential bias in poly(A) selection is partial recovery of mRNAs endonucleolytically cleaved during RNA isolation or poly(A) selection. The 5' fragments resulting from mRNA cleavage are not recovered by poly(A) selection, which causes a 3' bias in the resulting RNA-seq data (Nagalakshmi et al., 2008). Indeed, analyses of published RNA-seq datasets from ribosome-profiling studies revealed a severe 3' bias in poly(A)-selected RNA-seq reads, ranging from 19%–130% excess reads (Equation S15) (Figure S5). Because longer mRNAs have a higher probability of being cleaved, the abundances of longer mRNAs might be systematically underestimated by poly(A) selection (Table S3).

An alternative to poly(A) selection is rRNA depletion, which enriches mRNAs by removing rRNA using subtractive hybridization. A concern with subtractive hybridization is the potential depletion of mRNAs that either cross-hybridize to the oligonucleotides used to remove rRNA sequences or adhere to the solid matrix to which the oligonucleotides are attached. To investigate the extent to which unintended mRNA depletion occurs when using reagents sold for yeast RNA-seq library preparations, we subjected the same total RNA to each of three procedures: Dynabeads oligo(dT)₂₅ (Life Technologies), RiboMinus Yeast Transcriptome Isolation Kit (Life Technologies), or Ribo-Zero Yeast Magnetic Gold Kit (Epicenter). As a reference, we also generated an RNA-seq library from the total RNA that was not enriched for mRNA and thus contained primarily rRNA (90.2% of 199.7 million genome-mapping reads). We also note that we started with RNA extracted from the lysate that was used for ribosome-footprint profiling, as opposed to RNA extracted from whole cells as done in the original ribosome-profiling study (Ingolia et al., 2009). When comparing the 4,540 mRNAs for which we obtained at least 64 reads in our total RNA library, only the Ribo-Zero-treated sample faithfully recapitulated the mRNA abundances observed in total RNA ($R^2 = 0.98$; Figures 4A and S5). The poly(A)-selected and RiboMinus-treated

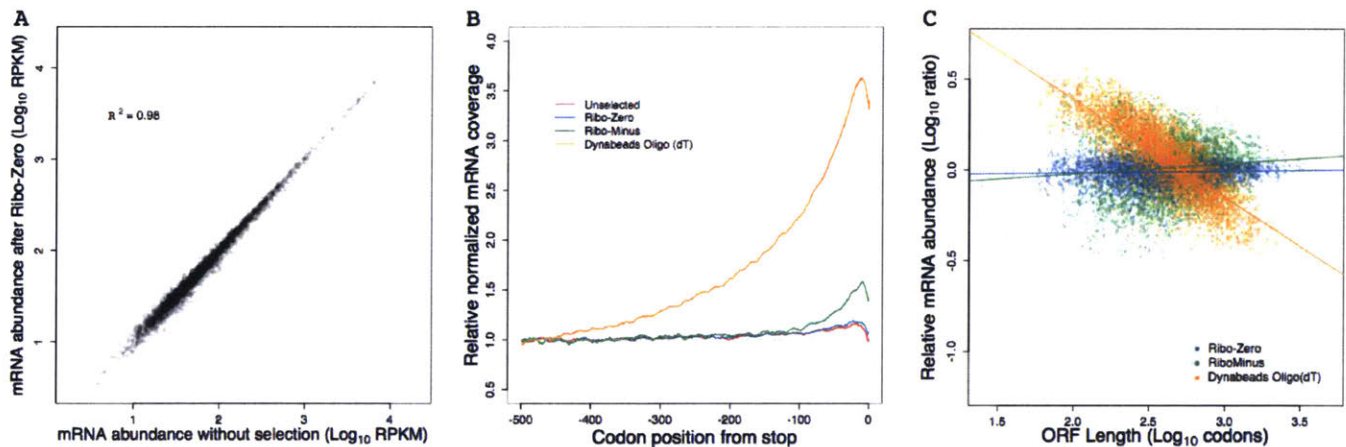


Figure 4. mRNA Enrichment Methods Can Bias mRNA Abundance Measurements

(A) mRNA abundances measured by RNA-seq of Ribo-Zero-treated RNA compared to those measured by RNA-seq of total unselected RNA. Pearson R^2 is indicated.

(B) Metagene analysis of RNA-seq read density in total unselected or mRNA-enriched RNA samples. Coding sequences were aligned by their stop codons, and RNA-seq reads were individually normalized by the mean reads within the ORF and then averaged with equal weight for each codon position across all ORFs (\bar{h}'_j in Equation S15).

(C) mRNA abundances for mRNA-enriched samples relative to total unselected RNA, as a function of ORF length.

See also Figures S5, S6, and S7 and Tables S3 and S4.

samples each had significantly lower correlations with total RNA ($R^2 = 0.85$ and $R^2 = 0.87$, respectively), indicating a skewed representation of the transcriptome. Compared to RNA-seq data from published ribosome-profiling studies, our Ribo-Zero-treated sample also exhibited the highest correlations with microarray-based estimates of mRNA abundances (Table S3).

As anticipated, the poly(A)-selected sample contained a strong 3' bias (Figure 4B), which caused a systematic underestimation of the abundances of longer genes (Figure 4C). After accounting for this strong bias in the poly(A)-selected sample, we did not detect a relationship between poly(A)-tail length and poly(A)-selection efficiency, suggesting that tail-length differences did not significantly contribute to the biases of poly(A)-selected RNA-seq data. For the RiboMinus-treated sample, cross-hybridization to the depletion probes might have skewed the mRNA abundances, which might have been largely avoided in the Ribo-Zero protocol because of its more stringent hybridization conditions. The RiboMinus-treated sample also had substantial rRNA contamination (44.5% of reads, originating primarily from the 5S rRNA).

Interestingly, the total-RNA and the Ribo-Zero datasets both contained a small 3' bias (Figure 4B), with median 3'/5' excess reads of 22% and 28%, respectively (Table S4). This bias was consistent with reports that yeast mRNAs are primarily degraded in the 5'-to-3' direction (Hu et al., 2009; Pelechano et al., 2015). The decay intermediates of this vectorial degradation process would contribute more reads toward the 3' ends of mRNAs, giving rise to the observed bias, especially when considering that our RNA samples were enriched for cytoplasmic RNA, which would diminish the countervailing vectorial mRNA synthesis process occurring in the nucleus. Nonetheless, the 3' biases in the total-RNA and Ribo-Zero datasets were smaller than those in poly(A)-selected samples, for which median 3'/5' excess

mRNA reads ranged from 42% to 275% (Table S4). Because Ribo-Zero treatment enabled deep coverage of the yeast transcriptome without substantially biasing mRNA abundances, we used mRNA abundances estimated from Ribo-Zero-treated RNA for all subsequent analyses.

A Narrow Range of Initiation Efficiencies in Log-Phase Yeast

Because protein synthesis is typically limited by the rate of translation initiation (Andersson and Kurland, 1990; Bulmer, 1991; Shah et al., 2013), we defined the initiation efficiency (IE) of each gene as its protein-synthesis rate divided by its mRNA abundance (Equation S27). Thus, the IE measure quantified the efficiency of protein production per mRNA molecule of a gene, in a typical cell. To facilitate comparisons with published datasets, we also calculated the translational efficiency (TE) of each gene, defined as its RPF density normalized by its mRNA abundance (Ingolia et al., 2009). Because TE is calculated based on the RPF density rather than the protein-synthesis rate, TE does not account for differential rates of elongation associated with the 5' ramp or codon identity. Nonetheless, IE and TE were highly correlated ($R = 0.951$; Figure S6A).

A wide range of IEs (or TEs) among genes would indicate that protein production is under strong translational control, whereas a narrow range would indicate that protein production is typically governed by mRNA abundances, and hence protein-synthesis rate is primarily controlled by mRNA transcription and decay. The first ribosome-profiling study suggested a large amount of translational control in yeast, with the range of TEs reported to span roughly 100-fold (Ingolia et al., 2009). Indeed, we found that the 1–99 percentile range of TEs in those data spanned 73-fold (Figure S6C). In contrast, the range of TEs observed in our data was narrower, with the 1–99 percentile spanning only

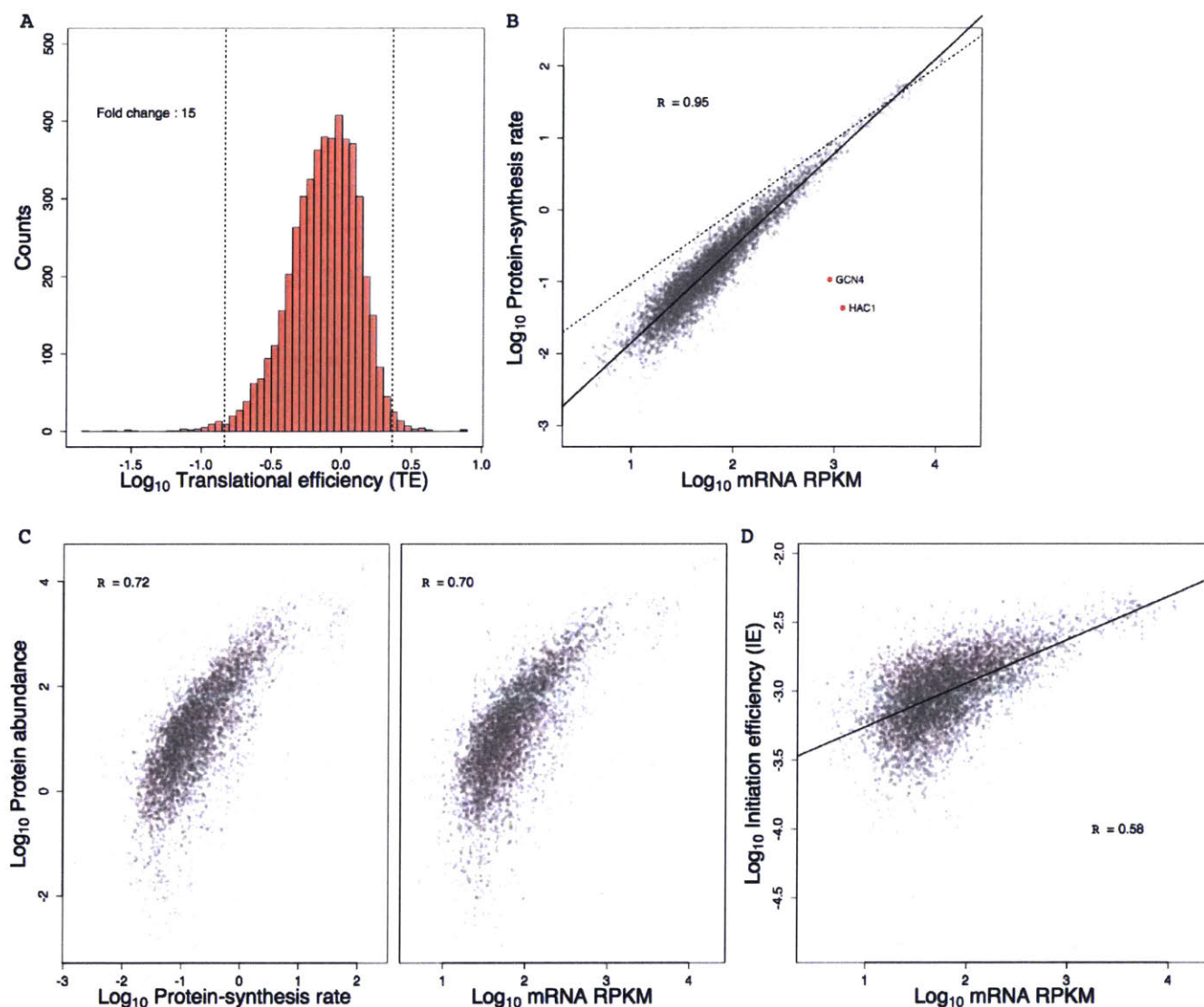


Figure 5. TEs and IEs Span a Narrow Range in Log-Phase Yeast Cells

(A) Distribution of TE measurements, with vertical dashed lines marking the first and 99th percentiles, and the fold change separating these percentiles indicated. All ORFs with at least 128 total reads between the ribosome-profiling and RNA-seq datasets were included (except *YCR024C-B*, which was excluded because it is likely the 3' UTR of *PMP1* rather than an independently transcribed gene).

(B) Relationship between estimated protein-synthesis rate and mRNA abundance for genes shown in (A). *GCN4* and *HAC1* (red points) were the only abundant mRNAs with exceptionally low protein-synthesis rates. The best linear least-squares fit to the data is shown (solid line), with the Pearson *R*. For reference, a one-to-one relationship between protein-synthesis rate and mRNA abundance is also shown (dashed line).

(C) Relationship between experimentally measured protein abundance (de Godoy et al., 2008) and either protein-synthesis rate (left) or mRNA abundance (right). The 3,845 genes from (A) for which protein-abundance measurements were available were included in these analyses. Pearson correlations are shown (*R*).

(D) Relationship between mRNA abundance and IE for genes shown in (A). The best linear least-squares fit to the data is shown, with the Pearson *R*.

See also Figures S8 and S9 and Table S5.

a 15-fold range (Figure 5A). Although the range of IEs was marginally wider than that of TEs (1–99 percentile spanning 21-fold; Figure S6B), it was still substantially smaller than the range of TEs initially reported (Ingolia et al., 2009). The relatively narrow range of IEs in our data was also reflected by the high correlation between mRNA abundance and protein-synthesis rate ($R = 0.948$; Figure 5B), supporting the conclusion that protein-synthesis rates are largely dictated by mRNA abundances (Csárdi et al., 2015). Interestingly, the slope of the regression between mRNA

and protein-synthesis rates was >1 on the log-scale, indicating that translation regulation mostly amplifies the effect of differential mRNA abundances rather than buffering it (Csárdi et al., 2015). Further indicating that mRNA abundance (when accurately measured) is a strong predictor of total protein production, mass-spectrometry-based measurements of steady-state protein abundance (de Godoy et al., 2008) correlated as well with mRNA abundances as they did with protein-synthesis rates (Figure 5C).

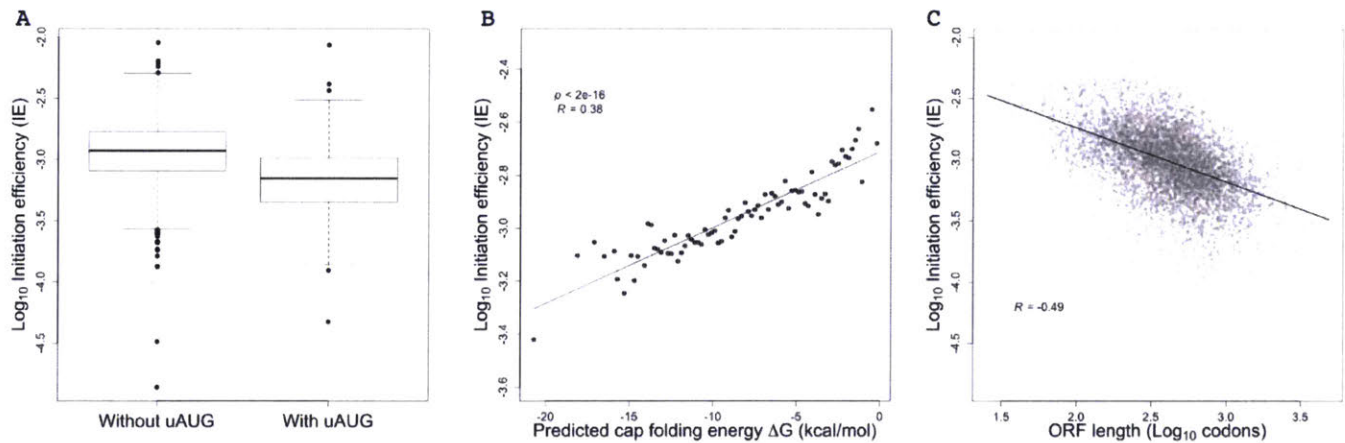


Figure 6. mRNA Sequence, Structure, and Length Correlate with IE

(A) Reduced IE values for genes with at least one upstream AUG (i.e., an AUG codon located within the annotated 5' UTR). The plots indicated the median (line), quartile (box) and first and 99th percentiles (whiskers) of the distributions.

(B) Inverse relationship between IE and the folding energy of predicted RNA secondary structure near the cap (Cap-folding energy). RNAfold was used to estimate folding energies for the first 70 nt of the mRNA. Gray bars indicate 1 SD of IE values for genes binned by predicted folding energy. The best linear least-squares fit to the data is shown (solid line), with the Pearson R .

(C) Inverse relationship between IE and ORF length. The best linear least-squares fit to the data is shown (solid line), with the Pearson R .

See also Figure S7.

When we examined the range of TEs in other published datasets, we also found more narrow ranges (as low as 11-fold from 1–99 percentiles) than that of Ingolia et al. (2009) (Figure S6C). However, the TEs in published datasets—which are all generated using poly(A)-selected mRNA—were not particularly well correlated with each other (Table S5). These discrepancies in TEs were largely due to differences in measured mRNA abundances, whereas the RPF abundances correlated almost perfectly (Table S5). Collectively, these results indicate that the amount of translational control in log-phase yeast has been overestimated due to inaccuracies in TE measurements, largely caused by challenges in accurately measuring mRNA levels.

We also noticed that the shape of the TE distribution from our data, which was asymmetric, differed from that of the Ingolia data, which is highly symmetric. In particular, in our data there were relatively few genes in the right tail of the distribution (Figure 5A; note the location of the mode closer to the 99th than the first percentile). This observation implied that mRNAs from very few genes contain elements that impart an exceptionally high initiation efficiency and are thereby “translationally privileged.” Rather, most mRNAs either initiate close to a maximum possible rate (likely set by the availability of free ribosomes or initiation factors) or contain features that modestly reduce the initiation rate.

To the extent that differences in IE were observed, the genes with lower IE tended to be expressed at lower mRNA levels, with IE increasing roughly linearly with mRNA expression levels (Figure 5D). These results were consistent with the notion that abundant mRNAs have undergone evolutionary selection to be efficiently translated (Sharp and Li, 1987; Andersson and Kurland, 1990; Plotkin and Kudla, 2011; Shah and Gilchrist, 2011). Interestingly, in the plots comparing protein-synthesis rate or IE with mRNA level, the points for 11 of the 12 highest expressed mRNAs fell below the regression lines (Figures 5B and 5D,

dashed lines), suggesting that the efficiency for the highest expressed mRNAs might have saturated.

Two notable outliers appeared in the comparison of mRNA abundances and synthesis rates (Figure 5B, red dots). These two, which corresponded to relatively abundant mRNAs with exceptionally low synthesis rates, were *HAC1* and *GCN4*. These are the two most well-known examples of translational control in log-phase yeast and are both involved in rapid stress responses, during which translational repression is relieved (Raggeegger et al., 2001; Mueller and Hinnebusch, 1986; Dever et al., 1992). The observation that *HAC1* and *GCN4* were the only abundant mRNAs that were strongly regulated at the translational level further emphasized that translational control only modestly influences the protein production of most yeast genes. Nevertheless, the tuning of synthesis rates via translational control can help maintain the proportional synthesis of the subunits of multiprotein complexes (Figures S6D–S6G; Supplemental Experimental Procedures).

Determinants of Initiation Efficiencies in Yeast

Next, we sought to identify sequence-based features that explain the variation in IE values that remained among genes after improving the RPF and mRNA measurements. First, we considered uORFS, which can inhibit translation by serving as decoys to prevent initiation at the start codons of bona fide ORFs (Zur and Tuller, 2013), as occurs for *GCN4* (Mueller and Hinnebusch, 1986; Dever et al., 1992), one of two genes with the greatest translational repression (Figure 5B). Using high-resolution 5' UTR annotations (Arribere and Gilbert, 2013), we identified upstream AUGs (uAUGs) in 303 out of the 2,549 genes that had reproducibly uniform transcription-start sites. Those genes containing uAUGs had significantly lower IEs than genes without uAUGs, even after controlling for 5' UTR lengths (Figure 6A; t test $p < 10^{-16}$). These results confirmed that a general feature of

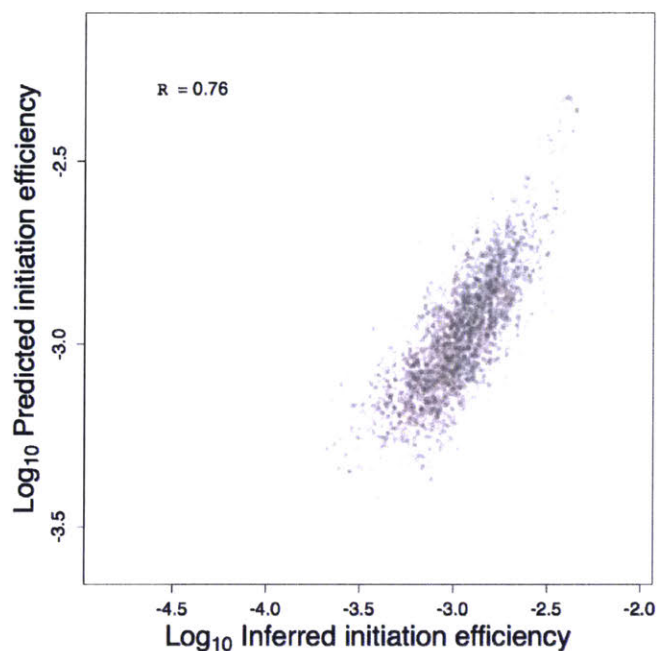


Figure 7. Sequence-Based Features of mRNAs Largely Explain Yeast IEs

Correspondence between predicted IEs and IEs inferred directly from the RPF and RNA-seq data. Initiation efficiencies were predicted using a multiple-regression model, based on mRNA abundance and sequence-based features of the 2,549 genes with empirically determined 5' UTRs. Shown is the Pearson R . See also Table S6.

uORFs is to decrease the translation of downstream ORFs, and that the presence of uAUGs can explain some of the variance in IEs (Arribere and Gilbert, 2013; Zur and Tuller, 2013).

Another feature that has been linked to differences in synthesis rates is mRNA secondary structure. Structure located near the 5' cap might interfere with binding of the eIF4F cap-binding complex, while structure within the 5' UTR could disrupt the scanning 40S ribosome. An open structure around the start codon might also be important for facilitating joining of the 60S subunit. Previous genome-wide structure analyses revealed a weak but significant inverse correlation between start-codon-proximal structure and TE (Kertesz et al., 2010), but the accessibility of the 5' UTR more generally was not reported, and the TE values used in those studies were affected by RNA-seq biases. For each mRNA with a single reproducible 5' end (Arribere and Gilbert, 2013), we predicted the accessibility of the 5' cap by calculating the predicted folding energy of the sequence spanning increasing distances from the cap. For all distances examined, we observed a significant correlation between predicted cap accessibility and IE (t test, $p < 10^{-6}$ for each window; Figures 6B and S7). This correlation rapidly increased with window length, approaching a maximum at 70–90 nt (Pearson correlation, $R \sim 0.37$ for windows 70–90 nt) and then steadily declined for larger windows (Figure S7), consistent with local folding of the 5' end determining cap accessibility. Together, these results confirmed that mRNAs with less-structured 5' UTRs tend to be initiated more efficiently (Godefroy-Colburn et al., 1985; Shah et al., 2013), which is consistent with eIF4F binding, 40S recruit-

ment, or scanning as influential regulatory steps during eukaryotic initiation. Notably, the correlations that we observed between predicted mRNA structure and translation were the largest that have been reported between these features in eukaryotes, which emphasized the utility of our accurate IE measurements and of predicting structure near the cap as opposed to more downstream regions.

Gene length has also been reported to correlate with translational efficiency. Although global polysome-profiling studies indicate strong anti-correlation between ORF length and ribosome density (Arava et al., 2003), analysis of published ribosome-footprint-profiling data revealed essentially no correlation (or even a positive correlation in some cases) between length and TE (Figure S7). In contrast, we observed a striking negative correlation in our IE (and correspondingly in our TE) data (Figures 6C and S7). Our IE measure already corrected for the elevated ribosome densities in the first 200 codons, and the negative correlation between ORF length and TE persisted even after removing the first 250 codons of each ORF, which further confirmed that this correlation was not caused by the 5' ramp (Figure S7). The discrepancy between our data and earlier ribosome-profiling datasets was likely due to the RNA-seq 3'-bias caused by poly(A) selection (Figures 4B and S5). Indeed, an anti-correlation between ORF length and TE was observed in most other datasets when we controlled for the 3' bias by estimating mRNA abundances based on mapped RNA-seq reads from only the 3' ends of genes (Figure S7). Together, these results showed that the original report of shorter mRNAs having relatively higher initiation efficiencies (Arava et al., 2003) is correct, even after accounting for the CHX-enhanced 5' ramp that confounded that analysis.

A Statistical Model that Predicts Initiation Efficiencies

Based on these results, we used multiple linear regression to build a model that considered number of uAUGs, predicted cap-proximal RNA-folding energy (and also GC content of the 5' UTR as another metric for structure), and lengths of the ORF and the 5' UTR to explain the variance in IE observed among genes. We also included an mRNA-abundance term in the model because IE is greater for more abundant mRNAs (Figure 5D). To identify the most informative features, we used Akaike's Information Criteria (AIC) for model selection and both step-up and step-down model-selection procedures (using the stepAIC function in the MASS package in R). The multiple regression model that best explained the variation in IE included all six variables, even after penalizing for model complexity (Figure 7; Table S6). The dominant explanatory variable was mRNA abundance, which alone accounted for $\sim 40\%$ of the variance in IE. Collectively, a model containing all six variables explained $\sim 58\%$ of the variance in IE. A model that excluded mRNA abundance, and therefore depended on only sequence-based features, still explained $\sim 39\%$ of the variance in IE. These results of our statistical modeling should help motivate mechanistic studies of how each of these mRNA features impacts translation.

DISCUSSION

We have shown that improved measurements of both mRNA abundances and RPFs can provide insights into the regulation

and dynamics of eukaryotic translation. The RPFs that we isolated and sequenced are indicative of a dynamic and heterogeneous elongation process, with ribosomes transiting along mRNA molecules at variable rates depending on the distance from the start codon, codon identity, and nascent polypeptide sequence.

What might explain the 5' ramp of ribosomes observed even in the absence of CHX pre-treatment (Figure 1C)? Codon usage accounted for about a third of it, but even the same codons were differentially occupied by ribosomes depending upon whether they occurred in the 5' or 3' ends of genes (Figure 1D), indicating that additional mechanisms must be involved. Although we cannot rule out ribosome drop-off as a contributing factor, we favor the idea that elongation is slower during the early phase of translation. Perhaps an initiation factor remains engaged with the 80S ribosome during early elongation, and the bound factor maintains the ribosome in a slower state until it stochastically dissociates from the ribosome within the first 200 codons. The eIF3 complex is a promising candidate for such a factor, as it binds the solvent-exposed face of the 40S ribosome (Siridechadilok et al., 2005) and can therefore bind to 80S ribosomes as well (Beznosková et al., 2013). Maintaining eIF3 on early elongating ribosomes might also facilitate re-initiation after translation of short uORFs (Szamecz et al., 2008; Zur and Tuller, 2013).

A practical finding of our studies is that the choice of mRNA enrichment method can have a significant impact on yeast mRNA-abundance measurements. rRNA depletion using the Ribo-Zero kit was the only method that enriched for mRNAs without introducing substantial and systematic biases (Figures 4A and S5). One caveat of rRNA depletion is that nascent pre-mRNAs that lack a poly(A) tail may also be recovered, which can inflate mRNA abundance measurements with respect to the pool of translatable mRNA molecules. This effect may be more pronounced in metazoans that contain long introns and correspondingly long transcription times. The extent to which poly(A)-selection biases affect metazoan mRNA abundance data and thereby influence TE measurements remains to be determined.

The initial report that TE spans a roughly 100-fold range across mRNAs in budding yeast spurred intensive investigation of the underlying TE determinants, with varying degree of success (Kertesz et al., 2010; Tuller et al., 2011; Charneski and Hurst, 2013; Zur and Tuller, 2013; Bentele et al., 2013; Rouskin et al., 2014). Our results showed that this apparently wide range of TEs is partly explained by inaccurate mRNA-abundance measurements. After identifying and minimizing this source of inaccuracy, we observed a narrower range of TEs and IEs (Figure 5A; Table S3), suggesting a more limited degree of translational control. The TE range that we observed in yeast resembled the range observed in mouse embryonic stem cells (Ingolia et al., 2011), suggesting that limited translational control is a general principle of gene regulation in rapidly dividing eukaryotic cells.

Using our IE measurements, we were able to generate a statistical model that explained a majority of the IE variance (Figure 7; Table S6). Based on this model, secondary structure within the 5' UTR, most especially cap-proximal structure, appears to be an important determinant of IE. These results are in agreement

with early mechanistic studies demonstrating that cap accessibility correlates with initiation efficiency (Godefroy-Colburn et al., 1985) and that stable 5' UTR secondary structures block the scanning ribosome (Kozak, 1986). One caveat of our structure analyses is that we used in silico prediction of mRNA structure, which does not always accurately capture the in vivo structure of mRNA (Rouskin et al., 2014). Further indicating the inadequacy of in silico predictions was the benefit of also including 5' UTR GC content as a feature in our model. Likewise, the inclusion of mRNA abundance might have helped compensate for the inadequacy of in silico structure predictions, as highly expressed genes have less predicted structure in 5' UTRs than do lowly expressed genes (Gu et al., 2010), and presumably these differences would be even greater when looking at actual 5' UTR structure. Therefore, mRNA structure presumably explains even more variation in IE than our analyses suggested.

We also found that longer ORFs tended to be more poorly translated in log-phase yeast, even after accounting for the 5' ramp (Figure 6C). Given that initiation occurs at the 5' ends of mRNAs, how might initiation rates be sensitive to ORF lengths? One possibility is that shorter mRNAs, which include ribosomal proteins and other housekeeping genes (Hurowitz and Brown, 2003), might be under selection for faster initiation rates by virtue of their high expression. However, our stepwise regression showed that ORF length was informative even after accounting for mRNA abundance. Another possibility is that the 5'-UTR-bound initiation machinery can sense and be affected by ORF length via the closed-loop structure. In eukaryotes, translating mRNAs are thought to adopt a pseudo-circularized structure in which the 5' and 3' ends are in close proximity, enhancing translation and mRNA stability (Christensen et al., 1987). Previous biochemical analysis of the closed loop in yeast extracts revealed that only short mRNAs adopt a stable closed-loop structure in vitro (Amrani et al., 2008), presumably due to the relatively short distance between the mRNA termini. If the same principle applies in vivo, then inefficient closed-loop formation of long mRNAs could explain their relatively low IEs.

EXPERIMENTAL PROCEDURES

Yeast Culture, Harvesting, and Lysate Preparation

S. cerevisiae strain BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) was grown at 30°C in 500 ml YPD to OD₆₀₀ 0.5. Cells were harvested by filtration using a Kontes Ultra-Ware Microfiltration Assembly with a Supor 450 Membrane Disc Filter that had been pre-wet with YPD. As the last liquid flowed through, the filtration apparatus was rapidly disassembled, cells were gently scraped off of the filter using a cell lifter, and the scraper was immediately submerged in a 50-ml conical tube filled with liquid nitrogen. Once the liquid nitrogen had boiled off, the pellet was stored in the conical tube at -80°C until lysis. To lyse cells under cryogenic conditions, the cell pellet was transferred into a pre-chilled mortar that was surrounded and filled with liquid nitrogen. The pellet was ground to a fine powder with a pre-chilled pestle, transferred into a 50-ml conical tube filled with liquid nitrogen, and after boiling off the liquid stored at -80°C. Crude lysate was prepared by briefly thawing the cell powder on ice for 1 min and then resuspending in 4 ml polysome lysis buffer (10 mM Tris-HCl [pH 7.4], 5 mM MgCl₂, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 100 mg/ml cycloheximide, 500 U/ml RNasin Plus RNase Inhibitor [Promega], cOmplete EDTA-free Protease Inhibitor Cocktail [Roche]). The lysate was centrifuged at 1,300 g for 10 min, and the supernatant was flash frozen in single-use aliquots.

RNA-Seq

Total RNA was extracted from an aliquot of frozen yeast lysate using TRI Reagent (Ambion) according to the manufacturer's protocol. Aliquots of the same sample were subjected to either no enrichment (the total RNA sample), poly(A) selection using 30 μ g total RNA and 100 μ l Dynabeads oligo(dT)₂₅ (Life Technologies) according to the manufacturer's instructions, rRNA depletion using 4 μ g total RNA and the RiboMinus Yeast Transcriptome Isolation Kit (Life Technologies) according to the manufacturer's instructions, and rRNA depletion using 10 μ g total RNA and the Ribo-Zero Gold Yeast rRNA Removal Kit (Illumina) according to the manufacturer's instructions. RNA samples were then diluted to 90 μ l with water and precipitated with 10 μ l 3 M NaCl, 30 μ g GlycoBlue (Life Technologies), and 250 μ l ethanol. RNA-seq was performed as described (Subtelny et al., 2014), using five cycles of PCR.

Ribosome Profiling

RPFs were isolated from an aliquot of frozen yeast lysate and sequenced on the Illumina HiSeq platform, as described (Subtelny et al., 2014). Detailed protocols for RNA-seq and ribosome profiling are available at <http://bartellab.wi.mit.edu/protocols.html>. RNase I treatment was performed using 0.2 U/ml lysate. Subtractive hybridization to remove contaminating rRNA fragments was using a mixture of three biotinylated oligonucleotides (Integrated DNA Technologies): 5'-GATCGGTGATTGTGCACCTC/3Bio/; 5'-CGC TTCATTGAATAAGTAAAG/3Bio/; 5'-GACGCCTTATTCGTATCCATC/3Bio/.

Analyses

Equations and detailed procedures for analyses are provided in Supplemental Experimental Procedures.

ACCESSION NUMBERS

Sequencing data have been deposited in the GEO database under accession number GEO: GSE75897.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2016.01.043>.

AUTHOR CONTRIBUTIONS

D.E.W., P.S., S.W.E., J.B.P., and D.P.B. designed the study. D.E.W., P.S., J.B.P., and D.P.B. wrote the manuscript, with help from the other authors. D.E.W. and S.W.E. prepared RNA-seq and ribosome-profiling libraries, respectively, under the supervision of D.P.B. D.E.W. performed initial data analysis. P.S. performed detailed data analyses and simulations with help from J.B.P. J.A.H. contributed new analytical tools.

ACKNOWLEDGMENTS

We thank J. Weissman and L. Lareau for helpful discussions, and the Whitehead Genome Technology Core for sequencing. This work was supported by the UCSF Program for Breakthrough Biomedical Research (funded in part by the Sandler Foundation, D.E.W.) and by NIH grants DP5OD017895 (D.E.W.) and GM061835 (D.P.B.), the Burroughs Wellcome Fund (J.B.P.), the David and Lucile Packard Foundation (J.B.P.), US Department of the Interior Grant D12AP00025 (J.B.P.), and US Army Research Office Grant W911NF-12-1-0552 (J.B.P.). D.P.B. is an Investigator of the Howard Hughes Medical Institute.

Received: July 6, 2015

Revised: November 17, 2015

Accepted: January 8, 2016

Published: February 11, 2016

REFERENCES

- Amrani, N., Ghosh, S., Mangus, D.A., and Jacobson, A. (2008). Translation factors promote the formation of two states of the closed-loop mRNP. *Nature* 453, 1276–1280.
- Andersson, S.G., and Kurland, C.G. (1990). Codon preferences in free-living microorganisms. *Microbiol. Rev.* 54, 198–210.
- Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 100, 3889–3894.
- Arribere, J.A., and Gilbert, W.V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res.* 23, 977–987.
- Artieri, C.G., and Fraser, H.B. (2014). Evolution at two levels of gene expression in yeast. *Genome Res.* 24, 411–421.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002). The Pfam protein families database. *Nucleic Acids Res.* 30, 276–280.
- Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Blöchl, N. (2013). Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol.* 9, 675.
- Beznošková, P., Cuchalová, L., Wagner, S., Shoemaker, C.J., Gunišová, S., von der Haar, T., and Valášek, L.S. (2013). Translation initiation factors eIF3 and HCR1 control translation termination and stop codon read-through in yeast cells. *PLoS Genet.* 9, e1003962.
- Brandman, O., Stewart-Ornstein, J., Wong, D., Larson, A., Williams, C.C., Li, G.W., Zhou, S., King, D., Shen, P.S., Weibezahn, J., et al. (2012). A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* 151, 1042–1054.
- Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335, 552–557.
- Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129, 897–907.
- Charneski, C.A., and Hurst, L.D. (2013). Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol.* 11, e1001508.
- Christensen, A.K., Kahn, L.E., and Bourne, C.M. (1987). Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. *Am. J. Anat.* 178, 1–10.
- Chu, D., and von der Haar, T. (2012). The architecture of eukaryotic translation. *Nucleic Acids Res.* 40, 10098–10106.
- Csárdi, G., Franks, A., Choi, D.S., Airoldi, E.M., and Drummond, D.A. (2015). Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. *PLoS Genet.* 11, e1005206.
- de Godoy, L.M., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Fröhlich, F., Walther, T.C., and Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* 455, 1251–1254.
- Dever, T.E., Feng, L., Wek, R.C., Cigan, A.M., Donahue, T.F., and Hinnebusch, A.G. (1992). Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. *Cell* 68, 585–596.
- Ding, Y., Shah, P., and Plotkin, J.B. (2012). Weak 5'-mRNA secondary structures in short eukaryotic genes. *Genome Biol. Evol.* 4, 1046–1053.
- Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R., and Weissman, J.S. (2013). Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *eLife* 2, e01179.
- Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., and Fitch, B. (2014). Measurement of average decoding rates of the 61 sense codons in vivo. *eLife* 3, 3.

- Gerashchenko, M.V., and Gladyshev, V.N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. *Nucleic Acids Res.* **42**, e134.
- Gerashchenko, M.V., Lobanov, A.V., and Gladyshev, V.N. (2012). Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc. Natl. Acad. Sci. USA* **109**, 17394–17399.
- Godefroy-Colburn, T., Ravelonandro, M., and Pinck, L. (1985). Cap accessibility correlates with the initiation efficiency of alfalfa mosaic virus RNAs. *Eur. J. Biochem.* **147**, 549–552.
- Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **4**, 117.
- Gu, W., Zhou, T., and Wilke, C.O. (2010). A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput. Biol.* **6**, e1000664.
- Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840.
- Guydosh, N.R., and Green, R. (2014). Dom34 rescues ribosomes in 3' untranslated regions. *Cell* **156**, 950–962.
- Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E., and Collier, J. (2009). Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature* **461**, 225–229.
- Hurowitz, E.H., and Brown, P.O. (2003). Genome-wide analysis of mRNA lengths in *Saccharomyces cerevisiae*. *Genome Biol.* **5**, R2.
- Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.* **15**, 205–213.
- Ingolia, N.T., Ghaemmghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223.
- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802.
- Jan, C.H., Williams, C.C., and Weissman, J.S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346**, 1257521.
- Jayaprakash, A.D., Jabado, O., Brown, B.D., and Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* **39**, e141.
- Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240.
- Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y., and Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103–107.
- Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M. (2007). A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**, 525–528.
- Kozak, M. (1986). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* **83**, 2850–2854.
- Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258.
- Lareau, L.F., Hite, D.H., Hogan, G.J., and Brown, P.O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife* **3**, e01257.
- Li, G.W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538–541.
- Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635.
- McManus, C.J., May, G.E., Spealman, P., and Shteyman, A. (2014). Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res.* **24**, 422–430.
- Mueller, P.P., and Hinnebusch, A.G. (1986). Multiple upstream AUG codons mediate translational control of GCN4. *Cell* **45**, 201–207.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349.
- Nedialkova, D.D., and Leidel, S.A. (2015). Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity. *Cell* **161**, 1606–1618.
- Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* **20**, 237–243.
- Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. *Cell* **161**, 1400–1412.
- Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42.
- Pop, C., Rouskin, S., Ingolia, N.T., Han, L., Phizicky, E.M., Weissman, J.S., and Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol. Syst. Biol.* **10**, 770.
- Qian, W., Yang, J.R., Pearson, N.M., Maclean, C., and Zhang, J. (2012). Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* **8**, e1002603.
- Richter, J.D., and Sonenberg, N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* **433**, 477–480.
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705.
- Riegsegger, U., Leber, J.H., and Walter, P. (2001). Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. *Cell* **107**, 103–114.
- Schneider-Poetsch, T., Ju, J., Elyer, D.E., Dang, Y., Bhat, S., Merrick, W.C., Green, R., Shen, B., and Liu, J.O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat. Chem. Biol.* **6**, 209–217.
- Shah, P., and Gilchrist, M.A. (2011). Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc. Natl. Acad. Sci. USA* **108**, 10231–10236.
- Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B. (2013). Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589–1601.
- Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295.
- Siridechadilok, B., Fraser, C.S., Hall, R.J., Doudna, J.A., and Nogales, E. (2005). Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **310**, 1513–1515.
- Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731–745.
- Sorefan, K., Pais, H., Hall, A.E., Kozomara, A., Griffiths-Jones, S., Moulton, V., and Dalmay, T. (2012). Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* **3**, 4.
- Sørensen, M.A., and Pedersen, S. (1991). Absolute in vivo translation rates of individual codons in *Escherichia coli*. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J. Mol. Biol.* **222**, 265–280.
- 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**, 66–71.
- Szamecz, B., Rutkai, E., Cuchalová, L., Munzarová, V., Herrmannová, A., Nielsen, K.H., Burela, L., Hinnebusch, A.G., and Valásek, L. (2008). eIF3a cooperates with sequences 5' of uORF1 to promote resumption of scanning by

- post-termination ribosomes for reinitiation on GCN4 mRNA. *Genes Dev.* **22**, 2414–2425.
- Thanaraj, T.A., and Argos, P. (1996). Ribosome-mediated translational pause and protein domain organization. *Protein Sci.* **5**, 1594–1612.
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354.
- Tuller, T., Veksler-Lublinsky, I., Gazit, N., Kupiec, M., Ruppin, E., and Ziv-Ukelson, M. (2011). Composite effects of gene determinants on the translation speed and density of ribosomes. *Genome Biol.* **12**, R110.
- Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J. Mol. Biol.* **180**, 549–576.
- Williams, C.C., Jan, C.H., and Weissman, J.S. (2014). Targeting and plasticity of mitochondrial proteins revealed by proximity-specific ribosome profiling. *Science* **346**, 748–751.
- Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., and Gough, J. (2009). SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* **37**, D380–D386.
- Xu, Y., Ma, P., Shah, P., Rokas, A., Liu, Y., and Johnson, C.H. (2013). Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature* **495**, 116–120.
- Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J.M., Sachs, M.S., and Liu, Y. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* **495**, 111–115.
- Zinshteyn, B., and Gilbert, W.V. (2013). Loss of a conserved tRNA anticodon modification perturbs cellular signaling. *PLoS Genet.* **9**, e1003675.
- Zur, H., and Tuller, T. (2013). New universal rules of eukaryotic translation initiation fidelity. *PLoS Comput. Biol.* **9**, e1003136.

Appendix D

Genome-Scale Networks Link Neurodegenerative Disease Genes to a-Synuclein through Specific Molecular Pathways

Vikram Khurana^{1,2,3,14,17}, Jian Peng^{1,4,14,15}, Chee Yeun Chung^{1,14,16}, Pavan K. Auluck¹, Saranna Fanning¹, Daniel F. Tardiff^{1,16}, Theresa Bartels¹, Martina Koeva^{1,5}, Stephen W. Eichhorn¹, Hadar Benyamini¹, Yali Lou¹, Andy Nutter-Upham¹, Valeriya Baru¹, Yelena Freyzon¹, Nurcan Tuncbag⁵, Michael Costanzo⁶, Bryan-Joseph San Luis⁶, David C. Schöndorf⁷, M. Inmaculada Barrasa¹, Sepehr Ehsani¹, Neville Sanjana^{8,9}, Quan Zhong¹⁰, Thomas Gasser⁷, David P. Bartel¹, Marc Vidal^{11,12}, Michela Deleidi⁷, Charles Boone⁶, Ernest Fraenkel⁵, Bonnie Berger³, Susan Lindquist^{1,13}

¹Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA

²Ann Romney Center for Neurologic Disease, Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

³Harvard Stem Cell Institute, Cambridge, MA 02138, USA

⁴Computer Science and Artificial Intelligence Laboratory and Department of Mathematics, MIT, Cambridge, MA 02139, USA

⁵Department of Biological Engineering, MIT, Cambridge, MA 02139, USA

⁶Banting and Best Department of Medical Research, University of Toronto, Toronto, ON M5G 1L6, Canada

⁷Department of Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), and Hertie-Institute for Clinical Brain Research, University of Tübingen, Tübingen, 72076, Germany

⁸Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

⁹New York Genome Center and Department of Biology, New York University, New York, NY 10013, USA

¹⁰Department of Biological Sciences, Wright State University, Dayton, OH 45435, USA

¹¹Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

¹²Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

¹³HHMI, Department of Biology, MIT, Cambridge, MA 02139, USA

The study was designed and supervised by V.K. and S.L. TransposeNet was conceived by V.K. and executed by J.P., with key contributions by M.K. and N.T., B.B., and E.F. and supervision by B.B. and E.F. C.Y.C. led the neuronal assay development efforts. P.K.A., S.F., D.F.T., T.B., S.W.E., H.B., Y.L., A.N.-U., V.B., Y.F., M.C., B.-J.S.L., D.C.S., M.I.B., S.E., N.S., Q.Z., T.G., D.P.B., M.V., M.D., and C.B. performed bench or computational experiments or supervised these efforts. The paper was written by V.K., C.Y.C., J.P., B.B., E.F., and S.L.

Published as:

Khurana, V., Chung, C.Y., Peng, J., Auluck, P.K., Tardiff, D.F., Fanning, S., Bartels, T., Koeva, M., Eichhorn, S.W., Benyamini, H., Lou, Y., Nutter-Upham, A., Baru, V., Freyzon, Y., Tuncbag, N., Costanzo, M., San-Luis, B., Schöndorf, D.C., Barrasa, M.I., Ehsani, S., Sanjana, N., Zhong, Q., Gasser, T., Bartel, D.P., Vidal, M., Deleidi, M., Boone, C., Berger, B., Fraenkel, E., Lindquist, S. (2016). Genome-Scale Networks Link Neurodegenerative Disease Genes to α -Synuclein through Specific Molecular Pathways. *Cell Systems* 4, 157-170.e14

Genome-Scale Networks Link Neurodegenerative Disease Genes to α -Synuclein through Specific Molecular Pathways

Vikram Khurana,^{1,2,3,14,17,*} Jian Peng,^{1,4,14,15} Chee Yeun Chung,^{1,14,16} Pavan K. Auluck,¹ Saranna Fanning,¹ Daniel F. Tardiff,^{1,16} Theresa Bartels,¹ Martina Koeva,^{1,5} Stephen W. Eichhorn,¹ Hadar Benyamini,¹ Yali Lou,¹ Andy Nutter-Upham,¹ Valeriya Baru,¹ Yelena Freyzon,¹ Nurcan Tuncbag,⁵ Michael Costanzo,⁶ Bryan-Joseph San Luis,⁶ David C. Schöndorf,⁷ M. Inmaculada Barrasa,¹ Sepehr Ehsani,¹ Neville Sanjana,^{8,9} Quan Zhong,¹⁰ Thomas Gasser,⁷ David P. Bartel,¹ Marc Vidal,^{11,12} Michela Deleidi,⁷ Charles Boone,⁶ Ernest Fraenkel,^{5,*} Bonnie Berger,^{3,*} and Susan Lindquist^{1,13}

¹Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA

²Ann Romney Center for Neurologic Disease, Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

³Harvard Stem Cell Institute, Cambridge, MA 02138, USA

⁴Computer Science and Artificial Intelligence Laboratory and Department of Mathematics, MIT, Cambridge, MA 02139, USA

⁵Department of Biological Engineering, MIT, Cambridge, MA 02139, USA

⁶Banting and Best Department of Medical Research, University of Toronto, Toronto, ON M5G 1L6, Canada

⁷Department of Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), and Hertie-Institute for Clinical Brain Research, University of Tübingen, Tübingen, 72076, Germany

⁸Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

⁹New York Genome Center and Department of Biology, New York University, New York, NY 10013, USA

¹⁰Department of Biological Sciences, Wright State University, Dayton, OH 45435, USA

¹¹Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

¹²Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

¹³HHMI, Department of Biology, MIT, Cambridge, MA 02139, USA

¹⁴Co-first author

¹⁵Present address: Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL 61801, USA

¹⁶Present address: Yumanity Therapeutics, Cambridge, MA 02139, USA

¹⁷Lead Contact

*Correspondence: vkhurana@bwh.harvard.edu (V.K.), fraenkel-admin@mit.edu (E.F.), bab@csail.mit.edu (B.B.)

<http://dx.doi.org/10.1016/j.cels.2016.12.011>

SUMMARY

Numerous genes and molecular pathways are implicated in neurodegenerative proteinopathies, but their inter-relationships are poorly understood. We systematically mapped molecular pathways underlying the toxicity of alpha-synuclein (α -syn), a protein central to Parkinson's disease. Genome-wide screens in yeast identified 332 genes that impact α -syn toxicity. To "humanize" this molecular network, we developed a computational method, TransposeNet. This integrates a Steiner prize-collecting approach with homology assignment through sequence, structure, and interaction topology. TransposeNet linked α -syn to multiple parkinsonism genes and druggable targets through perturbed protein trafficking and ER quality control as well as mRNA metabolism and translation. A calcium signaling hub linked these processes to perturbed mitochondrial quality control and function, metal ion transport, transcriptional regulation, and signal transduction. Parkinsonism gene interaction profiles spatially opposed in the network (*ATP13A2/PARK9* and *VPS35/PARK17*) were highly

distinct, and network relationships for specific genes (*LRRK2/PARK8*, *ATXN2*, and *EIF4G1/PARK18*) were confirmed in patient induced pluripotent stem cell (iPSC)-derived neurons. This cross-species platform connected diverse neurodegenerative genes to proteinopathy through specific mechanisms and may facilitate patient stratification for targeted therapy.

INTRODUCTION

Common neurodegenerative diseases result in the loss of distinct neuronal populations and abnormal accumulation of misfolded proteins. Synucleinopathies, including Parkinson's disease (PD), dementia with Lewy bodies, and multiple system atrophy, are associated with abnormal intracellular aggregation of alpha-synuclein (α -syn). Alzheimer's disease (AD) is associated with beta-amyloid ($A\beta$) and tau accumulation, while amyotrophic lateral sclerosis (ALS) is associated with altered localization and accumulation of TAR DNA-binding protein 43 (TDP-43). The richest source of hypotheses regarding the pathogenesis of these diseases has derived from neuropathology of postmortem brain. While providing pivotal insights, these observations are made decades after disease inception.



A revolution in human genetic analysis over the last 20 years has uncovered disease-causing mutations that connect protein misfolding to the neurodegenerative process. For instance, point mutations and gene multiplication at the α -syn (*SNCA*) locus lead to rare but early-onset, highly penetrant forms of PD and dementia. Common polymorphisms in regulatory regions of the *SNCA* locus that increase gene expression confer increased risk for late-onset PD (Fuchs et al., 2008; Nalls et al., 2014). These studies enabled the creation of animal and cellular disease models and enriched our understanding of disease mechanisms. However, with this knowledge, a new set of challenges has emerged.

First, seemingly disparate genes have been tied to particular clinical phenotypes. For example, parkinsonism is characterized by slowness (bradykinesia), rigidity, tremor, and postural instability. The most common form is PD, defined by α -syn pathology and loss of dopaminergic neurons. However, numerous other disease entities—tied to distinct genetic signatures and neuropathology—can lead to parkinsonism, demonstrating that there is not a simple correspondence among genotype, neuropathology, and clinical presentation (Martin et al., 2011; Shulman et al., 2011; Verstraeten et al., 2015). Those few genetic loci with parkinsonism as the primary clinical phenotype have been given a numeric “*PARK*” designation (for example, the *SNCA/PARK1* locus itself and *LRK2/PARK8*), but even mutations in the same gene can produce distinct neuropathology and diverse clinical presentations (Martin et al., 2011; Rajput et al., 2006; Shulman et al., 2011; Verstraeten et al., 2015). Understanding the inter-relationship between genetic risk factors for parkinsonism, and their relationship to α -syn itself, is vital for patient stratification and targeted therapeutic strategies.

Second, human genetic studies have sometimes produced ambiguous and controversial data. For rare variants, substantial recent genetic divergence of human populations may render traditional methods of cross-validation between different populations unfeasible (Nelson et al., 2012; Tennessen et al., 2012). Inconsistencies in the literature abound; for example, studies implicating *UCHL1* as “*PARK5*” and the translation initiation factor *EIF4G1* as “*PARK18*” have failed to reproduce. For common polymorphisms, the challenge is distinguishing between multiple candidate gene loci in linkage to a SNP. It is becoming clear that biological validation will be required to fully establish which genetic factors are causally related to disease processes and how (Casals and Bertranpetit, 2012).

One approach to validating candidate gene variants, and understanding their relationship to proteinopathy, is to systematically screen the entire genome to identify every gene that modifies proteotoxicity when overexpressed or deleted. This is achievable in baker’s yeast (*Saccharomyces cerevisiae*), a unicellular eukaryote of unparalleled genetic tractability. Yeast has proved highly informative for understanding the cytotoxicity induced by misfolded proteins (Khurana and Lindquist, 2010). This is not surprising, because human genetic data for neurodegenerative diseases heavily implicate cellular pathways that are among the most highly conserved in eukaryotic evolution, including protein homeostasis and quality control, protein trafficking, RNA metabolism, and mitochondrial function (Brás et al., 2015; Guerreiro et al., 2015).

Expressing toxic proteins relevant to neurodegeneration in yeast creates a robust and easily scorable growth and viability

defect amenable to genome-wide phenotypic screening. Toxicities of α -syn, A β , and TDP-43 have been screened by individually overexpressing each one of $\approx 5,500$ open reading frames (ORFs) that comprise the majority of the yeast genome (Khurana and Lindquist, 2010; Kim et al., 2014; Treusch et al., 2011; Yeger-Lotem et al., 2009). These screens have guided the discovery of cellular pathologies in neurons and animal models (Cooper et al., 2006; Dhungel et al., 2015; Khurana and Lindquist, 2010; Kim et al., 2014), provided important insights on the relationship of genetic modifier data to gene-expression analysis (Yeger-Lotem et al., 2009), and led to the identification of novel human disease genes (Elden et al., 2010). Recently, processes pinpointed by phenotypic screening in a yeast synucleinopathy model led to the discovery of cellular pathologies in induced pluripotent stem cell (iPSC)-derived neurons from patients with PD due to mutations at the α -syn locus (Chung et al., 2013). In that study, integrating high-throughput genetic and small-molecule screens identified genes and small molecules that could correct pathologies from yeast to neurons (Chung et al., 2013; Tardiff et al., 2013, 2014).

Here, we build genome-scale networks of α -syn and other proteotoxicities by combining a new computational approach with substantially broader yeast genetic screens. To discover meaningful molecular connections in yeast and patient-derived neurons, we develop a TransposeNet algorithm that (1) maps yeast hits to their human homologs by considering sequence, structure, and molecular interactions; (2) builds networks by linking yeast hits and hidden human genes through an optimization framework based on the prize-collecting Steiner forest algorithm (SteinerForest Ensemble); and (3) transposes molecular interactions across species from yeast to human, utilizing the high density of known molecular interactions in yeast to compensate for the relative sparseness of the human interactome. The networks linked many parkinsonism and neurodegenerative disease risk factors to α -syn toxicity through specific molecular pathways, most notably vesicle trafficking and mRNA metabolism.

RESULTS

SteinerForest Ensemble Networks Uncover Biological Connections between α -Syn Screen Hits

One conventional approach to creating a network from a gene list is to connect them via known genetic or physical protein-protein interactions. To illustrate this, we considered 77 genes that modify α -syn toxicity in our previous overexpression screen (Tables S1 and S2). Even with the rich yeast interactome, 30 hits were not incorporated into the network (Figure 1A, top). Moreover, some genes (“hubs”) occupied a central position in the network, not because of their importance to proteotoxicity but because they were connected to more genes. For example, *PMR1* is a hub that has 955 annotated interactions in BioGRID compared to the median of 70 interactions across the 77 modifiers (Figure 1A, top right; Table S2).

To build more inclusive networks, we adapted the “Prize-collecting Steiner Forest (PCSF) algorithm,” which connects gene or protein “nodes” through molecular interactions, or “edges” (Huang and Fraenkel, 2009; Tuncbag et al., 2013, 2016) (Figure 1A, bottom). Edges can include genetic or physical

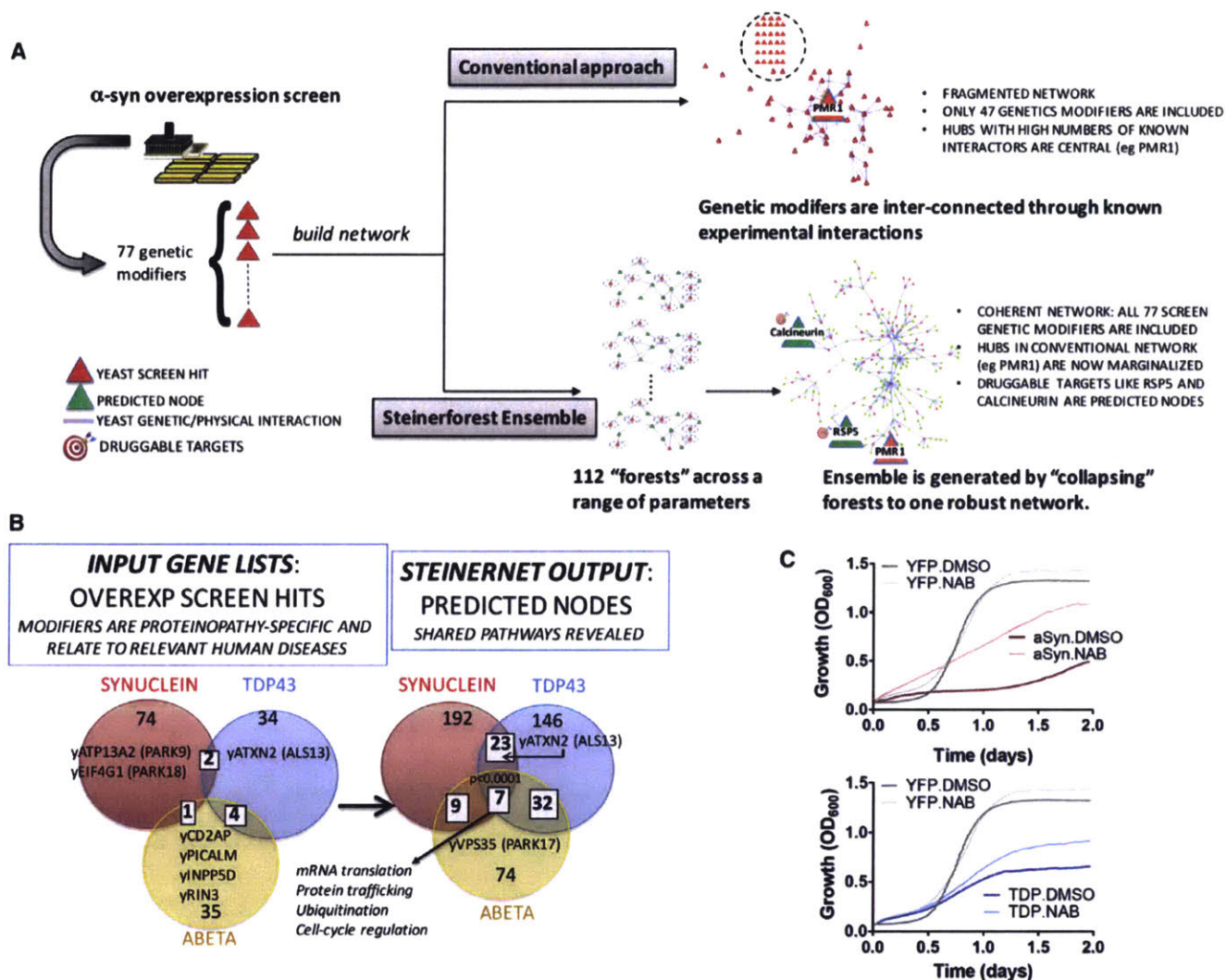


Figure 1. SteinerForest Ensemble Builds Proteotoxicity Networks from Yeast Genetic Screens and Uncovers Druggable Targets

(A) SteinerForest Ensemble methodology versus conventional approach. 77 genetic modifiers (“hits”) from a previous overexpression screen against α -syn toxicity are mapped to the yeast interactome. The conventional approach misses 30 genetic hits and overemphasizes “hub” genes like PMR1. SteinerForest Ensemble includes all 77 hits and predicts additional nodes of biological relevance, including the druggable targets Rsp5 and Calcineurin (Cnb1).

(B) Left: hits from three published overexpression screens for α -syn, A β , and TDP-43 proteotoxicities in yeast. Venn diagrams indicate the numbers of genetic modifiers recovered. Right: a comparison of the output SteinerForest Ensemble networks generated from inputting these three sets of screen hits (empirical p value for 1,000 similarly connected random networks is shown for triple-wise comparison).

(C) Growth curves demonstrating effects of a compound that activates Rsp5, NAB, on TDP-43 (bottom) and α -syn (top) toxicity. Yeast expressing yellow fluorescent protein (YFP), TDP-43(TDP), or α -syn were treated with 20 μ M (for TDP-43) or 10 μ M (for α -syn) NAB. Growth was monitored over time by optical density (OD) at 600 nm. Results are representative of three experiments.

interactions or annotated pathways from curated databases (Szkarczyk et al., 2015) and are refined by minimizing “cost.” Costs increase (1) when a “prized” node (an original hit from a genetic screen) is excluded, (2) when an “edge” connecting two nodes derives from a low-confidence interaction, or (3) when edges connect to hubs. To ensure that our PCSFs were not dependent on specific parameterization, we generated an ensemble of 112 individual forests with different algorithm parameters and created an averaged, or “collapsed,” representative network through a maximum spanning-tree algorithm (“SteinerForest Ensemble”; Figure 1A, bottom right).

To encompass the largest number of prized nodes while avoiding unlikely interactions, the PCSF method introduces

“predicted nodes,” which are proteins or genes not part of the original prized hit list, (Figure 1A, green triangles). Predicted nodes will occur between two nodes within the network. However, as the final network is a superposition of many different networks, these may be at the periphery in the final Ensemble output. Predicted nodes can add biological value, because any high-throughput screen will miss many true biological connections.

When we applied SteinerForest Ensemble to our previous α -syn overexpression screen data, the fragmented networks became more coherently connected. All 77 modifier genes were now incorporated in the network, (Figure 1A, bottom right; Tables S1 and Table S3; Data S1). By penalizing the exclusion

of genetic modifiers and the inclusion of hubs, the PCSF algorithm favored the biological context at the expense of hubs. To establish specificity of the network output, we generated ensembles of forests from 1,000 sets of 77 genes randomly chosen from the yeast genome with identical connectivity (degree distribution) to the α -syn modifier list. An empiric p value for each node (based on probability of occurring in a network by chance) was significant ($p = 0.025$, $SD = 0.021$).

Importantly, predicted nodes (Figure 1A, green triangles) included genetic modifiers of α -syn toxicity not hit in the original screen but uncovered through other studies, including Sec14 (phospholipase D) (Outeiro and Lindquist, 2003), and Pbp1 (yeast homolog of ataxin 2; see below and Figure 3). This network also identified two druggable targets: Cnb1 (Calcineurin subunit B) and Rsp5 (Figure 1A, bottom right). Cnb1 is targeted by FK506, a drug that ameliorates α -syn toxicity (Caraveo et al., 2014). Rsp5 is the target of a specific N-arylbenzimidazole (NAB) that protects against α -syn toxicity (Tardiff et al., 2013). The SteinerForest Ensemble methodology thus connects genetic screen hits through biologically relevant pathways, including druggable targets.

Cross-Comparison of α -Syn, TDP-43, and A β Proteinopathies Reveals Distinct and Shared Mechanisms

To cross-compare different proteinopathies, we examined previous A β and TDP-43 overexpression screens (Figure 1B; “yeast over-expression networks” in Table S1; Data S1) and found virtually no overlap (Figure 1B, left; Table S2). There was, however, reassuring overlap between the yeast genetic modifiers and disease genes associated with the human disorders, including putative parkinsonism genes recovered in the α -syn screen (*ATP13A2* [*PARK9*] and *EIF4G1* [*PARK18*]), AD risk factors in the A β screen (*PICALM*, *CD2AP*, *INPP5D*, and *RIN3*), and an ALS genetic risk factor (*ATXN2*) in the TDP-43 screen (Elden et al., 2010).

SteinerForest Ensembles from these screen hits revealed more biological overlap between these proteinopathies including protein trafficking, mRNA translation, ubiquitination, and cell-cycle genes (Tables S3 and S4; Figure 1B, right). There was also a crossover between genetic risk factors for distinct human diseases: the *ATXN2* homolog was a predicted node in the α -syn network (confirmed as a modifier of α -syn toxicity; Figures 3 and 4), and the homolog of the parkinsonism gene *VPS35* (*PARK17*) was a predicted node in the yeast A β network. *VPS35* encodes a key component of the retromer complex, and defective retromer function has been identified in AD brain and animal models (Small et al., 2005). These overlaps were unrelated to increasing the number of genes. Empirical p values for 1,000 similarly connected random networks were statistically significant, whether considered pairwise ($p \leq 0.002$) or triple-wise ($p \leq 0.001$).

One trafficking gene predicted to be a common node between all three proteinopathies was Rsp5, a ubiquitin ligase activated by NAB. Indeed, NAB was originally recovered in a small-molecule screen against TDP-43 proteinopathy in yeast. We utilized a sensitive bioscreen assay to test NAB on growth defects induced by these proteinopathies. NAB rescued all three proteinopathies as predicted by the network. It was most effective for α -syn (Figure 1C) and only rescued against

A β toxicity synergistically in combination with other compounds known to protect from A β toxicity (unpublished data). NAB failed to provide significant rescue for 20 unrelated toxic yeast strains (Figure S1).

TransposeNet Generates a “Humanized” Network

It would be desirable to identify connections between our yeast molecular networks to human genes, including human disease genes that have no straightforward homologs in yeast. We therefore developed TransposeNet, a suite of computational methods to “humanize” yeast molecular networks (Figure 2A).

The first step in TransposeNet is assignment of yeast-to-human homology by considering sequence similarity (BLAST and DIOPT scores; Hu et al., 2011), yeast-to-human structure alignments (using the *HHpred* tool; Söding et al., 2005), and incorporating network topology (Figure 2A, top left). Network topology assesses neighborhoods of genetic and physical molecular interactions around a given protein, positing “guilt-by-association” logic that the topological place within a network relates to biological function (Cho et al., 2016). Thus, sharing similar neighbors should be a factor in determining whether two proteins are homologs. The relative weight of each homology method was carefully tuned (see STAR Methods and Figure S2 for full details), providing a more comprehensive and unified protein homology score (Berger et al., 2013; Singh et al., 2008; Söding et al., 2005). The underlying framework that relates genes according to these different features is known as diffusion-component analysis (DCA). DCA has also been used as the core algorithm in Mashup, a tool for integrating multiple heterogeneous interactomes. More information can be found in the Method Details section of STAR Methods and in Cho et al., (2016).

Our method assigned 4,923 yeast proteins to human homologs and conversely predicted yeast homologs for 15,200 human proteins, a substantial improvement over BLAST (4,023 yeast to human and 7,248 human to yeast) or BLAST with HHpred (4,312 yeast to human and 9,577 human to yeast). Additionally, our method improved predictions as determined by gene ontology (GO) accuracy and Jaccard similarity scores (STAR Methods; Figure S3) and did not introduce false positives for high-confidence yeast-human protein pairs (EnsemblCompara; STAR Methods).

There is high conservation of core eukaryotic biology from yeast to man, and pivotal complementation studies in yeast have determined the functions of many genes in other species, including human (Osborn and Miller, 2007; Kachroo et al., 2015). On this basis, we used our homology tool to augment the human interactome with interactions inferred from the much richer yeast interactome. This was the key advance that enabled TransposeNet. Importantly, this cross-species “edge” transposition did not increase the rate of false-positive hits; rather, it substantially improved network performance. In fact, for our screen hits the PCSF-based SteinerForest Ensemble outperformed two alternative network-building methodologies, DAPPLE (Rossin et al., 2011) and PEXA (Tu et al., 2009) (STAR Methods; Figure S4).

In our humanized networks (indexed in Table S1; Figure 2A, right), each yeast gene (red triangle) is connected to one or more human homologs (blue circles) based on our homology-tool-generated score. SteinerForest Ensemble then interconnects

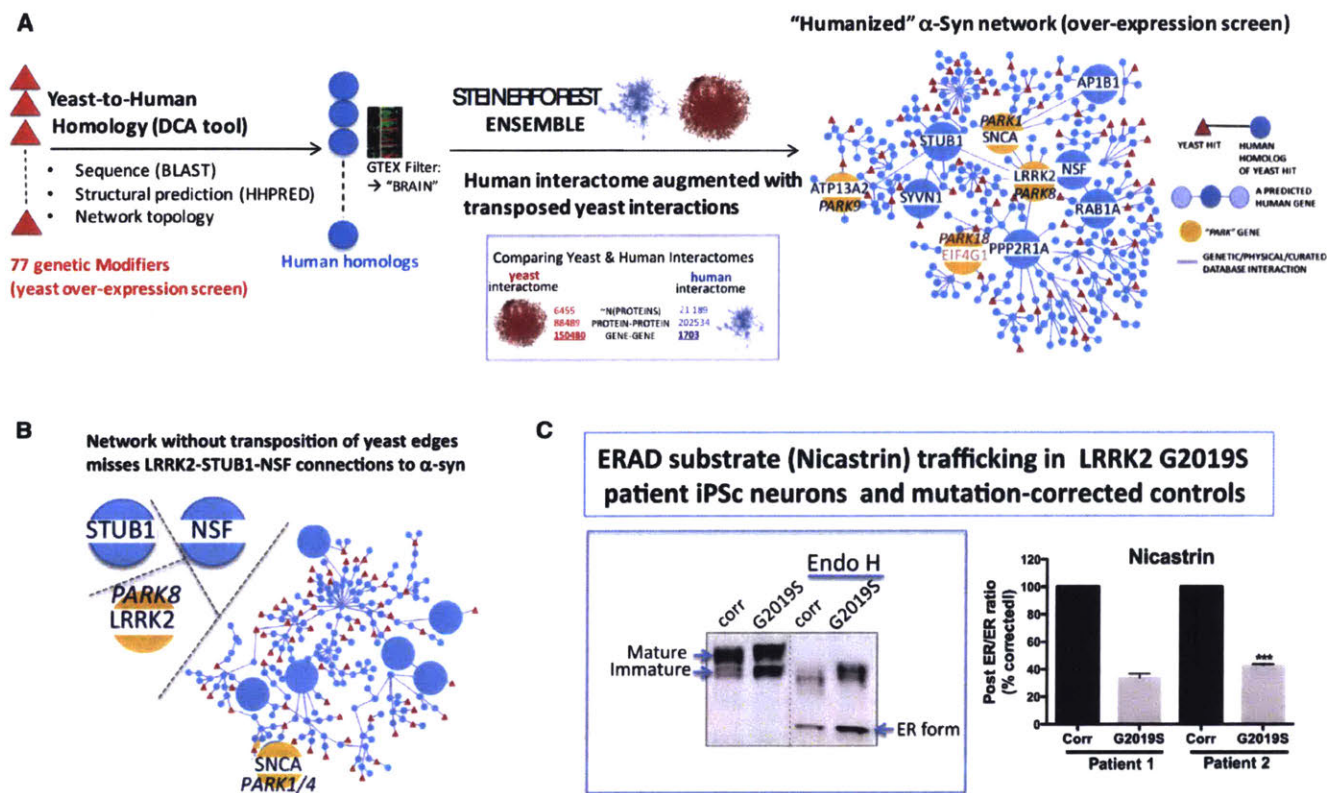


Figure 2. A "Humanized" TransposeNet Network Incorporates LRRK2 into the α -Syn Proteotoxicity Network

(A) A humanized network is generated from the 77 α -syn overexpression screen hits by TransposeNet. Each yeast gene (red triangle) is linked to its human homolog(s) (blue circle) by a weight proportional to the homology strength. Edges are weighted based on their experimental level of confidence. Certain nodes are enlarged for emphasis. LRRK2 is linked within network via NSF1 and STUB1. EIF4G1 is marked in red because it is a controversial PD gene. Inset: density of known molecular interactions in yeast and human (Biogrid: <https://wiki.thebiogrid.org/doku.php/statistics>). Abbreviations: DCA, diffusion component analysis; PARK, known "parkinsonism" gene. See the Supplemental Information for complete network.

(B) The effect on the humanized network of withholding yeast edge augmentation.

(C) Accumulation of Nicastrin in the endoplasmic reticulum (ER) in LRRK2^{G2019S} mutant iPSC-derived dopaminergic neurons compared to mutation-corrected control neurons. Endoglycosidase H (Endo H) removes post-ER glycosylation and reveals the ER form of Nicastrin, an ER-associated degradation substrate. The post ER-to-ER ratio was calculated using the ratio of the mature form over the deglycosylated ER form. Data represent mean \pm SEM (n = 2 for patient 1 and n = 3 for patient 2; ***p < 0.0001, two-tailed t test).

each resulting human gene or protein, through edges generated from the human interactome augmented with the humanized yeast molecular interactome. If a particular human homolog of a yeast genetic modifier had been implicated as a parkinsonism gene, a small inclusion weight is given. However, no special preference was given to any human disease genes other than close homologs of our yeast hits.

Humanized Network from Overexpression Screen Connects α -Syn to Other Human Disease Genes

We tested the humanized network approach on the 77 modifiers from the α -syn overexpression screen (" α -syn over-expression humanized network"; Tables S1, S9, and S11; Data S1; Figure 2A, right). Several predicted human nodes in the resultant humanized network had no obvious homologs in the yeast proteome, the most striking example being α -syn itself. α -syn was connected to endoplasmic reticulum (ER) quality control and protein trafficking modifiers through a predicted node Ap1b1 (Figure 2A, right), a component of the clathrin adaptor complex that localizes in the immediate vicinity of α -syn in neurons (Chung

et al., 2017, in this issue of *Cell Systems*). The emergence of α -syn in the humanized network strongly indicates that a functional, highly interconnected relationship between our original yeast genetic hits and α -syn is conserved from yeast to man.

LRRK2 and α -Syn Are Connected through ER Stress and Vesicle Trafficking

The kinase/GTPase LRRK2, another PD-gene-encoded protein without an obvious yeast homolog, was centrally incorporated into the humanized network (Figure 2A, right). We tested the robustness and specificity of this finding by computationally generating ensembles of humanized Steiner forests from 1,000 lists of genes that were randomly selected (matching the size of our original α -syn genetic modifier list). LRRK2 and α -syn (SNCA) occurred together in 72% of humanized networks generated through SteinerNet Ensemble from our input list (individually, SNCA appeared in 86% and LRRK2 in 76% of networks). Neither was incorporated in any of the humanized networks generated from A β or TDP-43 screen hits ("TDP-43"- and "A β "-over-expression humanized networks" in Table S1; Data S1). LRRK2

and α -syn appeared together in 0/1,000 of the randomly generated network ensembles. Without transposition of yeast interaction information into our networks, α -syn was peripherally placed and its connection to Ap1b1 (see above) was lost and LRRK2 was entirely absent (Figure 2B). Thus, the inclusion of LRRK2 and α -syn is robust, specific, and dependent upon augmentation of human networks with yeast interaction data.

LRRK2 was related to the humanized α -syn network through proteins involved in ER-to-Golgi trafficking (Nsf1 and Rab1a) and ER quality control (Stub1/Chip/Scar6, Sgk1, and Synv1), pathways previously implicated among many others in LRRK2 (Cho et al., 2014; Liu et al., 2012) and α -syn (Chung et al., 2013; Cooper et al., 2006)-induced toxicity. Our data pinpointed these pathways as key points of convergence. We previously showed that the A53T mutation and triplication of wild-type α -syn leads to pathologic accumulation of specific trafficked proteins in the ER of patient-derived neurons, including Nicastrin (Chung et al., 2013). Using previously generated LRRK2 mutant iPSC (Figure S5), we recapitulated this phenotype. As early as 4 weeks after initiating differentiation, Nicastrin accumulated in the ER of LRRK2^{G2019S} dopaminergic neurons compared to isogenic mutation-corrected controls (Figure 2C), phenocopying the previously described pathology in neurons derived from patients with α -syn mutations. Thus, the humanized α -syn network correctly predicted convergence of cellular pathologies in distinct forms of parkinsonism. A Nicastrin trafficking defect has also been demonstrated in LRRK2 knockout mouse fibroblasts (Cho et al., 2014), raising the possibility that the G2019S mutation may lead to deficiency of a LRRK2-related function in protein trafficking.

Genome-wide Pooled Overexpression and Deletion Screens against α -Syn Toxicity

For a more comprehensive view, we executed two additional genome-wide screens against α -syn toxicity: (1) a genome-wide deletion screen to identify nonessential genes that, when deleted, lead to an extreme sensitivity to low levels of α -syn that would otherwise not be toxic (Figure S6A number 2; Table S5); (2) a pooled screen in which the galactose-inducible overexpression library was transformed en masse into our α -syn HiTox strain (Figure S6A, number 3; Table S6). For pooled screens, we sequenced plasmid DNA to identify genes specifically over- or under-represented under selective conditions. These are putative suppressors and enhancers of toxicity, respectively. We compared a similarly transformed YFP strain as a control. Pooled screens are far more rapid, and theoretically more sensitive, than individually transforming each library plasmid into the α -syn strain and measuring growth.

These screens encompassed tests of approximately 10,000 potential genetic interactions (\square 5,500 overexpression, \square 4,500 deletion). After extensive validation of the hits (Figures S6C and S7B), we recovered 318 genetic modifiers. Very little overlap existed between the specific genes recovered in the deletion versus the overexpression screens (Figure 3A). However, we found considerable overlap in the biological pathways represented (see network analysis below). 16 modifiers have emerged in independent work from our laboratory (Caraveo et al., 2014; Chung et al., 2013) or were identified herein (Table S7). 14 of

these were distinct from our screen hits, leading to a total of 332 genetic modifiers of α -syn toxicity (Figure 3A).

Homologs of PARK and Other Neurodegeneration Genes Modify α -Syn Toxicity in Yeast

Modifiers of α -syn toxicity included homologs of many known genetic risk factors for parkinsonism and other neurodegenerative disease phenotypes (Figure 3B; Table S14). These included genes involved in vesicle trafficking (*yRAB7L1*, *yRAB39B*, *ySORL1*, *ySYNJ1/PARK20*, and *yVPS35/PARK17*), mRNA translation (*yATXN2* and *yEIF4G1/PARK18*), mitochondrial quality control and function (*yCHCHD2/10*), metal ion transport (*yATP13A2*), transcriptional regulation (*yATXN7*), metabolism (*yDHDDS*), and signaling (*yPDE8B* and *yPPP2R2B/ATXN12*), among others. Many of these genes, including those at so-called *PARK* loci, have been implicated in neuronal pathologies quite distinct from the α -syn pathology that defines PD. Their recovery in our screens suggested that mechanisms of neurotoxicity related to diverse neurodegenerative disease genes might be shared.

Of the 19 *PARK* loci, 9 have clear yeast homologs (Table S8). Four of these emerged in our screens: *yATP13A2 (PARK9) (YPK9* in yeast), *yVPS35 (PARK17) (VPS35)*, *yEIF4G1 (PARK18) (TIF4631, TIF4632)*, and *ySYNJ1 (PARK20) (INP53)*. A fifth putative *PARK* gene, *yRAB7L1 (PARK16) (YPT7)*, emerged as a genetic modifier when tested as a candidate (see below). The probability of recovering homologs of these genes by chance is low ($p = 0.00013$; hypergeometric test). None of these *yPARK* genes were modifiers in the $A\beta$ or TDP-43 overexpression screens (Table S2). These findings underscore the biological specificity of the α -syn screen hits in yeast.

TransposonNet Generates a Genome-Scale “Map” of α -Syn Toxicity

We applied TransposonNet to homologs of the 332 α -syn toxicity modifiers to generate a humanized network, or “map” (“Complete α -syn humanized network” in Tables S1, S10, and S11; Figures 3B and S8; Data S1). Multiple genes implicated in neurodegeneration emerged in this α -syn network by direct homology to yeast hits or as predicted network nodes (Figure 3B; Figure S8; Table S14).

We superimposed gene ontologies onto “branches” in our map (Figure 3B), and various vesicle-mediated transport processes dominated. Genetic risk factors associated with typical PD (*SNCA* itself, *LRRK2*, *RAB7L1*, and *VPS35*) were concentrated in the subnetwork enriched in vesicle trafficking genes (Figure 3B). In contrast, the majority of neurodegeneration genes associated with non-Lewy neuropathology, atypical parkinsonism, or non-parkinsonian neurodegenerative phenotypes (Table S14) were distant from the vesicle trafficking network. A full analysis of the biological processes enriched in the network branches is provided in Table S12. Notably, this humanized network elucidated the molecular context in which the previously identified druggable targets NEDD4 (Tardiff et al., 2013) and Calcineurin (Caraveo et al., 2014) impact α -syn toxicity, and identified the transcription factor NFAT, through which Calcineurin exerts its toxicity (Caraveo et al., 2014), as a hub (Figure 3B).

Furthermore, both α -syn itself and LRRK2 were predicted as nodes, just as in the overexpression network (Figure 2A). In the

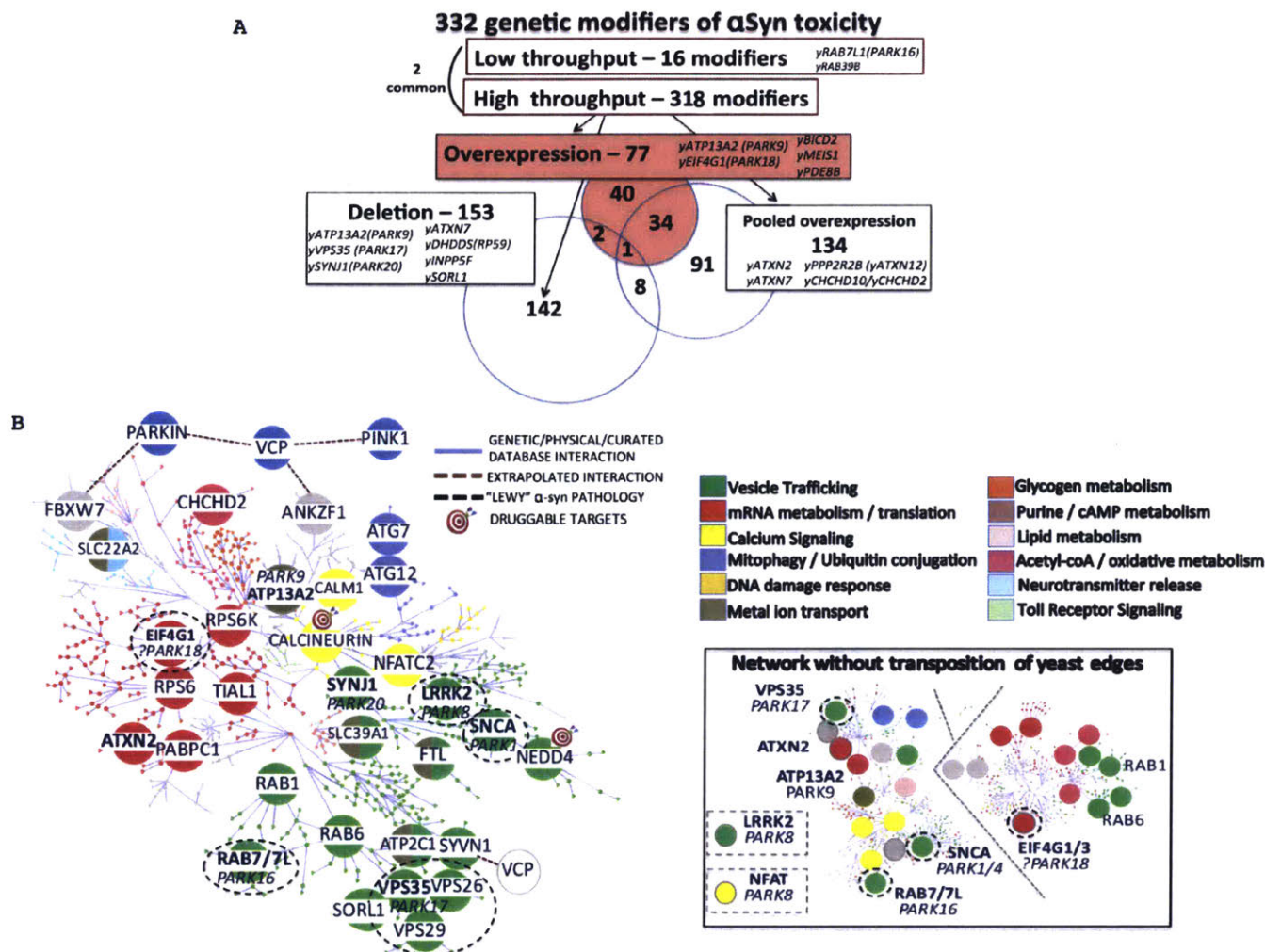


Figure 3. TransposeNet Builds Genome-Scale Molecular Network for α -Syn Toxicity from Genome-wide Deletion and Overexpression Yeast Screens

(A) Summary of genetic modifiers recovered in screens. 16 genetic modifiers (14 unique) from low-throughput investigations were also incorporated. Yeast homologs of genes linked to PD and other neurodegenerative disorders are listed. A “y” preceding the human gene name indicates “the yeast homolog of”. (B) A humanized network is generated from the 332 α -syn screen hits by TransposeNet. Genes of interest are enlarged, including multiple neurodegeneration-related disease genes (see also Figure S8 and Table S14). Gene ontology process enrichment within stems of the network are color-coded (full details in Table S12; gray portions were not enriched). Brown lines indicate extrapolated connections to VCP/Cdc48 through Vms1 (the yeast homolog of Ankzf1) and Hrd1 (the yeast homolog of Syv1) and from VCP to Parkin/PARK2 and Pink1/PARK6. A target symbol marks two druggable nodes, Calcineurin (Caraveo et al., 2014) and Nedd4 (a target of NAB; Tardiff et al., 2013). Inset: network without transposition of yeast edges. LRRK2 and NFAT are not included. Ontologically connected proteins (for example, Rab proteins) are dispersed.

ensemble of Steiner forests generated from our list of 332 modifiers, α -syn appeared in 100% and LRRK2 in 70%. In 1,000 random sets of 332 genes, even when we forced the incorporation of five $yPARK$ genes recovered in our genetic experiments ($yPARK9$, $yPARK16$, $yPARK17$, $yPARK18$, and $yPARK20$), α -syn and LRRK2 appeared together in only 0.6% of humanized networks. Thus, yeast modifiers of α -syn toxicity generated a specific humanized network in which the PD-associated proteins α -syn and LRRK2 emerged as critical network nodes.

TransposeNet generated a coherent network: 295 out of 332 yeast modifier genes were included in a single tree network (Table S10) with biologically intuitive “stems” comprising genes of similar ontology (Figure 3B). Networks generated from these 332 modi-

fiers without transposition of yeast interactome data produced three fragmented networks comprising 136, 2, and 122 yeast genes, respectively (Figure 3B, inset; Data S1). Genes that should be related biologically through involvement in common cellular processes were separated (Figure 3B). Moreover, LRRK2 and NFAT were not incorporated. Testable hypotheses, such as the relationship of EIF4G1 to ATXN2 (Figure 5), did not emerge because these proteins were not included in the same network. DAPPLE (Rossin et al., 2011) and PEXA (Tu et al., 2009) also produced highly fragmented or dense “hairball” networks of limited utility for hypothesis generation (Figure S9) and these algorithms did not incorporate critical nodes like LRRK2 (Figure S9). Thus, transposition of yeast networks to augment the human

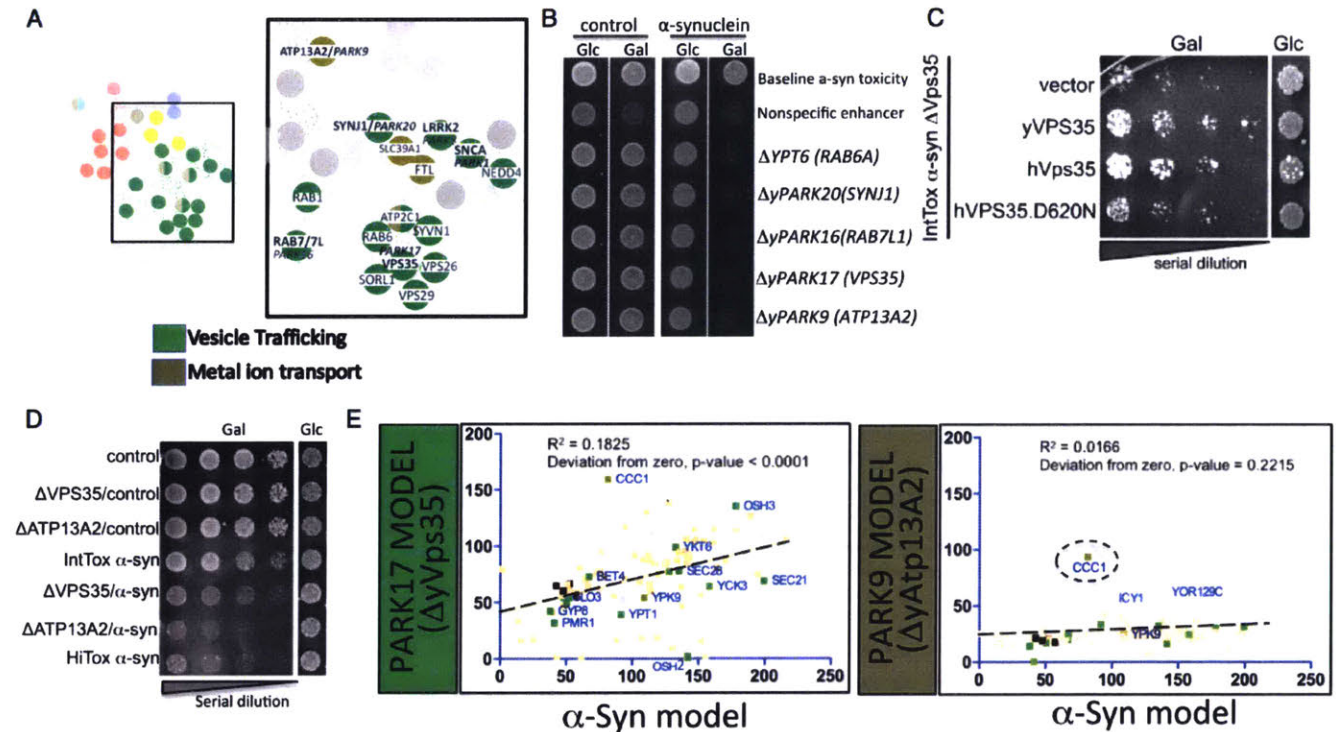


Figure 4. Genetic Dissection of Parkinsonism Susceptibility Genes Reveals Distinct Biology

(A) Vesicle trafficking subnetwork within the α -syn map (from Figure 3B) and location of *PARK9* (*ATP13A2*) outside of this subnetwork. Green, trafficking proteins; brown, metal ion transporters.

(B) Synthetic toxic interactions between trafficking genes and α -syn (spotting assays on agar plates). α -syn transgene is expressed from a galactose-inducible promoter (“on” in galactose [Gal]; “off” in glucose [Glc]). A “y” ahead of the human gene name indicates “the yeast homolog of”. Δ *GAL2* strain (“nonspecific enhancer”) serves as a (+) control because it grows less well on galactose (regardless of α -syn expression). The (□) control, a deletion (Δ *YMR191W*), has no deleterious effect in presence of α -syn (“baseline toxicity”).

(C) Expression of yeast *VPS35* (yVps35), human *VPS35* (hVps35), and human mutant (D620N) *VPS35* in α -syn-expressing *Vps35*-deleted “IntTox” cells (yeast spotting assay, showing serial 53 dilutions from left to right; transgenes are expressed from a galactose-inducible promoter).

(D and E) Cross-comparison of genetic interactors with similarly toxic HiTox α -syn, α -syn- Δ *VPS35*/*PARK17*, and α -syn- Δ *ATP13A2*/*PARK9* strains. Spotting assay demonstrates relative levels of toxicity among these three strains (D; 53 serial dilutions from left to right). In (E), data are shown on dot plots comparing the efficacy of 77 known α -syn modifiers (see Figure 1) in HiTox α -syn (x axis) versus Δ *PARK17*/ α -syn (y axis; left) Δ *PARK9*/ α -syn (y axis; right). Green, vesicle trafficking genetic modifiers; brown, metal ion transport modifiers. Axis scales represent growth relative to Mig1/Mig3-positive controls (100, black). Mig1/Mig3 overexpression represses the galactose promoter driving α -syn expression.

Each spot assay in this figure was repeated two to four times. The dot plot is representative of two experiments performed on separate days with biological replicates. Transformants were plated in quadruplicate for each experiment.

interactome created a coherent, biologically meaningful α -syn network.

An Endocytic and Retrograde Trafficking Subnetwork in the α -Syn Toxicity Map Incorporates Yeast Homologs of *RAB7L1* (*PARK16*) and *VPS35* (*PARK17*)

In the α -syn map, homologs of some parkinsonism genes coalesced in a sub-network around *YPT6*, the yeast homolog of *RAB6A* (Soper et al., 2011) (Figure 3B). Included were *YPT7*, *VTH1*, and *VPS35*, which encode proteins involved in endosomal trafficking. *YPT7* is a close homolog of *RAB7L1*, a leading candidate for the *PARK16* locus (MacLeod et al., 2013; Nalls et al., 2014), and also of the Mendelian parkinsonism risk factor *RAB39B* (Wilson et al., 2014). *VTH1* is a close yeast homolog of *SORL1*, an established AD risk modifier (Rogaeva et al., 2007) that encodes a protein involved in intracellular sorting (Nykjaer and Willnow, 2012). *VPS35* is homologous to the Mendelian risk factor for classic PD, *VPS35* (*PARK17*)

(Zimprich et al., 2011). *VPS35*, *VPS26*, and *VPS29* comprise the retromer complex that transports cargo from endosomal to Golgi compartments. In our accompanying study (Chung et al., 2017), we show that deletion of the *VSP26* and *VPS29* core retromer components strongly enhances α -syn toxicity. A fourth gene involved in Golgi-to-endosome and endocytic trafficking, *INP53*, is homologous to the Mendelian parkinsonism gene *SYNJ1* (*PARK20*) (Oligati et al., 2014). Deletion of any one of these genes was not toxic in a wild-type strain. However, deletion of any one of these genes in a strain expressing low (nontoxic) levels of α -syn produced a strong and synergistic growth defect (Table S5; Figures 4B and S10A). Importantly, ectopic expression of yeast or human *VPS35* rescued the toxicity induced by deleting *VPS35*, but expression of a disease-causing mutation (*VPS35-D620N*) did not (Figure 4C). Finally, y*RAB7L1* enhanced α -syn toxicity when deleted but rescued from this toxicity when overexpressed (Figure S10B).

The α -Syn Map Predicts Diverging Genetic Interaction Profiles in Δ PARK9 (ATP13A2)- and Δ PARK17 (VPS35)-Sensitized Yeast Models

To test functional consequences of being located in distinct subnetworks of our α -syn map, we compared VPS35 (PARK17) and ATP13A2 (PARK9). ATP13A2 is a type 5 P-ATPase implicated in cation transport and metal ion homeostasis (Kong et al., 2014; Park et al., 2014; Ramonet et al., 2012; Tsunemi and Krainc, 2014). Mutations in ATP13A2 lead to juvenile-onset parkinsonism or Kufor-Rakeb syndrome, which is distinct from PD (Schneider et al., 2010). *yATP13A2* suppressed α -syn toxicity in our overexpression screen (Figure 1B), and deletion of *yATP13A2* strongly enhanced α -syn toxicity (Figure 4B). In our humanized network, ATP13A2 was spatially distant from VPS35, lying well outside the vesicle trafficking subnetwork (Figures 3B and 4A). We asked whether this spatial separation reflected differences in underlying biology.

We generated three strains with similar toxicities (Figure 4D). In one strain, toxicity resulted from overexpression of α -syn (HiTox). In two other strains, mild toxicity induced by intermediate levels of α -syn expression was enhanced by deletion of yeast ATP13A2 (hereafter, Δ ATP13A2/ α -syn) or VPS35 (hereafter, Δ VPS35/ α -syn). These three yeast strains thus modeled cellular pathologies related to three forms of familial parkinsonism: two with typical α -syn pathology (PD related to α -syn multiplication and VPS35 (PARK17)-associated parkinsonism) and one with strikingly different pathology (PARK9 (ATP13A2)).

While Δ ATP13A2 is known to sensitize cells to metal ion stress (Kong et al., 2014), Δ VPS35 strains exhibit retrograde trafficking defects (Seaman et al., 1997), suggesting that Δ ATP13A2 and Δ VPS35 strains are differentially sensitized to α -syn toxicity. We asked whether our 77 α -syn overexpression screen hits differentially modified the toxicity of our Δ VPS35/ α -syn and Δ ATP13A2/ α -syn models.

We expressed these α -syn toxicity modifiers in each of the yeast models and monitored growth. For the α -syn HiTox and Δ VPS35/ α -syn models, 69 out of 77 genes overlapped (Figure 4E, left), correlating well with the similar pathology associated with these genetic forms of parkinsonism. Notably, the overlapping modifiers were enriched in vesicle trafficking genes (Table S13). In contrast, there were only 3 out of 77 modifiers in common between α -syn HiTox and Δ ATP13A2/ α -syn (Figure 4E, right). These were involved in iron and manganese homeostasis (CCC1) and actin cytoskeleton rearrangements (ICY1 and AFI1 (YOR129C)), respectively. Notably, metal ion homeostasis is strongly implicated in Kufor-Rakeb syndrome (Schneider et al., 2010) and its mammalian models (Park et al., 2014). Thus, neurodegenerative diseases that are genetically, clinically, and neuropathologically distinct may nonetheless share some key molecular pathologies that can be uncovered through genetic network analysis.

mRNA Translation Subnetwork Links α -Syn to PABPC1, EIF4G1, and ATXN2

In our overexpression screen against α -syn toxicity, TIF4632 (hereafter *yEIF4G1-2*) emerged as a suppressor. *yEIF4G1-2* is a yeast homolog of the translational initiation factor EIF4G1. The genome-wide deletion and pooled overexpression screens identi-

fied additional genetic modifiers related to mRNA translation, including initiation factors and multiple ribosomal subunits (Figures 3B and 5A; Tables S5 and S6). These included the homolog of human PABPC1, PAB1, the ATXN2 homolog PBP1, and the second EIF4G family homolog in yeast, TIF4631 (hereafter *yEIF4G1-1*). Enrichment of these hits in the pooled screen was confirmed by qPCR (Figure 5B, left), and overexpression of these genes suppressed α -syn toxicity in bioscreen (Figure 5B, right) and/or spot (Figure S11) growth assays. Genetic experiments in different proteinopathy models revealed that the effects of these modifiers on α -syn toxicity were specific (Figure S11). Thus, perturbation of mRNA translation in synucleinopathy was not simply a generic proteotoxic response.

Protein Translation Is Perturbed in PD-Patient-Derived Neurons

Because we recovered numerous genetic modifiers in the mRNA translation and mRNA processing pathways (Figures 3 and 5), we asked whether protein translation was perturbed in cellular synucleinopathy models, including PD-patient-derived neurons. Bulk changes in protein production were assessed by determining the rate at which S³⁵-radiolabeled methionine and cysteine are incorporated into protein over time after a brief "pulse labeling." Overexpression of α -syn in HEK cells and primary rat cortical neurons reduced the accumulation of S³⁵-Met/Cys without changing amino acid uptake (Figure S12). Similarly, in 6- to 8-week-old iPSC neurons harboring the α -syn^{A53T} mutation, S³⁵-Met/Cys incorporation into protein was reduced compared to isogenic mutation-corrected controls (Figure 5C). Thus, our α -syn screens and network analysis identified a strong effect of α -syn toxicity on bulk mRNA translation in cellular models of synucleinopathy. This effect was not attributable to a canonical ER stress response, because phosphorylation of EIF2A (Figures S12D and S13A) and XBP1 splicing (Figure S13B) was not altered in these neurons.

Conserved Genetic Interactions of ATXN2 and EIF4G1 from Yeast to Patient Neurons

We next tested whether human homologs of two translation factors that suppressed α -syn toxicity when overexpressed in yeast, ATXN2 and EIF4G1, could similarly reverse the protein translation defect in neurons. We generated transcription activator-like effector transcription factor (TALE-TF) constructs to transcriptionally upregulate neuronal isoforms of EIF4G1 and ATXN2 (Sanjana et al., 2012) (Figure 5D, left). These constructs were then delivered with an adeno-associated viral vector to differentiated α -syn^{A53T} iPSC-derived neuronal cultures.

10 days after transduction, endogenous EIF4G1 and ATXN2 mRNA levels increased by 0.4-fold, as measured by qPCR (Figure 5D, right). This increased expression substantially reversed the defect in bulk protein translation we had observed in α -syn^{A53T} neurons (Figure 5E). Overexpression of EIF4G1 increased translation in A53T neurons, but not in mutation-corrected controls. ATXN2 overexpression equally increased bulk translation in mutant and control cells (Figure 5E).

Thus, our cross-species molecular network predicted a biological interaction between α -syn and mRNA translation factors in PD-patient-derived neurons. These data strengthen the argument that perturbed mRNA translation is an important aspect of α -syn toxicity. Interestingly, we identified a strong signature of

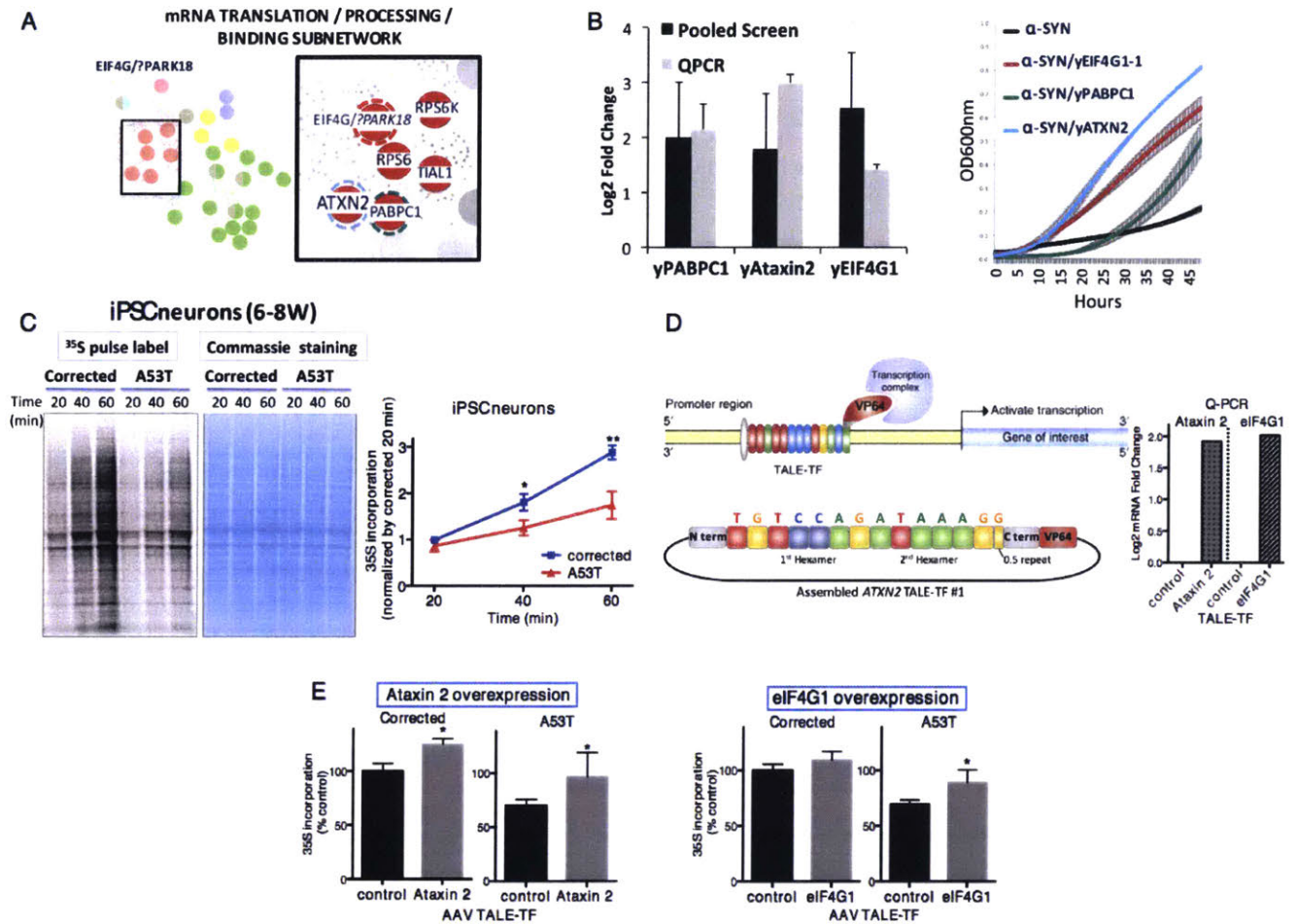


Figure 5. mRNA Translation Factors Impact α -Syn Toxicity from Yeast to Patient-Derived Neurons

(A) mRNA translation subnetwork in α -syn toxicity (from Figure 3B), including *ATXN2*, *EIF4G1* (*PARK18*), and *PABPC1*.
 (B) Effects of yPABPC1, yAtaxin2, and yEIF4G1-1 on α -syn toxicity (left: qPCR to verify enrichment of these genes in pooled over-expression screen; right: bioscreen growth assay). A “y” ahead of the human gene name indicates “yeast homolog of”.
 (C) Bulk mRNA translation in mutant α -syn^{A53T} iPS neurons compared to isogenic mutation-corrected control neurons as measured by ³⁵S-cysteine and ³⁵S-methionine incorporation over time (phosphorimager scan). Coomassie staining shows loading of protein. At right, the graph shows two subclones of the mutation-corrected line compared to α -syn^{A53T} cells (n = 4; *p < 0.05; **p < 0.01 (two-tailed t test)).
 (D) TALE-TFs were designed to elevate the endogenous levels of *ATXN2* or *EIF4G1* genes. These bind to the 5′ UTR of the target genes and recruit a transcriptional activator (Sanjana et al., 2012). qPCR indicates transcript levels after AAV-mediated TALE-TF delivery into A53T iPS neurons.
 (E) Effect of increasing endogenous *EIF4G1* or *ATXN2* levels on bulk translation in A53T neurons (n = 3).
 Data represent mean ± SEM. *p < 0.05 (two-tailed t test).

decreased translation of mRNA translation-related transcripts in ribosomal footprinting experiments of α -syn^{A53T} iPS-derived cortical neurons at 4 and 12 weeks of differentiation compared to isogenic mutation-corrected control neurons (Figures S14 and S15; Tables S16, S17, S18, and S19). Indeed, mRNA-related translation factors, ribonucleoproteins, and ribosomal proteins were enriched not only in our genetic and translational maps of α -syn toxicity but also among proteins in the immediate vicinity of α -syn in neurons (Figure 4 in accompanying manuscript, Chung et al., 2017; Figure S14A, “spatial α -syn map”). Moreover, a number of mRNA translation proteins directly complexed with α -syn. This convergence of genetic, translational, and spatial maps suggests the connection between α -syn toxicity and mRNA metabolism is deeply rooted in α -syn biology.

DISCUSSION

We describe a coherent, systems-level analysis of how α -syn misfolding and mistrafficking perturbs cell biology. Genome-wide screens identified modifiers of the toxic consequences of α -syn expression in baker’s yeast (*S. cerevisiae*). Our key computational contribution, TransposeNet, coupled richly annotated molecular interactions in yeast with a Steiner prize-collecting algorithm and a sophisticated cross-species homology tool to visualize the screen hits as a humanized molecular network. TransposeNet revealed that α -syn pathology is deeply connected to human genetic risk factors for parkinsonism and parsed out the molecular pathways through which these connections occur. We envisage TransposeNet as a valuable resource for the community, easily

generalizable to the modeling of any physiologic or pathologic process in genetically tractable organisms.

A pressing challenge in neurodegeneration is to determine whether genes associated with highly distinct pathologies, but that nevertheless converge on similar clinical phenotypes, are related at a molecular level or not. Our network tied α -syn not only to genes that cause classical PD (Ogaki et al., 2015) but also to genes that cause parkinsonism with different pathologies and genes associated with other neurodegenerative phenotypes altogether (Table S14). The relationships were highly specific to α -syn. Moreover, genes tied to classical PD or α -syn pathology (like RAB7L1, VPS35, and LRRK2) were concentrated in a vesicle-trafficking-associated subnetwork, while genes tied to “atypical” parkinsonism (like ATP13A2 and ATXN2) were in separate subnetworks. For a few examples, we experimentally validated the convergent and divergent molecular pathologies indicated by the spatial location on the map. Thus, LRRK2 and α -syn pathologies were connected at the level of perturbed protein trafficking, which was confirmed in patient-derived neurons. In another example, VPS35 and ATP13A2 exhibited highly distinct genetic modifier profiles in yeast. Other network and model-organism studies provide important support for our results, including connections between α -syn and LRRK2 (Cho et al., 2014; Liu et al., 2012), RAB7L1 and LRRK2 (MacLeod et al., 2013; Beilina et al., 2014), and VPS35 and α -syn (Dhungel et al., 2015).

For some genes connected to α -syn toxicity by our network, including *EIF4G1* (*PARK18*) and *CHCHD2*, human genetic data are limited or controversial (Funayama et al., 2015; Liu et al., 2015; Ogaki et al., 2015; Chartier-Harlin et al., 2011; Nuytemans et al., 2013). Another gene, *RAB7L1*, is one of two candidates in linkage with a parkinsonism-associated common gene variant (*PARK16*). Our analysis affirms a connection between such genes and α -syn proteinopathy and provides the biological context in which to place these interactions. For example, we make no claim that the translation factor *EIF4G1* should be designated a “PD gene.” However, *EIF4G1* and α -syn toxicity are connected in the context of an important and previously unrecognized direct effect of α -syn on mRNA biology and protein translation. This was confirmed by multiple hits in our genetic analysis (Figure 3) and in our mRNA translational profiling of neurons (Figure S14). Moreover, a connection to mRNA translation and metabolism was also confirmed in our spatial mapping of α -syn in neurons (Chung et al., 2017). This map revealed that α -syn is in the immediate vicinity and complexed to proteins involved in mRNA translation and protein trafficking, suggesting that these perturbations may be upstream or proximal events in α -syn toxicity. Interestingly, a connection is emerging between mRNA translation and other parkinsonism-related genes (Gehrke et al., 2015; Martin et al., 2014).

Finally, by identifying connections between druggable targets and gene networks, our approach provides a glimpse of how future treatments might be targeted to specific genetic lesions. We envisage that the inflexibility of a single clinical or pathologic diagnosis will yield to a more nuanced molecular diagnosis. In this scenario, genetic lesions will be matched to compound targets and confirmed in “personalized” cellular models in which combinatorial genetic lesions are introduced to reflect specific genetic risk and biology. Emerging genome-editing technologies will enable such models to be developed in patient-derived

cells, and for genome-wide screening to be carried out in these cells also (Hasson et al., 2013; Khurana et al., 2015; Shalem et al., 2014; Wang et al., 2014). These will unquestionably be welcome advances, but impressive developments will continue in simple model organisms. Variomic libraries in yeast, for example, now enable mutagenesis at single-amino-acid resolution across the entire yeast proteome (Huang et al., 2013), unlocking enormous potential for target identification in phenotypic screens. We envision multi-faceted, cross-species approaches will continue to reveal critical insights into many complex diseases and perhaps fulfill therapeutic promises in the post-genomics era.

STAR+ METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
 - B Yeast Strains
 - B Human iPSC Lines
 - B Human iPSC Generation and Differentiation into Midbrain Dopaminergic (DA) Neurons for LRRK2 Mutant Lines
 - B Human Pluripotent Stem Cell Culture for α -Syn Mutant Lines
 - B Primary Rat Cortical Cultures
- METHOD DETAILS
 - B Yeast-to-Human Homology
 - B Preprocessing of Interactomes
 - B Augmentation of Human Interactome with Yeast-to-Yeast Edges (for Humanized Networks Only)
 - B Prize-Collecting Steiner Forest Algorithm
 - B Node and Edge Setup for Yeast and Humanized Steiner Networks
 - B Parameter Tuning for Computational Pipelines
 - B Spotting Assays
 - B Screening against Known α -Syn Modifiers in Δ PARK17/ α -Syn and Δ PARK9/ α -Syn Strains
 - B Small Molecule (NAB2) Treatment
 - B Pooled α -Syn Overexpression Screen
 - B Pooled Screen-QPCR Verification
 - B Pooled Screen-Growth Curve Analysis
 - B Genome-Wide Deletion Screen (Synthetic Gene Array Methodology)
 - B Human iPSC Generation and Differentiation into Midbrain Dopaminergic (DA) Neurons for LRRK2 Mutant Lines
 - B Human Pluripotent Stem Cell Culture for α -Syn Mutant Lines
 - B Human Neural Induction by Embryoid Body (EB) Formation
 - B Human Cortical Neural Differentiation
 - B Primary Rat Cortical Cultures
 - B AAV-1 Transduction of iPS Neurons
 - B Antibodies
 - B Protein Labeling with 35 S-Methionine/-Cysteine

- B Free ³⁵S-Methionine/-Cysteine in the Cytosol
- B Cell Lysis and Endoglycosidase H Digestion
- B Western Blotting
- B TALE-TF Design
- B TALE-TF Assembly
- B Ribosomal Footprint Profiling
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- B Comparison with Existing Homology Prediction Approaches
- B Evaluation of PCSF and Humanized Steiner Networks
- B Statistical Methods and Data Analysis for Cell-Based Assays
- **DATA AND SOFTWARE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information includes 15 figures, 19 tables, and 1 data file and can be found with this article online at <http://dx.doi.org/10.1016/j.cels.2016.12.011>.

AUTHOR CONTRIBUTIONS

The study was designed and supervised by V.K. and S.L. TransposeNet was conceived by V.K. and executed by J.P., with key contributions by M.K. and N.T., B.B., and E.F. and supervision by B.B. and E.F. C.Y.C. led the neuronal assay development efforts. P.K.A., S.F., D.F.T., T.B., S.W.E., H.B., Y.L., A.N.-U., V.B., Y.F., M.C., B.-J.S.L., D.C.S., M.I.B., S.E., N.S., Q.Z., T.G., D.P.B., M.V., M.D., and C.B. performed bench or computational experiments or supervised these efforts. The paper was written by V.K., C.Y.C., J.P., B.B., E.F., and S.L.

ACKNOWLEDGMENTS

Research was supported by an HHMI Collaborative Innovation Award (V.K., C.Y.C., and S.L.), the JPB Foundation (V.K., C.Y.C., and S.L.), an American Brain Foundation and Parkinson's Disease Foundation Clinician-Scientist Development Award (V.K.), the Multiple System Atrophy Coalition (V.K.), the NIH/NIA (grant K01 AG038546, C.Y.C.), the Eleanor Schwartz Charitable Foundation (S.L.), the NIH (grants R01 NS 089076 and U54 NS 091046, E.F.; R01GM081871, B.B.; HG006061, M.V. and S.L.; and P50HG004233 and U01HG001715, M.V.). We thank Drs. Prathapan Thiru and George Bell, Bioinformatics and Research Computing (BaRC; Whitehead Institute) for expert assistance; Kristina Fischer for important technical assistance; Drs. Luke Whitesell, Philip de Jager, Mel Feany, Aftabul Haque, David Hill, Priyanka Narayan, and Linda Clayton for stimulating discussion and critical comments on the manuscript; and Dr. Julie Valastyan for yeast strain construction. We dedicate this manuscript to Dr Susan Lindquist, our incomparable mentor and colleague, who passed away while this manuscript was in its final stages of revision.

V.K., C.Y.C., D.F.T., and S.L. are scientific co-founders of Yumanity Therapeutics, a company developing neurodegenerative disease therapeutics.

Received: January 12, 2016

Revised: August 5, 2016

Accepted: December 14, 2016

Published: January 25, 2017

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.

Bailly-Bechet, M., Borgs, C., Braunstein, A., Chayes, J., Dagkessamanskaia, A., François, J.-M., and Zecchina, R. (2011). Finding undetected protein associations in cell signaling by belief propagation. *Proc. Natl. Acad. Sci. USA* 108, 882–887.

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S., et al. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* 7, 1017–1024.

Beilina, A., Rudenko, I.N., Kaganovich, A., Civiero, L., Chau, H., Kalia, S.K., Kalia, L.V., Lobbstaël, E., Chia, R., Ndukwe, K., et al.; International Parkinson's Disease Genomics Consortium; North American Brain Expression Consortium (2014). Unbiased screen for interactors of leucine-rich repeat kinase 2 supports a common pathway for sporadic and familial Parkinson disease. *Proc. Natl. Acad. Sci. USA* 111, 2626–2631.

Berger, B., Peng, J., and Singh, M. (2013). Computational solutions for omics data. *Nat. Rev. Genet.* 14, 333–346.

Brás, J., Guerreiro, R., and Hardy, J. (2015). SnapShot: Genetics of Parkinson's disease. *Cell* 160, 570–570.e1.

Caraveo, G., Auluck, P.K., Whitesell, L., Chung, C.Y., Baru, V., Mosharov, E.V., Yan, X., Ben-Johny, M., Soste, M., Picotti, P., et al. (2014). Calcineurin determines toxic versus beneficial responses to α -synuclein. *Proc. Natl. Acad. Sci. USA* 111, E3544–E3552.

Casals, F., and Bertranpetit, J. (2012). Genetics. Human genetic variation, shared and private. *Science* 337, 39–40.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* 27, 275–280.

Chartier-Harlin, M.-C., Dachsel, J.C., Vilarinho-Güell, C., Lincoln, S.J., Leprêtre, F., Hulihan, M.M., Kachergus, J., Milnerwood, A.J., Tapia, L., Song, M.-S., et al. (2011). Translation initiator EIF4G1 mutations in familial Parkinson disease. *Am. J. Hum. Genet.* 89, 398–406.

Cho, H.J., Yu, J., Xie, C., Rudrabhatla, P., Chen, X., Wu, J., Parisiadou, L., Liu, G., Sun, L., Ma, B., et al. (2014). Leucine-rich repeat kinase 2 regulates Sec16A at ER exit sites to allow ER-Golgi export. *EMBO J.* 33, 2314–2331.

Cho, H., Peng, J., and Berger, B. (2015). Diffusion Component Analysis: Unraveling Functional Topology in Biological Networks (RECOMB).

Cho, H., Berger, B., and Peng, J. (2016). Compact integration of multi-network topology for functional analysis of genes. *Cell Syst.* Published online November 22, 2016. <http://dx.doi.org/10.1016/j.cels.2016.10.017>.

Chung, C.Y., Khurana, V., Auluck, P.K., Tardiff, D.F., Mazzulli, J.R., Soldner, F., Baru, V., Lou, Y., Freyzon, Y., Cho, S., et al. (2013). Identification and rescue of α -synuclein toxicity in Parkinson patient-derived neurons. *Science* 342, 983–987.

Chung, C.Y., Khurana, V., Yi, S., Sahni, N., Loh, K.H., Auluck, P.K., Baru, V., Udeshi, N.D., Freyzon, Y., Carr, S.A., et al. (2017). In situ peroxidase labeling and mass-spectrometry connects alpha-synuclein directly to endocytic trafficking and mRNA metabolism in neurons. *Cell Syst.* 4, 242–250.

Cooper, A.A., Gitler, A.D., Cashikar, A., Haynes, C.M., Hill, K.J., Bhullar, B., Liu, K., Xu, K., Strathern, K.E., Liu, F., et al. (2006). Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. *Science* 313, 324–328.

Dhungel, N., Eleuteri, S., Li, L.-B., Kramer, N.J., Chartron, J.W., Spencer, B., Kosberg, K., Fields, J.A., Stafa, K., Adame, A., et al. (2015). Parkinson's disease genes VPS35 and EIF4G1 interact genetically and converge on α -synuclein. *Neuron* 85, 76–87.

Eiden, A.C., Kim, H.-J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armarkola, M., Geser, F., Greene, R., Lu, M.M., et al. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 466, 1069–1075.

- Fuchs, J., Tichopad, A., Golub, Y., Munz, M., Schweitzer, K.J., Wolf, B., Berg, D., Mueller, J.C., and Gasser, T. (2008). Genetic variability in the SNCA gene influences alpha-synuclein levels in the blood and brain. *FASEB J.* **22**, 1327–1334.
- Funayama, M., Ohe, K., Arno, T., Furuya, N., Yamaguchi, J., Saiki, S., Li, Y., Ogaki, K., Ando, M., Yoshino, H., et al. (2015). CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. *Lancet Neurol.* **14**, 274–282.
- Geetha, V., Di Francesco, V., Garnier, J., and Munson, P.J. (1999). Comparing protein sequence-based and predicted secondary structure-based methods for identification of remote homologs. *Protein Eng.* **12**, 527–534.
- Gehrke, S., Wu, Z., Klinkenberg, M., Sun, Y., Auburger, G., Guo, S., and Lu, B. (2015). PINK1 and Parkin control localized translation of respiratory chain component mRNAs on mitochondria outer membrane. *Cell Metab.* **21**, 95–108.
- Gietz, D., St Jean, A., Woods, R.A., and Schiestl, R.H. (1992). Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.* **20**, 1425.
- Gietz, R.D., Schiestl, R.H., Willems, A.R., and Woods, R.A. (1995). Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* **11**, 355–360.
- GTE Consortium (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585.
- Guerreiro, R., Brás, J., and Hardy, J. (2015). SnapShot: Genetics of ALS and FTD. *Cell* **160**, 798.e1.
- Hasson, S.A., Kane, L.A., Yamano, K., Huang, C.-H., Sliter, D.A., Buehler, E., Wang, C., Heman-Ackah, S.M., Hessa, T., Guha, R., et al. (2013). High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. *Nature* **504**, 291–295.
- Hu, Y., Rolfs, A., Bhullar, B., Murthy, T.V.S., Zhu, C., Berger, M.F., Camargo, A.A., Kelley, F., McCarron, S., Jepson, D., et al. (2007). Approaching a complete repository of sequence-verified protein-encoding clones for *Saccharomyces cerevisiae*. *Genome Res.* **17**, 536–543.
- Hu, Y., Flockhart, I., Vinayagam, A., Bergwitz, C., Berger, B., Perrimon, N., and Mohr, S.E. (2011). An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12**, 357.
- Huang, S.-S.C., and Fraenkel, E. (2009). Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Sci. Signal.* **2**, ra40.
- Huang, Z., Chen, K., Zhang, J., Li, Y., Wang, H., Cui, D., Tang, J., Liu, Y., Shi, X., Li, W., et al. (2013). A functional variomics tool for discovering drug-resistance genes and drug targets. *Cell Rep.* **3**, 577–585.
- Kachroo, A.H., Laurent, J.M., Yellman, C.M., Meyer, A.G., Wilke, C.O., and Marcotte, E.M. (2015). Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* **348**, 921–925.
- Khurana, V., and Lindquist, S. (2010). Modelling neurodegeneration in *Saccharomyces cerevisiae*: why cook with baker's yeast? *Nat. Rev. Neurosci.* **11**, 436–449.
- Khurana, V., Tardiff, D.F., Chung, C.Y., and Lindquist, S. (2015). Toward stem cell-based phenotypic screens for neurodegenerative diseases. *Nat. Rev. Neurol.* **11**, 339–350.
- Kim, J.-E., O'Sullivan, M.L., Sanchez, C.A., Hwang, M., Israel, M.A., Brennand, K., Deerinck, T.J., Goldstein, L.S.B., Gage, F.H., Ellisman, M.H., and Ghosh, A. (2011). Investigating synapse formation and function using human pluripotent stem cell-derived neurons. *Proc. Natl. Acad. Sci. USA* **108**, 3005–3010.
- Kim, H.-J., Raphael, A.R., LaDow, E.S., McGurk, L., Weber, R.A., Trojanowski, J.Q., Lee, V.M.-Y., Finkbeiner, S., Gitler, A.D., and Bonini, N.M. (2014). Therapeutic modulation of eIF2 α phosphorylation rescues TDP-43 toxicity in amyotrophic lateral sclerosis disease models. *Nat. Genet.* **46**, 152–160.
- Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588.
- Kong, S.M.Y., Chan, B.K.K., Park, J.-S., Hill, K.J., Aitken, J.B., Cottle, L., Farghaian, H., Cole, A.R., Lay, P.A., Sue, C.M., and Cooper, A.A. (2014). Parkinson's disease-linked human PARK9/ATP13A2 maintains zinc homeostasis and promotes α -Synuclein externalization via exosomes. *Hum. Mol. Genet.* **23**, 2816–2833.
- Kriks, S., Shim, J.-W., Piao, J., Ganat, Y.M., Wakeman, D.R., Xie, Z., Carrillo-Reid, L., Auyeung, G., Antonacci, C., Buch, A., et al. (2011). Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. *Nature* **480**, 547–551.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.
- Liu, G., Aliaga, L., and Cai, H. (2012). α -synuclein, LRRK2 and their interplay in Parkinson's disease. *Future Neurol.* **7**, 145–153.
- Liu, Z., Guo, J., Li, K., Qin, L., Kang, J., Shu, L., Zhang, Y., Wei, Y., Yang, N., Luo, Y., et al. (2015). Mutation analysis of CHCHD2 gene in Chinese familial Parkinson's disease. *Neurobiol. Aging* **36**, 3117.e7–3117.e8.
- MacLeod, D.A., Rhinn, H., Kuwahara, T., Zolin, A., Di Paolo, G., McCabe, B.D., Marder, K.S., Honig, L.S., Clark, L.N., Small, S.A., and Abeliovich, A. (2013). RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron* **77**, 425–439.
- Martin, I., Dawson, V.L., and Dawson, T.M. (2011). Recent advances in the genetics of Parkinson's disease. *Annu. Rev. Genomics Hum. Genet.* **12**, 301–325.
- Martin, I., Kim, J.W., Lee, B.D., Kang, H.C., Xu, J.-C., Jia, H., Stankowski, J., Kim, M.-S., Zhong, J., Kumar, M., et al. (2014). Ribosomal protein s15 phosphorylation mediates LRRK2 neurodegeneration in Parkinson's disease. *Cell* **157**, 472–485.
- Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., et al.; International Parkinson's Disease Genomics Consortium (IPDGC); Parkinson's Study Group (PSG) Parkinson's Research: The Organized GENetics Initiative (PROGENI); 23andMe; GenePD; NeuroGenetics Research Consortium (NGRC); Hussman Institute of Human Genomics (HIHG); Ashkenazi Jewish Dataset Investigator; Cohorts for Health and Aging Research in Genetic Epidemiology (CHARGE); North American Brain Expression Consortium (NABEC); United Kingdom Brain Expression Consortium (UKBEC); Greek Parkinson's Disease Consortium; Alzheimer Genetic Analysis Group (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993.
- Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.-A., Fraser, D., et al. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100–104.
- Nuytemans, K., Bademci, G., Inchausti, V., Dressen, A., Kinnamon, D.D., Mehta, A., Wang, L., Züchner, S., Beecham, G.W., Martin, E.R., et al. (2013). Whole exome sequencing of rare variants in EIF4G1 and VPS35 in Parkinson disease. *Neurology* **80**, 982–989.
- Nykjaer, A., and Willnow, T.E. (2012). Sortilin: a receptor to regulate neuronal viability and function. *Trends Neurosci.* **35**, 261–270.
- Ogaki, K., Koga, S., Heckman, M.G., Fiesel, F.C., Ando, M., Labbé, C., Lorenzo-Betancor, O., Moussaud-Lamodière, E.L., Soto-Ortolaza, A.I., Walton, R.L., et al. (2015). Mitochondrial targeting sequence variants of the CHCHD2 gene are a risk for Lewy body disorders. *Neurology* **85**, 2016–2025.
- Olgiati, S., De Rosa, A., Quadri, M., Crisculo, C., Breedveld, G.J., Picillo, M., Pappatà, S., Quarantelli, M., Barone, P., De Michele, G., and Bonifati, V. (2014). PARK2 caused by SYNJ1 homozygous Arg258Gln mutation in a new Italian family. *Neurogenetics* **15**, 183–188.
- Osborn, M.J., and Miller, J.R. (2007). Rescuing yeast mutants with human genes. *Brief. Funct. Genomics Proteomics* **6**, 104–111.
- Outeiro, T.F., and Lindquist, S. (2003). Yeast cells provide insight into alpha-synuclein biology and pathobiology. *Science* **302**, 1772–1775.
- Park, J.-S., Koentjoro, B., Veivers, D., Mackay-Sim, A., and Sue, C.M. (2014). Parkinson's disease-associated human ATP13A2 (PARK9) deficiency causes zinc dyshomeostasis and mitochondrial dysfunction. *Hum. Mol. Genet.* **23**, 2802–2815.

- Rajput, A., Dickson, D.W., Robinson, C.A., Ross, O.A., Dächsel, J.C., Lincoln, S.J., Cobb, S.A., Rajput, M.L., and Farrer, M.J. (2006). Parkinsonism, Lrrk2 G2019S, and tau neuropathology. *Neurology* 67, 1506–1508.
- Ramonet, D., Podhajski, A., Stafa, K., Sonnay, S., Trancikova, A., Tsika, E., Pletnikova, O., Troncoso, J.C., Glauser, L., and Moore, D.J. (2012). PARK9-associated ATP13A2 localizes to intracellular acidic vesicles and regulates cation homeostasis and neuronal integrity. *Hum. Mol. Genet.* 21, 1725–1743.
- Reinhardt, P., Schmid, B., Burbulla, L.F., Schöndorf, D.C., Wagner, L., Glatza, M., Höing, S., Hargus, G., Heck, S.A., Dhingra, A., et al. (2013). Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. *Cell Stem Cell* 12, 354–367.
- Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11, R25.
- Rogaeva, E., Meng, Y., Lee, J.H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C.T., Cheng, R., Hasegawa, H., et al. (2007). The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat. Genet.* 39, 168–172.
- Rolland, T., Taşan, M., Charlotiaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., et al. (2014). A proteome-scale map of the human interactome network. *Cell* 159, 1212–1226.
- Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., Cotsapas, C., and Daly, M.J.; International Inflammatory Bowel Disease Genetics Consortium (2011). Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* 7, e1001273.
- Sanjana, N.E., Cong, L., Zhou, Y., Cunniff, M.M., Feng, G., and Zhang, F. (2012). A transcription activator-like effector toolbox for genome engineering. *Nat. Protoc.* 7, 171–192.
- Schneider, S.A., Paisan-Ruiz, C., Quinn, N.P., Lees, A.J., Houlden, H., Hardy, J., and Bhatia, K.P. (2010). ATP13A2 mutations (PARK9) cause neurodegeneration with brain iron accumulation. *Mov. Disord.* 25, 979–984.
- Schöndorf, D.C., Aureli, M., McAllister, F.E., Hindley, C.J., Mayer, F., Schmid, B., Sardi, S.P., Valsecchi, M., Hoffmann, S., Schwarz, L.K., et al. (2014). iPSC-derived neurons from GBA1-associated Parkinson's disease patients show autophagic defects and impaired calcium homeostasis. *Nat. Commun.* 5, 4028.
- Seaman, M.N., Marcusson, E.G., Cereghino, J.L., and Emr, S.D. (1997). Endosome to Golgi retrieval of the vacuolar protein sorting receptor, Vps10p, requires the function of the VPS29, VPS30, and VPS35 gene products. *J. Cell Biol.* 137, 79–92.
- Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87.
- Shulman, J.M., De Jager, P.L., and Feany, M.B. (2011). Parkinson's disease: genetics and pathogenesis. *Annu. Rev. Pathol.* 6, 193–222.
- Singh, R., Xu, J., and Berger, B. (2008). Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. Natl. Acad. Sci. USA* 105, 12763–12768.
- Small, S.A., Kent, K., Pierce, A., Leung, C., Kang, M.S., Okada, H., Honig, L., Vonsattel, J.-P., and Kim, T.-W. (2005). Model-guided microarray implicates the retromer complex in Alzheimer's disease. *Ann. Neurol.* 58, 909–919.
- Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–8.
- Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., et al. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* 146, 318–331.
- Soper, J.H., Kehm, V., Burd, C.G., Bankaitis, V.A., and Lee, V.M.-Y. (2011). Aggregation of α -synuclein in *S. cerevisiae* is associated with defects in endosomal trafficking and phospholipid biosynthesis. *J. Mol. Neurosci.* 43, 391–405.
- 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, 66–71.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452.
- Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. *Genome Res.* 21, 2213–2223.
- Tardiff, D.F., Jui, N.T., Khurana, V., Tambe, M.A., Thompson, M.L., Chung, C.Y., Kamadurai, H.B., Kim, H.T., Lancaster, A.K., Caldwell, K.A., et al. (2013). Yeast reveal a “druggable” Rsp5/Nedd4 network of rare coding α -synuclein toxicity in neurons. *Science* 342, 979–983.
- Tardiff, D.F., Khurana, V., Chung, C.Y., and Lindquist, S. (2014). From yeast to patient neurons and back again: powerful new discovery platform. *Mov. Disord.* 29, 1231–1240.
- Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337, 64–69.
- Tong, A.H.Y., and Boone, C. (2006). Synthetic genetic array analysis in *Saccharomyces cerevisiae*. *Methods Mol. Biol.* 313, 171–192.
- Tong, A.H.Y., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., et al. (2004). Global mapping of the yeast genetic interaction network. *Science* 303, 808–813.
- Treusch, S., Hamamichi, S., Goodman, J.L., Matlack, K.E.S., Chung, C.Y., Baru, V., Shulman, J.M., Parrado, A., Bevis, B.J., Valastyan, J.S., et al. (2011). Functional links between A β toxicity, endocytic trafficking, and Alzheimer's disease risk factors in yeast. *Science* 334, 1241–1245.
- Tsunemi, T., and Krainc, D. (2014). Zn²⁺ dyshomeostasis caused by loss of ATP13A2/PARK9 leads to lysosomal dysfunction and alpha-synuclein accumulation. *Hum. Mol. Genet.* 23, 2791–2801.
- Tu, Z., Argmann, C., Wong, K.K., Mitnail, L.J., Edwards, S., Sach, I.C., Zhu, J., and Schadt, E.E. (2009). Integrating siRNA and protein-protein interaction data to identify an expanded insulin signaling network. *Genome Res.* 19, 1057–1067.
- Tuncbag, N., Braunstein, A., Pagnani, A., Huang, S.-S.C., Chayes, J., Borgs, C., Zecchina, R., and Fraenkel, E. (2013). Simultaneous reconstruction of multiple signaling pathways via the prize-collecting steiner forest problem. *J. Comput. Biol.* 20, 124–136.
- Tuncbag, N., Gosline, S.J.C., Kedaigle, A., Soltis, A.R., Gitter, A., and Fraenkel, E. (2016). Network-based interpretation of diverse high-throughput datasets through the Omics Integrator software package. *PLoS Comput. Biol.* 12, e1004879.
- Verstraeten, A., Theuns, J., and Van Broeckhoven, C. (2015). Progress in unraveling the genetic etiology of Parkinson disease in a genomic era. *Trends Genet.* 31, 140–149.
- Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 19, 327–335.
- Voevodski, K., Teng, S.-H., and Xia, Y. (2009). Finding local communities in protein networks. *BMC Bioinformatics* 10, 297.
- Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80–84.
- Wilson, G.R., Sim, J.C.H., McLean, C., Giannandrea, M., Galea, C.A., Riseley, J.R., Stephenson, S.E.M., Fitzpatrick, E., Haas, S.A., Pope, K., et al. (2014). Mutations in RAB39B cause X-linked intellectual disability and early-onset Parkinson disease with α -synuclein pathology. *Am. J. Hum. Genet.* 95, 729–735.
- Yeger-Lotem, E., Riva, L., Su, L.J., Gitler, A.D., Cashikar, A.G., King, O.D., Auluck, P.K., Geddie, M.L., Valastyan, J.S., Karger, D.R., et al. (2009). Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nat. Genet.* 41, 316–323.
- Zimprich, A., Benet-Pagès, A., Struhal, W., Graf, E., Eck, S.H., Offman, M.N., Haubenberger, D., Spielberger, S., Schulte, E.C., Lichtner, P., et al. (2011). A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am. J. Hum. Genet.* 89, 168–175.

STAR+ METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-Carboxypeptidase Y	Life Technologies	A66428
Rabbit anti-Nicastrin	Cell Signaling	3632; RRID: AB_2149581
phospho eIF2A	Cell Signaling	9721; RRID: AB_330951
total eIF2A	Cell Signaling	2103; RRID: AB_836874
LRRK2	Abcam	Ab133474
Experimental Models: Cell Lines		
Induced pluripotent stem cells (SNCA-A53T-1)	V.K. Lab, Chung et al. (2013)	N/A
Induced pluripotent stem cells (SNCA-A53T-CORR-1, SNCA-A53T-CORR-2)	V.K. Lab, Chung et al. (2013)	N/A
Induced pluripotent stem cells (LRRK2-G2019S-1, LRRK2-G2019S-2)	M.D. Lab, Reinhardt et al. (2013)	N/A
Induced pluripotent stem cells (LRRK2-G2019S-CORR-1, LRRK2-G2019S-CORR-2)	M.D. Lab, Reinhardt et al. (2013)	N/A
Experimental Models: Organisms/Strains		
Yeast strains: BY4741, deletion collection	C. Boone Lab, Tong and Boone (2006), Baryshnikova et al. (2010)	N/A
Yeast strain: W303 background, α -syn strains (NoTox), MATa <i>can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1</i>	V.K. Lab	N/A
Yeast strain: W303 background, α -syn strain (IntTox), MATa <i>can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1</i>	V.K. Lab	N/A
Yeast strain: W303 background, α -syn strain (HiTox), MATa <i>can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1</i>	V.K. Lab	N/A
Yeast strain: W303 background, α -syn/delta-VPS35, W303 MATa <i>can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1</i>	V.K. Lab	N/A
Yeast strain: W303 background, α -syn/delta-ATP13A2, MATa <i>can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1</i>	V.K. Lab	N/A
Recombinant DNA		
TALE cloning backbone	V.K. Lab	N/A
TALE-TF (Ataxin-2)	V.K. Lab	N/A
TALE (EIF4G1)	V.K. Lab	N/A
SNCA expression construct collection (GFP-tagged, untagged)	V.K. Lab	N/A
Software and Algorithms		
Transposenet Pipeline	web portal	http://transposenet.csail.mit.edu
DCA/Mashup	web portal	http://mashup.csail.mit.edu
PCSF	web portal	http://fraenkel-nsf.csbi.mit.edu/omicsintegrator/
Other		
Cytoscape Files	These are provided in the supplement (data files S20 through S30).	

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to the Lead Contact, Vikram Khurana, by email at vkhurana@bwh.harvard.edu.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Yeast Strains

For the deletion screen, strains were in the BY4741 background and have been described in detail elsewhere (Baryshnikova et al., 2010; Tong and Boone, 2006).

For all experiments except the deletion screen and validation, the yeast strains used were in the w303 background (MATa can1-100, his3-11,15, leu2-3,112, trp1-1, ura3-1, ade2-1). The vector control strain contained empty vector at the trp and ura loci (pAG304Gal, pAG306GAL). The NoTox α -syn strain contained α -syn fused to green fluorescent protein (α -syn-GFP) inserted at the his locus (pAG303Gal- α -syn-GFP). IntTox and HiTox α -syn strains contained multiple tandem copies of α -syn-GFP inserted at this and trp loci (pRS303GAL- α -syn-GFP, pRS304GAL- α -syn-GFP). IntTox strains have 4-5 copies of α -syn while HiTox cells have > 6 copies of α -syn. The Δ PARK17/ α -syn and Δ PARK9/ α -syn were generated by replacing the PARK17/VPS35 or PARK9/SPF1 gene loci in IntTox α -syn strains with a kanamycin resistance cassette (VPS35::kanMX or SPF1::kanMX).

Human iPSC Lines

iPSC from control individuals and PD patients carrying G2019S *LRRK2* along with isogenic gene-corrected controls were generated as previously described (Reinhardt et al., 2013). Skin biopsy, human dermal fibroblast culture, iPSC cell generation and mutation correction for the patient harboring the A53T mutation (α -syn^{A53T}) have been described previously (Cooper et al., 2006; Soldner et al., 2011). In that previous publication the A53T iPSC line was referred to as WIBR-IPS-SNCA^{A53T}. For all iPSC lines, informed consent was obtained from patients prior to cell donation using a written form, and the protocol was approved by the relevant institutional review board: for LRRK2 iPSC this was the Ethics Committee of the Medical Faculty and the University Hospital Tübingen (Ethik-Kommission der Medizinischen Fakultät am Universitätsklinikum Tübingen); for the A53T line, the IRB of the Boston University Medical Campus and the MIT Committee on the Use of Humans as Experimental Subjects.

Human iPSC Generation and Differentiation into Midbrain Dopaminergic (DA) Neurons for LRRK2 Mutant Lines

iPSC were differentiated into mDA neurons using a floor plate-based protocol with minor modifications (Kriks et al., 2011; Schöndorf et al., 2014). Differentiation was based on exposure to LDN193189 (100 nM, Stemgent) from days 0–11, SB431542 (10 mM, Tocris) from days 0–5, SHH C25II (100 ng/mL, R&D), purmorphamine (2 mM, EMD) and FGF8 (100 ng/mL, Peprotech) from days 1–7 and CHIR99021 (CHIR; 3 mM, Stemgent) from days 3–13. Cells were grown for 11 days on Matrigel (BD) in knockout serum replacement medium (KSR) containing DMEM, 15% knockout serum replacement, 2 mM L-glutamine and 10 μ M β -mercaptoethanol. KSR medium was gradually shifted to N2 medium starting on day 5 of differentiation. On day 11, media was changed to Neurobasal/B27/L-Glut containing medium (NB/B27; Invitrogen) supplemented with CHIR (until day 13) and with BDNF (brain-derived neurotrophic factor, 20 ng/ml; R&D), ascorbic acid (0.2 mM, Sigma), GDNF (glial cell line-derived neurotrophic factor, 20 ng/ml; R&D), TGF β 3 (transforming growth factor type β 3, 1 ng/ml; R&D), dibutyryl cAMP (0.5 mM; Sigma), and DAPT (10 μ M; Tocris) for 9 days. On day 18, cells were dissociated using Accutase (Innovative Cell Technology) and replated under high cell density conditions on dishes pre-coated with 15 μ g/ml polyornithine and 1 μ g/ml laminin in differentiation medium (NB/B27 + BDNF, ascorbic acid, GDNF, dbcAMP, TGF β 3 and DAPT). At DIV30, cells were collected and, after centrifugation, cell pellets were stored at -80°C until further analysis.

Human Pluripotent Stem Cell Culture for α -Syn Mutant Lines

Skin biopsy, human dermal fibroblast culture, iPSC cell generation and mutation correction for the patient harboring the A53T mutation (WIBR-IPS-^{A53T}) have been described previously (Cooper et al., 2006; Soldner et al., 2011). In that previous publication the A53T iPSC line was referred to as WIBR-IPS-SNCA^{A53T}.

Our pluripotent stem cell lines were initially maintained (5% O₂, 3% CO₂) on mitomycin C inactivated mouse embryonic fibroblast (MEF) feeder layers in hES medium [DMEM/F12 (Invitrogen) supplemented with 15% fetal bovine serum (FBS) (Hyclone), 5% KnockOut Serum Replacement (Invitrogen), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β -mercaptoethanol (Sigma) and 4 ng/ml FGF2 (R&D systems)]. Cultures were passaged every 5 to 7 days either manually or enzymatically with collagenase type IV (Invitrogen; 1.5 mg/ml). At around 50 passages prior to differentiation, lines were passaged to plates pre-coated with growth factor-reduced matrigel (BD Biosciences; 1:30 in DMEM:F12) and cultured (21% O₂, 5% CO₂) in mTESR-1 medium (Stem Cell Technologies), thereafter being passaged every 5 to 7 days enzymatically with dispase (Invitrogen; 1 mg/ml) until differentiation (at passage 40–90). For karyotyping, standard G-banding chromosomal analysis of cell lines was performed every 10–20 passages (Cell Line Genetics, Inc). We confirmed mycoplasma-negative status of our cultures every 2–4 weeks (MycoAlert, Lonza).

Primary Rat Cortical Cultures

All animal work was approved by the MIT Committee on Animal Care. Embryos were harvested by cesarean section from anesthetized pregnant Sprague-Dawley rats at embryonic day 18. Cerebral cortices were isolated and dissociated with Accumax (Innovative

Cell Technologies, Inc) digestion for 20 min at 37°C and triuration with Pasteur pipette. Poly-ornithine and laminin-coated 96 well plates were seeded with 4×10^4 cells respectively in neurobasal medium (Life Technologies) supplemented with B27 (Life Technologies), 0.5 mM glutamine, 25 μ M β -mercaptoethanol, penicillin (100 IU/ml) and streptomycin (100 μ g/ml). One third of the medium was changed every 3 to 4 days.

METHOD DETAILS

Yeast-to-Human Homology

Since yeast and human are evolutionarily distant species, to identify human homologs for yeast proteins, we developed a four-tiered meta-analysis pipeline. Our meta-analysis started at the sequence level, in which we first identify genes/proteins that are similar across yeast and humans. We then extend this analysis to the structural level, where we investigate the proteins that are structurally, and thus more distantly, similar across the species. Next, we identify proteins that are similar within each species by using a network-topology based approach. Finally, we introduce an approach to integrate similarity across sequence, structure and network topology. Details are as follows:

- 1) *Sequence Similarity*. To compute the sequence similarity between a yeast protein and a human protein, we used NCBI protein BLAST with the BLOSUM62 substitution matrix (Altschul et al., 1990, 1997). Sequence similarity was computed for all pairs of yeast proteins and human proteins. We used an E-value threshold = $1E-5$ to determine significance. We also used DIOPT (GTEx Consortium, 2013; Hu et al., 2011; Reinhardt et al., 2013; Söding et al., 2005), an integrative ortholog prediction web-server, to predict human orthologs for each yeast protein. We stored all filtered yeast-human protein pairs together with their BLAST E-values, bit scores and DIOPT scores.
- 2) *Evolutionary and Structural Similarity*. For each yeast and human protein, we applied PSI-BLAST to construct a multiple sequence alignment and build a profile hidden Markov model to encode a remote evolutionary signature. We then applied HHpred (Kriks et al., 2011; Robinson and Oshlack, 2010; Schöndorf et al., 2014; Söding et al., 2005; Voevodski et al., 2009), with the profile hidden Markov models and secondary structure annotations as input, to compare all pairs of yeast proteins and human proteins. As with the sequence similarity calculation, we also used an E-value = $1E-5$ threshold. We stored all filtered yeast-human protein pairs with their HHpred E-values and bit scores.
- 3) *Network Topology (Diffusion Component Analysis; DCA)*. The central idea behind our network topology approach is to try to capture functionally-related modules at the protein level, so that each node can be represented with a low-dimensional vector, instead of a single score, that captures homologous proteins in the network, along with conserved patterns of interactions. The eventual goal (see *Integrative Approach*, below) is to be able to compare low-dimensional representations of node vectors across species to yield information in other organisms. However, if we follow a straightforward PageRank-like approach (Cho et al., 2015; Tuncbag et al., 2016; Voevodski et al., 2009) to compute each node's vector, we get inaccuracies in functional similarity prediction due to network noise. Thus, using the intuition that compression decreases noise, we reduce the dimensionality of the vectors using sophisticated machine learning techniques. Our approach has been shown to reduce noise and be better able to extract topological network information such as functional similarity (Baillly-Bechet et al., 2011; Cho et al., 2015). The approach has recently been generalized into a method called Mashup (Cho et al., 2016).

More formally, let A denote the adjacency matrix of a (weighted) molecular interaction network $G = \delta V; EP$ with n nodes, each denoting a gene or a protein. Each entry B_{ij} in the transition probability matrix, which stores the probability of a transition from node i to node j , is computed as $B_{ij} = A_{ij} / \sum_k A_{jk}$. The diffusion algorithm is then defined as

$$s_j^{t+1} = \delta \sum_k p_{kj} B_{kj} + p e_i$$

until convergence, where p is the probability of restart, controlling the relative influence of local and global information in the network; e_i is a binary vector with $e_i = 1$ for node i itself and $e_j = 0$ for other nodes j . When the diffusion patterns of two nodes are similar to each other, it implies that they are in proximal locations in the network with respect to other nodes, which potentially suggest functional similarity. In practice, diffusion vectors obtained in this manner are still noisy, in part due to their high dimensionality as well as the noise and incompleteness of the original high-throughput network data. With the goal of noise and dimensionality reduction, we approximate each diffusion vector with a multinomial logistic model based on a latent vector representation of nodes that uses far fewer dimensions than the original vector. Specifically, we compute the probability assigned to node j in the diffusion vector of node i as:

$$s_{ij} = \frac{\exp(w_i^T x_j)}{\sum_k \exp(w_i^T x_k)}$$

where superscript T denotes vector transposition; w_i and x_i are low-dimension vectors. Each node is given two vector representations, w_i and x_i . We refer to w_i as the context feature and x_i as the node feature of node i , both capturing the intrinsic topological properties in the network. This multinomial logistic regression model is applied to model the relevance between a node and other nodes in a network, which can be modeled as a discrete distribution over all nodes in a network. To obtain

w and x vectors for all nodes, we optimize the KL-divergence (or relative entropy) between the diffusion vectors s_i and the model vectors ϕ_i :

$$\min_{w,x} \sum_i D_{KL}(s_i \| \phi_i)$$

Akin to PCA, which reveals the internal low-dimensional linear structure of matrix data that best explains the variance, this approach computes a low-dimensional vector-space representation for all genes such that the connectivity patterns in the network can be best explained. Comprehensive experiments showed that these low-dimensional vectors w and x are more accurate at identifying functional association within the network (Cho et al., 2016.; Tuncbag et al., 2013).

4) *Integrative Approach.* To compare proteins from yeast and human, we extended the above DCA method to consider the topology of both interactomes as well as the sequence/structural similarity between them. We converted the sequence and structure similarity scores to a probability distribution, and feature vectors of all pairs of nodes, including the sparse vector representations ones, were jointly computed by minimizing the Kullback-Leibler (KL) divergence between the relevance vectors and the parameterized multinomial distributions.

Formally, here we have two interactomes, G_Y for yeast and G_H for human. To capture the topological similarity within interactomes, we perform the described diffusion algorithm on G_Y and G_H separately and then obtain diffusion vectors s_i^Y for yeast protein i and s_j^H for human protein j . Similar to DCA on a single network, we also assign vectors $w_i^Y; x_i^Y$ for each yeast protein, and vectors $w_j^H; x_j^H$ for each human protein. To the sequence/structural similarity between obvious homologs, we normalize the BLAST bit scores between each yeast protein i and its human homologs j into a probability distribution as $b_{ij}^Y = \text{bit}_{ij}^Y / \sum_k \text{bit}_{ik}^Y$. Similarly we also normalize the BLAST bit scores between each human protein j and its yeast homologs i into a probability distribution as $b_{ji}^H = \text{bit}_{ji}^H / \sum_k \text{bit}_{jk}^H$. We likewise do the same normalization for HHpred bit scores as h_{ij}^Y and h_{ji}^H , and h_{ij}^Y and d_{ji}^H for DIOPT scores. Between each yeast protein i and human protein j , we approximate each normalized bit score distribution vector with a multinomial logistic model as:

$$\phi_{ij} = \frac{\exp(w_i^Y x_j^H)}{\sum_k \exp(w_i^Y x_k^H)}$$

Similar to the definition of ϕ_i for genes in the same molecular network, ϕ_{ij} captures the homologous similarity between a yeast gene and a human gene. In this way, although in different networks, yeast and human genes are represented in the same vector space.

Finally, we optimize an extended DCA objective function as:

$$\min_{w^Y, w^H, x^Y, x^H} \sum_{i \in V_Y} D_{KL}(s_i^Y \| \phi_i) + \sum_{j \in V_H} D_{KL}(s_j^H \| \phi_j) + a_{\text{Blast}} \sum_{i \in V_Y} D_{KL}(b_i^Y \| \phi_i) + a_{\text{HHpred}} \sum_{i \in V_Y} D_{KL}(h_i^Y \| \phi_i) + a_{\text{DIOPT}} \sum_{i \in V_Y} D_{KL}(d_i^Y \| \phi_i) + a_{\text{Blast}} \sum_{j \in V_H} D_{KL}(b_j^H \| \phi_j) + a_{\text{HHpred}} \sum_{j \in V_H} D_{KL}(h_j^H \| \phi_j) + a_{\text{DIOPT}} \sum_{j \in V_H} D_{KL}(d_j^H \| \phi_j)$$

where a_{Blast} , a_{HHpred} and a_{DIOPT} are parameters to tune the importance of each similarity component. Importantly, by optimizing these vectors, we integrate both molecular network connectivity and sequence similarity information into the same vector space for the purpose of comparison.

Here we used a greedy method to select these parameters. Specifically, we incrementally added each term and find the optimal or reasonable weight for the term, according to the functional concordance between the predicted yeast-human homology pairs. The details of the parameter selection procedure can be found in the “Parameter Tuning” section below. On the basis of the analyses included therein, we chose $a_{\text{Blast}} = 10$, $a_{\text{HHpred}} = 5$ and $a_{\text{DIOPT}} = 5$. Finally, we computed the integrated homologous association $p_{ij} = \phi_{ij} + \phi_{ji} / 2$ between yeast protein i and a human protein.

To find significant homology pairs, we computed p_{ij} for all yeast-human protein pairs and constructed the empirical background distribution. We used 0.0005 as the empirical p value cut-off to predict putative human homologs for yeast proteins and remove the homolog j if $p_{ij} < 0.5 \max_k p_{ik}$. The background distribution is generated by randomly pairing human and yeast genes. Utilizing this cutoff, there were 4923 yeast proteins with predicted human homologs, greatly improving the coverage of BLAST (4023 yeast proteins) and HHpred (4312 yeast proteins) (Figure S3A).

Preprocessing of Interactomes

We downloaded both yeast and human interactomes from the STRING v9.1 (string-db.org). In STRING, q_{ij} are the confidence values assigned for each edge in the interactome. We removed predicted interactions and re-calibrated the confidence for each interaction pair, such that $q_{ij} = 1 - \delta_{ij} / (q_{ij}^{\text{experimental}} + \delta_{ij} / q_{ij}^{\text{database}})$ with only “experimental” and “database” channels included. We also removed interaction pairs with low confidence $q_{ij} < 0.2$. After the preprocessing, we obtained a yeast interactome with 372026 interactions and 6164 proteins and a human interactome with 643822 interactions and 15317 proteins.

For the human networks, we also included two recently published high-quality binary human interactome datasets (11045 from high-quality re-curated binary interactions extracted from 7 public repositories; and 13944 from a recent yeast-2-hybrid experimental

dataset) (Geetha et al., 1999; Hu et al., 2011; Rolland et al., 2014). Since these interactions were unweighted, we needed to assign confidence scores for them. To estimate a good confidence value, we extracted all physical binary interactions from the BIOGRID database (v3.2.116) and computed the statistics of STRING confidence scores of these interactions. Since interactions from BIOGRID are mostly from high-throughput experiments and they are binary, we used the mean or median statistics to assign confidence scores for new binary interactions. The quantile statistics of STRING confidence scores of BIOGRID interactions were 25%: 0.391, 50%: 0.620 and 75%: 0.717. The average value of STRING confidence scores of BIOGRID interactions was 0.588. We thus considered it reasonable to assign a 0.6 confidence score for each unweighted binary interaction in these datasets.

As we were modeling neurodegenerative proteinopathies in the current work, we further pruned the human interactome to be brain-specific. To do so, we took GTEx gene expression dataset to only include genes appreciably expressed in brain (GTEx Consortium, 2013; Hu et al., 2011; Söding et al., 2005). Specifically, we normalized 357 GTEx brain RNA-seq datasets by the RPKM method (Robinson and Oshlack, 2010; Söding et al., 2005; Voevodski et al., 2009). We then filtered our human interactome such that only proteins with normalized brain expression level greater than (in at least one of 357 RNA-seq datasets) were included. In the end, our brain-specific interactome contained 369634 interactions and 10365 proteins.

Augmentation of Human Interactome with Yeast-to-Yeast Edges (for Humanized Networks Only)

Since genetic interactions are sparse in the human interactome, we used inferred homology to augment the human interactome by transferring edges from the yeast interactome. To do so, we added an edge between human proteins j and k if there exist a pair of yeast proteins i and l such that the integrated homologous association p_{ij} and p_{kl} satisfy $p_{ij} \square p_{kl} > 0.2$ (see definitions above). This threshold was chosen to make the augmented brain interactome attain a similar density ($\square 0.018$) to that of yeast interactome ($\square 0.019$) with 751282 interaction pairs transferred.

Prize-Collecting Steiner Forest Algorithm

We used the prize-collecting Steiner forest (PCSF) construction to analyze yeast networks and the augmented human-yeast network described above (Cho et al., 2015; Tuncbag et al., 2013, 2016.; Voevodski et al., 2009). For a network $G(V, E; c; p)$ of node (gene) set V and edge (interaction) set E (where $p(v) \in \mathbb{R}^+$ assigns a prize to each node $v \in V$, and $c(e) \in \mathbb{R}^+$ assigns a cost to each edge $e \in E$), the goal of PCSF is to find a set of trees $F \subseteq \mathcal{F}_F; E_F$ to minimize the following cost function:

$$f(F) = \sum_{v \in V} p(v) - b \sum_{v \in V} d(v) + \sum_{e \in E_F} c(e) + uk$$

where k is the number of connected components or trees in the forest F ; b is a parameter quantifying the trade-off between node prize and edge cost; $d(v)$ is the degree of node v ; m is a parameter to penalize hub nodes with a large number of neighbors in the network. In this way, the algorithm searches for a network of relatively high-confidence edges linking the experimental data.

To optimize the objective function $f(F)$, we introduced an extra root node v_0 into the network connected to each node $v \in V$ by an edge (v_0, v) with cost u . This step transforms the PCSF problem into a Prize-collecting Steiner Tree problem (PCST), which can be solved by a previously published message-passing-algorithm (Bailly-Bechet et al., 2011). After the tree solution was obtained, we removed node v_0 and all edges that point to it from the tree solution and obtained the forest solution. It is not hard to show that the tree solution is optimal for the above PCST if and only if the forest solution is optimal for the original PCSF. Although the message-passing algorithm is not guaranteed to find the optimal solution, it works very well in practice (Cho et al., 2015), and more importantly, it is substantially faster than linear programming approaches, which cannot handle large networks such as the yeast and human interactomes.

A computational difficulty of PCSF is how to tune the parameters b , u and m . Since b controls the scale of the prize values for nodes, we assigned a constant prize value (100) to each gene from our screens in our experiments. A perturbation of any parameter can potentially change the topology of the network structure, making the choice of parameters critical (Altschul et al., 1990, 1997; Ashburner et al., 2000; Tuncbag et al., 2013). Thus, instead of choosing a single set of parameters, we developed an ensemble approach to obtain the consensus network from multiple reasonable parameter settings.

To decide the range of parameters, we set the upper and lower bounds such that: the network solution of PCSF contained sufficient number of predicted proteins (which is half of the number of input prize genes); the network solution did not introduce hub nodes with more than 1000 neighbors in the input network. We discretized the range of the parameters into a grid and enumerated all possible parameter combinations for PCSF runs. For the yeast network, the range of b was $\{1; 2; 4; 6; 8; 10; 12g\}$; the range of u was $\{1; 2; 3; 4; 5; 6; 7; 8g\}$; the range of m was $\{0.001; 0.003g\}$. For the humanized network, the range of b was $\{4; 6; 8; 10; 12; 14; 16g\}$; the range of u was $\{3; 4; 5; 6; 7; 8; 9; 10g\}$; the range of m was $\{0.003; 0.005g\}$. We also injected edge noise for PCSF runs to test for robustness, using the default Gaussian noise setting in the PCSF program. After obtaining the solutions for each PCSF parameter setting, we computed the frequency of each possible edge appearing in the ensemble of all solutions. The frequency of an edge is a surrogate for the robustness of the edge across different parameter settings. Finally, we took as input the edges and their frequencies in the ensemble of all solutions and applied a maximum spanning tree algorithm to find the most robust, representative network.

To evaluate the significance of the selected nodes in the solution, we constructed a background distribution for each node by simulating the same PCSF and ensemble process using a random selection of the same number of yeast genes as input. We computed background distributions using random gene sets with identical degree distribution to that of the prize node lists. Specifically, we binned all yeast genes into four categories, each containing genes with degrees $[1-5]$, $[5-10]$, $[10-100]$ and $[> 100]$ respectively.

Random gene sets are then sampled without replacement from these categories such that the statistics of the degree distribution were identical to those of a prize node list. We then performed PCSF and generated 10000 random ensembles of forests from 1000 random sets to compute the empirical distributions of each node in the background.

To evaluate the significance of the overlaps of the forests relating to different proteinopathies (Figure 1D), we also calculated pairwise and triple-wise intersections of these random sets as background distributions. For example, we randomly paired the random ensembles generated for α -syn and random ensembles for TDP-43 and computed the distribution of the sizes of their overlaps. In this way, we constructed background distributions to evaluate the significance of the overlaps compared that simply caused by the increased size of the networks. Empirical p values are also computed. Similar to our previous results, all the pairwise overlaps were statistically significant ($p \leq 0.002$). For the triple-wise intersections, the p value was even more significant ($p \leq 0.001$).

Node and Edge Setup for Yeast and Humanized Steiner Networks

Aside from differences in parameterization (noted above), there were some important differences between the yeast networks and the “humanized” networks.

For the yeast networks (Figure 2), “prize nodes” were modifier hits from yeast genetic screens. Each prized node was assigned “100” as the arbitrary prize value. Edges for yeast networks were derived from STRING experimental and database edges. As described above, each edge was assigned a weight q_{ij} .

For the humanized networks (Figures 3 and 4), “prize nodes” were similarly defined as modifier hits from yeast genetic screens. Yeast-to-human edges were weighted by the strength of homology (p_{ij} above) between yeast proteins and their human homologs. On the humanized networks, these are the first-order links seen between the red triangles (which are hits from the screen) and blue circles (human homologs). If one of the clear human homologs of a yeast modifier was a known parkinsonism or neurodegenerative gene – for example, a *PARK* locus gene – an arbitrary reward of 0.5 was added to p_{ij} to favor inclusion of that node over other potential homologs. Finally, edges between human proteins in the humanized networks were derived from STRING, but also from other sources, as described in “Pre-processing of interactomes” and “Augmentation of human interactome with yeast-to-yeast edges” above.

Parameter Tuning for Computational Pipelines

Here, we provide analyses and guidelines for the parameters used in our paper.

Weights for BLAST, HHpred, and DIOPT in the DCA homology tool

Since it is impossible to select the optimal parameters without enumerating all possible combinations, we performed a greedy analysis for the parameter selection for the extended DCA objective function. Specifically, we incrementally added each term and found the optimal or reasonable weight for the term.

Since BLAST is the most sensitive method for sequence homology detection, we first explored a reasonable parameter interval for BLAST. We only retained the two network topology terms and the BLAST terms in the extended DCA objective function and enumerated alpha_BLAST from the set of {1,2,5,10, 20,100}. To evaluate the performance, we computed the average accuracy of Gene Ontology (GO) of the top 5 homologs predicted by our method. In Figure S2A it is readily seen that when the BLAST weight was too small (< 10), our method was not able to fully exploit the homology information from BLAST. When this weight was greater than or equal to 10, the predictive performance became saturated and only provided slight performance improvement over the original BLAST method. When the weight became too large ($= 100$), the predictive performance dropped and was identical to that of BLAST. This is because that the effect of network topology is diminished and our method simply reconstructed BLAST’s results. Thus, on the basis of the analysis we simply fixed the BLAST parameter to 10, although there might be better choices at extra computational cost by enumerating a larger and more refined set of possible values.

After we fixed the BLAST weight, we added the HHpred terms and performed the same analysis for HHpred weight. From the performance curve (Figure S2B), we observed that the optimal HHpred weight was around 5. This weight is smaller than BLAST weight, presumably at least in part because BLAST already captures most relevant homology information, while HHpred’s results extend BLAST by including extra remote sequence and structural homologs.

Finally, we fixed both BLAST and HHpred weights and performed the analysis for DIOPT weights (Figure S2C). For DIOPT, the performance difference was very small as long as the weight was not too large (< 20). This was consistent with the DIOPT database only providing a few additional sequence homologs missed by both BLAST and HHpred. For simplicity, we chose its weighting equal to 5 as well.

Significance threshold for BLAST and HHpred in the DCA homology tool

We chose $1E-5$ because it is a reasonably stringent threshold that is typically used for sequence homology or structure prediction (Geetha et al., 1999). Other choices of the threshold are possible but we believe that the results are not appreciably different from our setting. The following website and paper indicates $1E-5$ is a reasonably stringent cutoff for protein BLAST.

Reward to homologs of known Parkinson genes

The major reason why we added reward values to homologs of known Parkinson genes is that the prize-collecting Steiner forest (PCSF) algorithm is not guaranteed to include all prize nodes in the final network. In addition, our homology tool can sometimes assign similar scores to two homologs, one with known literature support, the other without. Although the PCSF algorithm itself is able to distinguish most correct homologs by considering the connectivity, we found that by rewarding well-known homologs the noise can be further reduced. The reward parameter 0.5 is chosen such that existing homologs of well-known Parkinson’s genes from

our screens are included in the final networks. It is obvious that larger reward values can have also the similar effect, but we didn't explore those choices because we hoped to not to overturn the effect of this reward heuristic in our pipeline.

Confidence threshold for existing interactomes and predicted links

The choice of confidence threshold for STRING is indeed a trade-off between false-positives and true-positives. A stringent threshold, e.g., 0.8, can reduce the number of false-positives but the truncated yeast and human interactomes appeared to be too sparse and disconnected. Such thresholds may work well for signaling pathways or other well-studied and localized biological pathways but we did not feel this was an appropriate approach for complex proteinopathies, where mechanisms are poorly understood (and casting a "broader net" seems more appropriate) and where the connections between seemingly disparate disease-relevant genes are not well understood. Thus, we selected 0.2 to only exclude very low-confidence interactions and still maintain the major connectivity of the interactomes.

Confidence score for new high-throughput binary interactomes

Since the new high-throughput binary interactomes are unweighted, we need to assign an appropriate score to merge them with STRING interactions. To estimate an appropriate confidence value, we extracted all physical binary interactions from the most recent BIOGRID database and computed the statistics of STRING confidence scores of these interactions. Since interactions from BIOGRID are mostly from high-throughput experiments and they are binary, we can use the mean or median statistics to assign confidence scores for new binary interactions. The quantile statistics of STRING confidence scores of BIOGRID interactions are 25%: 0.391, 50%: 0.620 and 75% 0.717. The mean value of STRING confidence scores of BIOGRID interactions is 0.588. We thus assigned 0.6 since it closely related to both the median and mean statistics, judging it a reasonable assignment for incorporating new high-throughput binary interactions into existing STRING database.

Parameters for prize-collecting Steiner forest algorithm (PCSF)

As noted above, we used an ensemble approach to avoid the problem of parameter selection. There is no obvious way to determine the effectiveness of a set of parameters for PCSF. Furthermore, since there are several parameters, enumeration of all combinations becomes computationally infeasible. To address this issue, as noted above, we selected a wide-range of possible parameters, ran PCSF with all parameter combinations and made an ensemble network from single networks generated from each parameter combinations. These parameters are chosen such that the final network can connect 80% prize nodes in the network. Our parameter range also excludes networks that are overly distorted by "greedy" hyperconnected hubs like ubiquitin. As noted in our methods section, we further tested robustness by injecting noise into the edge distribution. There is no question that there is an element of subjectivity here, as with any parameterized model but we have taken great pains to be as broad as we feel we possibly can. Ultimately, the purpose is to generate tenable hypotheses or to predict biologically meaningful interactions.

Spotting Assays

Yeast were cultured in synthetic media consisting of 0.67% yeast nitrogen base without amino acids (Fischer Scientific) supplemented with amino acids (MP Biomedicals) and 2% sugar. For most experiments, cells were first grown to mid-log phase in synthetic media containing glucose and then re-cultured overnight in synthetic media containing 2% raffinose. Mid-log phase cells were then diluted in synthetic media containing galactose. Typically, cells were induced for six hours at 30°C.

Each strain was diluted to a starting $OD_{600} = 1.0$ and serially diluted five-fold and then spotted on agar plates containing galactose (inducing) or glucose (control) plates.

Screening against Known α -Syn Modifiers in Δ PARK17/ α -Syn and Δ PARK9/ α -Syn Strains

The standard lithium acetate transformation protocol was adapted for use with 96-well plates (Cooper et al., 2006; Gietz et al., 1992, 1995). Following transformation, cells were grown to saturation in synthetic media with raffinose lacking uracil for selection of yeast transformed with the desired plasmid. Once at saturation, they were spotted onto synthetic media plates with either glucose or galactose. Following two days of growth, galactose and glucose plates were photographed and analyzed by eye. In parallel experiments, transformed yeast were rediluted to $OD_{600} = 0.01$ in 35 μ L of galactose media in 384-well plates. Growth in 384-plates was monitored by measuring the OD_{600} after 18, 24, and 48 hr of growth (Tecan safire²) giving a quantifiable measure of growth.

Small Molecule (NAB2) Treatment

Control, TDP-43 or α -syn yeast strains were grown to log-phase ($OD_{600} \approx 0.5$) in complete synthetic media containing raffinose (non-inducing). Cultures were then diluted to an OD_{600} of 0.01 (TDP-43 experiment) and 0.025 (α -syn experiment) in complete synthetic media containing 2% galactose to induce expression of the toxic protein. For NAB treatment, 10 μ M (for α -syn) or 20 μ M (for TDP-43) were added to the cultures and incubated in a Bioscreen instrument with intermittent shaking at 30°C for two days.

Pooled α -Syn Overexpression Screen

Pooled genetic screens were carried out in a YFP control strain and an α -syn strain. The yeast FLEXgene library representing most yeast open reading frames (Hu et al., 2007) was pooled from an arrayed bacterial library stock and grown to saturation in deep 96 well plates at 37°C. Cultures were pooled and plasmids isolated using QIAGEN maxi prep kits. The pooled FLEXgene library was then transformed en masse into either control YFP or α -syn-expressing yeast strains and selected on five square 15 cm solid agar plates lacking uracil for plasmid selection. Approximately 10^6 CFUs were obtained, representing an approximate 200-fold coverage of the $\approx 6,000$ yeast genes. Colonies were rinsed off of each plate, pooled, brought to 20% glycerol, aliquoted to individual use tubes (≈ 100 μ L), snap frozen in liquid nitrogen, and stored at $\approx 80^\circ\text{C}$.

Pooled screens were executed as follows. An aliquot of pooled yeast library was thawed on ice and diluted at three different concentrations into 3 × 30 mL flasks with S_{Raf}Ura (0.025, 0.05, and 0.1). After shaking at 30°C overnight, the culture with an OD₆₀₀ between 0.4 and 0.8 was selected to begin the pooled screen. Cultures were then diluted to an OD₆₀₀ of 0.1 in S_{Gal} Ura to induce expression of either YFP or α -syn. 50 OD units were kept as time zero and centrifuged, washed with water, and frozen. Cultures were then maintained in log phase growth for 24 hr, making appropriate dilutions when needed to maintain an OD₆₀₀ under 0.8. After this time, 50 OD units worth of culture were centrifuged, washed with water, and pellets frozen.

Plasmids were then isolated from yeast using QIAGEN minipreps with the following adaptations. Five minipreps were done per 50 OD units. Cell pellets were resuspended in buffer and lysed by bead beating with small acid-washed beads. Beads were removed and the lysate then taken through the conventional miniprep protocol. The purified plasmids from the five preps were then pooled. The yeast ORFs contained on the FLEXgene plasmids were then amplified using PCR primers that annealed to the attR Gateway sequences flanking the ORFs. HiFidelity Platinum Taq was used for amplification. 5 μ L DNA was used per 50 μ L reaction and four reactions were performed per sample. 30+ cycles with a 6' extension time was used to ensure amplification of longer ORFs. PCR product was purified using QIAGEN PCR columns. Two micrograms of PCR product was then sonicated, purified on QIAGEN Minelute PCR columns, and the OD₂₆₀ re-analyzed. This product was then used as input for library generation and sequencing by the Whitehead Institute Genome Technology Core. Illumina HiSeq platform was used to sequence approximately 120 million 40 bp single end reads.

Reads were mapped to the yeast ORFs sequences with bowtie (Langmead et al., 2009). We made a bowtie index with the DNA sequences of the yeast ORFs reported in Hu et al. (Hu et al., 2007), plus 903 ORFs that were present in SGD but were not included in the list of sequences from in Hu et al. Reads were mapped allowing 2 mismatches (-n 2) in the seed, seed length of 40 (-l 40), suppressing all alignments that map to more than one place (-m 1) and using "-best" and "-strata." Unmapped reads were trimmed with fastx_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/commandline.html) to remove the first 20 nt, and remapped with bowtie using the following parameters: "-n 0 -l 20 -best -strata -m 1". The number of reads mapping to each ORF was obtained parsing the output sam files. Differential expression analysis was done with the R package Noiseq (Tarazona et al., 2011). NOISeq is a nonparametric method to identify differentially expressed genes from count data. NOISeq calculates fold change values and probability of differential expression. The probability (P-val) of differential expression for each gene is derived from the joint distribution of fold-change differences (M)- absolute expression differences (D) values for all the genes within the Table Set.

A gene was selected for validation if it was: (A) up or down consistently in the two pooled α -syn screens ($|\log_2$ fold change| > 0.8 in both screens) except when neither experiment was associated with a P-val of > 0.5; (B) had an average fold change with absolute value of > 2.5 (regardless of P-val); (C) known modifiers from previous experimentation that had a fold-change in the pooled screen consistent with that source. Any gene with an $|\log_2$ fold change| > 1.0 in the YFP control (in the same direction as the putative suppressor or enhancer) was excluded, as well as genes associated with galactose metabolism that would be expected to alter expression of gal-inducible transgenes. Thresholds were guided by knowledge gained from our previous extensive characterization of the arrayed α -syn overexpression screen hits (see Figure 1). Put another way, our previous overexpression screen was used as a "gold standard" to analyze the pooled overexpression data.

Pooled Screen-QPCR Verification

Transformed cells generated from the pooled screen ("Pooled α -syn overexpression Screen" method) were thawed on ice and diluted in S_{Raf}-Ura to resulting ODs of approximately 0.03, 0.05 and 0.1. Cultures were grown at 30°C overnight and cultures with an OD of 0.4-0.8 were chosen for induction. These cultures were diluted to an OD of 0.1 in S_{Gal}-Ura. 50 OD units were stocked representing the time zero time point. Induced cultures were grown for 24 hr and 50 OD units were stocked representing the 24hr time point. Plasmids were isolated using the QIAGEN miniprep kit (27106) splitting the 50OD units for each time point in to 5 samples. Following cell resuspension in P1 buffer cells were lysed by bead beating using acid-washed beads. Following bead beating, beads were removed from samples and lysates subjected to the standard miniprep kit protocol. Resulting plasmids were pooled and used for QPCR analysis. The standard attF primer was used in combination with an orf specific reverse primer (sequence generated by Primer3 such that the product < 150bp in size) for QPCR analysis. Multiple negative controls used to normalize samples and positive controls were run on all QPCR plates. QPCR analysis was performed using technical triplicates of biological triplicates on the Applied Biosystems (7900HT) using the SYBR green fluorescence detection system (Applied Biosystems). The program for amplification comprised 40 cycles of 95°C for 15 s and 60°C for 1 min.

Pooled Screen-Growth Curve Analysis

Each individual putative modifier was overexpressed in the α -syn strain using the Flexgene overexpression library. Three independent Ura⁺ transformants were grown in S_{Raf}-Ura at 30°C overnight. Cultures were subcultured in S_{Raf}-Ura and at an OD of 0.4-0.8 were diluted in S_{Gal}-Ura for induction. Each isolate was set up in triplicate and growth was monitored every 15 min for approximately 60 hr.

Genome-Wide Deletion Screen (Synthetic Gene Array Methodology)

The method used was essentially as described previously (Baryshnikova et al., 2010; Tong and Boone, 2006). Briefly, deletion strains were pinned on to YPD+G418 plates. Query strains (α -syn and wild-type control) were grown in 5ml overnight cultures in YPD at 30°C and spread on YPD plates and grown overnight. Deletion strains were mated to each query strain by pinning together on YPD and

grown for 48hrs at 30°C. Resulting diploids were pinned to SD/MSG-Ura+G418 and grown for 2 days at 30°C. Cells were pinned to sporulation media plates and incubated at 23°C for 7 days. Spores were pinned to SD-His/Arg/Lys+canavanine+thialysine and grown for 2 days at 30°C. Cells were pinned to fresh SD-His/Arg/Lys+canavanine+thialysine and grown for 1 day at 30°C. Cells were pinned to SD/MSG-His/Arg/Lys+canavanine+thialysine + G418 and grown for 2 days at 30°C and then pinned to SD/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and grown for 2 days at 30°C. For the initial screen, cells were pinned both to SD/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and to Sgal/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and spot growth was monitored. For validation studies, cells were pinned to liquid SD/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and grown overnight at 30°C and then pinned both to SD/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and to Sgal/MSG-His/Arg/Lys/Ura +canavanine+thialysine+G418 and spot growth was monitored. Stock solutions (1000X) were prepared as follows: G418 200mg/ml, canavanine 50mg/ml, thialysine 50mg/ml. The method above was used for the initial screen and repeated, in duplicate, using 96-well plate format for validation of the initial screen hits.

Human iPSC Generation and Differentiation into Midbrain Dopaminergic (DA) Neurons for LRRK2 Mutant Lines

iPSC from control individuals and PD patients carrying G2019S *LRRK2* along with isogenic gene corrected controls were generated as previously described (Reinhardt et al., 2013). iPSC were differentiated into mDA neurons using a floor plate-based protocol with minor modifications (Kriks et al., 2011; Schöndorf et al., 2014). Differentiation was based on exposure to LDN193189 (100 nM, Stemgent) from days 0–11, SB431542 (10 mM, Tocris) from days 0–5, SHH C25II (100 ng/mL, R&D), purmorphamine (2 mM, EMD) and FGF8 (100 ng/mL, Peprotech) from days 1–7 and CHIR99021 (CHIR; 3 mM, Stemgent) from days 3–13. Cells were grown for 11 days on Matrigel (BD) in knockout serum replacement medium (KSR) containing DMEM, 15% knockout serum replacement, 2 mM L-glutamine and 10 μ M β -mercaptoethanol. KSR medium was gradually shifted to N2 medium starting on day 5 of differentiation. On day 11, media was changed to Neurobasal/B27/L-Glut containing medium (NB/B27; Invitrogen) supplemented with CHIR (until day 13) and with BDNF (brain-derived neurotrophic factor, 20ng/ml; R&D), ascorbic acid (0.2 mM, Sigma), GDNF (glial cell line-derived neurotrophic factor, 20 ng/ml; R&D), TGF β 3 (transforming growth factor type β 3, 1 ng/ml; R&D), dibutyryl cAMP (0.5 mM; Sigma), and DAPT (10 μ M; Tocris) for 9 days. On day 18, cells were dissociated using Accutase (Innovative Cell Technology) and replated under high cell density conditions on dishes pre-coated with 15 μ g/ml polyornithine and 1 μ g/ml laminin in differentiation medium (NB/B27 + BDNF, ascorbic acid, GDNF, dbcAMP, TGF β 3 and DAPT). At DIV30, cells were collected and, after centrifugation, cell pellets were stored at 4°C until further analysis.

Human Pluripotent Stem Cell Culture for α -Syn Mutant Lines

Skin biopsy, human dermal fibroblast culture, iPS cell generation and mutation correction for the patient harboring the A53T mutation (WIBR-IPS-A53T) have been described previously (Cooper et al., 2006; Soldner et al., 2011). In that previous publication the A53T IPS line was referred to as WIBR-IPS-SNCA^{A53T}.

Our pluripotent stem cell lines were initially maintained (5% O₂, 3% CO₂) on mitomycin C inactivated mouse embryonic fibroblast (MEF) feeder layers in hES medium [DMEM/F12 (Invitrogen) supplemented with 15% fetal bovine serum (FBS) (Hyclone), 5% KnockOut Serum Replacement (Invitrogen), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β -mercaptoethanol (Sigma) and 4 ng/ml FGF2 (R&D systems)]. Cultures were passaged every 5 to 7 days either manually or enzymatically with collagenase type IV (Invitrogen; 1.5 mg/ml). At around 50 passages prior to differentiation, lines were passaged to plates pre-coated with growth factor-reduced matrigel (BD Biosciences; 1:30 in DMEM:F12) and cultured (21% O₂, 5% CO₂) in mTESR-1 medium (Stem Cell Technologies), thereafter being passaged every 5 to 7 days enzymatically with dispase (Invitrogen; 1mg/mL) until differentiation (at passage 40–90). For karyotyping, standard G-banding chromosomal analysis of cell lines was performed every 10–20 passages (Cell Line Genetics, Inc.). We confirmed mycoplasma-negative status of our cultures every 2–4 weeks (MycAlert, Lonza).

Human Neural Induction by Embryoid Body (EB) Formation

A previously published protocol was used without modification (Chung et al., 2013; Hu et al., 2007; Kim et al., 2011). This protocol has been repeated here for completeness.

To initiate differentiation, on day 0 human ES or iPS cell colonies were pretreated for 30–60 min with 5 μ M Y-27632/ROCK inhibitor (Calbiochem), single cell-dissociated after 5–10min exposure to accutase (StemPro Accutase; Life Technologies) and then re-suspended in neural base (NB) medium, which is DMEM/F12 (GIBCO/Life Technologies) supplemented with N2 and B27. N2 and B27 supplements from Life Technologies and used at 1/2–1% and 1%–2%, respectively. Cells were plated in AggreWell 800 microwells (StemCell Technologies; priming and plating per manufacturer's protocol; 2.4x10⁶ cells were well) in NB medium supplemented with dual SMAD inhibitors (Chambers et al., 2009; Langmead et al., 2009) recombinant human Noggin (R&D Systems) at 200ng/mL and 10 μ M SB431542 (Tocris Bioscience), as well as 5 μ M Y-27632. Noggin and SB431542 remained in the medium at these concentrations throughout the neural differentiation protocol.

On day 1 medium was 1/2-changed. By day 2, well-formed neuralized EBs (NEBs) were typically observed in the AggreWells and transferred to Petri dishes (4 AggreWell wells/Petri dish) overnight, in NB medium. On day 4, NEBs were transferred to a dish coated with growth factor-reduced Matrigel (1:30 in DMEM:F12; BD Biosciences) for attachment. Y-27632 was omitted from this day onward. From day 5 to day 10, attached NEBs were additionally exposed to 20 ng/mL FGF2 (R&D Systems) and recombinant human Dkk1 at 200 ng/mL (R&D Systems). On day 10, neural rosettes were dissected (P20 pipette tip), incubated in accutase supplemented

with DnaseI (Sigma Aldrich) for 10 min at 37°C and gently dissociated to small cellular clumps and single cells. After washing, the rosettes were re-plated on plastic dishes pre-coated with poly-L-ornithine and laminin (BD Biocoat) at high density (200,000/cm²) in neural progenitor cell (NPC) medium, which is NB medium supplemented with 20 ng/mL FGF2. (Life Technologies), supplemented overnight with 10µm Y-27632. Typically, one Aggrewell 800 well provided enough NPCs for at least 1-2 6-wells at passage 0.

Thereafter, the surviving NPCs proliferated. Medium change was daily. They could be passaged up to 10 times before neural differentiation, and could successfully be freeze/thawed at early passage (p1 to p5) without compromising differentiation potential. Freezing medium was NPC medium with 10% FBS (Hyclone).

Human Cortical Neural Differentiation

A previously published protocol was used without modification (Chung et al., 2013; Hu et al., 2007; Kim et al., 2011). This protocol has been repeated here for completeness.

To begin neural differentiation, NPCs were dissociated with accutase and re-plated on matrigel-coated T75 flasks (CytoOne). The next, day medium was fully changed to Neural Differentiation (ND) medium, which is NB medium supplemented with recombinant human BDNF and GDNF (both at 10ng/mL; R&D Systems) and dibutyryl cyclic AMP (Sigma; 500µM), and without FGF-2. Thereafter, media was ½-changed every other day. On day 7-9, differentiating neurons were gently dissociated to single cell, resuspended in pre-chilled Hank's balanced salt solution (HBSS; GIBCO / Life Technologies) supplemented with 0.1% bovine serum albumin (GIBCO / Life Technologies). After a wash step, cells were plated on 6- or 24-well plastic plates pre-coated with poly-ornithine and laminin (BD Biocoat) for biochemical assays. Medium was ½-changed every 3 days for up to 12 weeks.

Primary Rat Cortical Cultures

Embryos were harvested by cesarean section from anesthetized pregnant Sprague-Dawley rats at embryonic day 18. Cerebral cortices were isolated and dissociated with Accumax (Innovative Cell Technologies, Inc) digestion for 20 min at 37°C and trituration with Pasteur pipette. Poly-ornithine and laminin-coated 96 well plates were seeded with 4x10⁴ cells respectively in neurobasal medium (Life Technologies) supplemented with B27 (Life Technologies), 0.5 mM glutamine, 25 µM β-mercaptoethanol, penicillin (100 IU/ml) and streptomycin (100 µg/ml). One third of the medium was changed every 3 to 4 days.

AAV-1 Transduction of IPS Neurons

Plasmids containing verified TALE-TFs were purified endotoxin-free (QIAGEN) and packaging into adeno-associated viruses serotype 1 (AAV-1) was conducted by the Gene Transfer Vector Core, Massachusetts Eye and Ear Infirmary/MEEI, Harvard Medical School (mini-scale production). A53T and mutation-corrected cortical neurons were aged for 4-7 weeks at a plating density of 0.25-0.75 3 10⁶ cells/cm². Cells were transduced with 30 µL of the mini scale produced MEEI AAV-1 titer, containing a single TALE-TF or the TALE cloning backbone alone, in 500 µL ND medium. ND medium was changed 12-16 hr post-transduction.

Antibodies

Mouse anti-Carboxypeptidase Y	Life Technologies A66428	Western blot	1:10 000
Rabbit anti-Nicastrin	Cell Signaling 3632	Western blot	1:1000
phospho eIF2A	Cell Signaling 9721	Western blot	1:1000
total eIF2A	Cell Signaling 2103	Western blot	1:1000
LRRK2	Abcam Ab133474	Western blot	1:500

Protein Labeling with ³⁵S-Methionine/-Cysteine

A53T and mutation-corrected cortical neurons were aged for 4-8 weeks at a plating density of 0.25-0.75 3 10⁶ cells/cm². Prior to the protein labeling the cortical neuronal cultures were kept in Neural Differentiation (ND) medium without methionine and cysteine for 90 min. ND medium was DMEM complemented with 1% (v/v) B-27, 0.5% (v/v) N-2 and 1% (v/v) GlutaMAX supplement, 1% (v/v) MEM non-essential amino acids, 1% (v/v) Penicillin-Streptomycin (all Life Technologies) as well as 10 ng/ml BDNF and GDNF (both R&D Systems) and 500 µM cAMP (Sigma-Aldrich). For protein labeling the neuronal cell cultures were incubated in ND medium supplemented with ³⁵S-methionine and -cysteine (Perkin Elmer) at a final concentration of 100 µCi/ml for various duration. After a quick wash with cold PBS, cells were lysed in a buffer containing 50 mM Tris-HCl and 2% (w/v) SDS, supplemented with protease inhibitor cocktail (Sigma-Aldrich). The samples were boiled at 100°C for 5 min and spun down at 10,000 g for 15 min. The supernatant was collected and the protein concentration was determined using BCA assay (Pierce, Thermo Fisher Scientific). ³⁵S labeled samples were run in 4%–12% Nupage Bis-Tris gel (Life Technologies). As a loading control, gels were stained with SimplyBlue SafeStain (Life Technologies), and destained by incubation in water. Thereafter, the gels were incubated in 11.2% (v/v) salicylic acid and 10% glycerol (v/v) for 15 min. The gels were dried and exposed to a phosphor screen (Fujifilm) for a minimum of 48 hr. The screen was scanned using the phosphorimager BAS-2500 (Fujifilm) and ³⁵S incorporation was determined by measuring the intensity of each lane (MultiGauge Analysis Software v2.2, Fujifilm).

Free ³⁵S-Methionine/-Cysteine in the Cytosol

Rat primary neurons overexpressing either GFP or α Syn-GFP were incubated with ³⁵S-methionine and -cysteine at 100 μ Ci/ml for various durations. After a quick wash with cold PBS, cells were lysed in RIPA buffer for 20 min on ice and the debris was removed by centrifugation. Proteins in the lysates were precipitated by adding 1 volume 100% TCA to 4 volume of lysate and incubate 10 min at 4°C. After centrifugation at 14K rpm for 10 min, supernatant was collected to measure a cytosolic pool of free ³⁵S-methionine/-cysteine. ³⁵S incorporation was determined by quantifying using an LS 6500 liquid scintillation counter (Beckman Coulter) with 5 μ L of the sample being immersed in 7 mL scintillation cocktail (National Diagnostics).

Cell Lysis and Endoglycosidase H Digestion

Cells were lysed in a buffer containing 20 mM HEPES, 150 mM NaCl, 10% (v/v) glycerol, 1 mM EGTA, 1.5 mM MgCl₂, 1% (v/v) Triton X-100, pH to 7.4, protease inhibitor cocktail (Sigma-Aldrich), and protein phosphatase inhibitor cocktail 1 and 2 (Sigma-Aldrich), and incubated in an ice/water slurry for 20 min, followed by 2 freeze-thaw cycles (□80°C/37°C, □1 min each). Supernatant was collected after ultracentrifugation at 100,000 g, 4°C, for 30 min. Protein concentration was determined using BCA assay (Pierce, Thermo Fisher Scientific). Endoglycosidase (Endo) H (New England Biolabs) digestion was performed based on the manufacturer's instructions. Briefly, 20–40 μ g bulk protein was assembled in 15.3 μ L reaction volume; 1.7 μ L denaturing buffer was added and samples were boiled for 10 min at 100°C. Then 2 μ L of G5 buffer and 1 μ L of Endo H or 1 μ L H₂O were added to the denatured reaction and incubated for 2 hr at 37°C.

Western Blotting

For protein trafficking after Endo H digestion, protein samples were denatured in sample buffer (20 mM Tris-Cl pH 6.8, 4% (v/v) glycerol, 180 mM 2-mercaptoethanol, 0.0003% (v/v) bromophenol blue and 2% (v/v) SDS), run in 10% Tris-glycine gel, and wet transferred with 20% methanol onto PVDF membranes (Bio-Rad). Blots were blocked in a 1:1 dilution of Odyssey blocking buffer (Li-Cor Biosciences) and PBS for 1 hr at room temperature, followed by incubation with primary antibodies in a 1:1 dilution of Odyssey blocking buffer (Li-Cor Biosciences) and PBS containing 0.1% Tween 20 (PBST) at 4°C overnight with gentle rocking. After three 5 min washes with PBST, blots were incubated with secondary antibodies such as anti-mouse or -rabbit IgG conjugated to IRDye 680 or 800 (1:10,000, Rockland) in a 1:1 dilution of Odyssey blocking buffer and PBST for 2 hr at room temperature. After three 5 min washes with PBST and two with water, blots were scanned using the Odyssey quantitative fluorescent imaging system (Li-Cor Biosciences) and bands were quantitated using Odyssey Software v2.1 (Li-Cor Biosciences).

For other western blots, samples were lysed in RIPA buffer and run in either 8 or 10% Nupage Bis-Tris gel (Life Technologies) and transferred using iBlot (Life Technologies). Blocking was in 5% nonfat dry milk in PBST. As for the secondary antibodies and chemiluminescent detection, anti-mouse, -rabbit IgG or avidin conjugated to HRP was used with SuperSignal West Pico chemiluminescent substrate (Thermo Fisher Scientific).

TALE-TF Design

TALE-TFs were designed to target between 200bp upstream (5') and 50bp downstream (3') of the transcription start site (TSS) of ATXN2 or EIF2G transcripts. Within these regions near the TSS, we identified DNaseI hypersensitive regions from human ventromedial prefrontal cortex samples (PMID: 22955617). Within these DNaseI HS regions, we designed 5 TALE-TFs for each transcript.

Each TALE-TF was designed to target a 14bp genomic sequence consisting of an initial thymidine (T) plus 12 full repeats and 1 half repeat. For each TALE-TF, the TALE repeats were cloned into an rAAV transfer plasmid using a PCR-based, Golden Gate cloning strategy as previously described (Konermann et al., 2015; Sanjana et al., 2012; Tarazona et al., 2011). The rAAV transfer plasmid contained the TALE backbone fused to the synthetic VP64 activator domain along with a 2A-linked EGFP that is cleaved during translation.

TALE-TF Assembly

14-mer transcription activator-like effector transcription factors (TALE-TFs) were constructed using Golden Gate cloning as described previously (Sanjana et al., 2012). For each gene, *ATXN2* and *EIF4G1* (transcript variant 7), five different TALE-TFs were designed with the 14 bp long target loci being located in the proximal promoter region (*ATXN2* TALE-TF #1: 5'-TGTCAGAA TAAAGG-3', #2: 5'-TGAACCTATGTTCC-3', #3: 5'-TGCCAGATTCAGGG-3', #4: 5'-TGGAGCGAGCGCCA-3', #5: 5'-TAGCTGGT CATGGT-3', *EIF4G1* TALE-TF #1: 5'-TGTCACGTGACGGG-3', #2: 5'-TGTGGCTGTACAGT-3', #3: 5'-TCAAAGTTCGGGAG-3', #4: 5'-TCGCGGAACAGAGA-3', #5: 5'-TCTCCTGCCTCAGC-3'). For each TALE-TF the correct sequence of the DNA-binding domain was verified by Sanger sequencing and all TALE-TF clones with non-silent mutations were excluded.

Ribosomal Footprint Profiling

For ribosome footprint profiling, 12-week old cells were treated with cycloheximide (100 μ g/mL) for 5 min at 37°C to stop translation elongation. Cells were washed twice with ice-cold 9.5 mM PBS, pH 7.3, containing 100 μ g ml⁻¹ cycloheximide, and lysed by adding lysis buffer (10 mM Tris-HCl, pH 7.4, 5 mM MgCl₂, 100 mM KCl, 2 mM dithiothreitol, 100 μ g ml⁻¹ cycloheximide, 1% Triton X-100, 500 U ml⁻¹ RNasin Plus, and protease inhibitor (1x complete, EDTA-free, Roche)), scrapping cells from the plate, and then triturating four times with a 26-gauge needle. After centrifuging the crude lysate at 1,300g for 10 min at 4°C, the supernatant was removed and flash-frozen in liquid nitrogen. The lysate was thawed on ice, after which ribosome profiling and mRNA-seq were performed as

described previously (Subtelny et al., 2014) using a detailed protocol available at <http://bartellab.wi.mit.edu/protocols.html>. The 4-week old cells were washed twice with 37°C growth media, then after removing the media by aspiration the plates were sealed and then plunged into liquid nitrogen. Cells were then lysed with lysis buffer as described above, but cycloheximide was excluded from all solutions including the sucrose gradients. After thawing on ice, a small amount of cycloheximide-free zebrafish RPF lysate was spiked into the 4-week old cell lysates (10-fold less based on A₂₆₀) prior to digestion with RNase I.

RPF and RNA-seq tags were mapped to the ORFs, as described previously (Subtelny et al., 2014). To account for the zebrafish reads present in the 4-week old samples, indexes comprising both the zebrafish and human genomes or transcriptomes were created and these data were mapped to the combined indexes. Only reads mapping uniquely were considered, and those mapping to zebrafish were excluded from the analysis.

Enriched pathways in the translational profiling for the 4-week and 12-week datasets were computed with the Gene Set Enrichment Analysis tool, available at the Broad Institute website (<http://software.broadinstitute.org/gsea/index.jsp>).

QUANTIFICATION AND STATISTICAL ANALYSIS

Comparison with Existing Homology Prediction Approaches

To evaluate the functional association between yeast proteins and the predicted human homologs, we computed the average accuracy of Gene Ontology (GO) of the top 5 homologs predicted by our method, HHpred and BLAST (Altschul et al., 1990, 1997; Ashburner et al., 2000; Tuncbag et al., 2013) (Figure S3B). We chose the top 5 homologs since yeast proteins often have more than one good human homolog. The accuracy of a homolog was calculated as the percentage of overlapped GO labels between the yeast protein and the putative homolog. We noted that the number of assigned GO labels per gene varied considerably between yeast and human proteomes, so that the GO accuracy metric favored predicted homologs with a large number of labels and query proteins with a small number of GO labels, potentially biasing the analysis. Furthermore, false positives were not considered by this metric. To address these issues, we computed the widely used Jaccard similarity score, which is the number of overlapping GO labels divided by the total number of unique GO labels of the yeast (or human) gene and its human (or yeast) homolog. BLAST's accuracy for 4023 yeast proteins was 31.1%. HHpred in conjunction with BLAST achieved of 32.6% for accuracy for 4312 yeast proteins. Our method obtained 31.6% accuracy for a significantly greater number, 4923, of yeast proteins. It also outperformed BLAST on 4023 yeast proteins with BLAST output (32.0% versus 31.1% accuracy and 25.2% versus 24.3% Jaccard similarity) and HHpred on 4312 proteins with HHpred output (34.1% versus 32.6% accuracy and 26.9% versus 24.9% Jaccard Similarity). The improvements over BLAST and HHpred were significant (paired t test p values < 0.01).

We then tested our method on finding yeast homologs for human proteins (Figures 3C and 3D). The improvement of the coverage over BLAST and HHpred was even more substantial than for generating human homologs from yeast proteins. Our method predicted homologs for 15200 proteins but BLAST and HHpred only covered a relatively small portion of human proteome (7248 and 9577 respectively). Accuracy metrics also favored the DCA method. Our method improved the predictive power over BLAST (57.6% versus 57% accuracy and 26% versus 26.6% Jaccard similarity) and HHpred (56% versus 54.9% accuracy and 25% versus 24.2% Jaccard similarity) on proteins which BLAST or HHpred can find yeast homologs on both GO accuracy and Jaccard similarity score. These comparisons were all statistically significant (all p values < 0.01 by paired t test).

We also compared our homology tool to the state-of-the-art Ensembl Compara method. Ensembl Compara identifies high confidence homolog pairs through phylogenetic tree-based clustering and analysis across multiple species. This sequence-based method sacrifices coverage for accuracy, and these pairs are considered a gold standard for traditional analyses (Vilella et al., 2009). We downloaded the Ensembl Compara v85, and mapped gene ids to the gene names used in our homology tool, identifying 5093 high-confidence yeast/human pairs for 2409 yeast genes. Among these pairs, there are three major categories: "one-to-one," "one-to-many" and "many-to-many." To evaluate our DCA homology tool, we checked whether it performed at least as well for high-confidence yeast/human pairs, whether predicted as one-to-one, one-to-many or many-to-many by Ensembl Compara. Since orthology relationships between human and yeast genes can be ambiguous due to their remote evolutionary distance, DCA and Ensembl Compara may predict different putative homologs, especially for the many-to-many case. For such cases, we also computed the GO accuracy as the percentage of overlapping GO labels between a yeast protein and the predicted homolog. For clear one-to-one pairs by Ensembl Compara, DCA differed in only 25 of 1040 genes. Of those 25 genes that differed, our method achieved comparable accuracy in ontology prediction (0.394) as compared to Ensembl Compara (0.388) based on ontology matching. There were 1518 entries in the "many2many" prediction category. For these, our method achieved a correct pairing (0.414) equivalent to Ensembl Compara (0.412). Finally, for the yeast genes in which a one-to-many correspondence was predicted, there were 2535 entries. Again, our method identified homologs by gene ontology (0.391) similar to Ensembl Compara (0.390). Among the differences, we observed most of them to be similar genes within the same family; moreover, these differences are not statistically significant. Thus, our approach does not disrupt homology prediction for high-confidence orthology pairs, a surrogate for false-positivity in the absence of any other gold standard yeast-to-human homolog pairing. From these results, we demonstrated that DCA provides comparable yeast-to-human accuracy as Ensembl Compara for the same input yeast genes.

Recently, Kachroo et al. (Kachroo et al., 2015) carefully tested 414 essential yeast genes for complementation by homologs that were clear by sequence. Thus, for each of these 414 yeast/human gene pairs, the complementation assay provided a binary and experimentally strong readout of homology. Kachroo et al. developed a method to predict which of these high confidence pairs were likely to be actual positive complementation pairs. They utilized more than 100 features, including careful manual curation of

sequence properties, network features, transcriptional and translational features, and expression abundances, to establish a predictive tool. They showed that this predictive tool could be trained on a subset of the experimentally tested yeast/human pairs to correctly identify functional replaceability in a separate test set. To demonstrate the effectiveness of integration of DCA, we trained a DCA-based classifier, using only sequence and network information, to predict true yeast-human complementation pairs. In particular, to check the predictive power of our DCA pipeline, we built a classifier based on the low-dimensional gene vectors obtained from our joint DCA learning pipeline. For each pair of yeast/human genes, we built features based on the gene vectors to consider their sequence similarity and topological roles in their molecular networks. These features, including element-wise product and difference and sum of two gene vectors, were used as input to a gradient boosted forest classifier. We tested whether this classifier, for our more elemental, automated DCA tool based on only sequence/network features, could be tuned to also predict the functional complementation between yeast and human. When we trained our DCA classifier via 5-fold cross-validation on the yeast/human pairs from Kachroo et al. we achieved a high rate of prediction accuracy (AUC = 0.82, SD = 0.08). This was comparable to the intricate, manual integrated method of Kachroo et al., demonstrating that our automated homology tool, based on only sequence and network topology, is sufficient for training a classifier for this specific homology task. It is worth noting that methods utilizing sequence-similarity alone, including BLAST and HHpred, performed considerably worse than DCA (0.70 and 0.69, respectively). It is clear that our DCA-based classifier, which effectively integrates network topology and sequence similarity, is just as effective as the method in Kachroo et al. that utilizes more than 100 features, thus overcoming the barrier of major time-consuming manual feature curation.

Evaluation of PCSF and Humanized Steiner Networks

We tested PCSF on two separate datasets and demonstrate vastly superior performance when compared to existing methods. For comparison, we identified two popular algorithms, DAPPLE (Rossin et al., 2011) and PEXA (Tu et al., 2009), and implemented them. Both methods take seed genes and identify subnetworks that span the seed genes to reveal possible functional interconnectedness of these genes. The first algorithm, DAPPLE, identifies significant direct and one-hop indirect edges in the human interactome to connect as many seed genes as possible. The second algorithm, PEXA, utilizes existing pathway annotations, such as KEGG or Reactome, to cover seed genes. Merging and pruning are then applied to link connected components and remove hanging genes. For these comparisons, we provided each algorithm with yeast-to-human homology links and injected yeast interaction edges into the human network, just as we provide for our PCSF method. For DAPPLE, we used the predicted dense network with significant one-hop indirect edges, since the sparse direct network is not able to identify hidden genes. We curated hits from 15 complete screens in yeast (Tong et al., 2004). In these screens, a gene is deleted as well as its genetic interactors or modifiers. We used these genetic modifiers as input for the network algorithms. The inactivated gene was hidden from the algorithm, and was used to evaluate the predicted network. Taking cues from previously-published methods (Yeger-Lotem et al., 2009), here we considered an algorithm successful in discovering the cellular response if the predicted hidden human genes were significantly enriched for specific gene ontology biological process terms attributed to the hidden inactivated yeast gene (hypergeometric test; p value < 0.01). We generated humanized networks with PCSF, and two alternative methods: DAPPLE (Rossin et al., 2011) and PEXA (Tu et al., 2009). For these screens, the success rate of PCSF was 47%, as compared to DAPPLE and PEXA which were 6.6% and 13%, respectively. These results suggest superior performance of PCSF over DAPPLE and PEXA.

To better understand the relevance of genes and predicted pathways recovered by PCSF, DAPPLE and PEXA, we designed a well-controlled simulation. To mimic genetic screens of perturbed pathways, we selected individual pathways from the well-known human pathway database KEGG and identified all genes in each pathway (Table S15). We then identified yeast homologs via stringent Ensembl one-to-one mapping. We treated those human genes with clear yeast homologs as “perturbed” and picked their homologs’ genetic interaction neighboring genes as hits from a “virtual yeast genetic screen.” Virtual screens like these minimize experimental noise as a confounding factor and enable cleaner evaluation of algorithm performance. Since we know the “true” pathway information, this method can be used to test the sensitivity and specificity of algorithms by quantifying how often “relevant” genes in the original KEGG pathway are recovered as predicted (non-seed) genes. We chose 50 KEGG pathways that had at least 5 human genes with clear yeast homologs and created 50 associated “virtual” screens for testing (Table S15). We used two performance metrics: precision, i.e., the percentage of predicted hidden genes shown in the original KEGG pathway, and recall, i.e., the percentage of the original KEGG genes shown as hidden nodes in the predicted pathway. Ideally, these values would be 100% for perfect predictions. For PCSF, the average precision and recall values are 63% and 74% resp. In contrast, for DAPPLE, the average precision and recall values are 6% and 47% resp., whereas for PEXA, they are 8% and 83% resp. The differences between three precision values are substantial: PCSF has much higher precision within very compact subnetworks, while both DAPPLE and PEXA predict huge “hair ball” networks with low precision. It is worth noting that PEXA has a very high recall value likely because it uses the KEGG pathways to build networks, and thus predictably has high recall (because the simulated screens here are generated from KEGG pathways); however, its precision metric is very low.

Further, we tested the effectiveness of injected yeast genetic interactions into networks through the simulated yeast genetic screens we generated, and cross-compare our PCSF method with the other algorithms, DAPPLE and PEXA. First, we tested performance by removing all injected yeast interactions. For PCSF, the average precision and recall values are 37% and 54% resp. For DAPPLE, the average precision and recall values are 8% and 27% resp. Compared to the precision and recall results (i.e., 63% and 74% for PCSF versus 6% and 47% for DAPPLE), it is clear that both PCSF and DAPPLE have much lower recall if yeast interactions are excluded. This analysis thus confirms with data that injection of yeast interactions into “humanized” networks provide key

connections between genetic modifiers to the perturbed genes. For PEXA, the average precision value is 9%, similar to that with yeast injection, whereas the recall rate is again predictably very high. Second, we tested the effects of randomly removing a portion of injected genetic interactions over 10 trials. The average precision and recall values are shown in Figure S4, demonstrating the relationship between the accuracy of these methods and the percentage of injected yeast interactions. A notable observation is that the performance becomes reasonable when >40% of interactions are injected. The performance of PEXA remains relatively unchanged because it utilizes the human KEGG pathway information in its algorithm, as noted above. In terms of false-positives and -negatives, there is clearly a trade-off between the different methods. PCSF works best for our current work, as PCSF identifies a small set of relevant genes for cost-effective experimental explorations.

Statistical Methods and Data Analysis for Cell-Based Assays

Sample sizes for all experimentation were chosen based on our previous extensive experience with the methods and assays in these studies. For most experiments in mammalian cells, robustness and consistency of the results are typically established after three biological replicates are analyzed. Unless otherwise stated in the figure legends, this was the standard number of replicates required for all experiments. For all human and rat cellular experiments, significance was then determined by appropriate statistical tests that are standard in the field. The two-tail t test was applied when there were only two conditions to compare within the experiments. One-Way ANOVA with a multiple comparisons post hoc test was performed when experiments include multiple conditions. Data points were excluded based on the following pre-established criteria: 1) errors were introduced to the particular sample while performing the experiments, 2) the values are greater or less than two standard deviation from the mean. For yeast spot assays, results were considered significant when three biological replicates (unless otherwise stated) demonstrated the same trend by eye. Methods used for Figure 5E are outlined in the figure legend. For the pooled screen yeast assay (Figures 4A and S6) detailed statistical methods for reads and cutoff thresholds are supplied above in the methods. The statistical methods for the computational analysis are described in detail in the methods sections above.

DATA AND SOFTWARE AVAILABILITY

All cytoscape files depicting networks are provided in the supplement and referred to in the text. The TransposeNet pipeline is described at <http://transposenet.csail.mit.edu>. The DCA/Mashup web portal is <http://mashup.csail.mit.edu>. The PCSF web portal is <http://fraenkel-nsf.csbi.mit.edu/omicsintegrator/>.

Curriculum vitae

Stephen W. Eichhorn

EDUCATION:

- Aug. 2010-May 2017 Massachusetts Institute of Technology, Cambridge, MA
Ph.D. Biology, Degree Expected 2017
Thesis advisor: David Bartel
“Regulatory consequences of mRNA poly(A)-tail length changes”
- Aug. 2006-May 2010 Reed College, Portland, OR
B.A. Biochemistry and Molecular Biology
Thesis advisor: Richard Goodman and Maryanne McClellan
“Identification of direct targets of microRNA 134, a brain-enriched microRNA”

RESEARCH EXPERIENCE:

- May 2011-May 2017 **Graduate Studies, MIT, Cambridge, MA**
Advisor: David Bartel
Regulatory consequences of mRNA poly(A)-tail length changes
- Summer 2010 **Summer Fellow, Vollum Institute, Portland, OR**
Advisor: Richard Goodman
Role of microRNA 134 during neuronal development
- Aug. 2009-May 2010 **Undergraduate Thesis, Reed College and Vollum Institute, Portland, OR**
Advisor: Richard Goodman and Maryanne McClellan
Biochemical identification of direct targets of microRNA 134

TEACHING EXPERIENCE:

- Spring 2014 Teaching Assistant, Physical Chemistry (7.10), MIT, Cambridge, MA
Spring, 2012 Teaching Assistant, Biotechnology II (7.16), MIT, Cambridge, MA
Spring 2009 Teaching Assistant, Organic Chemistry, Reed College, Portland, OR

AWARDS:

- May 2010 “Class of ’21” Award for Best Undergraduate Thesis, Reed College, OR
May 2010 Phi Beta Kappa

PUBLICATIONS:

Khurana, V., Chung, C.Y., Peng, J., Auluck, P.K., Tardiff, D.F., Fanning, S., Bartels, T., Koeva, M., **Eichhorn, S.W.**, Benyamini, H., Lou, Y., Nutter-Upham, A., Baru, V., Freyzon, Y., Tuncbag, N., Costanzo, M., San-Luis, B., Schöndorf, D.C., Barrasa, M.I., Ehsani, S., Sanjana, N., Zhong, Q., Gasser, T., Bartel, D.P., Vidal, M., Deleidi, M., Boone, C., Berger, B., Fraenkel, E., Lindquist, S. Genome-scale networks link diverse molecular pathways and neurodegenerative disease genes to alpha-synuclein. *Cell Systems*, 4, 157-170.e14.

Eichhorn, S.W., Subtelny, A.O., Kronja, I., Kwasnieski, J.C., Orr-Weaver, T.L., and Bartel, D.P. (2016). mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila* oocytes and early embryos. *eLife* 5:e16955.

Weinberg, D.E., Shah, P., **Eichhorn, S.W.**, Hussmann, J. Plotkin, J., and Bartel, D.P. (2016). Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep* 14, 1787-1799.

Eichhorn, S.W.*, Guo, H.*, McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S-h., Ghoshal, K., Villén, J., Bartel., D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell* 56, 104-115. **equal contribution*

Ludwig, L.S., Gazda, H.T., Eng, J.C., **Eichhorn, S.W.**, Thiru, P., Ghazvinian, R., George, T.I., Gotlib, J.R., Beggs, A.H., Sieff, C.A., Lodish, H.F., Lander, E.S., Sankaran, V.G. (2014). Altered translation of GATA1 in Diamond-Blackfan anemia. *Nat Med* 20, 748-753.

Kronja, I., Yuan, B., **Eichhorn, S.W.**, Dzek, K., Krijgsveld, J., Bartel, D.P., and Orr-Weaver, T.L. (2014). Widespread changes in the posttranscriptional landscape at the *Drosophila* oocyte-to-embryo transition. *Cell Rep* 7, 1495-1508.

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, 66-71. **equal contribution*

Chai, S., Cambronne, X.A., **Eichhorn, S.W.**, and Goodman, R.H. (2013). MicroRNA-134 activity in somatostatin interneurons regulates H-Ras localization by repressing the palmitoylation enzyme, DHHC9. *Proc Natl Acad Sci U S A* 110, 17898-17903.

ORAL PRESENTATIONS:

Eichhorn, S.W., Subtelny, A.O., Kronja, I., Orr-Weaver, T.L., and Bartel, D.P. mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in *Drosophila* oocytes and early embryos. Talk given at Translational Control, Cold Spring Harbor, NY, 2016.