Form and Function of Poly(A) Tails

by

Timothy J. Eisen

Sc. B. Chemistry (2011)
Brown University

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 2020

Signature of Author: _____

Timothy J. Eisen
Department of Biology
April 17, 2020

Certified by: _____

David P. Bartel
Professor of Biology
Thesis Supervisor

Accepted by: _____

Stephen P. Bell
Uncas and Helen Whitaker Professor of Biology
Co-Chair, Biology Graduate Committee

Form and Function of Poly(A) Tails

by

Timothy J. Eisen

Submitted to the Department of Biology on April 17, 2020
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Abstract

Central to mRNA metabolism is the poly(A)-tail, a stretch of adenosine nucleotides at the mRNA 3′ end. In this dissertation, I investigate the role of the tail in the dynamics of mRNA decay, and describe the predominant mechanisms of decay for thousands of mammalian mRNAs. Next, I examine the effects of microRNAs, which influence mRNA decay and perturb tail length dynamics. Finally, I describe a physiological context in which the tail helps to control translation: neurons of the mouse brain.

mRNA decay is tightly regulated in eukaryotes, determining the steady-state abundances and rates of accumulation of mRNAs. Despite this central role, the dynamics of decay have been described for only a handful of mRNAs. We determine these dynamics for thousands of endogenous mRNAs. Nascent mRNAs have reproducible and heterogeneous tail lengths just after they escape the nucleus. Once in the cytoplasm, most mRNAs are substrates for deadenylation, the rates of which vary by over 1000-fold, a range sufficiently large to capture the variation in mRNA decay rates. Surprisingly, once their tails become short, mRNAs decay at rates that also span a 1000-fold range. Moreover, these rates are coupled to their deadenylation rates, suggesting a concerted process of remodeling the mRNA–protein complex during decay.

MicroRNAs (miRNAs) are small RNAs that influence decay of mRNA targets by recruiting deadenylases. Despite this recruitment, we observe no changes to steady-state tail length for miRNA targets. Resolving this paradox, we find that miRNAs not only deadenylate their targets but also increase the decay rate of short-tailed target molecules. By enhancing both rates, miRNAs do not alter the distribution of tail lengths of target mRNAs but enhance the rate at which mRNAs traverse these lengths.

Neurons have unique requirements for translational control. We perform ribosome profiling in primary neuronal cultures and brain tissues from a mouse. mRNA poly(A)-tail lengths explain some (~5%) of the large variance in translational efficiency we observe, as does coding-sequence length, expression level, and codon composition. For some mRNAs, neuronal stimulation modifies tail lengths, and for a subset, transcription cannot explain these changes. A linear model that uses known determinants to predict translational efficiency explains only a portion (30–40%) of its variance, indicating the need for additional investigation of mechanisms of translation in neurons.

Thesis supervisor: David P. Bartel
Title: Professor of Biology

Acknowledgements

This dissertation is a product of many hours of experimentation, analysis, and contemplation. Not all of those hours reflect work that I performed, however, and I am honored to describe the contributions of the individuals who made this work possible.

My advisor Dave has mentored me throughout my graduate work. Sometimes, his guidance is direct, suggesting specific analyses or experiments. More often, though, he patiently discusses the interpretation of results with me, allowing both of us time to consider the most fruitful path forward. This careful approach to scientific inquiry allows wonderful discoveries to slowly become clear. When I started in the lab, I was impatient to generate data and interpret results, but in my attempts to emulate Dave, I have tried to learn the focus and precision required of a successful scientist.

Members of my thesis committee have provided support and direction for this work. Chris Burge and Gene-Wei Li, through their detailed questions during our meetings, have steered my modeling and analysis. Wendy Gilbert's prescience was always appreciated and I hope to develop this kind of scientific intuition, but so far it has eluded me. Finally, I thank Myriam Heiman. Myriam was my PI when I was a technician at the Broad and Picower institutes, and she has continued to mentor me throughout my graduate career, providing thoughtful advice. I've also had the opportunity to follow her lab from its beginnings and watch as it makes exciting discoveries.

My research is extremely collaborative. I had the privilege to work on two of the studies in this dissertation alongside two close friends and mentors, Alex Subtelny and Stephen Eichhorn. Stephen taught me the basics of RNA molecular biology and continues to inspire my work with his creativity. My scripts, data analysis, and imagination, would have been sluggish and foggy had it not been for two additional mentors, Sean McGeary and Kathy Lin. Sean taught me not to be afraid of the many areas of mathematics that might prove powerful and elegant solutions to biological problems. Kathy always helped me to think about my research creatively. I'd also like to thank Jamie Kwasnieski, my baymate for six years, who taught me to use computation and experiments side by side. We discussed our work endlessly, and the areas of this dissertation that benefited from her insights are innumerable. My collaborators outside the lab have also expanded my knowledge, interest, and excitement: Peter Bruno, Guy Bushkin, Lukas Chmatal, Erinc Hallacli, Sunny Kang, David Pincus, and Peter Tsvetkov.

I'd like to acknowledge members of the Bartel lab who have made the lab a wonderful place to work: Vikram Agarwal, Emir Avilés Pagán, Danny Briskin, Namita Bisaria, Grace Chen, Wenwen Fang, Matt Getz, Justin Gullingsrud, Glenn Li, Dan Lin, Gina Kelley, Elena Kingston, Ben Kleaveland, Thy Pham, Charlie Shi, Jarrett Smith, Michael Stubna, Peter Wang, and Coffee Xiang. Laura Resteghini has been enormously helpful administratively over the years, and I always enjoy when we share our latest travel adventures. Finally, Asia Stefano, our lab manager, has kept our lab together (both its equipment and people).

Many of the experiments in this dissertation involve high-throughput sequencing. This would not have been possible without the help of the Whitehead Genome Technology Core: Amanda

Chilaka, Stephen Mraz III, and Sumeet Gupta. Sumeet is exceptional in his knowledge of this set of experiments and his skill in reconfiguring Illumina sequencers for custom runs. Most importantly, he's always happy to discuss details of protocols and to try new experiments, many of which fail. The days that we try new types of sequencing are some of my favorite.

In addition to my scientific friends and collaborators, I'd also like to thank my roommates and friends for their support, particularly Matthew Mandelkern, Jonathan Phillips, Alissa Groisser, and Adrian Leanza.

My family has nurtured my interest in science for a long time, and I've been fortunate enough to have many scientific role models growing up. My parents are always excited to hear about what I've been working on, and have supported this work in many ways. My mom also discusses her science with me, which is a wonderful family tradition. Julia, my sister, is also a researcher, now pursuing medical training, and we've begun to discuss our science too. All four of my grandparents have also been very supportive and loving over the years. When I began my graduate work, I met weekly with my grandfather Herman to discuss our recent scientific interests. His words of encouragement were essential when all of my early projects were failing. Since then, I've shared the company of my grandmother Natalie and her curiosity. Finally, I would like to thank my partner Christi for her constant love. I look forward to our new lives together on the West Coast.

Table of Contents

Chapter 1. Introduction

Timothy J. Eisen[1,2,3]

[1]Howard Hughes Medical Institute, Cambridge, MA, 02142, USA
[2]Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

RNA Decay is Regulated in Eukaryotes

Gene expression awakens the genome. RNA, produced from genes, functions either as a vehicle for protein production, as is the case for mRNAs, or as the final functional output of the gene, as is the case for non-coding RNAs. In all cases, however, the abundance of the RNAs that are produced matters for their eventual function. Theory of chemical equilibria informs us that such abundance must be determined both by the rates of production and decay. If considering only steady-state abundance, both rates play crucial roles, but in the approach to steady-state, the rate at which these abundances change generally depends on decay rate, as RNAs that are long lived must first decay before they can be replaced.

Despite these first-principles assertions, one of two regimes might describe the biology of gene expression. In the first regime, the rates of decay for all RNAs have a narrow distribution, one that explains only a small fraction of the overall variation in RNA levels. The second regime has a uniform RNA production rate but a large variance in decay rate. These regimes make different predictions for cellular control of gene expression because the magnitude of the change in gene expression brought about by tuning a particular node differs based on the regime. In eukaryotic cells, evidence for regulation of production gave rise to the field of transcriptional control. Later, RNA decay rate was also found to vary, sometimes drastically, in different contexts and for different RNAs.

Evidence for regulation of RNA decay in eukaryotic systems mounted from early studies using induction systems for mRNAs. Kinetic measurements are required to disentangle the contribution of production and decay. Historically, these measurements could only be obtained for mRNAs that exhibited changes in expression in response to perturbation. In one such system, the *casein* mRNA was induced by addition of the hormone prolactin in cultured mammary gland

tissue from a rat (Guyette et al., 1979). Upon hormone addition, the transcription rate of this mRNA increases 2–4 fold, but this increase is not sufficient to explain the overall mRNA accumulation. The additional change in mRNA decay rate causes the half-life of *casein* to increase from 5.5 h to 92 h, far exceeding the fold change in transcription.

Subsequent studies began to dissect the mechanisms of mRNA turnover. For these studies, another induction system, involving the proto-oncogene *c-fos*, became paramount. *c-fos* is undetectable in cells from the mouse line NIH 3T3 when they are starved of serum. After re-addition of serum, the mRNA rapidly accumulates and can be detected with RNA blotting (Kruijer et al., 1984; Muller et al., 1984; Greenberg et al., 1986). Using this induction system, two groups were able to manipulate the rapid turnover of either heterologous human *c-fos* (Shyu et al., 1989) or mouse *c-fos* (Wilson and Treisman, 1988). The studies showed that AU-rich elements were important for the instability of the transcript, and insertion of an AU-rich element into an otherwise stable reporter mRNA could decrease its half-life (Shyu et al., 1989). The sufficiency of this element, and its location within the 3′ untranslated region (UTR) of the transcript, suggested a regulatory landscape of factors that influence mRNA decay rate through recognition of specific mRNA features.

Features of mRNAs That Relate to Decay

A eukaryotic mRNA contains nucleotide sequences that serve as a template for protein production, but these sequences are not enough for the mRNA to persist in the cell and recruit the translation machinery (Figure 1). Stability and translation are controlled in part by modifications at both the 5′ and 3′ ends of the mRNA. The two main modifications are the 7-methylguanosine cap at the 5′ end, connected to the second nucleotide by a noncanonical 5′–5′

triphosphate linkage, and the poly(A)-tail at the 3′ end, the latter of which will be the subject of this dissertation.

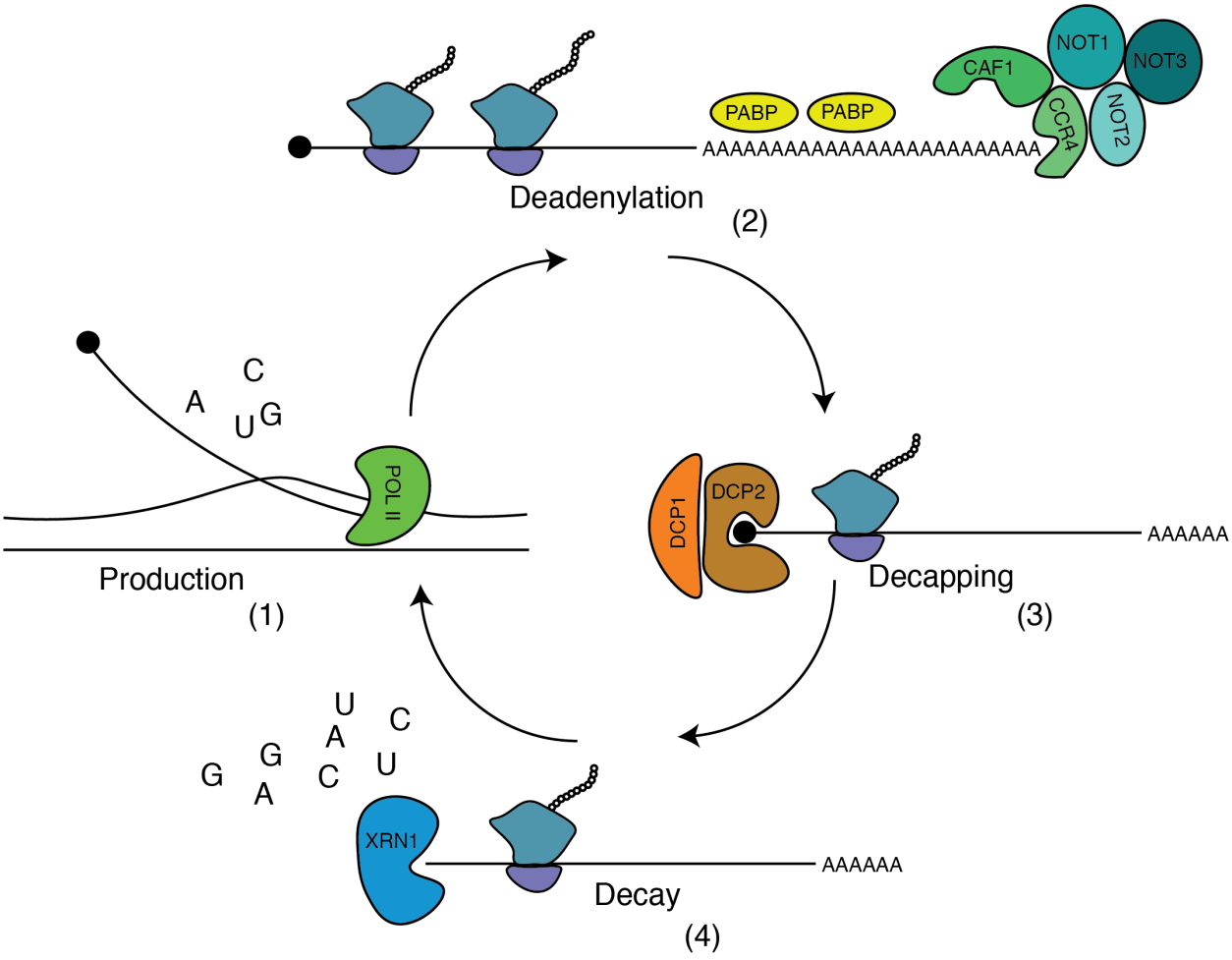


Figure 1. The Lifecycle of an mRNA
mRNAs are born in the nucleus through the processes of transcription, capping, splicing, cleavage, and polyadenylation (1). After export into the cytoplasm, the poly(A) tail is removed by the PAN2–PAN3 and CCR4–NOT complexes, with the CCR4–NOT complex being more consequential (2). Once the tail becomes short, the 5′ cap is removed by the DCP1–DCP2 complex, exposing a 5′ monophosphate (3). The loss of the cap makes the mRNA a substrate for XRN1, a 5′–3′ processive exonuclease that degrades the body of the mRNA (4). Some alternative decay pathways, such as degradation mediated by the LSM1–7 exosome complex, also require deadenylation.

Early reports identified an adenylate-rich stretch of RNA in mRNAs in the calf thymus and HeLa cells (Edmonds and Abrams, 1963; Darnell et al., 1971) and found it to be associated with mammalian ribosomes. A role for the tail in stability emerged later, however, in reports

showing that the length of the poly(A)-tail tracked with the age of bulk mRNA in both mammals (Sheiness and Darnell, 1973) and the slime mold *Dictyostelium* (Palatnik et al., 1979).

Subsequent studies of the poly(A)-tail shed light on mechanisms of mRNA decay in yeast. Examination of predominantly two mRNAs, *MFA2* and *PGK1* (Lowell et al., 1992; Muhlrad and Parker, 1992; Decker and Parker, 1993; Muhlrad et al., 1994), revealed that mRNAs are degraded in three phases (Figure 1). During the first phase, the tail, which is ~60 nt shortly after mRNAs enter the cytoplasm, shortens to ~10 nt. This process, known as deadenylation, involves the two exonuclease complexes CCR4–NOT and PAN2–PAN3, both of which are specific for adenosine, with CCR4–NOT being more consequential (Meyer et al., 2004). After the tail becomes short, poly(A)-binding protein (PAB1 in yeast or PABP in mammals) can no longer bind to the tail and its loss marks the second phase of decay in which the DCP1–DCP2 complex removes the 5′ cap in a process known as decapping (Chowdhury et al., 2007). In the third phase, the 5′ monophosphate, generated from decapping, allows the mRNA to become a substrate for XRN1 and perhaps other 5′–3′ exonucleases that degrade the mRNA body but have little activity on the cap structure. Some mRNAs are predominantly degraded by 3′–5′ exonucleases such as the cytoplasmic exosome (composed of a ring-like complex of LSM1–7 proteins), but this degradation also depends on tail-shortening, as the complex has little activity on adenylate-rich sequence (Schmid and Jensen, 2008).

Like those in yeast, mRNAs in mammals enter the cytoplasm with long tails that are then shortened over time (Yamashita et al., 2005). The PABP footprint in mammals is slightly larger than in yeast (Baer and Kornberg, 1983), and the tail length at which mRNAs are decapped and degraded spans the larger range of 10–60 nt. But despite similarities, more detailed studies of the

three-phase model in both yeast and mammals indicated differences between species and led to more precise but complex observations.

A major challenge in dissecting the model of decay lies in the relationship between the loss of PABP binding and decapping, which requires cross talk between the two ends of the mRNA. PABP antagonizes the decapping machinery by shielding it from decapping activators (Caponigro and Parker, 1995; Mangus et al., 2003; Goldstrohm and Wickens, 2008), proteins which can modulate the rate of decapping over three logs (Wurm et al., 2017). Further protection of the cap structure by the translation initiation complex, eIF4F, and the interaction between this complex and PABP, also modulate decay rates in some contexts and for some mRNAs. The relationship between translation initiation and mRNA decay, and an additional relationship between translation elongation and decay, will be discussed in more detail in subsequent sections.

Despite the nearly universal poly(A) stretch at the 3′ termini of eukaryotic mRNAs, mRNAs from replication-dependent histone genes do not have a tail (Mullen and Marzluff, 2008). Instead, their export, translation, and decay are coordinated by a conserved stem loop structure in their 3′ UTR (Marzluff et al., 2008), and their expression is synchronized and coincides with maximal production at the beginning of the S phase of the cell cycle. Just as PABP binds to the tail and coordinates translation and decay, stem loop binding protein (SLBP) binds to the stem loop and, through analogous pathways, coordinates translation by indirect recruitment of eIF4G using the adaptor SLIP1 (Cakmakci et al., 2008). Decay of the histone mRNAs at the end of S phase is also mediated by SLBP, which recruits terminal uridyl transferases that that add uridines to the mRNA 3′ ends (Zheng et al., 2003).

Terminal uridylation can be detected on mRNAs that do not code for histones as well. Terminal uridyl transferases generally add 1–10 uridines to the ends of mRNAs that either lack a tail or only have a short one (≲ 20 nt) (Chang et al., 2014; Lim et al., 2014). While details of the function of this modification await further inquiry, studies indicate that it stimulates decapping in a number of species, perhaps through recruitment of decapping activators (Mullen and Marzluff, 2008; Rissland and Norbury, 2009; Morozov et al., 2010; Chang et al., 2014; Lim et al., 2014). This stimulatory role is in keeping with observations that uridylation occurs mostly on mRNAs that are rapidly degraded (Rissland and Norbury, 2009; Morozov et al., 2010; Lim et al., 2014).

Decay-Rate Influencers

The canonical mRNA decay pathways collectively describe the turnover of mRNAs after their function in the cytoplasm is complete. But different mRNAs decay at different rates despite transiting through the same general pathways. The rates of each of the three nodes of decay (deadenylation, decapping, and decay of the mRNA body), can be tuned, sometimes changing the rate-determining steps of the overall process. Some mRNAs decay according to mechanisms bearing little resemblance to the general mechanism.

Early investigation of modulators of mRNA decay rate centered on AU-rich elements, found in the 3′ UTR and coding sequence of *c-fos* and other mRNAs (Wilson and Treisman, 1988; Shyu et al., 1989). Many RNA-binding proteins (RBPs) can bind these sequences, including tristetrapolin (TTP), a protein involved in the inflammatory response (Brooks and Blackshear, 2013). A short segment of the C terminus of TTP binds CNOT1, a subunit of the CCR4–NOT complex (Fabian et al., 2013). This interaction is necessary for the TTP-mediated

decay of target mRNAs. Indeed, recruitment of the CCR4–NOT complex is a common mechanism of modulating decay (Figure 2).

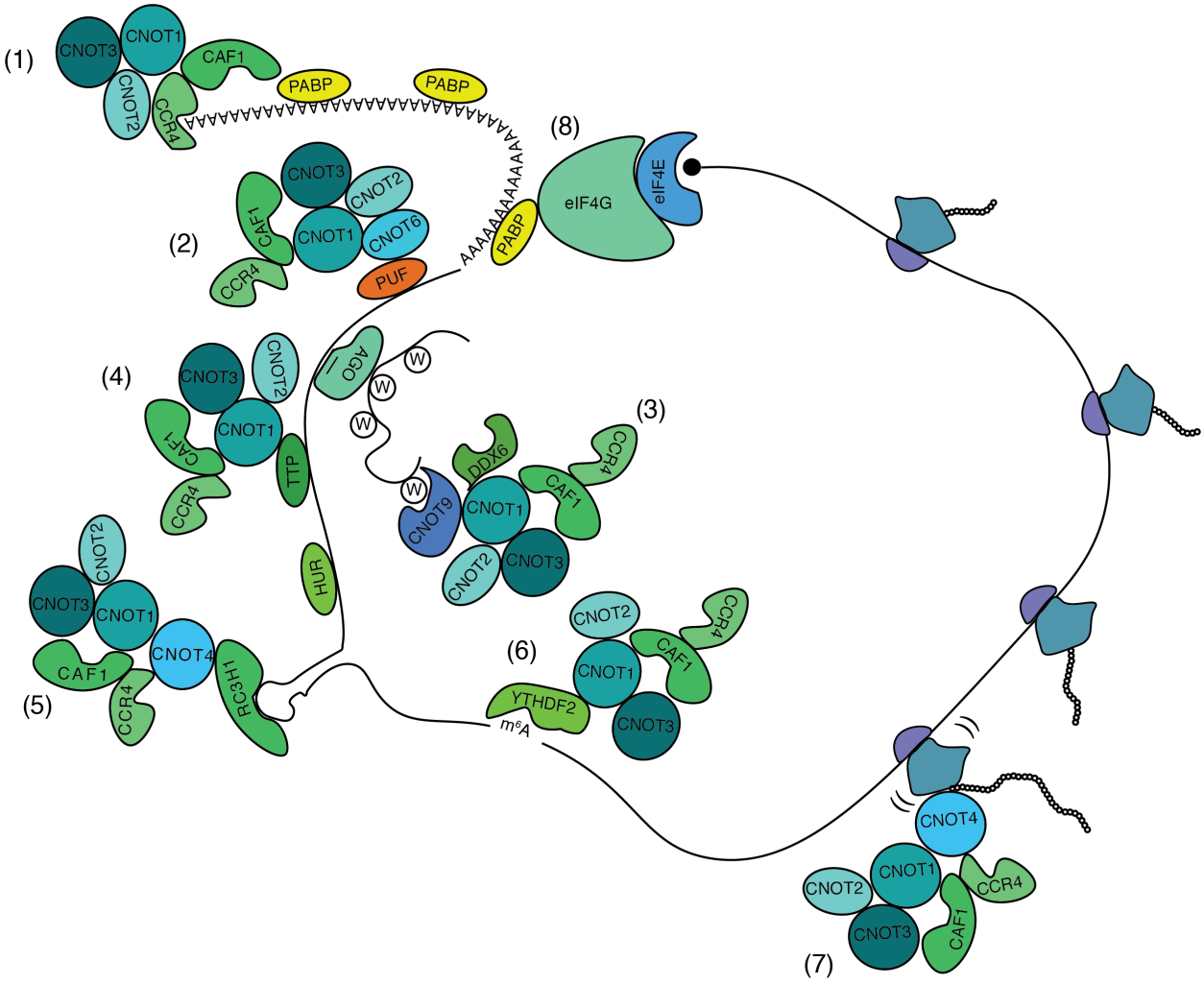


Figure 2. The Integrated Recruitment of Deadenylases
Regulation of mRNA deadenylation. Many cellular factors that regulate mRNA decay do so through recruitment of the deadenylase complexes, most often the CCR4–NOT complex (1). These regulators include the PUF proteins (2), microRNAs (3), TTP (4), Roquin (5), and YTHDF2 (6), among others. Ribosome pausing, if it results in ribosomes with both empty A and E sites, can recruit the deadenylase complex as well (7). The mRNA is shown as a loop because of the interactions between PABP and the eIF4F initiation complex (8). Protein names are from human.

Other examples of proteins that recruit deadenylases have also been described. These include Pumilio family (PUF) proteins (Van Etten et al., 2012), SMG5/7 (Muhlemann and Lykke-Andersen, 2010), GW182 (Fabian et al., 2011), BTG/TOB factors (Mauxion et al., 2009),

Roquin (Leppek et al., 2013), YTHDF2 (Du et al., 2016), and factors that bind GU-rich elements (Vlasova-St Louis and Bohjanen, 2011; Fabian et al., 2013). As most of these proteins directly interact with one or more subunits of the CCR4–NOT complex, it is tempting to hypothesize that it is the integration of these individual protein components, either through addition of discrete units of decay machinery or increased dwell time of the decay machinery, that create the broad landscape of mRNA decay (Figure 2). While most mechanisms of mRNA decay that have been described rely on factors that degrade the mRNA, there are also protective proteins, in addition to PABP. HuR has dual roles in stability and decay, for example, and in some contexts can antagonize the deadenylase complexes, leading to stabilization of AU-rich or AU-element containing mRNAs (Fan and Steitz, 1998).

A number of additional pathways have been described that regulate aberrant mRNAs in the cytoplasm or mRNAs that have been incorrectly processed in the nucleus. The best studied of these is nonsense-mediated decay (NMD). Classically, NMD targets mRNAs that contain a mutation in their coding sequence that aberrantly inserts a stop codon (premature termination codon or PTC), although mRNAs that are constitutive targets of NMD have also been described (Lykke-Andersen and Jensen, 2015). In yeast, the PTC is distinguished from a standard stop codon because of its distance from the poly(A)-tail (Amrani et al., 2004), but in mammalian transcriptomes there are additional mechanisms that contribute to this recognition. Splicing of the pre-mRNA in the nucleus deposits an exon-junction complex (EJC). The EJC aids in mRNA export to the cytoplasm but is removed in the first round of translation by the translocation of the ribosome. As an aberrant stop codon will cause translation to stop prematurely, downstream exon-exon junctions will still harbor EJCs, and the persistence of these EJCs causes recruitment of the NMD factors UPF1–3 (Lykke-Andersen and Jensen, 2015). Decay of the mRNA need not

coincide with the first round of translation, however (Hoek et al., 2019). Additional mRNA features are required for efficient NMD, including 3′ UTR length, number of remaining EJCs, and sequence composition (Hogg and Goff, 2010; Hoek et al., 2019).

Results are less consistent in pointing towards a definitive mechanism for degradation of the mRNA once it has been recognized as an NMD substrate. NMD, and a related pathway orchestrated by the protein Staufen that also requires UPF1, known as Staufen-Mediated Decay (SMD) (Park and Maquat, 2013), can proceed by recruiting the endonuclease SMG6, which cleaves the body of the mRNA near the PTC. This cleavage generates two fragments, each of which is a substrate for exonucleases. In a distinct but perhaps not mutually exclusive mechanism, recognition of the PTC results in recruitment of deadenylases (through the heterodimer SMG5– SMG7), enhancing mRNA decay through the general pathway of deadenylation followed by decapping and 5′–3′ exonuclease activity (Lykke-Andersen and Jensen, 2015). Additional NMD-mediated decay pathways may involve direct recruitment of the decapping machinery, as has been suggested because of interactions between UPF1 and DCP2, and additional interactions between UPF1 and a decapping-adaptor protein, PNRC2 (Loh et al., 2013). Whether certain transcripts are biased towards one of these decay pathways, and how this bias is determined, remains less clear.

MicroRNA-Mediated Repression

MicroRNAs (miRNAs) are one of the best-studied influencers of decay. These RNAs are small, 21–24 nt guides that direct the recruitment of the Argonaute protein (AGO) to target mRNAs (Bartel, 2018). MicroRNAs are conserved. One miRNA, miR-100, has been present throughout Eumetazoa, tracing its origins back prior to the split between Bilateria and Cnideria, over 600

million years ago (Grimson et al., 2008). Other extant miRNAs were present in early Bilateria. Interactions between miRNAs and their targets are also conserved: 90 miRNA families were present in the last common ancestor of human and fish, and each of these families has an average of >400 conserved target interactions, such that in aggregate these conserved miRNAs target over half of all protein-coding genes (Friedman et al., 2009). Furthermore, genetic studies have shown a range of phenotypes associated with miRNA knockouts in mice. Knockout of the miR-196 family, for example, results in changes to body plan, such as vertebral number, while knockout of the miR-181 family results in embryonic lethality (Fragoso et al., 2012; Wong et al., 2015; Bartel, 2018).

MicroRNAs function to repress the protein output of their targets by base-pairing interactions between the 5′ region of the miRNA, known as the seed (nucleotides 2–7), and the 3′ UTR of the mRNA. The affinity of this interaction specifies the extent of repression of target mRNAs (McGeary et al., 2019). Because these pairing interactions rely on sequence complementarity, they are generally more specific than interactions between RNAs and RNA-binding proteins (RBPs), as RBPs tend to have many mRNA targets containing binding motifs of low-information content (Dominguez et al., 2018). Despite this specificity, miRNAs tend to repress their targets modestly, with each interaction often affecting steady-state expression by less than 2-fold, leading to a view in which they sculpt, rather than silence, parts of the transcriptome (Bartel, 2009).

The features of this repression have been the subject of debate, although in mammalian cell lines, several studies have converged on a predominant mechanism. MicroRNA-mediated repression begins with the interaction between AGO and the target mRNA, mediated by the miRNA. AGO interacts with an additional protein, TNRC6 (GW182), which recruits the PAN2–

PAN3 and the CCR4–NOT deadenylase complexes (Chen et al., 2009; Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011; Christie et al., 2013; Huntzinger et al., 2013). Through this recruitment, miRNAs accelerate deadenylation of reporter mRNAs (Giraldez et al., 2006; Wu et al., 2006; Chen et al., 2009) and endogenous ones (Bazzini et al., 2012; Subtelny et al., 2014). The deadenylase complexes, in turn, interact with components of the decapping complex, such as EDC4 (Jonas and Izaurralde, 2015). This form of repression, proceeding with mRNA decay, accounts for most of the observed repression in post-embryonic cells (Guo et al., 2010; Eichhorn et al., 2014).

MicroRNA-mediated translational repression, a second form of repression, was identified earlier, and for many years was considered the predominant mode of miRNA action. Mechanisms of translational repression involve additional players and may be disparate. An important protein is DDX6, an RNA helicase that can enhance decapping and bind to the CCR4–NOT complex (Chen et al., 2014). Loss of DDX6 de-represses miRNA targets and some reporter mRNAs (Chu and Rana, 2006; Mathys et al., 2014). As a result of decapping, eIF4E is lost, resulting in decreased initiation. The deprotected mRNA is subsequently prone to decay, however, which may explain the linked nature of the two repression mechanisms, and the observation of an early phase of translational repression prior to decay (Eichhorn et al., 2014). In embryonic systems that do not exhibit mRNA decay, perhaps because of lack of decapping activity, alternative mechanisms of miRNA-mediated translational repression have been described and will be discussed in the next section.

Translation and Poly(A)-Tail Length

The poly(A)-tail is involved in mRNA translation as well as decay. Through the interaction of PABP and eIF4G, a scaffolding component of the eIF4F complex, the tail can influence activities surrounding the cap, such as decapping and translation initiation (Figure 2). The potential for interactions between the 5′ and 3′ ends of the mRNA has given rise to the closed-loop model of mRNA translation, which attempts to explain the observation that reporters lacking tails are translated poorly (Wakiyama et al., 2000). This model gained traction when direct interactions were observed between proteins known to interact with 5′ and 3′ mRNA termini (Borman et al., 2000). A study using atomic force microscopy showed that mRNAs can form circles in vitro, and circularization requires PABP and components of the initiation machinery (Wells et al., 1998). Despite these reports, direct evidence of closed-loop formation in vivo is lacking, and some literature casts doubt on the necessity of the closed loop for translation, or the phenomenon of circularization more generally. Depletion of eIF4G has only modest effects on translational efficiencies in yeast and the PABP–eIF4G interaction is not essential (Park et al., 2011a; Park et al., 2011b). In addition, a report that used fluorescent in situ hybridization (FISH) to label mRNA termini showed that mRNAs are more compact in the absence of ribosomes (upon addition of puromycin) than in their presence (Adivarahan et al., 2018), a paradoxical finding if ribosomes preferentially translate circularized mRNAs. This observation confirmed the notion that the loop is too simple a descriptor of the complex structural rearrangements that must occur during remodeling and translation of the mRNA–protein complex (mRNP).

Despite questions that remain concerning the conformation of an mRNA during translation, and the requirement of the poly(A)-tail for this conformation, poly(A)-tail length can determine an mRNA's translation in the specific contexts of some embryos and oocytes. A study

in the *Xenopus laevis* oocyte showed that translation of *c-mos* can be manipulated by removing the mRNA's endogenous tail and replacing it with a tail in trans by hybridizing a synthetic, injected RNA (Barkoff et al., 1998). The amount of protein produced, and the percentage of oocytes that underwent germinal vesicle breakdown in response to progesterone (a process for which *c-mos* is necessary and an indicator of meiotic maturation), increased as the tail length varied from 30 to 130 nt, leveling off at longer tail lengths. In another study in *Drosophila* embryos, injection of *bicoid* mRNA was only able to rescue mutant phenotypes when the injected mRNA had a tail length of 175 nt, confirming the requirement for cytoplasmic polyadenylation in translation of the message (Salles et al., 1994). The mRNA *nanos* does not undergo tail lengthening despite its translational regulation in the embryo, however (Salles et al., 1994).

Such studies showing an intimate relationship between tail length and translational efficiency suggested that coupling of the two might be more widespread, affecting more than just a handful of mRNAs, and extending beyond the embryonic and late-stage oocyte systems. Indeed, reports claim to observe such coupling in yeast (Beilharz and Preiss, 2007) and human fibroblasts (Burns and Richter, 2008). However, a more recent study combined ribosome profiling, which measures translational efficiencies of mRNAs from each gene, and PAL-seq (poly(A)-tail length sequencing or PAL-seq), a high-throughput tail-length profiling approach developed for the study, to investigate this coupling transcriptome-wide and across many contexts (Subtelny et al., 2014). These experiments and analyses confirmed and generalized a relationship between tail-length and translation in the early embryos of *Danio rerio* and *Xenopus laevis*, but could not find such coupling in other systems, including cell lines, yeasts, and mouse liver. Strikingly, the coupling decreased as the embryos aged: by 6 hours post fertilization (hpf)

in *D. rerio*, the dynamic range of translational efficiencies decreased, and tail length had little influence on translation. This coupling between translation and tail length was also observed in *Drosophila* ooyctes and early embryos (Eichhorn et al., 2016).

The embryonic switch in the nature of translational control, and the changing role of the poly(A) tail during this period in early development, suggested an explanation for previous inconsistencies concerning the mechanism of miRNA-mediated repression. Experiments in *D. rerio* embryos deficient for both the maternal and zygotic copies of *dicer*, a gene required for the biogenesis of miRNAs, showed that a miR-430, a highly expressed and developmentally regulated miRNA, causes translational repression that gives way to mRNA destabilization by 6 hpf (Bazzini et al., 2012). This stage of translational repression coincides with the period in which *D. rerio* embryos exhibit tail-length coupling (Subtelny et al., 2014). Indeed, injection of the exogenous miR-155 into the one-cell *D. rerio* embryo reproduced the early phase of translational repression at 2 hpf but showed dramatic tail shortening during this time. Both the translational repression and tail-shortening abate by 6 hpf (Subtelny et al., 2014). These experiments suggested that the immediate effect of miRNAs in the early and late stage embryos is the same: in both contexts, they act to repress by recruiting deadenylases. But the outcome of this recruitment depends on the role of the poly(A)-tail and whether its function is to control translation or decay. Additional questions remain, however, particularly concerning the influence of tail-length on mRNA decay. If miRNAs recruit deadenylases in the *D. rerio* 6 hpf embryo, it is unclear how they do so without shortening tails.

Codon Optimality in Yeast and Other Species

In theory, regulation of translation by either initiation or elongation might influence the amount of protein made from a message similarly. In practice, the magnitude of the effect depends on which step in translation is rate limiting, or whether repression can change which step is rate limiting. Classically, translation has been considered an initiation-limited process (Jackson et al., 2010), with features such as structure of the 5′ UTR and coding-sequence length most predictive of efficiency in yeast (Shah et al., 2013; Weinberg et al., 2016). Notable exceptions to this initiation-limited regime include regulation by FMRP in neurons (Qin et al., 2005; Darnell et al., 2011; Das Sharma et al., 2019), disease contexts that modify or mutate tRNAs (Ishimura et al., 2014), and amino-acid depletion (Wu et al., 2019a). Perturbations to the rate of translation elongation can have multiplicative effects on protein output, with recent observations linking the codon optimality of a message to its decay rate.

Codon optimality relies on the observation that the genetic code allows for redundancy, with multiple codons encoding the same amino acid. The cellular concentrations of the tRNAs that decode these codons help to dictate how readily a translating ribosome will have access to the correct decoder, and a transcript that mostly uses codons corresponding to highly abundant tRNAs is considered optimal. In 2015, Presnyak and colleagues, working in *S. cerevisiae*, generated measurements of mRNA decay rates using non-poly(A)–selected mRNA and used these measurements to show a relationship between a transcript's codon optimality and its half-life (Presnyak et al., 2015). In a striking series of experiments, the investigators showed that the decay rate of a reporter mRNA can vary by more than 10-fold depending on its percentage of optimal codons.

Subsequent studies expanded the scope of the link between mRNA decay and codon optimality. During the maternal-to-zygotic transition in *X. laevis*, *D. melanogaster*, *M. musculus*, and *D. rerio*, transcripts containing non-optimal codons are more destabilized than their optimal counterparts (Bazzini et al., 2016). In a human cell line, reporter mRNAs containing different coding sequences but the same UTRs exhibited a range of mRNA decay rates comparable to those of endogenous human mRNAs, although with little correlation between the endogenous and exogenous counterparts (Forrest et al., 2018; Wu et al., 2019b). Two studies observed that decay-rate effects might be due to differences in amino acid composition, not codon composition (Forrest et al., 2018; Hanson et al., 2018), as steric or electrostatic interactions between the amino acids and the ribosome-exit channel induce translational pauses (Lu and Deutsch, 2008; Kostova et al., 2017).

Mechanistically, the model describing the influence of codon optimality requires at least two lines of support. The first must provide evidence that codon composition affects translation elongation rate. The second must show that paused ribosomes engender mRNA decay. With regard to the first support, large changes in tRNA abundance, induced by amino acid starvation or by genetic ablation, can induce ribosome pauses at the cognate codons (Ishimura et al., 2014; Wu et al., 2019a). In physiological contexts, pausing is more difficult to observe, and studies have provided clues in the form of differential codon occupancy (Ingolia et al., 2011; Das Sharma et al., 2019). These pauses correlate inversely with tRNA abundance in yeast but not in mammalian cells (Ingolia et al., 2011; Weinberg et al., 2016).

More headway has been made on the second support. When paused, ribosomes exhibit a conformation in which a tRNA is absent from the A site (Hanson et al., 2018; Wu et al., 2019a). Without a tRNA in the A site, the CCR4–NOT component NOT5 binds to the ribosome through

its N-terminal domain inserted into the ribosome E site, which is also unoccupied (Buschauer et al., 2019). Loss of this interaction ablates the influences of codon optimality on mRNA decay. Interestingly, the predominant NOT5 complexes recovered in this structural study contain the initiator $tRNA_i^{Met}$ in the P site, suggesting that the role of the NOT5–ribosome interaction may be in late initiation rather than in elongation (Buschauer et al., 2019).

Paramount among the remaining questions for the codon optimality model of mRNA decay is the degree of its centrality. While the phenomenon clearly occurs, the fraction of the variance in mRNA decay that can be attributed to it remains unclear. Perhaps the most parsimonious interpretation at this time is to include codon optimality as a feature of an mRNA that contributes to its decay. Whether this feature contributes more than proteins or miRNAs that bind sequence elements to influence decay remains unknown. Most likely all currently known features will be required, as well as yet unknown ones, to understand the landscape of mRNA decay.

Poly(A)-Tail Length in Neurons

Despite their diversity, the embryonic systems that use the poly(A)-tail length to control translation have commonalities. All of these contexts exhibit mechanisms that can modify tail lengths post-transcriptionally. Cytoplasmic polyadenylation of mRNAs, with the *c-mos* mRNA serving as the paradigmatic example, was observed well before transcriptome-wide characterization of tail lengths. In *X. laevis* oocytes, *c-mos* is translated in a temporally restricted manner, with translation increasing after the oocyte is exposed to progesterone just before meiotic maturation (Richter, 1999). Other mRNAs, such as *ccnb1*, undergo temporally distinct activation, but their translation is also necessary for meiotic maturation.

Sequences in the 3′ UTR of some transcripts mediate cytoplasmic polyadenylation (de Moor and Richter, 1999; Charlesworth et al., 2013). These sequences include the polyadenylation signal (PAS) hexamer and the cytoplasmic polyadenylation element (CPE), an 8 nt U-rich sequence generally located upstream of the PAS, although sometimes overlapping it (Richter, 1999). The relative spacing of these elements with respect to one another, and with respect to 3′ terminus of the transcript, controls the timing and extent of polyadenylation (Pique et al., 2008). These elements also control cytoplasmic polyadenylation in *D. rerio*, but in *D. melanogaster*, no such element has been found. Instead, almost all mRNAs are substrates for polyadenylation by the cytoplasmic polyadenylase *wispy* (Eichhorn et al., 2016).

Careful interrogation of *cmos* and *ccnb1* in *X. laevis* oocytes injected with radiolabeled reporters revealed features of the mechanism of cytoplasmic polyadenylation. The CPE element mediates polyadenylation through association with the CPE-binding protein CPEB and the CPSF complex. These proteins recruit the scaffolding protein Symplekin and the polymerase Gld2, which catalyzes the nucleotide addition (Figure 3). They also recruit the deadenylase PARN, which opposes the action of Gld2, and the combined activities of these two enzymes, regulated by phosphorylation of CPEB, determine the tail length of the transcript during oocyte maturation (Kim and Richter, 2006).

In addition to posttranscriptional regulation of tail length, coupled embryonic and oocyte systems exhibit little mRNA decay (Subtelny et al., 2014). Little decapping activity, and perhaps additional factors, prevents transcript turnover until later in development. Translational regulation, by control of tail length and other means, is paramount when transcriptional repression cannot silence transcripts.

Control of tail-length in the cytoplasm is also shared by a non-developmental context: neurons of the rodent brain. A study noted that rat *Camk2a*, an mRNA known for its role in learning and memory, harbors two CPE elements, both of which are conserved to human (Wu et al., 1998). Furthermore, the cytoplasmic polyadenylation proteins CPEB and Gld2 are expressed in rat brain (Wu et al., 1998; Du and Richter, 2005), and six transcripts, including *Camk2a*, increase their tail length upon stimulation in cultured hippocampal neurons, as assessed by differential thermal elution from a poly(U)-sepharose column (Du and Richter, 2005). *Camk2a* also increases its tail length in response to light stimulation of the visual cortex (Wu et al., 1998).

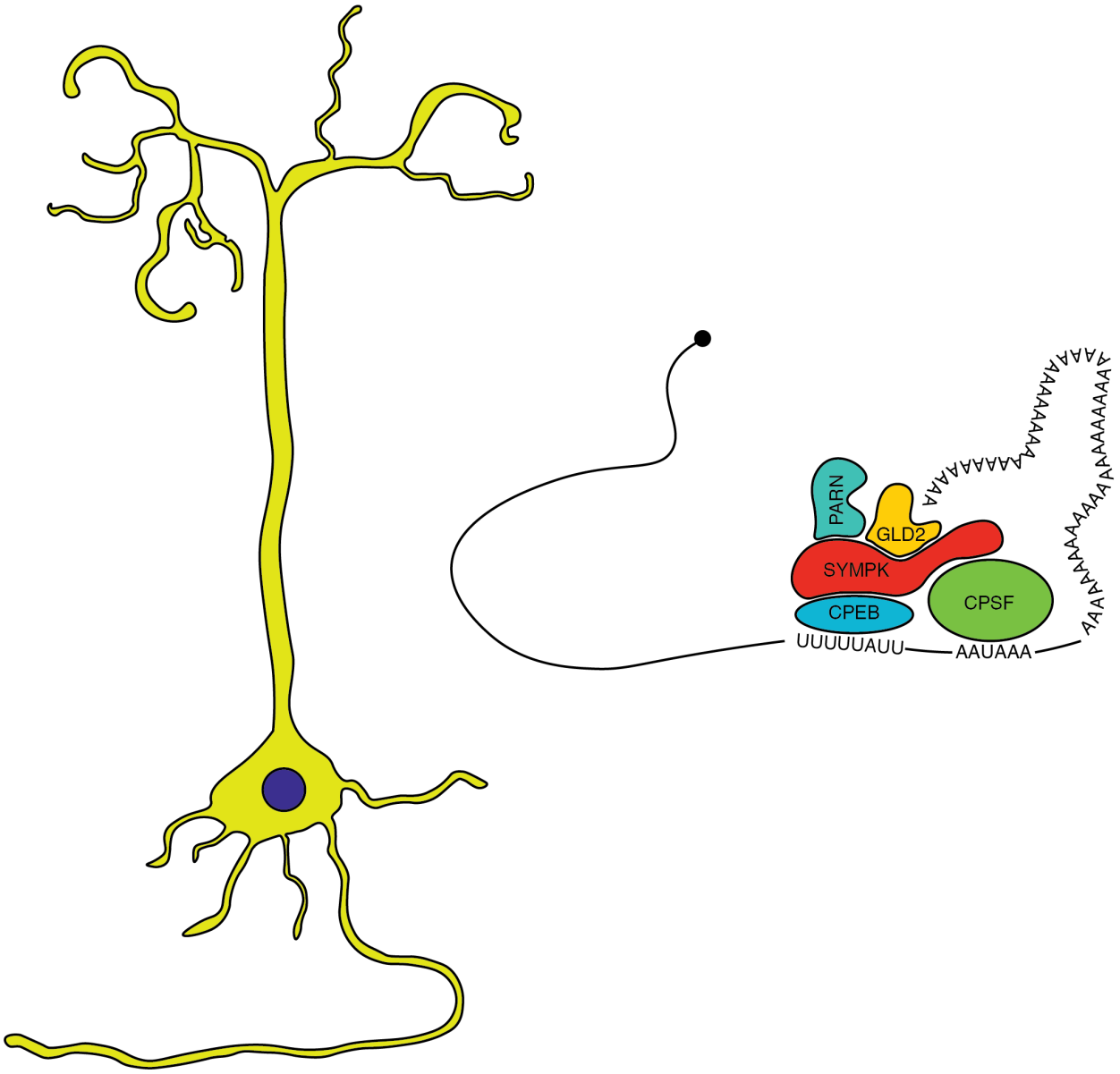


Figure 3. Cytoplasmic Polyadenylation in Neurons
(Left) Cytoplasmic polyadenylation in neurons has been hypothesized to occur in subcellular regions far from the nucleus, as these regions have additional requirements for translational control (Wu et al., 1998). The distal dendrites of the CA3 pyramidal neuron might provide such a region. (Right) Cytoplasmic polyadenylation in *Xenopus* oocytes is mediated by both the CPSF complex and the CPEB protein, which bind to distinct elements in the 3′ UTR. The deadenylase PARN and the poly(A)-polymerase GLD2 can both be recruited by CPEB but have opposing influences on tail length. When CPEB is phosphorylated, PARN is excluded from the complex, resulting in polyadenylation (Kim and Richter, 2006).

Neurons are transcriptionally active and exhibit mRNA decay, suggesting that if the role of cytoplasmic polyadenylation is to influence translation, these cells exist in a fundamentally different regime than coupled embryonic systems. An alternative hypothesis, suggested by the early studies of cytoplasmic polyadenylation in neurons (Wu et al., 1998), has received some support. In this hypothesis, cytoplasmic polyadenylation serves to increase translation for mRNAs that are localized to sites far from the nucleus (Figure 3). Transcriptional control in these circumstances is buffered by the temporal and spatial complexities of mRNA transport, but the requirements of neuronal activity are rapid.

In keeping with the hypothesis that cytoplasmic polyadenylation serves to control translation in the neuronal dendrites (and perhaps the axons as well), several studies have observed increased tail lengthening and translation in synaptosomes and synaptoneurosomes, biochemical fractionations that enrich for pieces of axons and dendrites that have broken from the rest of the axonal arbor and resealed (Huang et al., 2002; Dziembowska et al., 2012; Udagawa et al., 2012). While used extensively to study the biochemistry of the synapse, these preparations have biochemical purity that is often difficult to assess. This challenge stems primarily from three observations: synapse-localized proteins like PSD95 also tend to be highly expressed in the soma, synapse-localized RNAs tend to have modest synaptic enrichments, and electron micrographs of the preparations show complex and variable structures (Nagy and Delgado-Escueta, 1984; Chicurel et al., 1993; Bagni et al., 2000; Villasana et al., 2006; Westmark et al., 2011). Because the dendritic compartments are very small in comparison to the soma, with individual synapses expected to have at most a handful of mRNAs (Kosik, 2016), purifications have challenges resolving these signals. Despite these challenges, observations concerning cytoplasmic polyadenylation in whole-cell neuronal preparations, and in neurite-

enriched purifications, indicate that the phenomenon occurs. Whether it specifically controls translation in the dendrites is less clear.

Scope of the Thesis

Despite the importance of mRNA decay in determining accumulation rates and steady-state abundances, the metabolism of most mammalian mRNAs is uncharacterized. The second chapter of this thesis uses metabolic labeling and high-throughput tail-length measurements to determine metabolic rate parameters for thousands of mRNAs, including rate constants for deadenylation and short-tailed mRNA decay. Analyses of these rates constants confirm key aspects of the general three-step decay mechanism, with the vast majority of mRNAs exhibiting deadenylation-dependent decay. But this work also delineates the limits of this model, as deadenylation rates are not solely responsible for variation in decay rate. Variation in short-tailed mRNA decay rates among mRNAs from different genes also contributes to mRNA decay.

The third chapter of this thesis investigates the mechanism of miRNA-mediated decay. We resolve the paradox of how miRNAs can recruit deadenylases but leave tail-length unchanged. Building on insights in the second chapter, we find that, by enhancing short-tailed mRNA decay, miRNAs alter pre–steady-state tail lengths but leave steady-state tail lengths unchanged. These observations explain the lack of tail-length changes observed in the 6 hpf *D. rerio* embryo for targets of the injected miR-155 and provide more clarity in understanding the mechanism of miRNA-mediated repression in mammalian cells.

In the fourth chapter, this work investigates the role of the poly(A)-tail in neurons, using as motivation prior studies that indicate enhanced translation on mRNAs that have undergone cytoplasmic polyadenylation in response to neuronal stimuli. In keeping with these observations,

we observe a relationship between tail length and translational efficiency in neurons when investigating thousands of mRNAs. Non-neuronal cells may dampen the relationship in neuronal tissue, consistent with the lack of coupling between tail length and translation in most mature cell types (Subtelny et al., 2014). But when examined in either brain tissues or neuronal cultures, the coupling is modest, indicating that tail-length does not play a strong role in regulating translational efficiency. Work from this thesis shows that, in 3T3 cells, tail-length changes determine the dynamics of mRNA decay for almost all mRNAs; the fourth chapter suggests that determining these dynamics is also the primary function of the poly(A)-tail in neurons.

References

Adivarahan, S., Livingston, N., Nicholson, B., Rahman, S., Wu, B., Rissland, O.S., and Zenklusen, D. (2018). Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. Mol Cell *72*, 727-738 e725.

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature *432*, 112-118.

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

Bagni, C., Mannucci, L., Dotti, C.G., and Amaldi, F. (2000). Chemical stimulation of synaptosomes modulates alpha -Ca2+/calmodulin-dependent protein kinase II mRNA association to polysomes. J Neurosci *20*, RC76.

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.

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell *173*, 20-51.

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.

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.

Borman, A.M., Michel, Y.M., and Kean, K.M. (2000). Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. Nucleic Acids Res *28*, 4068-4075.

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.

Brooks, S.A., and Blackshear, P.J. (2013). Tristetraprolin (TTP): interactions with mRNA and proteins, and current thoughts on mechanisms of action. Biochim Biophys Acta *1829*, 666-679.

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.

Buschauer, R., Matsuo, Y., Chen, Y.-H., Alhusaini, N., Sweet, T., Sugiyama, T., Ikeuchi, K., Cheng, J., Matsuki, Y., Gilmozzi, A.*, et al.* (2019). The Ccr4-Not complex monitors the translating ribosome for codon optimality. bioRxiv.

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.

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.

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.

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., 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, Y., Boland, A., Kuzuoglu-Ozturk, D., Bawankar, P., Loh, B., Chang, C.T., Weichenrieder, O., and Izaurralde, E. (2014). A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. Mol Cell *54*, 737-750.

Chicurel, M.E., Terrian, D.M., and Potter, H. (1993). mRNA at the synapse: analysis of a synaptosomal preparation enriched in hippocampal dendritic spines. J Neurosci *13*, 4054-4063.

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.

Christie, M., Boland, A., Huntzinger, E., Weichenrieder, O., and Izaurralde, E. (2013). Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. Mol Cell *51*, 360-373.

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.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W.*, et al.* (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247-261.

Darnell, J.E., Wall, R., and Tushinski, R.J. (1971). An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA. Proc Natl Acad Sci U S A *68*, 1321-1325.

Das Sharma, S., Metz, J.B., Li, H., Hobson, B.D., Hornstein, N., Sulzer, D., Tang, G., and Sims, P.A. (2019). Widespread Alterations in Translation Elongation in the Brain of Juvenile Fmr1 Knockout Mice. Cell Rep *26*, 3313-3322 e3315.

de Moor, C.H., and Richter, J.D. (1999). Cytoplasmic polyadenylation elements mediate masking and unmasking of cyclin B1 mRNA. EMBO J *18*, 2294-2303.

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 Dev *7*, 1632-1643.

Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N.J., Van Nostrand, E.L., Pratt, G.A.*, et al.* (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. Mol Cell *70*, 854-867 e859.

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., and Wu, L. (2016). YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. Nat Commun *7*, 12626.

Du, L., and Richter, J.D. (2005). Activity-dependent polyadenylation in neurons. RNA *11*, 1340-1347.

Dziembowska, M., Milek, J., Janusz, A., Rejmak, E., Romanowska, E., Gorkiewicz, T., Tiron, A., Bramham, C.R., and Kaczmarek, L. (2012). Activity-dependent local translation of matrix metalloproteinase-9. J Neurosci *32*, 14538-14547.

Edmonds, M., and Abrams, R. (1963). Isolation of a Naturally Occuring Polyadenylate from Calf Thymus Nuclei. Journal of Biological Chemistry *238*, PC1186-PC1187.

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*.

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.

Fabian, M.R., Frank, F., Rouya, C., Siddiqui, N., Lai, W.S., Karetnikov, A., Blackshear, P.J., Nagar, B., and Sonenberg, N. (2013). Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. Nat Struct Mol Biol *20*, 735-739.

Fan, X.C., and Steitz, J.A. (1998). Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. EMBO J *17*, 3448-3460.

Forrest, M.E., Narula, A., Sweet, T.J., Arango, D., Hanson, G., Ellis, J., Oberdoerffer, S., Coller, J., and Rissland, O.S. (2018). Codon usage and amino acid identity are major determinants of mRNA stability in humans. bioRxiv.

Fragoso, R., Mao, T., Wang, S., Schaffert, S., Gong, X., Yue, S., Luong, R., Min, H., Yashiro-Ohtani, Y., Davis, M.*, et al.* (2012). Modulating the strength and threshold of NOTCH oncogenic signals by mir-181a-1/b-1. PLoS Genet *8*, e1002855.

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.

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.

Greenberg, M.E., Hermanowski, A.L., and Ziff, E.B. (1986). Effect of protein synthesis inhibitors on growth factor activation of c-fos, c-myc, and actin gene transcription. Mol Cell Biol *6*, 1050-1057.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.

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.

Guyette, W.A., Matusik, R.J., and Rosen, J.M. (1979). Prolactin-mediated transcriptional and post-transcriptional control of casein gene expression. Cell *17*, 1013-1023.

Hanson, G., Alhusaini, N., Morris, N., Sweet, T., and Coller, J. (2018). Translation elongation and mRNA stability are coupled through the ribosomal A-site. RNA *24*, 1377-1389.

Hoek, T.A., Khuperkar, D., Lindeboom, R.G.H., Sonneveld, S., Verhagen, B.M.P., Boersma, S., Vermeulen, M., and Tanenbaum, M.E. (2019). Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. Mol Cell *75*, 324-339 e311.

Hogg, J.R., and Goff, S.P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. Cell *143*, 379-389.

Huang, Y.S., Jung, M.Y., Sarkissian, M., and Richter, J.D. (2002). N-methyl-D-aspartate receptor signaling results in Aurora kinase-catalyzed CPEB phosphorylation and alpha CaMKII mRNA polyadenylation at synapses. EMBO J *21*, 2139-2148.

Huntzinger, E., Kuzuoglu-Ozturk, D., Braun, J.E., Eulalio, A., Wohlbold, L., and Izaurralde, E. (2013). The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res *41*, 978-994.

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.

Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). RNA function. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. Science *345*, 455-459.

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.

Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet *16*, 421-433.

Kim, J.H., and Richter, J.D. (2006). Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. Mol Cell *24*, 173-183.

Kosik, K.S. (2016). Life at Low Copy Number: How Dendrites Manage with So Few mRNAs. Neuron *92*, 1168-1180.

Kostova, K.K., Hickey, K.L., Osuna, B.A., Hussmann, J.A., Frost, A., Weinberg, D.E., and Weissman, J.S. (2017). CAT-tailing as a fail-safe mechanism for efficient degradation of stalled nascent polypeptides. Science *357*, 414-417.

Kruijer, W., Cooper, J.A., Hunter, T., and Verma, I.M. (1984). Platelet-derived growth factor induces rapid but transient expression of the c-fos gene and protein. Nature *312*, 711-716.

Leppek, K., Schott, J., Reitter, S., Poetz, F., Hammond, M.C., and Stoecklin, G. (2013). Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. Cell *153*, 869-881.

Lim, J., Ha, M., Chang, H., Kwon, S.C., Simanshu, D.K., Patel, D.J., and Kim, V.N. (2014). Uridylation by TUT4 and TUT7 marks mRNA for degradation. Cell *159*, 1365-1376.

Loh, B., Jonas, S., and Izaurralde, E. (2013). The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. Genes Dev *27*, 2125-2138.

Lowell, J.E., Rudner, D.Z., and Sachs, A.B. (1992). 3'-UTR-dependent deadenylation by the yeast poly(A) nuclease. Genes Dev *6*, 2088-2099.

Lu, J., and Deutsch, C. (2008). Electrostatics in the ribosomal tunnel modulate chain elongation rates. J Mol Biol *384*, 73-86.

Lykke-Andersen, S., and Jensen, T.H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Biol *16*, 665-677.

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.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet *9*, 843-854.

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.

Mauxion, F., Chen, C.Y., Seraphin, B., and Shyu, A.B. (2009). BTG/TOB factors impact deadenylases. Trends Biochem Sci *34*, 640-647.

McGeary, S.E., Lin, K.S., Shi, C.Y., Pham, T.M., Bisaria, N., Kelley, G.M., and Bartel, D.P. (2019). The biochemical basis of microRNA targeting efficacy. Science *366*.

Meyer, S., Temme, C., and Wahle, E. (2004). Messenger RNA turnover in eukaryotes: pathways and enzymes. Crit Rev Biochem Mol Biol *39*, 197-216.

Morozov, I.Y., Jones, M.G., Razak, A.A., Rigden, D.J., and Caddick, M.X. (2010). CUCU modification of mRNA promotes decapping and transcript degradation in Aspergillus nidulans. Mol Cell Biol *30*, 460-469.

Muhlemann, O., and Lykke-Andersen, J. (2010). How and where are nonsense mRNAs degraded in mammalian cells? RNA Biol *7*, 28-32.

Muhlrad, D., Decker, C.J., and Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. Genes Dev *8*, 855-866.

Muhlrad, D., and Parker, R. (1992). Mutations affecting stability and deadenylation of the yeast MFA2 transcript. Genes Dev *6*, 2100-2111.

Mullen, T.E., and Marzluff, W.F. (2008). Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. Genes Dev *22*, 50-65.

Muller, R., Bravo, R., Burckhardt, J., and Curran, T. (1984). Induction of c-fos gene and protein by growth factors precedes activation of c-myc. Nature *312*, 716-720.

Nagy, A., and Delgado-Escueta, A.V. (1984). Rapid preparation of synaptosomes from mammalian brain using nontoxic isoosmotic gradient material (Percoll). J Neurochem *43*, 1114-1123.

Palatnik, C.M., Storti, R.V., and Jacobson, A. (1979). Fractionation and functional analysis of newly synthesized and decaying messenger RNAs from vegetative cells of Dictyostelium discoideum. J Mol Biol *128*, 371-395.

Park, E., and Maquat, L.E. (2013). Staufen-mediated mRNA decay. Wiley Interdiscip Rev RNA *4*, 423-435.

Park, E.H., Walker, S.E., Lee, J.M., Rothenburg, S., Lorsch, J.R., and Hinnebusch, A.G. (2011a). Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1*PABP mRNPs in vivo. EMBO J *30*, 302-316.

Park, E.H., Zhang, F., Warringer, J., Sunnerhagen, P., and Hinnebusch, A.G. (2011b). Depletion of eIF4G from yeast cells narrows the range of translational efficiencies genome-wide. BMC Genomics *12*, 68.

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.

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.

Qin, M., Kang, J., Burlin, T.V., Jiang, C., and Smith, C.B. (2005). Postadolescent changes in regional cerebral protein synthesis: an in vivo study in the FMR1 null mouse. J Neurosci *25*, 5087-5095.

Richter, J.D. (1999). Cytoplasmic polyadenylation in development and beyond. Microbiol Mol Biol Rev *63*, 446-456.

Rissland, O.S., and Norbury, C.J. (2009). Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. Nat Struct Mol Biol *16*, 616-623.

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.

Schmid, M., and Jensen, T.H. (2008). The exosome: a multipurpose RNA-decay machine. Trends Biochem Sci *33*, 501-510.

Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B. (2013). Rate-limiting steps in yeast protein translation. Cell *153*, 1589-1601.

Sheiness, D., and Darnell, J.E. (1973). Polyadenylic acid segment in mRNA becomes shorter with age. Nat New Biol *241*, 265-268.

Shyu, A.B., Greenberg, M.E., and Belasco, J.G. (1989). The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. Genes Dev *3*, 60-72.

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.

Udagawa, T., Swanger, S.A., Takeuchi, K., Kim, J.H., Nalavadi, V., Shin, J., Lorenz, L.J., Zukin, R.S., Bassell, G.J., and Richter, J.D. (2012). Bidirectional control of mRNA translation and synaptic plasticity by the cytoplasmic polyadenylation complex. Mol Cell *47*, 253-266.

Van Etten, J., Schagat, T.L., Hrit, J., Weidmann, C.A., Brumbaugh, J., Coon, J.J., and Goldstrohm, A.C. (2012). Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. J Biol Chem *287*, 36370-36383.

Villasana, L.E., Klann, E., and Tejada-Simon, M.V. (2006). Rapid isolation of synaptoneurosomes and postsynaptic densities from adult mouse hippocampus. J Neurosci Methods *158*, 30-36.

Vlasova-St Louis, I., and Bohjanen, P.R. (2011). Coordinate regulation of mRNA decay networks by GU-rich elements and CELF1. Curr Opin Genet Dev *21*, 444-451.

Wakiyama, M., Imataka, H., and Sonenberg, N. (2000). Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for Xenopus oocyte maturation. Curr Biol *10*, 1147-1150.

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.

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.

Westmark, C.J., Hervey, C.M., Berry-Kravis, E.M., and Malter, J.S. (2011). Effect of Anticoagulants on Amyloid beta-Protein Precursor and Amyloid Beta Levels in Plasma. J Alzheimers Dis Parkinsonism *1*, 101.

Wilson, T., and Treisman, R. (1988). Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3' AU-rich sequences. Nature *336*, 396-399.

Wong, S.F., Agarwal, V., Mansfield, J.H., Denans, N., Schwartz, M.G., Prosser, H.M., Pourquie, O., Bartel, D.P., Tabin, C.J., and McGlinn, E. (2015). Independent regulation of vertebral number and vertebral identity by microRNA-196 paralogs. Proc Natl Acad Sci U S A *112*, E4884-4893.

Wu, C.C., Zinshteyn, B., Wehner, K.A., and Green, R. (2019a). High-Resolution Ribosome Profiling Defines Discrete Ribosome Elongation States and Translational Regulation during Cellular Stress. Mol Cell *73*, 959-970 e955.

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.

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 alpha-CaMKII mRNA at synapses. Neuron *21*, 1129-1139.

Wu, Q., Medina, S.G., Kushawah, G., DeVore, M.L., Castellano, L.A., Hand, J.M., Wright, M., and Bazzini, A.A. (2019b). Translation affects mRNA stability in a codon-dependent manner in human cells. Elife *8*.

Wurm, J.P., Holdermann, I., Overbeck, J.H., Mayer, P.H.O., and Sprangers, R. (2017). Changes in conformational equilibria regulate the activity of the Dcp2 decapping enzyme. Proc Natl Acad Sci U S A *114*, 6034-6039.

Yamashita, A., Chang, T.C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.Y., and Shyu, A.B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol *12*, 1054-1063.

Zheng, L., Dominski, Z., Yang, X.C., Elms, P., Raska, C.S., Borchers, C.H., and Marzluff, W.F. (2003). Phosphorylation of stem-loop binding protein (SLBP) on two threonines triggers degradation of SLBP, the sole cell cycle-regulated factor required for regulation of histone mRNA processing, at the end of S phase. Mol Cell Biol *23*, 1590-1601.

Chapter 2. The Dynamics of Cytoplasmic mRNA Metabolism

Timothy J. Eisen[1,2,3]*, Stephen W. Eichhorn[1,2,3]*, Alexander O. Subtelny[1,2,3]*,
Kathy S. Lin[1,2,3,4], Sean E. McGeary[1,2,3], Sumeet Gupta[2], and David P. Bartel[1,2,3]

[1]Howard Hughes Medical Institute, Cambridge, MA, 02142, USA
[2]Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
*These authors contributed equally to this work.

A.O.S., S.W.E., T.J.E., and D.P.B. conceived the project and designed the study. T.J.E., S.W.E., and A.O.S. performed the molecular experiments and analysis. T.J.E. performed the computational modeling with input from K.S.L. and S.E.M. S.W.E., S.G., and T.J.E. adapted PAL-seq for compatibility with current Illumina technologies. K.S.L. and S.W.E. wrote the analysis pipeline for determining tail-length measurements from PAL-seq data. A.O.S., S.W.E., and T.J.E. drafted the manuscript, and T.J.E. and D.P.B. revised the manuscript with input from the other authors.

Published as:
Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., Lin, K.S., McGeary, S.E., Gupta, S., and Bartel, D.P. (2020). The Dynamics of Cytoplasmic mRNA Metabolism. Mol Cell 77, 786-799 e710.

Summary

For all but a few mRNAs, the dynamics of metabolism are unknown. Here, we developed an experimental and analytical framework for examining these dynamics for mRNAs from thousands of genes. mRNAs of mouse fibroblasts exit the nucleus with diverse intragenic and intergenic poly(A)-tail lengths. Once in the cytoplasm, they have a broad (1000-fold) range of deadenylation rate constants, which correspond to cytoplasmic lifetimes. Indeed, with few exceptions, degradation appears to occur primarily through deadenylation-linked mechanisms, with little contribution from either endonucleolytic cleavage or deadenylation-independent decapping. Most mRNA molecules degrade only after their tail lengths fall below 25 nt. Decay rate constants of short-tailed mRNAs vary broadly (1000-fold) and are larger for short-tailed mRNAs that have previously undergone more rapid deadenylation. This coupling helps clear rapidly deadenylated mRNAs, enabling the large range in deadenylation rate constants to impart a similarly large range in stabilities.

Introduction

mRNAs corresponding to different genes are degraded at substantially different rates, with some mRNAs turning over in minutes and others persisting for days (Dölken et al., 2008). Different conditions or developmental contexts can modify these rates, resulting in the destabilization of previously stable mRNAs, or vice versa (Rabani et al., 2011). These rate changes influence the dynamics of mRNA accumulation and, ultimately, the steady-state abundance of mRNAs.

Many proteins that promote mammalian mRNA degradation also can recruit deadenylase complexes. These include Pumilio (Van Etten et al., 2012), SMG5/7 (Muhlemann and Lykke-Andersen, 2010), GW182 (Fabian et al., 2011), BTG/TOB factors (Mauxion et al., 2009), Roquin (Leppek et al., 2013), YTHDF2 (Du et al., 2016), and HuR, TTP, and other proteins that bind AU- and GU-rich elements (Vlasova-St Louis and Bohjanen, 2011; Fabian et al., 2013). That these diverse modifiers of mRNA stability converge on deadenylation suggests that differences in deadenylation rates might explain a substantial fraction of the variation observed in mRNA stability.

In the past, the dynamics of mRNA deadenylation have been examined on a gene-by-gene basis, involving pulsed expression and subsequent analysis of mRNA transcripts using RNase H to cleave the mRNA and RNA blots to probe for the poly(A)-tailed 3′ fragment. Because this procedure has been performed for only a handful of cellular mRNAs in yeast (Decker and Parker, 1993; Muhlrad et al., 1994; Hilgers et al., 2006) and mammals (Mercer and Wake, 1985; Wilson and Treisman, 1988; Shyu et al., 1991; Chen and Shyu, 1995; Gowrishankar et al., 2005), some fundamental questions, including the extent to which a global relationship exists between deadenylation rate and mRNA stability, have remained unanswered.

Here, we developed experimental and analytical tools for the global analysis of tail-length dynamics. Applying these tools to the mRNAs of cultured mouse fibroblasts generated a unique resource of initial cytoplasmic tail lengths, deadenylation rates, and decay parameters for mRNAs of thousands of individual genes, which in turn provided fundamental insights into cytoplasmic mRNA metabolism.

Results

Global Profiling of Tail-Length Dynamics

Two high-throughput methods, each with distinct advantages, were initially developed to profile poly(A)-tail lengths. One is PAL-seq (poly(A)-tail-length profiling by sequencing), which also reports the cleavage-and-polyadenylation site for each polyadenylated molecule (Subtelny et al., 2014), whereas the other is TAIL-seq, which can measure poly(A) tails that have been terminally modified with non-A residues (Chang et al., 2014; Lim et al., 2016). Here, we developed PAL-seq version 2 (v2), which combines these advantages and has the further benefit over both previous methods of more robust compatibility with contemporary Illumina sequencing platforms (Figure S1).

To observe tail-length dynamics of endogenous mRNAs, we employed a metabolic-labeling approach in which mRNAs of different age ranges were isolated and analyzed (Figure 1A). To initiate labeling, we added 5-ethynyl uridine (5EU) to 3T3 cells. After incubating for time periods ranging from 40 min to 8 h, cytoplasmically enriched lysates were collected, and RNA containing 5EU was isolated by virtue of the reactivity between the 5EU and an azide-bearing biotin tag. Poly(A)-tail lengths of captured mRNAs, as well as total-lysate mRNA, were measured using PAL-seq v2 (hereafter called PAL-seq). In parallel, we performed RNA-seq,

which measured mRNA abundance for each time interval. Spike-in of RNA standards with known tail lengths enabled estimates of recovery and measurement accuracy over a broad range of tail lengths, as well as absolute quantification of RNA measured by each method. These experiments were performed using each of two independently passaged 3T3 cell lines. Unless stated otherwise, figures show the results obtained for cell line 1. Nonetheless, the results of the two cell lines were reproducible at each time interval ($R_s \geq 0.81$ for mean tail-length measurements). Moreover, results from either PAL-seq v1, PAL-seq v2, or our implementation of TAIL-seq were highly correlated (Figures S2A–S2D; $R_s = 0.83–0.88$ for each of the two-way comparisons), which indicated that our conclusions were independent of the method used for tail-length profiling.

As expected if tail lengths become shorter over time in the cytoplasm (Sheiness and Darnell, 1973; Palatnik et al., 1979), mRNAs collected after the shortest labeling period (40 min) had the longest poly(A)-tail lengths, with median length of 133 nt (Figure 1B). As the average age of each labeled mRNA population increased with longer labeling periods, tail-length distributions shifted toward the steady-state distribution with respect to both length and abundance (Figure 1B). At each time interval, 10–20-nt tails preferentially possessed a 3′ terminal U (Figure S2E), although < 6.8% of tails had 3′ U residues in any sample, in keeping with previous reports on the fraction of short tails with terminal uridines at steady state (Chang et al., 2014; Lim et al., 2014). Analyses of mean poly(A)-tail lengths for mRNAs corresponding to thousands of individual genes showed that tails from mRNAs of essentially every gene shortened over time in the cytoplasm (Figures 1C and 1D).

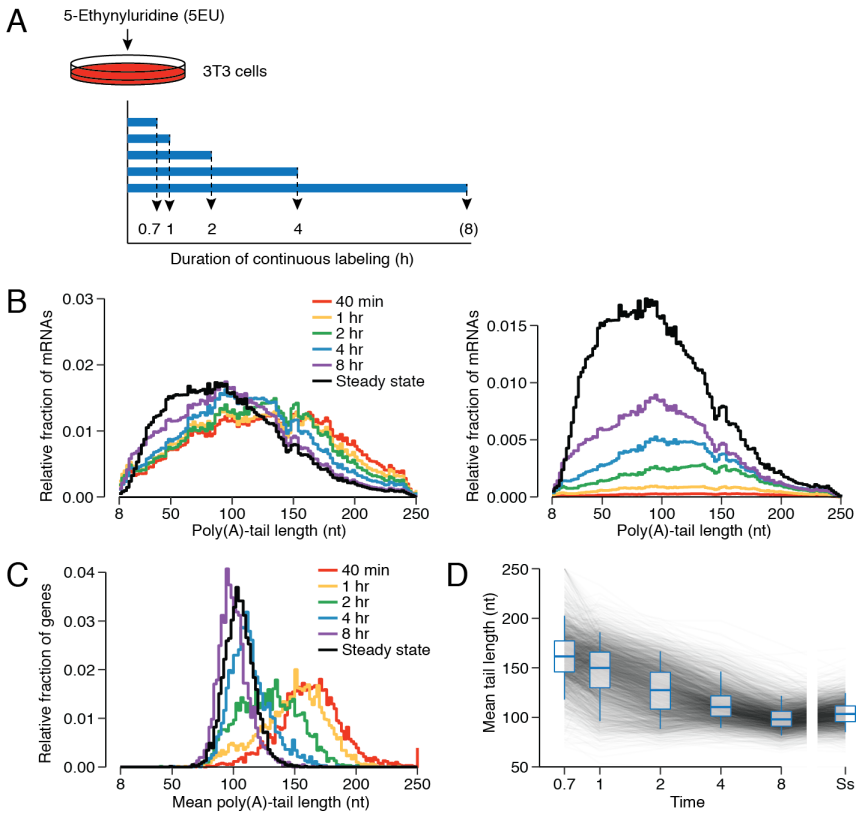


Figure 1. Global Tail-Length Dynamics of Mammalian mRNAs

(A) Schematic of 5EU metabolic labeling. Experiments were performed with two 3T3 cell lines designed to induce expression of either miR-155 or miR-1 (cell lines 1 and 2, respectively) but cultured without microRNA induction. The 8 is in parentheses because an 8-h labeling period was included for only one line (cell line 1). For simplicity, all subsequent figures show the results for cell line 1, unless stated otherwise. (B) Tail-length distributions of mRNA molecules isolated after each period of 5EU labeling (key). Left: distributions were normalized to each have the same area. Right: distributions were scaled to the abundance of labeled RNAs in each sample and then normalized such that the steady-state sample had an area of 1. The steady-state sample was prepared with unselected RNA from the 40-min time interval. Each bin is 2 nt; results for the bin with tail lengths ≥ 250 nt are not shown. (C) Distributions of mean poly(A)-tail lengths for mRNAs of each gene after the indicated duration of 5EU labeling. Values for all genes that passed the tag cutoffs for tail-length measurement at all time intervals were included (n = 3,048). Each bin is 2 nt. Genes with mean mRNA tail-length values greater than ≥ 250 nt were assigned to the 250-nt bin. (D) Tail lengths over time. Mean tail lengths for mRNAs from each gene (n = 3,048) are plotted along with box-and-whiskers overlays (line, median; box, 25th–75th percentiles; whiskers, 5th–95th percentiles). Ss, steady state. See also Figures S1 and S2.

Correspondence Between mRNA Half-life and Deadenylation Rate

After 2 h of labeling, a broad range of mean tail lengths was observed, as mean tail lengths for mRNAs of some genes approached their steady-state values, whereas those for others still resembled their initial values (Figure 1C). These different rates of approach to steady state presumably at least partly reflected differences in mRNA degradation rates, as short-lived mRNAs were expected to reach their steady-state abundance and poly(A)-tail length more rapidly than long-lived mRNAs.

To determine these degradation rates, we fit the yield of PAL-seq tags obtained for each gene at each time interval (normalizing to the spike-in controls) to the exponential function describing the approach to steady state, while also fitting a global offset to account for a delay between the time that 5EU was added and the time that labeled mRNAs appeared in the cytoplasm. This offset ranged from 27 to 36 min, depending on the experiment, a range consistent with single-gene measurements of the time required for mRNA transcription, processing, and export (Shav-Tal et al., 2004; Mor et al., 2010). Our half-life values (Table S1) correlated well with those previously reported for mRNAs of 3T3 cells growing in similar conditions (Schwanhäusser et al., 2011) (Figure S3A; R_s = 0.68–0.77), although our absolute values were substantially shorter (Figures S3B–S3D; median 2.1 h for mRNAs of the 3T3 cell line 1, as opposed to 9 h for previously reported values). This difference was attributable to potential divergence in the cell lines used in the two labs, as well as our focus on cytoplasmically enriched RNA and our absolute quantification of labeled RNA (enabled by spiking in standards).

Previous global analyses of the relationship between mRNA half-life and mean tail length have been limited to steady-state tail-length measurements, for which no positive relationship is observed (Subtelny et al., 2014), despite the established role of poly(A) tails in conferring

mRNA stability. Our current datasets, which provided the opportunity to make this comparison using half-life and tail-length measurements acquired from the same cells, reinforced this finding; we observed no positive relationship between mRNA half-life and mean steady-state tail length (Figure S3G; $R_s = -0.24$). This result held when incorporating results of PAL-seq implemented with direct ligation to mRNA 3′ termini, which better detected very short or highly modified tails (Figure 2A; $R_s = -0.02$). Indeed, the mean tail lengths of long-lived mRNAs, including those of ribosomal protein genes (RPGs), closely resembled tail lengths of short-lived mRNAs, including those of immediate-early genes (IEGs) (Figure 2A).

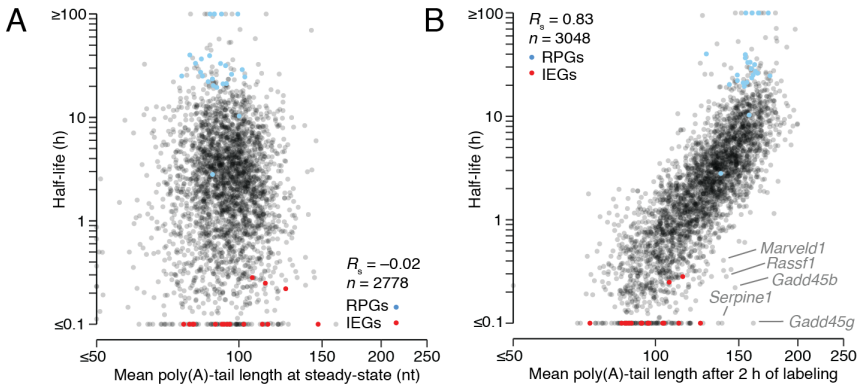


Figure 2. Correspondence Between mRNA Half-life and Deadenylation Rate
(A) Relationship between half-life and mean steady-state tail length of mRNAs in 3T3 cells. For mRNAs of each gene, standard PAL-seq data were used to determine the length distribution of tails ≥ 50 nt, and data generated from a protocol that used single-stranded ligation to the mRNA 3′ termini (rather than a splinted ligation to the tail) were used to determine both the length distribution of tails < 50 nt and the fraction of tails < 50 nt. Compared to the tail-length distribution generated by only standard PAL-seq data, this composite distribution better accounted for very short and highly modified tails. Nonetheless, using the standard PAL-seq data without this adjustment produced a similar result (Figure S3G). Results for mRNAs of ribosomal protein genes (RPGs) and immediate-early genes (IEGs) (Tullai et al., 2007) are indicated (blue and red, respectively). (B) Relationship between mRNA half-life and mean tail length of metabolically labeled mRNAs isolated after 2 h of labeling. Otherwise as in (A). See also Figures S3A–S3D and S3G.

A very different picture emerged when considering pre-steady-state tail-length measurements. After 2 h of labeling, half-life strongly corresponded to mean tail length (Figure 2B; $R_s = 0.83$). At this labeling interval, IEG mRNAs and other short-lived mRNAs had the shortest mean tail lengths, RPG mRNAs and other long-lived mRNAs had the longest mean tail lengths, and other mRNAs had mean tail lengths falling somewhere in between. The simplest explanation for this result is that the deadenylation rate dictates the stability of most mRNAs, and mean tail length at 2 h provides a proxy for deadenylation rate. Thus, slow deadenylation of long-lived mRNAs explains both why they have longer tails after 2 h of labeling and why they have such long half-lives, and rapid deadenylation of short-lived mRNAs explains why they have shorter tails after 2 h of labeling and why they have such short half-lives.

Several notable outliers had half-lives that were shorter than expected from their mean tail lengths in the 2-h sample, suggesting that their degradation and deadenylation rates were incongruous. *Rassf1*, *Serpine1*, and two *Gadd45* paralogs are known or suspected substrates for either nonsense-mediated decay (NMD) or other pathways that recruit UPF1 (Tani et al., 2012; Park and Maquat, 2013; Nelson et al., 2016). Another outlier, the *Marveld1* mRNA, has not yet been reported to interact with UPF1, but its protein product does interact with UPF1 in human cells and regulates UPF1 activity (Hu et al., 2013). Association with UPF1 can trigger endonucleolytic cleavage of mammalian mRNAs, which would decouple the rates of decay and deadenylation (Muhlemann and Lykke-Andersen, 2010), disrupting the relationship between half-life and tail length at intermediate labeling intervals. Nonetheless, the most notable feature of the outliers was their scarcity; the striking overall correspondence observed between half-life and mean tail lengths after 2 h of labeling implied that for the vast majority of endogenous

mRNA molecules of mouse fibroblasts, the rate of mRNA deadenylation largely determines the rate of degradation.

Initial Tail Lengths of Cytoplasmic mRNAs

Analysis of tail-length distributions for individual genes and the changes in these distributions over increased labeling intervals supported and extended the conclusions drawn from global analyses of abundances and mean tail lengths. This analysis confirmed that tail-length dynamics of mRNAs with short half-lives (e.g., *Metrnl*) substantially differed from those of mRNAs with longer half-lives (e.g., *Lsm1* and *Eef2*), with the short-lived mRNAs reaching their steady-state abundance and tail-length distribution much more rapidly (Figure 3). The stacked pattern of the distributions observed over increasing time intervals also illustrated that the longest-tailed mRNAs observed at steady state were essentially all recently transcribed, whereas the shortest-tailed mRNAs were mostly the oldest mRNAs (Figure 3).

Our tail-length data from short labeling periods provided the opportunity to examine the initial tail lengths of mRNAs soon after they entered the cytoplasm. The calculated 27–36-min delay in the appearance of labeled cytoplasmic mRNAs implied that most mRNAs isolated after 40 min of labeling were subject to cytoplasmic deadenylation for < 13 min. Thus, for all but the most rapidly deadenylated mRNAs, the tail lengths observed after 40 min of labeling should have approximated the tail lengths of mRNAs that first entered the cytoplasm.

Without data to the contrary, previous studies of tail-length dynamics have assumed that initial cytoplasmic tail lengths observed for mRNAs of one gene also apply to the mRNAs of all other genes. However, we observed substantial intergenic variation for average tail lengths at the shortest labeling period (Figures 1C, 3, and S3F), with the spread of the 5^{th}–95^{th} percentile values

at least that of steady state (112.2 ± 4.7 to 194.7 ± 6.0 nt for the 40-min samples and 84.8 ± 1.3 to 124.6 ± 2.1 nt for the steady-state samples; values ± SD), which suggested that mRNAs from different genes exit the nucleus with tails of quite different lengths. To examine whether

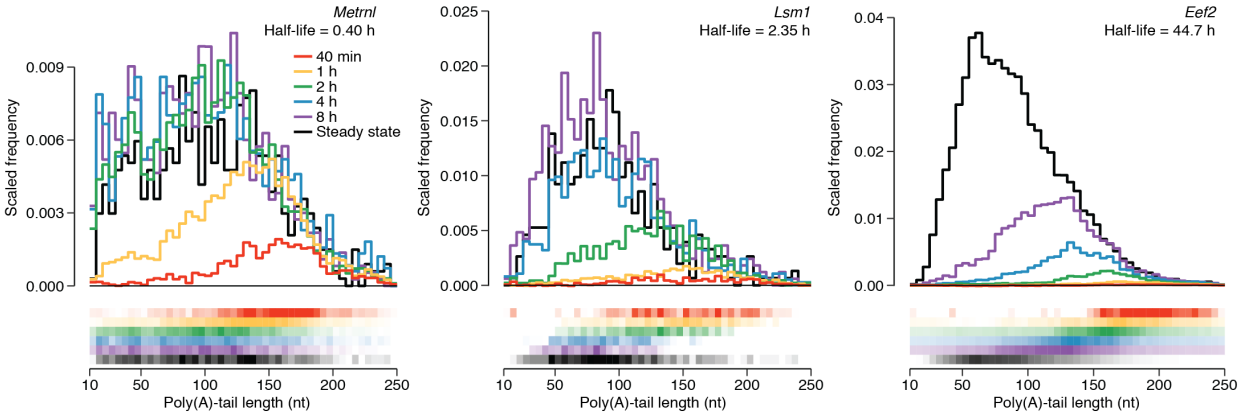


Figure 3. Tail-Length Dynamics of mRNAs with Different Half-Lives
Tail-length distributions for mRNAs from individual genes. For each time interval (key), the distribution is scaled to the abundance of labeled RNA in the sample (top), and the distribution is also represented as a heatmap (bottom), with the range of coloration corresponding to the 5th–95th percentiles of the histogram density. Each bin is 5 nt. Bins for tails < 10 nt are not shown because the splinted ligation to the tail used in the standard PAL-seq protocol depletes measurements for tails < 8 nt. Bins for tails ≥ 250 nt are also not shown. See also Figures S3F and S3H–S3J.

deadenylation occurring soon after nucleocytoplasmic export might have influenced this result, we focused on mRNAs with half-lives > 8 h. On average, mean tail lengths for these genes exhibited less than 4% change when comparing the 40-min and 1-h time intervals, implying that they also underwent little cytoplasmic deadenylation during the first 40 min of labeling. Average tail lengths observed at 40 min for mRNAs from these genes spanned a broad range, exceeding that observed at steady state (spread of the 5th–95th percentile values 128.3 ± 5.2 to 242.1 ± 16.1 nt for the 40-min samples and 81.0 ± 1.0 to 119.4 ± 1.4 nt for the steady-state samples; values ± SD), although these tail-length values observed at 40 min had little correspondence with those observed at steady state (R_s = 0.12).

When comparing mRNAs from the same gene, tail-length distributions were also quite broad for the newly exported mRNAs, as illustrated for mRNAs from three genes (Figure 3), and further demonstrated by the mean coefficient of variation (c.v.) of 0.41 for mRNAs of all measured genes (Figure S3H), compared to a c.v. of 0.20 for the 160-nt standard spiked into the 40-min sample. These c.v. values were reproducible between biological replicates and had little correspondence with mRNA half-life (Figures S3I and S3J). Although we cannot rule out the formal possibility that mRNA tails undergo exceedingly rapid and variable transient deadenylation immediately upon nuclear export, we interpret our results at short labeling periods to indicate that mRNAs exit the nucleus with considerable but reproducible intergenic and intragenic tail-length variability.

A Quantitative Model of mRNA Deadenylation and Decay

Our ability to isolate mRNAs of different age ranges for each gene and analyze their abundances and tail lengths (Figure 3) provided the unique opportunity to calculate the deadenylation rates and other metabolic rates and parameters for these mRNAs, thereby expanding the number of metabolically characterized mammalian mRNAs far beyond the four (*Mt1, Fos*, *Hbb*, and *IL8*) that have been examined using single-gene measurements (Mercer and Wake, 1985; Wilson and Treisman, 1988; Shyu et al., 1991; Gowrishankar et al., 2005). For each gene, the number of mRNA molecules with a given tail length is a function of (1) the rate of mRNA entering the cytoplasm, which in turn is a function of the rates of transcription, processing, and nucleocytoplasmic export; (2) the tail-length distribution of mRNA entering the cytoplasm; (3) the deadenylation rate; (4) the tail length below which the mRNA body is no longer protected from decay; and (5) the decay rate of the mRNA body (presumably preceded by decapping).

Therefore, we developed a mathematical model to determine, for mRNAs from thousands of genes, values for each of these parameters.

Our model was based on a system of differential equations that describe the rates of change of abundance of mRNA intermediates (Figure 4A; Table S2), an approach resembling that used to model the metabolism of RNAs from single-gene reporters (Cao and Parker, 2001; Jia et al., 2011). For each gene, transcription, nuclear processing, and export (hereafter abbreviated as "production") generates, with rate constant k_0, a distribution of initial poly(A)-tail lengths. Over time, deadenylation shortens the tail, one nucleotide at a time, with rate constant k_1. Decay of the mRNA body, with rate constant k_2, can occur alongside deadenylation and monotonically increases as the poly(A) tails get shorter. One interpretation of this deadenylation-dependent decay is that it represents decapping, followed by rapid degradation of the mRNA body. However, because we do not monitor cap status, our model was not designed to distinguish between decay mechanisms and is compatible with either 5′or 3′exonucleolytic decay of the mRNA body.

For individual mRNAs generated from the same gene, the production terms varied according to a negative binomial distribution—a distribution routinely used to model the probability of a failure after a series of successes (in our case, creating an mRNA of tail length n + 1 after successfully creating an mRNA of tail length n) (Figure 4A; Table S2). The decay rate constant (k_2) followed a logistic function, which accelerated as tails shortened. The two parameters of this function (m_d and v_d) were fit as global constants, while the scaling parameter (β) was fit to each gene (Table S2). Solving the differential equations of the model estimated both the tail-length distribution and the mRNA abundance at each time interval for mRNAs from each gene.

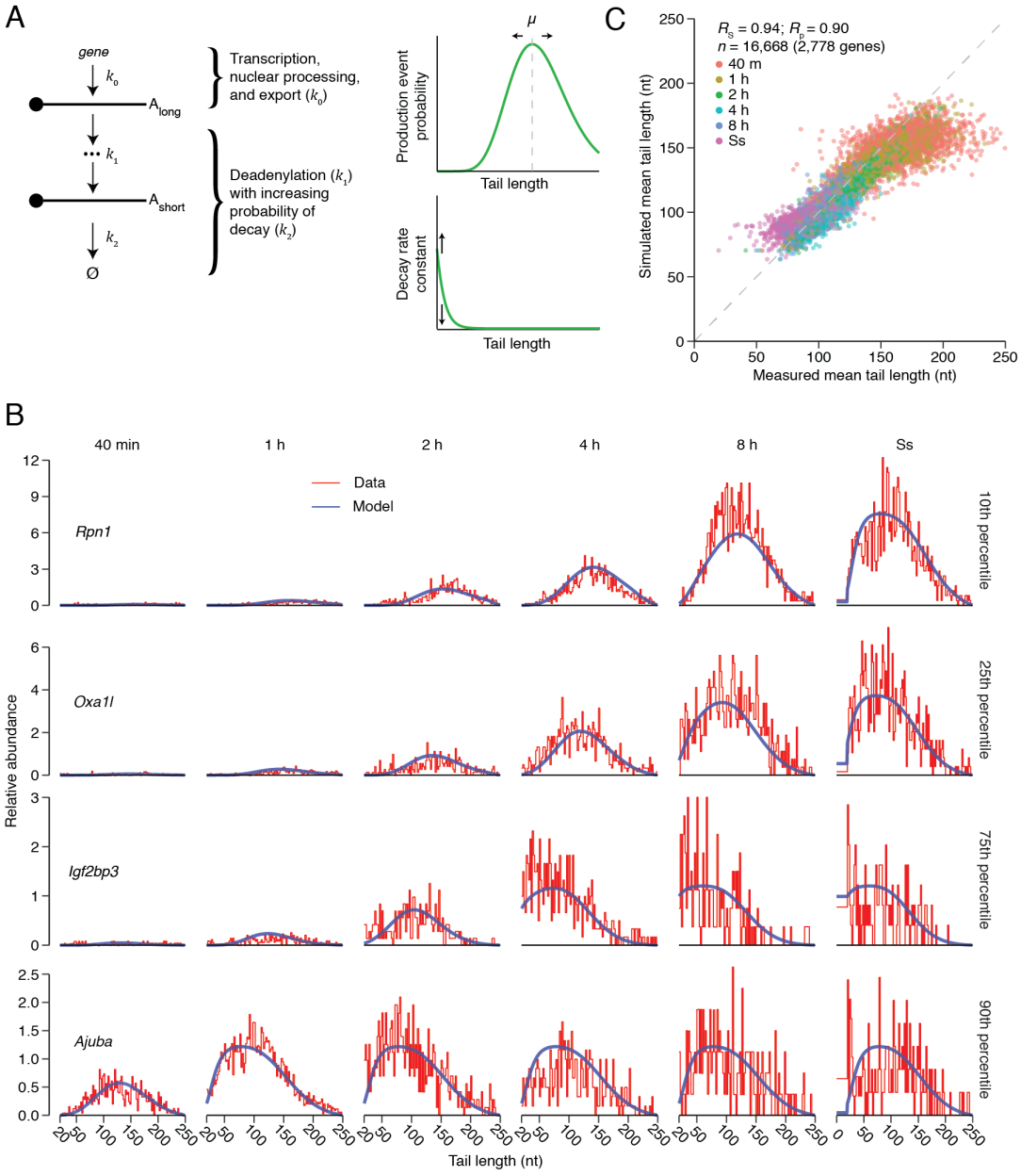


Figure 4. Computational Model of mRNA Deadenylation and Decay Dynamics

(A) Schematic of the computational model. k_0, k_1, and k_2 are terms for mRNA production, deadenylation, and decay, respectively, and Ø represents the loss of the mRNA molecule. The curves (right) indicate the distributions used to model probabilities of production and decay as functions of tail length. They are schematized using the globally fitted parameters (v_p, m_d, and v_d) that defined each distribution (Table S2). The parameter m_p controls the mean (μ) of the negative binomial distribution (top curve), whereas the decay rate constant, β, scales the decay distribution (bottom curve) (Table S2). (B) Correspondence between the model and the experimental data. Results for mRNAs of these four genes are shown as representative examples because their fits fell closest to the 10th, 25th, 75th, and 90th percentiles of the distribution of R^2 values for all genes that passed expression cutoffs in the PAL-seq datasets (Figure S4F; n = 2,778). For each time interval, the blue line shows the fit to the model, and the red line shows the distribution of observed

tail-length species, plotted in 2-nt bins and scaled to standards as in Figure 1B, right. Ss, steady state. (C) Correspondence between mean tail lengths generated from the model simulation and tail lengths measured in the metabolic labeling experiment. Shown for each gene are mean tail lengths for mRNAs at each time interval (key) from the simulation plotted with respect to the values observed experimentally. The discrepancy observed for some mRNAs at early time intervals was attributable to low signal for long-lived mRNAs at early times. The dashed line indicates y = x. See also Figures S4, S5A, and S5B and Tables S1 and S2.

Before arriving at the final version of the model (Figure 4A), we considered alternative models with varying levels of complexity. For example, building on the proposal that most mRNAs are substrates for both the PAN2–PAN3 and CCR4–NOT deadenylase complexes, with PAN2–PAN3 acting on tails > 110 nt and CCR4–NOT acting on shorter tails (Yamashita et al., 2005), we tested the performance of a model with two deadenylation rate constants, in which the transition between the two occurred at a tail length of 110 nt (Figure S4A). This model yielded residuals that were only marginally improved (Figure S4B), and for each mRNA the two deadenylation rates resembled each other (Figure S4C). A model in which the transition between the deadenylation rates occurred at 150 nt (Yi et al., 2018) yielded similar results (Figures S4D and S4E). These results indicated that, for endogenous mRNAs in 3T3 cells, either a single deadenylase complex dominates—as recently proposed for mRNAs with tail lengths ≤ 150 (Yi et al., 2018)—or both complexes act with indistinguishable kinetics. Thus, we chose not to implement a more complex model with two deadenylation rate constants.

Fitting the final version of the model to the tail-length and abundance measurements for mRNAs from thousands of genes yielded average initial tail lengths and rate constants for production, deadenylation, and deadenylation-dependent decay for each of these mRNAs (Table S2). The correspondence between the output of the model and the experimental measurements is illustrated for genes selected to represent different quantiles of fit based on the distribution of R^2 values (Figure 4B; Figure S4F). Mean tail-length values generated by the model corresponded

well to measured values (Figure 4C; $R_s = 0.94$, $R_p = 0.90$). Moreover, values fit for starting tail length, production, deadenylation, and deadenylation-dependent decay were reproducible between biological replicates and robust to parameter initialization as well as multinomial sampling (bootstrap analysis) (Figures S4G–S4J).

The Dynamics of Cytoplasmic mRNA Metabolism

Of the six yeast mRNAs and four mammalian mRNAs that have been metabolically characterized, the data for four yeast mRNAs and two mammalian mRNAs are of sufficient resolution to derive deadenylation rates. The two mammalian mRNAs, *Fos* and *Mt1*, have deadenylation rate constants that differ by 60-fold (20 and 0.33 nt/min, respectively) (Mercer and Wake, 1985; Shyu et al., 1991). Our analysis, which characterized the metabolism of 2,778 mRNAs, greatly expanded the set of mRNAs with measured deadenylation rates and showed that deadenylation rate constants of mammalian mRNAs can differ by > 1000-fold—as fast as > 30 nt/min and as slow as 1.8 nt/h (Figure 5A). Concordant with our direct analysis of the primary data, which indicated that most mRNAs degrade through a mechanism involving tail shortening (Figure 1F), mRNA half-lives corresponded strongly to deadenylation rate constants fit to our model ($R_s = -0.95$; Figure S5A).

Our model and its fitted parameters allowed us to compute the deadenylation-dependent decay rates at each tail length and thereby infer the tail lengths at which mRNAs were degraded (Figure 5B). This analysis indicated that nearly all decay of the mRNA body occurred after the tail lengths fell below 100 nt, which agreed with previous analyses of reporter genes (Yamashita et al., 2005). Decay accelerated as tail lengths fell below 50 nt (with > 92% of mRNAs decaying below this

length), a length less than the 54-nt footprint of two adjacent cytoplasmic poly(A)-binding protein (PABPC) molecules (Baer and Kornberg, 1983; Yi et al., 2018), but most mRNA

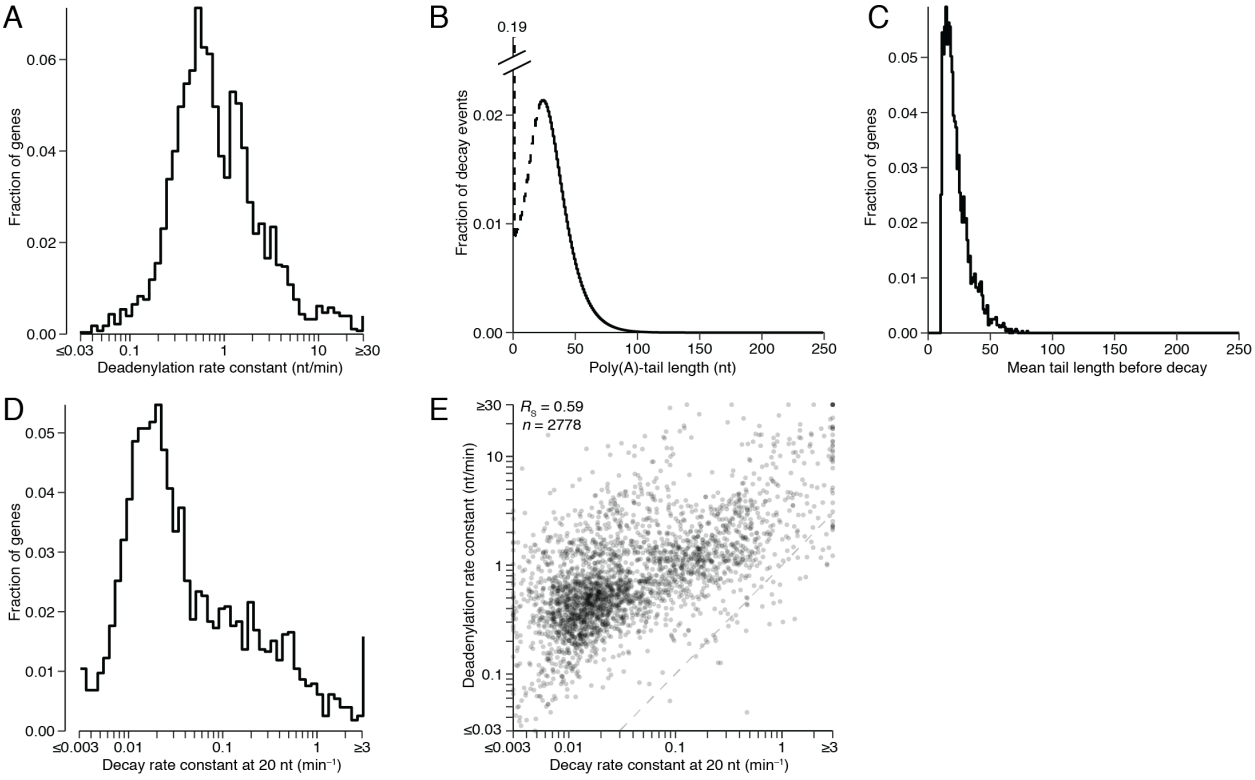


Figure 5. Dynamics of Cytoplasmic mRNA Metabolism

(A) Distribution of deadenylation rate constants (k_1 values), as determined by fitting the model to data for mRNAs from each gene (n = 2,778). (B) Tail lengths at which mRNAs decay, as inferred by the model. The model rate constants were used to simulate a steady-state tail-length distribution for each gene. The abundance of each mRNA intermediate was then multiplied by the decay rate constant k_2 to yield a distribution of decay events over all tail lengths. Plotted is the combined distribution for all mRNA molecules of all 2,778 genes. Results were indistinguishable when the distribution from each gene was weighted equally. Values for tails < 20 nt are shown as a dashed line because the model fit steady-state tail lengths < 20 nt as an average of the total abundance of tails in this region and, thus, did not provide single-nucleotide resolution for decay rates of these species. (C) Mean tail lengths at which mRNAs from each gene (n = 2,778) decayed, as inferred by the model. Otherwise, as in (B). (D) Distribution of decay rate constants (k_2 values) for mRNAs with 20-nt tail lengths, as determined by fitting the model to data for mRNAs from each gene (n = 2,778). (E) Correlation between the deadenylation rate constant (k_1) and the decay rate constant (k_2) at a tail length of 20 nt. The dashed line indicates y = x. See also Figure S4.

molecules (> 55%) did not decay until their tail lengths fell below 25 nt, a length less than the 27-nt footprint of a single PABPC molecule (Figure 5B).

When analyzing for mRNAs of each gene the mean tail length at which the mRNA body decays, the results generally concurred with those observed for all mRNAs combined, with mRNAs from most genes decaying at short mean tail lengths (Figure 5C; > 97% decaying at mean tail length < 50 nt and > 69% decaying at mean tail length < 25 nt). As expected, most mRNAs previously found to have discordant deadenylation and decay rates (Figure 1F) were also outliers in this analysis, with *Gadd45b* and *Marveld1* degrading at mean tail lengths of 62, and 59 nt, respectively. The estimates of mean tail lengths at which mRNAs decay together with initial tail lengths and deadenylation rate constants enabled estimates of the time required to reach the mean tail length of decay, which corresponded to lifetime slightly better than did the deadenylation rate constants on their own to half-life (Figures S5A and S5B; $R_s = -0.96$ and -0.95, respectively.)

Once tails reached a short length, the decay rate constants varied widely, with short-tailed mRNAs from some genes undergoing decay at rate constants > 1000-fold greater than those of short-tailed mRNAs from other genes (Figure 5D). *Fos*, a rapidly deadenylated mRNA, is degraded much faster upon reaching a short tail length than is *Hbb*, a less rapidly deadenylated mRNA (Shyu et al., 1991). More rapid degradation of short-tailed mRNAs that had been more rapidly deadenylated would help prevent the buildup of short-tailed isoforms of rapidly deadenylated mRNAs. However, such buildup sometimes does occur, as observed in *Drosophila* cells for three mRNAs characterized during heat shock (Dellavalle et al., 1994; Bönisch et al., 2007) and in mammalian cells for *Csf2* (Chen et al., 1995; Carballo et al., 2000), raising the question of the extent to which decay rates of short-tailed mRNAs are coupled to their

deadenylation rates. To answer this question, we examined the relationship between rate constants for deadenylation and those for decay of short-tailed mRNAs (the latter calculated for mRNAs with 20-nt tails). We found that more rapidly deadenylated mRNAs tended to be degraded more rapidly upon reaching short tail lengths (Figure 5E; $R_s = 0.59$).

A Modest Buildup of Short-Tailed Isoforms of Short-Lived mRNAs

Having found a strong tendency for more rapid clearing of mRNAs that had been more rapidly deadenylated, we investigated whether this phenomenon was able to prevent a large buildup of short-tailed isoforms of rapidly deadenylated mRNAs. For this investigation, we analyzed the steady-state dataset that incorporated results of PAL-seq implemented with direct ligation to mRNA 3′ termini, which better detected very short or highly modified tails. Despite the rapid decay of short-tailed mRNAs that had been more rapidly deadenylated, less-stable mRNAs generally did have a somewhat higher fraction of short-tailed transcripts (Figures 6A and S5C; $R_s = -0.56$). Nonetheless, the buildup of short-tailed isoforms of these unstable RNAs usually failed to exceed 30% of all transcripts (Figure 6A).

This preferential buildup of short-tailed isoforms of unstable RNAs was more clearly visualized in a meta-transcript analysis of the tail-length distribution at steady state. Short-lived mRNAs (half-lives < 20 min) had two peaks of short-tailed isoforms, a major peak centering at 7–15 nt and a minor peak at 0–1 nt, whereas long-lived mRNAs (half-lives > 10 h) were depleted of tails of < 20 nt (Figure 6B). Closer inspection of these two peaks revealed that these short-tailed isoforms of short-lived mRNAs were dramatically enriched in mono- and oligouridylated termini (Figures 6C, 6D, and S5D), consistent with studies showing that uridylation occurs preferentially on shorter tails and helps to destabilize mRNAs (Kwak and Wickens, 2007;

Rissland et al., 2007; Rissland and Norbury, 2009; Chang et al., 2014; Lim et al., 2014), and further indicating that uridylation occurs preferentially on short-lived mRNAs.

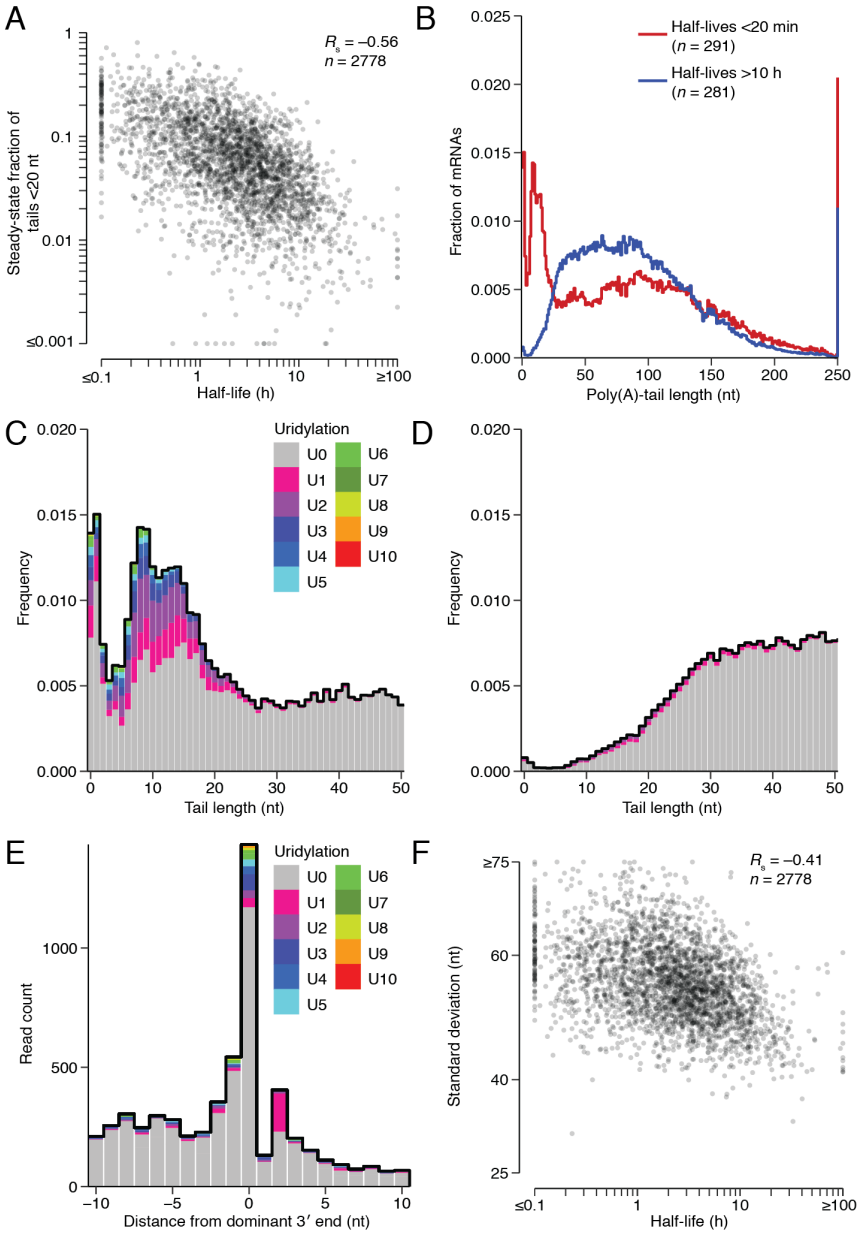


Figure 6. A Modest Buildup of Short-Tailed Isoforms of Short-Lived mRNAs

(A) Relationship between the steady-state fraction of tails < 20 nt and mRNA half-life. For mRNAs of each gene, the fraction of tails < 20 nt was calculated from a composite distribution generated as in Figure 2A, which accounted for very short and highly modified tails. (B) Metatranscript distributions of steady-state tail lengths of short- and long-lived mRNAs (red and blue, respectively), with mRNAs from each gene contributing density according to their abundance. Results were almost identical when mRNAs were weighted such that each gene contributed equally. This analysis used the composite distributions as in (A). (C) Uridylation of short-lived

mRNAs with short poly(A) tails. For mRNAs with half-lives < 20 min, the fraction of molecules with the indicated poly(A)-tail length at steady state is plotted, indicating for each tail length the proportion of tails appended with 0 through 10 U nucleotides (key). For mRNAs with poly(A)-tail length of 0, U residues were counted only if they could not have been genomically encoded. As poly(A) tails approached 20 nt, the ability to map reads with ≥ 3 terminal U residues diminished, but the ability to map reads with 1–2 terminal U residues was retained for poly(A) tails of each length. (D) Uridylation of long-lived mRNAs (half-lives > 10 h) with short poly(A) tails. Otherwise as in (C). (E) Distribution of tailless tags (regardless of mRNA half-life) as a function of their distance from the annotated 3′ end of the UTR. Tags with a terminal A (or with a terminal A followed by one or more untemplated U) were excluded, even if the A might have been genomically encoded. The proportion of tails appended with 0 through 10 U nucleotides is shown (key). (F) Relationship between the standard deviation of steady-state tail length and mRNA half-life. Otherwise as in (A). See also Figures S5C–S5J.

The observation of a 0–1-nt peak in the steady-state tail-length distribution prompted examination of fully deadenylated isoforms of mRNAs that were initially polyadenylated. Molecules without tails were often also missing the last few nucleotides of the 3′ UTR (Figures 6E and S5E), suggesting that after removing the tail, the deadenylation machinery (or some other 3′-to-5′ exonuclease) usually proceeds several nucleotides into the mRNA body. Analysis of mRNAs with tails indicated that, with few exceptions, the last nucleotide of the 3′ UTR was consistently defined (Figures S5F–S5H), which supported the idea that the missing nucleotides of tailless molecules had not been lost during the process of cleavage and polyadenylation. Analysis of the final dinucleotides of tailless tags revealed no consistent pattern after accounting for the genomic background, suggesting that other factors, such as proteins or more distal nucleotide composition, influence the position at which the exonuclease stops.

Despite their presence, the two peaks of short-tailed isoforms did not dominate the distribution, as most short-lived mRNAs (70%) had tails exceeding 30 nt (Figure 6B). Indeed, compared to long-lived mRNAs, these short-lived mRNAs also had modest enrichment for very long tails (> 175 nt) (Figures 6B, S5I, and S5J), perhaps due to an initial lag in assembling deadenylation machinery as mRNAs enter the cytoplasm, which would cause a relatively larger

fraction of short-lived mRNAs to exist in the cytoplasm prior to an initial encounter with a deadenylase. The increased fractions of both short-tail and long-tail isoforms for short-lived mRNAs led to broader overall tail-length distributions (Figure 6B) with increased standard deviations in tail length (Figure 6F; $R_s = -0.41$). Moreover, the increased fractions of shorter and longer isoforms offset each other when calculating mean tail length, leading to similar mean tail lengths for the short- and long-lived mRNAs (Figure S5K; median mean tail lengths = 89 and 92 nt, respectively), which contributed to the lack of correlation between half-life and mean tail length at steady state (Figure 2A). Most importantly, the low magnitude of the buildup supported our conclusion that for most mRNAs the steps of deadenylation and subsequent decay are kinetically coupled: short-tailed mRNAs that had previously undergone more rapid deadenylation are more rapidly degraded. This coupling prevents a large buildup of short-tailed isoforms of rapidly deadenylated RNAs, thereby enabling the large range in deadenylation rate constants to impart a similarly large range in mRNA stabilities.

Deadenylation and Decay Dynamics of Synchronous mRNA Populations

Our continuous-labeling experiments were designed to measure the dynamics of mRNA metabolism in an unperturbed cellular environment. However, this framework required deadenylation and deadenylation-dependent decay parameters to be inferred as mRNAs from each gene approached their steady-state expression levels and tail lengths, with their populations becoming progressively less synchronous, causing the signal for their end behavior to be diluted. For orthogonal measurements of these parameters, we performed a pulse-chase-like experiment that more closely resembled previous studies with single-gene reporters, in that it monitored synchronous populations of mRNAs from each gene. After a 1-h pulse of 5EU, 3T3 cells were

treated with actinomycin D (actD) to block transcription, and abundances and poly(A)-tail lengths of the mRNAs produced during the 5EU-labeling period were measured over the next 15 h, thereby revealing the behavior of synchronized mRNA populations as they age (Figure 7A).

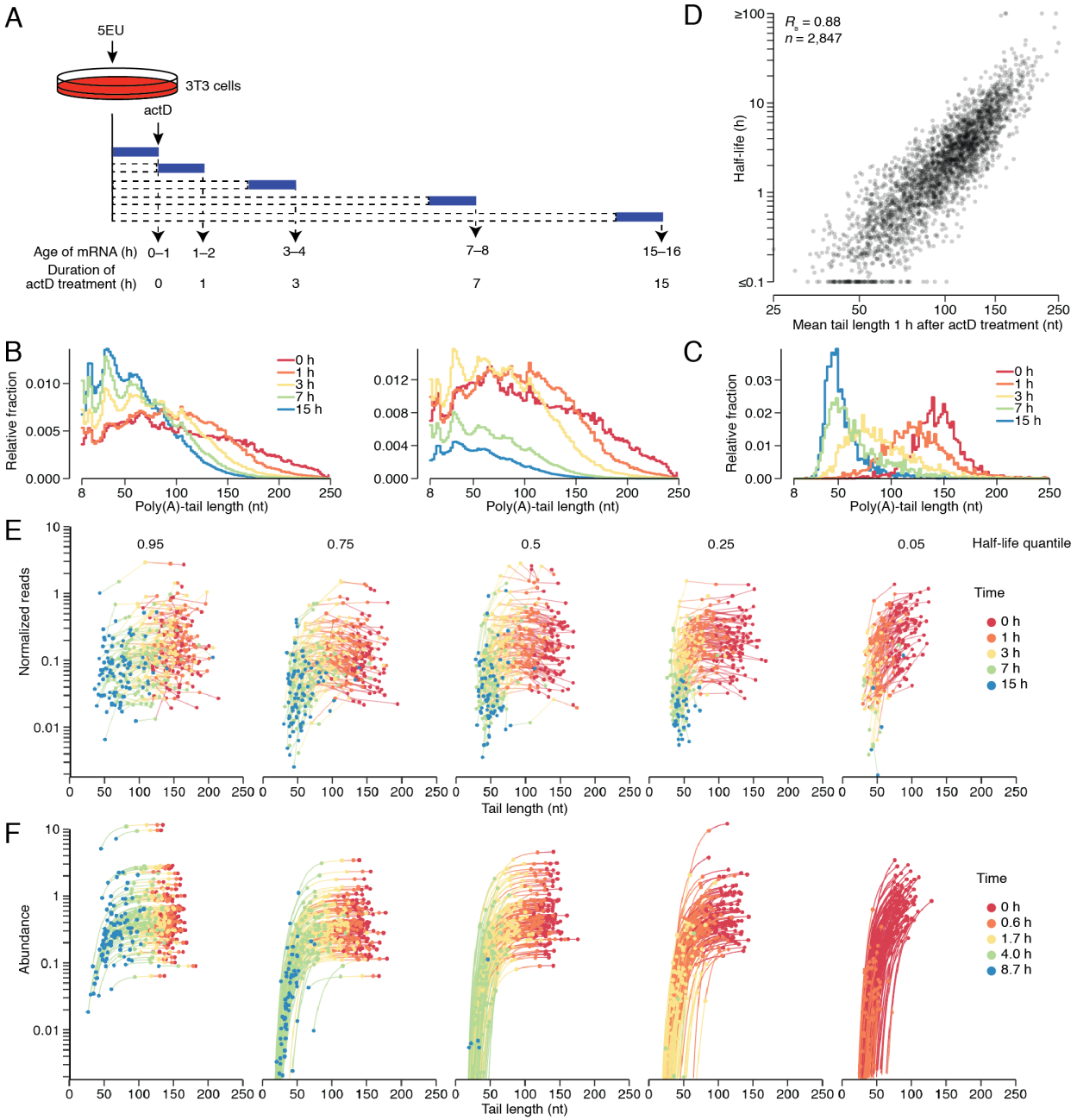


Figure 7. Deadenylation and Decay Dynamics of Synchronous mRNA Populations.
(A) Schematic of 5EU metabolic-labeling and actD treatments used to analyze synchronized cellular mRNAs. Cells from cell line 2 were treated for 1 h with 5EU, then treated with actD continuously over a time course spanning 15 h. (B) Tail-length distributions of labeled mRNA molecules observed at the indicated times after stopping transcription (key). Left: distributions

were normalized to all have the same area. Right: distributions were scaled to the abundance of labeled RNAs in each sample and then normalized such that the 0-h time interval had an area of 1. Each bin is 2 nt; results for the bins with tail lengths < 8 nt and ≥ 250 nt are not shown. At 0 h, 7% of the tails were still ≥ 250 nt, which helps explain why the density for the remainder of the tails fell below that observed at 1 h. (C) Distributions of mean poly(A)-tail lengths for labeled mRNAs of each gene after the indicated duration of transcriptional shutoff. Values for all mRNAs that passed the cutoffs for tail-length measurement at all time points were included (n = 2,155). Each bin is 2 nt. (D) Relationship between half-life and mean tail length of labeled mRNAs from each gene after 1 h of actD treatment. (E) Labeled mRNA abundance as a function of mean tail length over time. Results are shown for mRNAs grouped by half-life quantiles (95%, 75%, 50%, 25%, and 5%, left to right, with mRNAs in the 5% bin having the shortest half-lives). Each half-life bin contains 100 genes. mRNA abundance was determined from paired RNA-seq data. Each line connects values for mRNA from a single gene. (F) Simulation of mRNA abundance as a function of mean tail length over time. For each gene in (E), model parameters fit from the continuous-labeling experiment were used to simulate the initial production of mRNA and its mean tail length from each gene, as well as the fates of these mRNAs and mean tail lengths after production rates were set to 0. Results are plotted as in (E), but using a shorter time course (key) to accommodate the faster dynamics observed without actD. See also Figure S3E.

As expected, tail lengths of labeled mRNAs progressively decreased after transcriptional inhibition, with median lengths shortening from 123 to 51 nt over the course of the experiment (Figure 7B). Examination of mean tail lengths of mRNAs from each gene revealed a similar trend (Figure 7C). At later time points mean tail-length distributions peaked between 45 and 50 nt (Figure 7C), far below the 100–105-nt mode of the steady-state distribution, which included mRNAs of all ages (Figure 1C).

The actD treatment had some side effects. At later time points, a ~30-nt periodicity emerged in the single-molecule tail-length distributions (Figure 7B). Although such phasing of tail lengths, with a period resembling the size of a PABPC footprint, has been observed in mammalian cells following CCR4 knockdown (Yi et al., 2018) and in *C. elegans* (Lima et al., 2017), only subtle phasing was observed in unperturbed mammalian cells (Figure 6B). The more prominent periodicity observed after prolonged actD treatment was presumably the result of more dense packing of PABPC on poly(A) tails in the context of a diminishing mRNA pool. A

second side effect of actD treatment concerned mRNA half-lives, which increased from a median of 2.1 h in the continuous-labeling experiment to a median of 3.8 h in the transcriptional-shutoff experiment (Figure S3E). This increase was observed even for mRNAs with the shortest half-lives, which indicated that it occurred before actD could have influenced protein output, i.e., in less time than that required for mRNA nucleocytoplasmic export and translation. This result generalized previous observations concerning the effects of actD on reporter-mRNA stabilities (Chen et al., 1995).

Despite the side effects of actD, the rank order of mRNA half-lives determined from the transcriptional-shutoff experiment agreed well with that from the continuous-labeling experiment (Figure S3E; $R_s = 0.78$), indicating that the transcriptional-shutoff experiment captured key aspects of the unperturbed condition. In addition, mRNA half-lives calculated from the continuous-labeling experiment strongly corresponded to mean tail length observed 1 h after actD treatment (Figure 7D; note that 1 h after actD treatment was 2 h after 5EU labeling and thus most comparable to Figure 2B). Indeed, the strength of the correspondence between half-life and 1-h tail length ($R_s = 0.88$) further supported our conclusion that the vast majority of mRNAs are primarily degraded through deadenylation-linked mechanisms.

To further analyze the results of the transcriptional-shutoff experiment, we grouped mRNAs into cohorts based on their half-lives and monitored the abundance and average tail length of mRNAs from individual genes at each time point (Figure 7E). Regardless of mRNA half-life, tails initially shortened with little change in abundance until mean tail lengths fell below 100 nt. As expected based on the strong correspondence between half-life and 1-h tail length (Figure 7D), mRNAs with shorter half-lives underwent more rapid tail shortening (Figure 7E). Once mean tail lengths fell below 50 nt (implying that a substantial fraction of tails fell

below 25 nt), degradation accelerated. This acceleration was more prominent for mRNAs with shorter half-lives, which confirmed our conclusion that short-tailed mRNAs that had undergone more rapid deadenylation are also more rapidly degraded (Figure 7E).

To examine how well our model predicted this behavior, we used it to predict the results of the transcriptional-shutoff experiment, using the rate constants measured earlier from the continuous-labeling experiment. When simulating a shorter time course to account for the more rapid deadenylation and decay observed without actD, the results predicted by the model agreed well with the experimental observations (R_s = 0.93 and 0.61 for mean tail length and abundance, respectively; n = 11,273 values above the abundance threshold for 2,687 mRNAs), including the precipitous decline in abundance when mean tail lengths fell below 50 nt and the faster degradation of short-tailed mRNAs that had undergone faster deadenylation (Figure 7F). The striking correspondence between the predictions of the model, which had been trained on the continuous-labeling experiment, and the observations of the transcriptional-shutoff experiment validated the results and conclusions from both experiments as well as from our analytical framework.

Discussion

Previous studies provide information on deadenylation and degradation dynamics for four mammalian mRNAs and some derivatives, with deadenylation rates reported for two of these four (Mercer and Wake, 1985; Wilson and Treisman, 1988; Shyu et al., 1991; Chen et al., 1995; Gowrishankar et al., 2005; Yamashita et al., 2005). Our study provided a more comprehensive resource for deriving the principles of cytoplasmic mRNA metabolism. Initial analyses revealed unanticipated intra- and intergenic variability in initial tail lengths and indicated that almost all

endogenous mRNAs are degraded primarily through deadenylation-linked mechanisms, implying that the deadenylation rate of each mRNA largely determines its half-life with surprisingly little contribution from other mechanisms, such as endonucleolytic cleavage and deadenylation-independent decapping.

Mathematical modeling of our data expanded the known range in deadenylation rate constants from 60-fold to 1000-fold and showed that the link between deadenylation rate and decay generally operates at two levels. First, mRNAs with faster deadenylation rate constants more rapidly reach the short tail lengths associated with destruction of the mRNA body. With respect to the reason that short tail lengths trigger decay, our analyses support the prevailing view that loss of PABPC binding to the poly(A) tail enhances decay, with destabilization beginning as tails become too short for cooperative binding of a PABPC dimer and accelerating as tails become too short for efficient binding of a single PABPC molecule.

A more rapid approach to short-tailed isoforms is not the whole story. mRNAs with identical 20-nt tails but from different genes can have widely different decay rate constants (1000-fold). Moreover, there is a logic to these differences—a logic conferred by the second link between deadenylation rate and decay: mRNAs that had previously undergone more rapid deadenylation decay more rapidly upon reaching short tail lengths. The coherent regulation of deadenylation and short-tailed mRNA decay rates functionally integrates mRNA turnover into a single process to ensure that mRNAs that are rapidly deadenylated are also rapidly cleared from the cell. With respect to mechanism, perhaps changes that occur as mRNA–protein complexes are remodeled to enhance deadenylation also recruit the decapping machinery and its coactivators. Terminal uridylation, which is known to stimulate decapping (Rissland and Norbury, 2009; Morozov et al., 2010; Lim et al., 2014), may aid in this remodeling, as

uridylation was preferentially observed on rapidly deadenylated, short-lived mRNAs. Physical connections between the CCR4–NOT deadenylase complex and the decapping complex (Haas et al., 2010; Ozgur et al., 2010; Jonas and Izaurralde, 2015) as well as the intracellular colocalization of these complexes (Parker and Sheth, 2007) presumably also help coordinate deadenylation and short-tailed mRNA decay rates.

The large differences observed for both deadenylation and deadenylation-dependent decay rate constants of mRNAs from different genes raise the question of what mRNA features might specify these differences. MicroRNAs and other factors that help recruit deadenylase complexes typically bind to sites in 3′ UTRs, implying that these sites help to specify the differences (Mauxion et al., 2009; Muhlemann and Lykke-Andersen, 2010; Vlasova-St Louis and Bohjanen, 2011; Van Etten et al., 2012; Fabian et al., 2013; Leppek et al., 2013; Du et al., 2016; Bartel, 2018). However, global analyses of tandem UTR isoforms indicate that the magnitude of the differences conferred by 3′-UTR sequences in NIH 3T3 cells is relatively modest (Spies et al., 2013). Codon composition can also contribute to differences in mRNA stability, but this contribution explains only a small fraction of the variability observed for endogenous mRNAs of mammalian cells (Presnyak et al., 2015; Radhakrishnan et al., 2016; Forrest et al., 2018; Wu et al., 2019). Additional insight will be required to account more fully for the large differences in stabilities observed for different mRNAs. Our results indicate that the focus should be on sequences and processes that influence or correlate with deadenylation rates.

Our global observation that mRNAs typically degrade only after their tail lengths shorten extended to the mammalian transcriptome the notion that exponential decay is not fully appropriate for modeling mRNA degradation (Shyu et al., 1991; Cao and Parker, 2001; Trcek et al., 2011; Deneke et al., 2013). For the exponential model to be appropriate, an mRNA would

need to have the same probability of decaying at any point after entering the cytoplasm. In contrast, recently exported, long-tailed mRNAs typically underwent little if any decay, which supported the restricted-degradation model in which mRNAs are provided a discrete time window to function in the cytoplasm. During this window, the body of the mRNA is unaltered, but its age and lifespan are tracked and determined through the action of tail-length dynamics. Nonetheless, for some analyses we used the exponential model and referred to its decay parameter as "half-life" when fitting abundance changes over time because in those cases a more complex model did not provide additional insight, and using mRNA half-lives is still common practice in the field.

Despite the utility of our mathematical model, it did not capture some finer details of mRNA metabolism. For example, it was not designed to model the burst of deadenylation that typically accompanies the loss of each terminal PABPC molecule (Webster et al., 2018). However, when considering the aggregate behavior of multiple mRNAs from the same gene, these bursts become blurred, with some molecules in the burst phase and others between bursts. Accordingly, we fit a single, continuous deadenylation rate constant for the mRNAs of each gene. Likewise, we fit a single, continuous production rate constant for the mRNAs of each gene, despite the known burst behavior of transcription initiation when examined in single cells (Cai et al., 2008).

The uniform deadenylation rate constants of the model were also not suitable for capturing aspects of tail behavior that occurred as tails fell below 20 nt. For example, our analysis of steady-state data revealed buildups of isoforms of short-lived mRNAs at two tail-length ranges: 0–1 and 7–15 nt. A model with uniform deadenylation rate constants can potentially explain a peak at 0 nt but not one at an intermediate tail length, such as 7–15 nt.

Recognizing this limitation but still wanting to accurately account for the buildup of isoforms with tails < 20 nt observed for short-lived mRNAs, we fit the abundance of tails < 20 nt by averaging abundance over this length range and comparing this average to that predicted by the model—an approach that did not require additional parameters to model a buildup of 7–15-nt tails. Such parameters might be warranted if further study shows that the fate of mRNAs with 7–15-nt tails differs from that of mRNAs with 0-nt tails—studies that can be contemplated now that the existence of this buildup is known. Another aspect of mRNA metabolism remaining to be incorporated into a mathematical model is terminal uridylation, which was particularly prominent on short-tailed isoforms of short-lived mRNAs.

A recent study observed that cytoplasmic noncanonical poly(A) polymerases can extend tails, acting on longer-tailed mRNAs and adding mostly A residues but also sometimes generating a mixed tail including a G or another non-A nucleotide (Lim et al., 2018). Because most mRNAs with these mixed tails would not be detected by PAL-seq, these mRNAs would have appeared to have been degraded in our analysis. Thus, our observation of little-to-no degradation of long-tailed mRNAs indicated that, in 3T3 cells, mRNAs with mixed tails comprised only a small fraction of the mRNA molecules at any point in time and did not impact the overall conclusions of our study.

Although our current approach does not model all aspects of mRNA metabolism, there is every reason to believe that the broad behaviors observed in these initial analyses will continue to be observed in more detailed representations of mRNA metabolism. With the acquisition of suitable pre-steady-state data, the dynamics of tail-length changes in the 0–20-nt range, of terminal uridylation, and of cytoplasmic polyadenylation could be better characterized—ultimately enabling incorporation of these phenomena into a comprehensive model of mRNA

metabolism. Our methods and analytical framework offer inspiration as well as a foundation for these future efforts.

Acknowledgements

We thank J. Kwasnieski and other members of the Bartel lab for helpful discussions and the Whitehead Genome Technology Core for high-throughput sequencing. This research was supported by NIH grants GM061835 and GM118135 (D.P.B.) and an NSF Graduate Research Fellowship (T.J.E.). A.O.S. was supported by NIH Medical Scientist Training Program fellowship T32GM007753. D.P.B. is an investigator of the Howard Hughes Medical Institute.

Declaration of Interests

The authors declare no competing interests.

Methods

Experimental Model and Subject Details

Cell lines and cell culture

Clonal 3T3 cell lines engineered to express miR-155 (cell line 1) or miR-1 (cell line 2) upon doxycycline treatment were previously described (Eichhorn et al., 2014). Cells were grown at 37°C in 5% CO_2 in DMEM supplemented with 10% BCS (Sigma-Aldrich) and 2 μg/mL puromycin. 3T3 cells are male. Mycoplasma testing was performed and no contamination was observed.

Methods Details

Metabolic-labeling Time Courses

Cells from each line were plated onto 500 cm^2 plates at 6.6 million cells per plate and cultured for two days such that they reached ~70%–80% confluency, at which point growth media was supplemented with 5-ethynyl uridine (5EU, Jena Biosciences) (Jao and Salic, 2008) at a final concentration of 400 μM. After the desired labeling intervals cells were harvested (Figure 1A). Four plates were harvested for each 40 min time interval, three plates for each 1 h time interval, and two plates for each other time interval. A plate that had never received 5EU was harvested in parallel for each condition.

Cells were harvested at 4°C, washed twice with 50 mL ice-cold PBS, pH 7.3 containing 100 μg/mL cycloheximide and then used to prepare cytoplasmically enriched lysate as described (Subtelny et al., 2014). An aliquot of cleared lysate was flash frozen for use in ribosome profiling (Eisen et al., 2020), and the rest of the lysate was added to 5 volumes of TRI reagent (Ambion) and frozen at –80°C. Samples stored in TRI reagent were thawed at room temperature,

and RNA was purified according to the manufacturer's protocol and used for RNA-seq or PAL-seq v2.

RNA standards

Two sets of tail-length standards (set 1 and set 3, Table S3) were described previously (standard mix 2 and standard mix 1) (Subtelny et al., 2014). The other set of standards (set 2, Table S3) was prepared based on a 705 nt fragment of the *Renilla* luciferase mRNA, which was transcribed and gel purified as described (Subtelny et al., 2014) and then capped using a Vaccinia capping system (2000 µL reaction containing 500 µg RNA, 1000 U Vaccinia capping enzyme (NEB), 1X Capping Buffer (NEB), 0.1 mM S-adenosyl methionine, 0.5 mM GTP, 50 nM [α–^{32}P]-GTP, 2000 U SUPERaseIn (ThermoFisher) at 37°C for 1 h), monitoring the amount of incorporated radioactivity to ensure that capping was quantitative. Following the capping reaction, the 2′,3′ cyclic phosphate at the 3′ end was removed using T4 polynucleotide kinase (Subtelny et al., 2014). The capped, dephosphorylated product was joined by splinted ligation to each of seven different poly(A)-tailed barcode oligonucleotides (Subtelny et al., 2014). These seven 3′ ligation partners included 110 and 210 nt poly(A) oligonucleotides prepared as described (Subtelny et al., 2014), and five gel-purified synthetic oligonucleotides (IDT), one with a 10 nt poly(A) tract and the other four with a 29 nt poly(A) tract followed by either A, C, G, or U. Ligation products were gel purified, mixed in desired ratios, with the final ratios of the different-sized species confirmed by analysis on a denaturing polyacrylamide gel.

Short and long standards were used to monitor enrichment of 5EU-containing fragmented RNA or non-fragmented RNA, respectively. Short 5EU standards were prepared by in vitro transcription of annealed DNA oligos to produce a 30 nt and 40 nt RNA, with the latter

containing a single 5EU (Table S3). In vitro transcription was performed with the MEGAscript T7 transcription kit (ThermoFisher) according to the manufacturer's protocol, except UTP was replaced with 5-ethynyluridine-triphosphate (Jena Biosciences) when transcribing the 40 nt RNA. Long standards were prepared by in vitro transcription of sequences encoding firefly luciferase and GFP using the MEGAscript T7 transcription kit and 0.1 μM PCR product as the template. When transcribing *GFP* RNA, a 20:1 ratio of UTP to 5-ethynyluridine-triphosphate was used. Short and long standards were gel purified and stored at –80°C. Prior to use, a portion of each standard was cap-labeled and gel purified again, which enabled measurement of the recovery of the 5EU-containing standard relative to that of the uridine-only standard.

Three 28–30 nt RNAs (Table S3) were synthesized (IDT) for use as quantification standards in RNA-seq. These standards were gel purified, and 0.1 fmol of each was added to each sample immediately prior to library preparation.

Biotinylation of 5EU Labeled RNA

The RNA-seq libraries analyzed in this study were from fragmented RNAs, size selected to match ribosome-profiling libraries (Eisen et al., 2020). For these libraries, poly(A) RNA was purified from 50 μg total RNA of the 40 min, 1, 2, and 4 h samples and 25 μg total RNA of the 8 h sample using oligo(dT) Dynabeads (ThermoFisher) according to manufacturer's protocol. RNA was fragmented and 27–33 nt fragments were isolated as described (Subtelny et al., 2014) and short standards that monitored 5EU enrichment were added. Biotinylation was performed in the presence of a Cu(II) catalyst in a 20 μL reaction containing 50 mM HEPES, pH 7.5, 4 mM disulfide biotin azide (Click Chemistry Tools), 2.5 mM $CuSO_4$, 2.5 mM Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA, Sigma-Aldrich), and 10 mM sodium ascorbate

(Sigma-Aldrich), incubated at room temperature for 1 h. Reactions were stopped with 5 mM EDTA and then extracted with phenol–chloroform (pH 8.0). For the steady-state samples, 5 μg of RNA from the 40 min sample was poly(A) selected and fragmented, and 27–33 nt fragments were size selected and carried forward without enriching for 5EU.

For PAL-seq v2, long standards used to monitor 5EU enrichment and recovery were added to total RNA (using a 1:10 ratio of 5EU-containing standard to non-5EU-containing standard), and samples were click labeled as above in reactions with 2.5 μg/μL RNA. For samples from the cell line 1 time course, click reactions were performed with 500, 500, 250, 200, or 100 μg total RNA for the 40 min, 1 h, 2 h, 4 h, or 8 h samples. For samples from the cell line 2 time course, click reactions were performed with 800, 525, 350, or 200 μg total RNA for the 40 min, 1 h, 2 h, or 4 h, respectively. For both cell lines, the steady-state samples did not undergo biotinylation or pull-down.

Purification of Biotinylated RNA

For RNA-seq, Dynabeads MyOne Streptavidin C1 beads (ThermoFisher) for each set of samples were combined and batch washed, starting with 200 μL of beads per reaction. Beads were washed twice with 1X B&W buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl and 0.005% Tween-20), twice with solution A (0.1 M NaOH, 50 mM NaCl), twice with solution B (0.1 M NaCl), and then twice with water, using for each wash a volume equal to that of the initial bead suspension. Following the last wash, beads were resuspended in an initial bead volume of 1X high salt wash buffer (HSWB, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1 M NaCl, 0.01% Tween-20) supplemented with 0.5 mg/mL yeast RNA (ThermoFisher) and incubated at room temperature for 30 min with end-over-end rotation, again using a volume equal to that of the

initial bead suspension. Beads were then washed three times with 200 µL 1X HSWB per reaction and split for each reaction during the last wash. After the wash was removed, sample RNA resuspended in 200 µL 1X HSWB was added to blocked beads and incubated with end-over-end rotation at room temperature for 30 min. Beads were washed twice with 800 µL water at 50°C, incubating at 50°C for 2 min for each wash, and then twice with 800 µL 10X HSWB. RNA was eluted from beads by incubating with 200 µL 0.5 M tris(2-carboxyethyl)phosphine (TCEP, Sigma-Aldrich) at 50°C for 20 min with end-over-end rotation. The initial eluate was collected, and beads were resuspended in 150 µL water and eluted again, combining the two eluates for each sample. RNA from the eluate was then ethanol precipitated using linear acrylamide as a carrier.

Purifications of non-fragmented RNA were performed as above, except bead volumes were adjusted based on estimates of the amount of labeled RNA in each sample. For the cell line 1 samples, 292, 431, 410, 598, and 500 µL of beads were used for the 40 min, 1 h, 2 h, 4 h, and 8 h samples, respectively. For the cell line 2 samples, 467, 452, 575, and 598 µL streptavidin beads were used for the 40 min, 1 h, 2 h, and 4 h samples, respectively.

Pilot experiments designed to optimize the 5EU biotinylation and purification confirmed that RNAs containing at least one 5EU could be purified efficiently, with over 80% of a model RNA substrate containing a single 5EU becoming biotinylated in a 1 h reaction (Eisen et al., 2020). This high reaction efficiency was important for the RNA-seq samples, as RNA fragments from these libraries, generated to match ribosome-profiling samples (Eisen et al., 2020), were only ~30 nt long and estimated to typically contain at most a single 5EU. Indeed, for each of the three protocols, which started with either full-length RNA (PAL-seq) or fragmented RNA (RNA-

seq), metabolically labeled RNA was substantially enriched above background (Eisen et al., 2020).

PAL-Seq v2

This method starts with the same mRNA workup as the initial version of PAL-seq (Subtelny et al., 2014), except the design of the 3′ adaptor allows for ligation to tails ending with a uridine nucleotide, as implemented in an improved version of TAIL-seq (Lim et al., 2016). PAL-seq v2 also includes a primer-extension reaction that occurs on the Illumina flowcell, with the goal of extending the sequencing primer all of the way through the poly(A) tail, so that the first sequencing read identifies both the mRNA and its cleavage-and-polyadenylation site, as in PAL-seq v1 (Subtelny et al., 2014). After stripping the extended primer, the poly(A)-tail length is then measured by direct sequencing of the poly(A) tail, as in TAIL-seq (Chang et al., 2014) (Figure S1A).

We used RNA standards of defined tail lengths to monitor library preparation, sequencing, and the computational pipeline for improved versions of PAL-seq and our implementation of TAIL-seq. Depletion of long-tailed sequences was the most prevalent source of measurement error. For TAIL-seq, this depletion seemed highly dependent on the sequencing protocol, with the best results obtained on a HiSeq machine in high-output mode using the v3 reagent kit.

Steady-state RNA (25 μg of unselected RNA from the 40 min sample) or half of the RNA eluted from each 5EU-selected sample was used to prepare PAL-seq libraries. Tail-length standard mixes (1 ng of set 1 and 2 ng of set 2 for each 5EU-selected sample, and twice these amounts for the steady-state sample), and trace 5′-radiolabeled marker RNAs (Table S3) were

added to each sample to assess tail-length measurements and ligation outcomes, respectively. Polyadenylated ends including those with a terminal uridine were ligated to a 3′-biotinylated adaptor DNA oligonucleotide (1.8 μM) in the presence of two splint DNA oligonucleotides (1.25 μM and 0.25 μM for the U and A-containing splint oligos, respectively, Table S3) using T4 Rnl2 (NEB) in an overnight reaction at 18°C. Following 3′-adaptor ligation the RNA was extracted with phenol–chloroform (pH 8.0), precipitated, resuspended in 1X RNA T1 sequence buffer (ThemoFisher), heated to 50°C for 5 min and then put on ice. RNase T1 was then added to a final concentration of 0.006 U/μL, and the reaction was incubated at room temperature for 30 min, followed by phenol–chloroform extraction and RNA precipitation. Precipitated RNA was captured on streptavidin beads, 5′ phosphorylated, and ligated to a 5′ adaptor as described (Subtelny et al., 2014) but using a modified 5′ adaptor sequence (Table S3). Following reverse transcription using SuperScript III (Invitrogen) with a barcode-containing DNA primer, cDNA was purified as described (Subtelny et al., 2014), except a 160–810 nt size range was selected. Libraries were amplified by PCR for 8 cycles using Titanium Taq (Takara) polymerase according to the manufacturer's protocol with a 1.5 min combined annealing/extension step at 57°C. PCR-amplified libraries were purified using AMPure beads (Agencourt, 40 μL beads per 50 μL PCR, two rounds of purification) according to the manufacturer's instructions.

The use of a splinted ligation of the 3′ adaptor to the poly(A) tail had the advantage of specifically ligating to mRNAs without the need to deplete ribosomal or other abundant RNAs. However, this approach was not suitable for acquiring measurements for mRNAs with tails that were either very short (< 8 nt) or extended by more than one uridine, because such tails would ligate less efficiently (or not at all) when using a splinted ligation to the 3′ adaptor. To account for these mRNAs with either very short or highly modified tails, we implemented a protocol that

used single-stranded (ss) ligation and different mRNA enrichment steps to prepared libraries from steady-state RNA isolated from each of the two cell lines. For each sample, 5 μg of total RNA was depleted of rRNA using RiboZero Gold HMR (Illumina) and further depleted of the 5.8S rRNA by subtractive hybridization. Subtractive hybridization was performed by mixing 2x SSC buffer (3M sodium chloride, 300mM sodium citrate, pH 7.0), total RNA, and 4.8 μM of each 5.8S subtractive-hybridization oligo (Table S3) in a 50 μL reaction, heating the reaction to 70°C for 5 min, then cooling it at 1°C/min to 37°C to anneal the oligos to the RNA. During this cooling, 250 μL of Dynabeads MyOne Streptavidin C1 beads per sample (ThermoFisher) were washed twice with 1X B&W buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl and 0.005% Tween-20), twice with solution A (0.1 M NaOH, 50 mM NaCl), twice with solution B (0.1 M NaCl), and then resuspended in 50 μL of 2X B&W buffer. After cooling, the entire 50 μL RNA/oligo mixture was added to 50 μL of washed beads, then incubated at room temperature for 15 min with end-over-end rotation. The sample was then magnetized and the supernatant was withdrawn and precipitated by adding 284 μL of water, 4 μL of 5 mg/mL linear acrylamide, and 1 mL of ice-cold 96% ethanol. After resuspension, RNA was ligated to a 3′ adaptor containing four random-sequence nucleotides and an adenylyl group at its 5′ end (Table S3) in a 70 μl reaction containing 10 μM adaptor, 1X T4 RNA Ligase Reaction Buffer (NEB), 20 U/μL T4 RNA Ligase 2 truncated KQ (NEB), 0.3 U/μL SUPERaseIn (ThermoFisher), and 20% PEG 8000. The reaction was incubated at 22°C overnight and then stopped by addition of EDTA (3.5 mM final concentration after bringing the reaction to 400 μL with water). RNA was phenol–chloroform extracted, precipitated, and subsequent library preparation was as for the splinted-ligation libraries.

PAL-seq v2 libraries were sequenced on an Illumina HiSeq 2500 operating in rapid mode. Hybridization mixes were prepared with 0.375 fmol PCR-amplified library that had been denatured with standard NaOH treatment and brought to a final volume of 125 μL with HT1 hybridization buffer (Illumina, 3 pM library in final mix). Following standard cluster generation and sequencing-primer hybridization, two dark cycles were performed for the splinted-ligation libraries (i.e., two rounds of standard sequencing-by-synthesis in which imaging was skipped), which extended the sequencing primer by 2 nt, thereby enabling measurement of poly(A) tails terminating in non-adenosine bases. For the direct-ligation libraries, six dark cycles were performed instead of two, which extended the sequencing primer past the four random-sequence nucleotides in the 3′ adaptor and the last two residues of the tail.

Following the two dark cycles, a custom primer-extension reaction was performed on the sequencer using 50 μM dTTP as the only nucleoside triphosphate in the reaction. To perform this extension, the flow cell temperature was first set to 20°C. Then, 120 μL of universal sequencing buffer (USB, Illumina) was flowed over each lane, followed by 150 μL of Klenow buffer (NEB buffer 2 supplemented with 0.02% Tween-20). Reaction mix (Klenow buffer, 50 μM dTTP, and 0.1 U/μL Large Klenow Fragment, NEB) was then flowed on in two aliquots (150 μL and 100 μL). The flow-cell temperature was then increased to 37°C at a rate of 8.5°C per min and the incubation continued another 2 min after reaching 37°C. 150 μL of fresh reaction mix was then flowed in, and following a 2 min incubation, 75 μL of reaction mix was flowed in eight times, with each flow followed by a 2 min incubation. The reaction was stopped by decreasing the flow cell temperature to 20°C, flowing in 150 μL of quench buffer (Illumina HT2 buffer supplemented with 10 mM EDTA) and then washing with 75 μL of HT2 buffer. The flow cell was prepared for subsequent sequencing with a 150 μL and a 75 μL flow of HT1 buffer

(Illumina). 50 cycles of standard sequencing-by-synthesis were then performed to yield the first sequencing read (read 1). XML files used for this protocol are provided at https://github.com/kslin/PAL-seq.

The flow cell was stripped, a barcode sequencing primer was annealed, and seven cycles of standard sequencing-by-synthesis were performed to read the barcode. The flow cell was then stripped again, and the same primer as used for read 1 was hybridized and used to prime 250 cycles of standard sequencing-by-synthesis to generate read 2. Thus, each PAL-seq tag consisted of three reads: read 1, read 2, and the indexing (barcode) read. For cases in which a tag corresponded to a polyadenylated mRNA, read 1 was the reverse complement of the 3′ end of the mRNA immediately 5′ of the poly(A) tail and was used to identify the mRNA and cleavage-and-polyadenylation site of long-tailed mRNAs. The indexing read was used to identify the sample, and read 2 was used to measure poly(A)-tail length and identify the mRNA and cleavage-and-polyadenylation site of short-tailed mRNAs. The intensity files of reads 1 and 2 were used for poly(A)-tail length determination, along with the Illumina fastq files.

PAL-Seq v2 Data Analysis

Tail lengths for the splinted-ligation data were determined using a Gaussian hidden Markov model (GHMM) from the python2.7 package ghmm (http://ghmm.org/), analogous to the model used in TAIL-seq (Chang et al., 2014) and described in the next paragraph. Read 1 was mapped using STAR (v2.5.4b) run with the parameters '–alignIntronMax 1 –outFilterMultimapNmax 1 –outFilterMismatchNoverLmax 0.04 –outFilterIntronMotifs RemoveNoncanonicalUnannotated –outSJfilterReads', aligning to an index of the mouse genome built using mm10 transcript annotations that had been compressed to unique instances of each gene selecting the longest

transcript and removing all overlapping transcripts on the same strand (Eichhorn et al., 2014). The genome index also included sequences of the quantification spikes and the common portion of the poly(A)-tail length standards. The sequences that identified each RNA standard (the last 20 nt of each standard sequence, Table S3) were not aligned using STAR. Instead, the unix program grep (v2.16) was used to determine which reads matched each standard (allowing no mismatches), and these reads were added to the aligned reads from the STAR output. Tags corresponding to annotated 3′ UTRs of mRNAs were identified using bedtools (v2.26.0), and if the poly(A)-tail read (read 2) contained a stretch of ≥ 10 T residues (the reverse complement of the tail) in an 11-nt window within the first 30 nt, this read was carried forward for GHMM analysis. If read 2 failed to satisfy this criterion but began with ≥ 4 T residues, the tail length was called based on the number of contiguous T residues at the start of read 2; by definition, these tails were < 10 nt and thus easily determined by direct sequencing.

For each read 2 that was to be input into the GHMM a 'T signal' was first calculated by normalizing the intensity of each channel for each cycle to the average intensity of that channel when reading that base in read 1 and then dividing the thymidine channel by the sum of the other three channels. Sometimes a position in a read would have a value of 0 for all four channels. A read was discarded if it contained more than five such positions. Otherwise, the values for these positions were imputed using the mean of the five non-zero signal values upstream and downstream (ten positions total) of the zero-valued position. A three-state GHMM was then used to decode the sequence of states that occurred in read 2. It consisted of an initiation state (state 1), a poly(A)-tail state (state 2), and a non-poly(A)-tail state (state 3). All reads start in state 1. From state 1 the model can remain in state 1 or transition to state 2. From state 2 the model can

either remain in state 2 or transition to state 3. The model was initialized with the following transition probabilities:

$from \setminus to$	$state_1$	$state_2$	$state_3$
$state_1$	0.001	0.95	0.049
$state_2$	0.001	0.95	0.049
$state_3$	0.001	0.001	0.998

The initial emissions were Gaussian distributions with means of 100, 1, and −1 and variances of 1, 0.25 and 0.25, respectively. In general, the emission Gaussians for the model corresponded to the logarithm of the calculated T signal at each sequenced base in read 2. The initial state probabilities were 0.998, 0.001, and 0.001 for states 1, 2 and 3, respectively.

After initializing the model, unsupervised training was performed on 10,000 randomly selected PAL-seq tags, and then the trained model was used to decode all tags, with the number of state 2 cycles reporting the poly(A)-tail length for a tag. Only genes with ≥ 50 poly(A)-tail length measurements were considered for analyses involving mean poly(A)-tail lengths.

Analysis of PAL-Seq ss-Ligation Data

To account for mRNAs with very short tails or extensive terminal modifications, we implemented a version of PAL-seq that did not use splinted ligation. Tail lengths from these ss-ligation datasets, acquired for steady-state samples from both cell lines, were determined using a modified version of the PAL-seq analysis pipeline written for python3. The T-signal in this pipeline was modified to allow more accurate detection of mRNAs lacking tails. Instead of normalizing the intensity of each channel for each cycle to the average intensity of that channel when reading that base in read 1, the intensity of each channel was normalized to the average intensity of the channels for the other three bases in read 1. The intensity of the T channel was

then divided by the sum of the other channel intensities to calculate the T signal, and tails were called using the hmmlearn package (v0.2.0). Tags representing short tails, including short tails that ended with many non-A residues, were identified as those for which read 1 and read 2 mapped to the same mRNA 3′ UTR (usually ∼4% of the tags). Tail lengths for these tags were called without the use of the GHMM. Instead, their tail lengths were determined by string matching, allowing any number of untemplated U residues but no more than two G or C residues to precede the A stretch. Tags not identified as representing short-tails were analyzed using the GHMM, excluding from further analysis occasional outliers determined by the GHMM to have tails ≤ 8 nt.

Most of the tags that had either only a very short tail or no tail did not correspond to mRNA cleavage-and-polyadenylation sites. Therefore, to be carried forward in our analysis, short-tailed tags were required to have a 3′-most genome mapping position (as determined from read 1 but requiring that read 2 also map uniquely to the same 3′ UTR) that fell within a 10 nt window of a PAL-seq–annotated cleavage-and-polyadenylation site.

Although the single-stranded ligation protocol provided the opportunity to account for mRNAs with very short or highly modified tails, examination of the recovery of internal standards indicated that tags representing longer tails (≥ 100 nt) were not as well recovered in the datasets in which we implemented ss ligation. Therefore, for steady-state samples from each cell line, we generated composite tail-length distributions in which the ss-ligation dataset contributed to the distribution of tails < 50 nt, and the splinted-ligation dataset contributed to the distribution of tails ≥ 50 nt. For example, *Slc38a2* had 635 standard PAL-seq tags, 169 of which (∼27%) had tails < 50 nt, and this same gene had 703 ss-ligation PAL-seq tags, 393 of which (∼56%) had tails < 50 nt. The composite tail-length distribution replaced the 169 short-tailed splinted-ligation

PAL-seq tags with the 393 short-tailed ss-ligation PAL-seq tags, normalizing the latter cohort by a scaling factor. This scaling factor was determined from the ratio of the counts of the splinted-ligation tags with tail lengths between 30–70 nt (135 tags) to the counts of the corresponding tags in the ss-ligation dataset (153 tags).

3′-end annotations were generated from PAL-seq tags with tails ≥ 11 nt, using an algorithm previously developed for data from poly(A)-position profiling by sequencing (3P-seq) (Jan et al., 2011). Each PAL-seq read 1 that mapped (with at least 1 nt of overlap) to an annotated 3′ UTR (Eichhorn et al., 2014) was compiled by the genomic coordinate of its 3′-UTR nucleotide closest to the tail. The genomic coordinate with the most mapped reads was annotated as a 3′ end. All reads within 10 nt of this end (a 21 nt window) were assigned to this end and removed from subsequent consideration. This process was repeated until there were no remaining 3′ UTR-mapped reads. For each gene, the 3′-end annotations were used in subsequent analyses if they accounted for ≥ 10% of the 3′ UTR-mapping reads for that gene.

Documentation and code to calculate and analyze T signals and determine tail lengths are available for both the splinted-ligation and ss-ligation pipelines at https://github.com/kslin/PAL-seq.

TAIL-Seq

The 2 h time-interval TAIL-seq sample used for comparison with PAL-seq was prepared using the same library cDNA as was used for PAL-seq v2 libraries, but amplifying the library using different primers (Table S3). The first read of TAIL-seq involved sequencing the 3′ UTR from the gene body toward the tail, with the sequencing primer annealing to sequences added with the 5′ adaptor. This 5′ adaptor was an equimolar mixture of four sequences with different numbers of

nucleotides in between the primer binding site and the insert (Table S3) to ensure that highly abundant sequences (such as rRNA fragments) did not cause a large portion of the flow cell to fluoresce in a single channel. Amplification and purification were as for PAL-seq v2. Samples were sequenced with either a paired-end 50-by-250 run (2 h time-interval sample) using a HiSeq 2500 operating in normal mode using a v3 kit. Other Illumina sequencing chemistries (including v1, v2, and v4 kits run in rapid and normal modes) did not yield accurate tail-length measurements when used in paired-end mode. Analysis was as described for PAL-seq v2, except a five-state GHMM was used (Chang et al., 2014) to accommodate the difference in the nature of the T-signal output imparted by the different mode of sequencing. The five states were an initiation state, a poly(A) state, a poly(A) transition state, a non-poly(A) transition state, and a non-poly(A) state.

RNA-Seq

Fragmented poly(A)-selected RNAs were supplemented with three short quantification standards (Table S3), and then ligated to adapters, reverse-transcribed, and amplified to prepare the RNA-seq and ribosome-profiling libraries, respectively (Subtelny et al., 2014). These libraries were sequenced on an Illumina HiSeq 2500. For all RNA-seq data, only reads mapping to ORFs of annotated gene models (Eichhorn et al., 2014) were considered, excluding the first 50 nt of each ORF. The latter requirement which was implemented to match ribosome-profiling data of a concurrent study examining the effects of miRNAs (Eisen et al., 2020). A cutoff of ≥ 10 reads per million mapped reads (RPM) was applied to each gene in each sample.

Calculation of mRNA Half-Lives

Half-lives were estimated independently from both RNA-seq data and PAL-seq tag abundance. Prior to half-life fitting, mRNA abundances were normalized across time intervals based on the quantification standards added to each sample prior to library preparation.

Half-lives associated with the mRNAs from each gene *j* were determined by fitting to the equation

$$m_j(t_i) = \delta \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_i - t_{off})}\right) \quad (1)$$

in the case of the continuous-labeling experiment, or to the equation

$$m_j(t_i) = \frac{\alpha_j}{\beta_j} e^{-\beta_j t_i} + c_j \quad (2)$$

in the case of the transcriptional shutoff experiment, where $m_j(t_i)$ is the expression of mRNA *j* at time *t*, α_j is the rate constant for mRNA production, β_j is the rate constant for mRNA degradation, t_{off} is a global time offset, δ is a global scaling parameter to adjust the steady-state time point, and c_j is a baseline for the final expression of each gene in the transcriptional-shutoff experiment. Because the quantification standards were not applicable to the steady-state samples (as these samples did not undergo 5EU purification), the steady-state samples were normalized by a globally fitted constant (setting t_i to 100 h for this time interval).

Because the half-life fitting for the continuous-labeling experiment required the global parameters t_{off} and δ, half-lives for all genes needed to be fit simultaneously. Accordingly, we minimized the least-squares error loss function, $L_2(p)$:

$$L_2(p) = \sum_{i}^{I} \sum_{j}^{J} \left[\ln\left(m_{ij}(p)\right) - \ln\left(y_{ij}\right)\right]^2, \tag{3}$$

for the simulated number of normalized tags at time point *i* for gene *j,* m_{ij}, and the observed number, y_{ij}. The total number of time points and genes are denoted by *I* and *J*, respectively. L_2 depends on the parameters $p = (\alpha_1, \alpha_2, \ldots, \alpha_J, \beta_1, \beta_2, \ldots, \beta_J, t_{off}$, and $\delta)$. The optimization for α_j, β_j, t_{off}, and δ was performed using the "L-BFGS-B" method in the *optim* function in R.

To increase the efficiency of the optimization, we also calculated and implemented the analytical gradient of the model-and-loss function. This gradient computed the quantity $\frac{dL_2}{dp}$ which, when passed to the optimizer, decreased the number of iterations required to minimize the loss. This quantity was computed for each of the parameters as follows:

$$\frac{dL_2}{d\alpha_j} = \sum_{i}^{I} \zeta \frac{\delta}{\beta_j}\left(1 - e^{-\beta_j(t_i - t_{off})}\right) \tag{4.1}$$

$$\frac{dL_2}{d\beta_j} = \sum_{i}^{I} \zeta\left(\left(\frac{\alpha_j}{{\beta_j}^2}\right) e^{-\beta_j(t_i - t_{off})} - \left(\frac{\alpha_j}{{\beta_j}^2}\right) + \left(\frac{\alpha_j}{\beta_j}\right) t_i e^{-\beta_j(t_i - t_{off})} - \left(\frac{\alpha_j}{\beta_j}\right) t_{off} e^{-\beta_j(t_i - t_{off})}\right)(\delta) \tag{4.2}$$

$$\frac{dL_2}{dt_{off}} = -\sum_{i}^{I} \sum_{j}^{J} \zeta \delta \alpha_j e^{-\beta_j(t_i - t_{off})}, \tag{4.3}$$

$$\frac{dL_2}{d\delta} = \sum_{i}^{I} \sum_{j}^{J} \zeta \frac{\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_i - t_{off})}\right), \tag{4.4}$$

where ζ is the first component of the derivative of the loss function

$$\zeta = 2\left(\ln\left(\frac{\delta\alpha_j}{\beta_j}\left(1 - e^{-\beta_j(t_i - t_{off})}\right)\right) - \ln\left(y_{ij}\right)\right)\frac{\beta_j}{\delta\alpha_j\left(1 - e^{-\beta_j(t_i - t_{off})}\right)}. \quad (5.1)$$

δ is constrained such that its value is 1 for non-steady–state time points and only fit at steady state (i.e. fit when $t_i = t_{ss}$ and 1 otherwise).

The range of rate constants fit to this exponential model and the subsequent deadenylation model was bounded to reflect the lack of confidence in values of and differences between extreme outliers. Half-life values were bounded to fall between 6 min and 100 h, deadenylation rate constants were truncated to fall between 0.03 and 30 nt/min, decay rate constants at 20 nt were truncated to fall between 0.003 and 3 min^{-1}, and production rate constants were truncated to fall between 10^{-8} and 10^{-5} normalized reads/min. Reads were spike-normalized and background subtracted, as detailed in subsequent methods.

Model of mRNA metabolism

The model of mRNA production, deadenylation, and decapping (decay) was a system of differential equations

$$\frac{dA_l}{dt} = k_0 - (k_1 + k_2)A_l \quad (6.1)$$

$$\frac{dA_{l-1}}{dt} = k_0 + k_1A_l - (k_1 + k_2)A_{l-1} \quad (6.2)$$

$$\frac{dA_{l-2}}{dt} = k_0 + k_1A_{l-1} - (k_1 + k_2)A_{l-2} \quad (6.3)$$

$$\vdots \quad (6.4)$$

$$\frac{dA_0}{dt} = k_0 + k_1(A_1) - k_2(A_0)\,,$$

where A_l is an mRNA with tail length l, and k_0, k_1, and k_2 are rate constants that describe the production, deadenylation and decay rates, respectively. The final deadenylation product (A_0) has a deadenylation rate constant of zero, as it has no tail. The rate constants k_0 and k_2 are themselves functions of tail length (l), specified by the respective negative binomial and logistic functions

$$k_0(l) = \frac{\alpha\Gamma(v_p + l)}{l!\,\Gamma(v_p)}\left(\frac{m_p}{v_p + m_p}\right)^l\left(\frac{v_p}{v_p + m_p}\right)^{v_p} \tag{7.1}$$

$$k_2(l) = \frac{\beta}{\left(1+e^{-\frac{l-m_d}{v_d}}\right)}\,, \tag{7.2}$$

where the parameters α, β, v_p, m_p, m_d, v_d, are fitted parameters. The parameters α and β are scaling terms for production and decay distributions, respectively. The parameters m_p and m_d describe the expected value of those distributions, and v_p and v_d describe the spread.

Equations (6) were re-written as a linear, time-invariant (LTI) system (Dahleh et al., 2004)

$$\dot{x} = \begin{bmatrix} -k_1 - k_2 & 0 & \dots & 0 & 0 \\ k_1 & -k_1 - k_2 & & & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & & & -k_1 - k_2 & 0 \\ 0 & 0 & \dots & k_1 & -k_2 \end{bmatrix} x(t) + \begin{bmatrix} k_0 \\ k_0 \\ \vdots \\ k_0 \\ k_0 \end{bmatrix}, \tag{8}$$

or, more succinctly, as

$$\dot{x} = Cx(t) + D\,, \tag{9}$$

where the coefficient matrix, C, is specified by the coefficients of the differential equations (6), and the source vector D is specified by the production rate. C is a 251×251 matrix, whereas D,

$x(t)$, and $\dot{x}$ are 251×1 vectors. Although k_0 and k_2 do not depend on time, they do depend on tail length (Equations 7). In the case of the continuous-labeling experiment, $x(t = 0) = \mathbf{0}$. The transcriptional-shutoff experiment begins with $x(t = -1 \text{ h}) = \mathbf{0}$, but $x(t = 0)$ is determined by the values of the system after 1 h of simulation.

Equation (9) has the analytical solution

$$x(t) = C(e^{Ct} - I)D, \tag{10}$$

where I is the identity matrix and e^{Ct} is the matrix exponential of the coefficient matrix scaled by time. Both this analytical solution and numerical integrators (which do not require an analytical solution) can be used to compute the result. We found that numerical stability and computational efficiency were optimal when using the LSODE solver with parameters set for a banded Jacobian matrix in the *deSolve* package (v1.21) of R, with the model written in C and dynamically loaded into R. Increasing the number of allowed tail-length states from 250 to 300 had little effect on the resulting fitted rate constants but greatly increased computation time.

The model yielded abundances for each tail-length isoform at each time interval, using seven parameters for each gene, three of which were shared across all of the genes (Table S2). From these abundances, the residual sum of squares was computed from the corresponding standard-normalized PAL-seq datasets. Although the 250 nt tail length was modeled, measurements for this length were not available from PAL-seq v2, and thus were excluded from the fitting. Likewise, tail-lengths < 20 were modeled for all time intervals, but because the abundance of tail lengths < 20 nt was only available for steady state, these lengths were excluded from fitting all but the steady-state interval. As a result, for the continuous-labeling experiment using cell line 1, parameters for each gene were fit to 1400 data points (230 tail lengths × 5 time intervals + 250 for the steady-state), and for the experiment using the cell line 2, parameters for

each gene were fit to 1170 data points. The optimization was performed using the "L-BFGS-B" method in the *optim* function of R, or, in the case of the global fitting, using the "L-BFGS-B" method in the *NLopt* package (v1.0.4) of R.

A simple L_2 loss function skewed the fits to the time intervals that had larger values. A common solution to this problem is to fit to log-transformed values, but because our data were sparse, with many tail-length positions having zero tags, pseudo-counting to allow log-space fitting resulted in poor fits. Therefore, residuals were variance weighted using the loss function

$$L_2(p) = \sum_{i}^{I} \sum_{j}^{J} \sum_{k}^{K} \frac{(x_{ijk}(p) - y_{ijk})^2}{\mathrm{Var}(y_i)}, \tag{11}$$

where i, j, and k are the time-interval, gene, and tail length with I, J and K as the maximal values of time-intervals, genes, and tail lengths, and $\mathrm{Var}(y_i)$ is the variance of the dataset from time-interval i.

The model is constrained by the 0-tail-length species, which builds up when decay is slow with respect to deadenylation. Such a buildup was observed in the steady-state tail-length distribution of short-lived mRNAs but occurred primarily between 0 and 20 nucleotides (Figure 6C). Because of this discrepancy, a composite residual was calculated for the model and the data. Abundances for tails < 20 nt were averaged and this average was used to replace the abundances for each tail length < 20 nt for the steady-state data. In addition, when comparing the associated short tails from the data and the model, the residuals for tails < 20 nt were weighted by either 6- or 5-fold (cell lines 1 and 2, respectively) to account for opting not to fit to measurements for tails < 20 nt in the non-steady–state samples.

As with gene-specific parameters, global parameters v_p, m_d, and v_d were fit using pre-steady-state measurements of tails ranging from 20–249 nt. The composite steady-state tail-

length distributions of Figure 2A were also used, which constrained buildup of short-tailed mRNAs. Fitting was performed on subsets of 100 genes (selected randomly without replacement from genes with composite steady-state distributions, yielding 22 and 15 subsets for cell lines 1 and 2, respectively), including v_p, m_d, and v_d in the parameter vector. Median values of the global parameters (Table S1) were then used to fit each gene-specific parameter.

Bootstrap Analysis

Tags in the cell line 1 PAL-seq dataset were resampled 10 times with replacement and assigned to a gene and tail length based on a multinomial probability distribution generated from the counts for each tail length in the original dataset. These resampled datasets were then used for background subtraction, global parameter determination, and model fitting.

Background Subtraction for PAL-Seq Data

Although the efficacy of the 5EU purification enabled efficient enrichment of labeled RNAs at short time intervals (Eisen et al., 2020), we also modeled and corrected for residual background caused by non-specific binding of the unlabeled RNA to the streptavidin beads (Figure S2F).

We designed our background model under the assumption that the background in the time courses stems primarily from the capture of a fixed amount of non-5EU labeled mRNA during the 5EU purification. Accordingly, we subtracted a fraction (0.3%) of the steady-state data from each continuous-labeling dataset. This fraction of input sample was chosen such that at 40 min long-lived genes (half-life $\geq$ 8 h) had no mRNAs with tail lengths $\lessapprox$100 nt on average, but short-lived genes (half-life $\leq$ 30 min) were unaffected (Figure S2F). Likewise, we subtracted standard-normalized time-interval–matched input data from each transcriptional-inhibition

dataset, as actD influenced which unlabeled cellular mRNAs were available to contribute to the background. The fraction of each input sample to subtract was chosen such that at 0 h long-lived genes (half-life ≥ 8 h) had no mRNAs with tail lengths ≲100 nt on average, but short-lived genes (half-life ≤ 30 min) were unaffected. Genes were included in the final background-subtracted set only if the sum of their background-subtracted tag counts was ≥ 50 tags.

After background subtraction, PAL-seq datasets were scaled to each time interval by matching the total number of background-subtracted tags for all genes at all tail lengths to the total number of tags for all genes for the corresponding time interval in the RNA-seq data. The scaled PAL-seq data were then used to compute half-lives for each gene, scaling the steady-state sample using a globally fitted constant.

ActD Treatment

Cell line 2 was cultured as in the continuous-labeling experiments. We prepared 2, 2, 2, 3, and 4 500 cm^2 plates for the 0, 1, 3, 7 and 15 h time intervals, respectively. 5EU (400 μM final) was added to each plate (with one non-5EU plate prepared in parallel), and after 1 h actD (5 μg/mL final concentration, Sigma-Aldrich) was added. Cells were harvested as described for the continuous-labeling experiments, except that a quantitative spike RNA containing 5EU and corresponding to the chloramphenicol-resistance gene sequence (Table S3) was added to the lysis buffer at a concentration of 0.57 ng/mL, or 2 ng/plate. This RNA was prepared using an in vitro transcription reaction as above, with a 5EUTP-to-UTP ratio of 1:20.

Quantification and Statistical Analysis

Graphs were generated and statistical analyses were performed using R (R Core Team, 2019). Statistical parameters including the value of n, statistical test, and statistical significance (p value) are reported in the figures and their legends. No statistical methods were used to predetermine sample size.

Data and Code Availability

Raw and processed RNA-seq, PAL-seq, and TAIL-seq read data are available at the GEO, accession number GEO: GSE134660. Code for configuring an Illumina HiSeq 2500 machine for PAL-seq and for calculation of tail lengths from PAL-seq or TAIL-seq data are available at https://github.com/kslin/PAL-seq. Code for fitting kinetic models of abundance and tail length is available at https://github.com/timeisen/DynamicsOfCytoplasmicMrnaMetabolism.

References

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

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell *173*, 20-51.

Bönisch, C., Temme, C., Moritz, B., and Wahle, E. (2007). Degradation of hsp70 and other mRNAs in Drosophila via the 5' 3' pathway and its regulation by heat shock. J Biol Chem *282*, 21818-21828.

Cai, L., Dalal, C.K., and Elowitz, M.B. (2008). Frequency-modulated nuclear localization bursts coordinate gene regulation. Nature *455*, 485-490.

Cao, D., and Parker, R. (2001). Computational modeling of eukaryotic mRNA turnover. RNA *7*, 1192-1212.

Carballo, E., Lai, W.S., and Blackshear, P.J. (2000). Evidence that tristetraprolin is a physiological regulator of granulocyte-macrophage colony-stimulating factor messenger RNA deadenylation and stability. Blood *95*, 1891-1899.

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, C.Y., and Shyu, A.B. (1995). AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem Sci *20*, 465-470.

Chen, C.Y., Xu, N., and Shyu, A.B. (1995). mRNA decay mediated by two distinct AU-rich elements from c-fos and granulocyte-macrophage colony-stimulating factor transcripts: different deadenylation kinetics and uncoupling from translation. Mol Cell Biol *15*, 5777-5788.

Dahleh, M., Dahleh, M.A., and Verghese, G. (2004). Lectures on dynamic systems and control. A+ A *4*, 1-100.

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 Dev *7*, 1632-1643.

Dellavalle, R.P., Petersen, R., and Lindquist, S. (1994). Preferential deadenylation of Hsp70 mRNA plays a key role in regulating Hsp70 expression in Drosophila melanogaster. Mol Cell Biol *14*, 3646-3659.

Deneke, C., Lipowsky, R., and Valleriani, A. (2013). Complex degradation processes lead to non-exponential decay patterns and age-dependent decay rates of messenger RNA. PLoS One *8*, e55442.

Dölken, L., Ruzsics, Z., Radle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P.*, et al.* (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959-1972.

Du, H., Zhao, Y., He, J., Zhang, Y., Xi, H., Liu, M., Ma, J., and Wu, L. (2016). YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. Nat Commun *7*, 12626.

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.

Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., and Bartel, D.P. (2020). MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. Mol Cell *77*, 775-785 e778.

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.

Fabian, M.R., Frank, F., Rouya, C., Siddiqui, N., Lai, W.S., Karetnikov, A., Blackshear, P.J., Nagar, B., and Sonenberg, N. (2013). Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. Nat Struct Mol Biol *20*, 735-739.

Forrest, M.E., Narula, A., Sweet, T.J., Arango, D., Hanson, G., Ellis, J., Oberdoerffer, S., Coller, J., and Rissland, O.S. (2018). Codon usage and amino acid identity are major determinants of mRNA stability in humans. bioRxiv.

Gowrishankar, G., Winzen, R., Bollig, F., Ghebremedhin, B., Redich, N., Ritter, B., Resch, K., Kracht, M., and Holtmann, H. (2005). Inhibition of mRNA deadenylation and degradation by ultraviolet light. Biol Chem *386*, 1287-1293.

Haas, G., Braun, J.E., Igreja, C., Tritschler, F., Nishihara, T., and Izaurralde, E. (2010). HPat provides a link between deadenylation and decapping in metazoa. J Cell Biol *189*, 289-302.

Hilgers, V., Teixeira, D., and Parker, R. (2006). Translation-independent inhibition of mRNA deadenylation during stress in Saccharomyces cerevisiae. RNA *12*, 1835-1845.

Hu, J., Li, Y., and Li, P. (2013). MARVELD1 Inhibits Nonsense-Mediated RNA Decay by Repressing Serine Phosphorylation of UPF1. PLoS One *8*, e68291.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Comput Sci Eng *9*, 90-95.

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.

Jao, C.Y., and Salic, A. (2008). Exploring RNA transcription and turnover in vivo by using click chemistry. Proc Natl Acad Sci U S A *105*, 15779-15784.

Jia, H., Wang, X., Liu, F., Guenther, U.P., Srinivasan, S., Anderson, J.T., and Jankowsky, E. (2011). The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. Cell *145*, 890-901.

Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet *16*, 421-433.

Kwak, J.E., and Wickens, M. (2007). A family of poly(U) polymerases. RNA *13*, 860-867.

Leppek, K., Schott, J., Reitter, S., Poetz, F., Hammond, M.C., and Stoecklin, G. (2013). Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. Cell *153*, 869-881.

Lim, J., Ha, M., Chang, H., Kwon, S.C., Simanshu, D.K., Patel, D.J., and Kim, V.N. (2014). Uridylation by TUT4 and TUT7 marks mRNA for degradation. Cell *159*, 1365-1376.

Lim, J., Kim, D., Lee, Y.S., Ha, M., Lee, M., Yeo, J., Chang, H., Song, J., Ahn, K., and Kim, V.N. (2018). Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. Science *361*, 701-704.

Lim, J., Lee, M., Son, A., Chang, H., and Kim, V.N. (2016). mTAIL-seq reveals dynamic poly(A) tail regulation in oocyte-to-embryo development. Genes Dev *30*, 1671-1682.

Lima, S.A., Chipman, L.B., Nicholson, A.L., Chen, Y.H., Yee, B.A., Yeo, G.W., Coller, J., and Pasquinelli, A.E. (2017). Short poly(A) tails are a conserved feature of highly expressed genes. Nat Struct Mol Biol *24*, 1057-1063.

Mauxion, F., Chen, C.Y., Seraphin, B., and Shyu, A.B. (2009). BTG/TOB factors impact deadenylases. Trends Biochem Sci *34*, 640-647.

Mercer, J.F., and Wake, S.A. (1985). An analysis of the rate of metallothionein mRNA poly(A)-shortening using RNA blot hybridization. Nucleic Acids Res *13*, 7929-7943.

Mor, A., Suliman, S., Ben-Yishay, R., Yunger, S., Brody, Y., and Shav-Tal, Y. (2010). Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nat Cell Biol *12*, 543-552.

Morozov, I.Y., Jones, M.G., Razak, A.A., Rigden, D.J., and Caddick, M.X. (2010). CUCU modification of mRNA promotes decapping and transcript degradation in Aspergillus nidulans. Mol Cell Biol *30*, 460-469.

Muhlemann, O., and Lykke-Andersen, J. (2010). How and where are nonsense mRNAs degraded in mammalian cells? RNA Biol *7*, 28-32.

Muhlrad, D., Decker, C.J., and Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. Genes Dev *8*, 855-866.

Nelson, J.O., Moore, K.A., Chapin, A., Hollien, J., and Metzstein, M.M. (2016). Degradation of Gadd45 mRNA by nonsense-mediated decay is essential for viability. Elife *5*.

Oliphant, T.E. (2007). Python for scientific computing. Comput Sci Eng *9*, 10-20.

Ozgur, S., Chekulaeva, M., and Stoecklin, G. (2010). Human Pat1b connects deadenylation with mRNA decapping and controls the assembly of processing bodies. Mol Cell Biol *30*, 4308-4323.

Palatnik, C.M., Storti, R.V., and Jacobson, A. (1979). Fractionation and functional analysis of newly synthesized and decaying messenger RNAs from vegetative cells of Dictyostelium discoideum. J Mol Biol *128*, 371-395.

Park, E., and Maquat, L.E. (2013). Staufen-mediated mRNA decay. Wiley Interdiscip Rev RNA *4*, 423-435.

Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. Mol Cell *25*, 635-646.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V.*, et al.* (2011). Scikit-learn: Machine Learning in Python. J Mach Learn Res *12*, 2825-2830.

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.

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N.*, et al.* (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat Biotechnol *29*, 436-442.

Radhakrishnan, A., Chen, Y.H., Martin, S., Alhusaini, N., Green, R., and Coller, J. (2016). The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. Cell *167*, 122-132 e129.

Rissland, O.S., Mikulasova, A., and Norbury, C.J. (2007). Efficient RNA polyuridylation by noncanonical poly(A) polymerases. Mol Cell Biol *27*, 3612-3624.

Rissland, O.S., and Norbury, C.J. (2009). Decapping is preceded by 3' uridylation in a novel pathway of bulk mRNA turnover. Nat Struct Mol Biol *16*, 616-623.

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.

Shav-Tal, Y., Darzacq, X., Shenoy, S.M., Fusco, D., Janicki, S.M., Spector, D.L., and Singer, R.H. (2004). Dynamics of single mRNPs in nuclei of living cells. Science *304*, 1797-1800.

Sheiness, D., and Darnell, J.E. (1973). Polyadenylic acid segment in mRNA becomes shorter with age. Nat New Biol *241*, 265-268.

Shyu, A.B., Belasco, J.G., and Greenberg, M.E. (1991). Two distinct destabilizing elements in the c-fos message trigger deadenylation as a first step in rapid mRNA decay. Genes Dev *5*, 221-231.

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.

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.

Tani, H., Imamachi, N., Salam, K.A., Mizutani, R., Ijiri, K., Irie, T., Yada, T., Suzuki, Y., and Akimitsu, N. (2012). Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA Biol *9*, 1370-1379.

Trcek, T., Larson, D.R., Moldon, A., Query, C.C., and Singer, R.H. (2011). Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. Cell *147*, 1484-1497.

Tullai, J.W., Schaffer, M.E., Mullenbrock, S., Sholder, G., Kasif, S., and Cooper, G.M. (2007). Immediate-early and delayed primary response genes are distinct in function and genomic architecture. J Biol Chem *282*, 23981-23995.

Van Etten, J., Schagat, T.L., Hrit, J., Weidmann, C.A., Brumbaugh, J., Coon, J.J., and Goldstrohm, A.C. (2012). Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. J Biol Chem *287*, 36370-36383.

Vlasova-St Louis, I., and Bohjanen, P.R. (2011). Coordinate regulation of mRNA decay networks by GU-rich elements and CELF1. Curr Opin Genet Dev *21*, 444-451.

Webster, M.W., Chen, Y.H., Stowell, J.A.W., Alhusaini, N., Sweet, T., Graveley, B.R., Coller, J., and Passmore, L.A. (2018). mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. Mol Cell *70*, 1089-1100 e1088.

Wilson, T., and Treisman, R. (1988). Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3' AU-rich sequences. Nature *336*, 396-399.

Wu, Q., Medina, S.G., Kushawah, G., DeVore, M.L., Castellano, L.A., Hand, J.M., Wright, M., and Bazzini, A.A. (2019). Translation affects mRNA stability in a codon-dependent manner in human cells. Elife *8*.

Yamashita, A., Chang, T.C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.Y., and Shyu, A.B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol *12*, 1054-1063.

Yi, H., Park, J., Ha, M., Lim, J., Chang, H., and Kim, V.N. (2018). PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. Mol Cell *70*, 1081-1088 e1085.

Supplemental Figures and Legends

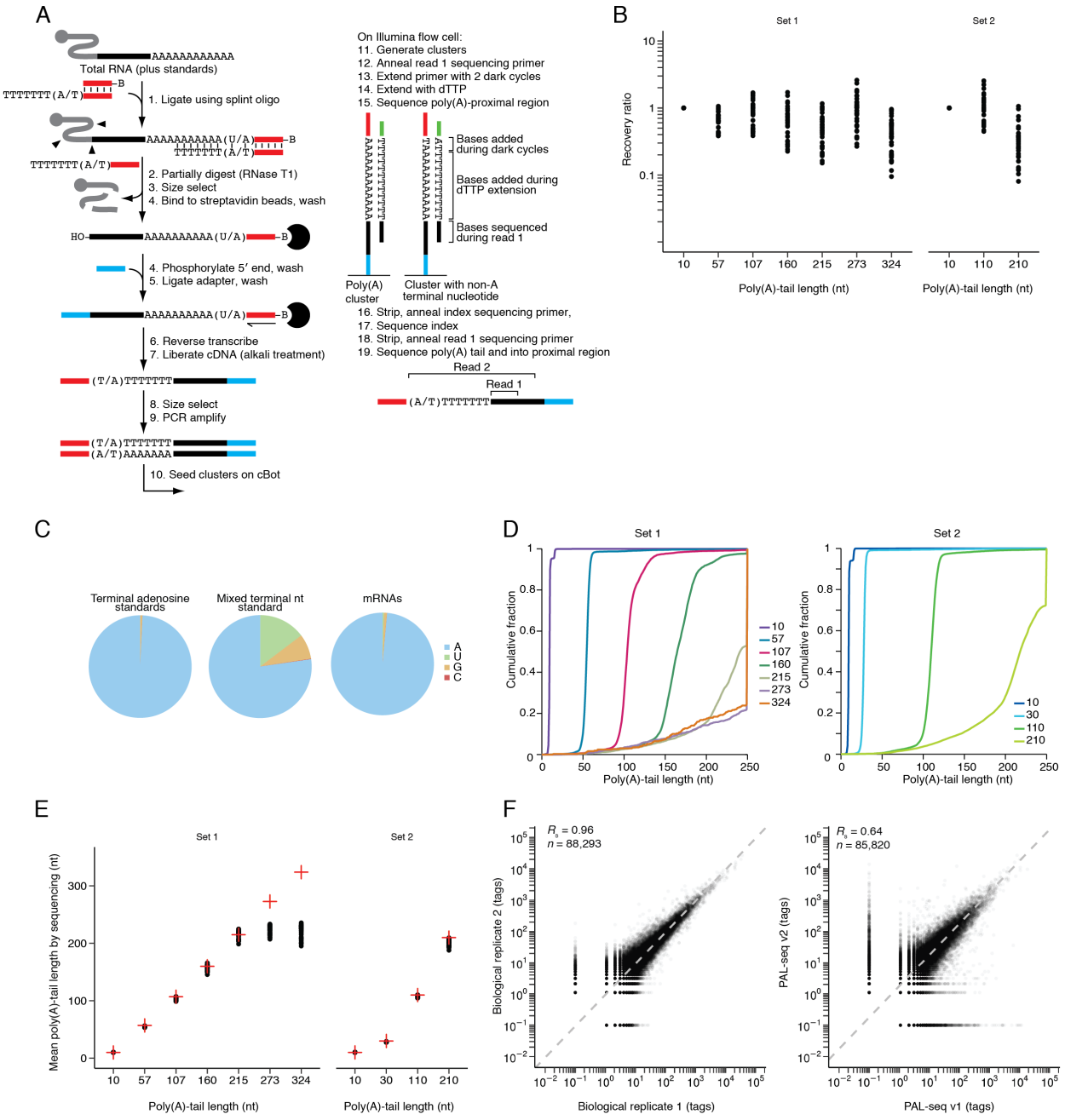


Figure S1. PAL-seq v2 Methodology and Benchmarking, Related to Figure 1

(A) Schematic of PAL-seq v2. The original version of PAL-seq (Subtelny et al., 2014) was modified to include an additional splint oligonucleotide capable of ligating to tails with a terminal U (step 1); two dark cycles prior to the primer-extension reaction (step 13), which prevented non-adenosine terminal residues from terminating the subsequent primer extension; primer extension through the tail with dTTP as the only nucleoside triphosphate (step 14); sequencing on a HiSeq

machine, with the opportunity for multiplexing (steps 16 and 17); an additional read using the read 1 sequencing primer (read 2), which collected sequence and intensity information used to call poly(A)-tail lengths, as in TAIL-seq (Chang et al., 2014) (steps 18 and 19). (B) Recovery of RNA standards. Before preparing libraries, two sets of RNA standards were added to each of the 34 RNA samples analyzed by PAL-seq v2 in this study and an accompanying study (Eisen et al., 2020). Set 1 contained seven RNAs with different tail lengths, and set 2 contained four RNAs with different tail lengths (Table S3). For each set of standards, the relative abundance of each standard in the final sequencing output was compared its relative abundance in the initial standard mixture, and this recovery ratio is plotted for each sample on a log scale. The relative recovery of standards varied somewhat, with no systematic bias that would indicate substantial depletion of poly(A)-tails of certain lengths. The 30 nt standard from set 2 was excluded from this analysis because it is an equal mixture of four different standards that end in a terminal A, C, G or U (Table S3), which was added to assess the ability to detect tails with a terminal U, as described in the next panel. (C) Terminal nucleotide compositions of RNAs with tail measurements ≥5 nt. Libraries were prepared using a 5:1 mixture of splint oligos that would hybridize perfectly to either the 3′ end of RNAs ending in eight adenosines or the 3′ end of RNAs ending in seven adenosines followed by a terminal uridine, respectively. Left: Terminal nucleotide composition of PAL-seq v2 tags from the RNA standards for which poly(A) tails were prepared using poly(A) polymerase and ATP. These standards were expected to terminate exclusively with adenosine. Middle: Terminal nucleotide composition of PAL-seq v2 tags from the synthetic 30 nt standard, which was prepared with a tail designed to have an equal mixture of terminal A, C, G, or U. Although the splint oligonucleotides perfectly matched the versions ending A and U, the terminal U was somewhat depleted compared to the terminal A, and a substantial fraction of terminal G was also captured, perhaps due to wobble-pairing between the T in the splint and the terminal G in the standard. Right: Terminal nucleotide composition of PAL-seq v2 tags from mRNAs. (D) The tail length distributions of the synthetic RNA standards, as measured by PAL-seq v2. Plotted is the cumulative distribution of poly(A)-tail lengths for each standard in the steady-state sample from cell line 1. The poly(A)-tail lengths measured by polyacrylamide gel electrophoresis (Subtelny et al., 2014) are indicated (key). (E) Mean poly(A)-tail lengths of the two sets of synthetic standards, as measured by PAL-seq v2. For each standard, the mean tail length in each of the 34 samples in this study and an accompanying study (Eisen et al., 2020), as measured using PAL-seq v2, is plotted (black points). Also plotted is the mean tail length of each standard, as determined using denaturing gels (red crosses). Tails that exceeded 250 nt were not expected to be measured accurately, because their length exceeded the length of the sequencing read used to measure the tail. (F) Comparison of the frequencies in which UTR 3′ ends were called by biological replicates of PAL-seq v2 (left) or the two different versions of PAL-seq (right). Each point is the number of tags that mapped to a genomic position, adding a pseudo-count of 0.1 tags. Dashed lines represent equivalency, after accounting for different sequencing depths.

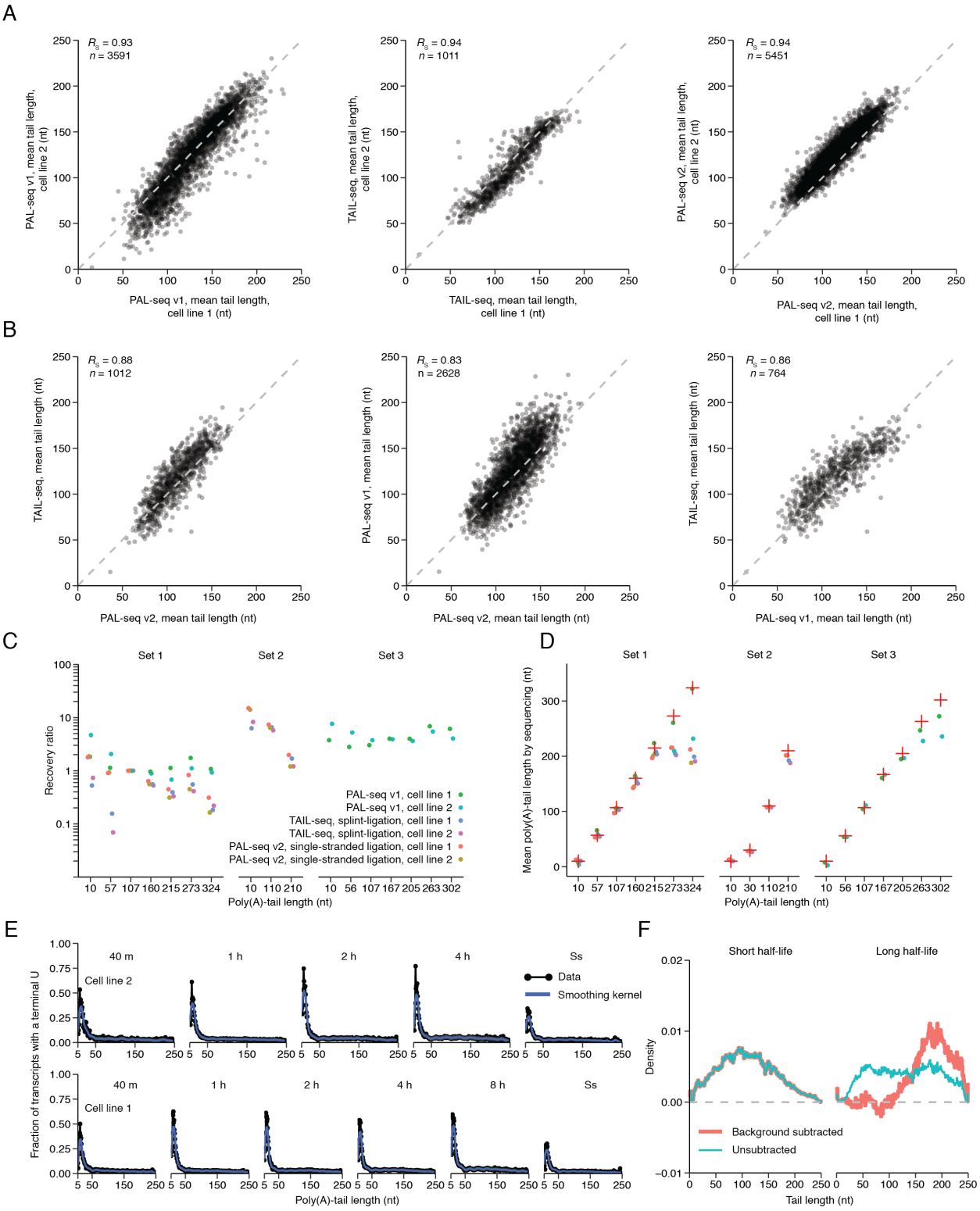


Figure S2. Reproducibility of PAL-seq v2, Related to Figure 1

(A) Comparisons of biological replicates for different library preparation and tail-profiling protocols. For each gene that passed a 50-tag cutoff, mean poly(A)-tail lengths after 2 h of continuous labeling are shown for PAL-seq (left panel), our implementation of TAIL-seq (Lim et

al., 2016), (middle panel), or PAL-seq v2 (right panel). Whereas the RNA examined using TAIL-seq and PAL-seq v2 datasets was isolated using 5EU labeling, the RNA examined using PAL-seq v1 dataset was isolated using 4-thiouridine labeling. The 2 h time interval was chosen for this analysis because its broad range of average tail lengths made it most suitable for comparing the results of different methods (Figure 1C). The dashed line represents y = x. (B) Comparisons between different tail-profiling protocols. Compared are mean tail lengths generated by TAIL-seq and PAL-seq v2 (left panel), PAL-seq v1 and PAL-seq v2 (middle panel), and TAIL-seq and PAL-seq v1 (right panel). Otherwise, as in (A). (C) Recovery of tail standards in the PAL-seq v1, TAIL-seq (splinted ligation) and PAL-seq v2 steady-state (single-stranded ligation) datasets. For analyses of the recovery of tail standards in PAL-seq v2 (splint-ligation) datasets, see Figure S1B. All six libraries contained seven standards from standard set 1; the TAIL-seq libraries contained three standards from standard set 2; and the PAL-seq v1 library contained seven standards from standard set 3. Tail lengths of the standards as determined by polyacrylamide-gel electrophoresis (Subtelny et al., 2014) are indicated (key) and shown as x-axis labels. The 30 nt standard from set 2 was excluded from this analysis because it was an equal mixture of four different standards that ended in A, C, G or U (Table S3) and was added to assess the ability to detect tails with a terminal U. The relative abundances of the standards in the sequencing data were quantified and compared to their relative starting abundance, and this recovery ratio is plotted for all samples. The values of each library were normalized to the abundance of the 107 nt standard in set 1. (D) Mean tail lengths of the standards shown in (C) and the 30 nt standard. Otherwise, as in (C). For analysis of mean tail lengths of the standards in PAL-seq v2 (splint-ligation) datasets, see Figure S1E. (E) Uridylation frequency as a function of tail length. The fraction of single uridine residues at the 3′ end of mRNA-mapping tags is plotted as a function of tail lengths ≥5 nt (black) along with a LOESS smoothing kernel (blue, with 5^{th}–95^{th} percent confidence intervals in grey) for either cell line 2 (top panels) or cell line 1 (bottom panels). These values were scaled by a factor of 5.23 to correct for the depletion of tags containing a terminal uridine, estimated from the ratio of tags mapping to the 30 nt standards terminating with either A or U, which had been added to the libraries at an equal molar ratio (Figure S1C). Uridine fractions corresponding to tail lengths ≥246 nt were combined into one bin at 246 nt. (F) Effects of background subtraction of PAL-seq data at the earliest (40 min) time interval. Distributions of the unsubtracted (blue) and background-subtracted (red) tail lengths for short-lived (half-life <30 min, n = 293 for both unsubtracted and background subtracted) and long-lived (half-life >8 h, n = 379) mRNAs. The background subtraction differentially affected the long half-life mRNAs, as these had a proportionally smaller amount of labeled relative to unlabeled RNA at short time intervals, and thus unlabeled RNAs contributed a larger fraction of their reads at these intervals.

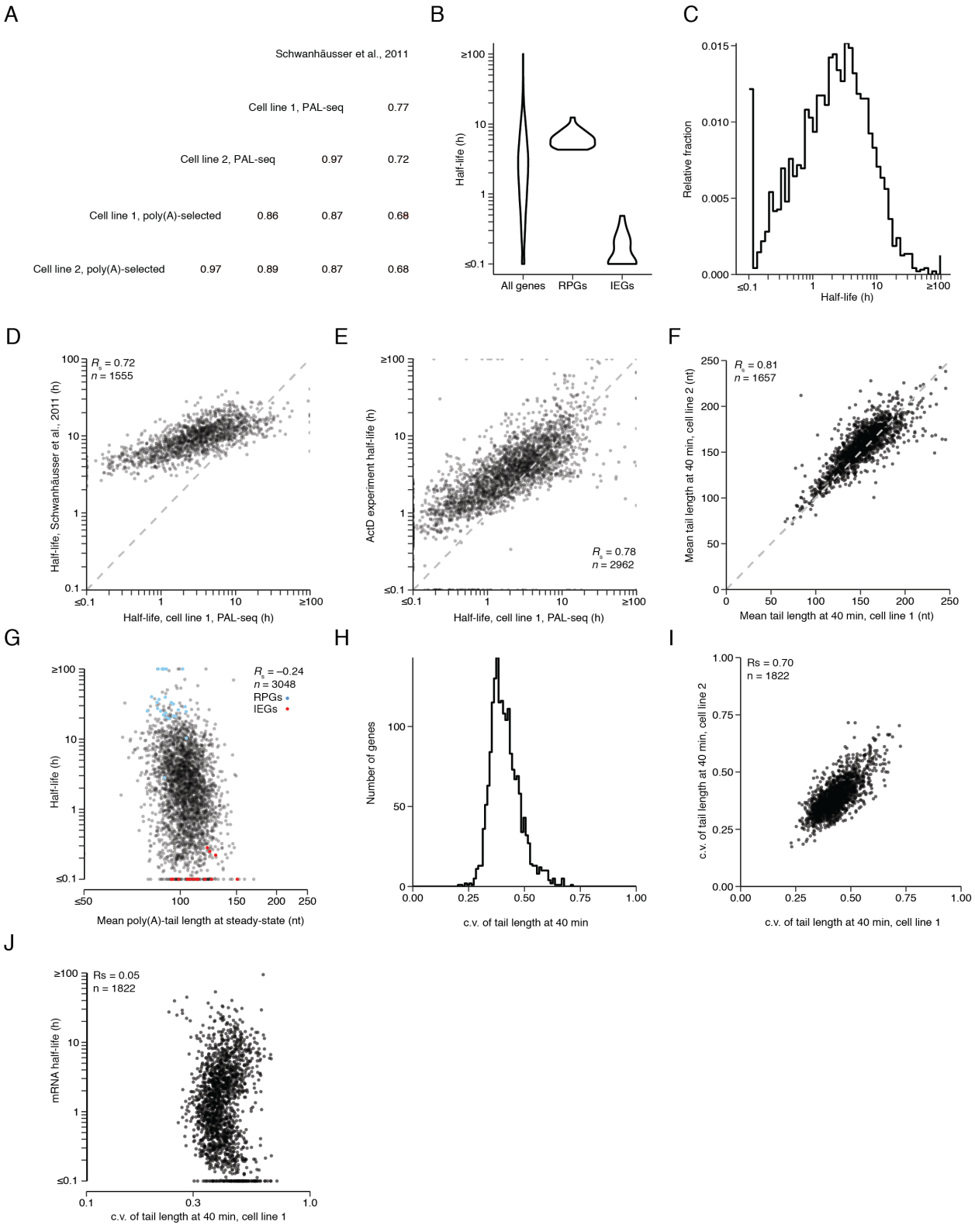

A
Schwanhäusser et al., 2011
Cell line 1, PAL-seq 0.77
Cell line 2, PAL-seq 0.97 0.72
Cell line 1, poly(A)-selected 0.86 0.87 0.68
Cell line 2, poly(A)-selected 0.97 0.89 0.87 0.68
B
Half-life (h)
All genes
RPGs
IEGs
C
Relative fraction
Half-life (h)
D
Half-life, Schwanhäusser et al., 2011 (h)
Half-life, cell line 1, PAL-seq (h)
E
ActD experiment half-life (h)
Half-life, cell line 1, PAL-seq (h)
F
Mean tail length at 40 min, cell line 2 (nt)
Mean tail length at 40 min, cell line 1 (nt)
G
Half-life (h)
Mean poly(A)-tail length at steady-state (nt)
RPGs
IEGs
H
Number of genes
c.v. of tail length at 40 min
I
c.v. of tail length at 40 min, cell line 2
c.v. of tail length at 40 min, cell line 1
J
mRNA half-life (h)
c.v. of tail length at 40 min, cell line 1

Figure S3. Half-life and Initial Tail-Length Measurements, Related to Figures 2 and 3
(A) Pairwise correlations (R_s) of half-life measurements. n = 4485, 4748, 3048, 1743, and 4658 genes for cell line 1 poly(A)-selected, cell line 2 poly(A)-selected, cell line 1 PAL-seq, cell line 2 PAL-seq, and Schwanhäusser et al. 2011 samples, respectively. (B) Distributions of half-lives for mRNAs from all genes (n = 3048), ribosomal-protein genes (RPGs, n = 31), or immediate-early genes (IEGs, n = 19) (Tullai et al., 2007) obtained using PAL-seq data from cell line 1. (C) Distribution of mRNA half-lives obtained using PAL-seq data from cell line 1 (n = 3048). (D) Comparison of published half-life measurements (Schwanhäusser et al., 2011) with those obtained from 5EU continuous labeling. Dashed line is $y = x$. (E) Comparison of half-life measurements from the transcriptional-shutoff experiment and those obtained from the continuous-labeling experiment. Dashed line is $y = x$. (F) Comparison of mean poly(A)-tail lengths of mRNAs isolated from cell line 2 after 40 min of labeling with those isolated from cell line 1 after 40 min of labeling. Dashed line is $y = x$. (G) Relationship between half-life and mean steady-state tail length of mRNAs in 3T3 cells. Tail lengths and half-lives were determined using only standard PAL-seq data, in which the adapter oligo was appended to the tail using splinted ligation. Otherwise, as in Figure 2A. (H) The distribution of c.v. values of tail lengths after 40 min of labeling for mRNAs from each gene. Each c.v. value is the average of two biological replicates. (I) Comparison of c.v. values of tail lengths after 40 min of labeling between two biological replicates. (J) Relationship between mRNA half-life and c.v. values of tail lengths after 40 min of labeling. Each c.v. value is the average of two biological replicates.

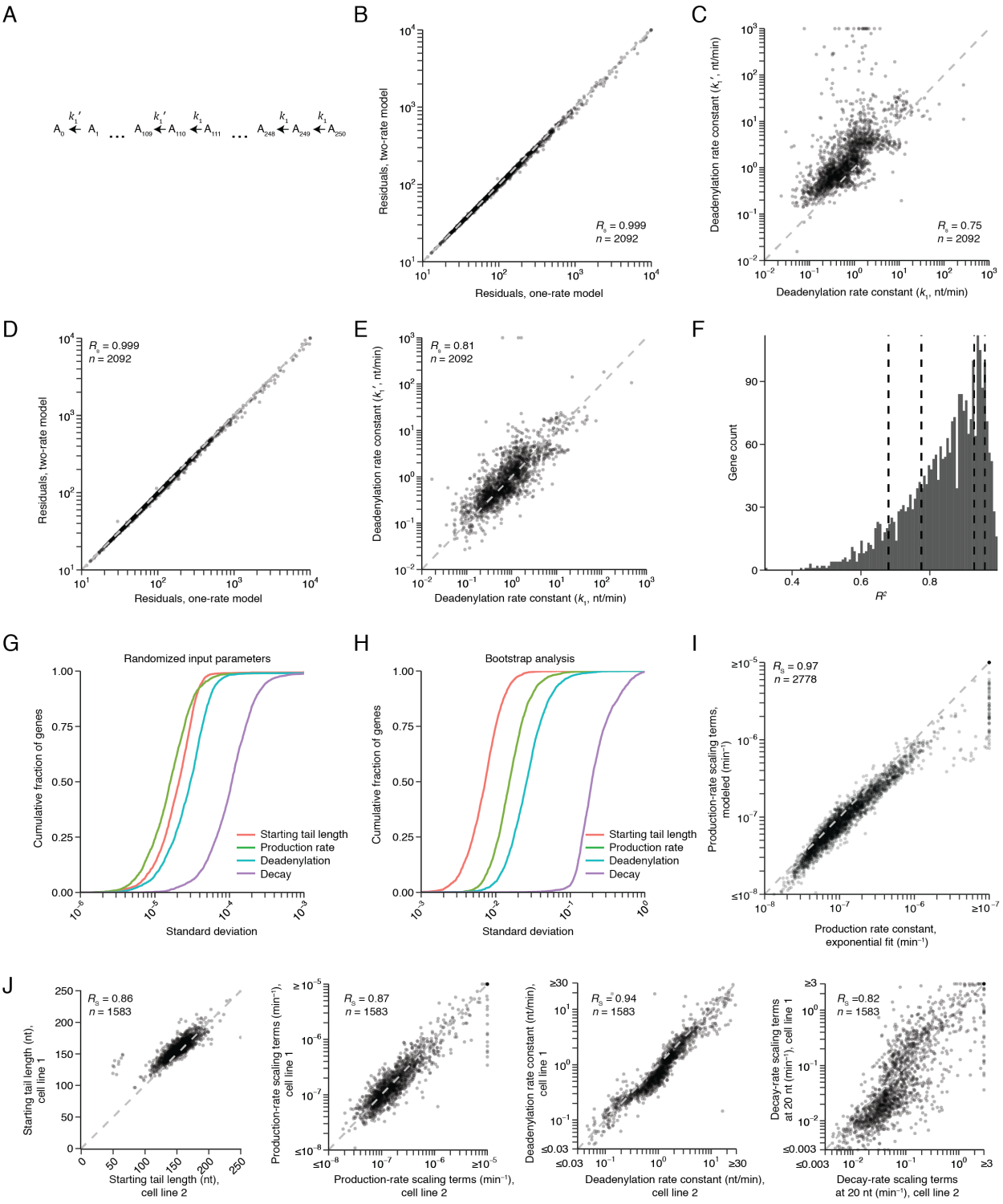


Figure S4. Model Development and Testing, Related to Figures 4 and 5

(A) Schematic of the model with two deadenylation rates. Deadenylation is parameterized with two rate constants, one that describes deadenylation of tail lengths >110 nt (k_1) and one that describes deadenylation of tails ≤110 nt (k_1'). The transition between these rates is determined by a generalized logistic function with a transition parameter arbitrarily set to 1 (a sharp transition).

Otherwise, as in Figure 4A. (B) Comparison of the residual sum of squares (RSS) between the model with two deadenylation rates (A) and the model with one deadenylation rate. Dashed line is $y = x$. (C) Relationship between the second (k_1') and the first (k_1) deadenylation rate constant fit for mRNAs of each gene using the model in (A). Dashed line is $y = x$. (D) Comparison of a model with two deadenylation rates in which the transition between the rates occurs at a tail length of 150 nt and the model with one deadenylation rate. Otherwise, as in (B). (E) Relationship between the second (k_1') and the first (k_1) deadenylation rate fit for mRNAs of each gene using the model in (D). Dashed line is $y = x$. (F) Distribution of R^2 values for all genes fit by the model (n = 2778). Dashed lines indicate the R^2 values of the four genes shown in Figure 4B. (G) Analysis of the robustness of fitted rate constants to input parameter identities. The distributions of s.d. values of rate constants for all fitted genes over 10 rounds of fitting with varying input parameters are displayed as empirical cumulative distributions. The input parameters were randomly selected from a uniform distribution bounded by the 10^{th} to 90^{th} percentiles of rate constants of all genes during a previous round of fitting. Using an unbounded randomized parameter selection resulted in larger variation but also larger final residuals. The s.d. values for 90% of genes were less than 3.7×10^{-5}, 3.5×10^{-5}, 5.8×10^{-5}, and 2.4×10^{-4} for rate constants for starting tail length, production, deadenylation, and decapping, respectively (with all parameters shown as s.d. of the $\log_{10}$ of the value). (H) Bootstrapping analysis of fitted rate constants. For each dataset, the total number of tags was resampled ten times based on a multinomial probability distribution specified by the original tag counts for every tail length position for each gene. These resampled datasets were then background subtracted and fit to the computational model. Shown for each fitted gene-specific parameter is the cumulative distributions of its c.v. values for all fitted genes. The s.d. for 90% of genes were less than 0.01, 0.03, 0.06, and 0.49, for rate constants for starting tail length, production, deadenylation, and decapping, respectively (with all parameters shown as s.d. of the $\log_{10}$ of the value). The greater variation observed for the decapping parameter reflected the relatively few data points effectively used for its fitting, as this parameter related primarily to the short-tailed region of the distribution. (I) Relationship between production rate from an exponential fit to the data and the production rate as determined by the computational model. Dashed line is $y = x$. (J) Model reproducibility across biological replicates. Plots show the relationship between mean starting tail lengths (m_p, left panel), production-rate scaling terms (α, middle left panel), deadenylation rate constants (δ, middle right panel) or decapping-rate scaling terms (β, right panel) for mRNAs from the same genes in the two cell lines. Dashed line is $y = x$.

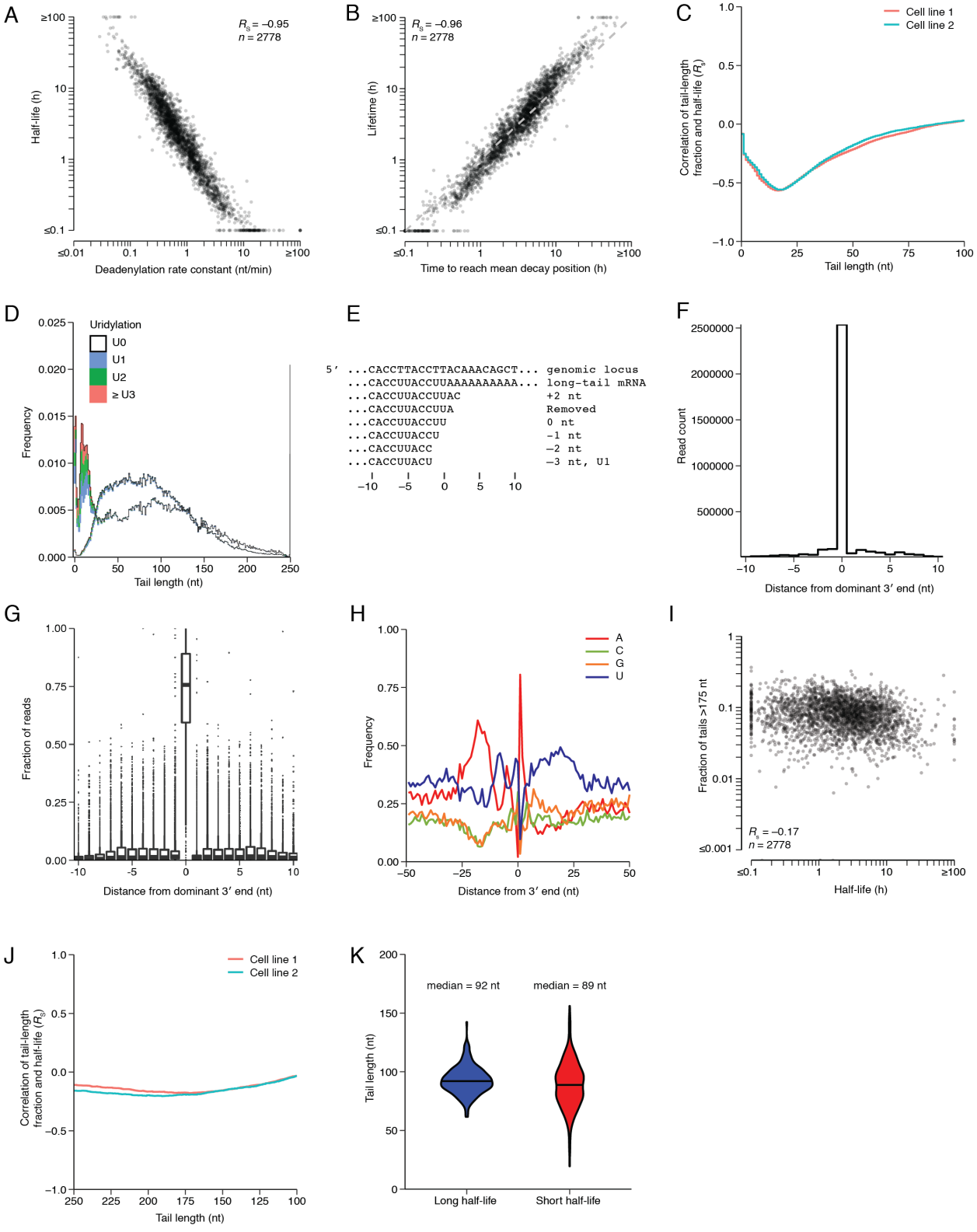

A
Half-life (h)
Deadenylation rate constant (nt/min)
R_S = –0.95
n = 2778
B
Lifetime (h)
Time to reach mean decay position (h)
R_S = –0.96
n = 2778
C
Cell line 1
Cell line 2
Correlation of tail-length fraction and half-life (R_S)
Tail length (nt)
D
Uridylation
U0
U1
U2
≥ U3
Frequency
Tail length (nt)
E
5′ ...CACCTTACCTTACAAACAGCT... genomic locus
...CACCUUACCUUAAAAAAAAAA... long-tail mRNA
...CACCUUACCUUAC +2 nt
...CACCUUACCUUA Removed
...CACCUUACCUU 0 nt
...CACCUUACCU -1 nt
...CACCUUACC –2 nt
...CACCUUACU –3 nt, U1
F
Read count
Distance from dominant 3′ end (nt)
G
Fraction of reads
Distance from dominant 3′ end (nt)
H
A
C
G
U
Frequency
Distance from 3′ end (nt)
I
Fraction of tails >175 nt
R_S = –0.17
n = 2778
Half-life (h)
J
Cell line 1
Cell line 2
Correlation of tail-length fraction and half-life (R_S)
Tail length (nt)
K
median = 92 nt
median = 89 nt
Tail length (nt)
Long half-life
Short half-life

Figure S5. Analyses of Deadenylation Rate Constants and Steady-State Tail Lengths and Modifications, Related to Figures 5 and 6

(A) Relationship between mRNA half-life, as determined by an exponential fit to abundance, and deadenylation rate constant, as determined by the model. Because the structure of the computational model enhanced the correspondence between half-life and deadenylation rate constants, analyses performed with primary data (Figure 2B and Figure 7D) provide a more accurate indication of the correspondence between mRNA half-life and deadenlylation rate. (B) Relationship between measured mRNA lifetime and mRNA lifetime inferred by the model. For mRNAs from each gene, the number of tail nucleotides separating the mean starting tail length (Figure S4J) and the mean tail length at decapping (Figure 5C) was multiplied by the deadenylation rate constant (Figure 5A). Measured lifetime (the inverse of the degradation rate from an exponential fit) was then compared to this inferred lifetime. The dashed line indicates y = x. (C) Relationship between the steady-state abundance of short-tailed transcripts and mRNA half-life. For each tail length from 1–100 nt, the fraction of mRNAs with tail lengths that fell below that position was calculated for each gene, and the relationship (R_s) between this fraction falling below the tail-length cutoff and half-life was determined as in Figure 6A. These 100 R_s values are plotted as a function of the tail-length cutoff used to classify short-tailed transcripts. This analysis started with the composite steady-state tail-length distribution generated for the analysis of Figure 2A, which accounted for very short and highly modified tails. (n = 2,778). (D) The distribution of terminal uridylation on short- and long-lived mRNAs at steady state. The tail-length distributions of Figure 6B are replotted and colored by uridylation frequency (key). (E) Examples of mRNA 3′-end isoforms plotted in Figure 6E. For the genomic locus corresponding to the dominant cleavage-and-polyadenylation site of *Actb*, several possible 3′-end isoforms lacking poly(A) tails are shown, along with distance from the dominant 3′ end and whether or not the isoform would be included in the plot of Figure 6E. Also depicted is a long-tail mRNA used to annotate the 3′ end of the UTR for the analyses of Figure 6E and Figures S6F–I. (F) Distribution of tags as a function of the distance between the inferred 3′ end of their UTR and the dominant 3′ end. Only tags with a poly(A)-tail longer than 30 nt were used in this analysis. (G) Fraction of tail-containing tags for each gene as a function of the distance between the inferred 3′ end of their UTR and the dominant 3′ end. As each dominant 3′ end was defined as the position represented by the most tags in a 21 nt window, in principle, no gene should have >50% of its tags at a position other than the dominant end. However, because dominant 3′ end annotations were determined using data from a separate experiment (standard PAL-seq with splinted ligation to the mRNA 3′ end), >50% of tags for some genes mapped to a position other than 0; these outliers represent discrepancies between biological replicates. (H) Nucleotide composition near cleavage-and-polyadenylation sites. For each tag mapping to within 10 nt of an annotated 3′ end of an mRNA 3′ UTR, the frequency of each genomic nucleotide is plotted as a function of the distance from the annotated 3′ end. The depletion of A at position 0 and its enrichment at position 1 were artifacts of 3′-end annotation because any A at the final nucleotide of a 3′ UTR was assigned to the poly(A) tail, even if that A might have been genomically encoded. (I) Relationship between the steady-state fraction of tails >175 nt and mRNA half-life. Otherwise as in Figure 6A. (J) Relationship between the steady-state abundance of long-tailed transcripts and mRNA half-life. For each tail length from 250–100, the fraction of mRNAs with tail lengths that fell above that position was calculated for each gene, and the relationship (R_s) between this fraction falling above the tail-length cutoff and half-life was determined as in (I) and plotted as in (C). (K) Mean tail lengths for mRNAs from each gene plotted in Figure 6B. Violin plots show distributions of short- and long-lived mRNAs, with the median of

each distribution shown as a horizontal line (and indicated above each group). Otherwise as in Figure 6B.

Supplemental Table, Table Titles, and Legends

Table S1. Parameters of the Computational Model, Related to Figure 4

Table of fitted parameters for the computational model. Staring tail lengths (m_p), production rate scaling terms (α), deadenylation rate constants (δ), and decapping rate scaling terms (β) were fit to the computational model. The global parameters v_p, m_d, and v_d were 16.27, 263.95, and 11.05 for cell line 1 and 15.5, 265.44, and 13.97 for cell line 2. mRNA half-lives were fit to an exponential model.

Table S2. Model Parameter Equations, Related to Figure 4

Parameter	Symbol	Distribution	Relevant equation	Parameter fit
Production rate	α	Negative binomial	$k_0(l) = \frac{\alpha\Gamma(v_p+l)}{l!\Gamma(v_p)}\left(\frac{m_p}{v_p+m_p}\right)^l\left(\frac{v_p}{v_p+m_p}\right)^{v_p}$	Gene specific
Starting tail length	m_p	″	″	Gene specific
Production rate spread	v_p	″	″	Global
Deadenylation	δ		$k_1 = \delta$	Gene specific
Decapping	β	Logistic	$k_2(l) = \frac{\beta}{(1+e^{-\frac{(l-m_d)}{v_d}})}$	Gene specific
Mean decapping tail length	m_d	″	″	Global
Decapping spread	v_d	″	″	Global

The equation defining production (k_0) has three fitted parameters: production rate (α) and starting tail length (m_p), which were both fit for each gene, and spread of the starting tail length distribution (v_p), which was fit globally. Deadenylation (k_1) was fit for each gene as a single parameter, deadenylation rate (δ). Decapping (k_2) uses the decapping rate (β) which was fit for each gene, as well as two globally fitted parameters, the decapping position (m_d) and the decapping rate spread (v_d).

Table S3. Oligonucleotides Used in This Study

Table S3: Oligonucleotides Used in This Study

RNA standards

Name	Sequence	Type	Source	Comment
30nt5EU	5′-7meG*-GACACAAGGACAGACUCACACGAGACAGAC-OH-3′	RNA	In vitro transcription	For 5-EU purification monitoring of RNA/RPF libraries. Asterisk denotes p32 label.
40nt5EU	5′-7meG*-GACGAACACAAGGACAGAG-5EU-GACACGAGACAGACGACAGC-OH-3′	RNA containing a single 5-EU	In vitro transcription	For 5-EU purification monitoring of RNA/RPF libraries. Asterisk denotes p32 label.
Rluc	5′-ppp-GGGCAAAUCAGGCAAAUCUGGUAAUGGUUCUUA UAGGUUACUUGAUCAUUACAAAUAUCUUACUGC AUGGUUUGAACUUCUUAAUUUACCAAAGAAGAU CAUUUUUGUCGGCCAUGAUUGGGGUGCUUGUUU GGCAUUUCAUUAUAGCUAUGAGCAUCAAGAUAA GAUCAAAGCAAUAGUUCACGCUGAAAGUGUAGU AGAUGUGAUUGAAUCAUGGGAUGAAUGGCCUGA UAUUGAAGAAGAUAUUGCGUUGAUCAAAUCUGA AGAAGGAGAAAAAAUGGUUUUGGAGAAUAACUU CUUCGUGGAAACCAUGUUGCCAUCAAAAAUCAU GAGAAAGUUAGAACCAGAAGAAUUUGCAGCAUA UCUUGAACCAUUCAAAGAGAAAGGUGAAGUUCG UCGUCCAACAUUAUCAUGGCCUCGUGAAAUCCC GUUAGUAAAAGGUGGUAAACCUGACGUUGUACA AAUUGUUAGGAAUUAUAAUGCUUAUCUACGUGC AAGUGAUGAUUUACCAAAAAUGUUUAUUGAAUC GGACCCAGGAUUCUUUUCCAAUGCUAUUGUUGA AGGUGCCAAGAAGUUUCCUAAUACUGAAUUUGU CAAAGUAAAAGGUCUUCAUUUUUCGCAAGAAGA UGCACCUGAUGAAAUGGGAAAAUAUAUCAAAUC GUUCGUUGAGCGAGUUCUCAAAAAUGAACAAUA AUUCUAGGAGCU-OH-3′	RNA	In vitro transcription	Common sequence of the standards, without the individual identifiers

Fluc	5′-7meG*- AUGGAAGACGCCAAAAACAUAAAGAAAGGCCCG GCGCCAUUCUAUCCGCUGGAAGAUGGAACCGCU GGAGAGCAACUGCAUAAGGCUAUGAAGAGAUAC GCCCUGGUUCCUGGAACAAUUGCUUUUACAGAU GCACAUAUCGAGGUGGACAUCACUUACGCUGAG UACUUCGAAAUGUCCGUUCGGUUGGCAGAAGCU AUGAAACGAUAUGGGCUGAAUACAAAUCACAGA AUCGUCGUAUGCAGUGAAAACUCUCUUCAAUUC UUUAUGCCGGUGUUGGGCGCGUUAUUUAUCGGA GUUGCAGUUGCGCCCGCGAACGACAUUUAUAAU GAACGUGAAUUGCUCAACAGUAUGGGCAUUUCG CAGCCUACCGUGGUGUUCGUUUCCAAAAAGGGG UUGCAAAAAAUUUUGAACGUGCAAAAAAAGCUC CCAAUCAUCCAAAAAAUUAUUAUCAUGGAUUCU AAAACGGAUUACCAGGGAUUUCAGUCGAUGUAC ACGUUCGUCACAUCUCAUCUACCUCCCGGUUUU AAUGAAUACGAUUUUGUGCCAGAGUCCUUCGAU AGGGACAAGACAAUUGCACUGAUCAUGAACUCC UCUGGAUCUACUGGUCUGCCUAAAGGUGUCGCU CUGCCUCAUAGAACUGCCUGCGUGAGAUUCUCG CAUGCCAGAGAUCCUAUUUUUGGCAAUCAAAUC AUUCCGGAUACUGCGAUUUUAAGUGUUGUUCCA UUCCAUCACGGUUUUGGAAUGUUUACUACACUC GGAUAUUUGAUAUGUGGAUUUCGAGUCGUCUUA AUGUAUAGAUUUGAAGAAGAGCUGUUUCUGAGG AGCCUUCAGGAUUACAAGAUUCAAAGUGCGCUG CUGGUGCCAACCCUAUUCUCCUUCUUCGCCAAA AGCACUCUGAUUGACAAAUACGAUUUAUCUAAU UUACACGAAAUUGCUUCUGGUGGCGCUCCCCUC UCUAAGGAAGUCGGGGAAGCGGUUGCCAAGAGG UUCCAUCUGCCAGGUAUCAGGCAAGGAUAUGGG CUCACUGAGACUACAUCAGCUAUUCUGAUUACA CCCGAGGGGGAUGAUAAACCGGGCGCGGUCGGU AAAGUUGUUCCAUUUUUUGAAGCGAAGGUUGUG GAUCUGGAUACCGGGAAAACGCUGGGCGUUAAU CAAAGAGGCGAACUGUGUGUGAGAGGUCCUAUG AUUAUGUCCGGUUAUGUAAACAAUCCGGAAGCG ACCAACGCCUUGAUUGACAAGGAUGGAUGGCUA CAUUCUGGAGACAUAGCUUACUGGGACGAAGAC GAACACUUCUUCAUCGUUGACCGCCUGAAGUCU CUGAUUAAGUACAAAGGCUAUCAGGUGGCUCCC GCUGAAUUGGAAUCCAUCUUGCUCCAACACCCC AACAUCUUCGACGCAGGUGUCGCAGGUCUUCCC GACGAUGACGCCGGUGAACUUCCCGCCGCCGUU GUUGUUUUGGAGCACGGAAAGACGAUGACGGAA AAAGAGAUCGUGGAUUACGUCGCCAGUCAAGUA ACAACCGCGAAAAAGUUGCGCGGAGGAGUUGUG UUUGUGGACGAAGUACCGAAAGGUCUUACCGGA AAACUCGACGCAAGAAAAAUCAGAGAGAUCCUC AUAAAGGCCAAGAAGGGCGGAAAGAUCGCCGUG AA-OH-3′	RNA	In vitro transcription	For 5-EU purification monitoring of RNA/RPF libraries. Asterisk denotes p32 label.

Chloram	5′-7meG-AUGGAGAAAAAAAUCACUGGAUAUACCACCGUUGAUAUAUCCCAAUGGCAUCGUAAAGAACAUUUUGAGGCAUUUCAGUCAGUUGCUCAAUGUACCUAUAACCAGACCGUUCAGCUGGAUAUUACGGCCUUUUUAAAGACCGUAAAGAAAAAUAAGCACAAGUUUUAUCCGGCCUUUAUUCACAUUCUUGCCCGCCUGAUGAAUGCUCAUCCGGAAUUCCGUAUGGCAAUGAAAGACGGUGAGCUGGUGAUAUGGGAUAGUGUUCACCCUUGUUACACCGUUUUCCAUGAGCAAACUGAAACGUUUUCAUCGCUCUGGAGUGAAUACCACGACGAUUUCCGGCAGUUUCUACACAUAUAUUCGCAAGAUGUGGCGUGUUACGGUGAAAACCUGGCCUAUUUCCCUAAAGGGUUUAUUGAGAAUAUGUUUUUCGUCUCAGCCAAUCCCUGGGUGAGUUUCACCAGUUUUGAUUUAAACGUGGCCAAUAUGGACAACUUCUUCGCCCCCGUUUUCACCAUGGGCAAAUAUUAUACGCAAGGCGACAAGGUGCUGAUGCCGCUGGCGAUUCAGGUUCAUCAUGCCGUCUGUGAUGGCUUCCAUGUCGGCAGAAUGCUUAAUGAAUUACAACAGUACUGCGAUGAGUGGCAGGGCGGGGCGUAA-OH-3′	RNA containing 5-EU	In vitro transcription	For quantitative spike normalization of actinomycin D libraries
GFP	5′-7meG*-AUGGUGAGCAAGGGCGCCGAGCUGUUCACCGGCAUCGUGCCCAUCCUGAUCGAGCUGAAUGGCGAUGUGAAUGGCCACAAGUUCAGCGUGAGCGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCUGUGCCCUGGCCCACCCUGGUGACCACCCUGAGCUACGGCGUGCAGUGCUUCUCACGCUACCCCGAUCACAUGAAGCAGCACGACUUCUUCAAGAGCGCCAUGCCUGAGGGCUACAUCCAGGAGCGCACCAUCUUCUUCGAGGAUGACGGCAACUACAAGUCGCGCGCCGAGGUGAAGUUCGAGGGCGAUACCCUGGUGAAUCGCAUCGAGCUGACCGGCACCGAUUUCAAGGAGGAUGGCAACAUCCUGGGCAAUAAGAUGGAGUACAACUACAACGCCCACAAUGUGUACAUCAUGACCGACAAGGCCAAGAAUGGCAUCAAGGUGAACUUCAAGAUCCGCCACAACAUCGAGGAUGGCAGCGUGCAGCUGGCCGACCACUACCAGCAGAAUACCCCCAUCGGCGAUGGCCCUGUGCUGCUGCCCGAUAACCACUACCUGUCCACCCAGAGCGCCCUGUCCAAGGACCCCAACGAGAAGCGCGAUCACAUGAUCUACUUCGGCUUCGUGACCGCCGCCGCCAUCACCCACGGCAUGGAUGAGCUGUACAAGUGA-OH-3′	RNA containing 5-EU	In vitro transcription	For 5-EU purification monitoring of RNA/RPF libraries. Asterisk denotes p32 label.
FragmentedRNAnorm1	5′-OH-GGCAUUAACGCGGCCGCUCUACAAUAGUGA-OH-3′	RNA	IDT	For quantitative-spike normalization of fragmented libraries
FragmentedRNAnorm2	5′-OH-GGCAUUAACGCGAACUCGGCCUACAAUAGU-OH-3′	RNA	IDT	For quantitative-spike normalization of fragmented libraries
FragmentedRNAnorm3	5′-OH-GAAAAAAAAAAAAAAUACAUUCCAAAAU-OH-3′	RNA	IDT	For quantitative-spike normalization of fragmented libraries

PAL-seq trace oligos				
Name	Sequence	Type	Source	Comment
32mer	5′-p*-GGCAUUAACGCGAACUCGGCCUACAAUAGUGA-OH-3′	RNA	IDT	For following the PAL-seq 3′-ligation. Asterisk denotes p32 label.
A29U	5′-p*-AAUGUCCGUUCAUUACUACUAAAAAAAAAAAAA AAAAAAAAAAAAAAAAU-OH-3′	RNA	IDT	For following the PAL-seq 3′-ligation. Asterisk denotes p32 label.
A65	5′-p*-UCCAUCAACUAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAA-OH-3′	RNA	IDT	For following the PAL-seq 3′-ligation. Asterisk denotes p32 label.
5.8S rRNA depletion oligos for ss-ligation PAL-seq				
Name	Sequence	Type	Source	Comment
5.8SRrnaSubHybMmu1	5′ ACGAGCCGAGTGATCCACCGC•Biotin 3′	DNA	IDT	
5.8SRrnaSubHybMmu2	5′ ATTCACATTAATTCTCGCAGC•Biotin 3′	DNA	IDT	
5.8SRrnaSubHybMmu3	5′ GCAAGTGCGTTCGAAGTGTCG•Biotin 3′	DNA	IDT	
PAL-seq and TAIL-seq oligos				
Name	Sequence	Type	Source	Comment
3pAdpater	5PhosAGATCGGAAGAGCACACGTCT-dT•Biotin-dT•Biotin-ddC-3'	DNA with a 5′ phosphate, two internal biotins, and a 3′ terminal dideoxycytidine	IDT	
U_Splint	TTCCGATCTATTTTTTT	DNA	IDT	
A_Splint	TTCCGATCTTTTTTTTT	DNA	IDT	
5pAdapter	AATGATACGGCGACCACCGAGATCTACACTCTT TCCCTACACGACGCTCTTCCGATCTrNrN*rNr NrNrNrNrN	DNA/RNA chimera	IDT	RNA bases are denoted r(N). * represents one of 4 phasing barcodes mixed in equal propotions (below)
RT_Primer	CAAGCAGAAGACGGCATACGAGATBBBBBBGTG ACTGGAGTTCAGACGTGTGCTCTTCCGA	DNA	IDT	BBBBBB represets the barcode, used only for sequencing cDNA with TAIL-seq
PAL_SeqPcrAmpFWD	CAAGCAGAAGACGGCATACGAGATBBBBBBCGT CTTTCCCTACACGACGCTCTTCC	DNA	IDT	BBBBBB represents the barcode, used for PAL-seq sequencing
PAL_SeqPcrAmpREV	AATGATACGGCGACCACCGAGATCTACACGTGA CTGGAGTTCAGACGTGTGCTCTTCCGA	DNA	IDT	
SeqPrimer1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC T	DNA	IDT	For PAL-seq v2 sequencing of reads 1 and 2
SeqPrimer2	AGATCGGAAGAGCGTCGTGTAGGGAAAGACG	DNA	IDT	For PAL-seq v2 sequencing of the index read
TailSeqPcrAmpFWD	AATGATACGGCGACCACCGAGATCTACACTCTT TCCCTACACGACGCTCTTCC	DNA	IDT	
TailSeqPcrAmpREV	CAAGCAGAAGACGGCATACGAGATBBBBBBGTG ACTGGAGTTCAGACGTGTGCTCTTCCGA	DNA	IDT	BBBBBB represents the barcode, used for TAIL-seq sequencing of PCR products.
3pAdapterSingleStrandedLig	5AmpNNNNAGATCGGAAGAGCACACGTCT-dT•Biotin-dT•Biotin-ddC-3'	DNA with a 5′ adenylyl group, two internal biotins, and a 3′ terminal dideoxycytidine	IDT	Note that that oligo was ordered with a 5' phosphate and then adenylylated.
TAIL-seq 5p Adapters				

Name	Sequence	Type	Source	Comment
5pAdapterPhase1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTrNrNr**Ar**NrNrNrNrNrN	DNA/RNA chimera	5pAdapter	If position 3 of the sequencing read is an A, sequencing of the molecule of interest begins at position 10; note that this is also the 5p adapter used for the single-stranded ligation experiments.
5pAdapterPhase2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNN**GA**NNNNNNN	DNA/RNA chimera	5pAdapter	If position 3-4 of the sequencing read is a GA, sequencing of the molecule of interest begins at position 11
5pAdapterPhase3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTrNrNr**CrGrAr**NrNrNrNrNrN	DNA/RNA chimera	5pAdapter	If position 3-5 of the sequencing read is an CGA, sequencing of the molecule of interest begins at position 12
5pAdapterPhase4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTrNrNr**UrCrGrAr**NrNrNrNrNrN	DNA/RNA chimera	5pAdapter	If position 3-6 of the sequencing read is an TCGA, sequencing of the molecule of interest begins at position 13

Barcode portion (3'-most 20nt, appended to Rluc sequence above) of PAL-seq and TAIL-seq RNA tail-length standards

Name	Sequence	Type	Source	Comment
A10_6	AAUGUCCGUUUCUCAAACUC-A10	RNA	Set 2	Not added to the same libraries as set 3. Poly(A)-tail length, 10 nt.
A30_6	AAUGUCCGUUCAUUACUACU-A29-A/U/C/G	RNA	Set 2	Not added to the same libraries as set 3. Poly(A)-tail length, 30 nt.
A110_6	AAUGUCCGUUACUCCAUCAU-A110	RNA	Set 2	Not added to the same libraries as set 3. Poly(A)-tail length, 110 nt.
A210_6	AAUGUCCGUUUACAAUCCCU-A210	RNA	Set 2	Not added to the same libraries as set 3. Poly(A)-tail length, 210 nt.
A10_7	AUCACAGCAUUCUCAAACUC-A10	RNA	Set 1 (AOS 2)	Subtelny et al., 2014. Poly(A)-tail length, 10 nt.
A50_7	AUCACAGCAUAUUCCAUCAC-A57	RNA	Set 1 (AOS 2)	Subtelny et al., 2014. Poly(A)-tail length, 57 nt.
A100_7	AUCACAGCAUAAAUUCCUCC-A107	RNA	Set 1 (AOS 2)	Subtelny et al., 2014. Poly(A)-tail length, 107 nt.
A150_7	AUCACAGCAUAACACUCUUC-A60	RNA	Set 1 (AOS 2)	Subtelny et al., 2014. Poly(A)-tail length, 60 nt.
A200_7	AUCACAGCAUCUAACUUACC-A215	RNA	Set 1 (AOS 2)	Subtelny et al., 2014. Poly(A)-tail length, 215 nt.
A250_7	AUCACAGCAUUACAUCCAUC-A273	RNA	Set 1 (AOS 2)	Subtelny et al., 2014. Poly(A)-tail length, 273 nt.
A300_7	AUCACAGCAUCUCACUAUAC-A324	RNA	Set 1 (AOS 2)	Subtelny et al., 2014. Poly(A)-tail length, 324 nt.
chrS_10	AAUGUCCGUUCCUACAACUU-A10	RNA	Set 3 (AOS 1)	Subtelny et al., 2014. Poly(A)-tail length, 10 nt.
chrS_50	AAUGUCCGUUUCCAUCAACU-A56	RNA	Set 3 (AOS 1)	Subtelny et al., 2014. Poly(A)-tail length, 56 nt.
chrS_100	AAUGUCCGUUACUCCAUCAU-A107	RNA	Set 3 (AOS 1)	Subtelny et al., 2014. Poly(A)-tail length, 107 nt.
chrS_150	AAUGUCCGUUAAAUCCCCUU-A167	RNA	Set 3 (AOS 1)	Subtelny et al., 2014. Poly(A)-tail length, 167 nt.
chrS_200	AAUGUCCGUUUACAAUCCCU-A205	RNA	Set 3 (AOS 1)	Subtelny et al., 2014. Poly(A)-tail length, 205 nt.
chrS_250	AAUGUCCGUUCCACUACUAU-A263	RNA	Set 3 (AOS 1)	Subtelny et al., 2014. Poly(A)-tail length, 263 nt.
chrS_300	AAUGUCCGUUAUACCUACCU-A302	RNA	Set 3 (AOS 1)	Subtelny et al., 2014. Poly(A)-tail length, 302 nt.

Chapter 3. MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs

Timothy J. Eisen[1,2,3]*, Stephen W. Eichhorn[1,2,3]*, Alexander O. Subtelny[1,2,3]*,
and David P. Bartel[1,2,3]

[1]Howard Hughes Medical Institute, Cambridge, MA, 02142, USA
[2]Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*These authors contributed equally to this work.

A.O.S., S.W.E., T.J.E., and D.P.B. conceived the project and designed the study. T.J.E., S.W.E., and A.O.S. performed the molecular experiments and analysis. T.J.E. performed the computational modeling. A.O.S., S.W.E., and T.J.E. drafted the manuscript, and T.J.E. and D.P.B. revised the manuscript with input from the other authors.

Published as: Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., and Bartel, D.P. (2020). MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. Mol Cell 77, 775-785 e778.

Summary

MicroRNAs (miRNAs) specify the recruitment of deadenylases to mRNA targets. Despite this recruitment, we find that miRNAs have almost no effect on steady-state poly(A)-tail lengths of their targets in mouse fibroblasts, which motivates the acquisition of pre-steady-state measurements of the effects of miRNAs on tail lengths, mRNA levels, and translational efficiencies. Effects on translational efficiency are minimal compared to effects on mRNA levels, even for newly transcribed target mRNAs. Effects on target mRNA levels accumulate as the mRNA population approaches steady state, whereas effects on tail lengths peak for recently transcribed target mRNAs and then subside. Computational modeling of this phenomenon reveals that miRNAs cause not only accelerated deadenylation of their targets but also accelerated decay of short-tailed target molecules. This unanticipated effect of miRNAs largely prevents short-tailed target mRNAs from accumulating despite accelerated target deadenylation. The net result is a nearly imperceptible change to the steady-state tail-length distribution of targeted mRNAs.

Introduction

The microRNA (miRNA) pathway targets mRNAs from most human genes (Friedman et al., 2009). Each miRNA associates with an Argonaute (AGO) protein to form a silencing complex. Within this complex, the miRNA pairs to regulatory sites within mRNAs (Bartel, 2018), whereas AGO interacts with TNRC6/GW182, which in turn recruits the PAN2–PAN3 and the CCR4–NOT deadenylase complexes, with the CCR4–NOT complex being more consequential (Chen et al., 2009; Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011; Christie et al., 2013; Huntzinger et al., 2013). This recruitment explains the observation that miRNAs accelerate deadenylation of reporter mRNAs (Giraldez et al., 2006; Wu et al., 2006; Chen et al., 2009) and is thought to underlie the widespread influence of miRNAs on post-transcriptional gene expression (Jonas and Izaurralde, 2015; Bartel, 2018).

The regulatory effects of miRNAs can differ, depending on the consequences of tail shortening in the posttranscriptional regulatory regime operating in the cell (Subtelny et al., 2014). For example, in early zebrafish embryos, where poly(A)-tail length influences translational efficiency but not mRNA stability, miRNAs cause reduced translational efficiency with little effect on mRNA stability (Bazzini et al., 2012; Subtelny et al., 2014). However, in gastrulating embryos, which are undergoing a developmental shift in the nature of translational control, such that poly(A)-tail length no longer influences translational efficiency but instead influences mRNA stability, miRNAs predominantly cause reduced mRNA stability (Bazzini et al., 2012; Subtelny et al., 2014). Likewise, in the post-embryonic systems that have been studied to date, mRNA destabilization is the principal mode of miRNA-mediated repression of endogenous messages, with a minor contribution from translational repression sometimes also detected (Baek et al., 2008; Hendrickson et al., 2009; Guo et al., 2010; Eichhorn et al., 2014).

This minor contribution from translational repression is not attributable to poly(A)-tail shortening, as poly(A)-tail length is not coupled with translational efficiency in these post-embryonic systems (Subtelny et al., 2014).

The lack of substantial effects of miRNAs on translational efficiency observed in post-embryonic systems at steady state does not rule out the idea that miRNAs might have a greater effect on translation soon after they encounter their targets. Investigations of this idea involving the dynamics of miRNA-mediated repression have yielded different results, depending upon the experimental approach. One approach has been to rapidly induce the expression of a reporter mRNA and measure its initial repression by a highly expressed constitutive miRNA (Bethune et al., 2012; Djuranovic et al., 2012), and the other approach has been to rapidly induce the expression of a miRNA and measure the repression of cellular mRNAs (Eichhorn et al., 2014). Translational repression seems to be much more prominent at early time points of the first approach compared to those of the second. One explanation for this discrepancy is that repression can be more readily detected at early time points if the miRNA is already at full strength at these time points. A second explanation is that reporter mRNAs might respond differently than endogenous cellular mRNAs. Analyses of the initial repression of newly transcribed endogenous mRNAs could distinguish between these possibilities.

Pre-steady-state analyses of miRNA effects might also provide mechanistic insight into miRNA-mediated mRNA destabilization. For instance, if the miRNA pathway acts independently of other decay pathways, then it would be expected to contribute less to the decay rate of rapidly decaying targets and, by extension, have greater influence on the expression level of long-lived mRNAs. Alternatively, if the miRNA pathway acts together with and reinforces other mRNA-decay pathways, then it would be expected to act more rapidly on mRNAs that are

already decaying more rapidly and contribute more equally to the destabilization of mRNAs regardless of their basal decay rate. Analyses of steady-state data seem to indicate that mRNAs with shorter half-lives are less susceptible to miRNA-mediated destabilization, which suggests at least some independence (Larsson et al., 2010). However, steady-state data cannot speak to the question of whether the miRNA pathway is acting more rapidly on some mRNAs than on others, and the results leave open the possibility of concerted action with other pathways.

We developed an experimental and computational framework to study the dynamics of deadenylation and decay of endogenous mRNAs of thousands of genes (Eisen et al., 2020). A conclusion of that study, performed in parallel with the current study, is that once deadenylation has proceeded to the point that mRNAs have short tails, these short-tailed mRNAs degrade at dramatically different rates. Moreover, mRNAs that have undergone more rapid deadenylation decay more rapidly upon reaching short tail lengths (Eisen et al., 2020). These results bring up the possibility that miRNA-mediated mRNA destabilization might have a second mechanistic component: in addition to the well-established role of promoting deadenylation, miRNAs might also cause more rapid decay of short-tailed target mRNAs. To investigate this possibility and to explore other questions that have awaited a pre-steady-state analysis of the global effects of miRNAs on endogenous mRNAs of the cell, we set out to investigate the effects of miRNAs on the dynamics of cytoplasmic mRNA metabolism, simultaneously examining their effects on tail lengths, translation, and decay of target mRNAs.

Results

Very Slight Effects of miRNAs on Bulk mRNA Poly(A)-Tail Lengths

Because miRNAs specify the recruitment of deadenylation machinery to their mRNA targets (Jonas and Izaurralde, 2015), we first examined the global effects of miRNAs on the steady-state tail lengths of mammalian mRNAs, making use of an improved version of poly(A)-tail length profiling by sequencing (PAL-seq) (Eisen et al., 2020). In this experiment, miR-1 was induced for 48 h in a clonal 3T3 cell line previously engineered to enable inducible miR-1 expression (Eichhorn et al., 2014), and tail lengths were profiled in these cells expressing miR-1 as well as cells grown in parallel without miR-1 induction. Despite robust reduction of both mRNA and ribosome-protected fragments (RPFs) of predicted miR-1 targets, as measured using RNA sequencing (RNA-seq) and ribosome-footprint profiling (Ingolia et al., 2009), respectively, the mean tail-length change for these targets was < 0.1 nucleotides (nt), and was not statistically significant (Figure 1A, left; $p = 0.83$). When focusing the analysis on a set of top-repressed targets, defined as those repressed by at least 25% in a prior experiment in which miR-1 was induced (Eichhorn et al., 2014), mean tail-length change was still < 0.1 nt (Figure 1A, right; $p = 0.90$). Essentially no change was also observed for predicted miR-155 targets after inducing miR-155 (Figure 1B; < 1 nt change for both predicted targets and top targets, $p = 0.35$ and 0.46, respectively).

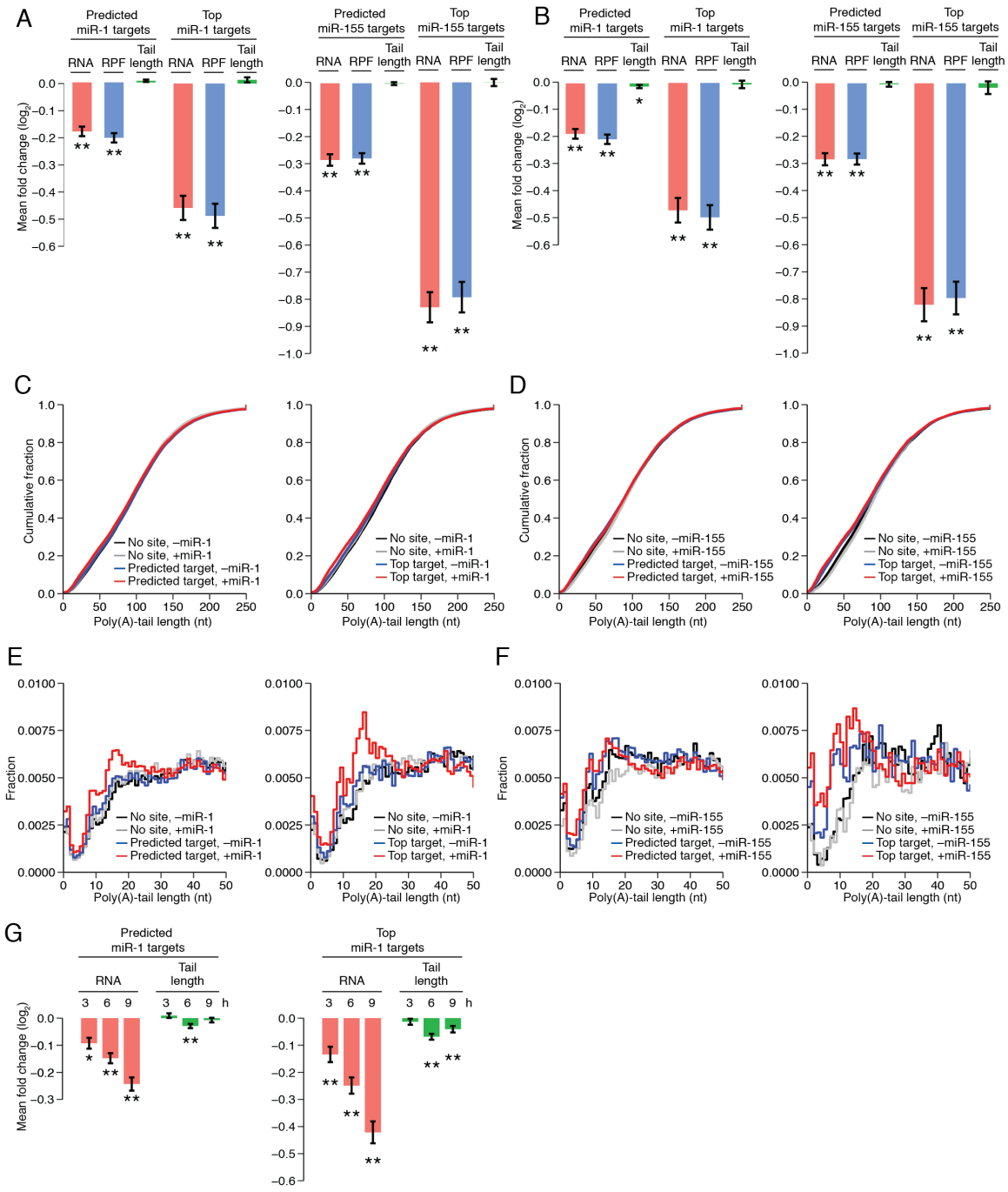


Figure 1. Very Small Effect of miRNAs on Steady-State Poly(A)-Tail Lengths

(A) Impact of miR-1 and miR-155 (left and right, respectively) on RNA abundance, RPF abundance, and mean poly(A)-tail length of their predicted targets (n = 657 and 466 for miR-1 and miR-155, respectively) and top predicted targets (n = 133 and 65, respectively). Values were normalized to those of mRNAs with no site to the induced miRNA (n = 2,170 and 3,501, respectively). Significant changes are indicated with asterisks below the bar (*$p \leq 0.05$; **$p \leq 0.001$, one-tailed t test). Error bars are the standard error of the mean (SEM), propagating the error for the SEM of the no-site mRNAs. (B) Impact of miR-1 and miR-155 on RNA abundance, RPF

abundance, and mean poly(A)-tail length of their predicted targets (n = 643 and 417 for miR-1 and miR-155, respectively) and top predicted targets (n = 130 and 57, respectively) when using a protocol that does not deplete very short or highly modified tails. For each gene, PAL-seq data were used to determine the length distribution of tails ≥ 50 nt, and data generated using a protocol without a splinted ligation were used to determine both the length distribution of tails < 50 nt and the fraction of tails < 50 nt. Values were normalized to those of mRNAs with no site to miR-1 or miR-155 (n = 2,131 and n = 3,138, respectively); otherwise, as in (A). (C) Very small influence of miR-1 on the steady-state tail-length distribution of its predicted targets. Left: cumulative distributions of poly(A)-tail lengths in miR-1–uninduced and miR-1–induced cells (–miR-1 and +miR-1, respectively) for predicted miR-1 targets (n = 643) and a set of no-site mRNAs with the same distribution of 3′ UTR lengths (n = 626). To ensure that the tail-length distributions were not unduly influenced by a few highly expressed mRNAs, values for mRNAs with more than the median number of poly(A) tags were down-weighted such that they contributed to the distribution as much as values from the mRNA with the median value. Right: as on the left, but plotting the results for top predicted targets of miR-1 (n = 130) and a set of no-site mRNAs matched for UTR length (n = 130). Results are from datasets used in (B), which avoided depletion of mRNAs with very short or highly modified tails. (D) Very small influence of miR-155 on the steady-state tail-length distributions of its predicted targets. Analysis is as in (C) but showing the effect of inducing miR-155 on tail lengths of its predicted targets (n = 417 and 57 for predicted and top targets, respectively) and matched controls (n = 417 and 57, respectively). (E) Influence of miR-1 on the steady-state tail-length distribution of short-tailed mRNAs. Analysis as in (C), except showing results for predicted targets (left) and top predicted targets (right) with tails ≤ 50 nt as a histogram. (F) Influence of miR-155 on the steady-state tail-length distributions of short-tailed mRNAs. Otherwise as in (E). (G) Reanalysis of data monitoring mRNA levels and tail lengths after transfecting miR-1 into HeLa cells (Chang et al., 2014). Changes in mRNA abundance of predicted miR-1 targets (n = 354) and top predicted miR-1 targets (n = 140) normalized to no-site controls (n = 908). For each time point, changes were normalized by the corresponding median $\log_2$ fold change of the 0-h time point. Otherwise as in (A).

These analyses used datasets generated from a splint ligation to the poly(A) tail, which was expected to be less efficient for mRNAs with tail lengths <8 nt, and although the protocol was able to detect mRNAs with tails modified with a single terminal U, albeit at reduced efficiency, it was not designed to detect mRNAs with tails with an oligo(U) or other modification (Eisen et al., 2020). Reasoning that this method might obscure miRNA-induced tail-length changes for mRNAs with very short or extensively modified tails, we performed the same experiments, except the splint ligation was replaced with single-stranded ligation. With these

additional data, mean tail length changes were all <1 nt, and statistically significant for only one of the four comparisons (Figure 1B, $p = 0.011$ for predicted targets of miR-1).

Having found that inducing expression of a miRNA caused essentially no detectable changes to the mean tail lengths of its predicted targets, we compared single-molecule tail-length distributions to query whether miRNAs altered these distributions in ways that did not affect mean values. Such differences would be expected if miRNAs preferentially direct the deadenylation of shorter tails, as might be inferred from reports that (1) miRNA function depends more on recruitment of the CCR4–NOT deadenylase complex than it does on recruitment of the PAN2–PAN3 deadenylase complex (Behm-Ansmant et al., 2006; Tritschler et al., 2010), and (2) the CCR4–NOT deadenylase complex acts at a later, more processive step of mRNA deadenylation than does the PAN2–PAN3 complex (Yamashita et al., 2005). Contrary to this expectation, induction of either miR-1 or miR-155 had almost no influence on the distribution of single-molecule tail-length measurements for all predicted targets or top-repressed targets of each miRNA (Figures 1C and 1D).

Analysis of tail-length distributions of mRNAs from different genes showed that for 116 of 1,060 predicted targets, induction of the miRNA shortened the tail-length distribution (Kolmogorov–Smirnov [K–S] test, $p < 0.01$, Figure S1). However, this group of 116 was not much larger than the 81 predicted targets for which induction of the miRNA lengthened the tail-length distribution. Moreover, for 83 of 1,041 no-site controls, induction of the miRNA also significantly shortened the tail-length distribution. Taken together, these analyses hinted at a signal, albeit of very low magnitude, for miRNA-dependent shortening of tail lengths. Indeed, closer inspection of the distributions in Figures 1C and 1D revealed a slight increase in short-tailed isoforms of predicted targets in the presence of the miRNA, which was most pronounced

at tail lengths of 7–20 nt (Figures 1E and 1F)—a range resembling that of the build-up generally observed for short-tailed isoforms of short-lived mRNAs (Eisen et al., 2020). However, for each comparison, these increases involved only a small fraction (0.001–0.021) of the target mRNAs.

Time-Resolved Measurements of miRNA-Mediated Changes

How then, might this low-magnitude effect observed for endogenous mRNAs be reconciled with both the established role for the poly(A) tail in stabilizing mRNAs (Goldstrohm and Wickens, 2008) and the established role for miRNAs in directing poly(A)-tail shortening of both reporter mRNAs in mammalian cells and endogenous mRNAs in early zebrafish embryos (Giraldez et al., 2006; Wu et al., 2006; Subtelny et al., 2014)? A clue to this riddle comes with the report of tail-length shortening measured 3, 6, and 9 h after miRNA transfection into HeLa cells (Chang et al., 2014). Further analyses of these tail-length changes revealed statistically significant, albeit small, tail-length shortening of predicted targets in this system in which the transcriptome is still adjusting to the recent transfection of a miRNA and has not yet reached a new steady state (Figure 1G, mean changes of 1.9 and 4.0 nt for all predicted targets and top predicted targets, respectively, 6 h after transfection). Perhaps measurements obtained at steady-state are not suitable for detecting miRNA effects on tails, and pre-steady-state measurements must be examined instead.

To understand how the miRNA pathway leverages the relationship between tail lengths and mRNA turnover, we set out to build an experimental and analytical framework for the global study of pre-steady-state miRNA-mediated changes in tail lengths and mRNA levels. To address lingering questions related to the dynamics of miRNA-mediated translational repression and the potential relationship between these dynamics and those of mRNA decay, we also designed the

experiments to detect early miRNA-mediated changes in translational efficiency. Our framework built upon the one that we concurrently developed to study the dynamics of cytoplasmic mRNA metabolism more generally (Eisen et al., 2020). Indeed, the cells used in that study had been engineered to have the ability to inducibly express either miR-1 or miR-155, which allowed data generated in the uninduced condition to be used in both that study and in the current study. In the other study, data generated in the uninduced condition were used to examine the dynamics of mRNA deadenylation and decay, whereas in the current study, those data provided the baseline for comparison to analogous datasets concurrently generated using the same cell lines grown for 48 h in the presence of doxycycline, which induced either miR-1 or miR-155, depending on the cell line. In these conditions, mRNAs emerged into a cytoplasm in which the induced miRNA was expressed at a level comparable to those of the highest-expressed endogenous miRNAs (Eichhorn et al., 2014).

Study of pre-steady-state mRNA dynamics and the influence of miRNAs on these dynamics required datasets reporting mRNA abundance, poly(A)-tail length, and translation for mRNAs of different ages. To generate these datasets, we performed continuous metabolic-labeling experiments in which the uridine analog 5-ethynyl uridine (5EU) was added to cells, and after different time intervals, cytoplasmic RNA was harvested for analysis of mRNA abundances, translational efficiencies, and tail lengths (Figure 2A). These analyses were each designed to report on only RNA that had been produced after 5EU addition, which was separated from pre-existing RNA by virtue of the alkyne moiety of 5EU; this moiety can be specifically and efficiently biotinylated using click chemistry, thereby enabling isolation of metabolically labeled RNA (Jao and Salic, 2008; Kwasnieski et al., 2019; Eisen et al., 2020). When accounting for the time required for 5EU incorporation, mRNA processing, and mRNA export, most

mRNAs isolated from the 40-min time interval had been in the cytoplasm for 0–10 min, whereas mRNAs isolated from longer time intervals had a correspondingly broader range of ages (Eisen et al., 2020). For example, most mRNAs isolated from the 1-h time interval had been in the cytoplasm for 0–30 min. For time-resolved tail-length profiling, 5EU-containing RNA was isolated prior to PAL-seq, whereas for time-resolved ribosome-footprint profiling, 5EU-containing RPFs were isolated and sequenced (Figure 2B). Likewise, to generate RNA-seq data comparable to the RPF data, we fragmented mRNA to the size of RPFs, and isolated and sequenced 5EU-containing mRNA fragments (Figure 2B). Generation of these RPF and RNA-seq datasets required purification of fragments containing a single 5EU, a task greatly facilitated by the high efficiency of 5EU biotinylation (Figure S2).

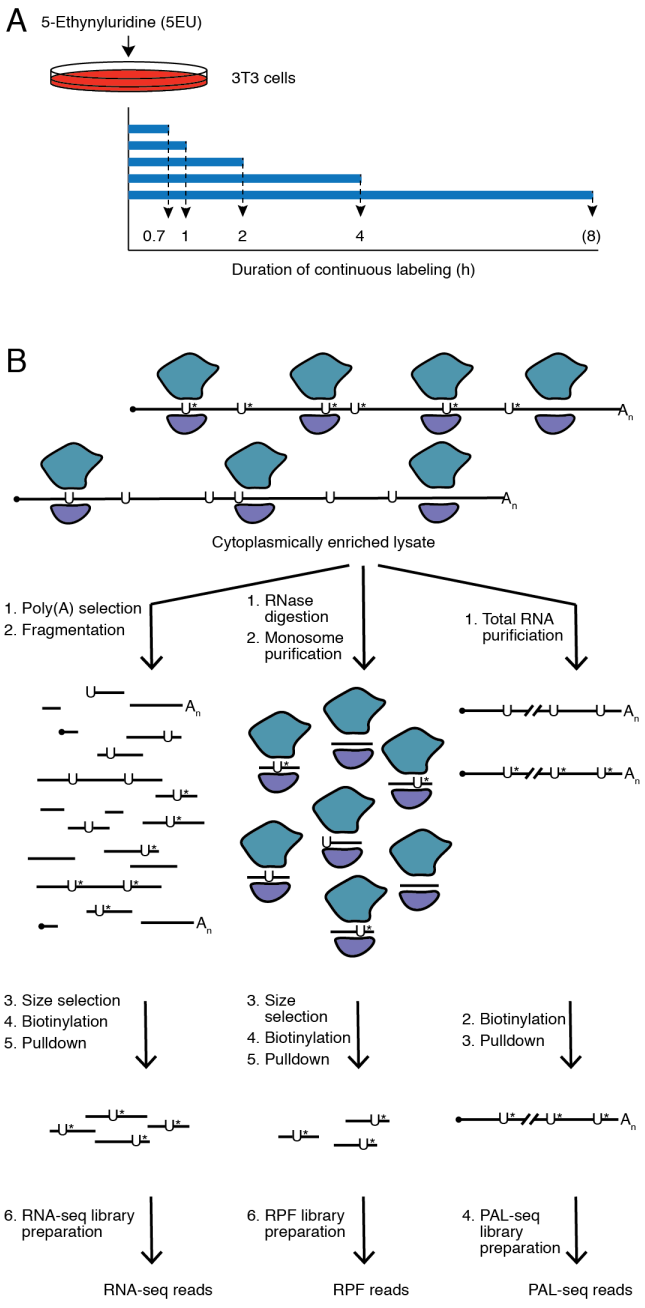


Figure 2. Profiling Dynamics of miRNA Effects

(A) Schematic of 5EU metabolic labeling. The time point in parentheses indicates that experiments with miR-155-inducible cells but not miR-1-inducible cells included an 8-h labeling period. (B) Schematic of the global assays applied to mRNAs labeled with 5EU (U*). All samples were harvested as cytoplasmically enriched lysates, and for each cell line and condition, all three libraries for each time interval were prepared from the same starting sample.

Effect of miRNAs on Tail-Lengths but not on Translational Efficiency of Young mRNAs

In cells in which miR-1 was induced, we observed significant decreases in both mRNA and RPFs for predicted miR-1 targets, and the magnitude of these decreases accrued with longer labeling periods (Figure 3A). When considering only the top predicted targets, repression was both more pronounced at each time period and reached statistical significance at earlier time periods (Figure 3A, bottom). Analogous results were observed upon miR-155 induction (Figure 3B). Translational repression, measured as changes in RPFs beyond those observed for mRNA abundance, was difficult to detect in either the miR-1 or the miR-155 experiment. Thus, miRNA-mediated translational repression of cellular mRNAs either does not occur in these cells or is too subtle to be reliably detected, even soon after the onset of miRNA-mediated regulation. The observation of robust mRNA destabilization with no sign of translational repression implied that early translational repression is not required to achieve destabilization of endogenous mRNAs.

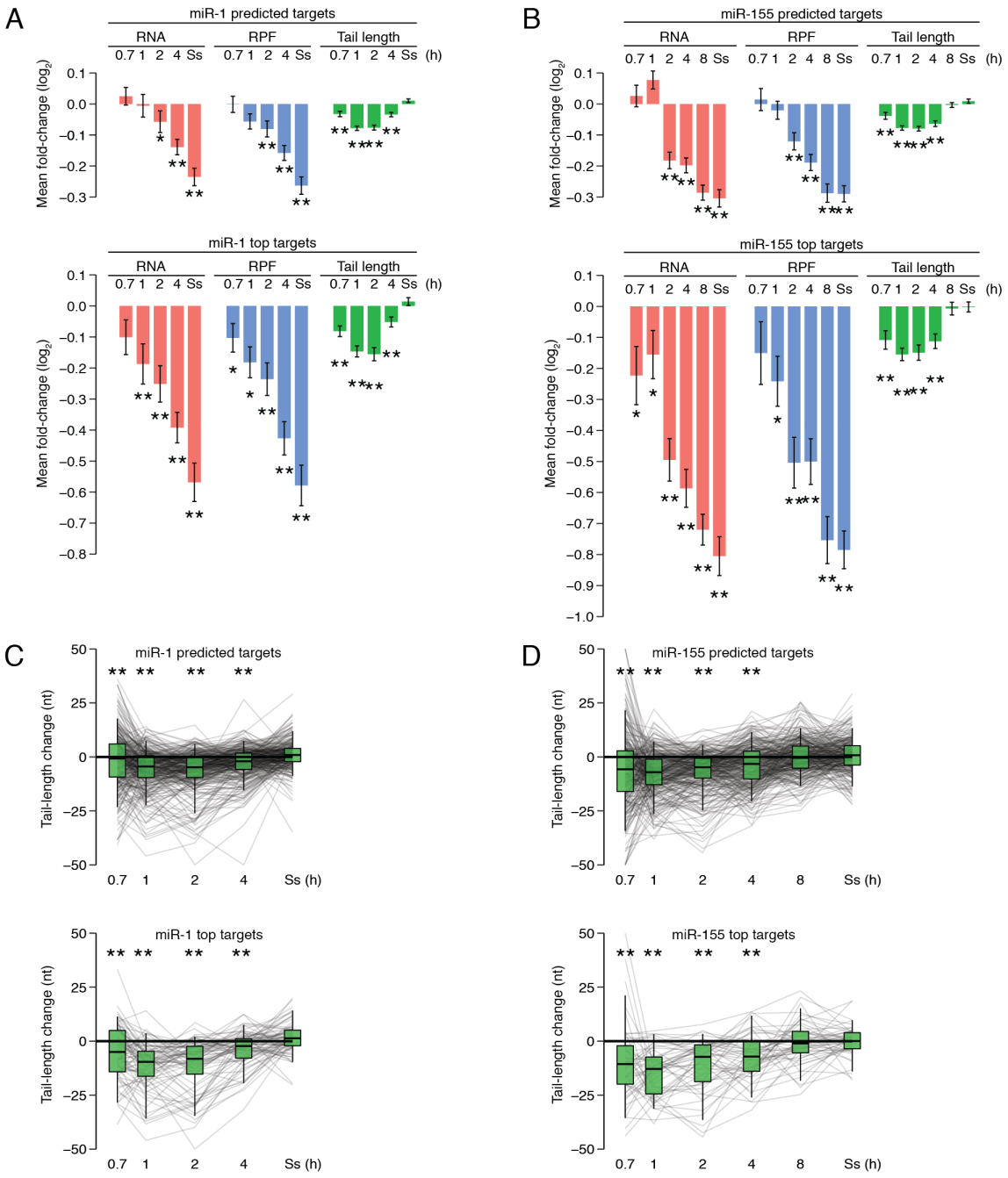


Figure 3. **Influence of miRNAs on the Dynamics of mRNA Metabolism**

(A) The dynamics of miR-1-mediated repression. Plotted are mean fold changes in mRNA abundance, RPF abundance, and mean poly(A)-tail length of mRNAs captured after labeling with 5EU for the indicated time intervals, comparing the values observed with and without miR-1 for predicted miR-1 targets (top) and top miR-1 targets (bottom), normalized to those observed for no-site mRNAs (n = 290, 70, and 684, respectively). Results are shown for mRNAs that passed the cutoffs for RNA, RPF, and tail-length measurements at all time intervals. Normalization, error bars, and significance testing were as in Figure 1A. Ss, steady state. (B) The dynamics of miR-

155-mediated repression. Analysis was of values for predicted miR-155 targets (top) and top miR-155 targets (bottom) normalized to those observed for no-site mRNAs (n = 251, 41, and 1,470, respectively); otherwise, as in (A). (C) Tail-length changes for predicted and top predicted targets of miR-1. Lines connect results for mRNAs from the same gene, and boxplots show the spread of the distribution (line, median; box, 25^{th}–75^{th} percentiles; whiskers, 5^{th}–95^{th} percentiles). Otherwise as in (A). (D) Tail-length changes for predicted and top predicted targets of miR-155. Otherwise as in (C).

In contrast to miRNA-mediated translational repression, miRNA-mediated tail shortening was readily detected in the pre-steady-state time periods, and moreover, it exhibited very different dynamics from those observed for mRNA and RPF changes. For the predicted targets of miR-1, statistically significant tail shortening was detectable after only 40 min of labeling, peaked between 1 and 2 h, and then abated such that no significant miRNA-mediated tail shortening was observed at steady state (Figure 3A, top). The same behavior was observed for the predicted targets of miR-155 with inclusion of the 8-h labeling interval, indicating that tail-length changes abated by this time (Figure 3B, top). Median tail-length changes observed after 1 h of labeling for predicted targets of miR-1 and miR-155 were 4.4 and 7.1 nt, respectively (Figures 3C and 3D). When considering only the top predicted targets for each miRNA, essentially the same dynamics were observed (Figures 3A and 3B, bottom) but with larger median changes of 9.6 and 12.8 nt observed after 1 h of labeling for top targets of miR-1 and miR-155, respectively (Figures 3C and 3D). In both experiments, statistically significant tail-length changes either preceded or occurred concurrently with statistically significant RNA and RPF changes (Figure 3).

An Analytical Framework for miRNA-mediated mRNA Dynamics

To explain the pattern of tail-length changes observed for miRNA targets, we used a computational model that describes the dynamics of deadenylation and decay for thousands of

mRNAs (Eisen et al., 2020). This model, which had been fit to the continuous-labeling data for tail lengths and mRNA abundances in the absence of miRNA induction, was used to simulate the effects of miRNAs over time by using three alternative approaches that each tested a different explanation for the fold change in steady-state abundance for the predicted targets (Figures 4A and 4B). One explanation tested was the prevailing notion that miRNAs act to destabilize mRNAs by causing more rapid deadenylation of their targets. To test this possibility, we increased the deadenylation rate constants to achieve the fold changes in steady-state abundance observed with miRNA induction and then used the model with and without these increased values to simulate the miRNA-mediated mRNA and tail-length changes observed in the continuous-labeling experiment, including a 35-min offset to account for nuclear processing and export. Although attributing the entire miRNA effect to increased rates of deadenylation resulted in simulated mRNA-abundance changes that roughly matched the observed changes, the simulated mean tail-length changes diverged from the observed measurements at later time intervals (Figures 4A and 4B, green). Attributing the entire miRNA effect to only increased rates of short-tailed mRNA decay resulted in even greater divergence between the simulated and observed tail-length changes, as more rapid decay of short-tailed mRNAs led to increased rather than decreased mean tail lengths (Figures 4A and 4B, red). By contrast, when miRNAs were allowed to increase the rates for both deadenylation and short-tailed mRNA decay, the best concordance was achieved; as for the observed mean tail-length changes, simulated changes peaked at short time periods, and then as short-tailed targets decayed more rapidly and a new

steady state was established, the effects on mean tail length disappeared (Figures 4A and 4B, blue).

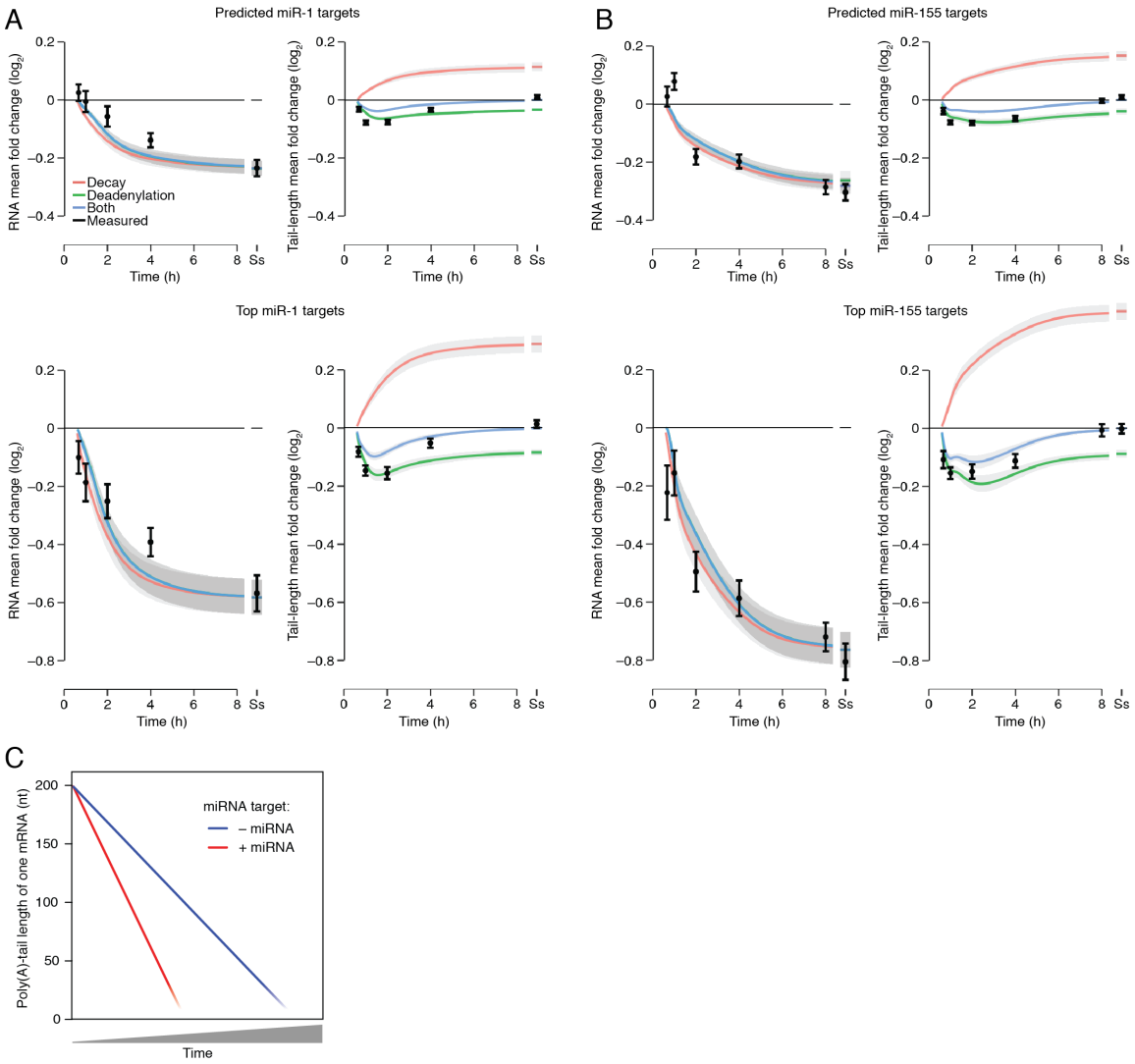


Figure 4. miRNAs Affect Deadenylation and Short-Tailed mRNA Decay

(A) Simulations of effects of miR-1 on its predicted targets (top panels) and top predicted targets (bottom panels). Effects of miRNA were simulated starting with fitted parameters describing initial tail lengths, production rates, deadenylation rates, and deadenylation-dependent decay rates of predicted targets over time without miRNA induction (Eisen et al., 2020). Effects of the miRNA were then modeled as changes in either the decay rate constant, the deadenylation rate constant, or both rate constants. These changes in rate constants were fit for mRNAs of each gene, minimizing the squared difference between the simulated and measured fold change in steady-state mRNA abundance. Using these fitted values, the mean fold changes in mRNA abundance (left) and mean tail lengths (right) were simulated and plotted over time with a 35-min offset to account for nuclear processing and export. For each simulation, the rate constant or combination of rate constants that was changed is indicated (key), with shading showing the SEM. Numbers of

predicted targets in each set are as in Figure 3A. Ss, steady state. (B) Simulations of effects of miR-155 on its predicted targets (top panels) and top predicted targets (bottom panels). Otherwise as in (A), with numbers of targets as in Figure 3B. (C) Schematic illustrating why miRNAs minimally affect steady-state tail lengths. In the absence of the miRNA, the mRNA exits the nucleus and undergoes deadenylation and then ultimately decay, once its tail becomes short (blue). In the presence of the miRNA, the mRNA target undergoes faster deadenylation, and once its tail is short, faster decay (red). With this more rapid decay of short-tailed isoforms, the targeted mRNA transits through the same distribution of tail lengths as it would have if it were not targeted—but just more rapidly. As a result, the weighted average of the populated tail-length states at steady state (the mean tail length) is unaffected, and the effect of the miRNA on tail length is revealed only when observing pre-steady–state kinetics.

These results indicate that miRNAs cause not only an increase in deadenylation rates of their targets but also an increase in rates at which short-tailed targets are degraded. The increased rate of short-tailed target degradation prevents a steady-state buildup of short-tailed isoforms; without this additional effect of miRNAs, the miRNA-mediated increase in deadenylation rate would lead to a corresponding increase in steady-state short-tailed isoforms, thereby decreasing steady-state mean tail length. With this additional effect of miRNAs, targeted mRNAs can transit through essentially the same distribution of tail lengths, spending the same proportion of time at each tail length in the presence of the miRNA as they would have in its absence (Figure 4C). Thus, this additional effect of miRNAs explains how miRNAs can cause accelerated target deadenylation with minimal influence on tail-length distributions (Figure 1) and why the effect of the miRNA on tail length is revealed only when examining the dynamics of mRNA metabolism.

Influence of mRNA Half-Life on the Dynamics of miRNA-Mediated Repression

One factor proposed to influence the impact of miRNAs on the expression of their targets is the basal decay rate of the target (Larsson et al., 2010). Indeed, if the induced miRNAs acted independently of the basal decay factors (which include other miRNAs), then the decay rate of a

miRNA target would be the sum of the basal decay rate of the target and the decay rate imposed by the induced miRNA, in which case, the targets with faster basal decay rates would experience smaller fold changes upon miRNA induction than targets with slower basal decay rates.

To examine this possibility, we looked at the relationship between steady-state repression of predicted targets and their basal half-lives. Although short-lived predicted targets of miR-1 appeared somewhat less repressed at steady state, the opposite trend was observed for predicted targets of miR-155 (Figure 5A, R_s = –0.13 and 0.15, respectively, with p = 0.007 and 0.008). This comparable repression of targets with short and long half-lives could not be explained by more effective target recognition of short-lived mRNAs, offsetting less effective repression of these mRNAs (Figure S3). These results did not support the model in which miRNAs independently add to the basal decay rates of their targets. Instead, they supported a model in which the miRNA silencing complex acts together with the basal decay factors, such that mRNAs with faster basal decay are no less influenced by miRNA targeting.

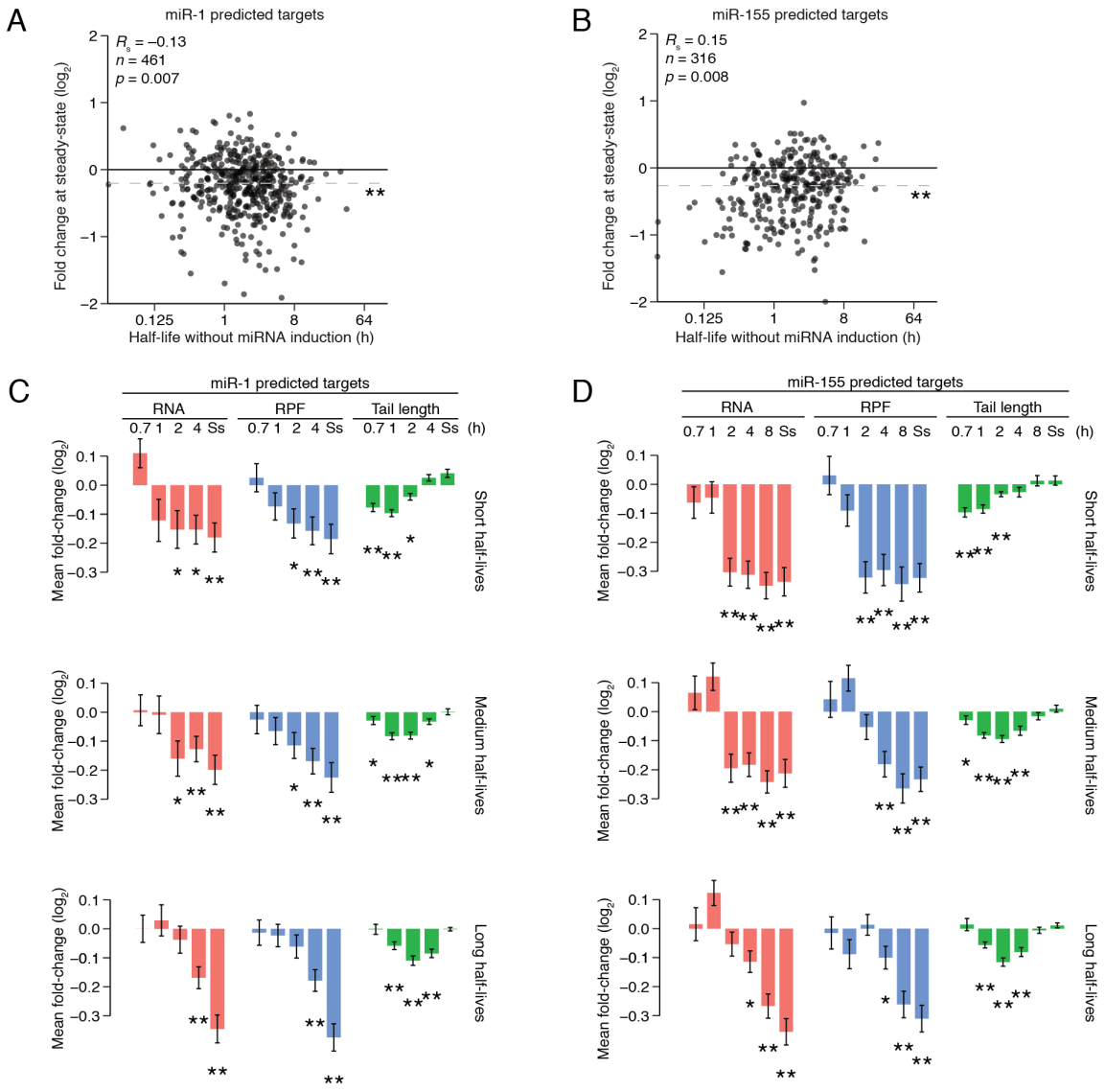


Figure 5. mRNA Half-life as a Determinant of miRNA-Mediated Repression

(A) The influence of basal half-lives on miR-1–mediated steady-state repression. miR-1–mediated changes in RNA abundance at steady state were compared to mRNA half-life in the uninduced condition. The median fold change of no-site mRNAs (n = 1,274) was subtracted from the fold changes of predicted targets (n = 461). Horizontal dashed gray lines denote the median values of each group, with asterisks denoting significance testing as in Figure 1A. The p value is the rank-order association test indicating the probability of a slope ≠ 0. (B) The influence of basal half-lives on miR-155–mediated steady-state repression. miR-155–mediated changes in RNA abundance at steady state were compared to mRNA half-life in the uninduced condition, with median fold change of no-site mRNAs (n = 1,862) subtracted from the fold change of the predicted targets (n = 316). Otherwise as in (A). (C) Dynamics of miR-1–mediated repression of predicted miR-1 targets split into three equally sized groups based on half-life. The groups have half-lives ranging from 0.01–0.985 h, 0.985–2.365 h, and 2.365–20 h, respectively, with 92 target genes in each group, based on half-life measurements from uninduced miR-155–inducible cells. For each group,

fold changes were normalized to those of a set of no-site mRNAs with the same range of half-lives (n = 183, 168, and 289, respectively). Otherwise, as in (A). (D) Dynamics of miR-155–mediated repression of predicted miR-155 targets split into three equally sized groups based on half-life. These groups have half-lives ranging from 0.01–1.03 h, 1.03–2.8 h, or 2.8–20 h, respectively, with 83 target genes in each group, and were normalized to a set of no-site mRNAs with the same range of half-lives (n = 368, 462, and 614, respectively). Otherwise as in (A). See also Figure S4.

Our conclusion that the induced miRNAs acted together with basal decay factors to further reduce the stability of mRNAs that were already short lived predicted that miRNA effects would be observed earlier for short-lived mRNAs than for long-lived mRNAs. Analysis of our pre-steady-state results confirmed this prediction. Steady-state RNA and RPF repression levels were reached substantially earlier for predicted targets of miR-1 with shorter basal half-lives, with maximal miRNA-mediated tail shortening also occurring earlier for these predicted targets (Figure 5C). Although the dynamics of miRNA-mediated deadenylation and decay came into sharper focus when grouping predicted targets with similar basal half-lives, we still did not observe compelling evidence for miRNA-mediated translational repression, even at early time intervals. Analogous results were observed for the destabilization and tail-length dynamics of the top predicted targets (Figure S4A) and for targets and top predicted targets of miR-155 after inducing this miRNA (Figures 5D and S4B). Collectively, these analyses demonstrated that the basal half-life of the mRNA target strongly influences the dynamics of miRNA-mediated repression, again implying connections between the basal and miRNA-recruited decay factors.

Discussion

Our experiments provided the opportunity to observe the early effects of miRNAs on translation and also to learn how miRNAs influence the dynamics of deadenylation and decay in mammalian cells. Although in postembryonic cells miRNAs mostly act to enhance the decay of

their mRNA targets (Guo et al., 2010; Eichhorn et al., 2014), translational repression can comprise a relatively greater fraction of the repression observed soon after a miRNA encounters an mRNA target. However, the magnitude of this phenomenon depends on the experimental approach. Substantial effects on translation are observed soon after inducing a reporter with sites to an expressed miRNA (Bethune et al., 2012; Djuranovic et al., 2012), whereas more modest effects on endogenous mRNAs are observed soon after inducing a miRNA (Eichhorn et al., 2014).

The results of our metabolic-labeling approach provided insight into the different effect sizes observed for the two previous experimental approaches. Like the miRNA-induction experiments, our metabolic-labeling experiments examined effects in 3T3 cells (indeed, in the same clonal lines as used in the previous miRNA-induction experiments) and monitored endogenous mRNAs rather than reporters. Like the reporter experiments, our metabolic-labeling experiments examined the early response to a preexisting, highly expressed miRNA, which presumably provides the best opportunity to observe the early effects of the miRNAs, which were already at full strength when the early measurements were taken. Despite this presumed greater sensitivity, our metabolic-labeling experiments were no better than the miRNA-induction experiments in detecting translational repression—indeed, they appeared to be somewhat worse. They did not detect convincing translational repression at any time interval (Figures 3 and 5), whereas the miRNA-induction experiments performed in the same cell lines seem to detect some translational repression after miR-155 induction (but not after miR-1 induction) (Eichhorn et al., 2014). This inability of our metabolic-labeling approach to detect translational repression at early time intervals indicated that the notion that translational repression constitutes a notable

component of the effect soon after a miRNA encounters a target mRNA does not apply to miRNA-mediated repression of endogenous mRNAs—at least not in NIH 3T3 cells.

In contrast to translational repression, miRNA-mediated enhanced deadenylation was observed at early time intervals. Indeed, it was reliably detected prior to or coincident with accelerated mRNA degradation—results that integrated deadenylation and previously observed effects of miRNA-mediated repression. The influence on poly(A)-tails peaked at intermediate time points and then disappeared at steady state. Explaining this disappearance was our discovery that miRNAs not only accelerate deadenylation of their regulatory targets but also accelerate the decay of these target mRNAs once their tails become shortened.

We speculate that accelerated decay of short-tailed mRNAs occurs through accelerated decapping. This accelerated decapping need not be through the direct contacts between TNRC6 (or AGO) and the decapping machinery or its coactivators. Instead, the changes that occur as the miRNA silencing complex helps remodel target mRNA-protein complexes to accelerate deadenylation might also recruit the decapping machinery or its coactivators. Indeed, physical connections between the CCR4–NOT deadenylase complex and the decapping complex (Haas et al., 2010; Ozgur et al., 2010; Jonas and Izaurralde, 2015) as well as the intracellular colocalization of these complexes (Parker and Sheth, 2007) could help link deadenylation with decapping.

Our concurrent analysis of the dynamics of cytoplasmic mRNA metabolism shows that mRNAs that undergo more rapid deadenylation generally undergo more rapid decay once their tails have become short (Eisen et al., 2020). This more rapid decay of short-tailed mRNAs that had previously been more rapidly deadenylated prevents a large buildup of short-tailed isoforms of rapidly deadenylated mRNAs, thereby enabling the broad range in mRNA deadenylation rates

to impart a similarly broad range in mRNA decay rates (Eisen et al., 2020). These results observed when examining mRNA metabolism more generally—not just when looking at the effects of miRNAs—suggest that processes of deadenylation and short-tailed–mRNA decay are coupled for not only miRNA targets but also targets of other degradative pathways. Moreover, our finding that miRNAs work in conjunction with and reinforce these other degradative pathways reveals another dimension of interconnectivity that further supports this emerging view of broad integration of post-transcriptional regulatory processes. In this view, targeting by an added miRNA would accelerate the coherently coupled processes of deadenylation and short-tailed–mRNA decay already occurring for each of the cellular mRNAs.

Acknowledgements

We thank J. Kwasnieski and other members of the Bartel lab for helpful discussions and the Whitehead Genome Technology Core for high-throughput sequencing. This research was supported by NIH grants GM061835 and GM118135 (D.P.B.) and an NSF Graduate Research Fellowship (T.J.E.). A.O.S. was supported by the NIH Medical Scientist Training Program fellowship T32GM007753. D.P.B. is an investigator of the Howard Hughes Medical Institute.

Declaration of Interests

The authors declare no competing interests.

Methods

Experimental Model and Subject Details

Cell Lines and Cell Culture

Clonal 3T3 cell lines engineered to express miR-155 (cell line 1) or miR-1 (cell line 2) upon doxycycline treatment were previously described (Eichhorn et al., 2014). Cells were grown at 37°C in 5% CO_2 in DMEM supplemented with 10% BCS (Sigma-Aldrich) and 2 μg/mL puromycin. 3T3 cells are male. Mycoplasma testing was performed and no contamination was observed.

Methods Details

Metabolic-Labeling Time Courses

Cells from each line were plated onto 500 cm^2 plates at 6.6 million cells per plate and cultured for two days such that they reached ~70%–80% confluency. Cells from each line were then re-plated in parallel for the uninduced and induced conditions, again at 6.6 million cells per 500 cm^2 plate, with the uninduced cells cultured in DMEM with 10% BCS, and the induced cells cultured in DMEM with 10% BCS and 2 μg/mL doxycycline. After 48 h, growth media was supplemented with 5-ethynyl uridine (5EU, Jena Biosciences) (Jao and Salic, 2008) at a final concentration of 400 μM, and cells were harvested after the desired labeling intervals (Figure 2A). Four plates were harvested for each 40 min time interval, three plates for each 1 h time interval, and two plates for each other time interval. A plate that had never received 5EU was harvested in parallel for each condition.

Cells were harvested at 4°C, washed twice with 50 mL ice-cold PBS, pH 7.3 containing 100 μg/mL cycloheximide and then used to prepare cytoplasmically enriched lysate as described (Subtelny et al., 2014). An aliquot of cleared lysate was flash frozen for use in ribosome profiling, and the rest of the lysate was added to 5 volumes of TRI reagent (Ambion) and frozen at –80°C. Samples stored in TRI reagent were thawed at room temperature, and RNA was purified according to the manufacturer's protocol and used for RNA-seq and PAL-seq. The non-5EU sample for the miR-1 cell line was labeled with 100 μM 4-thiouridine for 4 h prior to harvesting. 4-thiouridine has no known cross-reactivity with the reagents used to label 5EU.

RNA Standards

Two sets of tail-length standards (set 1 and set 3; Table S1) were described previously (standard mix 2 and standard mix 1) (Subtelny et al., 2014). The other set of standards (set 2; Table S1) was prepared based on a 705 nt fragment of the Renilla luciferase mRNA, which was transcribed and gel purified as described (Subtelny et al., 2014) and then capped using a Vaccinia capping system (2000 μL reaction containing 500 μg RNA, 1000 U Vaccinia capping enzyme (NEB), 1X Capping Buffer (NEB), 0.1 mM S-adenosyl methionine, 0.5 mM GTP, 50 nM [α–^{32}P]-GTP, 2000 U SUPERaseIn (ThermoFisher) at 37°C for 1 h), monitoring the amount of incorporated radioactivity to ensure that capping was quantitative. Following the capping reaction, the 2′,3′ cyclic phosphate at the 3′ end was removed using T4 polynucleotide kinase (Subtelny et al., 2014). The capped, dephosphorylated product was joined by splinted ligation to each of seven different poly(A)-tailed barcode oligonucleotides (Subtelny et al., 2014). These seven 3′-ligation partners included 110 and 210 nt poly(A) oligonucleotides prepared as described (Subtelny et al., 2014), and five gel-purified synthetic oligonucleotides (IDT), one with a 10 nt poly(A) tract and

the other four with a 29 nt poly(A) tract followed by either A, C, G, or U. Ligation products were gel purified, mixed in desired ratios, with the final ratios of the different-sized species confirmed by analysis on a denaturing polyacrylamide gel.

Short and long standards were used to monitor enrichment of 5EU-containing fragmented RNA or non-fragmented RNA, respectively. Short 5EU standards were prepared by in vitro transcription of annealed DNA oligos to produce a 30 nt and 40 nt RNA, with the latter containing a single 5EU (Table S1). In vitro transcription was performed with the MEGAscript T7 transcription kit (ThermoFisher) according to the manufacturer's protocol, except UTP was replaced with 5-ethynyluridine-triphosphate (Jena Biosciences) when transcribing the 40 nt RNA. Long standards were prepared by in vitro transcription of sequences encoding firefly luciferase and GFP using the MEGAscript T7 transcription kit and 0.1 μM PCR product as the template. When transcribing GFP RNA, a 20:1 ratio of UTP to 5-ethynyluridine-triphosphate was used. Short and long standards were gel purified and stored at –80°C. Prior to use, a portion of each standard was cap-labeled and gel purified again, which enabled measurement of the recovery of the 5EU-containing standard relative to that of the uridine-only standard.

Three 28–30 nt RNAs (Table S1) were synthesized (IDT) for use as quantification standards in RNA-seq and ribosome-profiling libraries. These standards were gel purified, and 0.1 fmol of each was added to each sample immediately prior to library preparation.

Biotinylation of 5EU Labeled RNA

For RNA-seq libraries, poly(A) RNA was purified from 50 μg total RNA of the 40 min, 1, 2, and 4 h samples and 25 μg total RNA of the 8 h sample and non-5EU-labeled sample using oligo(dT) Dynabeads (ThermoFisher) according to manufacturer's protocol. RNA was fragmented, 27–33

nt fragments were isolated as described (Subtelny et al., 2014), short standards that monitored 5EU enrichment were added. Biotinylation was performed in the presence of a Cu(II) catalyst in a 20 µL reaction containing 50 mM HEPES, pH 7.5, 4 mM disulfide biotin azide (Click Chemistry Tools), 2.5 mM $CuSO_4$, 2.5 mM Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA, Sigma-Aldrich), and 10 mM sodium ascorbate, incubated at room temperature for 1 h. Reactions were stopped with 5 mM EDTA and then extracted with phenol-chloroform (pH 8.0). For the steady-state samples, 5 µg of RNA from the 40 min sample was poly(A) selected and fragmented, and 27–33 nt fragments were size selected and carried forward without enriching for 5EU.

For ribosome profiling libraries, 800 µL aliquots of lysate were digested with 0.7 U/µL RNase I (Ambion) for 30 min at room temperature and then run on a sucrose gradient to purify monosomes (Subtelny et al., 2014). For each 40 min sample, monosomes from three aliquots were combined. For samples from all other time intervals, monosomes from two aliquots were combined, and for the non-5EU-labeled sample, only one aliquot was run. RPFs were released and purified (Subtelny et al., 2014), short standards used to monitor 5EU enrichment and recovery were added (using a 1:10 ratio of 5EU-containing standard to non-5EU-containing standard), and then click reactions were performed as above. For the steady-state samples, RPFs were isolated from one aliquot of the 40 min sample and carried forward without enriching for 5EU.

For PAL-seq v2, long standards used to monitor 5EU enrichment and recovery were added to total RNA (using a 1:10 ratio of 5EU-containing standard to non-5EU-containing standard), and samples were biotinylated as above in reactions with 2.5 µg/µL RNA. For samples from the uninduced and induced miR-1 time courses, biotinylation was performed with 800, 525,

350, 200, or 50 µg total RNA for the 40 min, 1 h, 2 h, 4 h, or non-5EU-labeled samples, respectively. For samples from the uninduced miR-155 time course, biotinylation was performed with 500, 500, 250, 200, or 100 µg total RNA for the 40 min, 1 h, 2 h, 4 h, or 8 h samples, and for samples from the induced miR-155 time course, biotinylation was performed with 400, 400, 250, 200, or 100 µg total RNA for the 40 min, 1 h, 2 h, 4 h, or 8 h samples, respectively. The non-fragmented miR-155 time course samples did not include a non-5EU-labeled sample (used solely to determine the background recovery in Figure S2).

Purification of 5EU Labeled RNA

For RNA-seq and ribosome-profiling samples, Dynabeads MyOne Streptavidin C1 beads (ThermoFisher) for each set of samples were combined and batch washed, starting with 200 µL of beads per reaction. Beads were washed twice with 1X B&W buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl and 0.005% Tween-20), twice with solution A (0.1 M NaOH, 50 mM NaCl), twice with solution B (0.1 M NaCl), and then twice with water, using for each wash a volume equal to that of the initial bead suspension. Following the last wash, beads were resuspended in an initial bead volume of 1X high salt wash buffer (HSWB, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1 M NaCl, 0.01% Tween-20) supplemented with 0.5 mg/mL yeast RNA (ThermoFisher) and incubated at room temperature for 30 min with end-over-end rotation, again using a volume equal to that of the initial bead suspension. Beads were then washed three times with 200 µL 1X HSWB per reaction and split for each reaction during the last wash. After the wash was removed, sample RNA resuspended in 200 µL 1X HSWB was added to blocked beads and incubated with end-over-end rotation at room temperature for 30 min. Beads were washed twice with 800 µL 50°C water, incubating at 50°C for 2 min for each wash, and then twice with

800 µL 10X HSWB. RNA was eluted from beads by incubating with 200 µL 0.5 M tris(2-carboxyethyl)phosphine (TCEP, Sigma-Aldrich) at 50°C for 20 min with end-over-end rotation. The initial eluate was collected, and beads were resuspended in 150 µL water and eluted again, combining the two eluates for each sample. RNA from the eluate was then ethanol precipitated using linear acrylamide as a carrier.

Purifications of non-fragmented RNA were performed as above, except bead volumes were adjusted based on estimates of the amount of labeled RNA in each sample. For the uninduced and induced miR-1 samples, 467, 452, 575, 598, and 250 µL streptavidin beads were used for the 40 min, 1 h, 2 h, 4 h, and non-5EU-labeled samples, respectively. For the uninduced miR-155 samples, 292, 431, 410, 598, and 500 µL of beads were used, and for the induced miR-155 samples, 234, 345, 410, 598, and 500 µL of beads were used for the 40 min, 1 h, 2 h, 4 h, and 8 h samples, respectively.

Pilot experiments designed to optimize the 5EU biotinylation and purification confirmed that RNAs containing at least one 5EU could be purified efficiently, with over 80% of a model RNA substrate containing a single 5EU becoming biotinylated in a 1 h reaction (Figure S2A). This high reaction efficiency was critical for both the ribosome-profiling and RNA-seq analyses, as RPFs and the RNA fragments from the paired RNA-seq libraries were only ~30 nt long and estimated to typically contain at most a single 5EU. Indeed, for each of the three protocols, which started with either full-length RNA (PAL-seq) or fragmented RNA (RNA-seq and ribosome-footprint profiling), captured labeled RNA was substantially enriched above background (Figures S2B and S2C).

PAL-Seq

We used an improved form of PAL-seq called PAL-seq v2, using RNA standards of defined tail lengths to monitor library preparation, sequencing, and the computational pipeline. Steady-state RNA (25 μg of unselected RNA from the 40 min sample) or half of the RNA eluted from each 5EU-selected sample was used to prepare PAL-seq libraries. Tail-length standard mixes (1 ng of set 1 and 2 ng of set 2 for each 5EU-selected sample, and twice these amounts for the steady-state sample), and trace 5′-radiolabeled marker RNAs (Table S1) were added to each sample to assess tail-length measurements and ligation outcomes, respectively. Polyadenylated ends including those with a terminal uridine were ligated to a 3′-biotinylated adaptor DNA oligonucleotide (1.8 μM) in the presence of two splint DNA oligonucleotides (1.25 μM and 0.25 μM for the U and A-containing splint oligos, respectively, Table S1) using T4 Rnl2 (NEB) in an overnight reaction at 18°C. Following 3′-adaptor ligation the RNA was extracted with phenol-chloroform (pH 8.0), precipitated, resuspended in 1X RNA T1 sequence buffer (ThemoFisher), heated to 50°C for 5 min and then put on ice. RNase T1 was then added to a final concentration of 0.006 U/μL and the reaction incubated at room temperature for 30 min followed by phenol chloroform extraction and precipitation. RNA was subsequently captured on streptavidin beads, 5′ phosphorylated, and ligated to a 5′ adaptor as described (Subtelny et al., 2014) but using a modified 5′ adaptor sequence (Table S1). Following reverse transcription using SuperScript III (Invitrogen) with a barcode-containing DNA primer, cDNA was purified as described (Subtelny et al., 2014), except a 160–810 nt size range was selected. Libraries were amplified by PCR for 8 cycles using Titanium Taq polymerase according to the manufacturer's protocol with a 1.5 min combined annealing/extension step at 57°C. PCR-amplified libraries were purified using

AMPure beads (Agencourt, 40 μL beads per 50 μL PCR, two rounds of purification) according to the manufacturer's instructions.

The use of a splinted ligation of the 3′ adaptor to the poly(A) tail had the advantage of specifically ligating to mRNAs without the need to deplete ribosomal or other abundant RNAs. However, this approach was not suitable for acquiring measurements for mRNAs with tails that were either very short (< 8 nt) or extended by more than one uridine, because such tails would ligate less efficiently (or not at all) when using a splinted ligation to the 3′ adaptor. To account for these mRNAs with either very short or highly modified tails, we implemented a protocol that used single-stranded (ss) ligation and different mRNA enrichment steps to prepare libraries from steady-state RNA isolated from miR-1 and miR-155 cell lines in uninduced and induced conditions. For each sample, 5 μg of total RNA was depleted of rRNA using RiboZero Gold HMR (Illumina) and further depleted of the 5.8S rRNA by subtractive hybridization. Subtractive hybridization was performed by mixing 2x SSC buffer (3 M sodium chloride, 300 mM sodium citrate, pH 7.0), total RNA, and 4.8 μM of each 5.8S subtractive-hybridization oligo (Table S1) in a 50 μL reaction, heating the reaction to 70°C for 5 min, then cooling at 1°C/min to 37°C to anneal the oligos to the RNA. During this cooling, 250 μL of Dynabeads MyOne Streptavidin C1 beads per sample (ThermoFisher) were washed twice with 1X B&W buffer (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl and 0.005% Tween-20), twice with solution A (0.1 M NaOH, 50 mM NaCl), twice with solution B (0.1 M NaCl), and then resuspended in 50 μL of 2X B&W buffer. After cooling, the entire 50 μL RNA/oligo mixture was added to 50 μL of washed beads, then incubated at room temperature for 15 min with end-over-end rotation. The sample was then magnetized and the supernatant was withdrawn and precipitated by adding 284 μL of water, 4 μL of 5 mg/mL linear acrylamide, and 1 mL of ice-cold 96% ethanol. After resuspension, RNA

was ligated to a 3′ adaptor containing four random-sequence nucleotides and an adenylyl group at its 5′ end (Table S1) in a 70 µl reaction containing 10 µM adaptor, 1X T4 RNA Ligase Reaction Buffer (NEB), 20 U/µL T4 RNA Ligase 2 truncated KQ (NEB), 0.3 U/µL SUPERaseIn (ThermoFisher), and 20% PEG 8000. The reaction was incubated at 22°C overnight and then stopped by addition of EDTA (3.5 mM final after bringing the reaction to 400 µL with water). RNA was phenol–chloroform extracted, precipitated, and subsequent library preparation was as for the splinted-ligation libraries.

PAL-seq v2 libraries were sequenced on an Illumina HiSeq 2500 operating in rapid mode. Hybridization mixes were prepared with 0.375 fmol PCR-amplified library that had been denatured with standard NaOH treatment and brought to a final volume of 125 µL with HT1 hybridization buffer (Illumina, 3 pM library in final mix). Following standard cluster generation and sequencing-primer hybridization, two dark cycles were performed for the splinted-ligation libraries (i.e., two rounds of standard sequencing-by-synthesis in which imaging was skipped), which extended the sequencing primer by 2 nt, thereby enabling measurement of poly(A) tails terminating in non-adenosine bases. For the direct-ligation libraries, six dark cycles were performed instead of two, which extended the sequencing primer past the four random-sequence nucleotides in the 3′ adaptor and then the last two residues of the tail.

Following the two dark cycles, a custom primer-extension reaction was performed on the sequencer using 50 µM dTTP as the only nucleoside triphosphate in the reaction. To perform this extension, the flow cell temperature was first set to 20°C. Then, 120 µL of universal sequencing buffer (USB, Illumina) was flowed over each lane, followed by 150 µL of Klenow buffer (NEB buffer 2 supplemented with 0.02% Tween-20). Reaction mix (Klenow buffer, 50 µM dTTP, and 0.1 U/µL units Large Klenow Fragment, NEB) was then flowed on in two aliquots (150 µL and

100 μL). The flow-cell temperature was then increased to 37°C at a rate of 8.5°C per min and the incubation continued another 2 min after reaching 37°C. 150 μL of fresh reaction mix was then flowed in, and following a 2 min incubation, 75 μL of reaction mix was flowed in eight times, with each flow followed by a 2 min incubation. The reaction was stopped by decreasing the flow cell temperature to 20°C, flowing in 150 μL of quench buffer (Illumina HT2 buffer supplemented with 10 mM EDTA) and then washing with 75 μL of HT2 buffer. The flow cell was prepared for subsequent sequencing with a 150 μL and a 75 μL flow of HT1 buffer (Illumina). 50 cycles of standard sequencing-by-synthesis were then performed to yield the first sequencing read (read 1). XML files used for this protocol are provided at https://github.com/kslin/PAL-seq.

The flow cell was stripped, a barcode sequencing primer was annealed, and seven cycles of standard sequencing-by-synthesis were performed to read the barcode. The flow cell was then stripped again, and the same primer as used for read 1 was hybridized and used to prime 250 cycles of standard sequencing-by-synthesis to generate read 2. Thus, each PAL-seq tag consisted of three reads: read 1, read 2, and the indexing (barcode) read. For cases in which a tag corresponded to a polyadenylated mRNA, read 1 was the reverse complement of the 3′ end of the mRNA immediately 5′ of the poly(A) tail and was used to identify the mRNA and cleavage-and-polyadenylation site of long-tailed mRNAs. The indexing read was used to identify the sample, and read 2 was used to measure poly(A)-tail length and identify the mRNA and cleavage-and-polyadenylation site of short-tailed mRNAs. The intensity files of reads 1 and 2 were used for poly(A)-tail length determination, along with the Illumina fastq files.

PAL-Seq v2 Data Analysis

Tail lengths for the splinted-ligation data were determined using a Gaussian hidden Markov model (GHMM) from the python2.7 package ghmm (http://ghmm.org/), analogous to the model used in TAIL-seq (Chang et al., 2014) and described in the next paragraph. Read 1 was mapped using STAR (v2.5.4b) run with the parameters '–alignIntronMax 1 –outFilterMultimapNmax 1 –outFilterMismatchNoverLmax 0.04 –outFilterIntronMotifs RemoveNoncanonicalUnannotated –outSJfilterReads', aligning to an index of the mouse genome built using mm10 transcript annotations that had been compressed to unique instances of each gene selecting the longest transcript and removing all overlapping transcripts on the same strand (Eichhorn et al., 2014). The genome index also included sequences of the quantification spikes and the common portion of the poly(A)-tail length standards. The sequences that identified each RNA standard (the last 20 nt of each standard sequence; Table S1) were not aligned using STAR. Instead, the unix program grep (v2.16) was used to determine which reads matched each standard (allowing no mismatches), and these reads were added to the aligned reads from the STAR output. Tags corresponding to annotated 3′ UTRs of mRNAs were identified using bedtools (v2.26.0), and if the poly(A)-tail read (read 2) contained a stretch of ≥ 10 T residues (the reverse complement of the tail) in an 11-nt window within the first 30 nt, this read was carried forward for GHMM analysis. If read 2 failed to satisfy this criterion but began with ≥ 4 T residues, the tail length was called based on the number of contiguous T residues at the start of read 2; by definition, these tails were < 10 nt and thus easily determined by direct sequencing.

For each read 2 that was to be input into the GHMM a 'T signal' was first calculated by normalizing the intensity of each channel for each cycle to the average intensity of that channel when reading that base in read 1 and then dividing the thymidine channel by the sum of the other

three channels. Sometimes a position in a read would have a value of 0 for all four channels. A read was discarded if it contained more than five such positions. Otherwise, the values for these positions were imputed using the mean of the five non-zero signal values upstream and downstream (ten positions total) of the zero-valued position. A three-state GHMM was then used to decode the sequence of states that occurred in read 2. It consisted of an initiation state (state 1), a poly(A)-tail state (state 2), and a non-poly(A)-tail state (state 3). All reads start in state 1. From state 1 the model can remain in state 1 or transition to state 2. From state 2 the model can either remain in state 2 or transition to state 3. The model was initialized with the following transition probabilities:

$from \setminus to$	$state_1$	$state_2$	$state_3$
$state_1$	0.001	0.95	0.049
$state_2$	0.001	0.95	0.049
$state_3$	0.001	0.001	0.998

The initial emissions were Gaussian distributions with means of 100, 1, and –1 and variances of 1, 0.25 and 0.25, respectively. In general, the emission Gaussians for the model corresponded to the logarithm of the calculated T signal at each sequenced base in read 2. The initial state probabilities were 0.998, 0.001, and 0.001 for states 1, 2 and 3, respectively.

After initializing the model, unsupervised training was performed on 10,000 randomly selected PAL-seq tags, and then the trained model was used to decode all tags, with the number of state 2 cycles reporting the poly(A)-tail length for a tag. Only genes with ≥ 50 poly(A)-tail length measurements were considered for analyses involving mean poly(A)-tail lengths.

Analysis of PAL-Seq SS-Ligation Protocol

To account for mRNAs with very short tails or extensive terminal modifications, we implemented a version of PAL-seq that did not use splinted ligation. Tail lengths from these ss-ligation datasets, acquired for steady-state samples from both cell lines, were determined using a modified version of the PAL-seq analysis pipeline written for python3. The T-signal in this pipeline was modified to allow more accurate detection of mRNAs lacking tails. Instead of normalizing the intensity of each channel for each cycle to the average intensity of that channel when reading that base in read 1, the intensity of each channel was normalized to the average intensity of the channels for the other three bases in read 1. The intensity of the T channel was then divided by the sum of the other channel intensities to calculate the T signal, and tails were called using the hmmlearn package (v0.2.0). Tags representing short tails, including short tails that ended with many non-A residues, were identified as those for which read 1 and read 2 mapped to the same mRNA 3′ UTR (usually ~4% of the tags). Tail lengths for these tags were called without the use of the GHMM. Instead, their tail lengths were determined by string matching, allowing any number of untemplated U residues but no more than two G or C residues to precede the A stretch. Tags not identified as representing short-tails were analyzed using the GHMM, excluding from further analysis occasional outliers determined by the GHMM to have tails ≤ 8 nt.

Most of the tags that had either only a very short tail or no tail did not correspond to mRNA cleavage-and-polyadenylation sites. Therefore, to be carried forward in our analysis, short-tailed tags were required to have a 3′-most genome mapping position (as determined from read 1 but requiring that read 2 also map uniquely to the same 3′ UTR) that fell within a 10 nt window of a PAL-seq–annotated cleavage-and-polyadenylation site.

Although the single-stranded ligation protocol provided the opportunity to account for mRNAs with very short or highly modified tails, examination of the recovery of internal standards indicated that tags representing longer tails (≥100 nt) were not as well recovered in the datasets in which we implemented ss ligation. Therefore, for steady-state samples from each cell line, we generated composite tail-length distributions in which the ss-ligation dataset contributed to the distribution of tails < 50 nt, and the splinted-ligation dataset contributed to the distribution of tails ≥ 50 nt. For example, *Slc38a2* had 635 standard PAL-seq tags, 169 of which (~27%) had tails < 50 nt, and this same gene had 703 ss-ligation PAL-seq tags, 393 of which (~56%) had tails < 50 nt. The composite tail-length distribution replaced the 169 short-tailed splinted-ligation PAL-seq tags with the 393 short-tailed ss-ligation PAL-seq tags, normalizing the latter cohort by a scaling factor. This scaling factor was determined from the ratio of the counts of the splinted-ligation tags with tail lengths between 30–70 nt (135 tags) to the counts of the corresponding tags in the ss-ligation dataset (153 tags).

3′-end annotations were generated from PAL-seq tags with tails ≥ 11 nt, using an algorithm previously developed for data from poly(A)-position profiling by sequencing (3P-seq) (Jan et al., 2011). Each PAL-seq read 1 that mapped (with at least 1 nt of overlap) to an annotated 3′ UTR (Eichhorn et al., 2014) was compiled by the genomic coordinate of its 3′-UTR nucleotide closest to the tail. The genomic coordinate with the most mapped reads was annotated as a 3′ end. All reads within 10 nt of this end (a 21 nt window) were assigned to this end and removed from subsequent consideration. This process was repeated until there were no remaining 3′ UTR-mapped reads. For each gene, the 3′-end annotations were used in subsequent analyses if they accounted for ≥ 10% of the 3′-UTR–mapping reads for that gene.

Documentation and code to calculate and analyze T signals and determine tail lengths are available for both the splinted-ligation and ss-ligation pipelines at https://github.com/kslin/PAL-seq.

RNA-Seq and Ribosome Profiling

Fragmented poly(A)-selected RNA and RPFs were supplemented with three short quantification standards (Table S1), and then ligated to adaptors, reverse-transcribed, and amplified to prepare the RNA-seq and ribosome-profiling libraries, respectively (Subtelny et al., 2014). These libraries were sequenced on an Illumina HiSeq 2500. For all RNA-seq and ribosome-profiling data, only reads mapping to ORFs of annotated gene models (Eichhorn et al., 2014) were considered, excluding the first 50 nt of each ORF. A cutoff of ≥ 10 reads per million mapped reads (RPM) was applied to each gene in each sample, with the exception of those samples generated from miRNA-induced cell lines, for which no cutoff was applied.

Calculation of miRNA-Mediated Repression

Secondary effects of expressing a miRNA can have a greater impact on mRNAs with longer 3′ UTRs relative to those with shorter 3′ UTRs (Agarwal et al., 2015), presumably because longer 3′ UTRs tend to contain more sites to other regulatory factors, including other miRNAs. As a result, 3′-UTR length differences can complicate the measurement of the repressive effects of an expressed miRNA. For this reason, we first normalized the fold-changes of all mRNAs based on their 3′ UTR length. The relationship between the fold-change for all mRNAs without a 6-nt seed-matched site to the induced miRNA in the entire transcript (no-site mRNAs) and 3′-UTR length was calculated using linear regression, and then the fold-changes of all mRNAs (with and

without a target site) were normalized by their 3′ UTR lengths such that the slope of the relationship between no-site mRNAs and 3′ UTR length was 0. We then compared normalized fold-changes for mRNAs containing at least one predicted miRNA target site in their 3′-UTR to those for the no-site mRNAs. For all mRNAs passing our expression threshold in the uninduced samples, we calculated the $\log_2$ fold-changes in mRNA abundance, RPF abundance, or poly(A)-tail length in samples from induced cells relative to the corresponding samples from uninduced cells. The repressive effect of the miRNA on a set of predicted miRNA targets was then calculated by subtracting the median-normalized fold-change for no-site mRNAs from the mean-normalized fold-change for a set of predicted targets. Top targets were defined using RPF measurements from a previous study (Eichhorn et al., 2014), choosing from among the predicted targets those with expression that decreased to ≤ 75% of their original expression after 12 hours of miRNA induction.

The cumulative-distribution plots each used a subset of all possible no-site mRNAs for display and analysis. The subsets were generated by choosing, for each site-containing mRNA, a no-site mRNA from the complete list of no-site mRNAs by random sampling (without replacement) such that the chosen no-site mRNA had a UTR length that fell within 20% of the length of the UTR of the site-containing mRNA. For each dataset, this sampling was performed 101 times and the control cohort yielding the median fold-change was chosen and used for plotting and analysis. The residual pool of no-site mRNAs was occasionally not sufficient to choose a no-site mRNA for each site-containing mRNA, which is why, for some figures, the no-site cohort had a slightly smaller n, particularly when the site-containing cohort had a large n.

Modeling miRNA Effects

To fit miRNA effects, we used model parameters for mRNAs from each gene, which had been fit for the uninduced miR-1 and miR-155 cell lines (Eisen et al., 2020) as a starting point for an additional round of model fitting. In this fitting, either the deadenylation rate constant, the rate constant for deadenylation-dependent decay of the mRNA body, or both of these rate constants were tuned in order to minimize the residual between the observed and simulated miRNA-induced fold-changes in mRNA abundance at steady state (calculated by summing the output of the model over all tail lengths ≤ 249). The optimization was performed using the "L-BFGS-B" method of the *optim* function in R.

Background Subtraction for PAL-Seq Data

Although the efficacy of the 5EU purification enabled efficient enrichment of labeled RNAs at short time intervals, we also modeled and corrected for residual background caused by non-specific binding of the unlabeled RNA to the streptavidin beads (Eisen et al., 2020).

We designed our background model under the assumption that the background in the time courses stems primarily from the capture of a fixed amount of non-5EU labeled mRNA during the 5EU purification. Accordingly, we subtracted a fraction (0.3%) of the steady-state data from each continuous-labeling dataset. This fraction of input sample was chosen such that at 40 min long-lived genes (half-life ≥ 8 h) had no mRNAs with tail lengths ≲100 nt on average, but short-lived genes (half-life ≤ 30 min) were unaffected (Eisen et al., 2020). Genes were included in the final background-subtracted set only if the sum of their background-subtracted tag counts was ≥ 50 tags.

After background subtraction, PAL-seq datasets were scaled to each time interval by matching the total number of background-subtracted tags for all genes at all tail lengths to the total number of tags for all genes for the corresponding time interval in the RNA-seq data. The scaled PAL-seq data were then used to compute half-lives for each gene, scaling the steady-state sample using a globally fitted constant.

Quantification and Statistical Analysis

Graphs were generated and statistical analyses were performed using R (R Core Team, 2019). Statistical parameters including the value of n, statistical test, and statistical significance (p value) are reported in the figures and their legends. No statistical methods were used to predetermine sample size.

Data and Code Availability

Raw and processed RNA-seq, ribosome profiling, and PAL-seq data are available at the GEO, accession number GSE134660. Code for configuring an Illumina HiSeq 2500 machine for PAL-seq and for calculation of tail lengths from PAL-seq data are available at https://github.com/kslin/PAL-seq. Code for fitting kinetic models of abundance and tail length are available at https://github.com/timeisen/DynamicsOfCytoplasmicMrnaMetabolism.

Supplemental Table Titles and Legends

Table S1. Oligonucleotides Used in This Study, Related to STAR Methods

This table is published as Table S3 in Chapter 2.

References

Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife *4*.

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.

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell *173*, 20-51.

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.

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.

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.

Christie, M., Boland, A., Huntzinger, E., Weichenrieder, O., and Izaurralde, E. (2013). Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. Mol Cell *51*, 360-373.

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.

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.

Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., Lin, K.S., McGeary, S.E., Gupta, S., and Bartel, D.P. (2020). The Dynamics of Cytoplasmic mRNA Metabolism. Mol Cell *77*, 786-799 e710.

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.

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.

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.

Haas, G., Braun, J.E., Igreja, C., Tritschler, F., Nishihara, T., and Izaurralde, E. (2010). HPat provides a link between deadenylation and decapping in metazoa. J Cell Biol *189*, 289-302.

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.

Huntzinger, E., Kuzuoglu-Ozturk, D., Braun, J.E., Eulalio, A., Wohlbold, L., and Izaurralde, E. (2013). The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res *41*, 978-994.

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.

Jao, C.Y., and Salic, A. (2008). Exploring RNA transcription and turnover in vivo by using click chemistry. Proc Natl Acad Sci U S A *105*, 15779-15784.

Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet *16*, 421-433.

Kwasnieski, J.C., Orr-Weaver, T.L., and Bartel, D.P. (2019). Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns. Genome Res *29*, 1188-1197.

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

Ozgur, S., Chekulaeva, M., and Stoecklin, G. (2010). Human Pat1b connects deadenylation with mRNA decapping and controls the assembly of processing bodies. Mol Cell Biol *30*, 4308-4323.

Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. Mol Cell *25*, 635-646.

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.

Tritschler, F., Huntzinger, E., and Izaurralde, E. (2010). Role of GW182 proteins and PABPC1 in the miRNA pathway: a sense of deja vu. Nat Rev Mol Cell Biol *11*, 379-384.

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, A., Chang, T.C., Yamashita, Y., Zhu, W., Zhong, Z., Chen, C.Y., and Shyu, A.B. (2005). Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol *12*, 1054-1063.

Supplemental Figures and Legends

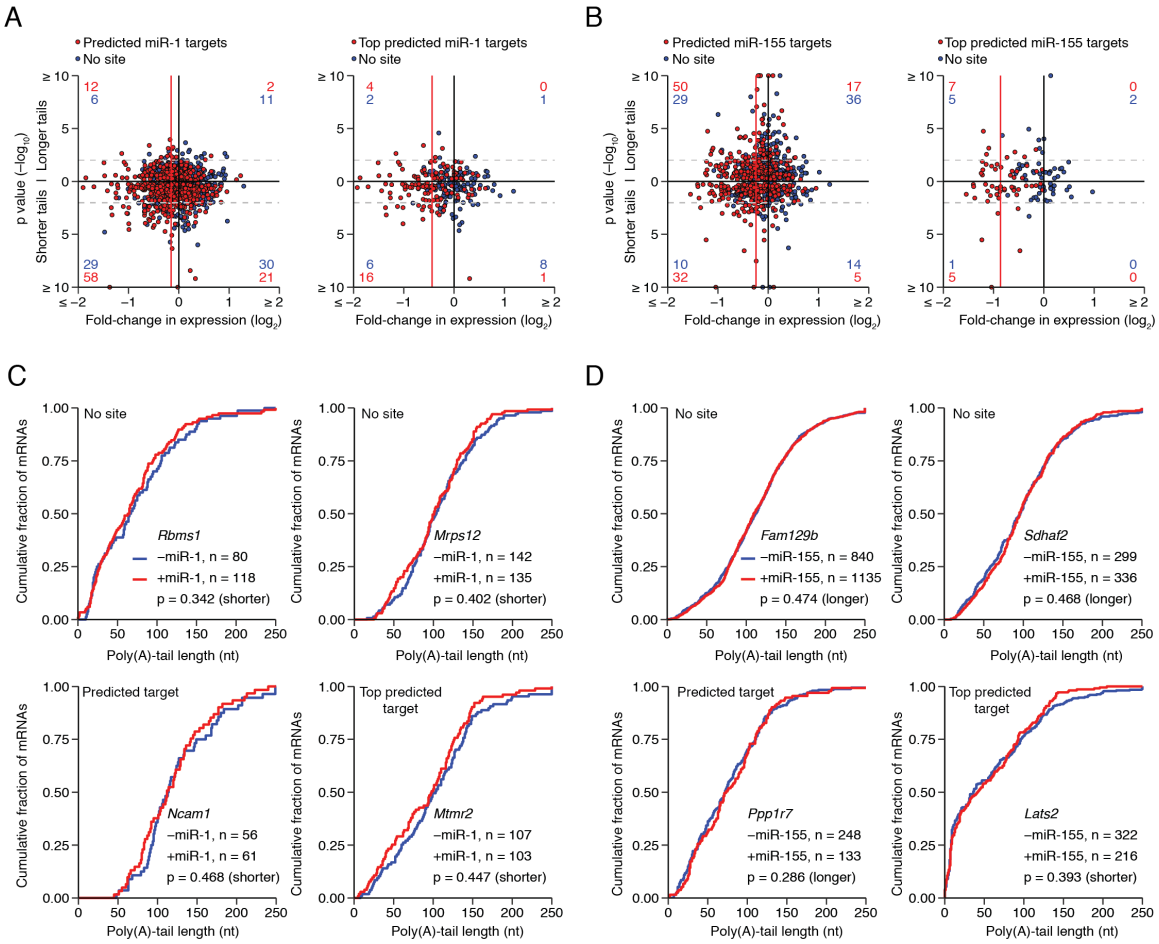


Figure S1. Very Small Effect of miRNAs on Steady-State Poly(A)-Tail Lengths When Examining mRNAs from Individual Genes, Related to Figure 1

(A) Influence of miR-1 on tail-length distributions of mRNAs from individual genes. On the left, K–S tests were performed for each predicted target of miR-1 (n = 643, red) and a corresponding set of no-site mRNAs (n = 624, blue). The p values ($-\log_{10}$) from these tests are plotted as a function of the miR-1–mediated fold-change in expression ($\log_2$), with placement above or below the *x* axis indicating the direction of the more significant p value in a one-tailed test, with values above the axis indicating a more significant increase in tail lengths with induction of the miR-1 and values below the axis indicating a more significant decrease in tail lengths. Horizontal dashed lines indicate p values of 0.01; for each quadrant, the number of predicted-target and no-site genes with p values below this threshold is indicated (red and blue, respectively). The median of the fold-change in expression for predicted targets of miR-1 after subtracting the median fold-change for the no-site controls is also indicated (red line). The analysis on the right is the same but considering the top predicted targets and their no-site controls (n = 130 and 130, respectively). (B) Influence of miR-155 on tail-length distributions of mRNAs from individual genes. Analysis is as

in (A), except for predicted targets of miR-155 and their no-site controls (n = 417 and 417, respectively) and top predicted targets of miR-155 and their no-site controls (n = 57 and 57, respectively). (C) Typical influence of miR-1 on tail-length distributions. Tail-length distributions observed with and without miRNA induction (red and blue, respectively) are plotted for mRNAs with the median p values in (A). *Rbms* (top left) was the mRNA from the no-site cohort of the predicted miR-1 targets with the median p value; *Mrps12* (top right) was the mRNA from the no-site cohort of the top predicted miR-1 targets with the median p value; *Ncam1* was the predicted miR-1 target with the median p value, and *Mtmr2* was the top predicted miR-1 target with the median p value. The p values, calculated in (A), are shown along with the numbers of tags in each distribution (key). (D) Typical influence of miR-155 on tail-length distributions. Tail-length distributions observed with and without miRNA induction (red and blue, respectively) are plotted for mRNAs with the median p values in (B). Otherwise, this panel is as in (A).

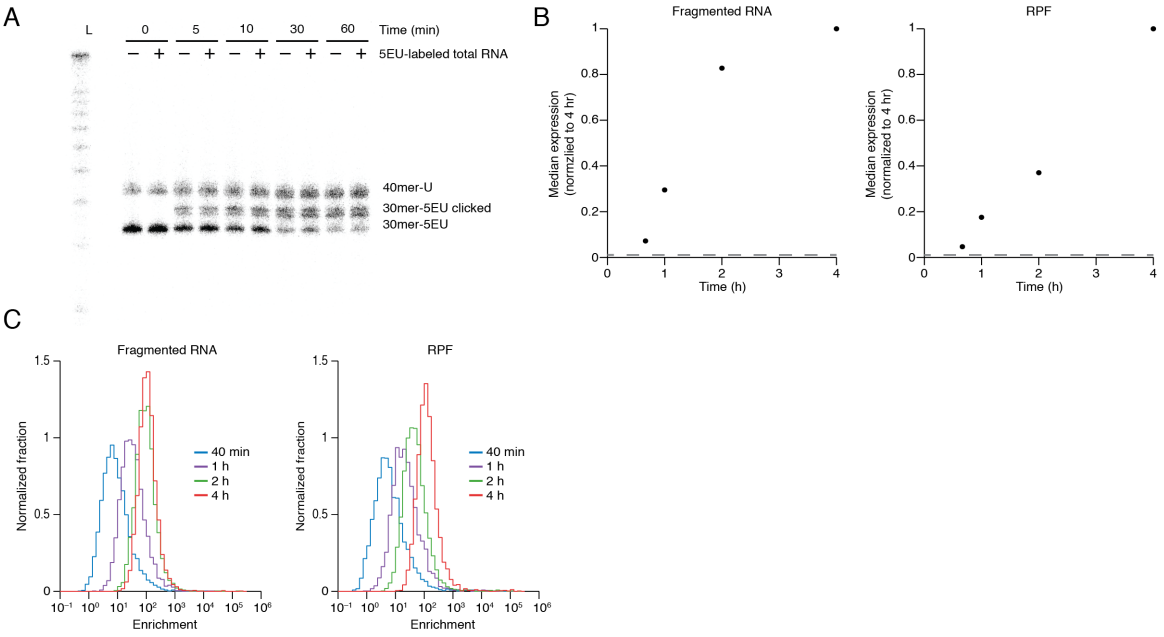


Figure S2. Enrichment of 5EU–Labeled RNA, Related to Figure 2

(A) Biotinylation of a 5EU–containing RNA. A 40 nt RNA containing a single uridine and a 30 nt RNA containing a single 5EU were mixed with or without an additional 50 µg of 5EU–labeled total RNA and then reacted with disulfide biotin azide for the indicated amount of time. Decreased mobility was observed upon biotinylation, with ~80% of the 5EU–containing RNA shifting to a slower mobility (30mer-5EU clicked) after 1 h and no detectable formation of a slower-mobility product from the 40 nt RNA that lacked 5EU. The clicked-labeled form of the RNA tended to run as a doublet, but reduced to a single band upon treatment with TCEP (not shown). Note that although in this pilot experiment the 40 nt RNA lacks 5EU and the 30 nt RNA contains it, the samples prepared for sequencing used a 30 nt RNA lacking 5EU and a 40 nt RNA containing it so that the 5EU containing RNA would be removed after size selection. (B) Abundance normalization of sequencing reads. mRNA was fragmented using either alkali (left) or as ribosome protected fragments (RPFs, right), and then 5EU-containing fragments were isolated and used to make libraries. To evaluate the background in labeled samples, we also attempted to isolate 5EU-containing fragments of both types starting with RNA from cells that had never received 5EU. All samples contained RNA standards that were used to scale mRNA abundance measurements. Median-scaled expression is plotted for both sets of libraries in the miR-1 line after normalizing to that of the 4 h sample for mRNAs passing cutoffs for RNA-seq and RPF analyses. No cutoff and a pseudo-count of 0.1 reads was applied to the unlabeled samples used to assess background. The normalized background expression is plotted as a dashed grey line. (C) Enrichment of labeled RNAs in fragmented RNA and RPF samples. Absolute mRNA abundance in each sample was normalized to expression in the corresponding unlabeled background sample, and the enrichment ratios for mRNA from each gene are plotted as a histogram. An enrichment of 1 would indicate as many reads in the labeled sample as in the background sample. For all mRNAs with enrichments >1, most reads came from labeled RNA; otherwise, as in (B).

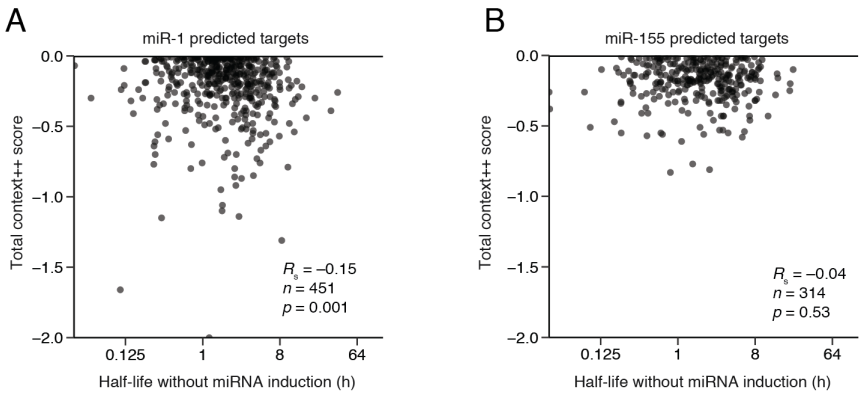


Figure S3. Half-life Analyses of miRNA Top Targets, Related to Figure 5

(A) The relationship between basal half-lives and predicted miR-1 target-site efficacy. Total context++ scores (Agarwal et al., 2015) for predicted targets of miR-1 were plotted with respect to half-life in the uninduced condition. mRNAs with predicted sites that do not have context ++ scores (n = 10) have sites in mRNA isoforms that fell below the TargetScan7 cutoff for isoform abundance. Otherwise as in Figure 5A. (B) The relationship between basal half-lives and predicted miR-155 target-site efficacy. As in (A), except plotting the miR-155 predicted targets in Figure 5B and the respective context++ scores. Two mRNAs had sites in isoforms that fell below the TargetScan7 cutoff for isoform abundance.

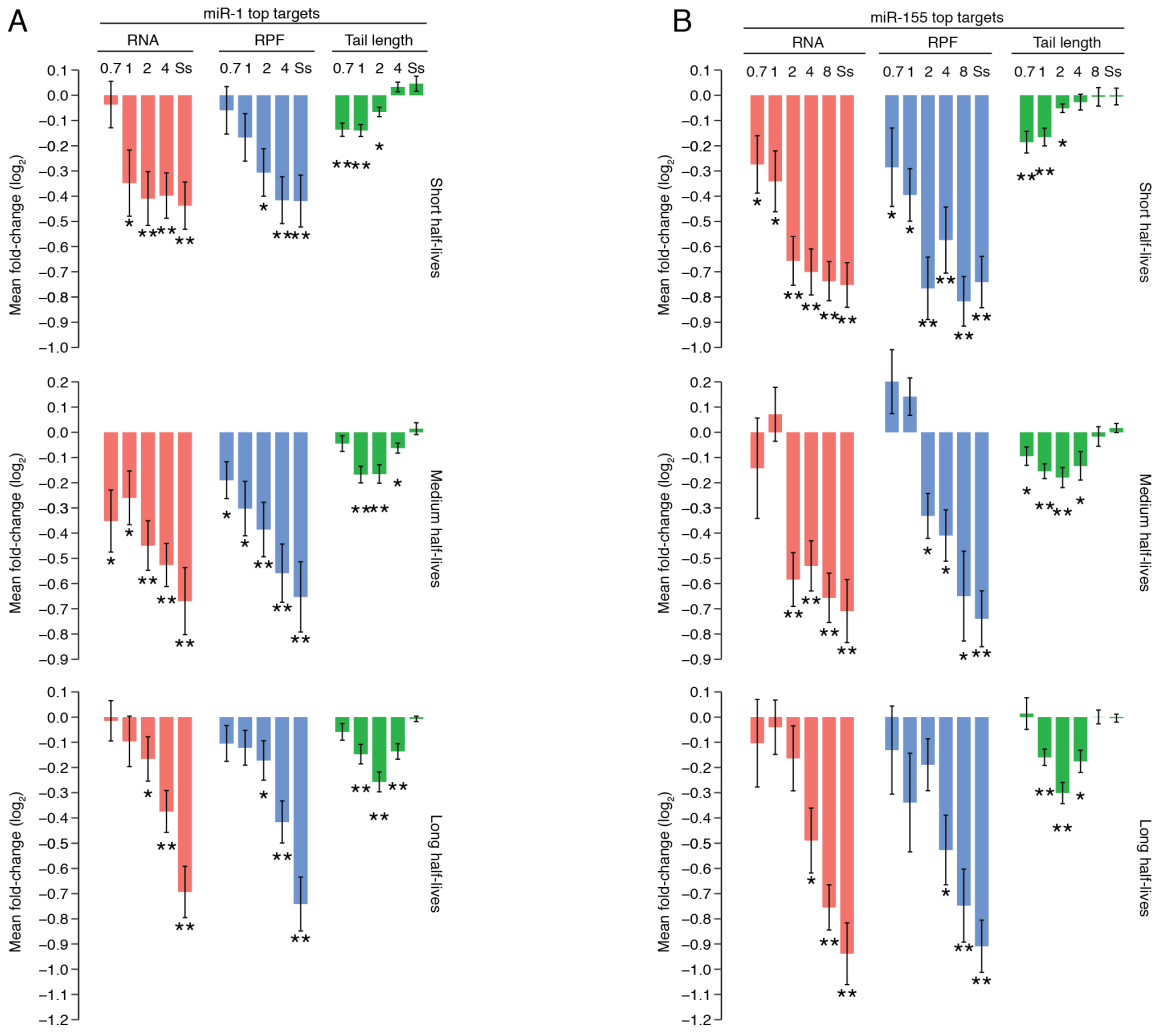


Figure S4. Half-life Analyses of miRNA Top Targets, Related to Figure 5

(A) The dynamics of miR-1–mediated repression for top targets. miR-1-mediated changes in RNA and RPF abundance and mean poly(A)-tail length for top miR-1 targets that were split into three groups (n = 25, 17 and 24 for short, medium and long half-life targets, respectively, using the same no-site genes and half-life ranges as in Figure 5C). Only mRNAs passing cutoffs for half-life measurements and RNA, RPF, and poly(A)-tail length measurements at all time points were analyzed. (B) The dynamics of miR-155–mediated repression for top targets. miR-155–mediated changes in RNA and RPF abundance and mean poly(A)-tail length for top miR-155 targets that were split into three groups (n = 19, 11 and 11 for short, medium and long half-life targets, respectively). Half-life cutoffs and numbers of no-site genes as in Figure 5D. Otherwise, as in (A).

Chapter 4. Poly(A)-Tail Length has a Significant but Modest Effect on Translational Efficiency in Neurons

Timothy J. Eisen[1,2,3], and David P. Bartel[1,2,3]

[1]Howard Hughes Medical Institute, Cambridge, MA, 02142, USA
[2]Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Unpublished.

Summary

Neurons require unique mechanisms of translational control to modify gene expression programs rapidly in response to stimuli. These mechanisms are reflected in the broad range of translational efficiencies (TE) that can be observed using ribosome profiling. Here, we investigate the landscape of translational control in neurons and neuronal tissues to determine features of mRNAs that explain this broad range. Building on reports that the length of the poly(A) tail influences TE, we profiled poly(A)-tail lengths transcriptome-wide using poly(A)-tail length sequencing (PAL-seq) and TAIL-seq. These data show that tail lengths change in response to stimulation in the visual cortex of light-stimulated mice, and a subset of these changes occur without corresponding changes to expression level, perhaps through cytoplasmic polyadenylation. Despite these changes, tail length explains only ~5% of the variance in TE in primary-culture neurons, and even less in neuronal tissues. Larger determinants of TE include coding-sequence length, expression level, and codon composition. Overall, our linear model of translation explains 30–40% of the variance in TE, indicating that the factors contributing to the majority of this variance remain unknown.

Introduction

In the biochemically tractable cells that are often used for high-throughput biochemical studies, transcriptional regulation accounts for a large fraction of the variance in gene expression (Schwanhausser et al., 2011; Li et al., 2014). Global techniques such as RNA sequencing, ribosome profiling, and mass spectrometry have shown that the relationship between mRNA levels and protein levels is a close one, and indeed if one also considers the fraction of the genome that is transcriptionally silent, this relationship improves still further. Conclusions such as these, despite their robustness in well-studied cells, such as the mouse NIH 3T3 cell and the human A549, HeLa, and HEK293T cells, do not generalize to every tissue, particularly those with rapid temporal or precise spatial requirements for protein expression.

In the brain, the signals of learning and memory are integrated throughout rapidly-responding networks that alter their proteome in response to stimuli (Costa-Mattioli et al., 2009). Decades of study of these expression changes have enticed researchers to query the role of translation at a transcriptome-wide scale, curious as to whether the paradigms developed for cell lines can be applied to the less uniform and more delicate neuronal systems. Studies using ribosome profiling in neurons and brain tissues have indeed found an impressive variance in translational regulation, with mRNAs from some genes exhibiting an average of 1000-fold greater translational efficiency than others (Cho et al., 2015; Hornstein et al., 2016; Blair et al., 2017; Zappulo et al., 2017; Liu et al., 2018; Umegaki et al., 2018; Biever et al., 2019; Das Sharma et al., 2019; Rodriguez et al., 2019; Simbriger et al., 2020). Why these mRNAs are better translated, however, and what features of the mRNAs induce ribosome recruitment or avoidance, is not yet clear. Examples of motifs, such as the TOP motif (Thoreen et al., 2012), or proteins, such as FMRP (Qin et al., 2005), contribute to the landscape of translational control, but by

themselves are insufficient to explain the vast diversity of translational efficiency in mammalian neurons and brain tissues.

Additional phenomena that modify mRNAs also control translation, although the scope of their effect has not been investigated on a transcriptome-wide scale in neurons. Cytoplasmic polyadenylation, which extends the length of the poly(A)-tail of substrate mRNAs (Richter, 1999), is one such phenomenon. The ability of this extension to control translation was originally observed in *Xenopus* oocytes (Fox and Wickens, 1990; de Moor and Richter, 1997). A motif involved in cytoplasmic polyadenylation, the cytoplasmic polyadenylation element (CPE), is also found in a highly-expressed, activity-induced neuronal mRNA, *Camk2a*, and can cause poly(A)-tail lengthening of this mRNA in response to stimulation, both in cultured neurons and brain tissues (Wu et al., 1998; Dziembowska et al., 2012). Additional mRNAs may also be substrates for cytoplasmic polyadenylation in neurons, and screens to identify them, relying on differential thermal elution from poly(U)-sepharose columns, have revealed tens to hundreds more mRNAs, many containing the canonical CPE motif (Du and Richter, 2005; Udagawa et al., 2012). The impact of tailing is less clear, and questions remain about the extent of molecules affected, the extent of tail lengthening, the effect of the phenomenon on translation, and whether it is both cell-wide or distal-dendrite specific in its subcellular localization (Wu et al., 1998; Du and Richter, 2005).

In this study, we combined transcriptome-wide measurements of translation and poly(A)-tail length in mouse primary-culture neurons and brain tissues in order to investigate the mRNA features underlying translation, and whether control of poly(A)-tail length plays a central role in the activity-induced translation of mRNAs. We found a small but significant relationship between the length of the tail and the translational efficiency of mRNAs, but this relationship

was dwarfed by other features of the mRNAs, such as the length and composition of the coding sequence and 5′ UTR. After stimulation, we observe robust changes in tail lengths for mRNAs from immediate-early genes and others. Some of these changes cannot be explained by concomitant increases in expression, suggesting that while some mRNAs experience increased tail length because of their recent transcription, cytoplasmic polyadenylation modifies tail lengths of others.

Results

Light Stimulation Induces Poly(A)-Tail Length Changes in the Visual Cortex

We used visual stimulus of the primary visual cortex to investigate the relationship between stimulation and changes to mRNA levels and poly(A)-tail lengths. This paradigm is frequently used as a controlled physiological stimulus, as it generates rapid and widespread alterations to large portions of the visual cortex (Majdan and Shatz, 2006; Mardinly et al., 2016). In this study, we used tissue from an experiment investigating the diversity of cellular responses to stimulation using single-cell RNA sequencing (Hrvatin et al., 2018). The experiment harvested visual-cortex tissue from 8 mice between 6 and 7 weeks old that had been housed in the dark for one week prior to the dissection. Four of these mice were exposed to light for 1 h prior to dissection. The samples were used to prepare libraries for tail-length profiling using PAL-seq (Eisen et al., 2020b) (Figure 1A).

Addition of tail-length standards of known length to these libraries allowed quantification of recovery and tail-length estimation for standards of different lengths. Modest depletion of long-tailed standards was observed for tail lengths ≳160 nt, with some libraries exhibiting worse depletion than others (Figure S1A). Despite these issues, tail lengths were called accurately

(Figure S1B), and biological replicates showed reasonable agreement between the eight mice ($R_S \geq 0.65$ for all samples without stimulation, Figure S1C).

The PAL-seq tag yields of these data allowed investigation of experience-induced gene expression changes in the visual cortex tissue. mRNAs with significant changes mostly increased expression, and many were known or suspected immediate-early genes (Figure 1B). When considering the tail lengths for all mRNAs from all samples, stimulation did not change the shape of the distribution and only subtly affected the mean tail length (weighted mean of all tail lengths of 87.8 and 89.8 for unstimulated and stimulated samples, respectively, Figure 1C), and only modestly affected the distribution of tail lengths when considering the mean length from mRNAs from each gene (average mean tail length of 85.2 and 89.5 nt for unstimulated and stimulated samples, respectively, Figure 1D). Despite the modest global changes, several mRNAs significantly increased tail lengths upon stimulation, with the average mean tail length increase across the four stimulated replicates exceeding 30 nt for one mRNA (*Egr3*) but more often increasing by ~20 nt (Figure 1E) for those mRNAs with significant tail-length changes.

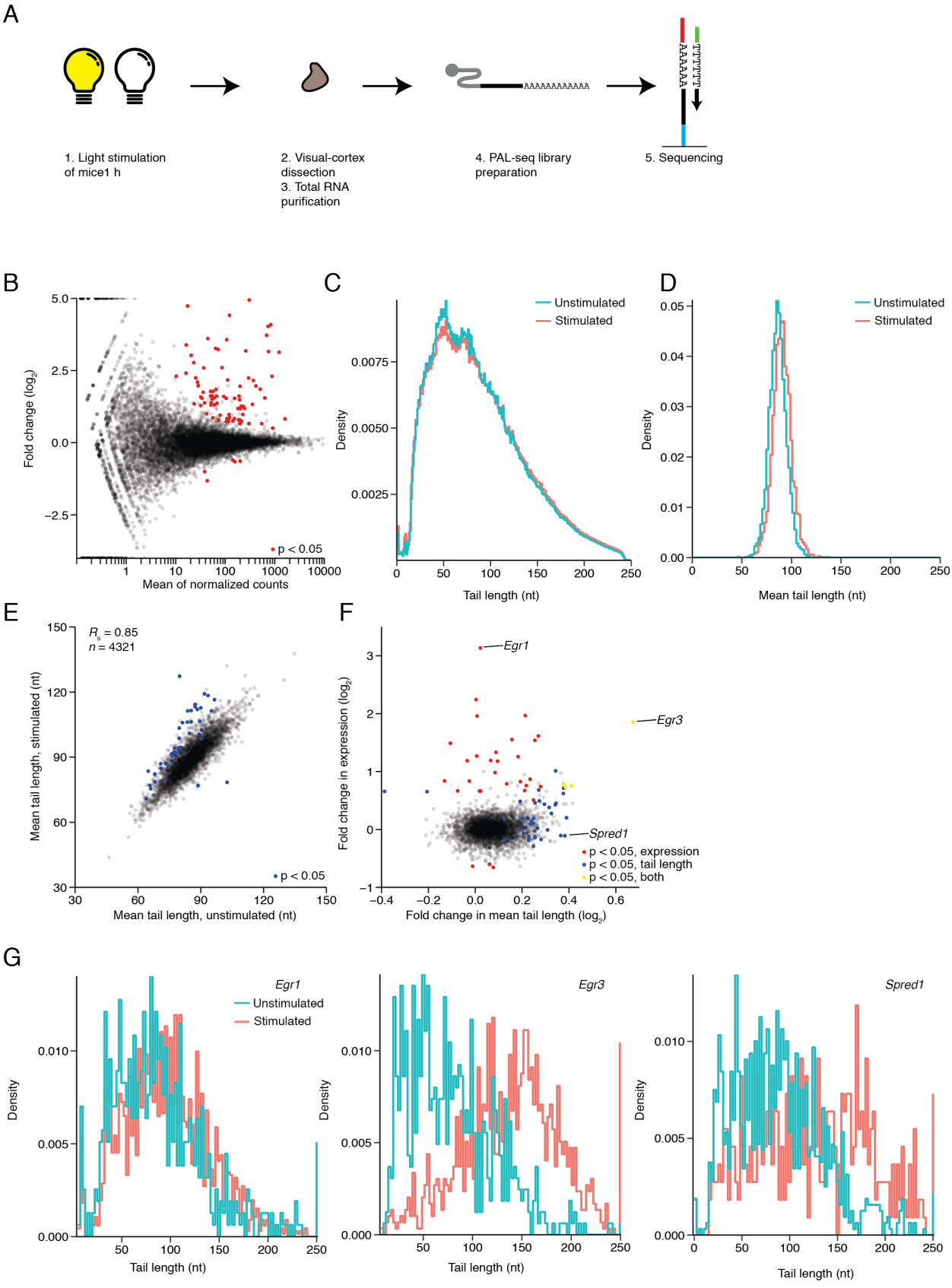

A
1. Light stimulation of mice1 h
2. Visual-cortex dissection
3. Total RNA purification
4. PAL-seq library preparation
5. Sequencing
AAAAAAAAAAAA
AAAAAA
TTTTTT
B
Fold change (log2)
Mean of normalized counts
p < 0.05
C
Unstimulated
Stimulated
Density
Tail length (nt)
D
Unstimulated
Stimulated
Density
Mean tail length (nt)
E
Rs = 0.85
n = 4321
Mean tail length, stimulated (nt)
Mean tail length, unstimulated (nt)
p < 0.05
F
Egr1
Egr3
Spred1
Fold change in expression (log2)
Fold change in mean tail length (log2)
p < 0.05, expression
p < 0.05, tail length
p < 0.05, both
G
Egr1
Egr3
Spred1
Unstimulated
Stimulated
Density
Tail length (nt)

Figure 1. Light Stimulation Changes Poly(A)-Tail Lengths in the Visual Cortex
(A) Schematic of light stimulation. Mice between 6 and 7 weeks old were housed in the dark for 1 week and the experimental cohort was exposed to light for 1 h prior to dissection and library preparation (Hrvatin et al., 2018). (B) Changes in gene expression between the light-stimulated and unstimulated cohorts. For each gene, the fold change in expression is plotted as a function of the mean number of normalized PAL-seq tags. mRNAs with significant changes in gene expression, as determined by DESeq (Anders and Huber, 2010), are plotted in red ($p < 0.05$, negative binomial test, $n = 4$ mice per cohort). (C) Distribution of tail lengths for all reads mapping to annotated genes, shown for the unstimulated and stimulated cohorts (key) n = 8,797,239 and 6,236,196 tags for the unstimulated and stimulated samples, respectively. (D) Mean tail lengths for each gene, plotted as a histogram. For each gene that passed a 50-read cutoff for each of the four samples from one cohort, the mean tail length from each sample was calculated and then the four means were averaged. n = 4,552 and 4,635 genes for the unstimulated and stimulated distributions, respectively. (E) Comparison of mean tail lengths for mRNAs from each gene in the unstimulated and stimulated cohorts. mRNAs that have four mean tail-length values that differ significantly in the two cohorts are shown in blue ($p < 0.05$, t test, $n = 4$ values in each cohort). (F) Comparison of stimulation-induced fold changes in tail length and expression. Colors and significance testing as in (B) and (E) (key). (G) Histograms of tail-length distributions for mRNAs labeled in (F). Distributions were normalized to all have unit area.

Tail-length increases for mRNAs from individual genes could stem from predominantly two mechanisms. One possibility is that, because stimulation increases expression of many mRNAs (Spiegel et al., 2014)(Figure 1B), some of these mRNAs may have failed to achieve a new steady-state when the tissue was harvested, 1 h after light stimulation. In this scenario, mean tail length increases would stem predominantly from the nascent population of mRNAs that have just been transcribed, which generally exhibit longer tails (Eisen et al., 2020b). An alternative scenario is that mature mRNAs become substrates for cytoplasmic polyadenylation upon stimulation (Wu et al., 1998). The former hypothesis predicts that changes in tail length should accompany changes in expression, while the latter does not require changes in expression, although expression changes may still accompany tail-length changes for a different population of mRNAs from the same gene. To investigate these relationships, we plotted the fold-change in expression compared with the fold-change in mean tail length (Figure 1F). This analysis revealed that mRNAs tend to behave in different ways, with some mRNAs (n = 4, such as *Egr3*)

exhibiting both large increases in tail length and expression, and others exhibiting only increases in expression with no significant tail-length change (n = 34, such as *Egr1*). This second cohort might have achieved a new steady-state tail length at a higher expression level. Still other mRNAs (n = 49, such as *Spred1*) exhibited changes in tail length with little accompanying expression change. Investigation of the individual tail-length distributions of these mRNAs supported these observations (Figure 1G). Presence or absence of a CPE motif (Fox and Wickens, 1990) (analysis as in (Subtelny et al., 2014)) did not provide more insight into tail length increases (Figure S2).

Primary-Culture Neurons Exhibit Tail-Length Changes in Response to Stimulants

The primary visual cortex, like all brain tissues, is composed of many cells of distinct lineages, including many non-neuronal cells (Hrvatin et al., 2018). Because of this, our previous conclusions regarding the relationships between tail length and expression are not specific to neurons, and indeed are most likely an integration of many signals. To assess tail-length changes in neurons, we measured tail lengths from primary cultured neurons dissected from the cortex of early postnatal animals, treated with cytosine arabinoside (ara-C, a potent poison of diving cells), and grown for 14 days in vitro (DIV14). These cells were then stimulated with either BDNF (Ghosh et al., 1994) or glutamate and glycine to mimic long-term potentiation (LTP) induction (Lu et al., 2001) (Figure 2A). Stimulation was performed for 30 min prior to harvest. Both treatments resulted in induction of immediate-early genes, but the BDNF stimulation provided more robust and larger induction (Figure 2B–C, left panels).

Tail-length changes sometimes accompanied induction (Figure 2B–C, middle panels), in keeping with analysis of the light-induction experiment (Figure 1). Inter-sample correlations

improved in the culture experiment, however, perhaps because these samples are mixtures of cells from many animals (usually ~15 pups per culture set, ~5 pups per sample), rather than a single animal, or because the culture and treatment methods mitigate some of the biological diversity among live mice. The samples treated with either glutamate and glycine or BDNF exhibited many mRNAs with concordance between tail-length changes and expression changes upon stimulation, indicating that these mRNAs are captured prior to their return to steady-state levels and a nascent population is providing the tail-length increases. Some mRNAs, however, like the light-stimulation system, exhibited changes in expression without changes in tail length (e.g. *Btg2* in the glutamate and glycine condition), while others exhibited changes in tail length without changes in expression (e.g. *Pnrc1* in both conditions). Despite the three classes of mRNA tail-length behaviors observed in response to stimulation in both the tissue and primary culture experiments, many of the mRNAs that increased either tail length or expression or both did not do so consistently across the two types of experiments, indicating differences in the nature of the signals in the two systems and confirming the notion that in vivo gene expression can differ markedly from in vitro models.

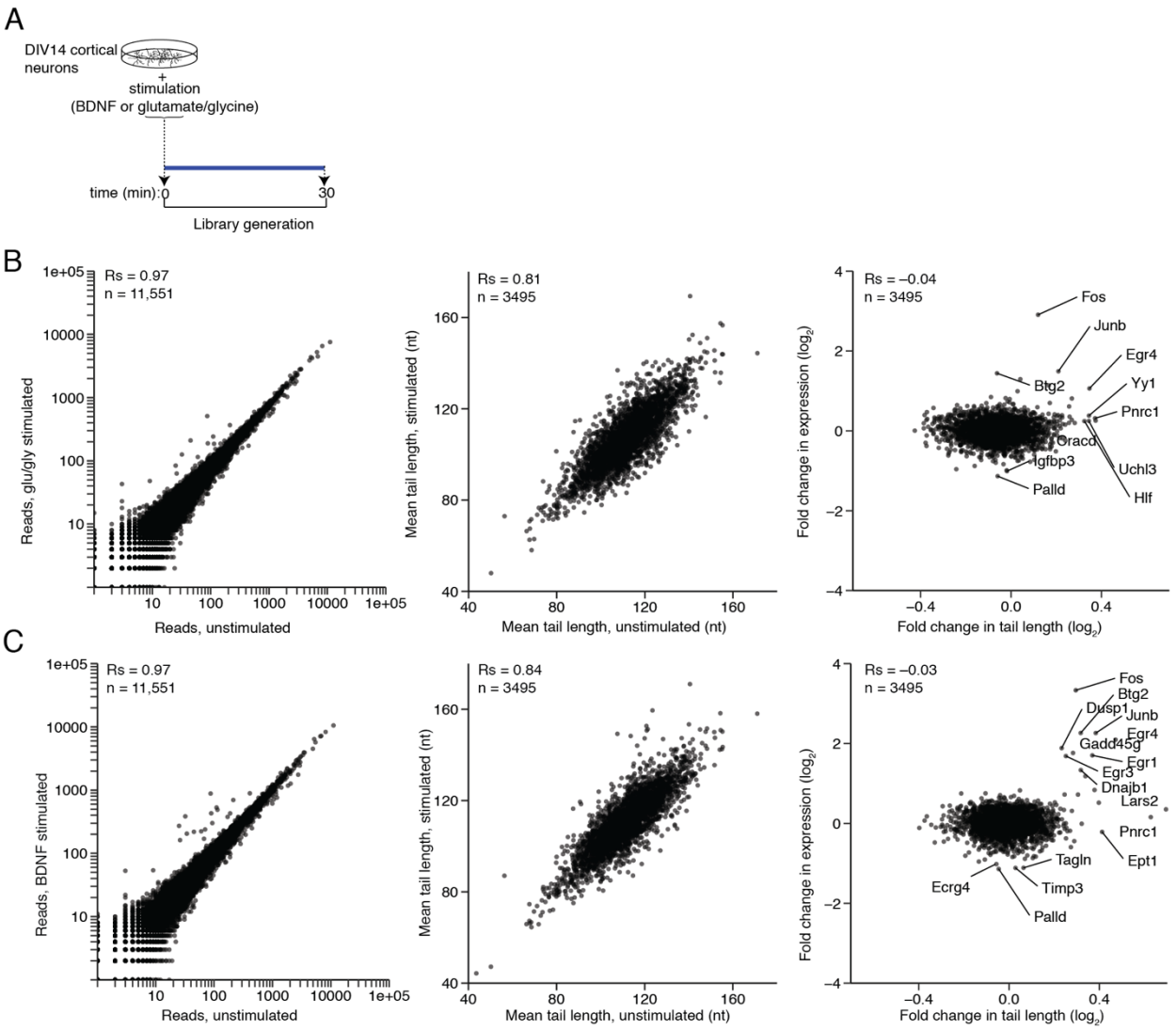


Figure 2. Tail Length and Expression Changes During Stimulation for Primary Cortical Cultures

(A) Schematic of two culture experiments. Primary mouse cortical cultures were harvested at DIV14 after stimulating for 30 minutes with either glutamate and glycine or BDNF. (B) Expression, tail length, and the relationship between them during glu/gly treatment. (Left) Correlation between the number of PAL-seq tags after 30 min of glu/gly stimulation and the number of tags in the untreated control. (Middle) Correlation between the mean tail-length changes after 30 min of glu/gly treatment and the untreated control for the subset of mRNAs in the left panel that passed a 50-read expression cutoff. (Right) Relationship between the change in expression and the change in tail length, normalized to the untreated control. Labels (RefSeq gene symbols) were added for all points with a $\log_2$ fold change in expression of > 1.3 or < –0.1 or in tail length of > 0.3 using the ggrepel package (Slowikowski, 2019). (C) As in (B), except plotting the data for the sample stimulated for 30 min with BDNF, using the matched DIV 14 culture as a control. Labels were added for all points exceeding a $\log_2$ fold change in expression of > 1.3 or < –0.1 or in tail length of > 0.4.

An Extensive Range of Translational Efficiencies

mRNA poly(A)-tails serve different functions in different contexts. In the embryonic systems of the frog and fish, and in the oocyte (and activated egg) of the fly, tail length is coupled to translational efficiency, with mRNAs with longer tails translated better than mRNAs with shorter ones (Subtelny et al., 2014; Eichhorn et al., 2016). Such systems are marked by little if any mRNA decay. As development continues, a new paradigm is established in which translation is no longer influenced by the length of the tail (Subtelny et al., 2014). This uncoupled state is a feature of all cell lines and adult tissues studied so far, including the mouse samples of liver tissue and 3T3 cells. The length of the poly(A)-tail serves a different role in these uncoupled states: the tail is a platform to chart deadenylation rate, which determines the rate of decay of the mRNA (Eisen et al., 2020b). Neurons exhibit mRNA decay, and thus are unlike any coupled system previously investigated. However, they also exhibit cytoplasmic polyadenylation, a feature of coupled systems, with increased translational efficiency a suggested outcome of tail lengthening for several mRNAs (Wu et al., 1998; Du and Richter, 2005; Dziembowska et al., 2012; Udagawa et al., 2012). By measuring both tail-length and translational efficiency from the same samples, we were able to investigate the extent of coupling between tail length and translation in neuronal cells and tissues.

We performed ribosome profiling and matched fragmented RNA-seq of the primary cortical culture samples, in addition to a primary hippocampal tissue sample, a primary cortical tissue sample, and an embryonic hippocampal tissue sample. We also performed ribosome profiling on three additional primary cortical cultures used for studies of microRNA effects. The ribosome profiling data from these samples showed strong enrichment for coding sequence

(CDS, Figure S3A), as well as 3-nt periodicity across the CDS (Figure S3B), as expected (Ingolia et al., 2009; Guo et al., 2010). Comparison of our datasets with nine datasets from six studies of neurons from the literature (Cho et al., 2015; Hornstein et al., 2016; Zappulo et al., 2017; Liu et al., 2018; Das Sharma et al., 2019; Gonatopoulos-Pournatzis et al., 2020) showed reasonable agreement, although a substantial range in correlation depending on the sample and study ($R_s = 0.16$ to $R_s = 0.71$ for samples from this study compared with the literature, Figure S3C). A similar range in agreement was observed when only considering datasets from the literature ($R_s = –0.07$ to $R_s = 0.57$, Figure S3C).

Ribosome profiling of the primary cortical culture samples supported the notion that translational efficiency is regulated in neuronal contexts. The measurements ranged over 300-fold, and these values were reproducible across multiple samples (Figure 3A and Figure S3C). Despite this large range, the stimulus-induced changes to translational efficiency were modest and dwarfed by the differences between mRNAs from different genes (Figure 3A). Hierarchical clustering of translational efficiencies from all ribosome-profiling datasets in this study showed strong relationships between samples from primary cortical culture, but substantial differences between these samples and brain tissues (Figure 3B).

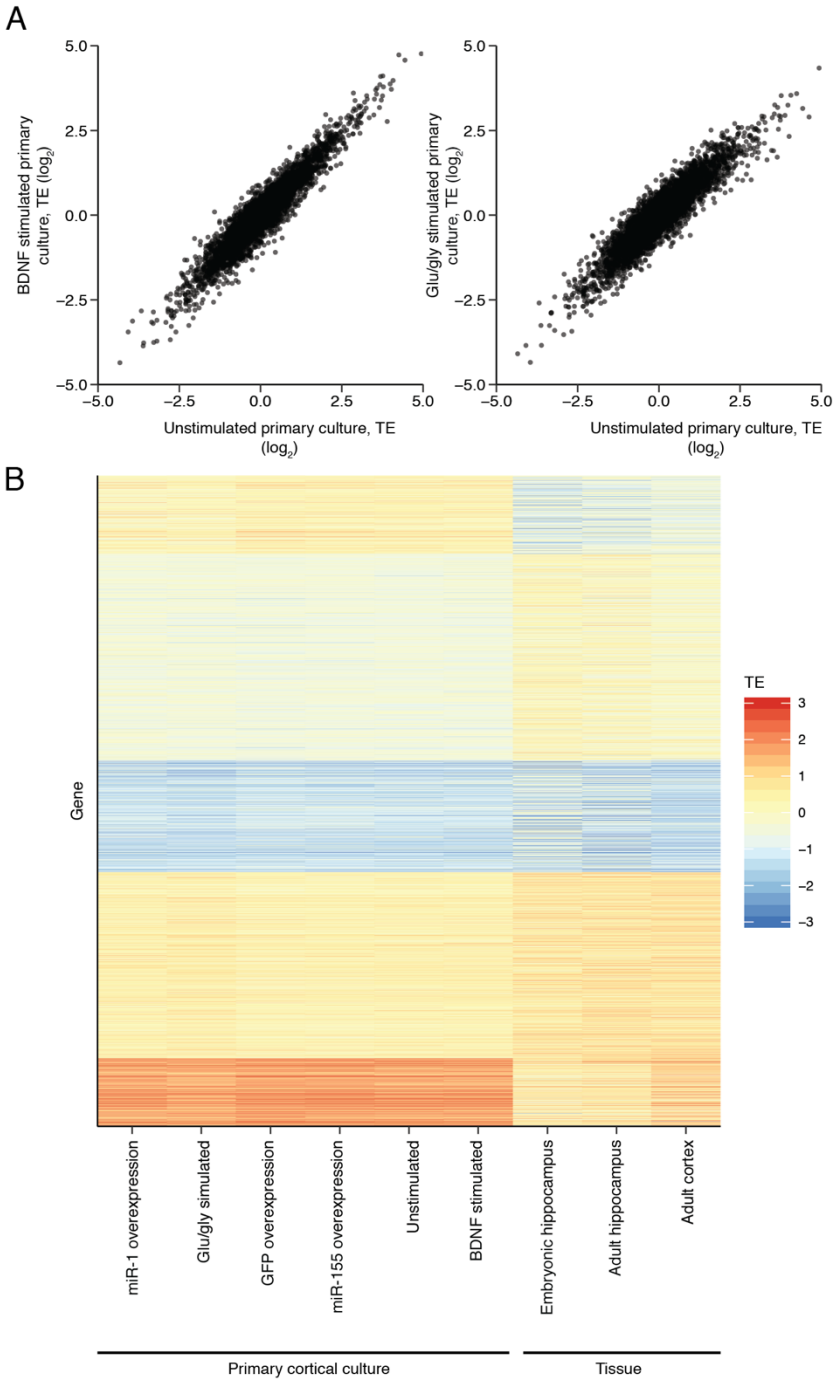


Figure 3. Translational Efficiencies in Neuronal Cultures and Tissues
(A) Correlation between translational efficiency (TE) in stimulated and unstimulated cortical cultures for the BDNF (left), and glutamate and glycine (right) samples. TE was calculated as the number of ribosome-protected fragments normalized to fragmented RNA-seq reads for ORFs of mRNAs from all genes that exceeded an RNA-seq expression threshold of 10 RPM. In each sample, TE values (log_2) were median centered by subtracting the median TE value of that sample from each value (Eichhorn et al., 2016). (B) Heatmap showing the relationship between TEs for

all mRNAs measured in each of the 9 samples in this study. For each sample, TEs were z-score normalized and then clustered using k-means into 5 clusters. Gene index is plotted as a function of sample and colored by TE, thresholding any TE values < –3 or > 3. mRNAs from n = 2,835 genes passed cutoffs in all samples.

A Modest Coupling Between Translational Efficiency and Poly(A)-Tail Length

While the dynamic range of the translational efficiency measurements proved expansive, even comparable to the translational efficiencies ranges in the early embryonic systems of the frog and fish, the ranges in mean tail length were much narrower than previously investigated coupled systems (Subtelny et al., 2014). The relationships between poly(A)-tail lengths and translational efficiencies for mRNAs from the six primary cortical culture samples had a modest relationship with translational efficiencies (R_s = 0.19–0.24, Figure 4A), indicating mean poly(A)-tail length in these cultures explained a small amount of the variance in translational efficiencies. The tissue samples, by contrast, had lower and more variable correlations overall (R_s = 0.06–0.19, Figure 4B), perhaps due to the heterogenous nature of these samples.

A coupling between tail length and TE that exists exclusively in neuronal cells might result from features of the neuronal cytoplasm that allow this coupling, such as a regime of limiting PABP concentration, for example. Alternatively, mRNAs expressed in neurons might contain sequences that allow translational regulation by tail length. In the latter scenario, considering only transcripts that are exclusively expressed in neurons, rather than ubiquitously expressed or exclusively expressed in non-neuronal cells, might enhance the relationship between tail length and translational efficiency. To investigate this possibility, we used an RNA-seq dataset of sorted populations of neurons and non-neuronal cells to subset our data (Zhang et al., 2014). Considering only mRNAs with neuronal-biased expression failed to increase the correlation between tail length and TE for most samples, suggesting that, as a class, neuronal-

biased mRNAs do not exhibit more coupling. (Figure S4). Subsetting the data based on a study that determined mRNAs highly enriched in the hippocampal synaptic neuropil, a region enriched for dendrites, axons and glia but depleted of neuronal somata, did not improve the correlations either (Cajigas et al., 2012).

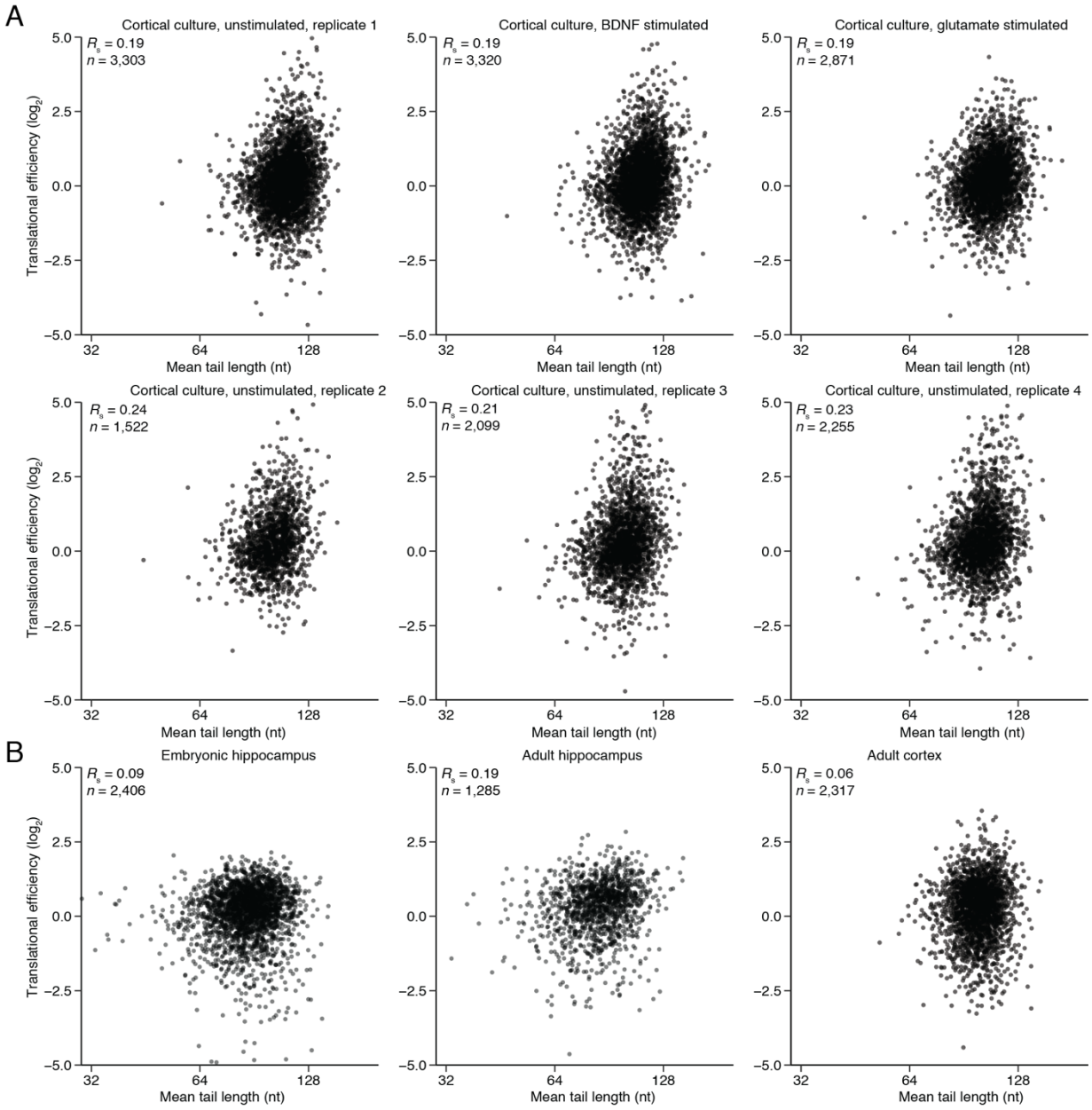


Figure 4. Tail Length and Translational Efficiency in Primary Cortical Cultures
(A) Correlation between TE and mean tail length for mRNAs from each gene that passed cutoffs for both measurements. Plotted are all correlations for six primary cultures used in this study. (B) Correlation between TE and mean tail for tissue samples. Otherwise as in (A).

MicroRNAs Decrease mRNA Target Abundance

MicroRNAs (miRNAs), 21–24 nt guide RNAs, direct recruitment of the effector protein Argonaute (AGO) to repress target mRNAs (Bartel, 2018). They predominantly act by destabilizing these targets in most cell lines and tissues (Guo et al., 2010; Eichhorn et al., 2014). In neurons, studies have detailed the role of miRNAs but have reached different conclusions regarding mechanism. When knocked out, miR-128, a highly expressed miRNA, induces phenotypes related to neuronal overexcitability and epilepsy, altering mRNA levels in the striatum (Tan et al., 2013). In the retina, miRNAs of the miR-183/96/182 cluster induce changes to mRNA levels (Krol et al., 2010). At the synapse, however, miRNAs are reported to mediate translational repression with relatively little change in mRNA levels, with the proposal that the transience of this mode of regulation might allow more rapid expression of mRNA targets (Banerjee et al., 2009; Schratt, 2009; Sambandan et al., 2017).

We established a lentiviral-based gene-delivery system to investigate miRNA activity in neuronal cultures. Primary cortical cultures were infected with a lentivirus expressing one of two exogenous miRNAs, miR-1 or miR-155, from the 3′ UTR of *GFP* (Figure 5A). Almost all neurons were fluorescent upon infection (Figure S5A), although perhaps the molecules contributing to this fluorescence failed to generate miRNAs, as the processing steps required for generating a miRNA would cleave a portion of the *GFP* 3′ UTR. The cultures also expressed the exogenous miRNAs, however (Figure S5B). Neither miRNA reached the levels of miR-124, a highly-expressed and endogenous miRNA in neurons, indicating that the exogenous miRNAs

were not expressed at levels exceeding physiological ones. RNA-sequencing of samples from the cultures showed that mRNA targets were repressed and that site-type hierarchy, with some target-sites more efficacious than others (Bartel, 2009), was maintained (Figure S5C–D).

Translational repression due to miRNAs was not observed in primary neuronal cultures. When investigating all predicted targets of either miR-1 or miR-155, most of the changes to ribosome-protected fragment levels could be attributed to changes in RNA abundance, indicating no additional change to ribosome-protected fragment levels beyond those captured by RNA-seq (Figure 5B, left and middle bars). The same observation held true when considering only the top-predicted targets, defined as mRNAs repressed by either miR-1 or miR-155 in a separate study (Eichhorn et al., 2014), despite the increased repression observed for these targets (Figure 5C, left and middle bars). These results support previous conclusions regarding miRNA mechanism when examined for neurons or neuronal tissues without subcellular fractionation.

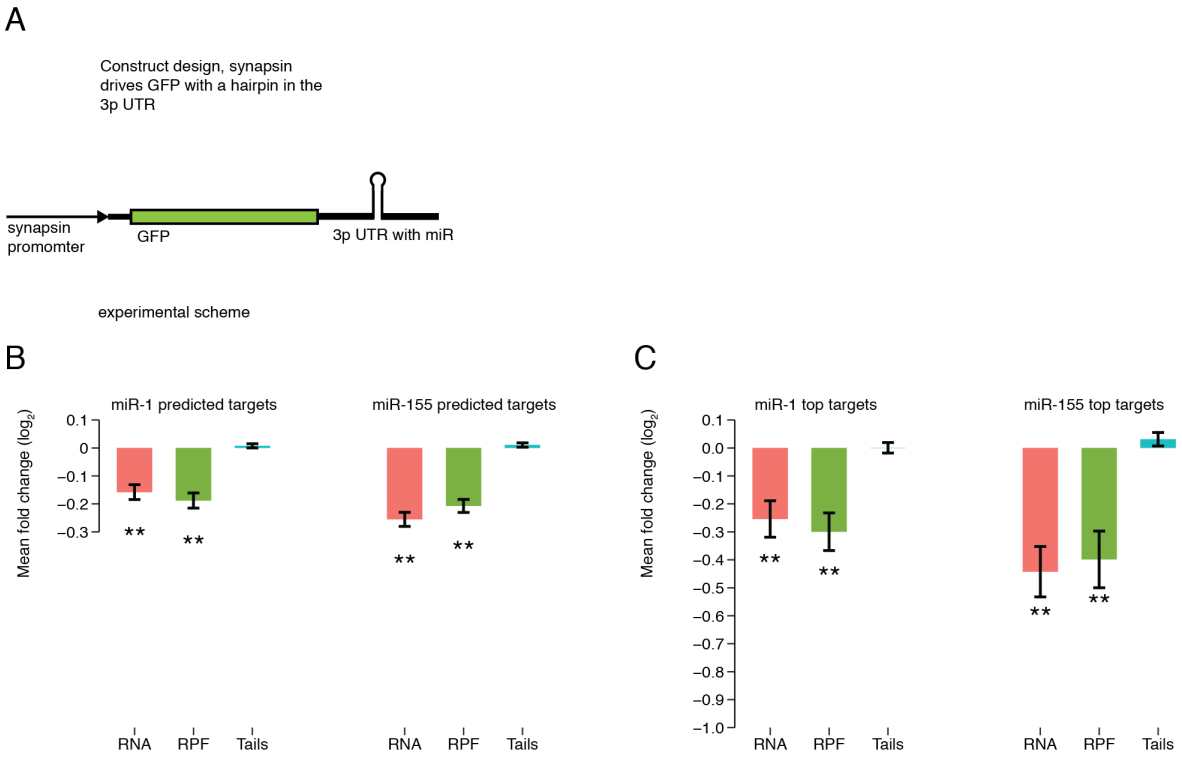


Figure 5. MicroRNA-Mediated Repression in Cultured Neurons from Mouse Cortex
(A) Schematic of the miRNA-transduction construct. Pre-miR-1 or pre-miR-155, along with 100 nt of flanking sequence, were cloned as the 3′ UTR of GFP. These constructs were packaged into lentivirus in HEK293FT cells, and purified virus was used to transduce primary culture neurons at DIV5, and neurons were harvested at DIV14. (B) Impact of miR-1 (left) and miR-155 (right) on RNA levels, ribosome protected fragments, and mean tail lengths for predicted targets of miR-1 ($n = 270$ and 203 for miR-1 and miR-155, respectively). Values were normalized to those of mRNAs with no site to the transduced miRNA ($n = 745$ and 1017 for no-site cohorts for miR-1 and miR-155, respectively). Significant changes are indicated with asterisks below the bar (*$p < 0.05$; **$p < 0.001$, one-tailed t test). (C) Analysis as in (B), but considering only top predicted targets ($n = 39$ and 17 for miR-1 and miR-155, respectively). These values were normalized to the same cohort of no-site mRNAs in (B).

Like translational repression, the ability of miRNAs to elicit changes in tail length depends on context. In the early zebrafish embryo, miRNAs predominantly induce translational repression of mRNA targets (Bazzini et al., 2012). This difference in mechanism relates to the predominant mechanism of regulatory control: in the embryonic system, unlike in cell lines or mature tissues (Guo et al., 2010; Eichhorn et al., 2014), translational control, mediated by tail length changes, rather than transcriptional control, takes precedent, as little mRNA decay occurs during this developmental stage. As development continues, mean tail lengths for mRNAs exhibit increasingly narrower ranges of values, and the miRNA-mediated tail-shortening that previously caused translational repression gives way to mRNA destabilization (Subtelny et al., 2014).

Mature cell types and tissues exhibit miRNA-mediated changes to RNA levels but nearly imperceptible changes to steady-state tail lengths (Chang et al., 2014; Eisen et al., 2020a). Although they accelerate deadenylation, miRNAs also accelerate short-tailed decay of their mRNA targets, allowing an equivalent distribution of mRNA tail lengths with or without the miRNA. Without a shift in the distribution of tail lengths, miRNA-mediated repression does not result in a change to steady-state tail length (Eisen et al., 2020a). This explanation is consistent

with our results in primary neurons, in which neither miR-1 nor miR-155 altered tail lengths of their respective target mRNAs (Figure 5B–C, right bars). Indeed, the neuronal system, unlike the early zebrafish embryo, exhibits mRNA decay.

A Linear Model of Translational Efficiency

The expansive range of translational efficiency in neurons and neuronal tissues, combined with the datasets measuring tail length, prompted us to determine features of mRNAs that contribute to translation in neurons. A previous study investigated features in human embryonic-stem-cell derived neuronal cultures using polysome profiling, and concluded that CDS length, 3′ UTR length, and 5′ UTR length influenced translation most consistently across the samples tested, with CDS length correlating positively with translation and UTR length correlating negatively (Blair et al., 2017). Our analysis integrated these features and others to build a predictive model of translational efficiency from ribosome profiling data. To do so, we chose two samples to compare: an unstimulated primary cortical culture and an adult cortical tissue sample, reasoning that these two samples represent distinct classes (Figure 3B). The model construction was similar to another study investigating the determinants of mRNA stability (Spies et al., 2013).

The most important features for predicting translational efficiency in adult cortex and cortical culture were CDS length and expression, respectively (Figure 6A–B). While observed previously, the correlation with CDS length is stronger in the adult cortex than other systems described (Blair et al., 2017), including the culture system, where it had very little effect (Figure 6D). The final model for adult cortex yielded an R^2 of 0.31 (Figure 6C). In cultured neurons, the model was less explanatory, with an overall R^2 of 0.23 (Figure 6D). Notable features in the

culture system included the GC content of the CDS ($R^2 = 0.06$) and tail length ($R^2 = 0.03$), the latter of which was poorly predictive in the cortex.

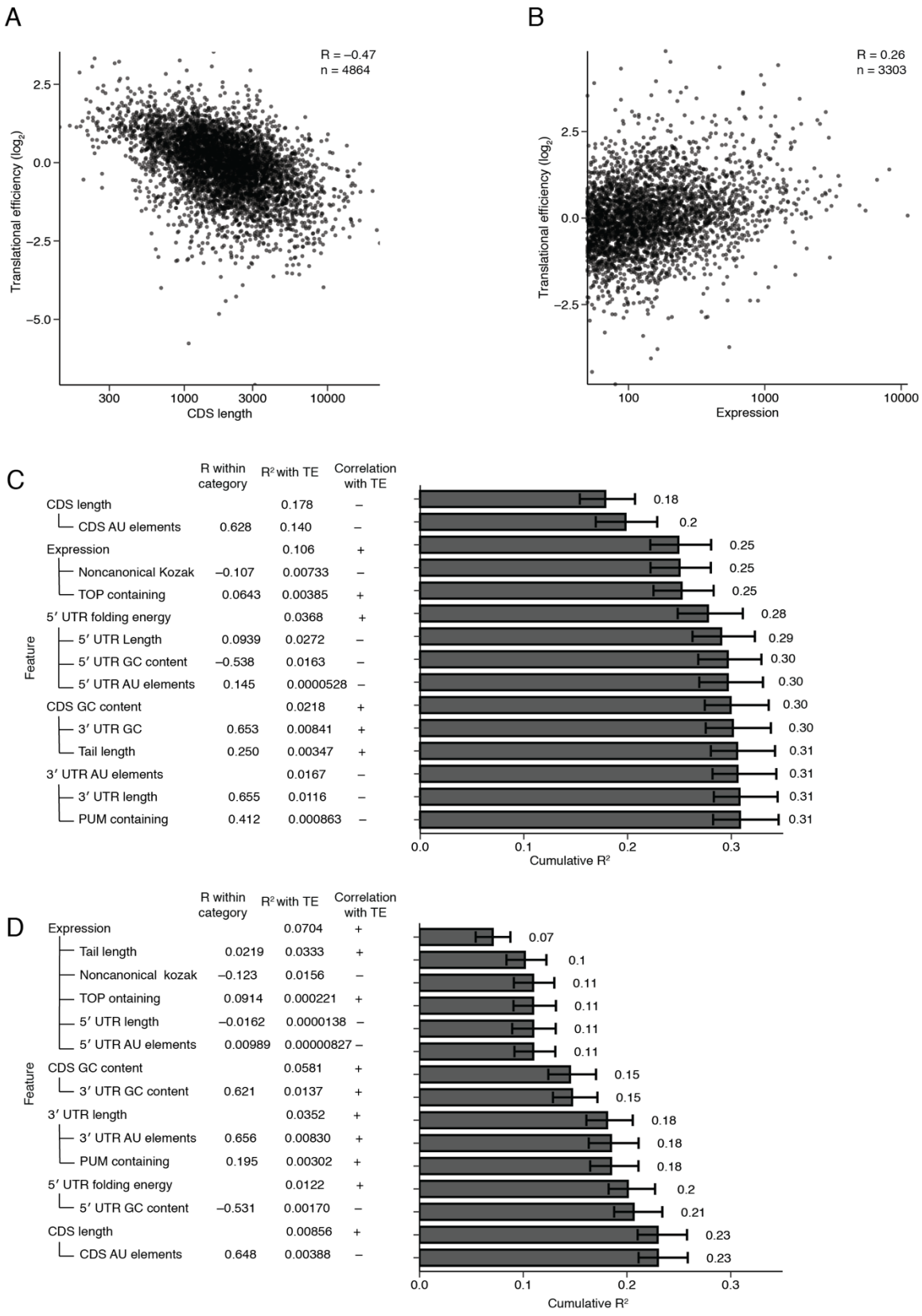

A
R = –0.47
n = 4864
Translational efficiency (log2)
2.5
0.0
–2.5
–5.0
300
1000
3000
10000
CDS length
B
R = 0.26
n = 3303
Translational efficiency (log2)
2.5
0.0
–2.5
100
1000
10000
Expression
C
R within category
R² with TE
Correlation with TE
CDS length 0.178 –
CDS AU elements 0.628 0.140 –
Expression 0.106 +
Noncanonical Kozak –0.107 0.00733 –
TOP containing 0.0643 0.00385 +
5′ UTR folding energy 0.0368 +
5′ UTR Length 0.0939 0.0272 –
5′ UTR GC content –0.538 0.0163 –
5′ UTR AU elements 0.145 0.0000528 –
CDS GC content 0.0218 +
3′ UTR GC 0.653 0.00841 +
Tail length 0.250 0.00347 +
3′ UTR AU elements 0.0167 –
3′ UTR length 0.655 0.0116 –
PUM containing 0.412 0.000863 –
Feature
0.18
0.2
0.25
0.25
0.25
0.28
0.29
0.30
0.30
0.30
0.30
0.31
0.31
0.31
0.31
0.0
0.1
0.2
0.3
Cumulative R²
D
R within category
R² with TE
Correlation with TE
Expression 0.0704 +
Tail length 0.0219 0.0333 +
Noncanonical kozak –0.123 0.0156 –
TOP ontaining 0.0914 0.000221 +
5′ UTR length –0.0162 0.0000138 –
5′ UTR AU elements 0.00989 0.00000827 –
CDS GC content 0.0581 +
3′ UTR GC content 0.621 0.0137 +
3′ UTR length 0.0352 +
3′ UTR AU elements 0.656 0.00830 +
PUM containing 0.195 0.00302 +
5′ UTR folding energy 0.0122 +
5′ UTR GC content –0.531 0.00170 –
CDS length 0.00856 +
CDS AU elements 0.648 0.00388 –
Feature
0.07
0.1
0.11
0.11
0.11
0.11
0.15
0.15
0.18
0.18
0.18
0.2
0.21
0.23
0.23
0.0
0.1
0.2
0.3
Cumulative R²

Figure 6. A Linear Model Quantifying Features of Translation in Neurons
(A) Correlation between the cortical tissue TE and its largest explanatory feature, coding-sequence length. (B) Correlation between the DIV14 cortical culture control TE and its largest explanatory feature, expression level. Expression levels are read counts from PAL-seq tags. (C) Sequence features used to predict TE in adult cortex (n = 2,210 mRNAs). These features are arranged in approximate order of importance in the linear model. Features are grouped into cohorts that have significant collinearity in order to avoid overinterpretation. The Pearson R relationship between features within a category and the most-predictive feature is shown. The relationship between each individual feature is also shown, along with the sign of the relationship and its effect on the cumulative model of TE. Error bars denote a 95% confidence interval on each feature after bootstrapping the data 1000 times using the *boot* package in R. Note that the R^2 values differ slightly from the R values in panel A because the model uses only the subset of mRNAs for which tail length was measured. (D) Analysis as in (C), except showing the results for a model trained on data from primary cortical cultures instead of cortical tissue (n = 3,145 mRNAs).

Inspired by reports of regulation of translation elongation in neurons (Darnell et al., 2011; Das Sharma et al., 2019), we investigated whether CDS composition might be an important feature of translational efficiency. The mRNAs expressed in cortical culture neurons tended to have codons that clustered largely into two categories, with two codons either mutually exclusive or found together (Figure S6A). When comparing the frequency of each codon in a particular coding sequence with the translational efficiency of that transcript, some codons tended to be associated with better translation, and some with worse, with all six primary culture samples showing similar trends (Figure 7A). The codon associated with the second-worst translation, AGA encoding arginine, was previously observed to induce ribosome pausing the C57BL6J mouse strain (the strain used in this study) due to a single nucleotide polymorphism in the *n-Tr20* tRNA, an isodecoder for arginine (Ishimura et al., 2014). The observation that this codon is associated with poor translation is paradoxical because pausing would be expected to increase RPFs for these codons, but perhaps mRNAs that are translated well are under pressure to use this codon less. A second codon correlated with poor translation, AAA encoding lysine, is associated with ribosome 'sliding' and reduced translation in *E. coli* and *S. cerevisiae* (Koutmou et al.,

2015), and sometimes causes stalling (Kostova et al., 2017). This is in contrast to another codon, AAG also encoding lysine, which was not translated poorly in that study, consistent with the effect of these two codons on translational efficiency here. Three other codons poorly translated in our analysis, TTA, ATA, and CTA, were previously shown to be rarely used in HEK293T cells and generally excluded from the ribosomal-protein genes, a well-translated cohort in that system (Saikia et al., 2016).

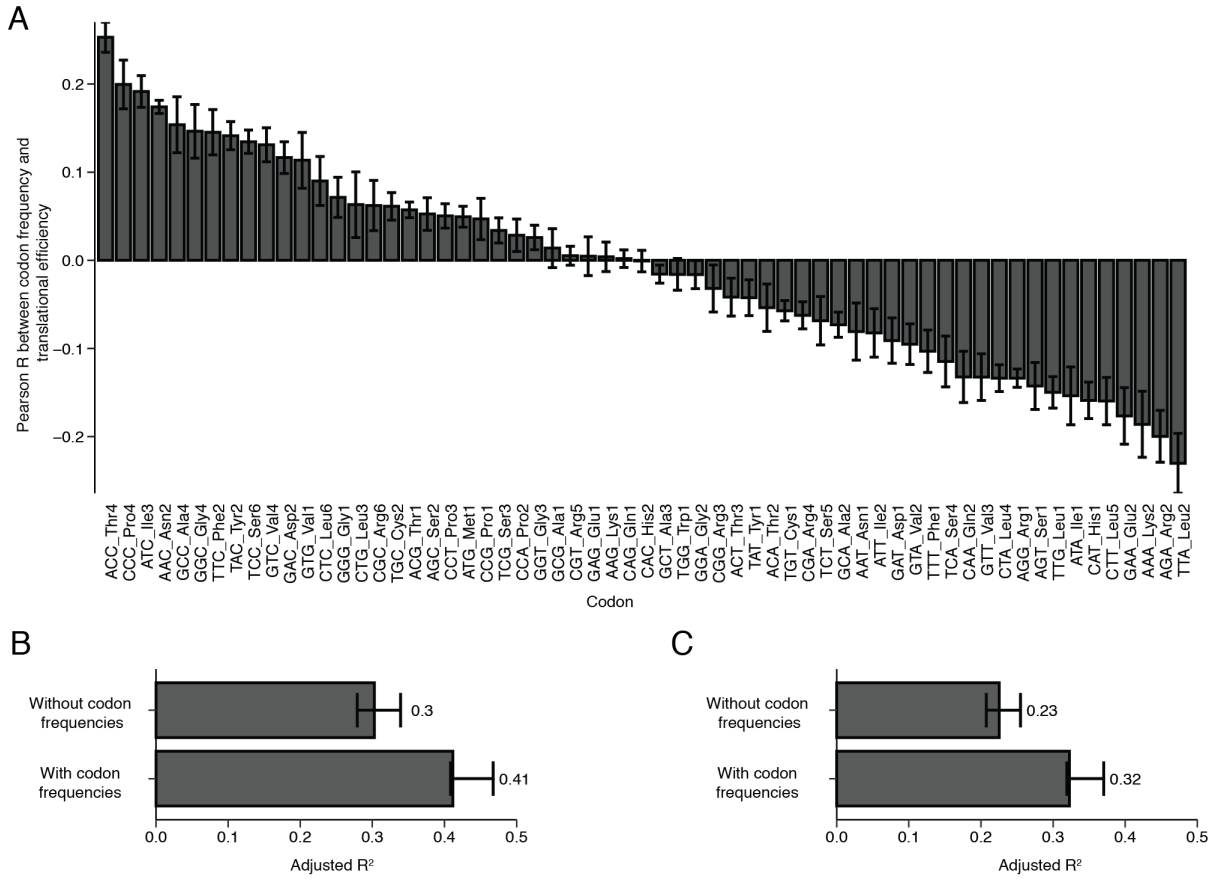


Figure 7. Codon Composition Increases the Predictions of a Linear Model of TE
(A) Some codons correlate with TE. The relationship between codon frequency and TE is plotted for each of the 61 codons. Error bars represent standard deviation for the R values for all 6 primary culture samples used in this study. (B) Codon composition increases the predictive efficacy of a linear model of TE in adult cortex. The full linear model in Figure 6C was re-trained with the 61 codon frequencies, adjusting the R^2 to account for the number of model features. Bar plots show this model along with a model that lacks the codon frequency parameters. Error bars as in Figure 6C. (C) Codon composition increases the predictive efficacy of a linear model of TE in primary cortical culture. Analysis as in (B).

Lys and arg codons code for positively charged amino acids, which can have electrostatic interactions with the ribosome exit tunnel, influencing elongation rate (Lu and Deutsch, 2008). To examine whether we could observe this influence, we re-analyzing our CDS composition, considering amino-acids instead of codons (Figure S6B). In this analysis, Lys and Arg were associated with decreased translation, but the magnitude of these correlation was lower than when considering codons, perhaps because some Lys and Arg codons (e.g. CGC) are favorably associated with translation. Furthermore, glutamate also tends to be associated with worse translational efficiency, despite its negative charge.

Encouraged by these modest but reproducible relationships, we underwent a second round of model building in which we incorporated codon composition into the linear framework. Because each codon was represented as its own feature, we used an adjusted R^2 to offset the increased R^2 that would be achieved by chance by incorporating the 61 additional features. Even with this adjustment, addition of the codon composition features significantly increased the predictive power of both the primary culture and adult cortex models (Figure 7B–C), in keeping with hypotheses concerning regulation of translation elongation rate in neurons.

Discussion

One goal of this study was to build a predictive framework for translational control in mammalian neurons. Our R^2 value, and the features that we have collected, indicate that the majority of the variance in translational regulation between mRNAs from different genes has yet to be explained. Some of this variation will be attributable to experimental noise, but the reproducibility of our measurements, and their general agreement with similar studies from the

literature, suggest that we are missing major features of mRNAs that control their translation. One such feature is structure. In this study, GC content and computational folding with RNAfold are used as a proxy for structural accessibility, but finer accounting for structure—and finer positioning of structures at defined positions with respect to start codons, stop codons or transcript ends, will increase the power of our predictions.

With respect to tail length, our study supports the notion that the length of the poly(A)-tail influences translation in neurons. Indeed, the complete lack of correlation (or slightly negative correlation) observed between translational efficiency and tail length in the mouse and human samples previous profiled (Subtelny et al., 2014) implies that the relationship observed here is unique among tissues of the adult mouse. Despite this signal, it is modest, explaining less than 6% of the variance in translation for the most correlated samples. A modest relationship suggests several interpretations.

One possibility is that the coupling that we observe in neuronal cultures might be strengthened with the ability to capture and correctly quantify long-tailed mRNAs. This would expand the dynamic range of tail length measurements beyond those described here, allowing analysis of mRNA populations spanning tail-length ranges not previously observed. The median tail lengths in our samples are already some of the longest reported in the literature, and perhaps use of long-read sequencing or approaches that do not require reverse transcription and PCR, such as Oxford Nanopore Sequencing, might allow capture of longer ones.

A second possibility is that there exists a subcellular context in which this relationship might be improved. The initial studies that characterized this relationship for *Camk2a* and a handful of other mRNAs did so while arguing for a dendrite-specific role for this mechanism (Wu et al., 1998; Du and Richter, 2005; Dziembowska et al., 2012; Udagawa et al., 2012). Many

aspects of this hypothesis are intriguing because the different concentrations of decay machinery, PABP, and translation machinery present in the dendrites leave open the possibility of a fundamentally different regulatory regime. Perhaps mRNA decay might be substantially slower in the distal regions of the neuron, as a parallel to the coupled embryonic systems. Indeed, calculations based on dendritic volume suggest that concentrations of most macromolecules differ between the dendrite and soma (Kosik, 2016). Tail-length mediated control of translation could serve to tune translational output on rapid timescales, while mRNA transcription, processing, transport to the dendrites, and local translation might take too long.

Experimental validation of the local tail-length regulation hypothesis requires biochemical purification of the dendrites. So far, many studies of RNAs in synapses rely on synaptosome or synaptoneurosome preparations. These are biochemical purifications of the synaptic bouton, which can respond to stimulation even after purification and are widely used to investigate questions that microscopy cannot yet answer about the biology of the dendrite. The enrichments with these preparations are modest when assessed using western blotting for PSD95 or other postsynaptic proteins (sometimes with 2–4 fold) because these proteins are expressed highly in the soma as well. Electron-microscopy–based examination of synaptosomes indicates the purified samples are complex and variable (Nagy and Delgado-Escueta, 1984; Chicurel et al., 1993; Bagni et al., 2000; Villasana et al., 2006; Westmark et al., 2011). Using RNAs as markers for enrichment is more challenging still, as a bona fide synapse-specific RNA has yet to be identified. Many mRNAs studied, while enriched in the dendrite when compared to other mRNAs that could be in the dendrite, have the majority of their molecules in the cytoplasm. Because of these challenges, the modest coupling between tail length and translation observed in this study is unlikely to strengthen with standard biochemical methods to enrich neuronal

processes. Further investigation of the local regulatory regime awaits techniques that can measure these elusive signals.

Acknowledgements

We thank Sinisa Hrvatin and Michael Greenberg for a kind gift of samples from the mouse visual cortex. We also thank Valeriya Baru for guidance on neuronal culture preparation, the Whitehead Genome Technology Core for high-throughput sequencing, and Sean McGeary and members of the Bartel lab for helpful discussions. This research was supported by NIH grant GM118135 (D.P.B.) and an NSF Graduate Research Fellowship (T.J.E.). D.P.B. is an investigator of the Howard Hughes Medical Institute.

Declaration of Interests

The authors declare no competing interests.

Methods

Cell lines and cell culture for transformed lines

HEK293FT cells (ThermoFisher) were used for packaging lentivirus. These cells were grown at 37°C in 5% CO_2 in DMEM supplemented with 10% FBS (Takara). HEK293 cells are female with a complex karyotype. No mycoplasma testing was performed.

Methods Details

Mouse Husbandry

Mice were group-housed (fewer than 5 mice per cage, co-housing females or sibling males) in a 12 hr light/dark cycle (light between 07:00 and 19:00) in a temperature-controlled room (21.1 ± 1.1°C) at the Whitehead Institute for Biomedical Research with food and water ad libitum. Mice were maintained according to protocols approved by the Massachusetts Institute of Technology Committee on Animal Care. Euthanasia was performed by CO_2 inhalation except for neonates, where rapid decapitation was used. Mouse ages are indicated in the figure legends. Sex was not determined for embryos or neonatal pups. For analyses of older animals (non-neonatal), only male mice were used.

Primary Cortical Culture

Primary neurons were cultured as described (Beaudoin et al., 2012) with slight modifications. The six samples used in this study were prepared in two batches of three samples from two litters per batch. Cortices were dissected from male and female P0–P1 mouse pups in ice-cold dissection medium (Hanks Balanced Salt Solution, ThermoFisher, supplemented with 1 mM sodium pyruvate, 0.1% glucose, and 10 mM HEPES, pH 7.3) and washed twice with ice-cold

dissection medium. Samples were trypsinized by incubating half of the cortices in each of two 15 mL conical tubes containing 5 mL dissection medium and 0.25% trypsin for 20 min at 37°C. DNase (0.1% final) was added and the sample was incubated for 2–5 min to prevent clumping. Tissue was washed again in room-temperature dissection medium at which point the medium exchanged for 10 mL of plating medium (Basal Medium Eagle, ThermoFisher, supplemented with 20% FBS, 0.45% glucose, 1 mM sodium pyruvate, 2 mM GlutaMAX, ThermoFisher, and 100 U/mL penicillin/streptomycin), taking care not to allow the tissue to become exposed to air. The tissue was then triturated in two 15 mL conical tubes with an FBS-coated 5 mL serological pipet in 3 rounds, performing no more than 10 triturations per round and transferring the dissociated cells to a new 15 mL conical tube between each round. Dissociated cells were passed through a 70 μm cell strainer, counted, and plated in plating medium at ~100,000 cells/cm^2 on 15 cm tissue culture plates coated with poly-D-lysine (50 μg/mL). Cells were only cultured if viability, assessed by Trypan Blue staining, exceeded 80% upon plating. Plating medium was exchanged for maintenance medium (Neurobasal, ThermoFisher, supplemented with 1X B27 Serum-Free supplement, and 2 mM GlutaMAX), between 2–6 h after plating. Two days after plating, cells were treated with 5 μM AraC for 24 h and then maintenance media was fully exchanged. Neurons were fed using half-media exchanges every 3 days thereafter. All neurons in this study were harvested at DIV 14 (with the day of plating considered DIV 1).

Lentiviral Production and Transduction

The day prior to transfection, ~18,000 293FT cells/cm^2 were plated onto 15 cm dishes, with each dish preparing a different virus. The following day a transfection mix was prepared with 9 μg of shuttle vector, 9 μg of pCMV-dR8.91 packaging plasmid (a gift from Jonathan Weissman), and 1

μg of the envelope plasmid pMD2.G (a gift from Didier Trono; Addgene #12259) in 1 mL of Optimem (ThermoFisher) and 125 μL of FuGene HD (Promega). The mix was incubated for 15 minutes at room temperature, then brought up to volume with 20 mL of Optimem containing 10% FBS and used to refeed the cells. The following day the cells were refed with 30 mL of DMEM containing 10% FBS and grown for two more days, at which point the medium was collected and cleared of cell debris by filtering through a 0.45 μM PES syringe filter, using two filters per supernatant to prevent clogging. The ~30 mL of viral supernatant was concentrated by addition of 10 mL of Lenti-X Concentrator (Takara), overnight incubation at 4°C, and centrifugation at 1,500g for 45 min in a swinging-bucket rotor in a 50 mL conical tube. After discarding the supernatant, the pellet was resuspended in 1 mL of ice-cold PBS, aliquoted into 100 μL aliquots, flash frozen, and stored at –80°C. Neurons were transduced by infecting with 800 μL of virus at DIV 5 (4/5th of the total virus) in a 15 cm plate.

Small-RNA Blots

Small-RNA blots were performed as described (Kleaveland et al., 2018). Briefly, total RNA (1 μg) from miRNA-transduced neuronal cultures was denatured with 2X formamide loading dye (ThermoFisher) and resolved on a 15% urea-polyacrylamide gel. RNA was then transferred onto an Amersham Hybond-NX neutral nylon membrane (GE Healthcare) using a semi-dry transfer apparatus (BioRad). Membranes were crosslinked with 0.466 g EDC (N-(3-dimethylaminopropyl)-N0-ethylcarbodiimide; ThermoFisher) dissolved in 15 mL of 125 mM 1-methylimidazole for 1 h at 60°C and then pre-incubated with ULTRAhyb Oligo Buffer (ThermoFisher) for 30 minutes at 42°C with rotation. 5′ radiolabeled probe was then added and the membrane was hybridized overnight with rotation at 42°C. The next morning, membranes

were rinsed twice with low-stringency buffer (2X SSC and 0.1% SDS), then incubated for 30 min under rotation at 42°C, rinsed twice with high-stringency buffer (0.1X SSC and 0.1% SDS), and incubated for 30 min under rotation at 42°C. The blots were then exposed to a phosphorimaging screen for several hours or overnight. Signal was detected using the Typhoon FLA 7000 phosphorimager (GE Healthcare Life Sciences).

In order to hybridize additional probes, membranes were stripped of the original probe. To do so, they were incubated four times with 100 mL of boiling stripping buffer (0.1% SDS, 0.1x SSC) for 10 minutes each time with vigorous shaking. Membranes were rinsed with water to remove residual SDS, exposed to a phosphorimager screen to check the residual signal, and then pre-hybridized and hybridized as above.

Linear Model

Multiple linear regression was performed as described previously by using the core R packages (Spies et al., 2013). Prior to modeling, variables were log transformed if indicated, then Z-score normalized. To determine features that were highly correlated with one another, all pairwise correlations were clustered using the core R command *hclust* into five clusters. For the Kozak sequence feature, the canonical sequence was defined as GCCACC in the last 6 nt of the 5′ UTR and the hamming distance between this sequence and the observed sequence was used as a linear variate. 5′ UTR folding energy was determined by using RNAfold to calculate the minimum free energy structure for the first 70 nt of the 5′ UTR (Lorenz et al., 2011). The last 70 nt were used because of a previous report showing that the folding energy of this region had the strongest relationship with translational efficiency (Weinberg et al., 2016). 3′ UTR lengths were

determined from PAL-seq annotations using the eight light-stimulation visual cortex PAL-seq datasets.

Light Stimulation of Mouse Visual Cortex

Total RNA samples from the visual cortex of light-stimulated mice were a kind gift from Sinisa Hrvatin and Michael Greenberg and used to prepare sequencing libraries for a separate study (Hrvatin et al., 2018). Briefly, mice were housed under standard light cycles (6:00–18:00) until aged between 6 and 8 weeks, at which time they were placed in complete darkness for 7 d. Four mice (the stimulated cohort) were then exposed to light for 1 h prior to isoflurane anesthesia. Eyes were then enucleated and animals were transcardially perfused with 25 mM sodium bicarbonate per liter, 12 mM glucose, 1.25 mM monobasic sodium phosphate, 7.5 mM magnesium chloride hexahydrate, 2.5 mM potassium chloride, 10 mM HEPES, 110 mM choline chloride, 13 mM ascorbic acid, and 3.9 mM g pyruvic acid. This perfusion buffer contained the following additives, designed to inhibit residual stimulation or transcription: 1 μM TTX (Sigma), 100 μM AP-V (ThermoFisher), 5 μg/mL actinomycin D (Sigma) and 10 μM triptolide (Sigma). Following perfusion, visual cortex V1 was microdissected. RNA was isolated using the RNeasy Mini (Qiagen) procedure according to the manufacturer's instructions, unlike the other RNA in this study, which was isolated using TRIzol (ThermoFisher).

Tissue Preparation

Hippocampi from pups from embryonic day 16 were dissected individually into Eppendorf tubes and flash frozen. The adult hippocampus from a male mouse was dissected and flash frozen.

Adult cortex from a male mouse was flash frozen. Samples were stored at -80°C until lysate preparation.

To prepare lysate, frozen tissue was transferred to a 2 mL Dounce homogenizer filled with ribosome-profiling lysis buffer (10 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 100 mM KCl, 1 % Triton-X 100, 2 mM DTT, 100 µg/mL cycloheximide, 500 U/mL RNasein Plus (Promega), and cOmplete mini EDTA-free protease inhibitor tablets (Roche, 1 tablet/10 mL)) (Subtelny et al., 2014). Samples were then homogenized with 10 strokes of pestle A followed by 10 strokes of pestle B, taking care not to introduce bubbles into the buffer. Following homogenization, the sample was transferred to two Eppendorf tubes, centrifuged at 1,300g for 10 min, and the supernatant was aliquoted and flash frozen at -80°C for use in ribosome profiling, RNA sequencing, and tail sequencing.

PAL-Seq v2

PAL-seq v2 was performed for the eight samples dissected from the visual cortex and the adult and embryonic hippocampal tissue samples (10 samples in total). Library preparation was as described previously (Eisen et al., 2020b), and summarized as follows. 20–30 µg of total RNA was used to prepare PAL-seq libraries. Tail-length standard mixes (1 ng of set 1 and 2 ng of set 2), and trace 5′-radiolabeled marker RNAs were added to each sample to assess tail-length measurements and ligation outcomes, respectively. Polyadenylated ends, including those with a terminal uridine, were ligated to a 3′-biotinylated adaptor DNA oligonucleotide (1.8 µM) in the presence of two splint DNA oligonucleotides (1.25 µM and 0.25 µM for the U and A-containing splint oligos, respectively) using T4 Rnl2 (NEB) in an overnight reaction at 18°C. Following 3′-adaptor ligation the RNA was extracted with phenol–chloroform (pH 8.0), precipitated,

resuspended in 1X RNA T1 sequence buffer (ThemoFisher), heated to 50°C for 5 min and then put on ice. RNase T1 was then added to a final concentration of 0.006 U/μL, and the reaction was incubated at room temperature for 30 min, followed by phenol–chloroform extraction and RNA precipitation. Precipitated RNA was captured on streptavidin beads, 5′ phosphorylated, and ligated to a 5′ adaptor as described (Subtelny et al., 2014) but using a modified 5′ adaptor sequence. Following reverse transcription using SuperScript III (Invitrogen) with a barcode-containing DNA primer, cDNA was purified as described (Subtelny et al., 2014), except a 160–810 nt size range was selected. Libraries were amplified by PCR for 8 cycles using Titanium Taq (Takara) polymerase according to the manufacturer's protocol with a 1.5 min combined annealing/extension step at 57°C. PCR-amplified libraries were purified using AMPure beads (Agencourt, 40 μL beads per 50 μL PCR, two rounds of purification) according to the manufacturer's instructions.

PAL-seq v2 libraries were sequenced on an Illumina HiSeq 2500 operating in rapid mode. Hybridization mixes were prepared with 0.375 fmol PCR-amplified library that had been denatured with standard NaOH treatment and brought to a final volume of 125 μL with HT1 hybridization buffer (Illumina, 3 pM library in final mix). Following standard cluster generation and sequencing-primer hybridization, two dark cycles were performed for the splinted-ligation libraries (i.e., two rounds of standard sequencing-by-synthesis in which imaging was skipped), which extended the sequencing primer by 2 nt, thereby enabling measurement of poly(A) tails terminating in non-adenosine bases. For the direct-ligation libraries, six dark cycles were performed instead of two, which extended the sequencing primer past the four random-sequence nucleotides in the 3′ adaptor and the last two residues of the tail.

Following the two dark cycles, a custom primer-extension reaction was performed on the sequencer using 50 μM dTTP as the only nucleoside triphosphate in the reaction. To perform this extension, the flow cell temperature was first set to 20°C. Then, 120 μL of universal sequencing buffer (USB, Illumina) was flowed over each lane, followed by 150 μL of Klenow buffer (NEB buffer 2 supplemented with 0.02% Tween-20). Reaction mix (Klenow buffer, 50 μM dTTP, and 0.1 U/μL Large Klenow Fragment, NEB) was then flowed on in two aliquots (150 μL and 100 μL). The flow-cell temperature was then increased to 37°C at a rate of 8.5°C per min and the incubation continued another 2 min after reaching 37°C. 150 μL of fresh reaction mix was then flowed in, and following a 2 min incubation, 75 μL of reaction mix was flowed in eight times, with each flow followed by a 2 min incubation. The reaction was stopped by decreasing the flow cell temperature to 20°C, flowing in 150 μL of quench buffer (Illumina HT2 buffer supplemented with 10 mM EDTA) and then washing with 75 μL of HT2 buffer. The flow cell was prepared for subsequent sequencing with a 150 μL and a 75 μL flow of HT1 buffer (Illumina). 50 cycles of standard sequencing-by-synthesis were then performed to yield the first sequencing read (read 1). XML files to configure a HiSeq 2500 for this protocol are provided at https://github.com/kslin/PAL-seq.

The flow cell was stripped, a barcode sequencing primer was annealed, and seven cycles of standard sequencing-by-synthesis were performed to read the barcode. The flow cell was then stripped again, and the same primer as used for read 1 was hybridized and used to prime 250 cycles of standard sequencing-by-synthesis to generate read 2. Thus, each PAL-seq tag consisted of three reads: read 1, read 2, and the indexing (barcode) read. For cases in which a tag corresponded to a polyadenylated mRNA, read 1 was the reverse complement of the 3′ end of the mRNA immediately 5′ of the poly(A) tail and was used to identify the mRNA and cleavage-and-

polyadenylation site of long-tailed mRNAs. The indexing read was used to identify the sample, and read 2 was used to measure poly(A)-tail length and identify the mRNA and cleavage-and-polyadenylation site of short-tailed mRNAs. The intensity files of reads 1 and 2 were used for poly(A)-tail length determination, along with the Illumina fastq files.

TAIL-Seq

The primary cortical culture samples and the adult cortex sample (7 samples in total) were prepared using TAIL-seq (Chang et al., 2014). The library preparation for this protocol was the same as for PAL-seq v2, but the resulting cDNA was amplified using different primers and sequenced using a different protocol. The first read of TAIL-seq involved sequencing the 3′ UTR from the gene body toward the tail, with the sequencing primer annealing to sequences added with the 5′ adaptor. This 5′ adaptor was an equimolar mixture of four sequences with different numbers of nucleotides in between the primer binding site and the insert to ensure that highly abundant sequences (such as rRNA fragments) did not cause a large portion of the flow cell to fluoresce in a single channel. Amplification and purification were as for PAL-seq v2. Samples were sequenced with a paired-end 50-by-250 run using a HiSeq 2500 operating in normal mode using a v3 kit. Analysis was as described for PAL-seq v2, except a five-state GHMM was used (Chang et al., 2014) to accommodate the difference in the nature of the T-signal output imparted by the different mode of sequencing. The five states were an initiation state, a poly(A) state, a poly(A)-transition state, a non-poly(A) transition state, and a non-poly(A) state.

PAL-Seq v2 Data Analysis

Tail lengths for the splinted-ligation data were determined using a Gaussian hidden Markov model (GHMM) from the python2.7 package ghmm (http://ghmm.org/), analogous to the model used in TAIL-seq (Chang et al., 2014) and described in the next paragraph. Read 1 was mapped using STAR (v2.5.4b) run with the parameters '–alignIntronMax 1 –outFilterMultimapNmax 1 –outFilterMismatchNoverLmax 0.04 –outFilterIntronMotifs RemoveNoncanonicalUnannotated –outSJfilterReads', aligning to an index of the mouse genome built using mm10 transcript annotations that had been compressed to unique instances of each gene selecting the longest transcript and removing all overlapping transcripts on the same strand (Eichhorn et al., 2014). The genome index also included sequences of the quantification spikes and the common portion of the poly(A)-tail length standards. The sequences that identified each RNA standard (the last 20 nt of each standard sequence) were not aligned using STAR. Instead, the unix program grep (v2.16) was used to determine which reads matched each standard (allowing no mismatches), and these reads were added to the aligned reads from the STAR output. Tags corresponding to annotated 3′ UTRs of mRNAs were identified using bedtools (v2.26.0), and if the poly(A)-tail read (read 2) contained a stretch of ≥ 10 T residues (the reverse complement of the tail) in an 11-nt window within the first 30 nt, this read was carried forward for GHMM analysis. If read 2 failed to satisfy this criterion but began with ≥ 4 T residues, the tail length was called based on the number of contiguous T residues at the start of read 2; by definition, these tails were < 10 nt and thus easily determined by direct sequencing.

For each read 2 that was to be input into the GHMM a 'T signal' was first calculated by normalizing the intensity of each channel for each cycle to the average intensity of that channel when reading that base in read 1 and then dividing the thymidine channel by the sum of the other

three channels. Sometimes a position in a read would have a value of 0 for all four channels. A read was discarded if it contained more than five such positions. Otherwise, the values for these positions were imputed using the mean of the five non-zero signal values upstream and downstream (ten positions total) of the zero-valued position. A three-state GHMM was then used to decode the sequence of states that occurred in read 2. It consisted of an initiation state (state 1), a poly(A)-tail state (state 2), and a non-poly(A)-tail state (state 3). All reads start in state 1. From state 1 the model can remain in state 1 or transition to state 2. From state 2 the model can either remain in state 2 or transition to state 3. The model was initialized with the following transition probabilities:

$from \setminus to$	$state_1$	$state_2$	$state_3$
$state_1$	0.001	0.95	0.049
$state_2$	0.001	0.95	0.049
$state_3$	0.001	0.001	0.998

The initial emissions were Gaussian distributions with means of 100, 1, and −1 and variances of 1, 0.25 and 0.25, respectively. The emission Gaussians for the model corresponded to the logarithm of the calculated T signal at each sequenced base in read 2. The initial state probabilities were 0.998, 0.001, and 0.001 for states 1, 2 and 3, respectively.

After initializing the model, unsupervised training was performed on 10,000 randomly selected PAL-seq tags, and then the trained model was used to decode all tags, with the number of state 2 cycles reporting the poly(A)-tail length for a tag. Only genes with ≥ 50 poly(A)-tail length measurements were considered for analyses involving mean poly(A)-tail lengths.

Documentation and code to calculate and analyze T signals and determine tail lengths are available for both the TAIL-seq and PAL-seq pipelines, for the splinted and single-stranded ligation, at https://github.com/kslin/PAL-seq.

Annotation of 3′ Ends

3′-end annotations were generated from PAL-seq tags from the eight light-stimulation datasets. All data from these samples were combined and mRNAs with tails ≥ 11 nt were used for annotation, using an algorithm previously developed for data from poly(A)-position profiling by sequencing (3P-seq) (Jan et al., 2011). Each PAL-seq read 1 that mapped (with at least 1 nt of overlap) to an annotated 3′ UTR (Eichhorn et al., 2014) was compiled by the genomic coordinate of its 3′-UTR nucleotide closest to the tail. The genomic coordinate with the most mapped reads was annotated as a 3′ end. All reads within 10 nt of this end (a 21 nt window) were assigned to this end and removed from subsequent consideration. This process was repeated until there were no remaining 3′ UTR-mapped reads. For each gene, the 3′-end annotations were used in subsequent analyses if they accounted for ≥ 10% of the 3′ UTR-mapping reads for that gene.

RNA-Seq and Ribosome Profiling

Total RNA was either poly(A)-selected (as for the adult hippocampus and embryonic hippocampus samples) or Ribozero-depleted (all six primary cortical culture samples and the adult cortex sample). For ribosome profiling libraries, 300–600 μL aliquots of lysate were digested with 0.3 U/μL RNase I (Ambion) for 30 min at room temperature and then run on a 10–50% sucrose gradient to purify monosomes (Subtelny et al., 2014). RNAs from both RNA-seq and ribosome profiling were then size-selected, ligated to adapters, reverse-transcribed, and amplified (Subtelny et al., 2014). These libraries were sequenced on an Illumina HiSeq 2500. For all RNA-seq and ribosome profiling data, only reads mapping to ORFs of annotated gene models

(Eichhorn et al., 2014) were considered, excluding the first 50 nt of each ORF. A cutoff of ≥ 10 reads per million mapped reads (RPM) was applied to each gene in each RNA-seq sample.

Calculation of miRNA-Mediated Repression

Secondary effects of expressing a miRNA can have a greater impact on mRNAs with longer 3′ UTRs relative to those with shorter 3′ UTRs (Agarwal et al., 2015), presumably because longer 3′ UTRs tend to contain more sites to other regulatory factors, including other miRNAs. As a result, 3′-UTR length differences can complicate the measurement of the repressive effects of an expressed miRNA. For this reason, we first normalized the fold-changes of all mRNAs based on their 3′ UTR length as in (Eisen et al., 2020a). The relationship between the fold-change for all mRNAs without a 6-nt seed-matched site to the induced miRNA in the entire transcript (no-site mRNAs) and 3′-UTR length was calculated using linear regression, and then the fold-changes of all mRNAs (with and without a target site) were normalized by their 3′ UTR lengths such that the slope of the relationship between no-site mRNAs and 3′ UTR length was 0. We then compared normalized fold-changes for mRNAs containing at least one predicted miRNA target site in their 3′-UTR to those for the no-site mRNAs. For all mRNAs passing our expression threshold in the GFP-overexpression sample, we calculated the $\log_2$ fold-changes in mRNA abundance, RPF abundance, or poly(A)-tail length in samples from neurons transduced with either miR-155 or miR-1 compared to neurons transduced with GFP. The repressive effect of the miRNA on a set of predicted miRNA targets was then calculated by subtracting the median-normalized fold-change for no-site mRNAs from the mean-normalized fold-change for a set of predicted targets. Top targets were defined using RPF measurements from a previous study (Eichhorn et al., 2014), choosing from among the predicted targets those with expression that decreased to ≤ 75% of their original expression after 12 hours of miRNA induction.

References

Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife *4*.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Bagni, C., Mannucci, L., Dotti, C.G., and Amaldi, F. (2000). Chemical stimulation of synaptosomes modulates alpha -Ca2+/calmodulin-dependent protein kinase II mRNA association to polysomes. J Neurosci *20*, RC76.

Banerjee, S., Neveu, P., and Kosik, K.S. (2009). A coordinated local translational control point at the synapse involving relief from silencing and MOV10 degradation. Neuron *64*, 871-884.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell *173*, 20-51.

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.

Beaudoin, G.M., 3rd, Lee, S.H., Singh, D., Yuan, Y., Ng, Y.G., Reichardt, L.F., and Arikkath, J. (2012). Culturing pyramidal neurons from the early postnatal mouse hippocampus and cortex. Nat Protoc *7*, 1741-1754.

Biever, A., Glock, C., Tushev, G., Ciirdaeva, E., Langer, J.D., and Schuman, E.M. (2019). Monosomes actively translate synaptic mRNAs in neuronal processes. bioRxiv.

Blair, J.D., Hockemeyer, D., Doudna, J.A., Bateup, H.S., and Floor, S.N. (2017). Widespread Translational Remodeling during Human Neuronal Differentiation. Cell Rep *21*, 2005-2016.

Cajigas, I.J., Tushev, G., Will, T.J., tom Dieck, S., Fuerst, N., and Schuman, E.M. (2012). The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. Neuron *74*, 453-466.

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.

Chicurel, M.E., Terrian, D.M., and Potter, H. (1993). mRNA at the synapse: analysis of a synaptosomal preparation enriched in hippocampal dendritic spines. J Neurosci *13*, 4054-4063.

Cho, J., Yu, N.K., Choi, J.H., Sim, S.E., Kang, S.J., Kwak, C., Lee, S.W., Kim, J.I., Choi, D.I., Kim, V.N.*, et al.* (2015). Multiple repressive mechanisms in the hippocampus during memory formation. Science *350*, 82-87.

Costa-Mattioli, M., Sossin, W.S., Klann, E., and Sonenberg, N. (2009). Translational control of long-lasting synaptic plasticity and memory. Neuron *61*, 10-26.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W.*, et al.* (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247-261.

Das Sharma, S., Metz, J.B., Li, H., Hobson, B.D., Hornstein, N., Sulzer, D., Tang, G., and Sims, P.A. (2019). Widespread Alterations in Translation Elongation in the Brain of Juvenile Fmr1 Knockout Mice. Cell Rep *26*, 3313-3322 e3315.

de Moor, C.H., and Richter, J.D. (1997). The Mos pathway regulates cytoplasmic polyadenylation in Xenopus oocytes. Mol Cell Biol *17*, 6419-6426.

Du, L., and Richter, J.D. (2005). Activity-dependent polyadenylation in neurons. RNA *11*, 1340-1347.

Dziembowska, M., Milek, J., Janusz, A., Rejmak, E., Romanowska, E., Gorkiewicz, T., Tiron, A., Bramham, C.R., and Kaczmarek, L. (2012). Activity-dependent local translation of matrix metalloproteinase-9. J Neurosci *32*, 14538-14547.

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*.

Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., and Bartel, D.P. (2020a). MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. Mol Cell *77*, 775-785 e778.

Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., Lin, K.S., McGeary, S.E., Gupta, S., and Bartel, D.P. (2020b). The Dynamics of Cytoplasmic mRNA Metabolism. Mol Cell *77*, 786-799 e710.

Fox, C.A., and Wickens, M. (1990). Poly(A) removal during oocyte maturation: a default reaction selectively prevented by specific sequences in the 3' UTR of certain maternal mRNAs. Genes Dev *4*, 2287-2298.

Ghosh, A., Carnahan, J., and Greenberg, M.E. (1994). Requirement for BDNF in activity-dependent survival of cortical neurons. Science *263*, 1618-1623.

Gonatopoulos-Pournatzis, T., Niibori, R., Salter, E.W., Weatheritt, R.J., Tsang, B., Farhangmehr, S., Liang, X., Braunschweig, U., Roth, J., Zhang, S.*, et al.* (2020). Autism-Misregulated eIF4G Microexons Control Synaptic Translation and Higher Order Cognitive Functions. Molecular Cell.

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.

Hornstein, N., Torres, D., Das Sharma, S., Tang, G., Canoll, P., and Sims, P.A. (2016). Ligation-free ribosome profiling of cell type-specific translation in the brain. Genome Biol *17*, 149.

Hrvatin, S., Hochbaum, D.R., Nagy, M.A., Cicconet, M., Robertson, K., Cheadle, L., Zilionis, R., Ratner, A., Borges-Monroy, R., Klein, A.M.*, et al.* (2018). Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. Nat Neurosci *21*, 120-129.

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.

Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). RNA function. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. Science *345*, 455-459.

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.

Kleaveland, B., Shi, C.Y., Stefano, J., and Bartel, D.P. (2018). A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. Cell *174*, 350-362 e317.

Kosik, K.S. (2016). Life at Low Copy Number: How Dendrites Manage with So Few mRNAs. Neuron *92*, 1168-1180.

Kostova, K.K., Hickey, K.L., Osuna, B.A., Hussmann, J.A., Frost, A., Weinberg, D.E., and Weissman, J.S. (2017). CAT-tailing as a fail-safe mechanism for efficient degradation of stalled nascent polypeptides. Science *357*, 414-417.

Koutmou, K.S., Schuller, A.P., Brunelle, J.L., Radhakrishnan, A., Djuranovic, S., and Green, R. (2015). Ribosomes slide on lysine-encoding homopolymeric A stretches. Elife *4*.

Krol, J., Busskamp, V., Markiewicz, I., Stadler, M.B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H.J., Schubeler, D.*, et al.* (2010). Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. Cell *141*, 618-631.

Li, J.J., Bickel, P.J., and Biggin, M.D. (2014). System wide analyses have underestimated protein abundances and the importance of transcription in mammals. PeerJ *2*, e270.

Liu, B., Li, Y., Stackpole, E.E., Novak, A., Gao, Y., Zhao, Y., Zhao, X., and Richter, J.D. (2018). Regulatory discrimination of mRNAs by FMRP controls mouse adult neural stem cell differentiation. Proc Natl Acad Sci U S A *115*, E11397-E11405.

Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol Biol *6*, 26.

Lu, J., and Deutsch, C. (2008). Electrostatics in the ribosomal tunnel modulate chain elongation rates. J Mol Biol *384*, 73-86.

Lu, W., Man, H., Ju, W., Trimble, W.S., MacDonald, J.F., and Wang, Y.T. (2001). Activation of synaptic NMDA receptors induces membrane insertion of new AMPA receptors and LTP in cultured hippocampal neurons. Neuron *29*, 243-254.

Majdan, M., and Shatz, C.J. (2006). Effects of visual experience on activity-dependent gene regulation in cortex. Nat Neurosci *9*, 650-659.

Mardinly, A.R., Spiegel, I., Patrizi, A., Centofante, E., Bazinet, J.E., Tzeng, C.P., Mandel-Brehm, C., Harmin, D.A., Adesnik, H., Fagiolini, M.*, et al.* (2016). Sensory experience regulates cortical inhibition by inducing IGF1 in VIP neurons. Nature *531*, 371-375.

Nagy, A., and Delgado-Escueta, A.V. (1984). Rapid preparation of synaptosomes from mammalian brain using nontoxic isoosmotic gradient material (Percoll). J Neurochem *43*, 1114-1123.

Parras, A., Anta, H., Santos-Galindo, M., Swarup, V., Elorza, A., Nieto-Gonzalez, J.L., Pico, S., Hernandez, I.H., Diaz-Hernandez, J.I., Belloc, E.*, et al.* (2018). Autism-like phenotype and risk gene mRNA deadenylation by CPEB4 mis-splicing. Nature *560*, 441-446.

Qin, M., Kang, J., Burlin, T.V., Jiang, C., and Smith, C.B. (2005). Postadolescent changes in regional cerebral protein synthesis: an in vivo study in the FMR1 null mouse. J Neurosci *25*, 5087-5095.

Richter, J.D. (1999). Cytoplasmic polyadenylation in development and beyond. Microbiol Mol Biol Rev *63*, 446-456.

Rodriguez, C.M., Chun, S.Y., Mills, R.E., and Todd, P.K. (2019). Translation of upstream open reading frames in a model of neuronal differentiation. BMC Genomics *20*, 391.

Saikia, M., Wang, X., Mao, Y., Wan, J., Pan, T., and Qian, S.B. (2016). Codon optimality controls differential mRNA translation during amino acid starvation. RNA *22*, 1719-1727.

Sambandan, S., Akbalik, G., Kochen, L., Rinne, J., Kahlstatt, J., Glock, C., Tushev, G., Alvarez-Castelao, B., Heckel, A., and Schuman, E.M. (2017). Activity-dependent spatially localized miRNA maturation in neuronal dendrites. Science *355*, 634-637.

Schratt, G. (2009). microRNAs at the synapse. Nat Rev Neurosci *10*, 842-849.

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.

Simbriger, K., Amorim, I.S., Chalkiadaki, K., Lach, G., Jafarnejad, S.M., Khoutorsky, A., and Gkogkas, C.G. (2020). Monitoring translation in synaptic fractions using a ribosome profiling strategy. J Neurosci Methods *329*, 108456.

Slowikowski, K. (2019). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.1.

Spiegel, I., Mardinly, A.R., Gabel, H.W., Bazinet, J.E., Couch, C.H., Tzeng, C.P., Harmin, D.A., and Greenberg, M.E. (2014). Npas4 regulates excitatory-inhibitory balance within neural circuits through cell-type-specific gene programs. Cell *157*, 1216-1229.

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.

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.

Tan, C.L., Plotkin, J.L., Veno, M.T., von Schimmelmann, M., Feinberg, P., Mann, S., Handler, A., Kjems, J., Surmeier, D.J., O'Carroll, D.*, et al.* (2013). MicroRNA-128 governs neuronal excitability and motor behavior in mice. Science *342*, 1254-1258.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature *485*, 109-113.

Udagawa, T., Swanger, S.A., Takeuchi, K., Kim, J.H., Nalavadi, V., Shin, J., Lorenz, L.J., Zukin, R.S., Bassell, G.J., and Richter, J.D. (2012). Bidirectional control of mRNA translation and synaptic plasticity by the cytoplasmic polyadenylation complex. Mol Cell *47*, 253-266.

Umegaki, Y., Brotons, A.M., Nakanishi, Y., Luo, Z., Zhang, H., Bonni, A., and Ikeuchi, Y. (2018). Palladin Is a Neuron-Specific Translational Target of mTOR Signaling That Regulates Axon Morphogenesis. J Neurosci *38*, 4985-4995.

Villasana, L.E., Klann, E., and Tejada-Simon, M.V. (2006). Rapid isolation of synaptoneurosomes and postsynaptic densities from adult mouse hippocampus. J Neurosci Methods *158*, 30-36.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184-2185.

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.

Westmark, C.J., Hervey, C.M., Berry-Kravis, E.M., and Malter, J.S. (2011). Effect of Anticoagulants on Amyloid beta-Protein Precursor and Amyloid Beta Levels in Plasma. J Alzheimers Dis Parkinsonism *1*, 101.

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 alpha-CaMKII mRNA at synapses. Neuron *21*, 1129-1139.

Zappulo, A., van den Bruck, D., Ciolli Mattioli, C., Franke, V., Imami, K., McShane, E., Moreno-Estelles, M., Calviello, L., Filipchyk, A., Peguero-Sanchez, E*., et al.* (2017). RNA localization is a key determinant of neurite-enriched proteome. Nat Commun *8*, 583.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N*., et al.* (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci *34*, 11929-11947.

Supplemental Figures and Legends

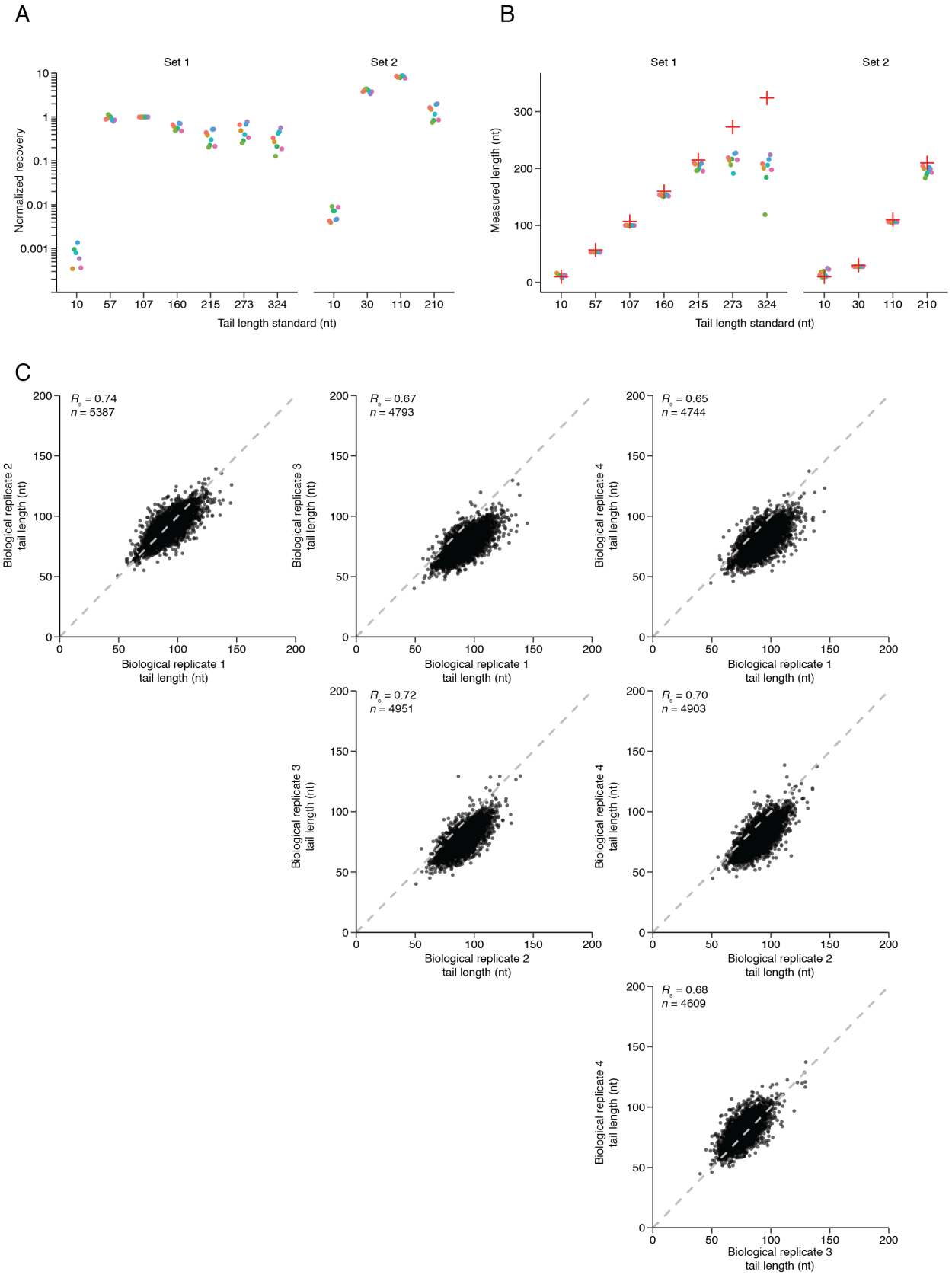


Figure S1. PAL-seq Standards and Reproducibility, Related to Figure 1

(A) Two sets of PAL-seq v2 library standards were added to total RNA prior to each library preparation. The recovery of each standard in each of the 8 PAL-seq datasets is normalized to the 107 nt standard. Recovery was adjusted for expected values based on quantification of the original spike mixture. (B) Mean tail length measurements of the synthetic standards of each standard in (A). For each standard in each dataset, the mean tail length measured from PAL-seq is plotted on the y axis. The tail length measured from PAGE is plotted on the x-axis, with red crosses denoting expected measurements. Standards with tail lengths > 250 nt were not expected to be measured accurately because their tail length exceeded the length of the sequencing read used to measure the tail. (C) Pairwise correlations between tail length measurements for each biological replicate in the condition in which the mice had not been exposed to light (n = 4). Shown are mean tail lengths for mRNAs from each gene exceeding the 50-tag cutoff in both samples. Dashed line is y = x.

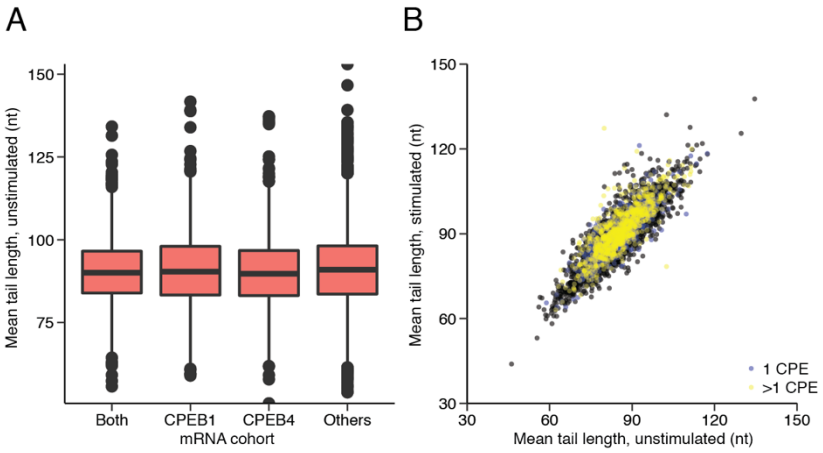


Figure S2. No Evidence for Longer Tail Lengths of CPE-Containing mRNAs, Related to Figure 1

(A) CPEB-bound mRNAs do not have longer tails. RNA immunoprecipitation followed by sequencing (RIP-seq) data for CPEB1, CPEB4, or their intersection (Parras et al., 2018) from the mouse striatum were used to bin mean tail lengths from one unstimulated visual cortex sample. CPEB-bound mRNAs do not tend to have longer tail lengths. n = 622 mRNAs bound by only CPEB1, 1215 mRNAs bound by only CPEB4, 1064 mRNAs bound by both CPEB1 and CPEB4, and 5198 mRNAs not bound by either. Line, median. Box, 25th and 75th percentiles. Whiskers, 1.5x interquartile range. Data beyond the whiskers is plotted as points. Using the other visual cortex datasets did not change this result. (B) mRNAs containing CPE motifs do not become polyadenylated in response to light stimulation. Mean tail lengths for the unstimulated and stimulated cohort (n = 4 mice per cohort) are plotted. mRNAs in blue contain one CPE motif, defined as one stretch of 12 contiguous uridines in their 3′ UTR, allowing one non-uridine anywhere in this stretch (Subtelny et al., 2014). Points in yellow contain more than one CPE motif.

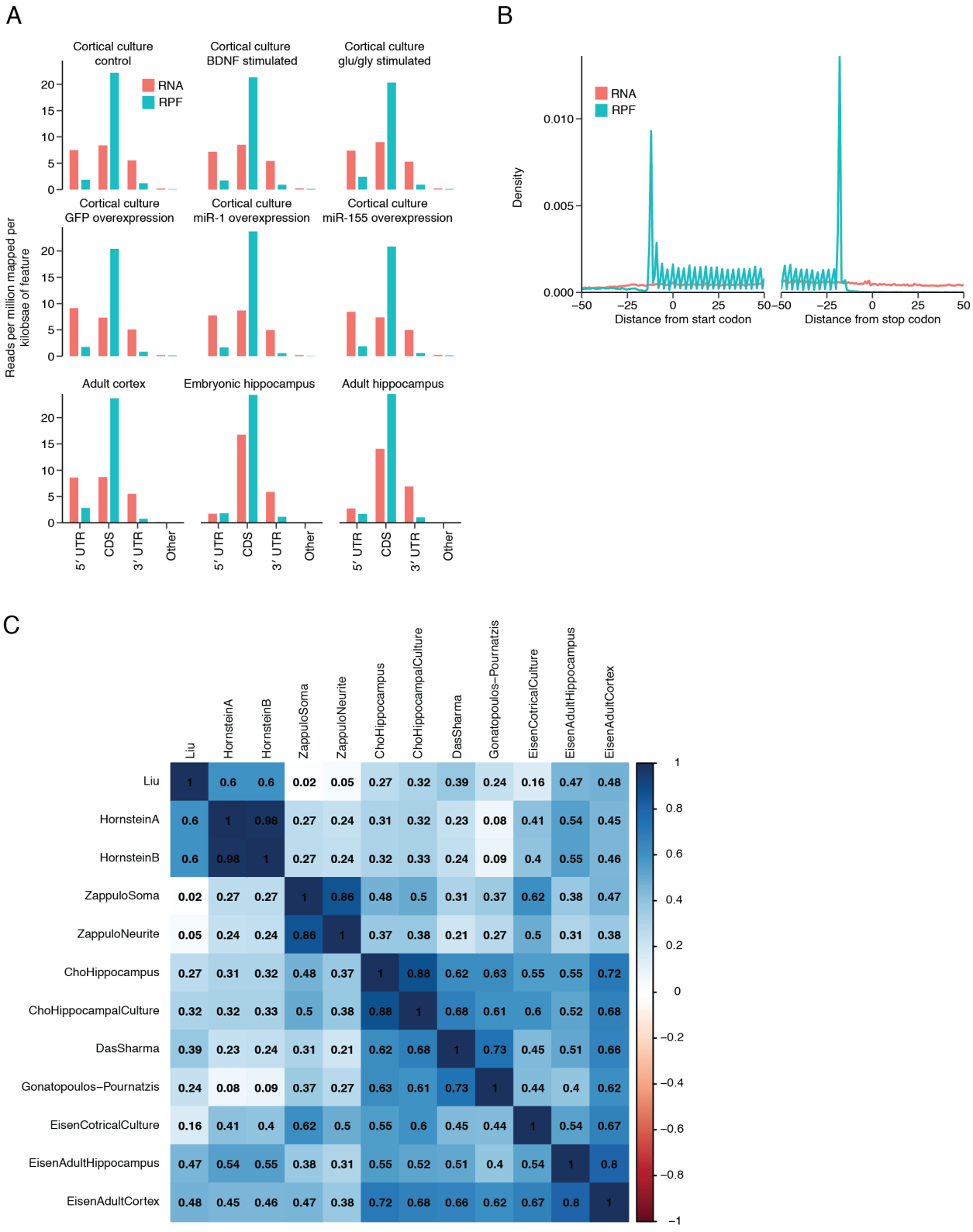

A
Cortical culture control
Cortical culture BDNF stimulated
Cortical culture glu/gly stimulated
Cortical culture GFP overexpression
Cortical culture miR-1 overexpression
Cortical culture miR-155 overexpression
Adult cortex
Embryonic hippocampus
Adult hippocampus
RNA
RPF
Reads per million mapped per kilobase of feature
5′ UTR
CDS
3′ UTR
Other
B
RNA
RPF
Density
Distance from start codon
Distance from stop codon
C
Liu
HornsteinA
HornsteinB
ZappuloSoma
ZappuloNeurite
ChoHippocampus
ChoHippocampalCulture
DasSharma
Gonatopoulos–Pournatzis
EisenCotricalCulture
EisenAdultHippocampus
EisenAdultCortex

Figure S3. Assessment of Ribosome Profiling Data, Related to Figure 3

(A) Distribution of reads across the coding sequence and UTRs for the nine ribosome profiling and matched RNA-seq datasets in this study. For each dataset, the fraction of reads per million mapped reads per kilobase of either the coding sequence or UTRs is displayed for the RNA-seq data (red) and the ribosome profiling data (green). Reads that map to introns or 10 kb upstream or downstream of the transcription start and end, respectively, comprise the other category. Analysis was performed using RSeQC software (Wang et al., 2012). (B) Periodicity in ribosome profiling data. The expected 3 nt periodicity is observed in the ribosome profiling but not the RNA-seq data. Shown is a meta-transcript from the primary cortical culture control sample generated by calculating the distance between the 5′-most nucleotides of gene-mapping reads and the start (left plot) or end (right plot) of coding sequences. (C) Agreement between TE measurements from different studies. Pairwise spearman correlations of translational efficiencies ($\log_2$) between each dataset are plotted, along with a color indicating their strength. Data are from the following brain or neuron-derived samples in mouse: Cho et al. profiled primary hippocampal culture and tissue (Cho et al., 2015); Liu et al. profiled adult neural stem cells (Liu et al., 2018); Das Sharma et al. profiled mouse cortex (Das Sharma et al., 2019); Hornstein et al. profiled two biological replicates from brain (Hornstein et al., 2016); Zappulo et al. profiled somata and neurites from neurons induced from embryonic stem cells (Zappulo et al., 2017); Gonatopoulos-Pournatzis profiled embryonic stem-cell derived neurons. Also shown are the cortical culture, adult hippocampus, and adult cortex samples from this study. The order of the samples is the result of hierarchical clustering using the *hclust* command in R. n = 3,577 mRNAs passed cutoffs in each dataset and are included in the comparison.

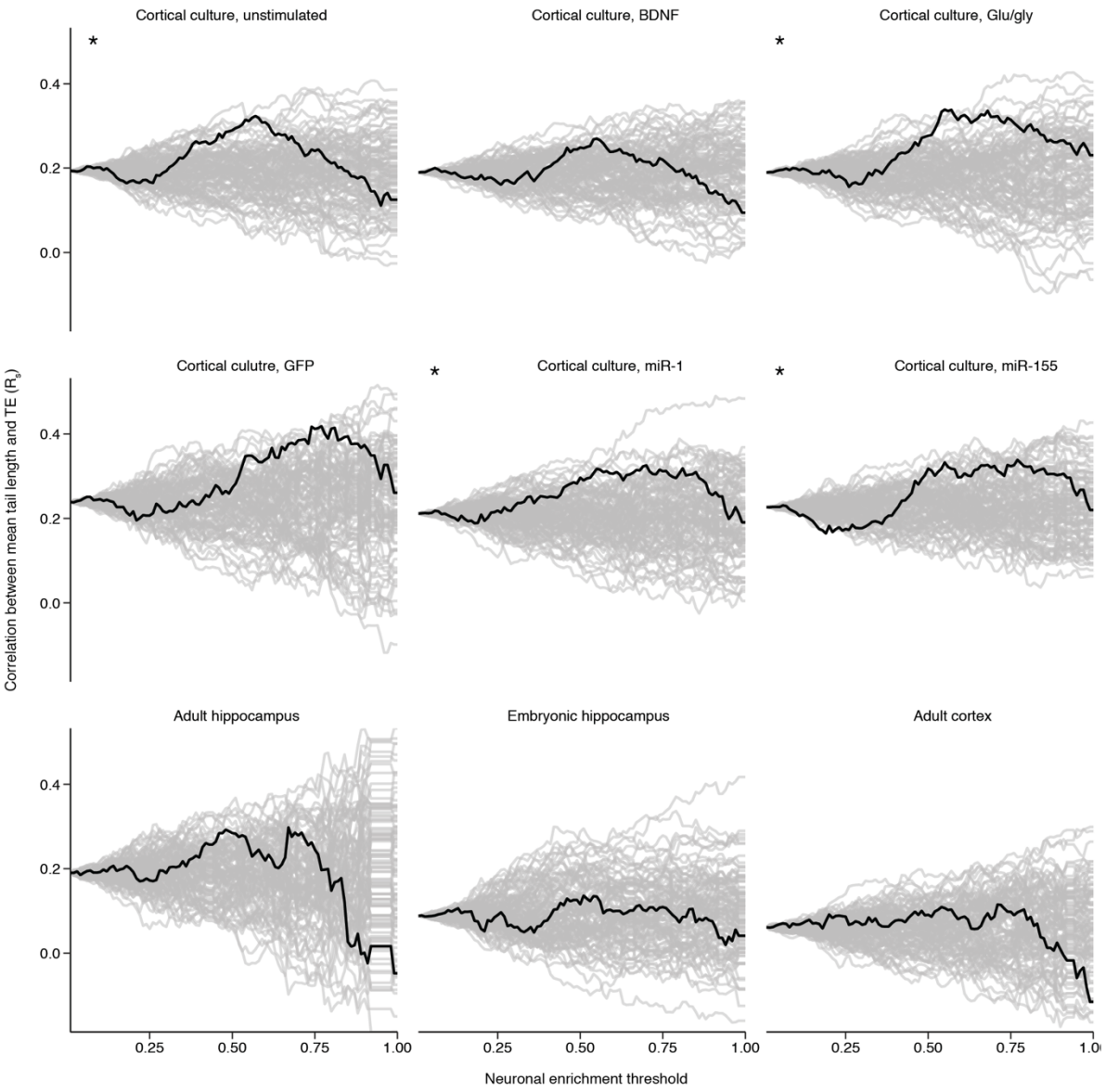


Figure S4. Neuronal mRNAs Do Not Exhibit Stronger Correlations Between Tail Length and TE, Related to Figure 4

For each of the 9 samples for which tail length and TE were measured, a subset of mRNAs were considered based on their biased expression in neurons. To determine this subset, mRNAs were ranked by the RPKM ratio of neuronal expression to non-neuronal expression (Zhang et al., 2014). These rankings were binned into 100 quantiles, and the fraction of these quantiles considered is shown on the x axis, with larger values indicating increasing stringency of neuronal-biased expression. The y axis is the Spearman correlation between tail length and TE for the mRNAs considered. These correlation coefficients are highly susceptible to sample size, and as such 100 random subsets of the original data with the same sample size at each threshold are shown in grey, with the subsetting based on enrichment shown in black. Asterisks denote significance using an empirical $p < 0.05$ based on the 50^{th} quantile.

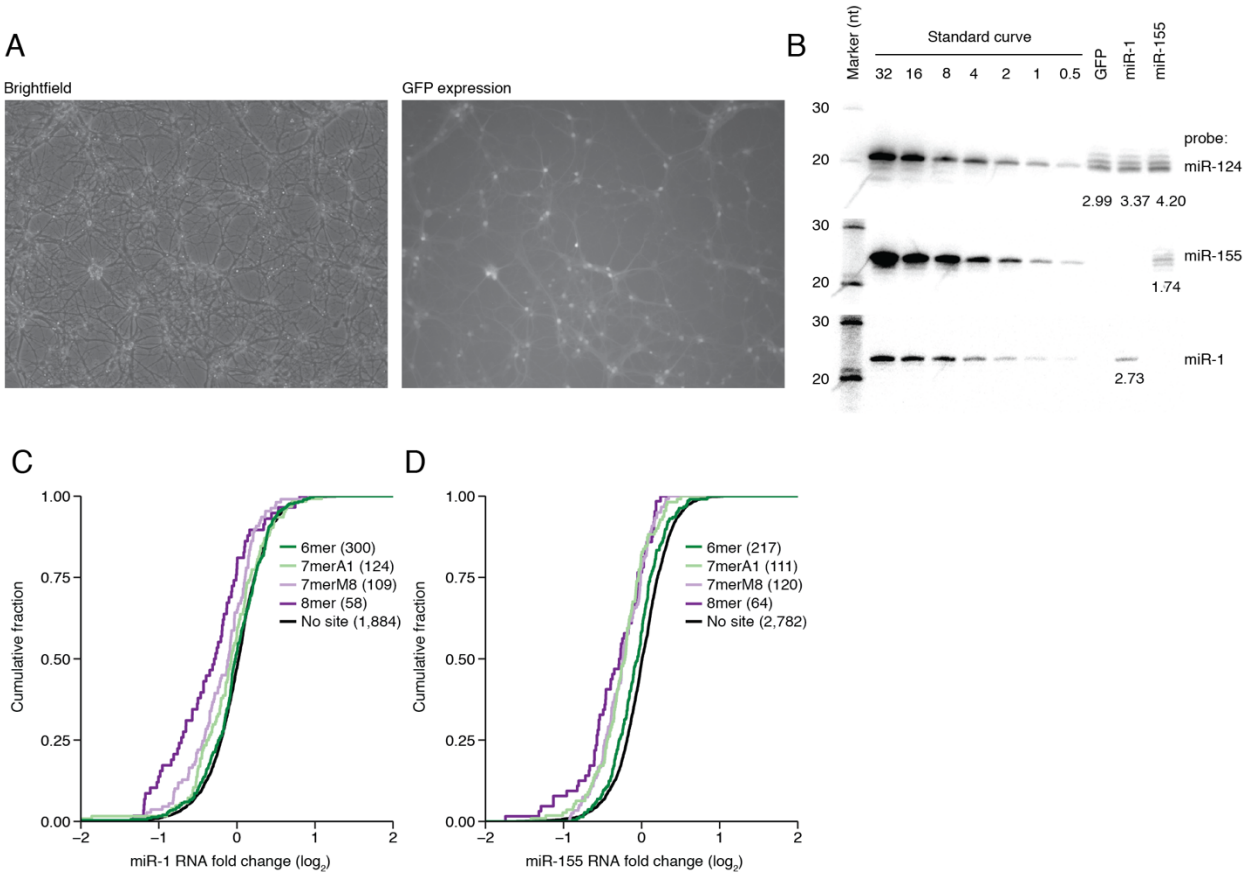


Figure S5. Exogenous miRNAs Are Expressed in Neurons, Related to Figure 5

(A). Almost every neuron expresses lentivirus. A representative brightfield (left) and fluorescent (488 nm, right) micrograph of the neuronal cultures transduced with miRNA-expressing virus. (B) miRNAs are expressed in neurons. Shown is an RNA blot for miR-124 (top) with a standard curve (fmol/µg) and 1 µg of total RNA loaded from either GFP-expressing, miR-1-expressing, or miR-155-expressing cultured neurons. This blot was stripped and re-probed for either miR-1 or miR-155. Quantification based on the standard curve (in fmol) is below the band. For miR-124 and miR-155, all bands were included in the quantification. (C) Site-type hierarchies are maintained for miR-1 in neurons. mRNAs were classified as to whether they lacked a site throughout their entire sequence or whether they contained one of the four canonical target-site types (Bartel, 2009). The fold change in RNA-seq reads between the miR-1-transduced and GFP-transduced primary cultures was calculated and plotted as a CDF, binned by site type. Numbers of targets are shown in parentheses. (D) Site-type hierarchies are maintained for miR-155 in neurons. Analysis as in (C), except plotting data for primary neurons transduced with the miR-155 virus instead of the miR-1 virus, and considering only seed-match sites to miR-155.

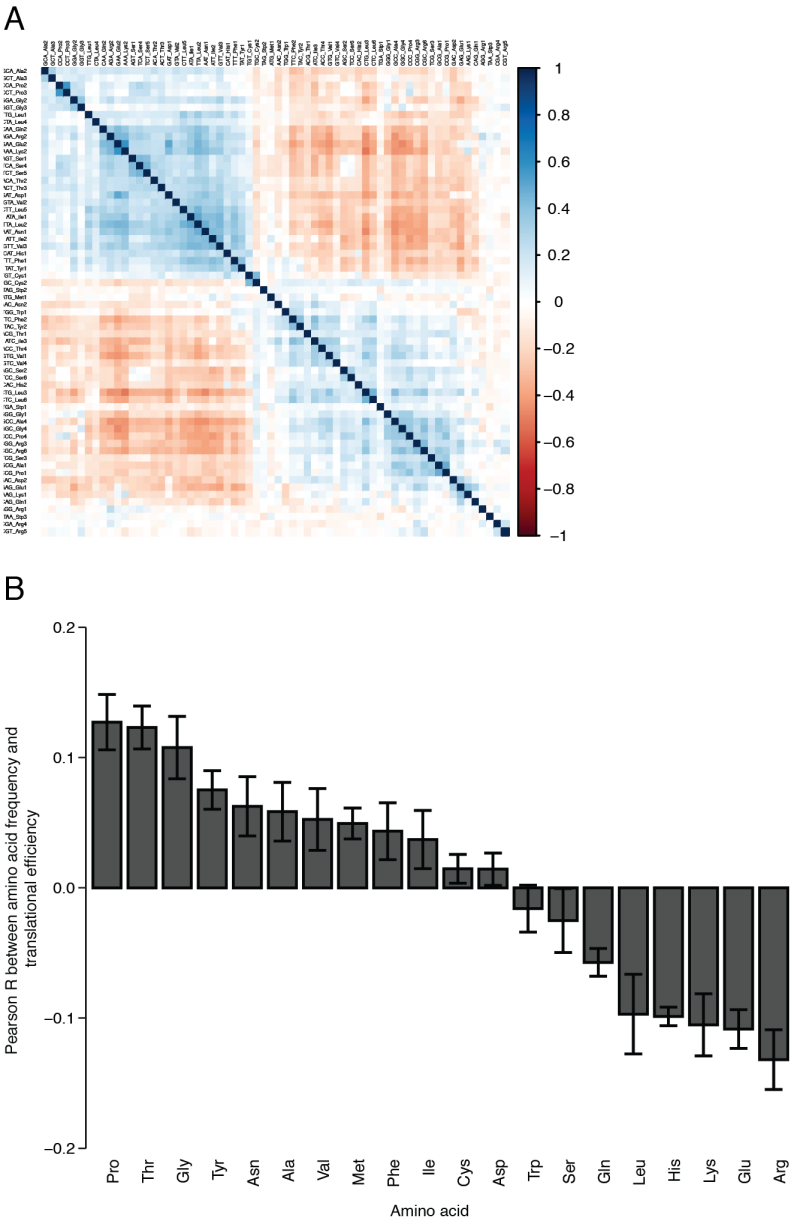


Figure S6. Frequencies of Some Amino Acids Correlate With TE, Related to Figures 6–7
(A) Codon composition across mRNAs expressed in DIV14 primary cortical culture neurons. For each of the 61 coding codons, its frequency in mRNAs that passed cutoffs for expression (n = 5,128) was determined. The Pearson correlation between these frequencies was calculated for every pair of codons and the resulting data were clustered using Euclidean distance. Positive values indicate that two codons are more likely to appear in the same transcript, while negative values indicate that two codons are more likely to be mutually exclusive. (B) The TE and amino acid frequency were compared for all measured transcripts in each of the six primary culture samples. The resulting Pearson R value is plotted as a bar plot, with error bars denoting the standard deviation between different cortical culture samples. Correlations are ranked by the mean R value. Otherwise as in Figure 7B.

Chapter 5. Future Directions

Timothy J. Eisen[1,2,3]

[1]Howard Hughes Medical Institute, Cambridge, MA, 02142, USA
[2]Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

The Dynamics of Cytoplasmic mRNA Metabolism

Our experiments and analysis of poly(A)-tail lengths of ageing mRNAs in mammalian cells have provided the basis for understanding the dynamics of mRNA decay. We find deadenylation rates that span 1000-fold and form the basis for mRNA decay rates. After mRNA tail lengths become short, the decay rates of these mRNAs vary by a similar 1000-fold range, indicating that additional mechanisms governing either decapping or decay of the mRNA body contribute to mRNA decay rate overall. These short-tailed decay rates are tuned to the deadenylation rates, with faster deadenylation predictive of more rapid decay of short-tailed isoforms. Collectively, these observations suggest that mRNA decay is an integrated process, with deadenylation and short-tailed decay coupled, a coupling that ensures the uninterrupted flow of intermediates through the decay pathway.

How the rate of short-tailed mRNA decay is coupled to the rate of deadenylation remains unclear. Known physical interactions between the CCR4–NOT complex and the DCP1–DCP2 complex, mediated by EDC4 and other proteins, link the two complexes (Haas et al., 2010). In addition, the two activities of decapping and deadenylation are colocalized subcellularly in processing bodies (Parker and Sheth, 2007). From these observations, one model is that additional recruitment of the deadenylase complex to mRNAs that are rapidly degraded also recruits the decapping machinery. However, this model poses several challenges. One such challenge is that our results do not indicate a change in substrate preference for decapping complexes active on rapidly degraded mRNAs: long-lived and short-lived mRNAs appear to be degraded at similar tail lengths. Thus, for the increased-recruitment model to be appropriate, additional decapping complexes would be present but inactive during the rapid deadenylation, and only when tail length passes the decapping threshold, perhaps due to loss of PABP, they

would become active. This posits a poised decapping model, where the decapping complex is concentrated on young, long-tailed and short-lived mRNAs.

In these studies, we did not investigate the directionality of deadenylation-dependent mRNA decay. Previous analyses of some rapidly degraded mRNAs revealed that they are substrates for NMD (Bicknell et al., 2012). Perhaps, for these cleaved mRNAs, we are observing deadenylation of their 3′ fragments. If this generalizes to other mRNAs, the coupling between deadenylation and decapping might be mediated by two simultaneous but largely independent phenomena that achieve the same goal of degrading the message. Recruitment of the decapping complex need not be enhanced by rapid deadenylation.

A technically challenging but exciting direction would be to extend the time-resolved sequencing experiments performed here to include 5′-end status. This could be achieved by additional biochemical separation of capped and uncapped mRNAs, or by long-read sequencing. Oxford Nanopore Technology can measure poly(A)-tail lengths alongside full-length transcript information, albeit at lower coverage, a higher mutation rate, and poorer quality tail-length measurements (Krause et al., 2019; Workman et al., 2019). These data might shed light on the fraction of mRNAs of different tail lengths that have intact 5′ ends.

In addition to questions related to mechanism, the scope of the coupling between tail length and short-tailed decay has yet to be addressed. Thus far, detailed studies of mRNA decay have been confined to biochemically tractable biological systems such as yeast, 3T3 cells, and human cell lines. These systems have provided enormous insight into mRNA decay dynamics but might prove poor correlates of in vivo contexts. A systematic study of steady-state tail lengths of many tissues in a mouse, for example, has not been undertaken. Ideally, this study would make these measurements using a technique that does not limit poly(A)-tail lengths to 250

nt, because for some tissues, much longer tail lengths have been reported (Wu et al., 1998). The technique would also not require mRNAs to have a poly(A)-tail, as different tissues might contain different fractions of tail-less mRNAs, which could provide insights into the relative rates of deadenylation and decay in those contexts. Finally, such a study would require kinetics. In vivo metabolic labeling has been performed successfully in *Drosophila* using temporally and spatially restricted expression of uracil phosphoribosyltransferase (UPRT), an enzyme that is required to incorporate 4-thiouridine into RNA (Miller et al., 2009; Gay et al., 2013). Additional studies used the approach to label RNAs in several tissues of the mouse including the intestine and adipose tissue (Matsushima et al., 2018), sperm (Sharma et al., 2018), and hippocampus (Chatzi et al., 2016). Using transient expression of UPRT or transient dosing of 4-thiouridine would allow estimates of mRNA turnover and tail length dynamics in tissues, but the lack of a meaningful steady-state might make subsequent analyses difficult. Such challenges can be overcome, however, by using short, length-invariant labeling intervals (Scheetz et al., 2000; Rabani et al., 2011).

The studies concerning mRNA decay thus far have been restricted to cytoplasmic mRNAs. While the bulk of RNA degrades in the cytoplasm, many quality control pathways for aberrant nuclear RNA have been described. Two complexes are the exosome and the Trf4–Air2–Mtr4 (TRAMP) complex. The exosome is a ring-like complex that, along with Rrp6, functions in rRNA-precursor processing (Houseley et al., 2006; Vanacova and Stefl, 2007) as well as quality control of many types of RNA. It degrades RNA by threading it through a central channel to its exonucleolytic subunit. The TRAMP complex provides specificity for the nuclear exosome by adding short oligo(A) tails to molecules that are destined for degradation (Jia et al., 2011). These activities, combined with the polyadenylation activity that occurs on almost all pre-mRNAs after

cleavage, suggest a dynamic landscape of tail lengths and mRNA processing. Investigating these dynamics for the milieu of endogenous nuclear RNA will provide new insight into nuclear RNA turnover.

Despite the discovery of many cytoplasmic and nuclear factors that influence mRNA decay, our understanding of the variance in deadenylation rate is still incomplete. A linear model that uses features of the mRNA, such as protein-binding motifs, miRNA-target sites, structure, and codon optimality to predict decay rate accurately would indicate a thorough understanding of the cytoplasmic players contributing to mRNA decay. One prerequisite problem is understanding the variance in miRNA targeting, where there has been headway (McGeary et al., 2019). Another study captured some of the differences in mRNA decay rate, with CDS exon junction density as the most predictive feature (Spies et al., 2013). We are also encouraged that recent efforts to predict mRNA expression level from sequence have captured a majority of the variance (Agarwal and Shendure, 2018), but further work is required to interpret the complex networks of proteins and RNAs that cause mRNA decay.

MicroRNA Cause Accelerated Decay of Short-Tailed Target mRNAs

Prior to our work, investigators appreciated the interaction between TNRC6 and the CCR4–NOT complex, and, by extension, the interaction between CCR4–NOT and DCP1–DCP2. But miRNAs do not alter poly(A)-tail lengths of their targets either 6 hpf in the developing Zebrafish embryo (Subtelny et al., 2014) or in steady-state 3T3 cells (Eisen et al., 2020). This paradox may be explained by a concomitant increase in both the deadenylation rate and the short-tailed decay rate. When rapidly deadenylated molecules are more rapidly cleared from the cell, these molecules fail to drive down the mean tail length. The increase in short-tailed decay rate need

not be specific to miRNAs. Indeed, because connections exist between the deadenylation and decay machinery more generally, it may be difficult to accelerate one rate without the other.

The observation that miRNAs fail to alter steady-state tail length does not support one of the mechanisms proposed for miRNA-mediated translational repression. In this mechanism, deadenylation due to miRNAs decreases PABP occupancy, which in turn decreases the affinity of eIF4E for the cap. As the steady-state fraction of short-tailed mRNAs is unchanged when they become miRNA targets, PABP occupancy would be unaltered. In pre-steady state conditions, however, miRNAs change tail length. Despite these tail length changes, we fail to observe significant translational repression at any time interval, suggesting that the fraction of short-tailed mRNAs is either not sufficient to produce significant translational repression, or short-tailed mRNAs are still efficiently translated (Subtelny et al., 2014; Eisen et al., 2020). A study in *Drosophila* S2 cells using reporters for miRNA activity observed substantial translational repression but no changes to pre-steady state or steady-state tail lengths (Djuranovic et al., 2012), proposing tail-length independent mechanisms of translational repression.

A technical achievement of our study was the ability to perform translation measurements using ribosome profiling on nascent mRNAs. This advance was made possible by the 5-ethynyl uridine labeling system, which allowed purification of RNA fragments containing a single labeled uridine. The previous thiouridine labeling systems (Dolken et al., 2008; Duffy et al., 2015) were too inefficient to achieve substantial signal above background for molecules containing one labeled uridine. Despite this advance, the earliest time interval examined was 40 minutes, which more closely resembled a 13-minute time interval because of the amount of time required for mRNAs to be transcribed, processed and exported, and we used only cytoplasmically enriched RNA (Eisen et al., 2020). An exciting future direction from these

experiments involves a more detailed study of time-resolved translation of mRNAs as they age. This would require earlier time intervals to observe an early phase of translation, when only a few ribosomes are bound to the mRNA just after it has exited the nucleus. Such data might provide insight into NMD, which is thought to target nascent mRNAs that retain exon-junction complexes after the pioneering round of translation. Literature is in conflict as to whether NMD occurs immediately after the pioneering round (Lykke-Andersen and Jensen, 2015; Hoek et al., 2019), but the timing of this translation event has not been investigated for endogenous mRNAs.

A related direction would be to recover aged mRNAs at late time intervals in a pulse-chase–like experiment. These data might speak to whether an mRNA's age dictates its ability to be translated. Perhaps as an mRNA lives in the cytoplasm for longer periods, it is more likely to develop impediments to translation that reduce its output, leading it to decay.

Poly(A)-Tail Length has a Significant but Modest Effect on Translational Efficiency in Neurons

An observation that our study confirmed is the dramatic range of translational efficiencies that neurons exhibit, with a 1000-fold difference between the best and the worst translated mRNAs (Hornstein et al., 2016). Our linear model was able to explain a small but significant fraction of this variance, but the two systems that we used, primary mouse cultured neurons and brain tissues, exhibited differences. In both contexts, expression is an important feature, suggesting that mRNAs are optimized for both high expression and high translational output. In tissues from the brain, coding sequence length inversely correlates with translational efficiency, with short coding sequences translated best. In all samples, however, we observed that codon composition greatly increases the predictive accuracy of our linear model. This finding is in keeping with

reports suggesting extensive ribosome stalling (Das Sharma et al., 2019) and elongation-rate determinants of translational efficiency in neurons (Darnell et al., 2011).

In cultured cortical neurons, we observe a modest but consistent relationship between poly(A)-tail length and translational efficiency, explaining ~5% of the variance, depending on the sample. The compelling future direction from our study is to understand this effect. At least two possibilities exist for such a modest relationship: (1) cis-acting elements that regulate translation but vary between genes dwarf the effects of tail length, or (2) this signal is driven by a subset of mRNAs in a specific subcellular compartment. In the former case, experiments that compare mRNAs from the same gene and measure tail lengths, stratifying based on translational efficiency, should observe a stronger relationship. We performed experiments to investigate this hypothesis using polysome profiling but failed to see signal for a relationship between tail length and translation. The same method could not observe a strong relationship in the coupled system of the *Xenopus* oocyte, however, indicating that technical challenges of polysome profiling, such as its limited dynamic range of translation efficiency measurements and imprecise assignment of ribosome numbers to mRNAs, clouded these results. Alternative methods for selecting a subset of highly translating mRNAs exist, however. Translating ribosome affinity purification can purify translating mRNAs using an EGFP tag on a large ribosomal subunit protein (Heiman et al., 2008), with many mouse lines harboring labels of different cell populations in the brain. Other ribosomal tagging strategies have also proven fruitful (Sanz et al., 2009; Simsek et al., 2017). Perhaps these strategies would allow discovery of the tail length differences between translating and non-translating pools for neuronal populations in vivo.

The more pressing issue concerning this study is whether the relationship between tail length and translational efficiency might increase when considering only mRNA in the distal

regions of the neuron, such as the synapse or part of the dendrite. Biochemical purifications of these regions have been challenging, either because of the low enrichments that are often observed, for example in synaptoneurosome preparations, or because of the small amounts material recovered, which is in keeping with the small volume and numbers of mRNAs that have been hypothesize to exist in dendrites (Kosik, 2016). A recent report used microdissection of the rat synaptoneuropil, a region of the hippocampus largely devoid of neuronal somata, and performed ribosome profiling (Biever et al., 2020). They observed that monosomes, rather than polysomes, are more frequently found translating mRNAs in this region as compared to the cell bodies, in keeping with the lower frequency of ribosomes in the dendrites. These samples would be interesting to use for examining the role of poly(A)-tail length in translational efficiency.

Conclusion

Previous study of poly(A)-tail length in mammalian cells was limited to either the dynamics of a couple of mRNAs or the transcriptome-wide measurement of mRNAs at steady-state. By combining these two approaches, we were able to measure rate constants for many steps in the metabolism of thousands of mRNAs. Subsequent analysis of these rate constants revealed insights into mRNA decay and miRNA-mediated repression. These rate measurements generate more questions than answers though. Future studies are required to describe the concerted mechanism of mRNP remodeling during decay, the role of the tail in different in vivo contexts, the dynamics of translational efficiency, and the cohort of mRNAs for which tail-length controls translation in neurons. New techniques that describe full-length transcripts, require less material, and capture mRNAs as they age will certainly be of use.

References

Agarwal, V., and Shendure, J. (2018). Predicting mRNA abundance directly from genomic sequence using deep convolutional neural networks. bioRxiv.

Bicknell, A.A., Cenik, C., Chua, H.N., Roth, F.P., and Moore, M.J. (2012). Introns in UTRs: why we should stop ignoring them. Bioessays *34*, 1025-1034.

Biever, A., Glock, C., Tushev, G., Ciirdaeva, E., Dalmay, T., Langer, J.D., and Schuman, E.M. (2020). Monosomes actively translate synaptic mRNAs in neuronal processes. Science *367*.

Chatzi, C., Zhang, Y., Shen, R., Westbrook, G.L., and Goodman, R.H. (2016). Transcriptional Profiling of Newly Generated Dentate Granule Cells Using TU Tagging Reveals Pattern Shifts in Gene Expression during Circuit Integration. eNeuro *3*.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W.*, et al.* (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247-261.

Das Sharma, S., Metz, J.B., Li, H., Hobson, B.D., Hornstein, N., Sulzer, D., Tang, G., and Sims, P.A. (2019). Widespread Alterations in Translation Elongation in the Brain of Juvenile Fmr1 Knockout Mice. Cell Rep *26*, 3313-3322 e3315.

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.

Dolken, L., Ruzsics, Z., Radle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P.*, et al.* (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959-1972.

Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. Mol Cell *59*, 858-866.

Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., and Bartel, D.P. (2020). MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. Mol Cell *77*, 775-785 e778.

Gay, L., Miller, M.R., Ventura, P.B., Devasthali, V., Vue, Z., Thompson, H.L., Temple, S., Zong, H., Cleary, M.D., Stankunas, K.*, et al.* (2013). Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. Genes Dev *27*, 98-115.

Haas, G., Braun, J.E., Igreja, C., Tritschler, F., Nishihara, T., and Izaurralde, E. (2010). HPat provides a link between deadenylation and decapping in metazoa. J Cell Biol *189*, 289-302.

Heiman, M., Schaefer, A., Gong, S., Peterson, J.D., Day, M., Ramsey, K.E., Suarez-Farinas, M., Schwarz, C., Stephan, D.A., Surmeier, D.J.*, et al.* (2008). A translational profiling approach for the molecular characterization of CNS cell types. Cell *135*, 738-748.

Hoek, T.A., Khuperkar, D., Lindeboom, R.G.H., Sonneveld, S., Verhagen, B.M.P., Boersma, S., Vermeulen, M., and Tanenbaum, M.E. (2019). Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. Mol Cell *75*, 324-339 e311.

Hornstein, N., Torres, D., Das Sharma, S., Tang, G., Canoll, P., and Sims, P.A. (2016). Ligation-free ribosome profiling of cell type-specific translation in the brain. Genome Biol *17*, 149.

Houseley, J., LaCava, J., and Tollervey, D. (2006). RNA-quality control by the exosome. Nat Rev Mol Cell Biol *7*, 529-539.

Jia, H., Wang, X., Liu, F., Guenther, U.P., Srinivasan, S., Anderson, J.T., and Jankowsky, E. (2011). The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. Cell *145*, 890-901.

Kosik, K.S. (2016). Life at Low Copy Number: How Dendrites Manage with So Few mRNAs. Neuron *92*, 1168-1180.

Krause, M., Niazi, A.M., Labun, K., Torres Cleuren, Y.N., Muller, F.S., and Valen, E. (2019). tailfindr: alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. RNA *25*, 1229-1241.

Lykke-Andersen, S., and Jensen, T.H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Biol *16*, 665-677.

Matsushima, W., Herzog, V.A., Neumann, T., Gapp, K., Zuber, J., Ameres, S.L., and Miska, E.A. (2018). SLAM-ITseq: sequencing cell type-specific transcriptomes without cell sorting. Development *145*.

McGeary, S.E., Lin, K.S., Shi, C.Y., Pham, T.M., Bisaria, N., Kelley, G.M., and Bartel, D.P. (2019). The biochemical basis of microRNA targeting efficacy. Science *366*.

Miller, M.R., Robinson, K.J., Cleary, M.D., and Doe, C.Q. (2009). TU-tagging: cell type-specific RNA isolation from intact complex tissues. Nat Methods *6*, 439-441.

Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. Mol Cell *25*, 635-646.

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N*., et al.* (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat Biotechnol *29*, 436-442.

Sanz, E., Yang, L., Su, T., Morris, D.R., McKnight, G.S., and Amieux, P.S. (2009). Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. Proc Natl Acad Sci U S A *106*, 13939-13944.

Scheetz, A.J., Nairn, A.C., and Constantine-Paton, M. (2000). NMDA receptor-mediated control of protein synthesis at developing synapses. Nat Neurosci *3*, 211-216.

Sharma, U., Sun, F., Conine, C.C., Reichholf, B., Kukreja, S., Herzog, V.A., Ameres, S.L., and Rando, O.J. (2018). Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm. Dev Cell *46*, 481-494 e486.

Simsek, D., Tiu, G.C., Flynn, R.A., Byeon, G.W., Leppek, K., Xu, A.F., Chang, H.Y., and Barna, M. (2017). The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. Cell *169*, 1051-1065 e1018.

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.

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.

Vanacova, S., and Stefl, R. (2007). The exosome and RNA quality control in the nucleus. EMBO Rep *8*, 651-657.

Workman, R.E., Tang, A.D., Tang, P.S., Jain, M., Tyson, J.R., Razaghi, R., Zuzarte, P.C., Gilpatrick, T., Payne, A., Quick, J.*, et al.* (2019). Nanopore native RNA sequencing of a human poly(A) transcriptome. Nat Methods *16*, 1297-1305.

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 alpha-CaMKII mRNA at synapses. Neuron *21*, 1129-1139.

Appendix A. Affinity Inequality among Serum Antibodies That Originate in Lymphoid Germinal Centers

Myungsun Kang[1,2] *, Timothy J. Eisen[3,4] *, Ellen A. Eisen[5†], Arup K. Chakraborty[1,2,6,7,8,9†], Herman N. Eisen[3,10††]

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
[2]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
[4]Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
[5]Environmental Health Sciences, School of Public Health, University of California, Berkeley, California, United States of America
[6]Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
[7]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
[8]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America
[9]Ragon Institute of the Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard Medical School, Cambridge, Massachusetts, United States of America,
[10]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

*These authors contributed equally to this work.
[†]Correspondence: arupc@mit.edu (AKC); eeisen@berkeley.edu (EAE)
[††]Deceased.

Published as: Kang, M., Eisen, T.J., Eisen, E.A., Chakraborty, A.K., and Eisen, H.N. (2015). Affinity Inequality among Serum Antibodies That Originate in Lymphoid Germinal Centers. PLoS One 10, e0139222.

PLOS | ONE

RESEARCH ARTICLE

Affinity Inequality among Serum Antibodies That Originate in Lymphoid Germinal Centers

Myungsun Kang[1,2☯], Timothy J. Eisen[3,4☯], Ellen A. Eisen[5]*, Arup K. Chakraborty[1,2,6,7,8,9]*, Herman N. Eisen[3,10†]

1 Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **2** Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **3** Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **4** Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **5** Environmental Health Sciences, School of Public Health, University of California, Berkeley, California, United States of America, **6** Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **7** Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **8** Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **9** Ragon Institute of the Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard Medical School, Cambridge, Massachusetts, United States of America, **10** Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

† Deceased.
☯ These authors contributed equally to this work.
* arupc@mit.edu (AKC); eeisen@berkeley.edu (EAE)

Citation: Kang M, Eisen TJ, Eisen EA, Chakraborty AK, Eisen HN (2015) Affinity Inequality among Serum Antibodies That Originate in Lymphoid Germinal Centers. PLoS ONE 10(10): e0139222. doi:10.1371/journal.pone.0139222

Editor: Sebastian D. Fugmann, Chang Gung University, TAIWAN

Received: June 6, 2015

Accepted: September 10, 2015

Published: October 7, 2015

Data Availability Statement: All relevant data are within the paper.

Funding: AKC and MSK are supported by the Ragon Institute of MGH, MIT, & Harvard (ragoninstitute.org) and the Samsung Scholarship Foundation (http://www.ssscholarship.com/), respectively. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. TJE is supported by National Science Foundation Graduate Research Fellowship under Grant No. 1122374. The funders had no role in study design, data collection and analysis, decision to

Abstract

Upon natural infection with pathogens or vaccination, antibodies are produced by a process called affinity maturation. As affinity maturation ensues, average affinity values between an antibody and ligand increase with time. Purified antibodies isolated from serum are invariably heterogeneous with respect to their affinity for the ligands they bind, whether macromolecular antigens or haptens (low molecular weight approximations of epitopes on antigens). However, less is known about how the extent of this heterogeneity evolves with time during affinity maturation. To shed light on this issue, we have taken advantage of previously published data from Eisen and Siskind (1964). Using the ratio of the strongest to the weakest binding subsets as a metric of heterogeneity (or affinity inequality), we analyzed antibodies isolated from individual serum samples. The ratios were initially as high as 50-fold, and decreased over a few weeks after a single injection of small antigen doses to around unity. This decrease in the effective heterogeneity of antibody affinities with time is consistent with Darwinian evolution in the strong selection limit. By contrast, neither the average affinity nor the heterogeneity evolves much with time for high doses of antigen, as competition between clones of the same affinity is minimal.

publish, or preparation of the manuscript. EAE and HNE received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

The strength of bonds formed by antibodies (Abs) with antigens (Ags) is one of the critical determinants of immune responses against pathogens. Ags are generally proteins and structurally so complex that much of what was first learned about Ab-Ag bonding, and how that evolves as affinity maturation ensues, is based upon Ab binding of small molecules that closely approximate the sites on protein Ags (epitopes) that are recognized by Abs. Called haptens, these small molecules bind specifically to Ag binding sites but themselves are not immunogenic. The strengths of the bonds that haptens form with Abs extend over about a million-fold range (10^3 or 10^4–10^{10} M^{-1}) [1,2]. This range encompasses the bond strengths measured for authentic protein Ags binding with Abs—whether monoclonal Abs or average values for heterogeneous (polyclonal) populations of purified Abs isolated from serum.

Long before Ab affinity could be measured it was known that immune sera to bacteria, red blood cells and proteins cross-reacted with structures that resembled the inciting antigen (the immunogen) and that after adsorption or precipitation of all Abs to cross-reacting (heterologous) structures, the remaining Abs could still react with the immunogen. These findings were attributed to a diversity of serum Abs that could react with different components of the complex immunogens. In 1936, however, Landsteiner and van der Scheer showed that antisera raised against an immunogen having a chemically defined epitope (azophenylsuberanilic acid) could be exhaustively adsorbed with various cross-reacting alternate forms of the epitope and the remaining Abs could still react with the homologous epitope; hence their conclusion that Abs to a singular epitope "are not entirely uniform but vary in specificity to some degree" [3].

Studies of Ab binding to haptens confirmed and extended previous work on more complex immunogens. Haptens used either to inhibit specific precipitation of Abs from antisera by haptenated Ags (hapten-inhibition) [4], or to bind directly, in absence of any Ag, to purified Abs [5], demonstrated variability in Ab binding. When ligands were added incrementally to Abs at a constant concentration in hapten inhibition experiments, the resulting binding curves were nearly always non-linear, as though apparent equilibrium constants, determined for each point, decreased with increasing concentration of the ligand. Such non-linearity was attributed to the variability inherent in the combined free energy of Abs with hapten in heterogeneous anti-arsanilate antiserum, which could be accounted for by a Gaussian error function [4]. There were misgivings about the quantitative nature of this approach because in hapten-inhibition of specific precipitation the composition of soluble complexes—of Ab, Ag, and hapten—were unknown [6]. However, Karush found that there was good agreement between hapten-Ab interactions measured directly by equilibrium dialysis and the theoretical binding curves based upon an assumed Gaussian distribution of free energy of hapten-Ab binding [7,8]. Thus binding of a hapten to a population of cognate Ab molecules isolated from serum could be characterized by two constants: i) the average bond-strength (K_0, the mean equilibrium association constant or intrinsic affinity), and ii) an index of heterogeneity with respect to affinity (sigma, σ).

In response to most immunogens, Abs made initially have low affinity and those made later have progressively higher affinity [9,10]. This progression, or affinity maturation, arises from events that take place in germinal centers (GC), small clusters of cells in secondary lymphoid tissues including lymph nodes, spleen, and Peyer's patches on intestinal mucosae [11]. Much of what we know about how processes in GCs lead to higher affinity Abs was learned subsequent to the first descriptions of affinity maturation. Each GC is formed by a few antigen (Ag)-stimulated naïve B cells [12] which, in GC, express the antigen-activated cytidine deaminase, AID, that causes mutations in the variable regions of the H and L chains of the Ab expressed by that B cell [13]. The diverse population of B cells thus generated express different Ag-

binding receptors (BCR). These cells are then selected against the antigen, which is displayed on follicular dendritic cells (FDC) in the GC [14–16]. Cells with BCRs that have high affinity for Ag bind it more readily than those with low affinity BCRs and are thus more likely to receive survival signals [17]. BCRs are also endocytic receptors [18]. Hence, cells with high affinity BCRs are more capable than those with low affinity BCRs to endocytose the Ag and present it as peptide-MHC-II complexes. These peptide-MHC-II complexes can engage with T cell receptors on the surface of cognate CD4 helper T cells in the GC [19]. The different B cells compete with each other for limiting numbers of T helper cells; B cells that internalize more antigens have a competitive advantage. Receipt of signals from T helper cells has been shown to be the key gatekeeper of B cell survival [20]. Inability to bind to Ag sufficiently strongly or receive T cell help results in apoptosis [21]. A few selected B cells emerge from the GC and differentiate into memory cells and antibody-secreting plasma cells, but most are recycled for further rounds of mutation and selection [22]. Thus, in vaccinated or infected individuals, Darwinian evolution occurs in a relatively short time-scale to generate antibodies with increasingly higher affinity as time ensues.

Advances in microscopy of live cells in intact lymph nodes have provided remarkable visualization of the cellular dynamics underlying the evolutionary events in GC [17,20,23]. Implicit in these observations, and in some of the mathematical models that characterize GC reactions, is an expectation that the Abs produced in a given immune response evolve over time to become relatively homogeneous and of high affinity [24–27]. Previously, however, affinity heterogeneity was found to increase over time, not to decrease [28]. Resolving this disparity is of interest as the extent of heterogeneity may have implications for the evolution of cross-reactive antibodies against highly mutable and persistent virus infections such as HIV-1 and HCV.

To examine the disparity between expected and observed changes in the heterogeneity of affinity of Abs undergoing affinity maturation, we re-analyzed previously published data of Eisen and Siskind [28] on the binding of haptens by serum Abs. Our analysis is based on affinity measurements of purified Abs obtained from rabbit sera. The protocol from the original paper collected sera over several weeks after dosing with a range of the hapten-bearing immunogens and characterized each collected population by average equilibrium constant (K_0) and sips heterogeneity index (a). In view of the overall approximately million-fold range in intrinsic affinity values measured for Ab-hapten and Ab-Ag reactions in general, our re-analysis addresses three specific questions. 1. How large is the diversity of Ab affinities, henceforth referred to as affinity inequality, in bleeds from individual animals or serum pools from a few individuals (rabbits)? 2. Does affinity inequality change over time after initiation of the response to the injected Ag? 3. To what extent is inequality affected by the quantity of Ag introduced?

To measure affinity inequality, we have relied primarily on affinity ratios between the strongest binding 5% and the weakest binding 5% of each purified Ab population. In response to relatively small doses of the immunogen, the heterogeneity indices have been found to increase over time [29,30]. Eisen and Siskind (26) originally interpreted their observation that heterogeneity increases over time to suggest that maintenance of Ab heterogeneity during the immune response is evolutionarily advantageous. But, we now show here, at low Ag dose, the extent of heterogeneity changes more slowly than the increase in the average affinity. As a result, the effective affinity inequality decreases. At high doses of Ag, however, affinity inequality is greater and more persistent. This finding suggests that, at low Ag dose, the Darwinian evolution process of affinity maturation is in the strong selection limit, while this is not the case when Ag dose is high.

Methods

The data we analyze are drawn from studies in which rabbits were injected with "haptenated" proteins 2, 4- dinitrophenyl (DNP)-bovine-γ-globulin. The quantities of Ag administered varied from 5mg to 250mg given as a single injection via footpads. Hapten-specific Abs were precipitated (or adsorbed) from serum by hapten-protein conjugates in which the protein component differed from (and did not cross-react with) the protein in the immunogen. Purified Abs recovered from the precipitates (or adsorbates) corresponded generally to ≈ 30–90% of total precipitable Abs in sera. Binding of haptens, ε-DNP-L-lysine, at various concentrations by fixed amounts of purified Abs were measured by equilibrium dialysis or fluorescence energy transfer (at 2–30°C).

The average equilibrium constants (K_0) were obtained as the reciprocal of the free ligand concentration when half the Ab binding sites were occupied [4,31]. From titrations of the hapten-Ab binding, an index of heterogeneity with respect to affinity could be obtained from the generalized adsorption isotherm [6,32]

$$\frac{r}{n} = \frac{(K_0 c)^a}{1 + (K_0 c)^a} \tag{1}$$

or more conveniently when expressed in alternative form [33] as

$$\log\left(\frac{r}{n-r}\right) = a\log(K_0) + a\log(c) \tag{2}$$

where r is mols hapten bound per Ab molecule, n is the number of binding sites per Ab molecule of the IgG type (2), c is the free ligand concentration, and *a* is an index of heterogeneity with respect to K_0 (affinity). By plotting $\log\left(\frac{r}{n-r}\right)$ vs. $\log(c)$, the data fall on a straight line whose slope is *a*, the Sips index of heterogeneity with respect to K_0.

The Sips and Gaussian probability distributions are virtually congruent [32] and affinity inequality was measured by converting values for the Sips heterogeneity index (*a*) to sigma (σ), the corresponding index of heterogeneity of the Gaussian distribution. The conversion was based upon the assumption that variations in ΔG described by the Sips distribution

$$N(\Delta G) = \frac{1}{\pi} \times \frac{\sin(\pi a) e^{[(\frac{a}{RT})(\Delta G^\circ - \Delta G)]}}{1 + 2\cos(\pi a)\, e^{[(\frac{a}{RT})(\Delta G^\circ - \Delta G)]} + e^{[(\frac{2a}{RT})(\Delta G^\circ - \Delta G)]}} \tag{3}$$

are equivalent to those described by the Gaussian distribution,

$$W(\Delta G) = \frac{1}{\sigma\sqrt{\pi}} \times e^{-[(\Delta G^\circ - \Delta G)^2/(RT\sigma)^2]} \tag{4}$$

when ΔG_0, the maximum free energy values for hapten-Ab bonding, are equivalent such that $N(\Delta G^\circ) = W(\Delta G^\circ)$, thus satisfying $\sigma = 2\sqrt{\pi}(\cot \pi a/2)$.

The extent of affinity inequality among Ab subpopulations in each of the purified serum Ab preparations, which are described by the Gaussian distributions, can be measured as a ratio between the average bond strength of the strongest 5% subset and that of the weakest 5% subset. For each serum sample with average equilibrium constant (K_0) and Gaussian heterogeneity index (σ), the weighted average of equilibrium binding constant (a) of the weakest 5% Ab was

calculated as

$$\frac{\int_{10^4}^{a_{5th\ percentile}} ae^{-(a-K_0)^2/\sigma^2}\ da}{\int_{10^4}^{a_{5th\ percentile}} e^{-(a-K_0)^2/\sigma^2}\ da}$$

where Abs with affinity lower than $a_{5th\ percentile}$ correspond to the lowest 5% of the population.

Similarly, the average of the strongest 5% population was calculated as

$$\frac{\int_{a_{95th\ percentile}}^{10^{10}} ae^{-(a-K_0)^2/\sigma^2}\ da}{\int_{a_{95th\ percentile}}^{10^{10}} e^{-(a-K_0)^2/\sigma^2}\ da}$$

where Abs with affinity higher than $a_{95th\ percentile}$ correspond to the highest 5% of the population. Equilibrium constants for Ab binding of ligands that most closely approximate the epitope in the immunogen are rarely, if ever, found to be less than 10^4 M^{-1} or greater than 10^{10} M^{-1}, whether measured with monoclonal Abs or heterogeneous populations of purified serum Abs [1,2]. These limits were therefore taken as the boundaries of the truncated Gaussian distribution. The affinity inequalities are essentially unchanged if the boundaries are taken as 10^3–10^{10} M^{-1} instead of 10^4–10^{10} M^{-1}. The 5th and the 95th percentiles of each population were calculated by first transforming the truncated Gaussian distribution to its error function, and then inverting it to calculate the respective percentiles using Matlab software.

Results

Listed in Table 1 are average equilibrium constants (K_0) for the binding of various Ab isolates to ε-DNP-L-lysine. ε-DNP-L-lysine approximates the principal epitope of the relevant immunogen DNP-BGG, bovine-γ-globulin with DNP groups linked to ε-amino groups of multiple lysine residues. The Abs were isolated from serum obtained at various times (2–8 weeks) from 19 rabbits injected once with various amounts of the immunogen in water-in-oil emulsion as "incomplete" Freund's adjuvant. For those that received the smallest dose (5mg, Fig 1A), the mean affinity (K_0) rose progressively and heterogeneity with respect to affinity also increased; Values of *a* in the Sips distribution tended to decrease and Gaussian σ to increase. However, the mean affinity (K_0) increased faster. Hence, the coefficient of variance (σ/K_0 ratio) decreased progressively over the 2–8 week period. Correspondingly, the ratios between the strongest and the weakest binding subsets also decreased: at the earliest time (2 weeks) this ratio was largest (the highest affinity subsets exhibited ≈13-55-fold higher K_0 than that of the weakest subsets) and it then fell progressively to where at 8 weeks the ratio was close to 1.0. Thus, although the affinity heterogeneity index (*a* or σ) tended to increase, affinity inequality decreased over time indicating a decreasing effective diversity in Abs.

When higher doses (50, 100 or 250mg) of the same immunogen were used to initiate responses, the mean K_0 values were also initially low and affinity heterogeneity values were high (Table 2). But in contrast to the responses to the 5mg dose, in 13 of 14 rabbits that received the higher doses, these initial parameters did not change over the 2–8 week period. It is especially notable that affinity inequality remained high and changed little over at least 8 weeks. It is possible that the one exception, rabbit # 9, may not have received the full 50 mg dose.

Discussion

In view of the wide (≈million-fold) range of affinity values and the affinity heterogeneity indices previously measured for Ab-Ag and Ab-hapten reactions [1,2], one may have expected that

Table 1. Affinity inequality of antibodies after single injection of 5mg dose of the immunogen in rabbits.

Dose (mg)	RabbitNo.	Weeks fromantigeninjection	Avg. equilibriumconstant (K_0)(M^{-1} X 10^6)	Heterogeneity index		Coefficient ofvariation ($\frac{\sigma}{K_0}$)	Average ofthe lowest 5thpercentile (M^{-1} X 10^6)	Average ofthe highest 5th percentile (M^{-1} X 10^6)	Affinityinequality
				Sips (a) (M^{-1} X 10^6)	Gaussian ($\sigma^{*\dagger}$)(M^{-1} X 10^6)				
5	#1	2	**0.60**	0.70	1.8	**3.0**	0.070	3.4	**49**
"		5	**32**	0.50	3.5	**0.11**	27	37	**1.4**
5	#2	2	**1.6**	0.70	1.8	**1.1**	0.15	4.3	**29**
"		5	**27**	0.70	1.8	**0.067**	24	30	**1.3**
5	#3	2	**0.32**	0.70	1.8	**5.6**	0.059	3.2	**54**
"		5	**1.6**	0.60	2.6	**1.6**	0.14	5.5	**39**
"		8	**20**	0.30	7.0	**0.35**	9.9	30	**3.0**
5	#4	2	**1.0**	0.80	1.2	**1.2**	0.10	2.7	**27**
"		5	**5.9**	0.60	2.6	**0.44**	2.2	9.7	**4.4**
"		8	**250**	0.30	7.0	**0.028**	240	260	**1.1**
5	#5	2	**0.78**	0.90	0.56	**0.72**	0.12	1.6	**13**
"		5	**1.5**	0.50	3.5	**2.4**	0.14	7.0	**50**
"		8	**80**	0.50	3.5	**0.044**	75	85	**1.1**

* σ derived from sips to Gaussian distribution transformation by Karush (1964) equation

† Standard deviation is σ/ $\sqrt{2}$

Antibodies were isolated from serum after single injection of 2, 4-DNP-bovine-γ-globulin as incomplete Freund's adjuvant in rabbits. Numbers are rounded up to two significant figures (except for data from rabbit # 4 at 8 weeks).

doi:10.1371/journal.pone.0139222.t001

the affinity values of Abs to a defined epitope, isolated from a given serum sample, would also extend over a wide range, and indeed they do. But, importantly, the affinity inequality, measured as the ratio of the strongest and weakest binding Ab subsets, exhibits a somewhat unexpected behavior. In a comprehensive set of changes over time in response to various Ag doses, the affinity of the strongest subset was only about 10-to 30- fold higher than the weakest in early bleeds taken two weeks after a small dose of Ag injection. This limited inequality then decreased progressively to the point where at eight weeks after immunization the difference between these subsets was essentially not detectable. The observed progressive reduction in the effective affinity inequality is in accord with the view that GC B cells are in competition for limiting amounts of Ag, and evolution occurs in the strong selection limit. Note, however, that from the data (collected almost 50 years ago) we cannot comment on the evolution of the heterogeneity of clones during GC reactions. Following low Ag dose (e.g. 5mg in Table 1, Fig 1A), B cells with high affinity BCR are preferentially selected over B cells with low affinity BCR to mature into Ab-secreting plasma cells and memory B cells. For high Ag doses, there is a surfeit of Ag, and even B cells with low affinity BCR can also be stimulated to internalize sufficient antigen and receive T cell help from a potentially less limiting amount of cognate T helper cells. Thus, most B cells can mature into Ab producing plasma cells and memory B cells. High Ag dose corresponds to conditions of weak selection. While there is a dearth of information on how the actual levels of Ag on FDC depends upon dose, the two quantities are likely to be

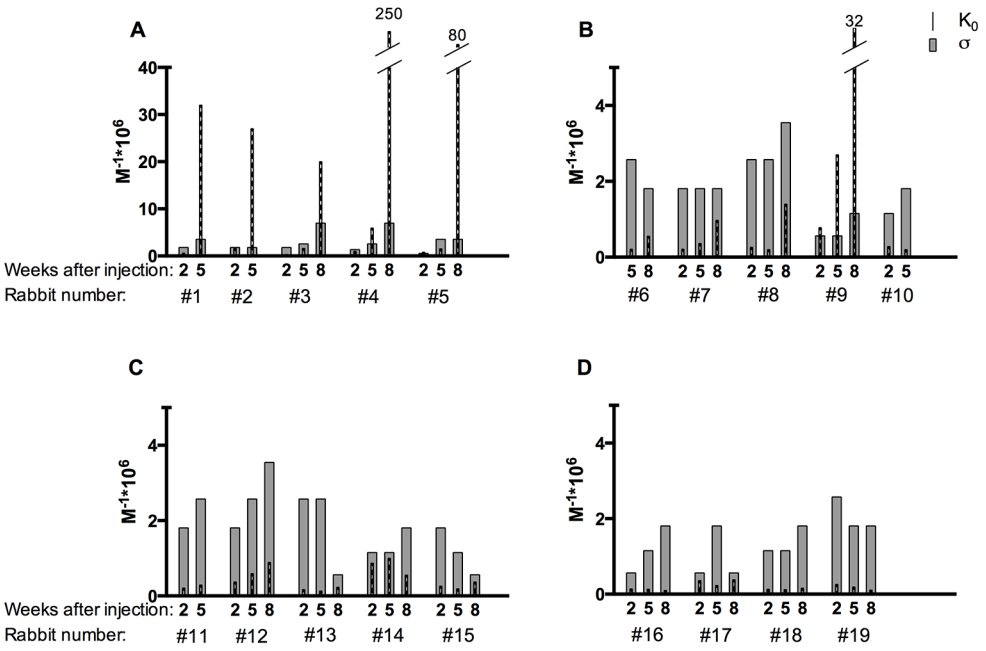


Fig 1. Mean affinity and Gaussian deviation of antibodies over time. (A) Mean affinity (K_0) and Gaussian deviation (σ) (calculated from Sips to Gaussian transformation, Karush (1964)) of antibodies isolated from serum 2–8 weeks after injection of 5mg of the immunogen (2,4-DNP-bovine-γ-globulin) as incomplete Freund's adjuvant in rabbits. (B) As in (A), except 50mg injection dose. (C) As in (A), except 100mg injection dose. (D) As in (A), except 250mg injection dose.

doi:10.1371/journal.pone.0139222.g001

proportional. In Table 1, Table 2 and Fig 1 the dose effect is strikingly clear because the same Ag was given to all rabbits in the same way.

Though high Ag doses, like those in Table 2, are not now used in clinical vaccines, and rarely used experimentally, they are nevertheless of interest because they provide a glimpse of the range of affinities of BCR on the naïve B cells that initially respond to immunogens. Thus in rabbits that received 100–250mg Ag (Table 2), the affinities of their serum Abs were on average $\approx 10^5$ M^{-1} and ranged from $\approx 10^4$–10^5 M^{-1} at the lower level to $\approx 10^6$ M^{-1} at the upper level. Values in the former range are similar to those reported for IgM monoclonal antibodies whose intrinsic univalent affinities have been carefully measured [8,34]. These low affinity levels may well be characteristic of BCRs on the naïve IgM+ naïve B cells that are first activated by immunogens and initiate formation of GC [17,20].

In the haptenated proteins used as immunogens for studies of Ab-hapten interactions and affinity maturation, there typically are many haptenic groups per protein molecule (e.g. ≈50 DNP groups attached to a protein molecule with ≈70 lysine residues). It has therefore been suggested that the haptenated epitopes are actually diverse, despite having a common haptenic group, and that this diversity could account for the affinity heterogeneity observed with a simple ligand, such as ε-DNP-L-lysine, which only approximates the actual epitopes. This possibility has been evaluated with an immunogen, ε-41-DNP-ribonuclease, in which the DNP group was attached to the epsilon amino group of lysine 41; Abs from rabbits immunized with this mono-epitope Ag were just as heterogeneous with respect to affinity for ε-DNP-L-lysine as those Abs elicited with conventional haptenated Ags [35]. The Abs elicited with another single-epitope immunogen, DNP-lysine attached to the single SH group of papain, were also heterogeneous in terms of their diverse L chains, though their affinities were not determined [36].

Table 2. Affinity inequality of antibodies after injection of high dose of the immunogen in rabbits.

Dose (mg)	Rabbit No.	Weeks from antigen injection	Avg. equilibrium constant (K_0) (M^{-1} X 10^6)	Heterogeneity index		Coefficient of variation ($\frac{\sigma}{K_0}$)	Average of the lowest 5th percentile (M^{-1} X 10^6)	Average of the highest 5th percentile (M^{-1} X 10^6)	Affinity inequality
				Sips(a) *(M^{-1} X10^6)*	Gaussian ($\sigma^{*\dagger}$) *(M^{-1} X 10^6)*				
50	#6	5	0.21	*0.60*	2.6	**12**	0.072	4.4	**61**
"		8	0.55	0.70	1.8	**3.3**	0.067	3.4	**51**
50	#7	2	0.21	0.70	1.8	**8.6**	0.055	3.1	**56**
"		5	0.36	0.70	1.8	**5.0**	0.060	3.2	**53**
"		8	0.97	0.70	1.8	**1.9**	0.090	3.7	**41**
50	#8	2	0.26	0.60	2.6	**9.9**	0.074	4.4	**59**
"		5	0.20	0.60	2.6	**13**	0.072	4.4	**61**
"		8	1.4	0.50	3.5	**2.5**	0.14	6.9	**49**
50	#9	2	0.78	0.90	0.56	**0.72**	0.12	1.6	**13**
"		5	2.7	0.90	0.56	**0.21**	1.9	3.5	**1.8**
"		8	32	0.80	1.2	**0.036**	30	34	**1.1**
50	#10	2	0.28	0.80	1.2	**4.1**	0.044	2.1	**48**
"		5	0.20	0.70	1.8	**9.0**	0.055	3.1	**56**
100	# 11	2	0.21	0.70	1.8	**8.6**	0.055	3.1	**56**
"		5	0.29	0.60	2.6	**8.9**	0.075	4.5	**60**
100	#12	2	0.37	0.70	1.8	**4.9**	0.060	3.2	**53**
"		5	0.59	0.60	2.6	**4.4**	0.085	4.7	**55**
"		8	0.89	0.50	3.5	**4.0**	0.11	6.4	**58**
100	#13	2	0.17	*0.60*	2.6	**15**	0.071	4.4	**62**
"		5	0.13	0.60	2.6	**20**	0.070	4.4	**63**
"		8	0.23	0.90	0.56	**2.4**	0.030	1.2	**40**
100	#14	2	0.87	0.80	1.2	**1.3**	0.081	2.7	**33**
"		5	1.0	0.80	1.2	**1.2**	0.096	2.8	**29**
"		8	0.55	0.70	1.8	**3.3**	0.067	3.4	**51**
100	#15	2	0.26	0.70	1.8	**6.9**	0.057	3.2	**56**
"		5	0.19	0.80	1.2	**6.1**	0.041	2.1	**51**
"		8	0.37	0.90	0.56	**1.5**	0.039	1.3	**33**
250	# 16	2	0.14	0.90	0.56	**4.0**	0.027	1.1	**41**
"		5	0.13	0.80	1.2	**8.9**	0.040	2.1	**53**
"		8	0.10	0.70	1.8	**18**	0.052	3.1	**60**
250	#17	2	0.36	0.90	0.56	**1.6**	0.039	1.3	**33**
"		5	0.23	0.70	1.8	**7.9**	0.056	3.2	**57**
"		8	0.38	0.90	0.56	**1.5**	0.040	1.3	**33**
250	#18	2	0.13	0.80	1.2	**8.9**	0.040	2.1	**53**
"		5	0.12	0.80	1.2	**9.6**	0.039	2.1	**54**
"		8	0.16	0.70	1.8	**11**	0.054	3.1	**57**
250	#19	2	0.26	0.60	2.6	**9.9**	0.074	4.5	**61**
"		5	0.19	0.70	1.8	**9.5**	0.055	3.1	**56**
"		8	0.11	0.70	1.8	**16**	0.053	3.1	**58**

* σ derived from sips to Gaussian distribution transformation by Karush (1964) equation

† Standard deviation is σ/ $\sqrt{2}$

Antibodies were isolated from serum after single injection of 2, 4-DNP-bovine-γ-globulin in 50, 100 or 250mg as incomplete Freund's adjuvant in rabbits. Numbers are rounded up to two significant figures.

doi:10.1371/journal.pone.0139222.t002

It is important to note that affinity has its limits as a measure of diversity: Ab molecules with the same affinity for a given ligand could differ markedly in on- and off-rates and have different paratopes and cross-reactivities with structural variants of a ligand. In contrast to the historic data presented here, which was collected by equilibrium dialysis, recent techniques such as mutation analyses in Ig genes and Ig repertoire sequencing can trace the breadth of Abs raised against Ag as the immune response progresses, quantifying how the diversity of mutations correlates with affinity [37]. Thus, the micro-evolution in GC towards less affinity inequality seen in responses to low Ag dose (Table 1, Fig 1) tells us more about forces driving B cell selection in GCs than about the extent of cross reactions and protective efficacy of affinity matured serum Abs. Indeed, the broadly-neutralizing monoclonal Abs, such as VRC01, isolated from persons persistently infected with HIV-1, are widely cross-reactive with structural variants of the epitope in many different strains of the virus despite being homogenous with respect to their affinity for a viral epitope [38]. In addition, low intrinsic affinity IgM (immunoglobulin M), which have ten binding sites per molecule, can bind strongly with high avidity to Ags that have multiple repeats of closely spaced cognate epitopes as, for example, on pneumococcal polysaccharides and influenza virus hemagglutinin. These IgM Abs have been found to protect mice against otherwise lethal influenza virus infection [39].

The measure of affinity inequality used here is based upon the assumption that a normal probability distribution function (Sips or Gaussian) is an apt description of the diversity of bond-strengths or equilibrium constants that characterize the binding of epitopes by a heterogeneous population of Ab molecules isolated from serum by immune precipitation or adsorption. This assumption is supported by agreement between theoretical curves based upon normal distribution functions and measured K values [6–8,28]. It does not mean, of course, that other distribution functions may not also be applicable. However, the only report of non-Gaussian distributions of which we are aware is based on Ab affinity measurements made with crude globulin fractions of antisera [40]; the skewed distributions bordering on bimodality may have been due to low affinity Abs produced by B cells that differentiate into Ab secretors outside of GCs, or even perhaps to non-Ig proteins in the crude globulin fractions [41].

It may well be that the affinity diversity of the Abs produced by a small number of plasma cells emerging from a single lymph node or a few GC would not correspond to a normal distribution. Nevertheless, the average affinity values (K_0) considered here (Tables 1 and 2, Fig 1) were obtained by analyzing samples of Ag-specific precipitable Abs that each contained about 10^{20} ~10^{21} purified Ab molecules under approximation that all Abs are 7S-IgG. Thus, these data represent the average of the output of a large number of plasma cells arising from a great many GCs. Given these conditions, normal distributions (Sips or Gaussian) remain the most useful means for describing the affinity diversity of serum Abs that recognize a particular epitope.

Acknowledgments

HNE had many helpful discussions with Dr. Gabriel Victora during the preparation of this manuscript.

Author Contributions

Analyzed the data: MSK TJE EAE AKC HNE. Wrote the paper: MSK TJE EAE AKC HNE.

References

1. Foote J, Eisen HN. Kinetic and affinity limits on antibodies produced during immune responses. Proc Natl Acad Sci U S A. 1995 Feb 28; 92(5):1254–6. PMID: 7877964

2. Batista FD, Neuberger MS. Affinity dependence of the B cell response to antigen: a threshold, a ceiling, and the importance of off-rate. Immunity. 1998 Jun; 8(6):751–9. PMID: 9655489
3. Landsteiner K. On cross reactions of immune sera to azoproteins.J Exp Med. 1936 Mar 1; 63(3):325–39. PMID: 19870475
4. Pauling L, Pressman D, Grossberg AL. The Serological Properties of Simple Substances. VII. A Quantitative Theory of the Inhibition by Haptens of the Precipitation of Heterogeneous Antisera with Antigens, and Comparison with Experimental Results for Polyhaptenic Simple Substances and for Azop. J Am Chem Soc. American Chemical Society; 1944 May; 66(5):784–92.
5. Eisen HN, Karush F. The Interaction of Purified Antibody with Homologous Hapten. Antibody Valence and Binding Constant. J Am Chem Soc. American Chemical Society; 1949 Jan; 71(1):363–4. PMID: 18108974
6. Nisonoff A, Pressman D. Heterogeneity and average combining constants of antibodies from individual rabbits. J Immunol. American Association of Immunologists; 1958 Jun 1; 80(6):417–28. PMID: 13549727
7. Karush F. The Interaction of Purified Antibody with Optically Isomeric Haptens 1,2. J Am Chem Soc. American Chemical Society; 1956 Nov; 78(21):5519–26.
8. Karush F, Marks R. The preparation and properties of purified anti-hapten antibody. J Immunol. American Association of Immunologists; 1957 Apr 1; 78(4):296–303. PMID: 13429102
9. Gray EE, Cyster JG. Lymph node macrophages. 2013;4:424–36.
10. Rajewsky K. Clonal selection and learning in the antibody system. Nature. 1996; 381:751–8. PMID: 8657279
11. Nieuwenhuis P, Opstelten D. Functional anatomy of germinal centers. Am J Anat. 1984; 170:421–35. PMID: 6383007
12. Kroese FG, Wubbena a S, Seijen HG, Nieuwenhuis P. Germinal centers develop oligoclonally. Eur J Immunol. 1987 Jul; 17(7):1069–72. PMID: 3301368
13. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell. 2000; 102:553–63. PMID: 11007474
14. Mandel TE, Phipps RP, Abbot A, Tew JG. The follicular dendritic cell: long term antigen retention during immunity. Immunol Rev. 1980; 53:29–59. PMID: 6162778
15. Chen LL, Frank AM, Adams JC, Steinman RM. Distribution of horseradish peroxidase (HRP)-anti-HRP immune complexes in mouse spleen with special reference to follicular dendritic cells. J Cell Biol. 1978; 79:184–99. PMID: 701370
16. Tew JG, Phipps RP, Mandel TE. The maintenance and regulation of the humoral immune response: persisting antigen and the role of follicular antigen-binding dendritic cells as accessory cells. Immunol Rev. 1980; 53:175–201. PMID: 7009404
17. Victora GD, Nussenzweig MC. Germinal centers. Annu Rev Immunol. 2012 Jan; 30:429–57. doi: 10.1146/annurev-immunol-020711-075032 PMID: 22224772
18. O'Neill SK, Veselits ML, Zhang M, Labno C, Cao Y, Finnegan A, et al. Endocytic sequestration of the B cell antigen receptor and toll-like receptor 9 in anergic cells. Proc Natl Acad Sci U S A. 2009 Apr 14; 106(15):6262–7. doi: 10.1073/pnas.0812922106 PMID: 19332776
19. Pape KA, Catron DM, Itano AA, Jenkins MK. The humoral immune response is initiated in lymph nodes by B cells that acquire soluble antigen directly in the follicles. Immunity. 2007 Apr; 26(4):491–502. PMID: 17379546
20. Allen CDC, Okada T, Tang HL, Cyster JG. Imaging of germinal center selection events during affinity maturation. Science. 2007 Jan 26; 315(5811):528–31. PMID: 17185562
21. Liu YJ, Joshua DE, Williams GT, Smith CA, Gordon J, MacLennan IC. Mechanism of antigen-driven selection in germinal centres. Nature. Jan; 342(6252):929–31.
22. Meyer-Hermann M, Mohr E, Pelletier N, Zhang Y, Victora GD, Toellner K-M. A theory of germinal center B cell selection, division, and exit. Cell Rep. 2012 Jul 26; 2(1):162–74. doi: 10.1016/j.celrep.2012.05.010 PMID: 22840406
23. Shulman Z, Gitlin AD, Targ S, Jankovic M, Pasqual G, Nussenzweig MC, et al. T follicular helper cell dynamics in germinal centers. Science. 2013; 341:673–7. doi: 10.1126/science.1241680 PMID: 23887872
24. Meyer-Hermann ME, Maini PK, Iber D. An analysis of B cell selection mechanisms in germinal centers. Math Med Biol. 2006 Sep; 23(3):255–77. PMID: 16707510
25. Zhang J, Shakhnovich EI. Optimality of mutation and selection in germinal centers. PLoS Comput Biol. 2010 Jun; 6(6):e1000800. doi: 10.1371/journal.pcbi.1000800 PMID: 20532164

26. Meyer-Hermann M, Deutsch a, Or-Guil M. Recycling probability and dynamical properties of germinal center reactions. J Theor Biol. 2001 Jun 7; 210(3):265–85. PMID: 11397129

27. Shannon M, Mehr R. Reconciling repertoire shift with affinity maturation: the role of deleterious mutations. J Immunol. 1999 Apr 1; 162(7):3950–6. PMID: 10201914

28. Eisen HN, Siskind GW. Variations in affinities of antibodies during the immune response. Biochemistry. 1964; 3:996–1008. PMID: 14214095

29. Fujio H, Karush F. Antibody Affinity. II. Effect of Immunization Interval on Antihapten Antibody in the Rabbit*. Biochemistry. American Chemical Society; 1966 Jun 1; 5(6):1856–63. PMID: 5963428

30. Parker CW, Yoo TJ, Johnson MC, Godt SM. Fluorescent Probes for the Study of the Antibody-Hapten Reaction. I. Binding of the 5-Dimethylaminonaphthalene-1-sulfonamido Group by Homologous Rabbit Antibody*. Biochemistry. American Chemical Society; 1967 Nov 1; 6(11):3408–16. PMID: 6073028

31. Karush F, Sonenberg M. Long Chain Alkyl Sulfates. Anal Chem. American Chemical Society; 1950 Jan; 22(1):175–7.

32. Sips R. On the Structure of a Catalyst Surface. J Chem Phys. AIP Publishing; 1948 Dec 22; 16(5):490.

33. Karush F. Advances in Immunology Volume 2. Advances in Immunology. Elsevier; 1963. 1–40 p.

34. Fields BA, Goldbaum FA, Dall'Acqua W, Malchiodi EL, Cauerhff A, Schwarz FP, et al. Hydrogen bonding and solvent structure in an antigen-antibody interface. Crystal structures and thermodynamic characterization of three Fv mutants complexed with lysozyme. Biochemistry. American Chemical Society; 1996 Dec 3; 35(48):15494–503.

35. Eisen HN. The immune response to a simple antigenic determinant. Harvey Lect. UNITED STATES; 1966; 60:1–34. PMID: 5338066

36. Brenneman L, Singer SJ. The generation of antihapten antibodies with electrophoretically homogeneous L chains. Proc Natl Acad Sci U S A. 1968 May; 60(1):258–64. PMID: 5241528

37. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat Biotechnol. 2014 Feb; 32(2):158–68. doi: 10.1038/nbt.2782 PMID: 24441474

38. Klein F, Diskin R, Scheid JF, Gaebler C, Mouquet H, Georgiev IS, et al. Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. Cell. 2013; 153:126–38. doi: 10.1016/j.cell.2013.03.018 PMID: 23540694

39. Throsby M, van den Brink E, Jongeneelen M, Poon LLM, Alard P, Cornelissen L, et al. Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. PLoS One. 2008 Jan; 3(12):e3942. doi: 10.1371/journal.pone.0003942 PMID: 19079604

40. Werblin TP, Siskind GW. Distribution of antibody affinities: Technique of measurement. Immunochemistry. 1972 Oct; 9(10):987–1011. PMID: 4117541

41. Boyd WC, Bernard H. Quantitative Changes in Antibodies and Globulin Fractions in Sera of Rabbits Injected with Several Antigens. J Immunol. American Association of Immunologists; 1937 Aug 1; 33 (2):111–22.

Appendix B. A Subset of Platinum-Containing Chemotherapeutic Agents Kills Cells by Inducing Ribosome Biogenesis Stress

Peter M. Bruno[1,2,] Yunpeng Liu[1,2,] Ga Young Park[3], Junko Murai[4], Catherine E. Koch[1,2], Timothy J. Eisen[2,5], Justin R. Pritchard[1,2], Yves Pommier[4], Stephen J. Lippard[1,3] & Michael T. Hemann[1,2]

[1]The Koch Institute for Integrative Cancer Research at MIT, Cambridge, Massachusetts, USA.
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[3]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
[4]Developmental Therapeutics Branch and Laboratory of Molecular Pharmacology, Center for Cancer Research, National Institutes of Health, Bethesda, Maryland, USA.
[5]Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA.
Correspondence should be addressed to M.T.H. (hemann@mit.edu) or S.J.L. (lippard@mit.edu).

Published as: Bruno, P.M., Liu, Y., Park, G.Y., Murai, J., Koch, C.E., Eisen, T.J., Pritchard, J.R., Pommier, Y., Lippard, S.J., and Hemann, M.T. (2017). A subset of platinum-containing chemotherapeutic agents kills cells by inducing ribosome biogenesis stress. Nat Med 23, 461-471.

nature medicine

A subset of platinum-containing chemotherapeutic agents kills cells by inducing ribosome biogenesis stress

Peter M Bruno[1,2], Yunpeng Liu[1,2], Ga Young Park[3], Junko Murai[4], Catherine E Koch[1,2], Timothy J Eisen[2,5], Justin R Pritchard[1,2], Yves Pommier[4], Stephen J Lippard[1,3] & Michael T Hemann[1,2]

Cisplatin and its platinum analogs, carboplatin and oxaliplatin, are some of the most widely used cancer chemotherapeutics. Although cisplatin and carboplatin are used primarily in germ cell, breast and lung malignancies, oxaliplatin is instead used almost exclusively to treat colorectal and other gastrointestinal cancers. Here we utilize a unique, multi-platform genetic approach to study the mechanism of action of these clinically established platinum anti-cancer agents, as well as more recently developed cisplatin analogs. We show that oxaliplatin, unlike cisplatin and carboplatin, does not kill cells through the DNA-damage response. Rather, oxaliplatin kills cells by inducing ribosome biogenesis stress. This difference in drug mechanism explains the distinct clinical implementation of oxaliplatin relative to cisplatin, and it might enable mechanistically informed selection of distinct platinum drugs for distinct malignancies. These data highlight the functional diversity of core components of front-line cancer therapy and the potential benefits of applying a mechanism-based rationale to the use of our current arsenal of anti-cancer drugs.

The use of cisplatin in the clinic began more than 45 years ago in the absence of understanding of the cellular and molecular mechanisms that underlie its efficacy[1]. Nonetheless, cisplatin has become a component of treatment regimens for at least 18 distinct tumor types[2]. However, cisplatin-induced side effects and the emergence of resistance to treatment led to the development of two derivatives, carboplatin and oxaliplatin, which have also seen considerable clinical use for a wide array of cancers. Interestingly, oxaliplatin has a side-effect profile that differs from those of cisplatin and carboplatin, and it is used in colorectal and other gastrointestinal cancers, against which cisplatin and carboplatin have minimal efficacy. The decision to use oxaliplatin to treat colorectal cancer was motivated primarily by its activity against colorectal cancer cell lines, rather than because of any rationale involving its mechanism of action[3,4]. Although the assumption has been that oxaliplatin, as does cisplatin, kills cells by eliciting a DNA-damage response, no satisfactory explanation for the unique clinical use and side-effect profile of oxaliplatin has been identified. Here we demonstrate that oxaliplatin acts through a fundamentally distinct mechanism of action relative to that of cisplatin, and we propose that these agents should be used in a mechanism-targeted manner for the treatment of cancer.

RESULTS

Diverse mechanisms of action for platinum compounds

To examine the mechanism of action of cisplatin and its platinum analogs, we used an RNAi-based functional genetic strategy to predict the mechanism of cytotoxic drug action[5–7]. This methodology has the advantages of being mammalian, isogenic and unbiased by dosage effects resulting from export or metabolism. Additionally, it has previously been used to characterize the mechanism of action of other metal-based anti-cancer agents[8–15]. It is based on a fluorescence competition assay that uses lymphoma cells partially infected with eight short hairpin RNAs (shRNAs) that target distinct genes encoding proteins with known or putative roles in cell-death signaling pathways: p53 (*Trp53*), Chk2 (*Chek2*), Chk1 (*Chek1*), ATR (*Atr*), ATX (*Smg1*), DNAPKcs (*Prkdc*), Bok (*Bok*) and Bim (*Bcl2l11*). The shRNA-bearing cells either enrich or deplete relative to the uninfected population, depending on the survival advantage or disadvantage conferred by a given shRNA (**Fig. 1a**). The combined responses of these cells to different drugs constitute drug 'signatures.' Signatures of all classes of clinically used cytotoxic agents have been generated and assembled into a reference set separated into eight distinct drug categories, which are based on the constituents' shared molecular mechanism of action (**Supplementary Table 1**). A new drug signature can then be classified by a probabilistic K–nearest neighbors algorithm to determine whether a drug belongs to a class in the reference set or requires a new category not represented therein (**Fig. 1b**).

To eliminate dosage or potency effects from confounding the RNAi signatures, all agents were administered at a concentration that killed 80–90% (lethal dose $(LD)_{80–90}$) of the cells at 48 h. $LD_{80–90}$ concentrations varied greatly from one compound to the next (**Supplementary Fig. 1a**). However, through atomic-absorption spectroscopy, we determined that, for cisplatin analogs representing low, medium and high potency, the

[1]The Koch Institute for Integrative Cancer Research at MIT, Cambridge, Massachusetts, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [3]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [4]Developmental Therapeutics Branch and Laboratory of Molecular Pharmacology, Center for Cancer Research, National Institutes of Health, Bethesda, Maryland, USA. [5]Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. Correspondence should be addressed to M.T.H. (hemann@mit.edu) or S.J.L. (lippard@mit.edu).

Received 8 August 2016; accepted 23 January 2017; published online 27 February 2017; doi:10.1038/nm.4291

amount of platinum required inside the cells for killing corresponded to their respective $LD_{80–90}$ values (**Supplementary Fig. 1b**). For instance, pyriplatin treatment at $LD_{80–90}$ led to much more intracellular platinum than the other $LD_{80–90}$ treatments, indicating that each molecule of pyriplatin is relatively less toxic. Thus, differences in cellular uptake were controlled effectively by using $LD_{80–90}$ concentrations.

Next, to examine the mechanisms of action of cisplatin, carboplatin, and oxaliplatin, we obtained their RNAi signatures. Cisplatin and carboplatin were both predicted to be DNA cross-linkers (**Fig. 1c**,**d**, **Supplementary Table 2**, and **Supplementary Fig. 2a**). Previously, the three US Food and Drug Administration (FDA)-approved platinum-based drugs were understood to function primarily as DNA-damaging agents that form intra- and inter-strand cross-links. Intrastrand cross-links are removed largely by excision repair, whereas interstrand cross-links require homologous recombination (HR) and interstrand cross-link (ICL) repair. The latter generate single- or double-stranded breaks in the process[16]. However, despite the fact that oxaliplatin forms 1,2-intra-strand and other cross-links on DNA, like cisplatin and carboplatin, we found that oxaliplatin was most similar to compounds that inhibit transcription or translation (transcription–translation inhibitors) (**Fig. 1c**,**d**, **Supplementary Table 2** and **Supplementary Fig. 2a**). These results may begin to explain why oxaliplatin has a different cytotoxicity profile and clinical application than cisplatin or carboplatin. Additionally, RNAi signatures of cisplatin and oxaliplatin, when taken in combination with 5-fluorouracil (5-FU)—the primary drug with which oxaliplatin is paired—preserved these mechanistic differences (**Supplementary Fig. 3**). Prior analysis of NCI-60 data similarly concluded that oxaliplatin acts in a manner distinct from that of cisplatin and carboplatin, although no cellular function was identified as being responsible for the difference[17]. Notably, in this prior study, compounds clustered strictly by structure, which is indicative of the NCI-60 methodology's emphasis on drug metabolism and transport over mechanism of action[18].

Interestingly, phenanthriplatin, a monofunctional and highly potent platinum(II) compound, also classified as a transcription–translation inhibitor (**Fig. 1c**,**d**, **Supplementary Table 2** and **Supplementary Fig. 2c**). Because phenanthriplatin is incapable of making DNA cross-links, yet also classifies as a transcription and translation inhibitor, like oxaliplatin, the ability of oxaliplatin to form cross-links on the DNA might be irrelevant to its mechanism of action. We went on to characterize seven additional platinum compounds and discovered that most also classified as DNA cross-linkers or transcription–translation inhibitors (**Supplementary Fig. 2b**,**c** and **Supplementary Table 2**). Curiously, two monofunctional platinum agents, acriplatin and pyriplatin, were found to have mechanisms of action not represented in our reference set (**Supplementary Fig. 2d** and **Supplementary Table 2**). This result suggests that potential mechanisms of action for platinum compounds extend beyond the scope of anti-cancer agents in current clinical use. Furthermore, these signature predictions are maintained in all permutations of leave-one-out cross-validation of the drugs in the reference set (**Supplementary Table 3**). Additionally, we used an indicator of structural similarity, the Tanimoto coefficient[19], to cluster the compounds. Hierarchically, we found that structural clustering was unable to recapitulate RNAi-signature-based clustering, regardless of whether the compounds were clustered by their native structure or their anticipated structure once inside the cell (**Supplementary Fig. 4**). Thus, mechanism cannot be correctly predicted on the basis of structure alone.

To more thoroughly examine the differences in RNAi signatures used to classify these molecules, we performed a detailed analysis of all of their signatures. The most notable differences included decreased resistance with shChk2 and decreased sensitivity with shChk1 for the transcription–translation-inhibitor-like compounds relative to the DNA-cross-linker-like compounds (**Fig. 1e** and **Supplementary Table 4**). As another means of visualizing the data, we used principal-component analysis (PCA) to represent the variance of our data in fewer dimensions. After plotting all of our tested platinum analogs with canonical transcription–translation inhibitors, DNA-cross-linking agents and topoisomerase II (Top2) poisons, we saw that the transcription–translation inhibitors separated from DNA cross-linkers along the first principal component (PC1) (**Fig. 1e**). Upon examining the variable contributions that made up PC1, we saw that shChk2 contributed most strongly among the hairpins. We also identified shChk2 as the greatest contributor to the distinction between these two sets of drugs in a $p185^{+}$ *BCR-Abl* $Cdkn2a^{Arf-/-}$ mouse model of acute lymphoblastic leukemia (ALL)[20] (**Supplementary Fig. 2e**,**f**).

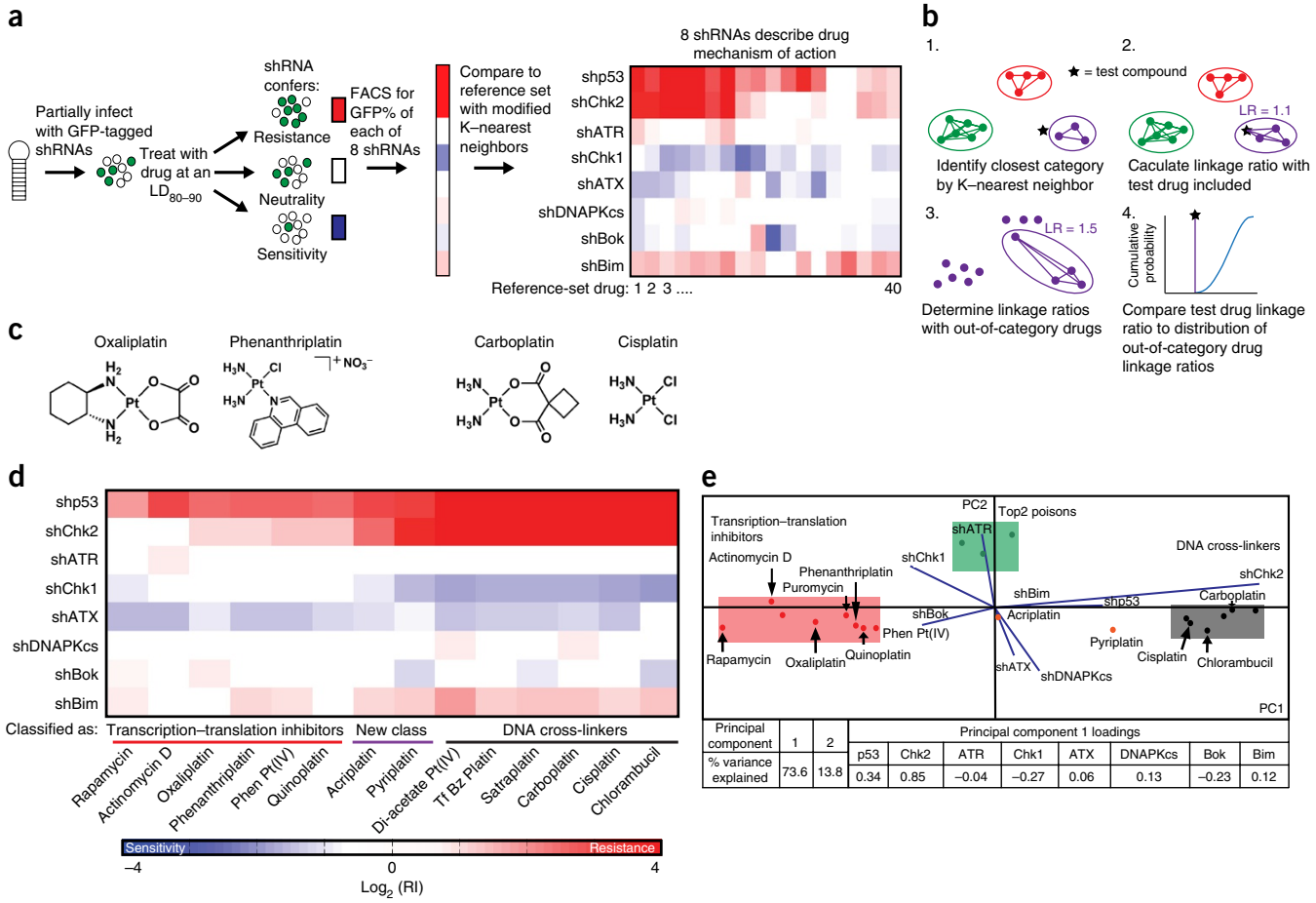
DNA-damage response affects response to cisplatin but not oxaliplatin

To confirm the RNAi-signature data using a parallel approach, we examined drug response in the avian DT40 cell line[21]. Here 40 different DT40 cell lines, each with a different gene knockout related to DNA-damage repair and tolerance, were dosed with five different platinum agents. In agreement with our RNAi signatures, the DT40 knockouts showed distinct sensitivities to oxaliplatin and phenanthriplatin relative to the other three platinum agents (**Fig. 2a**,**b**, and **Supplementary Tables 5** and **6**). In particular, loss of genes involved in HR (*XRCC2*, *XRCC3* and *BRCA2*) and ICR (*FANCC*, *FANCD2* and *FANCG*) showed the greatest differences between the two categories of platinum agents. The relative lack of sensitivity of HR- and ICR-deficient cells to oxaliplatin suggests that it, as with phenanthriplatin, fails to form intra- and inter-strand cross-links. Interestingly, genes necessary for replication bypass (*POLZ* and *PCNA*) were critical for all of the platinum derivatives. This result suggests that oxaliplatin and phenanthriplatin treatment create lesions on the DNA that are toxic only in the absence of normal replication-bypass machinery. In addition, these results were recapitulated using RNAi against several genes related to DNA-damage repair and/or tolerance in *Eμ-Myc* $Cdkn2a^{Arf-/-}$ lymphoma cells and in *Eμ-Myc* $Trp53^{-/-}$ lymphoma cells (**Supplementary Fig. 5a**,**b**). Thus, relative drug sensitivities in the context of DT40 knockout cells support the RNAi-based category classifications.

Dependence on checkpoint kinases stratifies platinum agents

Given the importance of the two cell-cycle checkpoint kinases, Chk1 and Chk2, in discriminating between the two mechanistic classes of platinum drug action and DNA-damage response signaling, we decided to first confirm that these distinctions are relevant *in vivo*. To do this, we conducted a cell-competition experiment using the *Eμ-Myc* $Cdkn2a^{Arf-/-}$ lymphoma cells that were partially infected with GFP-tagged shChk2 and then tail-vein injected into syngeneic recipient mice. Tumors from untreated, cisplatin-, oxaliplatin- or phenanthriplatin-treated mice were then analyzed for GFP percentage. As the *in vitro* data predicted, shChk2-containing cells were significantly enriched as compared to uninfected cells in mice treated with cisplatin, but not in mice treated with oxaliplatin or phenanthriplatin (**Fig. 3a**). These results suggest that dependence on Chk2 activity, a key mediator of the canonical DNA-damage response, represents a primary distinction between the mechanistic classifications of DNA cross-linkers and transcription–translation inhibitors.

Subsequently, we examined the cell-cycle profiles of cells treated with phenanthriplatin, oxaliplatin and cisplatin for 12 h at $LD_{80–90}$.



Principal component	1	2
% variance explained	73.6	13.8

Principal component 1 loadings								
p53	Chk2	ATR	Chk1	ATX	DNAPKcs	Bok	Bim
0.34	0.85	–0.04	–0.27	0.06	0.13	–0.23	0.12

Figure 1 RNAi signatures identify a spectrum of platinum-drug activities. (**a**) A schematic representation of our signature-based approach. *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells are partially infected with GFP-tagged shRNAs targeting the indicated eight genes. The individual pools of eight shRNA-containing cells are treated with drug to achieve 80–90% killing at 48 h. Relative GFP% is measured at 72 h by flow cytometry, and the subsequent shRNA signature is compared to our reference set using our modified K–nearest neighbors algorithm. (**b**) A schematic representation of our modified K–nearest neighbors algorithm. First, the test compound's nearest reference-set category is identified by Euclidian K–nearest neighbor's analysis. Second, the linkage ratio (LR) is determined by dividing the pairwise distances of the category containing the new drug by the category without the new drug. Third, the linkage ratios are then calculated for all out-of-category drugs as if they were members of the category in question. This generates a background distribution of negative-control linkage ratios. Fourth, the linkage ratio of the new drug is compared to the linkage ratio distribution of the negative controls to obtain a *P* value. If $P > 0.05$, then the drug is classified as belonging to a 'new class' and having a mechanism of action not represented in the reference set. (**c**) On the right, cisplatin and carboplatin, two of the three clinically approved platinum agents classify as DNA cross-linkers. On the left, oxaliplatin, the third clinically approved platinum agent, and phenanthriplatin, a monofunctional platinum agent, both classify as transcription–translation inhibitors. (**d**) A heat map showing platinum-compound signatures, with reference compounds rapamycin, actinomycin D and chlorambucil included. Rapamycin and actinomycin D represent transcription–translation inhibitors, whereas chlorambucil represents DNA cross-linkers. The other agents are labeled according to their respective classification. (**e**) Principal-component analysis of the platinum compounds, Top2 poisons, DNA cross-linkers and transcription–translation inhibitors. Tables show the percentage variance explained by each principal component, as well as the principal component 1 loadings depicting the percentage contribution of each of the eight shRNAs for the PCA shown above. Shaded boxes represent the approximate space the category occupies in the PCA and are meant only to aid visualization, not to be used for category classification.

Oxaliplatin and phenanthriplatin induced a G1 cell-cycle arrest, whereas cisplatin arrested cells in the S and G2/M phases (**Fig. 3b**). We obtained similar results with the three drugs after 24 h of treatment in human lung adenocarcinoma and colorectal cell lines, A549 and LoVo, respectively (**Supplementary Fig. 6**). To determine the mechanistic basis for these cell-cycle differences, we examined signaling pathways that may be engaged following the induction of the DNA-damage response. As shown by western blot, p21 (encoded by *CDKN1A*) protein is activated more quickly in response to oxaliplatin and phenanthriplatin than to cisplatin (**Supplementary Fig. 7a**). Additionally, knockdown of p21 sensitizes cells to oxaliplatin and phenanthriplatin treatment but elicits resistance to cisplatin (**Supplementary Fig. 7b**).

Next, to gain further insight into DNA-damage signaling in response to these compounds, we examined the p53-activating kinase, Chk2. Chk2 is activated in response to double-strand breaks, whereby it goes on to phosphorylate p53 on serine 20 (serine 18 in mice), which relieves MDM2 (*MDM2*) inhibition of p53 (ref. 22). Thus, treatment of cells with DNA-damaging agents selects for cells harboring Chk2 or p53 hairpins. Consequently, mechanistic characteristics of DNA-cross-linking agents should be discernible by examining canonical markers of DNA damage. We therefore tested for γ-H2AX (phospho-ser139 of *H2afx*), a canonical marker of DNA damage, and phospho-ser18 p53 by western blot at 12 h, with and without hairpins targeting Chk1 and Chk2, for cisplatin, oxaliplatin and

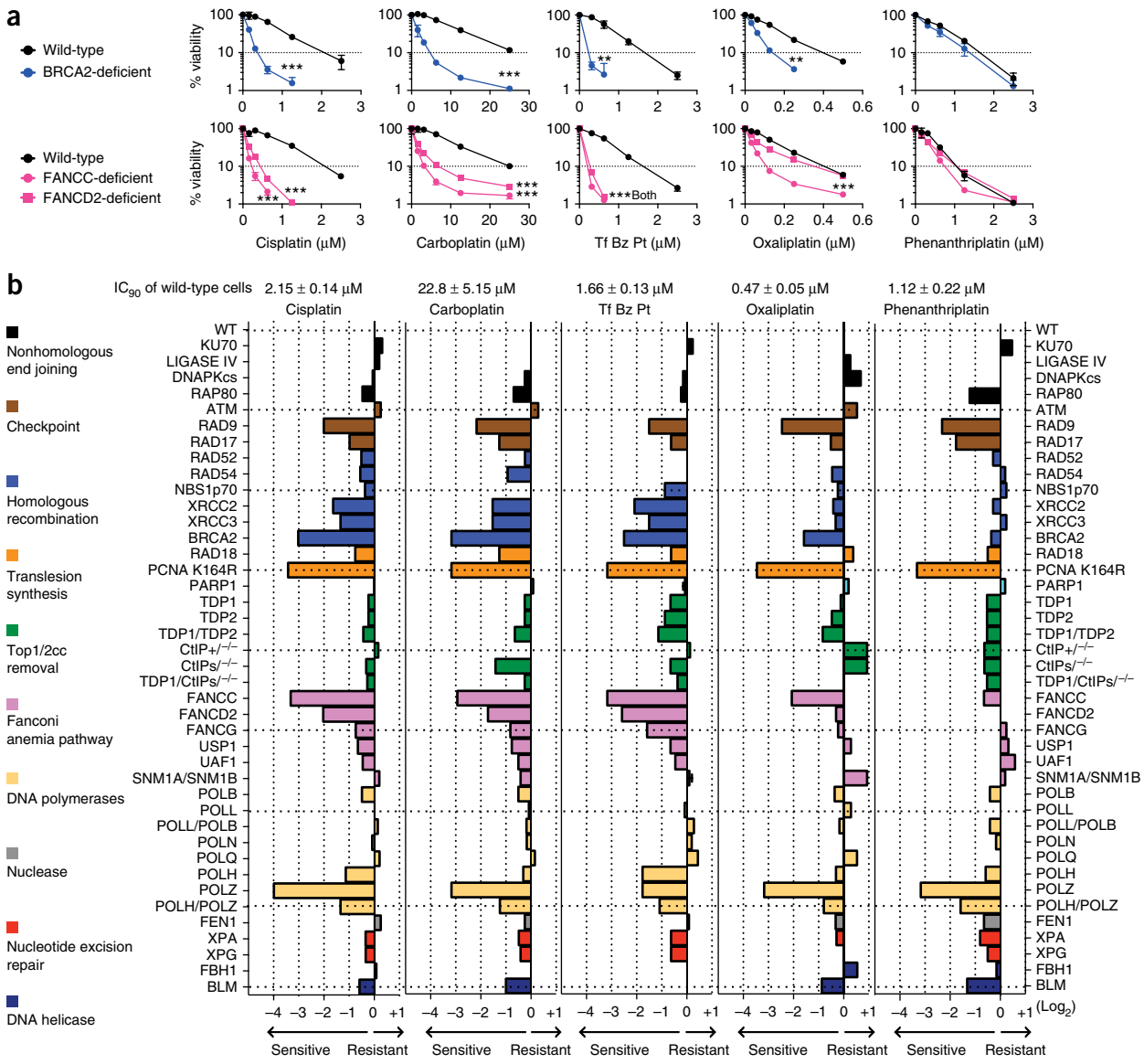


Figure 2 Sensitivity profiles of the indicated platinum drugs on a panel of the repair-deficient DT40 mutants. (**a**) Viability curves of the indicated cell lines after continuous treatment for 72 h with the indicated drugs. Error bars represent the s.d. ($n = 3$). $P < 0.05$ (*), $P < 0.01$ (**) or $P < 0.001$ (***) by unpaired two-tailed Student's *t*-test on the highest doses only. (**b**) Relative sensitivity of all of the DT40 mutant cell lines. A negative score and a positive score indicates that the cells are sensitive and resistant to the drug, respectively. The score is $\log_2$ of the difference between IC_{90} (inhibitory concentration, 90%) values. IC_{90} of wild-type cells are shown at the top of each panel ($n = 8$–10). $n = 3$ for all others. The bars are colored according to the main DNA-repair function of deficient gene(s). Black, nonhomologous end joining (NHEJ); brown, checkpoint; blue, homologous recombination (HR); orange, translesion synthesis (TLS); aqua, PARP1; green, removal of topoisomerase I or topoisomerase II cleavage complex; pink, Fanconi anemia (FA) pathway; light orange, DNA polymerase; gray, nuclease; red, nucleotide excision repair (NER); dark blue, DNA helicase.

phenanthriplatin. We observed that treatment with oxaliplatin and phenanthriplatin resulted in γ-H2AX signal, but that this effect was not dependent on Chk2, as it was for cisplatin (**Fig. 3c**,**e**). All three drugs also elicited total p53 induction and phosphorylation of p53; however, this behavior was dependent on Chk2 for cisplatin (**Fig. 3d**). Additionally, the same was observed for both γ-H2AX and p53 at 4 h, when comparing cisplatin and phenanthriplatin with and without shChk2 (**Supplementary Fig. 8a**,**b**). Moreover, we observed upregulation of the p53 transcriptional targets *Puma* and *Noxa* following phenanthriplatin treatment, concomitant with the increase in p53 levels seen by western blot (**Fig. 3d**,**f**). Subsequently, we examined γ-H2AX and phospho-ser18 p53 at and before 4 h. Phenanthriplatin, and to a lesser degree, oxaliplatin, induced γ-H2AX, phospho-ser18 p53 and total p53 accumulation, sooner and to a greater degree than cisplatin (**Supplementary Fig. 8c**,**d**). Phenanthriplatin and oxaliplatin also caused more rapid cell death than cisplatin and doxorubicin (**Supplementary Fig. 9**). We confirmed that cisplatin-induced phosphorylation of serine 20 of p53 was Chk2-dependent in the human LoVo colorectal cell line (**Supplementary Fig. 10a**). Furthermore, transcript levels of the pro-apoptotic gene *NOXA* were increased following treatment with all platinum agents tested in multiple human cell lines (**Supplementary Fig. 10c**–**e**). Taken together, the early activation of apoptosis, along with the early appearance, persistence and

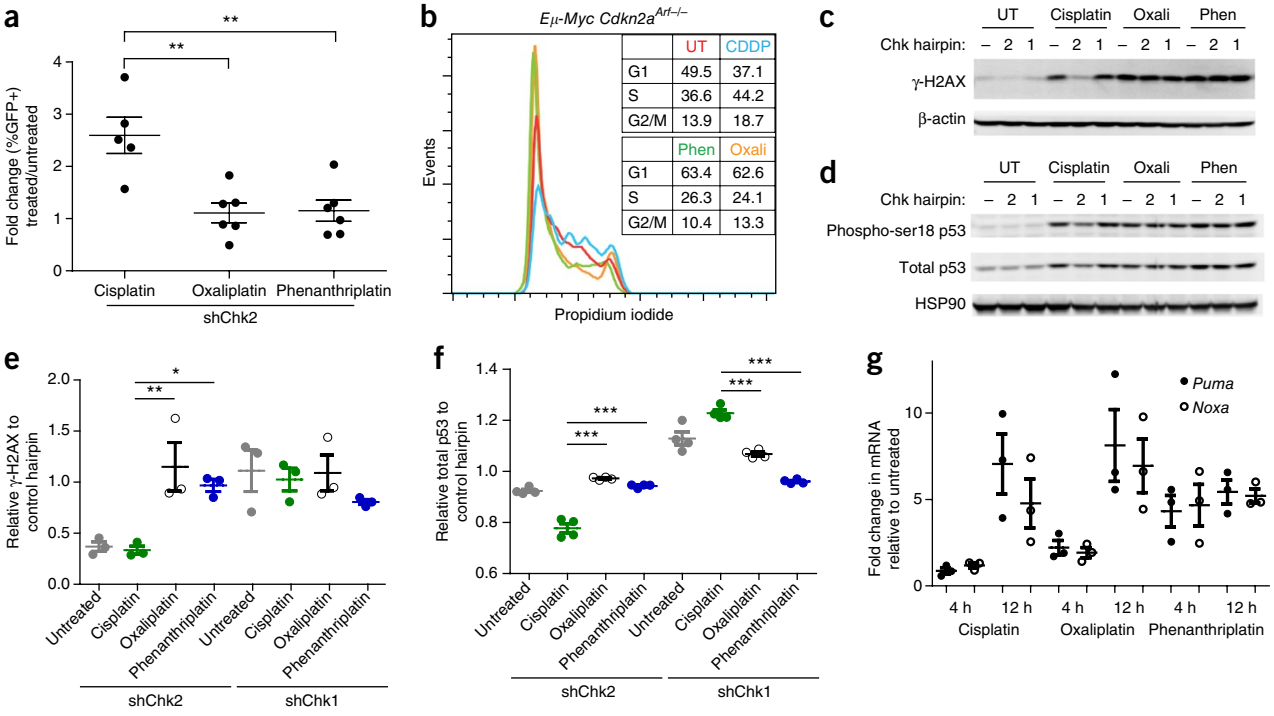


Figure 3 Phenanthriplatin and oxaliplatin exhibit distinct differences from cisplatin in cell-cycle profiles, γ-H2AX and p53 signaling in *Eμ-Myc Cdkn2a*$^{Arf-/-}$ cells. (**a**) Results of a GFP competition assay *in vivo*. Fold change in GFP% was assessed relative to untreated mice after *Eμ-Myc Cdkn2a*$^{Arf-/-}$ tumor cell harvesting. $^{**}P < 0.01$ by a two-tailed Mann–Whitney test. Mean ± s.e.m. is depicted. $n = 5$ for cisplatin and $n = 6$ for oxaliplatin and phenanthriplatin. (**b**) Cell-cycle profiles resulting from 12 h treatment with cisplatin, phenanthriplatin or oxaliplatin. UT, untreated; CDDP, cisplatin; Phen, phenanthriplatin; Oxali, oxaliplatin. (**c**) Western blot for γ-H2AX after cisplatin, oxaliplatin or phenanthriplatin treatment with or without shChk2 or shChk1 at 12 h. (**d**) Western blot for phospho-ser18 and total p53 after cisplatin, oxaliplatin or phenanthriplatin treatment with or without shChk2 or shChk1 at 12 h. (**e**) Densitometry quantification of **c**. Data are mean ± s.e.m. via three independent quantification results. $^{*}P < 0.05$, $^{**}P < 0.01$ for each group relative to cisplatin shChk2 or shChk1 by one-way ANOVA with Dunnett's multiple-comparison test. (**f**) Densitometry quantification of **d**. Data are mean ± s.e.m. via four independent quantification results. $^{***}P < 0.001$ for each group, relative to cisplatin shChk2 or shChk1 by one-way ANOVA with Dunnett's multiple-comparison test. (**g**) Results of qPCR analysis conducted for *Puma* and *Noxa* after 4 h and 12 h of cisplatin, phenanthriplatin or oxaliplatin treatment. Data are represented as mean ± s.e.m. $n = 3$ from independent experiments from independent cultures of cells.

Chk2 independence of γ-H2AX and phospho-ser18 p53, suggest a mechanism of cell death induced by oxaliplatin and phenanthriplatin that does not rely on canonical DNA-strand-break signaling.

Oxaliplatin does not induce a DNA-damage response

Considering the aforementioned rapid induction of γ-H2AX signal and early apoptosis following oxaliplatin and phenanthriplatin treatment, we examined γ-H2AX via immunofluorescence to distinguish DNA-damage-related foci from pan-nuclear γ-H2AX; the latter is indicative of apoptosis[23,24]. We observed, as expected, that cisplatin-treated cells had γ-H2AX foci characteristic of a DNA-damage response. However, similar to untreated cells, oxaliplatin- and phenanthriplatin-treated cells did not have γ-H2AX foci (**Fig. 4a–d**). Quantification of γ-H2AX signal showed that at both 4 h and 8 h, nearly all of the γ-H2AX signal in oxaliplatin- and phenanthriplatin-treated cells was derived from pan-nuclear γ-H2AX (**Fig. 4c**). Additionally, in the human LoVo cell line, we confirmed that oxaliplatin-treated cells had substantially fewer γ-H2AX foci as compared to cisplatin-treated cells (**Supplementary Fig. 10f**).

To further explain the absence of a DNA-damage response following treatment with oxaliplatin and phenanthriplatin, comet assays were performed 6 h after treatment. Comet assays test for the presence of smaller-weight DNA fragments produced following DNA breaks. A neutral comet assay, most sensitive to double-strand-DNA breaks, indicated that oxaliplatin and phenanthriplatin treatment results in significantly fewer double-strand breaks than untreated or cisplatin-treated mouse lymphoma or human breast cancer cells (**Fig. 4e** and **Supplementary Fig. 10g**). Interestingly, an alkaline comet assay, sensitive to both single-strand and double-strand breaks, indicated that only oxaliplatin treatment yielded significantly fewer DNA breaks than either no treatment or cisplatin treatment (**Fig. 4f**). In the alkaline comet assay, S-phase cells yield longer tails than G1 or G2 cells. This result is consistent with having a large proportion of cells in S phase, as in the untreated condition, particularly relative to oxaliplatin and phenanthriplatin treatment[25].

Oxaliplatin and phenanthriplatin induce ribosome biogenesis stress

Our RNAi signatures predicted that the mechanism of action of oxaliplatin and phenanthriplatin would involve transcription or translation inhibition. To further explore this possibility, we first measured the amount of platinum on RNA and DNA after 3 h of cisplatin, oxaliplatin and phenanthriplatin treatment at $LD_{80–90}$. Atomic-absorption spectroscopy revealed that all three compounds were present at appreciable amounts on both nucleic acids (**Fig. 5a**). Interestingly, oxaliplatin treatment yielded the least amount of platinum on either nucleic acid—potentially because it has the lowest $LD_{80–90}$ concentration.

Defects in ribosome biogenesis can rapidly induce cell death in a p53-dependent and DNA-damage-independent manner[26]. Given the similarity between this type of cell death and that induced by oxaliplatin and phenanthriplatin treatment, we examined ribosomal RNA

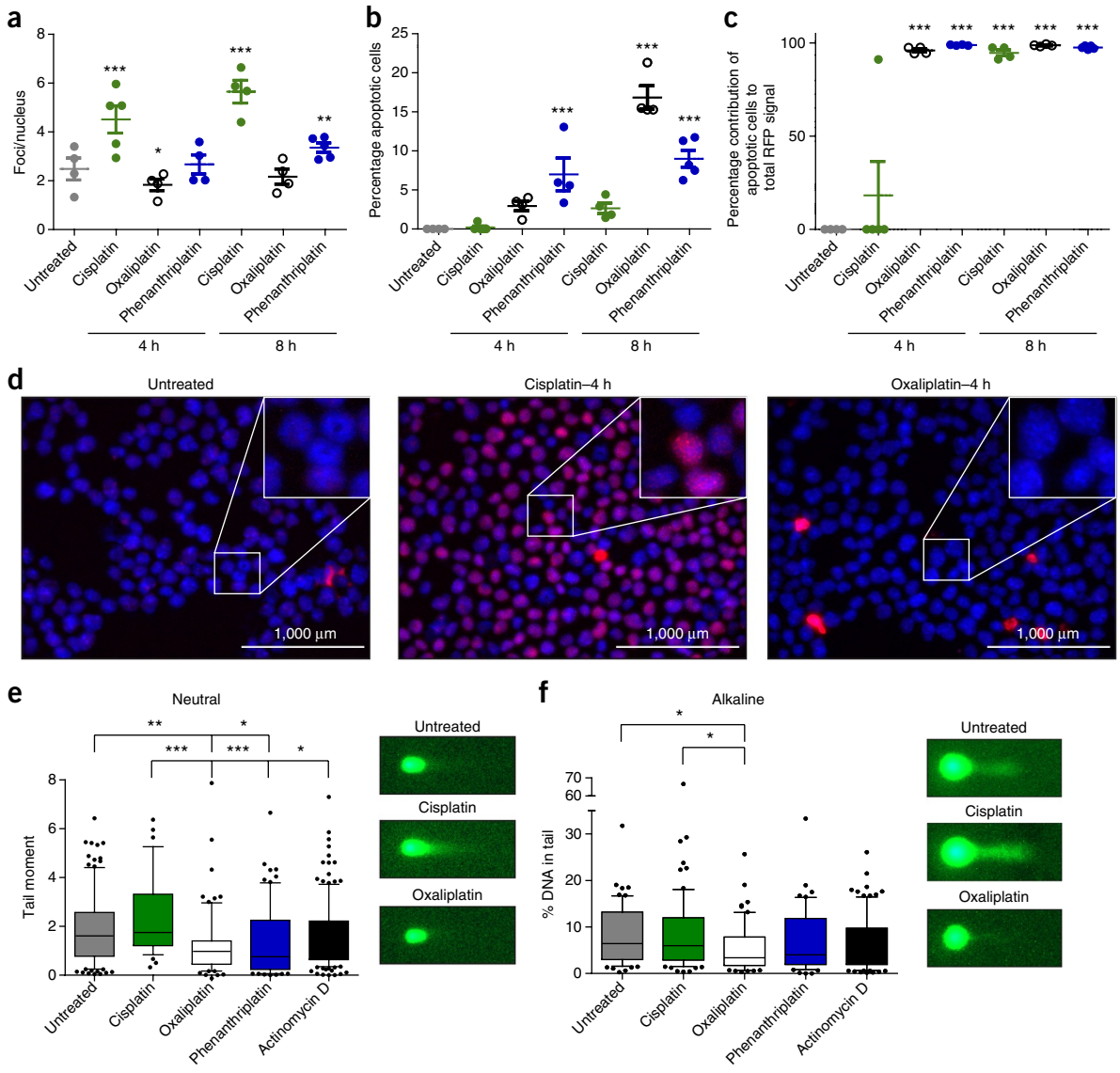


Figure 4 Immunofluorescence of γ-H2AX and comet assays reveal lack of DNA damage resulting from oxaliplatin and phenanthriplatin treatment in *Eμ-Myc Cdkn2a*$^{Arf-/-}$cells. (**a**) Foci per nucleus for each condition at both 4 h and 8 h after treatment. Data are represented as mean ± s.e.m. for each field. *$P < 0.05$, **$P < 0.01$ or ***$P < 0.001$ for each group relative to untreated by one-way ANOVA with Dunnett's multiple-comparison test. (**b**) Apoptotic cells identified via pan-nuclear γ-H2AX as a percent of total nuclei. Data are represented as mean ± s.e.m. for each field. ***$P < 0.0005$ for each group relative to untreated by one-way ANOVA with Dunnett's multiple-comparison test. (**c**) The sum of the integrated intensity of pan-nuclear γ-H2AX divided by the total RFP signal for each field. Data are represented as mean ± s.e.m. for each field. ***$P < 0.0005$ for each group relative to untreated by one-way ANOVA with Dunnett's multiple-comparison test. For **a**–**c**, four fields were analyzed for each condition except the 4-h cisplatin and 8-h phenanthriplatin conditions, which had five each. (**d**) Representative images of γ-H2AX immunofluorescence staining. Insets are 2.5× magnified. Blue, DAPI staining of DNA; red, γ-H2AX. (**e**) Quantification of tail moment after performing a neutral comet assay 6 h after indicated drug treatment. Box center line represents the mean, and box limits are quartiles 1 and 3, and whiskers show 10th and 90th percentiles. Right, representative images from each untreated, cisplatin-treated and oxaliplatin-treated cells. *$P < 0.05$, **$P < 0.01$ or ***$P < 0.001$ for each group relative to untreated or cisplatin-treated cells, by one-way ANOVA with Dunnett's multiple-comparison test. Number of comets analyzed for untreated, cisplatin-, oxaliplatin-, phenanthriplatin- and actinomycin-D-treated cells were 117, 37, 77, 76 and 143, respectively. (**f**) Quantification of percentage of DNA in tail after performing a neutral comet assay 6 h after indicated drug treatment. Box center line represents the mean, and box limits are quartiles 1 and 3, and whiskers show 10th and 90th percentile. Right, representative images from untreated, cisplatin-treated and oxaliplatin-treated cells. *$P < 0.05$ for each group relative to untreated or cisplatin by one-way ANOVA with Dunnett's multiple-comparison test. Number of comets analyzed for untreated, cisplatin-, oxaliplatin-, phenanthriplatin- and actinomycin-D-treated cells were 67, 77, 64, 50 and 91, respectively.

(rRNA) synthesis following drug treatment. Indeed, we observed that within 30 min, pre-rRNA was decreased by nearly 50% in response to oxaliplatin and phenanthriplatin treatment (**Fig. 5b**). At later time points, pre-rRNA was upregulated many-fold by oxaliplatin and phenanthriplatin exposure—a phenotype paralleled by actinomycin D, a known inducer of ribosome biogenesis stress[26], but not by cisplatin (**Fig. 5b**). Importantly, RNA polymerase II transcript levels were unaffected by treatment with any of the platinum agents (**Supplementary Fig. 11**). Finally, to demonstrate on a functional genetic level that ribosome biogenesis stress was relevant to cell death caused by

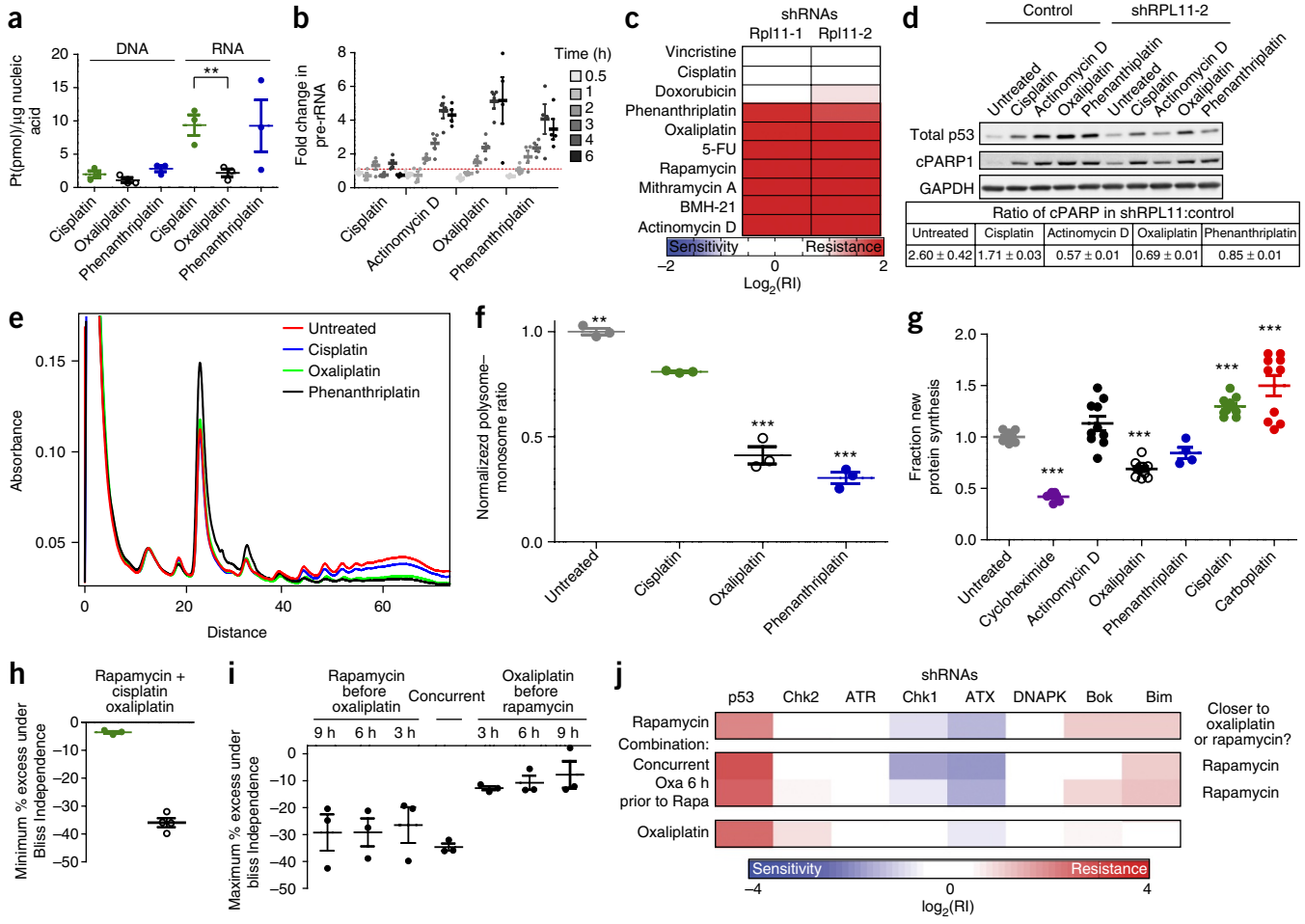
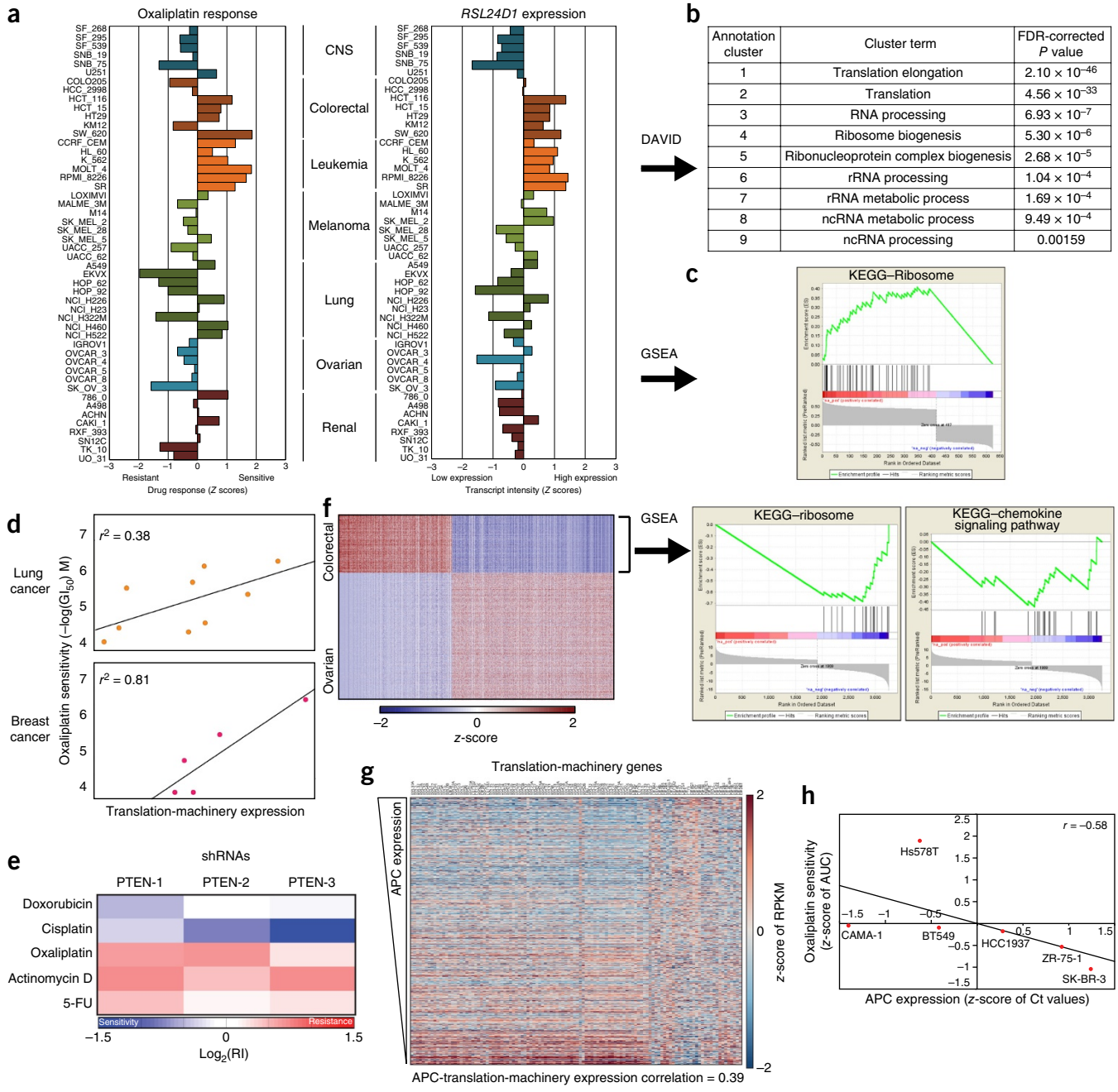


Figure 5 Oxaliplatin and phenanthriplatin induce ribosome biogenesis stress. (**a**) Platinum per µg of nucleic acid, as determined by atomic-absorption spectroscopy as a result of 3 h of treatment of *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells with cisplatin, oxaliplatin or phenanthriplatin at a concentration required to achieve $LD_{80–90}$ at 48 h. Data are represented as mean ± s.e.m. from three independent doses and cultures. $^{**}P < 0.005$ by a two-tailed Student's *t*-test. (**b**) Percentage of pre-ribosomal RNA at various time points after treatment of *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells, as determined by qPCR and normalized to *Gapdh*. Dashed red line indicates no change. Data are represented as mean ± s.e.m. from four independent doses and cultures, except 3-h treatment of all conditions, which had five. (**c**) A heat map depicting enrichment or depletion of two validated hairpins against RPL11 as a result of drug treatment of *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells. (**d**) Top, western blot for total p53, cleaved PARP (cPARP) and GAPDH 12 h after treatment with or without a hairpin against RPL11 in of *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells. Bottom, densitometry quantification of cPARP. Shown is the ratio of cPARP in the shRPL11-2 condition relative to control for each treatment. Data are represented as mean ± s.e.m. from three independent densitometry quantification results. (**e**) Representative polysome gradient after 6 h of treatment with the indicated agents in *Eμ-Myc Cdkn2a*$^{Arf-/-}$ cells. (**f**) Ratio of polysomes to monosomes, normalized to the untreated condition derived from the quantification of the area under the curve (AUC) of monosome and polysome fractions. Data are represented as mean ± s.e.m. from three technical dosing replicates. $^{**}P < 0.01$ or $^{***}P < 0.001$ for each group relative to cisplatin by one-way ANOVA with Dunnett's multiple-comparison test. (**g**) Fraction of newly synthesized protein relative to the untreated condition for *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells at 9 h as measured by O-propargyl puromycin incorporation. Data are represented as mean ± s.e.m. $^{***}P < 0.001$ for each group relative to untreated by one-way ANOVA with Dunnett's multiple-comparison test. All conditions were conducted with two independent doses on independent cultures. Number of fields analyzed was 11 for untreated, 10 for actinomycin D, oxaliplatin, cisplatin and carboplatin, 7 for cycloheximide and 4 for phenanthriplatin. (**h**) Minimum percentage excess propidium iodide (PI)-negative of *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells under Bliss Independence, a control model of additivity, for the combination of rapamycin with either cisplatin or oxaliplatin. Data are represented as mean ± s.e.m. from three independent doses on independent cultures. (**i**) Minimum percentage excess PI-negative cells under Bliss Independence, for the combination of rapamycin and oxaliplatin for which co-dosing was staggered, as shown. Data are represented as mean ± s.e.m. from three independent doses on independent cultures. (**j**) RNAi signatures for rapamycin oxaliplatin, the combination of the two doses simultaneously or oxaliplatin 6 h before rapamycin. Both combination signatures were more similar to rapamycin than oxaliplatin.

oxaliplatin and phenanthriplatin, we examined the role of RPL11 (*Rpl11*) in drug-induced cell death. Ribosome biogenesis stress results in excess subunits of RPL11 that then bind MDM2 and block the ability of MDM2 to bind p53 (ref. 26). shRNA-mediated knockdown of RPL11 in *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells and A549 human lung adenocarcinoma cells confers resistance to agents that induce ribosome biogenesis stress, including actinomycin D, rapamycin, mithramycin A, BMH-21 and 5-FU (**Fig. 5c** and **Supplementary Fig. 12a–c**). shRPL11 also induced resistance to oxaliplatin or phenanthriplatin treatment. Furthermore, we examined total p53 and cleaved poly-ADP ribose polymerase I (cPARP) levels with or without an RPL11 hairpin 12 h after treating with the platinum agents and actinomycin D (**Fig. 5d** and **Supplementary Fig. 12d**). We saw that RPL11 knockdown led to diminished total p53 and cPARP levels in

cells treated with actinomycin D, oxaliplatin or phenanthriplatin, but not with cisplatin. This effect is particularly strong when accounting for p53 and cPARP induction caused by knockdown of an essential component of the ribosome such as RPL11. Thus, the ribosome-biogenesis-stress pathway is a central mediator of oxaliplatin and phenanthriplatin cytotoxicity.

The localization of nucleolar proteins is altered in a characteristic manner in response to different types of cellular stress[27], including



Annotation cluster	Cluster term	FDR-corrected *P* value
1	Translation elongation	2.10×10^{-46}
2	Translation	4.56×10^{-33}
3	RNA processing	6.93×10^{-7}
4	Ribosome biogenesis	5.30×10^{-6}
5	Ribonucleoprotein complex biogenesis	2.68×10^{-5}
6	rRNA processing	1.04×10^{-4}
7	rRNA metabolic process	1.69×10^{-4}
8	ncRNA metabolic process	9.49×10^{-4}
9	ncRNA processing	0.00159

Figure 6 Evidence for sensitization to oxaliplatin in 'translation-addicted' cell lines and primary tumors. (**a**) Graphs of *z*-scores depicting relative resistance or sensitivity to oxaliplatin for various cell lines (left) or relative abundance of *RSL24D1* transcript levels (right). (**b**) Gene-ontology terms identified by DAVID as being significantly enriched among the 417 genes whose expression significantly correlated to oxaliplatin sensitivity. (**c**) The sole KEGG pathway identified by GSEA as being significantly enriched among the 417 genes whose expression was significantly correlated with oxaliplatin sensitivity. (**d**) The correlation of the translation-machinery metagene, an average of ~120 translation-related gene transcripts, to oxaliplatin sensitivity from the NCI-60 for lung and breast cancer cell lines. (**e**) A heat map depicting enrichment or depletion of three validated hairpins targeting PTEN as a result of drug treatment of *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells. (**f**) Left, a heat map representing the relative expression levels of all genes that are differentially expressed in ovarian cancer relative to colorectal cancer, by an absolute fold change of at least $\log_2(2.5)$. Right, the two nondigestion/metabolism-related KEGG pathways identified by GSEA as being significantly enriched among the genes whose expression was lower in ovarian cancer relative to colorectal cancer. (**g**) A heat map of expression of translation-machinery genes from the breast cancer TCGA data set, ranked by APC expression. (**h**) Breast cancer cell line sensitivity to oxaliplatin is correlated with APC expression. Each dot is labeled according to the name of the cell line and represents the mean of three independent replicates of both the dosing and qPCR from the same culture.

impaired translation. After 8 h of treatment of *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells with the platinum agents and actinomycin D, we observed differences in the localization of nucleophosmin (NPM) and fibrillarin (FBL), two proteins normally localized to the nucleolus (**Supplementary Fig. 13a**). Cisplatin treatment did not result in any cells with nucleolar (FBL) foci, whereas actinomycin D, oxaliplatin and phenanthriplatin treatment resulted in a modest increase relative to no treatment (**Supplementary Fig. 13b**). More notably, NPM was found predominantly outside the nucleus for roughly 80% of cisplatin-treated cells, but not in the other conditions (**Supplementary Fig. 13c**). The unique localization of nucleolar proteins among platinum agents further indicates that oxaliplatin and phenanthriplatin have a mechanism of action that is distinct from cisplatin and more similar to that of actinomycin D.

Next, we sought to gain more insight into the influence that oxaliplatin may exert on translation machinery. To do so, we first examined the polysome profiles of *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells after 6 h of treatment with cisplatin, oxaliplatin and phenanthriplatin. Cells treated with oxaliplatin or phenanthriplatin had significantly fewer polysomes relative to monosomes when compared to untreated or cisplatin-treated cells (**Fig. 5e**,**f** and **Supplementary Fig. 14**). Thus, the global translation machinery is markedly perturbed by oxaliplatin treatment. Next, to demonstrate more directly the potential disruptive effect that oxaliplatin treatment had on translation, we quantified nascent protein synthesis by means of a click-chemistry compatible puromycin analog[28]. We observed in multiple cancer cell lines that oxaliplatin disrupts protein synthesis as early as 9 h after treatment (**Fig. 5g** and **Supplementary Fig. 15**).

If oxaliplatin impairs ribosomal function, we would expect to see an effect of oxaliplatin treatment on the efficacy of known translation inhibitors. To examine this possibility, we treated cells with a combination of rapamycin, an MTOR inhibitor and known ribosome-biogenesis-stress inducer, and either cisplatin or oxaliplatin. We then measured synergy or antagonism by comparing to the Bliss Independence additivity model[29]. We saw that the combination of cisplatin and rapamycin was additive, as indicated by the near-zero deviation from Bliss Independence (**Fig. 5h**). However, the combination of oxaliplatin and rapamycin was highly antagonistic. Interestingly, this antagonism was not present if oxaliplatin was dosed before rapamycin, which indicates an epistatic relationship between the mechanisms of action of the two drugs (**Fig. 5i**). To further study this relationship, we examined RNAi signatures to understand how the mechanisms of action varied in combination. Interestingly, the signature was more similar to rapamycin alone than oxaliplatin alone, whether dosed simultaneously or dosed with oxaliplatin 6 h before rapamycin (**Fig. 5j**). This result suggests that active translation must be occurring for oxaliplatin to exert its cytotoxic effects.

Ribosome biogenesis stress sensitizes cells to oxaliplatin

Next, we sought to determine whether ribosome biogenesis stress was a central determinant of oxaliplatin efficacy in human cells and across diverse tumor types. To perform this analysis, we used the NCI-60 human cancer cell line database, which includes gene-expression data for each cell line and its oxaliplatin treatment response. Using the CellMiner NCI-60 database query tool (https://discover.nci.nih.gov/cellminer/)[30], we were able to probe for the genes for which expression in naive cells was correlated with sensitivity to oxaliplatin treatment. For example, *RSL24D1*, the gene most highly correlated to treatment sensitivity, is shown in comparison to the drug-response profile of oxaliplatin (**Fig. 6a** and **Supplementary Fig. 16**). By performing gene-ontology annotation via DAVID on the list of 417 genes whereby expression was significantly correlated with oxaliplatin sensitivity, we identified an enrichment of terms related to translation, the ribosome and rRNA (**Fig. 6b**)[31,32]. Additionally, gene-set enrichment analysis (GSEA) on the same set of genes identified "ribosome" as the only significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (**Fig. 6c**)[33]. Performing similar analysis for cisplatin at the same threshold used for oxaliplatin yielded only one gene, *SFLN11*. Whereas the previous analysis examined gene expression relative to the entire set of cell lines, we wanted to see whether the expression of translation-machinery was correlated with oxaliplatin sensitivity within particular cancer types. Indeed, we saw that in both lung and breast cancer cell lines, higher translation machinery expression was correlated with an increase in oxaliplatin sensitivity (**Fig. 6d**).

We reasoned that this increased expression of translation machinery represents a 'translation addiction,' and that ribosome-biogenesis-stress inducers kill by depriving the cell of translation machinery. Thus, a further increase of translation-machinery expression would buffer against translation-machinery-depriving treatments such as inducers of ribosome biogenesis stress. To examine this possibility, we knocked down PTEN (encoded by *PTEN*) to upregulate the MTOR pathway and thus elevate translation-machinery expression (**Supplementary Fig. 17**). Indeed, upon knocking down PTEN, the cells were rendered more resistant to the known ribosome-biogenesis-stress inducers actinomycin D and 5-FU, as well as to oxaliplatin (**Fig. 6e**). Interestingly, we also saw that PTEN knockdown sensitized cells to cisplatin.

We next wanted to explore how results these might be relevant to the clinic. Given the frequent use of oxaliplatin for the treatment of colorectal cancer, we examined whether translation addiction could underlie oxaliplatin efficacy in this disease, as opposed to cancers for which oxaliplatin shows little efficacy. We compared gene expression between all available colorectal and ovarian tumor samples from the The Cancer Genome Atlas (TCGA)[34,35]. We then performed GSEA on all differentially expressed genes between the two cancer types. On the basis of genes that are upregulated in colorectal cancer relative to ovarian cancer, GSEA identified the 'ribosome' as one of the most enriched pathways that distinguishes colorectal cancer from ovarian cancer (**Fig. 6f** and **Supplementary Table 7**). GSEA did not identify any notable pathway enrichment from genes upregulated in ovarian cancer relative to colorectal cancer (**Supplementary Table 8**). Collectively, these data argue that the efficacy and clinical utility of oxaliplatin is derived from its ability to induce ribosome biogenesis stress.

Finally, we wanted to determine whether we could identify correlates of oxaliplatin response in cancers that are not typically thought to respond to oxaliplatin. Prior studies have indicated that APC (*APC*), a key negative regulator of the WNT pathway, loss in colorectal cancer causes a 'translation addiction' that is necessary for tumorigenesis[36]. Thus, we examined WNT-pathway genes and their correlation with translation-metagene expression in the TCGA expression databases[37]. Interestingly, we found that *APC* expression was significantly correlated with translation-metagene expression in several cancers, including breast and lung (**Fig. 6g** and **Supplementary Fig. 18**). We chose breast cancer for further examination because the NCI-60 breast cancer cell lines had the strongest correlation between translation-machinery expression and oxaliplatin sensitivity. Using a panel of breast cancer cell lines, we confirmed that *APC* expression was correlated with oxaliplatin sensitivity (**Fig. 6h**). Therefore, *APC* expression may represent a marker of oxaliplatin sensitivity in breast cancers, and potentially other cancers as well.

DISCUSSION

Our findings have important implications for the evaluation of small molecules and their derivatives in the clinic. First, in addition to potential changes in drug pharmacodynamics and pharmacokinetics, alterations in the nature of the ligands in platinum complexes have profound implications for primary mechanisms of action. Thus, platinum drugs might not function interchangeably with their derivatives in cancer regimens. This phenomenon explains an observed lack of efficacy for oxaliplatin in the treatment of malignancies conventionally treated by cisplatin, as well as the unanticipated and poorly understood value of oxaliplatin as a treatment for colorectal cancer[38–41]. Second, changes in small-molecule structure can alter the molecular determinants of chemotherapeutic response. For example, instead of cisplatin, oxaliplatin has been tested as a front-line treatment for breast and non-small-cell lung cancer (NSCLC)[42–47]. Furthermore, recent sequencing data have shown that greater than 10% of breast and non-small-cell lung cancers harbor inactivating mutations in *ATM* and *CHEK2* (refs. 48,49). Our data suggest that these mutant tumors may have differential responses to cisplatin and oxaliplatin. Thus, tumor mutations might represent important determinants of susceptibility to related platinum drugs.

Additionally, our data suggest that a 'translation addiction' in colorectal cancer is responsible for the effectiveness of oxaliplatin. Notably, other studies have indirectly associated the translation machinery with the mechanism of oxaliplatin-mediated cell killing. In one study, three oxaliplatin-resistant sublines were generated, and microarrays were performed to compare the resistant and parental cell lines[50]. In each pair of cell lines, a substantial portion of the differentially expressed transcripts between the resistant and parental lines corresponded to genes composing the ribosome. Another study generated an oxaliplatin-sensitivity predictor on the basis of NCI-60 oxaliplatin dose responses and cell line gene-expression data[51], which enabled prediction of the responses to oxaliplatin in both cell lines *in vitro* as well as with patient colorectal cancer xenografts. Additionally, others used microarray technology to identify gene-expression signatures that best predicted outcomes in clinical colorectal cancer cases[52–57]. Despite these unbiased approaches, their signatures consisted of ribosome components. Although this implies a role of the translation machinery in colorectal cancer, none of the aforementioned studies made such a connection.

MTOR-dependent translation elongation plays a crucial part in colorectal cancer tumorigenesis[36]. For example, APC-deficient lesions increase protein synthesis, such that the progression of tumorigenesis is prevented by rapamycin. Thus, translation addiction accompanies colorectal cancer progression, making rapamycin an attractive therapeutic option. Additionally, our data indicate that APC and the translation machinery are linked in other cancer types. However, despite numerous attempts, rapamycin has not attained a solid foothold in cancer treatment, owing in part to cell-intrinsic resistance mechanisms, such as compensatory upregulation of PI3K signaling[58], and the development of rapamycin-insensitive *MTOR* mutations[59]. In cancer types for which rapamycin has failed in clinical trials, oxaliplatin may be viewed as a preferable choice for the inhibition of translation. Moreover, given the role of tumor-intrinsic WNT signaling in suppressing anti-tumor T cell immune responses[60], therapies, such as oxaliplatin, that are more effective in WNT-activated tumors present a potential alternative to immunotherapy.

Although previous studies have identified differences in cellular response between cisplatin and oxaliplatin, ours is the first to identify causal links between oxaliplatin, ribosome biogenesis stress and cell death. For instance, oxaliplatin creates fewer cross-links per base than cisplatin, yet retains its cytotoxicity[61]. We suggest that the ability of oxaliplatin to cross-link DNA might be of questionable relevance, for we do not see activation of the DNA-damage response. However, oxaliplatin-modified DNA could still lead to the inhibition of rRNA synthesis, which would ultimately be responsible for ribosome biogenesis stress. In addition, one can envision scenarios in which oxaliplatin-modified rRNA, mRNA and/or protein could all potentially contribute to ribosome biogenesis stress. A DNA-damage-response-independent mechanism of cell death is consistent with previous observations that cisplatin- or oxaliplatin-resistant cell lines often do not exhibit cross-resistance[3]. Finally, an analysis of the effects of various chemotherapeutics at a range of concentrations on rRNA synthesis and processing found that, at high enough concentrations, many agents—including cisplatin and oxaliplatin—perturb this process[62]. However, this study did not ascribe a causal link between ribosome biogenesis perturbation and cell death. Thus, although many chemotherapeutics may inhibit ribosome biogenesis at high drug concentrations, this behavior is unlikely to be their primary mechanism of action. Our findings demonstrate that, at clinically relevant and lower concentrations, oxaliplatin causes ribosome biogenesis stress, whereas cisplatin does not. Thus, our work demonstrates that perturbation of ribosome biogenesis is functionally important for oxaliplatin, but not cisplatin, to mediate cell death.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

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

ACKNOWLEDGMENTS

This work was supported by the Koch Institute Frontier Research Program through the Michael (1957) and Inara Erdei Fund and the Kathy and Curt Marble Cancer Research Fund, by the Koch Institute Support (core) Grant P30-CA14051 from the National Cancer Institute, by National Cancer Institute Grant CA034992 (S.J.L.), by the Integrative Cancer Biology Program grant #U54-CA112967-09 and by the Center of Cancer Research, the Intramural Program of the National Cancer Institute, NIH (Z01 BC006150-19). M.T.H. is the Chang and Eisen Associate Professor of Biology, C.E.K. was supported by award Number T32GM007753 from the National Institute of General Medical Sciences and G.Y.P. was supported by a Misrock Postdoctoral Fellowship. The authors would also like to thank the Koch Institute Swanson Biotechnology Center for technical support, specifically G. Paradis of the Flow Cytometry Core Facility. We thank J. Wilson for providing platinum compounds ($[Pt(tfbz)(NH_3)_2](NO_3)$) and $[Pt(acac)(NH_3)_2](SO_4)_{0.5}$) and D. Bartel for advice and discussion regarding translation and ribosome translation experiments. The authors also thank G. Walker, A. Koehler, E. Bent, C. Braun, E. Kreidl and B. Zhao for comments and discussion on the paper and N. Fenouille, H. Criscione and F. Lam for technical assistance. The authors thank S. Takeda and M. Takata (Kyoto University, Japan) for providing us with the mutant DT40 cell lines used in this study. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

AUTHOR CONTRIBUTIONS

P.M.B., Y.L., G.Y.P., T.J.E., J.R.P., Y.P., S.J.L. and M.T.H. conceived the idea for the research, designed experiments and interpreted data. P.M.B., Y.L. and C.E.K. performed experiments. P.M.B. and Y.L. performed bioinformatic analyses. J.M. performed DT40 sensitivity profiles. T.J.E. performed polysome gradient profiling. P.M.B., S.J.L. and M.T.H. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Kelland, L. The resurgence of platinum-based cancer chemotherapy. *Nat. Rev. Cancer* **7**, 573–584 (2007).
2. Chu, E. & v DeVita Jr, V.T. *Physician's Cancer Chemotherapy Drug Manual* (Jones & Bartlett Learning Oncology, Sudbury, 2008).
3. Rixe, O. *et al.* Oxaliplatin, tetraplatin, cisplatin, and carboplatin: spectrum of activity in drug-resistant cell lines and in the cell lines of the National Cancer Institute's Anticancer Drug Screen panel. *Biochem. Pharmacol.* **52**, 1855–1865 (1996).
4. Machover, D. *et al.* Two consecutive phase II studies of oxaliplatin (L-OHP) for treatment of patients with advanced colorectal carcinoma who were resistant to previous treatment with fluoropyrimidines. *Ann. Oncol.* **7**, 95–98 (1996).
5. Jiang, H., Pritchard, J.R., Williams, R.T., Lauffenburger, D.A. & Hemann, M.T. A mammalian functional-genetic approach to characterizing cancer therapeutics. *Nat. Chem. Biol.* **7**, 92–100 (2011).
6. Pritchard, J.R. *et al.* Defining principles of combination drug mechanisms of action. *Proc. Natl. Acad. Sci. USA* **110**, E170–E179 (2013).
7. Pritchard, J.R., Bruno, P.M., Hemann, M.T. & Lauffenburger, D.A. Predicting cancer drug mechanisms of action using molecular network signatures. *Mol. Biosyst.* **9**, 1604–1619 (2013).
8. Suntharalingam, K. *et al.* Bidentate ligands on osmium(VI) nitrido complexes control intracellular targeting and cell death pathways. *J. Am. Chem. Soc.* **135**, 14060–14063 (2013).
9. Suntharalingam, K. *et al.* A breast cancer stem cell-selective, mammospheres-potent osmium(VI) nitrido complex. *J. Am. Chem. Soc.* **136**, 14413–14416 (2014).
10. Suntharalingam, K. *et al.* Necroptosis-inducing rhenium(V) oxo complexes. *J. Am. Chem. Soc.* **137**, 2967–2974 (2015).
11. Awuah, S.G., Zheng, Y.-R., Bruno, P.M., Hemann, M.T. & Lippard, S.J.A. Pt(IV) pro-drug preferentially targets indoleamine-2,3-dioxygenase, providing enhanced ovarian cancer immuno-chemotherapy. *J. Am. Chem. Soc.* **137**, 14854–14857 (2015).
12. Boodram, J.N. *et al.* Breast cancer stem cell potent copper(II)-non-steroidal anti-inflammatory drug complexes. *Angew. Chem. Int. Ed.Engl.* **55**, 2845–2850 (2016).
13. Cressey, P.B. *et al.* The potent inhibitory effect of a naproxen-appended cobalt(III)-cyclam complex on cancer stem cells. *ChemBioChem* **17**, 1713–1718 (2016).
14. Zheng, Y.-R. *et al.* Mechanistic studies of the anticancer activity of an octahedral hexanuclear Pt(II) cage. *Inorganica Chim. Acta* **452**, 125–129 (2016).
15. Barnes, J.C. *et al.* Using an RNAi signature assay to guide the design of three-drug-conjugated nanoparticles with validated mechanisms, *in vivo* efficacy, and low toxicity. *J. Am. Chem. Soc.* **138**, 12494–12501 (2016).
16. Zamble, D.B., Mu, D., Reardon, J.T., Sancar, A. & Lippard, S.J. Repair of cisplatin—DNA adducts by the mammalian excision nuclease. *Biochemistry* **35**, 10004–10013 (1996).
17. Fojo, T. *et al.* Identification of non-cross-resistant platinum compounds with novel cytotoxicity profiles using the NCI anticancer drug screen and clustered image map visualizations. *Crit. Rev. Oncol. Hematol.* **53**, 25–34 (2005).
18. Alvarez, M. *et al.* Generation of a drug resistance profile by quantitation of mdr-1/P-glycoprotein in the cell lines of the National Cancer Institute Anticancer Drug Screen. *J. Clin. Invest.* **95**, 2205–2214 (1995).
19. Willett, P., Barnard, J.M. & Downs, G.M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **38**, 983–996 (1998).
20. Williams, R.T., Roussel, M.F. & Sherr, C.J. Arf gene loss enhances oncogenicity and limits imatinib response in mouse models of Bcr-Abl-induced acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. USA* **103**, 6688–6693 (2006).
21. Maede, Y. *et al.* Differential and common DNA repair pathways for topoisomerase I- and II-targeted drugs in a genetic DT40 repair cell screen panel. *Mol. Cancer Ther.* **13**, 214–220 (2014).
22. Jiang, H. *et al.* The combined status of ATM and p53 link tumor development with therapeutic response. *Genes Dev.* **23**, 1895–1909 (2009).
23. Rogakou, E.P., Nieves-Neira, W., Boon, C., Pommier, Y. & Bonner, W.M. Initiation of DNA fragmentation during apoptosis induces phosphorylation of H2AX histone at serine 139. *J. Biol. Chem.* **275**, 9390–9395 (2000).
24. Bonner, W.M. *et al.* GammaH2AX and cancer. *Nat. Rev. Cancer* **8**, 957–967 (2008).
25. Olive, P.L. & Banáth, J.P. The comet assay: a method to measure DNA damage in individual cells. *Nat. Protoc.* **1**, 23–29 (2006).
26. Golomb, L., Volarevic, S. & Oren, M. p53 and ribosome biogenesis stress: the essentials. *FEBS Lett.* **588**, 2571–2579 (2014).
27. Boulon, S., Westman, B.J., Hutten, S., Boisvert, F.M. & Lamond, A.I. The nucleolus under stress. *Mol. Cell* **40**, 216–227 (2010).
28. Liu, J., Xu, Y., Stoleru, D. & Salic, A. Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin. *Proc. Natl. Acad. Sci. USA* **109**, 413–418 (2012).
29. Bliss, C.I. The calculation of microbial assays. *Bacteriol. Rev.* **20**, 243–258 (1956).
30. Reinhold, W.C. *et al.* CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res.* **72**, 3499–3511 (2012).
31. Huang, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
32. Huang, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
33. 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).
34. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
35. Muzny, D.M. *et al.*; Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
36. Faller, W.J. *et al.* mTORC1-mediated translational elongation limits intestinal tumour initiation and growth. *Nature* **517**, 497–500 (2015).
37. Koboldt, D.C. *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
38. Raymond, E., Chaney, S.G., Taamma, A. & Cvitkovic, E. Oxaliplatin: a review of preclinical and clinical studies. *Ann. Oncol.* **9**, 1053–1071 (1998).
39. Raymond, E., Lawrence, R., Izbicka, E., Faivre, S. & Von Hoff, D.D. Activity of oxaliplatin against human tumor colony-forming units. *Clin. Cancer Res.* **4**, 1021–1029 (1998).
40. Shelley, M.D., Burgon, K. & Mason, M.D. Treatment of testicular germ-cell cancer: a cochrane evidence-based systematic review. *Cancer Treat. Rev.* **28**, 237–253 (2002).
41. Goldberg, R.M. *et al.* A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *J. Clin. Oncol.* **22**, 23–30 (2004).
42. Raez, L.E. *et al.* Efficacy and safety of oxaliplatin and docetaxel in patients with locally advanced and metastatic non-small-cell lung cancer (NSCLC). *Lung Cancer* **53**, 347–353 (2006).
43. Atmaca, A. *et al.* A randomised multicentre phase II study with cisplatin/docetaxel vs oxaliplatin/docetaxel as first-line therapy in patients with advanced or metastatic non-small cell lung cancer. *Br. J. Cancer* **108**, 265–270 (2013).
44. Scagliotti, G.V. *et al.* Pemetrexed combined with oxaliplatin or carboplatin as first-line treatment in advanced non-small cell lung cancer: a multicenter, randomized, phase II trial. *Clin. Cancer Res.* **11**, 690–696 (2005).
45. Yardley, D.A. *et al.* A phase II trial of oxaliplatin and trastuzumab in the treatment of HER2-positive metastatic breast cancer. *Cancer Invest.* **28**, 865–871 (2010).
46. Njiaju, U.O. *et al.* Capecitabine and oxaliplatin in combination as first- or second-line therapy for metastatic breast cancer: a Wisconsin Oncology Network trial. *Cancer Chemother. Pharmacol.* **71**, 613–618 (2013).
47. Guerrero, A. *et al.* Phase I/II study of biweekly vinorelbine and oxaliplatin as first-line treatment in patients with metastatic breast cancer. *Anticancer Drugs* **22**, 283–289 (2011).
48. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107–1120 (2012).
49. Meijers-Heijboer, H. *et al.* Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat. Genet.* **31**, 55–59 (2002).
50. Samimi, G. *et al.* cDNA microarray-based identification of genes and pathways associated with oxaliplatin resistance. *Cancer Chemother. Pharmacol.* **55**, 1–11 (2005).
51. Kim, M.K. *et al.* Characterization of an oxaliplatin sensitivity predictor in a preclinical murine model of colorectal cancer. *Mol. Cancer Ther.* **11**, 1500–1509 (2012).
52. Bertucci, F. *et al.* Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene* **23**, 1377–1391 (2004).
53. Arango, D. *et al.* Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. *Gastroenterology* **129**, 874–884 (2005).
54. Barrier, A. *et al.* Stage II colon cancer prognosis prediction by tumor gene expression profiling. *J. Clin. Oncol.* **24**, 4685–4691 (2006).
55. Yamasaki, M. *et al.* The gene expression profile represents the molecular nature of liver metastasis in colorectal cancer. *Int. J. Oncol.* **30**, 129–138 (2007).
56. Bandrés, E. *et al.* A gene signature of 8 genes could identify the risk of recurrence and progression in Dukes' B colon cancer patients. *Oncol. Rep.* **17**, 1089–1094 (2007).
57. Fritzmann, J. *et al.* A colorectal cancer expression profile that includes transforming growth factor beta inhibitor BAMBI predicts metastatic potential. *Gastroenterology* **137**, 165–175 (2009).
58. Zoncu, R., Efeyan, A. & Sabatini, D.M. mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat. Rev. Mol. Cell Biol.* **12**, 21–35 (2011).
59. Voss, M.H. *et al.* Tumor genetic analyses of patients with metastatic renal cell carcinoma and extended benefit from mTOR inhibitor therapy. *Clin. Cancer Res.* **20**, 1955–1964 (2014).
60. Spranger, S., Bao, R. & Gajewski, T.F. Melanoma-intrinsic β-catenin signalling prevents anti-tumour immunity. *Nature* **523**, 231–235 (2015).
61. Woynarowski, J.M. *et al.* Oxaliplatin-induced damage of cellular DNA. *Mol. Pharmacol.* **58**, 920–927 (2000).
62. Burger, K. *et al.* Chemotherapeutic drugs inhibit ribosome biogenesis at various levels. *J. Biol. Chem.* **285**, 12416–12425 (2010).

ONLINE METHODS

shRNA constructs. All shRNAs used were in the pMSCV-LTR-miR30-SV40-GFP (MLS) retroviral vector, with the exception of RPL11 in TMPVIR[63,64]. Transfection and infection were performed as previously described[6]. Unless shown, all knockdowns were previously validated for knock-down and off-target effects[5]. shRNA target sequences are listed in **Supplementary Table 9**.

GFP competition assays. *Eμ-Myc Cdkn2a*$^{Arf-/-}$, *Eμ-Myc p53*$^{-/-}$ and p185+ *BCR-Abl Cdkn2a*$^{Arf-/-}$ leukemia cells were infected with GFP-tagged shRNAs such that 15–25% of the population were GFP positive. One-eighth of a million cells in 250 μl of B-cell media (BCM) were then seeded into 24-well plates. For wells that would remain untreated as a control, only 1/16th of a million cells were seeded. Next, 250 μl of drug-containing media was added to the cells. After 24 h, 300 μl of cells from untreated wells were removed and replaced by 300 μl of fresh BCM. All wells then received 500 μl of BCM before being placed in the incubator for another 24 h. At 48 h, cells transduced with the control vector, MLS, were checked for viability via flow cytometry on a BD FACScan using propidium iodide (PI) as a live/dead marker. Untreated wells then had 700 μl of cells removed and replaced with 700 μl of fresh media, followed by a further 1 ml of fresh media. Wells for which the drug had killed 80–90% of cells (LD_{80-90}) were then diluted further by adding 1 ml of BCM. Finally, at 72 h, all wells for which an LD_{80-90} had been achieved, as well as the untreated samples, were run via flow cytometry to determine GFP%. For shRPL11 experiments, the above procedures were followed except in the presence of 1 μg/ml of doxycycline, including 24 h of pre-incubation of the hairpin-containing cells in doxycycline. For all competition assays, at least three separate biological replicates were performed. Here we define biological replicates as a fresh thaw that was infected, selected or sorted if necessary, then run through the competition assay.

RNAi signatures. Drugs were dosed to achieve an LD_{80-90} in *Eμ-Myc Cdkn2a*$^{Arf-/-}$ cells by propidium iodide exclusion, as determined by flow cytometry at 48 h. GFP enrichment/depletion was then determined by flow cytometry at 72 h. Linkage ratios and *P* values were generated as described previously[5–7]. All flow cytometry was conducted using a BD FACScan. For all RNA signatures, at least three separate biological replicates were performed. Here we define biological replicates as a fresh thaw of *Eμ-Myc Cdkn2a*$^{Arf-/-}$ cells that were infected, selected or sorted if necessary, and then run through the signature assay.

Code availability. Parameters and code for the modified K–nearest neighbors (K-NN) algorithm for determining drug mechanism of action can be found in previous publications[5,7].

Drug doses. Unless otherwise noted, all drugs were administered to *Eμ-Myc Cdkn2a*$^{Arf-/-}$, *Eμ-Myc p53*$^{-/-}$ and p185+ *BCR-Abl Cdkn2aArf*$^{-/-}$, leukemia cells at a concentration to achieve an LD_{80-90} and 48 h, as assessed via flow cytometry on a BD FACScan using propidium iodide as a live/dead marker. For time-point experiments, a portion of each sample was cultured until 48 h post-treatment to ensure that LD_{80-90} was achieved. If not, that sample was discarded and not included in further analysis. Similarly rigorous equipotent dosing was followed for adherent cell lines by first dosing over a large concentration range for 96 h and analyzing cell number via CellTiter-Glo (Promega). Then, for future equipotent doses, only, concentrations that achieved an IC_{80} at 96 h were used. For AUC calculations, whole-dose response curves were analyzed at 96 h and analyzed via MatLab 2015a (MathWorks). All drug doses were performed at least once from three independent cultures.

Cell cycle. 1–2 million cells were collected per sample after treatment with a concentration of drug required to achieve an LD_{80-90} at 48 h for *Eμ-Myc Cdkn2a*$^{Arf-/-}$ or IC_{80} at 96 h for A549 and LoVo. They were washed in PBS and then resuspended in 1 ml of PBS + 2% FBS (FBS). Cells were then fixed in 70% ethanol overnight at −20 °C. Cells were then rehydrated with 5 ml of PBS, spun out of ethanol and then washed in PBS. Next, 1 ml of PI staining solution (3.8 mM sodium citrate, 50 μg/ml of PI in PBS) and 20 μl of RNAse A were used to resuspend the pellet. Cells were then incubated overnight at 4 °C before FACS on a BD FACScan. Cell-cycle profiles were analyzed using FlowJo V10.

RNA isolation and expression analysis. Total RNA was isolated via QIAshredder and Qiagen RNeasy Mini Kit, and cDNA was generated via M-MLV Reverse Transcriptase with RNaseOUT (ThermoFisher Scientific) and random hexamers. RT–qPCR was performed with Fast SYBER Green Mastermix and StepOnePlus Real-Time PCR System (Applied Biosystems) with three biological replicates, each with two technical replicates. Primer sequences listed in **Supplementary Table 10**.

Western blotting. Cell pellets were washed in PBS and frozen before being lysed in RIPA buffer. They were then boiled after the addition of Laemmli sample buffer and run on an SDS–PAGE gel. From the gel, they were transferred to Millipore immobilin-P membranes. The antibodies used were mouse anti-TRP53 (NCL-p53-505, Novocastra; 1:500), human anti-TP53 (2524, Cell Signaling; 1:1,000), anti-phospho TRP53 (Ser 15) (9284, Cell Signaling; 1:1,000), anti-phospho TP53 (Ser 20) (9287, Cell Signaling; 1:1,000), anti-γ-H2AX (20E3, Cell Signaling; 1:1,000), anti-CDKN1A (sc-6246, Santa Cruz, 1:200), anti-cPARP1 (9544, Cell Signaling, 1:1,000), anti-GAPDH (sc-32233, Santa Cruz, 1:200), anti-ACTB (13E5, Cell Signaling; 1:1,000), anti-HSP90 (68/Hsp90, Becton Dickinson; 1:6,000). See **Supplementary Fig. 19** for full western blots and repetitions.

Cell culture. *Eμ-Myc Cdkn2a*$^{Arf-/-}$, *Eμ-Myc Trp53*$^{-/-}$ and p185+ *BCR-Abl Cdkn2a*$^{Arf-/-}$ primary murine tumor cells were cultured according to established protocols[65]. DT40 cells were cultured at 37 °C with 5% CO_2 in RPMI-1640 medium with glutamine (11875, Invitrogen, Carlsbad, CA) supplemented with 1% chicken serum (16110-082, Invitrogen, Carlsbad, CA), 1% M β-mercaptoethanol (M-3148, Sigma-Aldrich, St. Louis, MO), penicillin–streptomycin (15140-122, Invitrogen) and 10% FBS (100-106, Gemini Bio-Products, West Sacramento, CA). A549, SW480, LoVo, CAMA-1, Hs578T, ZR-75-1, BT549, SK-BR-3 and HCC1937 cells were obtained from ATCC (CCL-185, CCL-228, CCL-229, HTB-21, HTB-126, CRL-1500, HTB-122, HTB-30 and CRL-2336) and cultured according to ATCC guidelines. All cell lines, except those obtained and used directly from ATCC, were tested for mycoplasma and found to be negative (MycoAlert, Lonza). No authentication was performed.

Cellular uptake of platinum. Two million *Eμ-Myc Cdkn2a*$^{Arf-/-}$ cells were dosed at concentrations required to achieve an LD_{80-90} at 48 h or at 5 μM in three biological replicates. After 3 h, they were washed three times in PBS, pelleted and frozen. Atomic-absorption spectroscopic analysis was then performed as previously described[66].

Structure-based hierarchical clustering. The ChemmineR R package was used to calculate Tanimoto coefficients and perform hierarchical clustering.

DT40 knockout-line drug-sensitivity analysis. The detailed information about mutant DT40 cell lines used in this study was shown previously[21]. To measure the sensitivity, cells were continuously exposed to various concentrations of drugs for 72 h in triplicate. Cell survival was determined using the ATPlite 1-step kit (PerkinElmer) and an EnVision 2104 Multilabel Reader. To evaluate the relative cellular sensitivity of each mutant to wild-type cells, logarithm sensitivity curves were used to determine the concentration that causes 90% reduction of ATP activity (IC_{90}). The IC_{90} value of each mutant was divided by that of wild-type cells on the same plate. Finally, the quotient was converted to a logarithmic scale (base 2). Pairwise distances were calculated via the 'pdist' function with the Euclidian distance metric in Matlab 2013b.

Immunofluorescence staining. *Eμ-Myc Cdkn2a*$^{Arf-/-}$ cells were dosed at an LD_{80-90} and after 4 h or 8 h of treatment placed on slides via cytospin. LoVo cells were dosed at IC_{80} and fixed after 12 h or treatment. Cells were then dried and stained with anti-γ-H2AX (20E3, Cell Signaling; 1:500) and DAPI or anti-NPM (B0556, Sigma-Aldrich, 1:2,000), anti-FBN (ab5821, Abcam, 1:500), and DAPI. AlexFluor 488 or 555 was used as a secondary and then samples were mounted with Prolong Gold. An EVOS FL Auto microscope was used for imaging. Image analysis was performed using CellProfiler on the original images.

Comet assays. *Eμ-Myc Cdkn2a*$^{Arf-/-}$ cells were dosed at an LD_{80-90} and, after 6 h, were washed and prepared for either alkaline or neutral comet assays. Comet-tail

doi:10.1038/nm.4291

analysis was performed using CaspLab on images that were all equally level adjusted in Adobe Photoshop. ZR-75-1 cells were dosed at an IC_{80} and, after 24 h, were washed and prepared for neutral comet assays. CellProfiler was used to analyze ZR-75-1 comets. The Trevigen CometAssay Kit was used for both cell lines.

***In vivo* competition assay.** 2 million *Eμ-Myc Cdkn2a*$^{Arf-/-}$ cells were partially infected with shChk2 and tail-vein injected into 8-week-old female C57BL/6 mice. Cisplatin, oxaliplatin or phenanthriplatin in sterile saline were administered via intraperitoneal injection at disease presentation (7 mg/kg, 5 mg/kg or 3 mg/kg, respectively), and mice were killed and tumors harvested upon relapse, 2–3 d after treatment. Disease presentation and relapse were defined as the presence of palpable lymph node burden. Fold change in GFP% was assessed relative to untreated mice after tumor cell harvesting by flow cytometry on a BD FACScan. Each treatment group had five or six mice. This was chosen independently of statistical power analysis. No randomization or blinding methods were used to determine treatment branches. The experiment was performed in accordance with MIT's Committee on Animal Care (CAC) before execution.

Combination doses and Bliss Independence. *Eμ-Myc Cdkn2a*$^{Arf-/-}$ lymphoma cells were treated with drug and analyzed as detailed above, except with two drugs in combination, in a 3 × 5–dose response matrix in which the outer wells were only exposed to a single drug[6]. The single-drug wells were then used to calculate the expected Bliss Independence values of the combination wells. The expected Bliss Independence values were then subtracted from the observed values. The minimum or maximum value in the matrix was then chosen as the minimum or maximum percentage excess under/over Bliss Independence.

Statistical analysis. Two-sided, nonparametric tests were performed unless otherwise specified. If two conditions were compared, Mann–Whitney *U* test was used. If more than one condition was compared to a control condition, then one-way ANOVA with Dunnett's multiple-comparison test was used. If a *t* test was used, the data were first confirmed to fit a normal distribution via the 'kstest' function in Matlab 2015a. See figure legends for details on types of replicate and *n* values used.

Principal-component analysis. Principal-component analysis (PCA) was used to provide another means of visualizing our data, and was not used to classify new compounds. This was done via Matlab 2015a. The 'pca' function was used with data from **Supplementary Table 4** to generate the necessary input variables for the 'biplot' function.

Bioinformatic analysis. CellMiner was used to obtain NCI-60 oxaliplatin sensitivity-expression level correlation values[30]. Benjamini–Hochberg correction was then used to eliminate genes above an FDR threshold of 0.05. For the TCGA tumor-expression analysis, RNA-seq gene-level raw-count data for colorectal and ovarian cancer tumor samples that went through the same pipeline for generation of expression levels (RPKM values) (193 and 420 samples respectively) were downloaded from the TCGA data portal. The Bioconductor edgeR package in R was used to perform normalization and analysis for differential expression between these two cancer types[67]. We obtained 3,247 significantly differentially expressed genes between these two cancer types, with a $\log_2$-fold change in normalized tag counts greater than 2.5 (corresponding to a Benjamini–Hochberg corrected *P* value of 5.8×10^{-3} from a negative binomial generalized linear model fit). Among these genes, 1,909 and 1,338 were significantly upregulated and downregulated, respectively, in ovarian cancer as compared to colorectal cancer. Significantly correlated (NCI-60) or differentially expressed (TCGA) genes were then loaded in GSEA, and the pre-ranked list tool was used[33].

For both NCI-60 and TCGA data sets, a translation metagene was obtained by averaging the expression of 120 genes known to be involved in translation for each. To demonstrate correlation between oxaliplatin sensitivity and translation, we plotted the GI_{50} values (log scale) against the metagene-expression levels for lung and breast cancer cell lines in the NCI-60 catalog. To examine associations between the WNT-pathway activity and expression levels of the translation machinery, we calculated Pearson-correlation coefficients of the translation-metagene expression with that of known WNT-pathway genes in the Reactome database for breast cancer samples in the TCGA database. A heat map of expression values of translation-machinery genes, ordered by decreasing APC expression values, was then plotted to visualize such correlations.

Ribosome profiling. 1 million cells for each of three replicates were harvested by centrifugation at 300*g* for 5 min, resuspended in ice-cold PBS containing 100 μg/ml cycloheximide and centrifuged again. Cell pellets were flash-frozen in liquid nitrogen. To prepare the lysate, cell pellets were thawed in lysis buffer (10-mM Tris-HCl, pH 7.4, 5-mM $MgCl_2$, 1% Triton X-100, 100-mM KCl, 0.02 U/μl Superase-IN, 2-mM DTT and 100-μg/ml cycloheximide), passed four times through a 26-gauge needle and cleared by centrifugation at 1,300*g* for 10 min at 4 °C. The supernatant was loaded onto a 10–50% w/v linear sucrose gradient containing 20-mM HEPES-KOH, pH 7.4, 5-mM $MgCl_2$, 100-mM KCl, 2-mM DTT, 100-μg/ml cycloheximide and 0.02-U/μl Superase-IN. Gradients were centrifuged for 2 h at 36,000 rpm and 4 °C and then fractionated by upward displacement using a Biocomp Gradient Station with continuous absorbance monitoring at 254 nm. Traces were quantified using Fiji. Dosing of replicates was performed as technical replicates from the same culture that had been split into three the previous day. Profiling was conducted in two separate rounds, one set of samples together and the other two together in consecutive weeks.

Data-availability statement. The data that support the findings of this study are available from the corresponding author upon reasonable request.

63. Dickins, R.A. *et al.* Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nat. Genet.* **37**, 1289–1295 (2005).
64. Zuber, J. *et al.* Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nat. Biotechnol.* **29**, 79–83 (2011).
65. Zhao, B., Pritchard, J.R., Lauffenburger, D.A. & Hemann, M.T. Addressing genetic tumor heterogeneity through computationally predictive combination therapy. *Cancer Discov.* **4**, 166–174 (2014).
66. Park, G.Y., Wilson, J.J., Song, Y. & Lippard, S.J. Phenanthriplatin, a monofunctional DNA-binding platinum anticancer drug candidate with unusual potency and cellular activity profile. *Proc. Natl. Acad. Sci. USA* **109**, 11987–11992 (2012).
67. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

Appendix C. Transient Hsp90 Suppression Promotes a Heritable Change in Protein Translation

Peter Tsvetkov[1]*, Zarina Brune[1], Timothy J. Eisen[1], Sven Heinrich[1], Greg A. Newby[1], Erinc Hallacli[1], Can Kayatekin[1], David Pincus[1], Susan Lindquist[1,2,3,Ψ]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
[2]Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[3]Howard Hughes Medical Institute, Cambridge, MA 02139, USA
[Ψ]Deceased
*correspondence: ptsvetko@broadinstitute.org

Unpublished: Tsvetkov, P., Brune, Z., Eisen, T.J., Heinrich, S., Newby, G.A., Hallacli, E., Kayatekin, C., Pincus, D., and Lindquist, S. (2018). Transient Hsp90 suppression promotes a heritable change in protein translation. bioRxiv, 366070.

Transient Hsp90 suppression promotes a heritable change in protein translation

Peter Tsvetkov[1*], Zarina Brune[1], Timothy J. Eisen[1], Sven Heinrich [1], Greg A. Newby[1], Erinc Hallacli[1], Can Kayatekin[1], David Pincus[1], Susan Lindquist[1,2,3, Ψ].

[1] Whitehead Institute for Biomedical Research, Cambridge, MA 02142
[2] Massachusetts Institute of Technology, Cambridge, MA 02142, USA
[3] Howard Hughes Medical Institute, Cambridge, MA 02139, USA.
[Ψ] Deceased
[*] correspondence: ptsvetko@broadinstitute.org

The heat shock protein 90 (Hsp90) chaperone functions as a protein-folding buffer and plays a unique role promoting the evolution of new heritable traits. To investigate the role of Hsp90 in modulating protein synthesis, we screened more than 1200 proteins involved in mRNA regulation for physical interactions with Hsp90 in human cells. Among the top hits was CPEB2, which strongly binds Hsp90 via its prion domain, reminiscent of the prion-like regulation of translation of *Aplysia* CPEB. In a yeast model of CPEB prion-dependent translation regulation, transient inhibition of Hsp90 amplified CPEB2 prion activity and resulted in persistent translation of the CPEB reporter. Remarkably, inhibition of Hsp90 was sufficient to induce a heritable change in protein translation that persisted for 30 generations, even in the absence of exogenous CPEB. Although we identified a variety of perturbations that enhanced translation of the reporter, only Hsp90 inhibition led to persistent activation. Thus, transient loss of Hsp90 function leads to the non-genetic inheritance of a novel translational state. We propose that, in addition to sculpting the conformational landscape of the proteome, Hsp90 promotes phenotypic variation by modulating protein synthesis.

Introduction:

Across evolution many of the molecular pathways underlying complex multicellular processes such as development and neuronal synapse formation evolved from basic environmental signaling circuits in unicellular organisms like yeast [1, 2]. The regulation of protein translation is a key hub that integrates extracellular signals into a cellular response [3, 4]. Indeed, the requirement for specific translation regulating factors to integrate environmental cues and induce specific mRNA translation is critical in development, stress response and neuronal function [4, 5], and also for the survival of single-celled organisms.

The members of the cytoplasmic polyadenylating element binding (CPEB) family of RNA-binding proteins act as regulators of development [6, 7] and synaptic protein synthesis [8, 9], aiding in synaptic remodeling and the persistence of behavioral memories [10, 11]. The CPEB homologs can be divided into two evolutionarily conserved groups. The first group includes human CPEB1 and its orthologs from *Xenopus* and *Drosophila (Orb1)* [12]. The second group includes human CPEB2-4, the neuronal *Aplysia* CPEB and the *Drosophila* ortholog, Orb2, all of which have been shown to regulate synaptic function. This group of CPEBs possess a glutamine-rich N-terminal domain thought to comprise a functional prion domain (PrD). The ability of neuronal CPEBs to activate translation in response to synaptic activity depends largely on the prion domain [9, 13-19]; in the aggregated prion form, the neuronal CPEB is active. In the soluble form, it is dormant [13, 19]. The prion properties of *Aplysia* neuronal CPEB were initially described in the context of the yeast model using a CPEB translation reporter as a read-out of activity [13]. When expressed in yeast, *Aplysia* CPEB maintains its function as a translational regulator and employs a prion mechanism to create stable, finely tuned, and self-perpetuating activity states controlling translation [13, 20]. While the prion-based activation of CPEB is evolutionarily conserved,

little is known concerning the mechanism that induces CPEB structural switching.

In fungi, the chaperone machinery largely regulates protein folding and prion switching. Transient alteration in the chaperone machinery function can induce prion protein conformational switching, and this conformation persists for many generations [21-23]. The best-characterized regulator of prion switching in fungi is the Hsp104 chaperone. Transient perturbation of Hsp104 chaperone function either by the introduction of chemical or genetic Hsp104 inhibitors or by overexpression of the wild-type Hsp104 has been demonstrated to regulate the conformational switching of different prions and induce a heritable epigenetic trait [21-23]. Despite its critical role in protein folding homeostasis in fungi, Hsp104 has no known metazoan ortholog, suggesting that other chaperones have subsumed its function. For example, the chaperone Hsp70 also mediates some prion switching events in fungi [24-26] and is conserved in metazoans. Another highly abundant chaperone is Hsp90, which is evolutionarily conserved from bacteria to mammals. The role of Hsp90 in the cell extends far beyond the maintenance of protein homeostasis. For example, the ability of Hsp90 to bind a repertoire of metastable protein clients also enables Hsp90 to facilitate the evolution of diverse new biological phenotypes in different organisms [27-29]. Recent evidence suggests that the Hsp90 chaperone plays a role in prion switching as well [30-32].

In this work, using a proteomic approach, we reveal that the Hsp90 chaperone is a regulator of persistent CPEB translation activation. We then used a reporter of CPEB translational activation in yeast [13, 20] to elucidate an endogenous yeast mechanism through which transient Hsp90 suppression induces mRNA translation that persists for many generations. This work demonstrates that Hsp90 can regulate inheritance of a particular state of translational activity, elucidating an additional dimension to the role of Hsp90 as a major regulator of phenotypic diversity.

Results

Exploring the interactions of translation regulating proteins with the Hsp90 chaperone

We set out to explore how HSP90 regulates translation by characterizing the protein:protein interaction between HSP90 and a large library consisting of more than 1200 proteins that were shown to regulate translation or were associated with translation regulating processes based on Pfam nucleotide binding domains and previous experimental characterization [33, 34] (sup table 1). To quantify their interaction with Hsp90, we used LUMIER (luminescence-based mammalian interactome mapping) [35] (Fig 1A). In this assay, each protein from the library (bait) was Flag-tagged and overexpressed in HEK293T cells stably over-expressing HSP90 fused to Renilla luciferase. Following immunoprecipitation of the flag epitope, the protein-protein interaction was determined by the intensity of luminescence (Figure 1A). The distribution of interaction strength across all 1247 proteins (some with multiple isoforms) examined revealed that translation-related proteins bound HSP90 rather infrequently, less overall than previously described for transcription factors [35] (Figure 1B). We re-sequenced the plasmids encoding the proteins interacting with HSP90 and validated the Hsp90 interaction for these proteins. The HSP90 interacting proteins can be categorized as mRNA-binding, tRNA, rRNA and ribosomal biogenesis as well as other miscellaneous proteins (Figure 1C).

Interestingly, two CPE element binding proteins, CPEB1 and CPEB2, were among the strong Hsp90-binding proteins. We validated the interaction of both CPEB1 and CPEB2 with Hsp90 by performing a direct immunoprecipitation assay (Fig 1D, S1). The binding of CPEB2 to Hsp90 was dependent on its predicted prion domain as CPEB2 lacking the N-terminal region containing the prion domain (ΔNCPEB2) lost the ability to interact with HSP90. Although other CPEB family members have well-categorized interacting proteins, fewer binding partners have been described for human CPEB2. We therefore adopted a proteomic approach and examined the protein interaction landscape of CPEB2 in human cells. Full-length CPEB2 and CPEB2 lacking the PrD (ΔNCPEB2) were expressed in HEK293T cells and lysates immunoprecipitated with anti-flag. Co-precipitating interactors were assessed by mass spectrometry (Fig 1E Sup table 2). Four major

functional categories of CPEB2 interacting proteins were revealed: chaperones, mitochondrial proteins, cell structure proteins and mRNA processing proteins (Fig 1E). Consistent with our immunoblot assay, the mass spectrometry results show that the Hsp90 interaction depends on the CPEB2 PrD (Fig 1E, PrD-dependent interactions in red). Indeed, the Hsp90-CPEB2 interaction shows the highest dependence on the PrD, but is not the only protein with this characteristic (Sup table 2). Thus, CPEB2 strongly binds Hsp90 in a PrD-dependent manner.

Prion-like characteristics of the human CPEB2 protein

We determined whether human CPEB proteins had prion-like properties when expressed in yeast. The four human CPEB paralogs all harbor homologous RNA-recognition motifs (RRMs) but possess distinct N-terminal regions (Fig. 2A). Overexpression of CPEB1, CPEB2 and CPEB3 was toxic to yeast cells (Fig. 2B). Curiously, this was not due to the prion domain; indeed, removal of the CPEB2 PrD increased CPEB2 overexpression toxicity (Fig. 2C). However, this trend was reversed when the PrD was deleted from CPEB3 and the homologous deletion had little effect on the toxicity of CPEB4, suggesting unique functions and/or interactions for each PrD within each CPEB paralog (Fig. 2C).

To examine the ability of the CPEB2 PrD to maintain a distinct heritable structural state, we exploited the modular nature of prion domains and the well-characterized [PSI^+] phenotype in yeast. The prion domain of the Sup35 protein can be replaced by a prion domain of interest, producing a protein chimera that can form heritable phenotypes as does the wild type [PSI^+]. The prion properties of such a chimera (PrD-Sup35C) can be visualized in yeast strains harboring an *ade1* nonsense allele. In the non-prion state these cells form red colonies and cannot grow on media lacking adenine. In the prion state, the chimeric protein aggregates and sequesters the Sup35 C-terminal translation termination activity, resulting in a read-through of the pre-mature stop codon of the ade1-14 allele, which leads to production of a functional Ade1 (nonsense suppression). In this [PRION+] state, the cells can grow in the absence of adenine and produce white colonies (Fig. 2D) [21].

To test the prion properties of the human CPEB2 protein, we overexpressed the CPEB2 PrD-SUP35C chimeric protein in cells lacking endogenous Sup35. In this isogenic background, the PrDCPEB2-Sup35C cells acquired two distinct, heritable states: exhibition of differential pigment accumulation when grown on rich media and growth on media lacking adenine (Fig. 2E, S2A). Furthermore, CPEB2 overexpressed in human cells exhibits SDS-resistant aggregation that is dependent on its prion domain (Fig. S2C). Thus, human CPEB2 exhibits many of the prion like characteristics previously described for the *Aplysia* CPEB [9, 13, 15, 20, 36]. The CPEB3 PrD exhibited prion properties similar to those of CPEB2 (Fig. S2A-B).

The hallmark of a prion phenotype is the ability to self-template and persist over many generations. In the context of prion-like proteins, their transient overexpression should be sufficient to induce a persistent phenotype [37]. The *Aplysia* neuronal CPEB, when expressed in yeast, can induce persistent activation of a translation reporter depending on its structural state [13]. The translation reporter has a cytoplasmic polyadenylation element (CPE) in its 3′ UTR, which mediates increased β-gal (β-galactosidase) translation in response to CPEB binding [8, 15]. In the soluble state, CPEB is inactive, whereas in the prion aggregated form, CPEB induces activation of the translation reporter [13, 20]. We therefore examined the ability of the different CPEBs to induce persistent translation activation of the CPEB translation reporter upon transient overexpression in yeast. Expression of the CPEB proteins was induced for 6 hours and then the induction was switched off. The relative translation activation was measured after cells were grown for ~ 7 generations in the absence of CPEB production (Fig. 2F). Transient overexpression of human CPEB2 in yeast had the strongest persistent activation of the CPEB translation reporter (Fig. 2G).

Transient Hsp90 inhibition induces persistent translation activation

The strong interaction of CPEB2 with Hsp90 prompted us to explore whether perturbation of Hsp90 activity could alter the

prion-like CPEB-mediated translation regulation in yeast. We explored if transient Hsp90 inhibition would induce the prion-like properties of CPEB2 and promote a persistent increase in translation of the reporter even when the Hsp90 inhibitor is absent. Cells overexpressing CPEB2 were grown in the presence or absence of the Hsp90 inhibitor (radicicol) for 10 generations and then grown on media free of Hsp90 inhibitors. Transient inhibition of Hsp90 was sufficient to induce a persistent effect on CPEB2-mediated β-gal translation when cells were grown in liquid media (Fig. 3A) and on solid media (Fig. 3B). Thus, suppression of Hsp90 activity augments the CPEB2-mediated translation activation in a self-sustaining manner.

Endogenous activation of the CPEB translation reporter in yeast in the absence of exogenous CPEB

Strikingly, transient Hsp90 inhibition had a significant effect on the persistent translation of the CPEB translation reporter even in the absence of exogenous CPEB (Fig. 4A). Although in the absence of exogenous CPEB the activation of the translation reporter was lower it still exhibited persistent high activity upon transient Hsp90 inhibition. This suggests the existence of an endogenous yeast mechanism that is affected by Hsp90 and regulated in a fashion similar to the regulation of exogenous CPEB. Suppression of Hsp90 activity for ten generations induced a persistent translation output in cells that was observable after growth on solid media without Hsp90 inhibitor for approximately 15 generations, as indicated after replica-plating onto x-gal plates (Fig 4B). Thus, even in the absence of exogenous CPEB, inhibition of Hsp90 induced a persistent translation output in cultures grown in either liquid or on plates.

We next explored the effect on the mRNA levels of the CPEB translation reporter (β-gal) upon treatment with an Hsp90 inhibitor and the persistent effect of this exposure ("memory state") compared to control. To do so, we initially performed quantitative PCR on the β-gal open-reading frame on both total RNA and the poly-A selected fraction. In both cases, no significant changes in mRNA levels were observed under any of the conditions (Fig 4C-D). We then performed polysomal fractionation followed by specific probing of the CPEB reporter (Figure 4E-F). In the memory state, β-gal mRNA shifted to the heavier polysomal fractions (Fig 4F), indicating greater translational throughput. Thus, the β-gal expression induced following transient Hsp90 inhibition is, indeed, due to increased translation with no detectable increase in the levels of β-gal mRNA.

In the process of these experiments, it came to our attention that the CPEB translation reporter used in the original report [13] included a predicted stem-loop structure in the 5′UTR that repressed its translation (LacZ-CPE). This structure was absent in the control reporter used in the same study (SYM-CAM) (Fig. S3A-F) [13]. To rule out the possibility that this stem loop might contribute to the CPEB-mediated translational activation, we made several adjustments to the original CPEB reporter. First, we removed the stem loop to eliminate the 5′UTR suppression. Second, we replaced the strong constitutive promoter with a weaker constitutive promoter (TDH3 promoter to SUP35 promoter) to decrease the level of transcription. Third, we replaced the long CPE-containing 3′UTR (Fig. S3A) with a short CPE element taken from the ADH1 3′UTR from *C.albicans*. Transient Hsp90 suppression was effective in inducing the persistent activation of β-gal in both the original and modified versions of the CPEB translation reporters (Fig. S3D-H). Thus, the induction of β-gal translation is not affected by altering promoters, or translation initiation inhibition in the form of a stem loop in the 5′UTR, suggesting that the regulation indeed arises from the 3′UTR.

Translation activation induced by transient Hsp90 inhibition is reversible

In yeast, prion conformational switching is a rare event and has a low frequency of spontaneous reversion [38]. Reasoning that the activation of the CPEB translation reporter would also show trans-generational persistence, we tracked the persistence of reporter translation induction by transient Hsp90 inhibition to determine its reversibility. We treated cells with an Hsp90 inhibitor or DMSO as a control, grew out single colonies from both groups in the absence of the Hsp90 inhibitor and quantified the levels of the translation reporter for a span of

almost 40 generations (schematic representation in Fig. 5A). Persistent activation of translation was significant after 20 generations and then gradually declined. Persistent activation remained significant for up to 30 generations, but not after 40 generations (Fig. 5B-C). Exploring the distribution of single colonies (Fig. 5C) further emphasized that the induction of translation and persistence were not uniform across clones despite all having been derived from a single clone prior to Hsp90 inhibition. The loss of translational memory after 40 generations indicates that this phenomenon is distinct from canonical prions as these would be expected to persist indefinitely in the lineage of most activated cells.

Aggregation of the 3′UTR-processing protein Hrp1 induces the CPEB translation reporter

Based on our finding that transient Hsp90 inhibition results in persistent activation of the CPEB translation reporter, we hypothesized that other modifications of protein homeostasis might elicit translational activation. Hrp1, a component of the 3′UTR-processing complex, was previously suggested to be the functional CPEB homolog in yeast [20]. Like CPEB, Hrp1 contains PrD and RRM domains and was recently shown to be highly aggregation prone [39]. To determine whether induced aggregation of Hrp1 promoted activation of the CPEB translation reporter, we assessed the effect of expressing an aggregation-prone mutant of Hrp1 (PY-mutant, [40]). Indeed, overexpression of the Hrp1 PY-mutant resulted in detectable aggregates (Fig. S4A). We further introduced the Hrp1 PY-mutant into the yTRAP (yeast transcriptional reporting of aggregating proteins) reporter system [39] (Fig. 6A). As expected, the expression of wt Hrp1 resulted in increased fluorescence compared to the PY-mutant (Fig. 6B), which in the yTRAP system indicates that the wt protein has greater solubility. This change in yTRAP signal was not due to decreased levels of the protein (Fig. S4B). Next, we introduced the CPEB translation reporter into the Hrp1 yTRAP strains to simultaneously monitor the Hrp1 aggregation state and the levels of translation. Under these conditions, the strains expressing wt Hrp1 exhibited a basal level of translation reporter activity, while the strains expressing the mutant aggregated form of Hrp1 exhibited heightened translational output as indicated by increased β-gal activity (Fig. 6C). These results suggest that the aggregation of a component in the 3′UTR processing machinery can mediate the induction of translation of the CPEB reporter.

Perturbation of vacuole signaling induces the CPEB translation reporter but does not lead to translation memory

A more detailed time course analysis of translation reporter activation following Hsp90 inhibition revealed that activation peaked during the beginning of an alteration in growth rate associated with the yeast diauxic shift (Fig. 7A). Surprisingly, after removal of the Hsp90 inhibitor, this synchronous translation activation persisted, suggesting that both the original induction of translation and the persistent induction of translation depend on a specific cell state (Fig. 7A, bottom). One characteristic of the diauxic shift is that, during this time, the vacuole is the source of much of the nutrient signaling [41, 42]. To explore whether vacuolar signaling is involved in regulating the CPEB translation reporter, we genetically deleted various components of vacuolar pumps and measured the effects on translation output (Fig. 7B). Interestingly, genetic suppression of Vma1, a critical component of the V-ATPase complex, increased translation of the CPEB reporter (Fig. 7B). Hsp90 inhibition did not induce apparent Vma1 aggregation or mislocalization (Fig. S5A-C) but rather resulted in reduced protein levels of Vma1 (Fig. S5C). Deletion of other components of the V-ATPase complex also induced the translation reporter, suggesting that the effect on translation is not specifically due to regulation of Vma1 but rather due to impairment in V-ATPAse functionality (Fig. S5D). Further examination of organelle morphology following Hsp90 inhibition revealed some mitochondrial structure abnormalities (Fig. S5E), suggesting that translation activation induced by Hsp90 inhibition might involve the vacuolar-mitochondrial axis.

The V-ATPase function of acidifying the vacuole can affect a large subset of vacuole functions. For instance, many vacuolar pumps depend on the proton gradient generated by the V-ATPase complex. To identify signaling components downstream of Vma1, we over-

expressed specific vacuolar pumps in the background of the Vma1 deletion to determine if any would reverse the β-gal activation observed in that strain. Overexpression of the vacuolar iron/copper transporter Fet5 completely eliminated the Vma1 deletion-induced β-gal activation (Fig. 7C). Fet5 is a multi-copper oxidase that plays a role in iron transport [43] and recently was shown to regulate iron recycling during diauxic shift [44]. Culturing cells in media depleted of iron and copper was sufficient to increase the expression of β-gal from the CPEB translation reporter and the effect was additive to radicicol activation (Fig. S5F). Thus, signals associated with iron homeostasis that arise from the vacuole and are potentially mediated by Fet5 affect the translation of the β-gal CPE-reporter.

If Hsp90 inhibition induces the persistent translation of the CPEB reporter solely through Vma1 suppression, we would expect a similar outcome with transient suppression of the V-ATPase complex. Indeed, the V-ATPase inhibitor Concanamycin A (ConcA), like the Hsp90 inhibitor, induced activation of the translation reporter (Fig. 7D). However, persistent translation activation was only achieved with Hsp90 suppression (Fig. 7D). Thus, Hsp90 inhibition is specific and unique in inducing and promoting a heritable change in protein translation.

Discussion

In this work, we establish an intriguing phenomenon of Hsp90-regulated, non-genetically inherited translation memory. We first established a strong link between the human CPE element binding protein 2 (CPEB2) translation regulation and the Hsp90 chaperone by surveying the protein-protein interactions of HSP90 with a large library of RNA-binding proteins. We discovered a strong protein-protein interaction between CPEB2 and HSP90 that is mediated by the prion-like domain of CPEB2. The functionality of this protein-protein interaction is revealed upon Hsp90 inhibition: even transient suppression of Hsp90 is sufficient to induce the activation of the CPEB translation reporter in the yeast model. This translation activation is persistent and gradually declines over 40 generations. Moreover, transient suppression of Hsp90 induces the persistent translation activation of the CPEB translation reporter even in the absence of exogenous CPEB overexpression, suggesting yeast possess an endogenous CPEB-like molecule that mediates translational memory. Although other perturbations, including iron-copper depletion and V-ATPase deletions, can affect the translational reporter, Hsp90 inhibition was the only identified stimulus that triggered persistent activation over many generations of growth. Thus, another Hsp90 interactor likely acts like CPEB2 to induce the translation of mRNAs.

The evolutionarily conserved family of CPEB proteins can be subdivided by the presence or absence of a prion domain. In *Aplysia*, *Drosophila* and mice, the prion domain has distinct functions in promoting CPEB activity [9, 13-18, 36]. In humans, there are four homologs of CPEB; three (CPEB2-4) contain a predicted prion-like domain and one does not (CPEB1). Using the translation assay originally designed to characterize the prion properties of the *Aplysia* CPEB [13], we show that the expression of human CPEB2 yields the strongest heritable activation of translation. The prion domains of both CPEB2 and CPEB3 enable prion-like inheritance. However, the function of these domains seems to be specific and unique in the context of the whole protein as its absence increases the toxicity of CPEB2 but reduces the toxicity of CPEB3. Moreover, our proteomic analysis revealed that the PrD of CPEB2 is crucial for mediating specific protein-protein interactions, including interactions with the chaperone Hsp90, proteins involved in mRNA translation regulation, tubulin and mitochondrial translocation proteins (TIM8 and TIM13). Interestingly, CPEB3 protein-protein interaction fall largely within the same functional interaction as for CPEB2 however different PrD-dependent protein-protein interaction was observed for CPEB3.

The strong interaction between the PrD of CPEB and Hsp90, along with the profound induction of the translational memory upon Hsp90 inhibition suggests that Hsp90 is a key mediator of the prion-like phenotypes of CPEB. CPEB2 overexpression was sufficient to promote the activation of translation as previously shown for the *Aplysia* CPEB [13]. However, the Hsp90 inhibition further enhanced this activation of translation in a heritable fashion. The surprising

finding is that suppression of Hsp90, even in the absence of exogenous CPEB, can induce a persistent and inherited translation phenotype in yeast. Although this activation is lower in magnitude than the one observed in the presence of CPEB, it is nevertheless sufficient to promote a heritable translation phenotype that declines gradually over 40 generations.

We believe these data are best explained by the existence of an endogenous yeast RNA-binding protein with prion-like properties with similar structure–function characteristics as described for the CPEB proteins [16-20, 45], and for simplicity will be termed [CPEB] prion. The enhancement of the [CPEB] phenotype by overexpression of exogenous CPEB and induced Hrp1 aggregation support a model of altered 3′ UTR regulation. Although the aggregated Hrp1 induced the translation, we have no additional evidence to support the fact that Hrp1 is the yeast [CPEB] prion. However, the aggregation of an element from the 3′UTR processing complex could potentially induce the sequestration or co-aggregation of the actual [CPEB] prion as reflected by the induction of translation by aggregated Hrp1. Many of the yeast RNA-binding proteins contain segments resembling prion-domains and have the ability to form stable aggregates [21, 46-48]. In some cases, as shown for Whi3 and Rim4, aggregation is an integral part of their biological function [45, 49, 50]. Supporting the possibility that the [CPEB] prion phenotype is mediated by a protein-based mechanism mediated by an RNA-binding protein.

In our model, upon transient Hsp90 suppression, the endogenous yeast prion transitions from the [cpeb] translation inactive state to the [CPEB] prion state that self-templates additional functional aggregates resulting in persistent translation activation of specific mRNAs. This prion-like phenomenon replicates enough to provide 30 generations of phenotypic consequences without the exponential dilution that would occur for non-prion assemblies. However, ultimately this prion-like state is not stable and through either its own instability or the action of Hsp90, is eliminated after additional generations. The [CPEB] prion activity following Hsp90 inhibition peaks at the diauxic shift and is also strongly affected by the V-ATPase and iron-copper signaling. This might suggest that the induction of the [CPEB] prion protein expression or stabilization (or post-translational modification) is achieved at the diauxic shift stage and is regulated by lysosomal signaling. Perturbing this regulation alters the activation of the CPEB translation reporter but does not result in heritable [CPEB] prion state with persistent translation activity as observed in the case of Hsp90 suppression (Fig. 8). Thus, Hsp90 inhibition is unique in promoting the [CPEB] prion persistent translation activation revealing yet another mechanism through which this powerful chaperone enhances phenotypic variation.

Acknowledgments

This work is dedicated to the special times and people of the Lindquist lab and to Susan Lindquist that enabled us to flourish in this amazing environment that tragically came to an end. Thanks to all the Lindquist lab members that stuck around the hard times and helped in many ways beyond the work itself, and the Lindquist lab alumni that reached out and gave much support. Special thanks to Linda Clayton for the constructive reading and comments and the general support over all the years. We would like to thank Gerald Fink for his supervision and constructive comments, Dirk Landgraf for the endogenously tagged fluorescent yeast strains and Kausik Si for the human yeast codon optimized CPEB PrD plasmids. P. Tsvetkov was supported by EMBO Fellowship ALTF 739-2011 and by the Charles A. King Trust Postdoctoral Fellowship Program. This work was supported by the Mathers foundation. S.L. was an investigator of the Howard Hughes Medical Institute.

References :

1. Sarto-Jackson I, Tomaska L. How to bake a brain: yeast as a model neuron. Curr Genet. 2016;62(2):347-70. Epub 2016/01/20. doi: 10.1007/s00294-015-0554-2. PubMed PMID: 26782173.
2. Khurana V, Lindquist S. Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? Nat Rev Neurosci. 2010;11(6):436-49. Epub 2010/04/29. doi: 10.1038/nrn2809. PubMed PMID: 20424620.
3. Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. Nature reviews Molecular cell biology. 2010;11(2):113-27. Epub 2010/01/23. doi: 10.1038/nrm2838. PubMed PMID: 20094052; PubMed Central PMCID: PMCPMC4461372.
4. Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell. 2009;136(4):731-45. Epub 2009/02/26. doi: 10.1016/j.cell.2009.01.042. PubMed PMID: 19239892; PubMed Central PMCID: PMCPMC3610329.
5. Harding HP, Novoa I, Zhang Y, Zeng H, Wek R, Schapira M, et al. Regulated translation initiation controls stress-induced gene expression in mammalian cells. Molecular cell. 2000;6(5):1099-108. Epub 2000/12/07. PubMed PMID: 11106749.
6. Richter JD. CPEB: a life in translation. Trends in biochemical sciences. 2007;32(6):279-85. doi: 10.1016/j.tibs.2007.04.004. PubMed PMID: 17481902.
7. Mendez R, Richter JD. Translational control by CPEB: a means to the end. Nature reviews Molecular cell biology. 2001;2(7):521-9. doi: 10.1038/35080081. PubMed PMID: 11433366.
8. Wu L, Wells D, Tay J, Mendis D, Abbott MA, Barnitt A, et al. CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of alpha-CaMKII mRNA at synapses. Neuron. 1998;21(5):1129-39. Epub 1998/12/18. PubMed PMID: 9856468.
9. Si K, Giustetto M, Etkin A, Hsu R, Janisiewicz AM, Miniaci MC, et al. A neuronal isoform of CPEB regulates local protein synthesis and stabilizes synapse-specific long-term facilitation in aplysia. Cell. 2003;115(7):893-904. PubMed PMID: 14697206.
10. Costa-Mattioli M, Sossin WS, Klann E, Sonenberg N. Translational control of long-lasting synaptic plasticity and memory. Neuron. 2009;61(1):10-26. Epub 2009/01/17. doi: 10.1016/j.neuron.2008.10.055. PubMed PMID: 19146809; PubMed Central PMCID: PMCPMC5154738.
11. Richter JD, Klann E. Making synaptic plasticity and memory last: mechanisms of translational regulation. Genes & development. 2009;23(1):1-11. Epub 2009/01/13. doi: 10.1101/gad.1735809. PubMed PMID: 19136621.
12. Hake LE, Mendez R, Richter JD. Specificity of RNA binding by CPEB: requirement for RNA recognition motifs and a novel zinc finger. Molecular and cellular biology. 1998;18(2):685-93. PubMed PMID: 9447964; PubMed Central PMCID: PMC108779.
13. Si K, Lindquist S, Kandel ER. A neuronal isoform of the aplysia CPEB has prion-like properties. Cell. 2003;115(7):879-91. PubMed PMID: 14697205.

14. Keleman K, Kruttner S, Alenius M, Dickson BJ. Function of the Drosophila CPEB protein Orb2 in long-term courtship memory. Nature neuroscience. 2007;10(12):1587-93. doi: 10.1038/nn1996. PubMed PMID: 17965711.
15. Si K, Choi YB, White-Grindley E, Majumdar A, Kandel ER. Aplysia CPEB can form prion-like multimers in sensory neurons that contribute to long-term facilitation. Cell. 2010;140(3):421-35. Epub 2010/02/11. doi: 10.1016/j.cell.2010.01.008. PubMed PMID: 20144764.
16. Kruttner S, Stepien B, Noordermeer JN, Mommaas MA, Mechtler K, Dickson BJ, et al. Drosophila CPEB Orb2A mediates memory independent of Its RNA-binding domain. Neuron. 2012;76(2):383-95. doi: 10.1016/j.neuron.2012.08.028. PubMed PMID: 23083740; PubMed Central PMCID: PMC3480640.
17. Majumdar A, Cesario WC, White-Grindley E, Jiang H, Ren F, Khan MR, et al. Critical role of amyloid-like oligomers of Drosophila Orb2 in the persistence of memory. Cell. 2012;148(3):515-29. doi: 10.1016/j.cell.2012.01.004. PubMed PMID: 22284910.
18. Drisaldi B, Colnaghi L, Fioriti L, Rao N, Myers C, Snyder AM, et al. SUMOylation Is an Inhibitory Constraint that Regulates the Prion-like Aggregation and Activity of CPEB3. Cell reports. 2015;11(11):1694-702. Epub 2015/06/16. doi: 10.1016/j.celrep.2015.04.061. PubMed PMID: 26074071; PubMed Central PMCID: PMCPMC5477225.
19. Khan MR, Li L, Perez-Sanchez C, Saraf A, Florens L, Slaughter BD, et al. Amyloidogenic Oligomerization Transforms Drosophila Orb2 from a Translation Repressor to an Activator. Cell. 2015;163(6):1468-83. Epub 2015/12/08. doi: 10.1016/j.cell.2015.11.020. PubMed PMID: 26638074; PubMed Central PMCID: PMCPMC4674814.
20. Heinrich SU, Lindquist S. Protein-only mechanism induces self-perpetuating changes in the activity of neuronal Aplysia cytoplasmic polyadenylation element binding protein (CPEB). Proceedings of the National Academy of Sciences of the United States of America. 2011;108(7):2999-3004. doi: 10.1073/pnas.1019368108. PubMed PMID: 21270333; PubMed Central PMCID: PMC3041084.
21. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell. 2009;137(1):146-58. doi: 10.1016/j.cell.2009.02.044. PubMed PMID: 19345193; PubMed Central PMCID: PMC2683788.
22. Chernoff YO, Lindquist SL, Ono B, Inge-Vechtomov SG, Liebman SW. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. Science. 1995;268(5212):880-4. PubMed PMID: 7754373.
23. Halfmann R, Jarosz DF, Jones SK, Chang A, Lancaster AK, Lindquist S. Prions are a common mechanism for phenotypic inheritance in wild yeasts. Nature. 2012;482(7385):363-8. doi: 10.1038/nature10875. PubMed PMID: 22337056; PubMed Central PMCID: PMC3319070.
24. Higurashi T, Hines JK, Sahi C, Aron R, Craig EA. Specificity of the J-protein Sis1 in the propagation of 3 yeast prions. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(43):16596-601. doi: 10.1073/pnas.0808934105. PubMed PMID: 18955697; PubMed Central PMCID: PMC2575465.
25. Hines JK, Li X, Du Z, Higurashi T, Li L, Craig EA. [SWI], the prion formed by the chromatin remodeling factor Swi1, is highly sensitive to alterations in Hsp70 chaperone system activity. PLoS genetics. 2011;7(2):e1001309. doi: 10.1371/journal.pgen.1001309. PubMed PMID: 21379326; PubMed Central PMCID: PMC3040656.

26. Suzuki G, Shimazu N, Tanaka M. A yeast prion, Mod5, promotes acquired drug resistance and cell survival under environmental stress. Science. 2012;336(6079):355-9. doi: 10.1126/science.1219491. PubMed PMID: 22517861.
27. Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. Nature. 2002;417(6889):618-24. doi: 10.1038/nature749. PubMed PMID: 12050657.
28. Rutherford SL, Lindquist S. Hsp90 as a capacitor for morphological evolution. Nature. 1998;396(6709):336-42. doi: 10.1038/24550. PubMed PMID: 9845070.
29. Whitesell L, Santagata S, Mendillo ML, Lin NU, Proia DA, Lindquist S. HSP90 empowers evolution of resistance to hormonal therapy in human breast cancer models. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(51):18297-302. doi: 10.1073/pnas.1421323111. PubMed PMID: 25489079; PubMed Central PMCID: PMC4280614.
30. Chakrabortee S, Kayatekin C, Newby GA, Mendillo ML, Lancaster A, Lindquist S. Luminidependens (LD) is an Arabidopsis protein with prion behavior. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(21):6065-70. Epub 2016/04/27. doi: 10.1073/pnas.1604478113. PubMed PMID: 27114519; PubMed Central PMCID: PMCPMC4889399.
31. Hou F, Sun L, Zheng H, Skaug B, Jiang QX, Chen ZJ. MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. Cell. 2011;146(3):448-61. doi: 10.1016/j.cell.2011.06.041. PubMed PMID: 21782231; PubMed Central PMCID: PMC3179916.
32. Moosavi B, Wongwigkarn J, Tuite MF. Hsp70/Hsp90 co-chaperones are required for efficient Hsp104-mediated elimination of the yeast [PSI(+)] prion but not for prion propagation. Yeast. 2010;27(3):167-79. Epub 2009/12/17. doi: 10.1002/yea.1742. PubMed PMID: 20014008.
33. Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Molecular cell. 2012;46(5):674-90. Epub 2012/06/12. doi: 10.1016/j.molcel.2012.05.021. PubMed PMID: 22681889.
34. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell. 2012;149(6):1393-406. Epub 2012/06/05. doi: 10.1016/j.cell.2012.04.031. PubMed PMID: 22658674.
35. Taipale M, Krykbaeva I, Koeva M, Kayatekin C, Westover KD, Karras GI, et al. Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. Cell. 2012;150(5):987-1001. Epub 2012/09/04. doi: 10.1016/j.cell.2012.06.047. PubMed PMID: 22939624; PubMed Central PMCID: PMCPMC3894786.
36. Raveendra BL, Siemer AB, Puthanveettil SV, Hendrickson WA, Kandel ER, McDermott AE. Characterization of prion-like conformational changes of the neuronal isoform of Aplysia CPEB. Nat Struct Mol Biol. 2013;20(4):495-501. Epub 2013/02/26. doi: 10.1038/nsmb.2503. PubMed PMID: 23435382; PubMed Central PMCID: PMCPMC5518672.
37. Chakrabortee S, Byers JS, Jones S, Garcia DM, Bhullar B, Chang A, et al. Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. Cell. 2016;167(2):369-81 e12. Epub 2016/10/04. doi: 10.1016/j.cell.2016.09.017. PubMed PMID: 27693355; PubMed Central PMCID: PMCPMC5066306.
38. Lancaster AK, Bardill JP, True HL, Masel J. The spontaneous appearance rate of the yeast prion [PSI+] and its implications for the evolution of the evolvability properties of the [PSI+]

system. Genetics. 2010;184(2):393-400. Epub 2009/11/18. doi: 10.1534/genetics.109.110213. PubMed PMID: 19917766; PubMed Central PMCID: PMCPMC2828720.

39. Newby GA, Kiriakov S, Hallacli E, Kayatekin C, Tsvetkov P, Mancuso CP, et al. A Genetic Tool to Track Protein Aggregates and Control Prion Inheritance. Cell. 2017;171(4):966-79 e18. Epub 2017/10/24. doi: 10.1016/j.cell.2017.09.041. PubMed PMID: 29056345; PubMed Central PMCID: PMCPMC5675731.

40. Lange A, Mills RE, Devine SE, Corbett AH. A PY-NLS nuclear targeting signal is required for nuclear localization and function of the Saccharomyces cerevisiae mRNA-binding protein Hrp1. J Biol Chem. 2008;283(19):12926-34. Epub 2008/03/18. doi: 10.1074/jbc.M800898200. PubMed PMID: 18343812; PubMed Central PMCID: PMCPMC2442336.

41. Murphy JP, Stepanova E, Everley RA, Paulo JA, Gygi SP. Comprehensive Temporal Protein Dynamics during the Diauxic Shift in Saccharomyces cerevisiae. Mol Cell Proteomics. 2015;14(9):2454-65. Epub 2015/06/17. doi: 10.1074/mcp.M114.045849. PubMed PMID: 26077900; PubMed Central PMCID: PMCPMC4563728.

42. Zampar GG, Kummel A, Ewald J, Jol S, Niebel B, Picotti P, et al. Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. Mol Syst Biol. 2013;9:651. Epub 2013/04/04. doi: 10.1038/msb.2013.11. PubMed PMID: 23549479; PubMed Central PMCID: PMCPMC3693829.

43. Spizzo T, Byersdorfer C, Duesterhoeft S, Eide D. The yeast FET5 gene encodes a FET3-related multicopper oxidase implicated in iron transport. Mol Gen Genet. 1997;256(5):547-56. Epub 1997/12/31. PubMed PMID: 9413439.

44. Horie T, Kawamata T, Matsunami M, Ohsumi Y. Recycling of iron via autophagy is critical for the transition from glycolytic to respiratory growth. J Biol Chem. 2017;292(20):8533-43. Epub 2017/03/23. doi: 10.1074/jbc.M116.762963. PubMed PMID: 28320861; PubMed Central PMCID: PMCPMC5437257.

45. Berchowitz LE, Kabachinski G, Walker MR, Carlile TM, Gilbert WV, Schwartz TU, et al. Regulated Formation of an Amyloid-like Translational Repressor Governs Gametogenesis. Cell. 2015;163(2):406-18. Epub 2015/09/29. doi: 10.1016/j.cell.2015.08.060. PubMed PMID: 26411291; PubMed Central PMCID: PMCPMC4600466.

46. Maharana S, Wang J, Papadopoulos DK, Richter D, Pozniakovsky A, Poser I, et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science. 2018;360(6391):918-21. Epub 2018/04/14. doi: 10.1126/science.aar7366. PubMed PMID: 29650702.

47. Harrison AF, Shorter J. RNA-binding proteins with prion-like domains in health and disease. Biochem J. 2017;474(8):1417-38. Epub 2017/04/09. doi: 10.1042/BCJ20160499. PubMed PMID: 28389532; PubMed Central PMCID: PMCPMC5639257.

48. King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain research. 2012;1462:61-80. Epub 2012/03/27. doi: 10.1016/j.brainres.2012.01.016. PubMed PMID: 22445064; PubMed Central PMCID: PMCPMC3372647.

49. Carpenter K, Bell RB, Yunus J, Amon A, Berchowitz LE. Phosphorylation-Mediated Clearance of Amyloid-like Assemblies in Meiosis. Dev Cell. 2018;45(3):392-405 e6. Epub 2018/05/09. doi: 10.1016/j.devcel.2018.04.001. PubMed PMID: 29738715; PubMed Central PMCID: PMCPMC5944619.

50. Caudron F, Barral Y. A super-assembly of Whi3 encodes memory of deceptive encounters by single cells during yeast courtship. Cell. 2013;155(6):1244-57. Epub 2013/12/10. doi: 10.1016/j.cell.2013.10.046. PubMed PMID: 24315096.
51. Yang X, Boehm JS, Yang X, Salehi-Ashtiani K, Hao T, Shen Y, et al. A public genome-scale lentiviral expression library of human ORFs. Nat Methods. 2011;8(8):659-61. Epub 2011/06/28. doi: 10.1038/nmeth.1638. PubMed PMID: 21706014; PubMed Central PMCID: PMCPMC3234135.
52. Taipale M, Tucker G, Peng J, Krykbaeva I, Lin ZY, Larsen B, et al. A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. Cell. 2014;158(2):434-48. Epub 2014/07/19. doi: 10.1016/j.cell.2014.05.039. PubMed PMID: 25036637; PubMed Central PMCID: PMCPMC4104544.
53. Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP. Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature. 2014;508(7494):66-71. Epub 2014/01/31. doi: 10.1038/nature13007. PubMed PMID: 24476825; PubMed Central PMCID: PMCPMC4086860.
54. Landgraf D, Huh D, Hallacli E, Lindquist S. Scarless Gene Tagging with One-Step Transformation and Two-Step Selection in Saccharomyces cerevisiae and Schizosaccharomyces pombe. PLoS One. 2016;11(10):e0163950. Epub 2016/10/14. doi: 10.1371/journal.pone.0163950. PubMed PMID: 27736907; PubMed Central PMCID: PMCPMC5063382.
55. Zheng X, Krakowiak J, Patel N, Beyzavi A, Ezike J, Khalil AS, et al. Dynamic control of Hsf1 during heat shock by a chaperone switch and phosphorylation. Elife. 2016;5. Epub 2016/11/11. doi: 10.7554/eLife.18638. PubMed PMID: 27831465; PubMed Central PMCID: PMCPMC5127643.
56. Teste MA, Duquenne M, Francois JM, Parrou JL. Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. BMC Mol Biol. 2009;10:99. Epub 2009/10/31. doi: 10.1186/1471-2199-10-99. PubMed PMID: 19874630; PubMed Central PMCID: PMCPMC2776018.

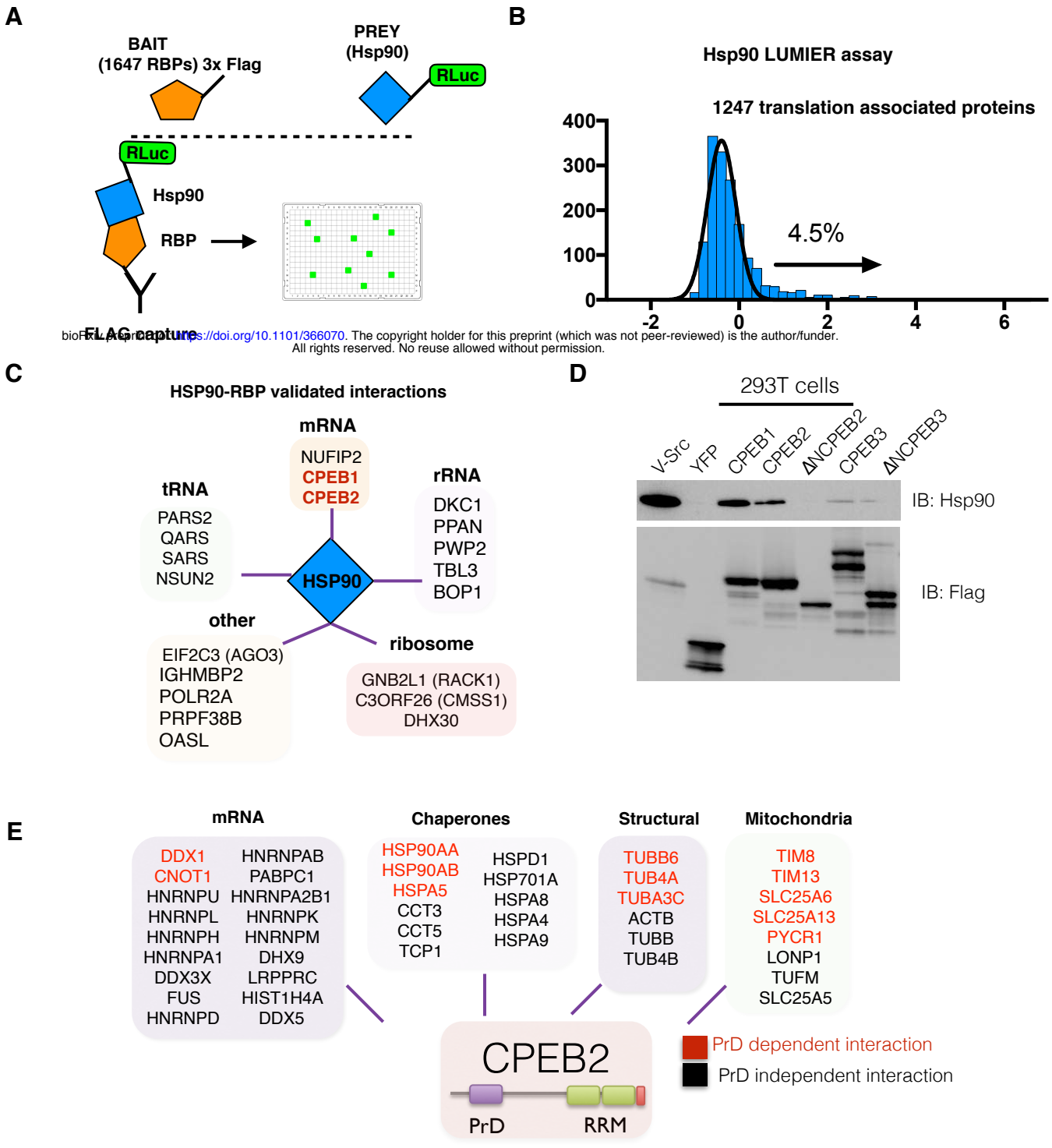
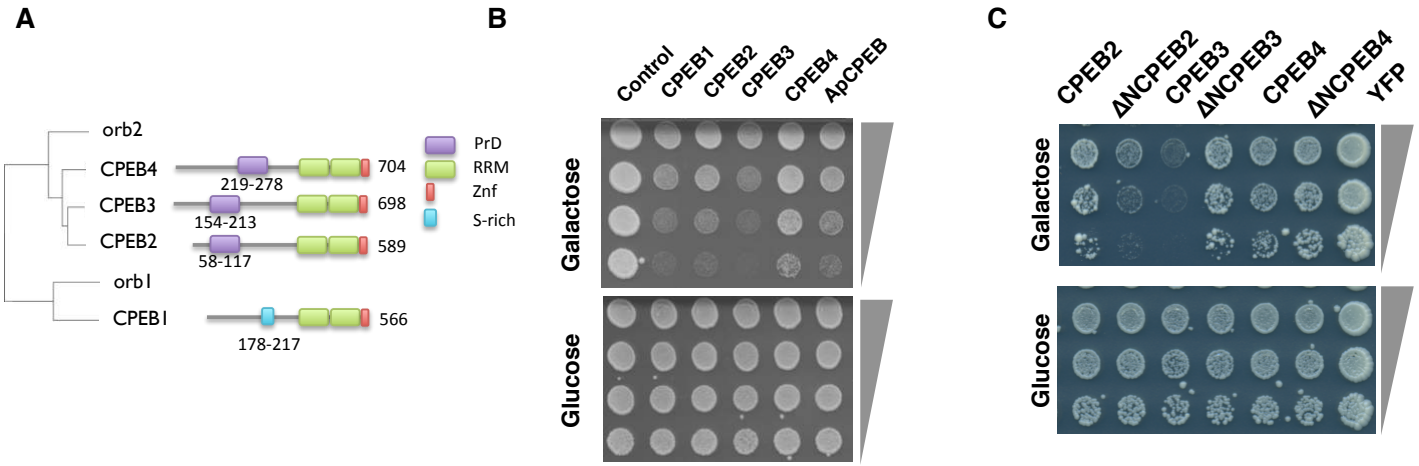


Figure 1. Proteomic approaches reveal a strong interaction of HSP90 with the CPE element binding protein 2 (CPEB2).
(A) Schematic representation of the LUMIER assay for high throughput characterization of protein-protein interactions. 1647 Flag-tagged bait proteins (RBPs) are over expressed in 293T cells over expressing Hsp90 tagged with renila (Prey). Flag capture is performed in lates and the degree of protein-protein interaction is assessed by luminescence intensity .
(B) Distribution of the Z-scores of all examined RBP interactions with Hsp90. 4.5% of the 1247 unique proteins (1647 with different isoforms) analyzed RBPs have a Z-score > 1.5.
(C) Schematic of the classification of hits from the LUMIER assay that were validated to strongly bind HSP90 (raw data in Fig. S1). The CPE-element binding (CPEB) proteins are marked in red.
(D) Flag-tagged CPEB1-3 proteins, CPEB2-3 lacking their PrD (DNCPEBs) and GFP and v-Src (used as controls) were over expressed in 293T cells and lysates immunoprecipitated with anti-flag beads. Interaction with the endogenous HSP90 was evaluated by immunoblotting with the indicated antibodies.
(E) Schematic representation of proteins shown to associate with CPEB2 by mass spectrometry when over expressed in 293T cells. Proteins are classified according to the most abundant categories (Chaperones, Mitochondrial, Structural or mRNA-associated proteins). In red are the specific protein-protein interactions that were lost when the PrD domain of CPEB2 was deleted.



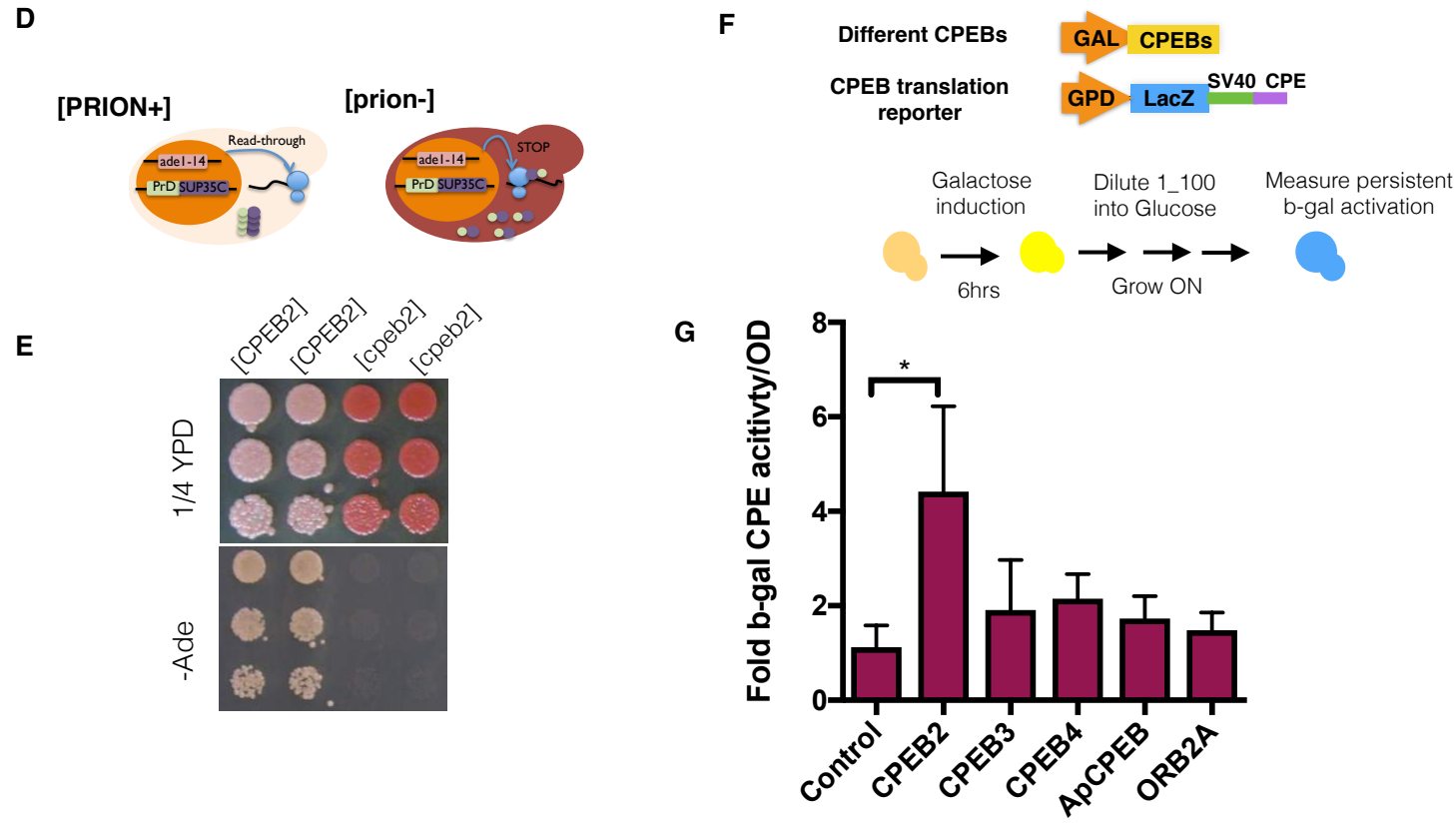


Figure 2. Characterizing human CPEB prion-like properties and Hsp90 interactions.
(A) Phylogenetic tree representation of the human CPEBs, *Drosophila* (Orb2) and predicted yeast (Hrp1) orthologs. The functional domains (RRM- green, Zinc finger- red, Serine-rich-turquoise and PrD- purple) are shown on the four human CPEB homologs.
(B-C) Yeast harboring Gal promoter full length (B) or PrD deleted (C) CPEBs tagged with mKate were grown overnight and then spotted onto expression-inducing plates (Galactose, upper) and uninducing plates (Glucose, lower) with 4-fold serial dilutions.
(D) Schematic representation of the PrD-Sup35C assay. Cells in the [PRION+] state will read through the ade 1-14 mutation to produce a functional Ade product. This enables growth on plates lacking adenine. The resulting colonies are white due to lack of accumulation of red pigment.
(E) Sup35C assay of the CPEB2 PrD. Shown is the ability of two distinct heritable conformations of the CPEB2PrD-Sup35C , white [CPEB2] and the red [cpeb2], to grow on medium lacking adenine and ¼ YPD.
(F-G) Schematic representation of the experiment (F). 426 CPEBs under the control of the Gal promoter and tagged with mKate were introduced into the W303 alpha GEM strain. These yeast were then mated to a W303 strain expressing the CPEB β-gal translation reporter. The overexpression of CPEBs was achieved by incubating the cells ON in raffinose and then diluting them to an OD of 0.1 in 2% galactose. After 6 hours, the cells were diluted 1:100 into glucose-containing media and grown overnight (~ 7 generations). The OD600 and β-gal activities were then measured (G). The mean and SD of activity is presented, unpaired t-test was used for statistical analysis.

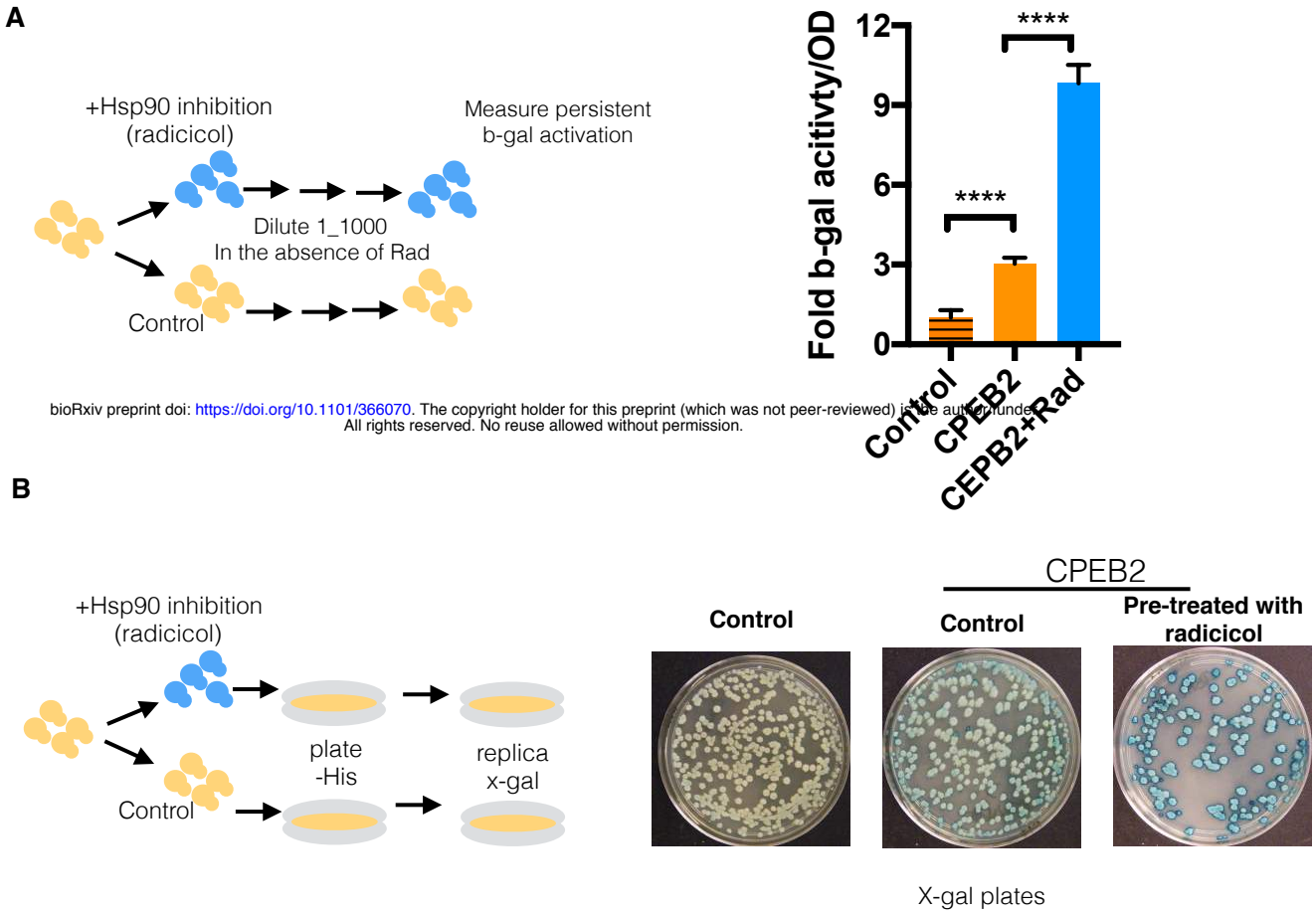


Figure 3. Transient suppression of Hsp90 induces a persistent translation activation.
(A) Cells were grown overnight in media with or without 5uM radicicol. The cells were then diluted 1:1000 and grown for an additional 24 hours in media free of Hsp90 inhibitors to assess the persistent effect of transient Hsp90 inhibition on translation as measured by β-gal activity. The mean and SD of activity is presented, unpaired t-test was used for statistical analysis. **** $p<0.0001$
(B) CPEB2-overexpressing cells that harbor the CPEB translation reporter were treated with 10uM radicicol overnight and then plated on plates lacking radicicol. Cells were further replica plated on x-gal-containing plates to visualize the β-gal activity. Cells harboring the translation reporter but without CPEB2 overexpression were used as a control.

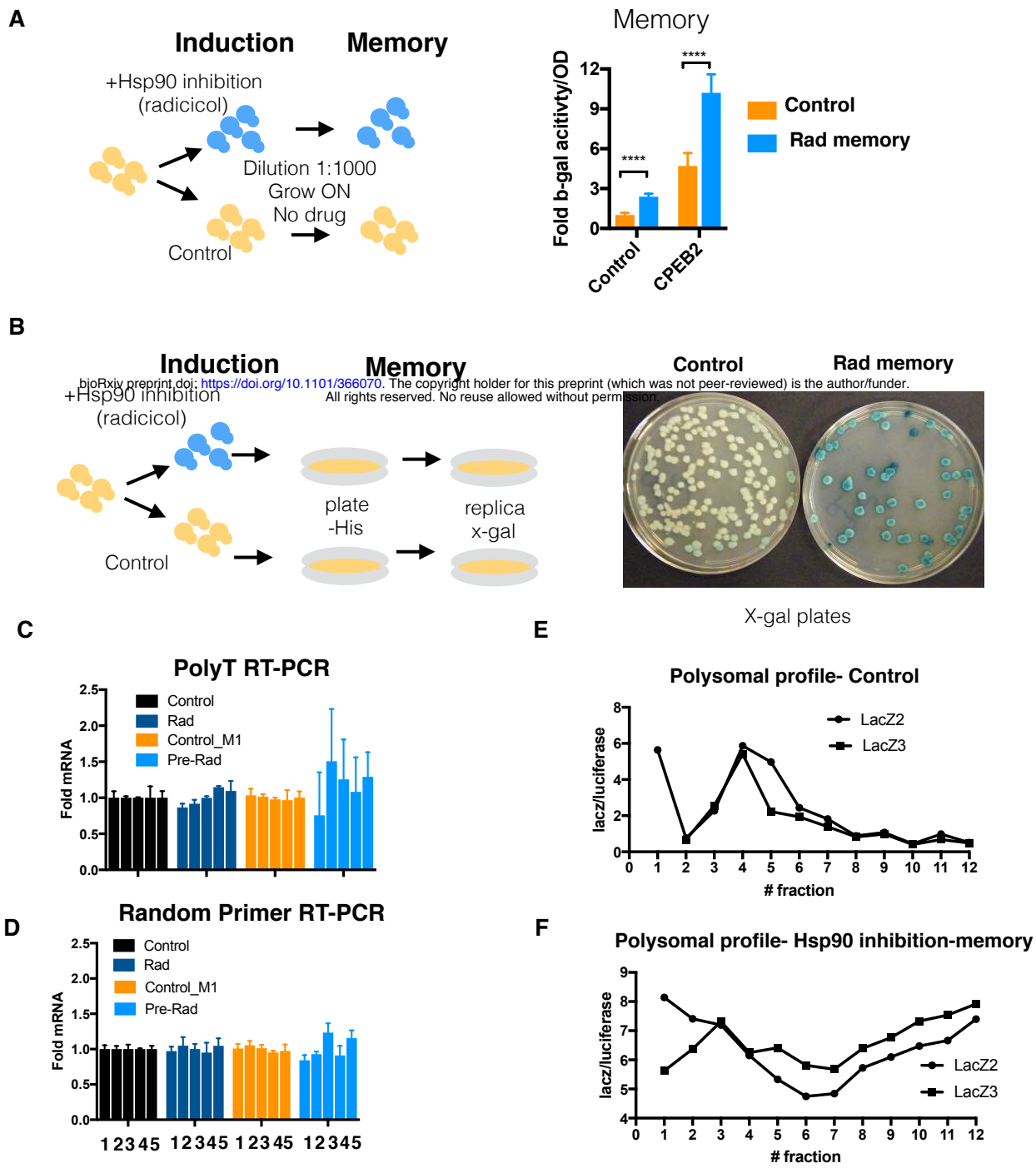


Figure 4. Transient suppression of Hsp90 activity induces a persistent translation activation in the absence of exogenous CPEB over expression.

(A) Control and CPEB2-overexpressing cells were grown overnight with and without 5uM radicicol. Cells were then diluted 1:1000 and grown 24 hours in media lacking Hsp90 inhibitors The persistent effect of transient Hsp90 inhibition on translation was examined by measuring the β-gal activity. Results represent the mean values +- SD. Paired t-test analysis was conducted **** $p < 0.001$

(B) Cells were grown in liquid in the presence or absence of 5 uM radicicol and then plated on plates devoid of the Hsp90 inhibitor. After cells formed colonies, the plates were replica plated onto x-gal-containing plates to assess the persistence of β-gal activity.

(C-D) The relative levels of β-gal open reading frame mRNA in the poly-A selected mRNA pool PolyT RT-PCR) (C) and total RNA pool (Random Primer RT-PCR) (D) in cells that were either treated with radicicol (rad) or control (Control) and then diluted to grow 10 generations in the absence of the drug (Pre-Rad) and (control-M1), respectively.

(E-F) The distribution of β-gal mRNA across the polysomal fractions of control cells and cells that grew out from a 1:1000 dilution of a 10uM radiciciol-treated culture (Hsp90 inhibition-memory). LacZ2 and LacZ3 are two distinct primer sets used to amplify the β-gal mRNA and normalized to Luciferase (that was spiked into each fraction prior to analysis) .

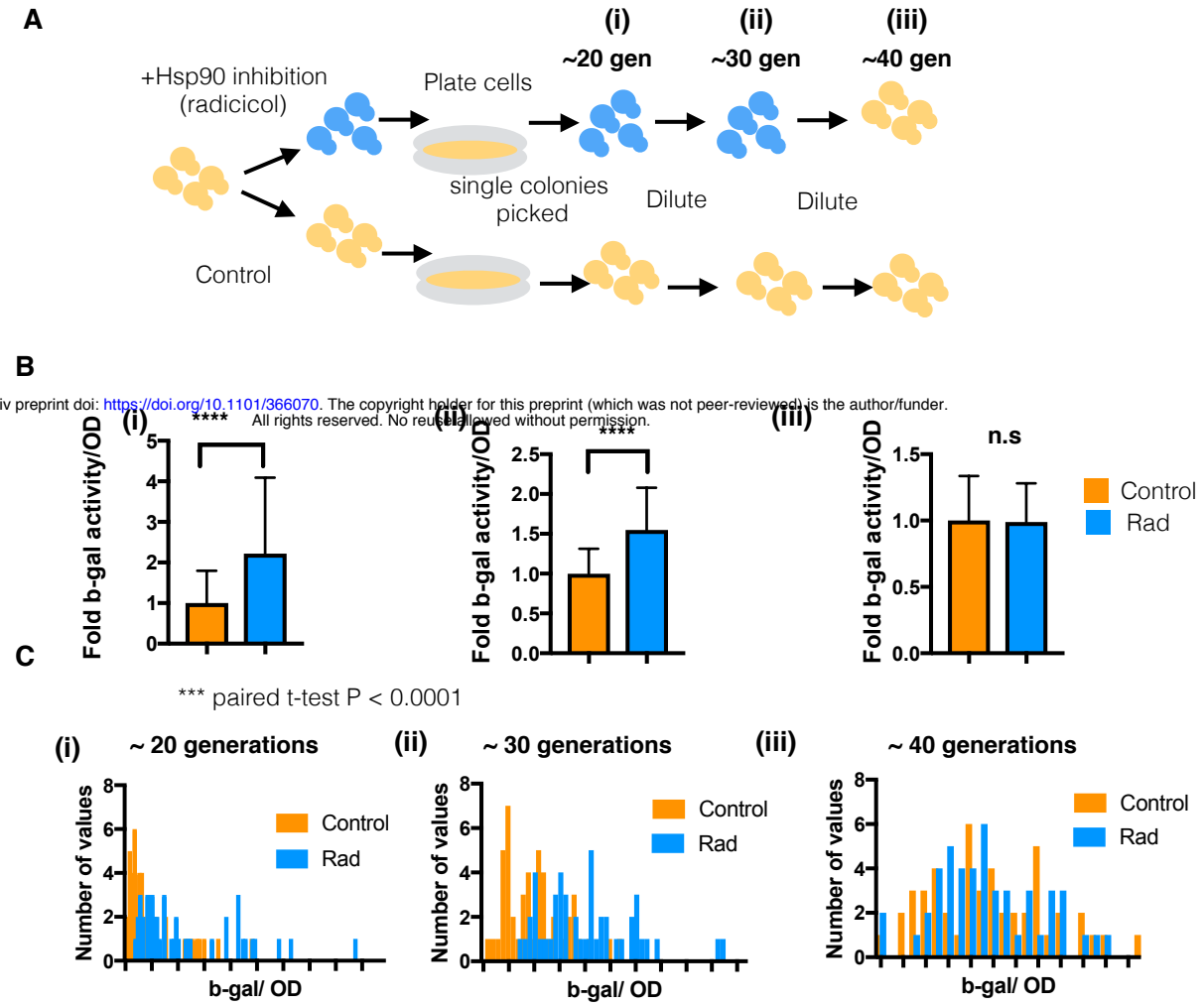


Figure 5. Endogenous activation of the CPEB translation reporter in yeast in the absence of exogenous CPEB

Cells were grown in the presence or absence of 10uM radicicol and then plated on plates devoid of the Hsp90 inhibitor. After cells formed colonies, the plates were replica plated onto x-gal-containing plates to assess the persistence of β-gal activity.

(A-C) Cells were grown in the presence or absence of 10uM radicicol and then plated on plates devoid of the Hsp90 inhibitor. 48 individual colonies were picked and β-gal activity assessed after overnight growth (i). β-gal activity was further measured after dilution (1:1000) and an additional 24-hour growth period (ii), and again after sequential dilution (1:1000) and another 24-hour growth period (iii). Paired t-tests were conducted for statistical analysis. The schematic representation of the experiment (A), the median +- SD of the β-gal activity (B) and the overall frequency distribution of the individual colonies examined (C) are presented for the different time points. *** p<0.001.

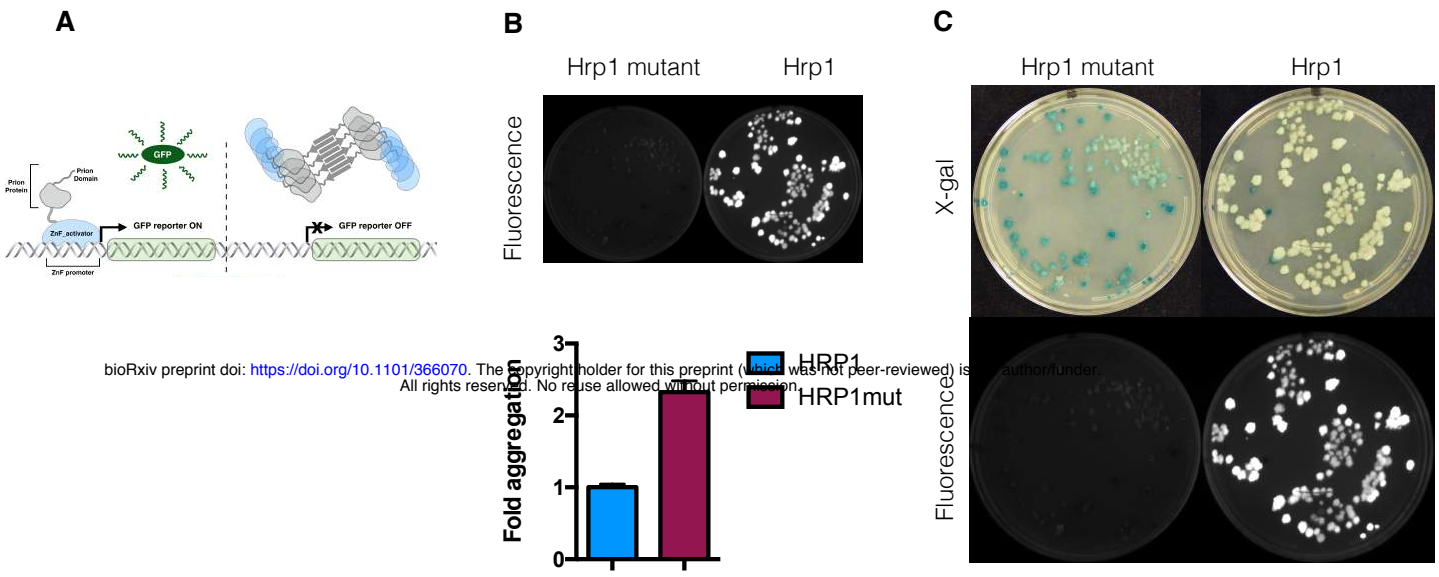


Figure 6. Aggregation of the 3'UTR-processing protein Hrp1 induces the CPEB translation reporter
(A) Schematic of yTRAP sensor mechanism
(B) Visualization (top) and quantification (bottom) of fluorescence in cells harboring the wild-type (wt) and PY521/2AA mutant (mut) Hrp1 yTRAP sensor.
(C) Cells harboring the wild-type or mutant Hrp1 yTRAP sensor and the CPEB translation reporter were replica plated onto x-gal plates and the fluorescent intensity and β-gal activity are shown.

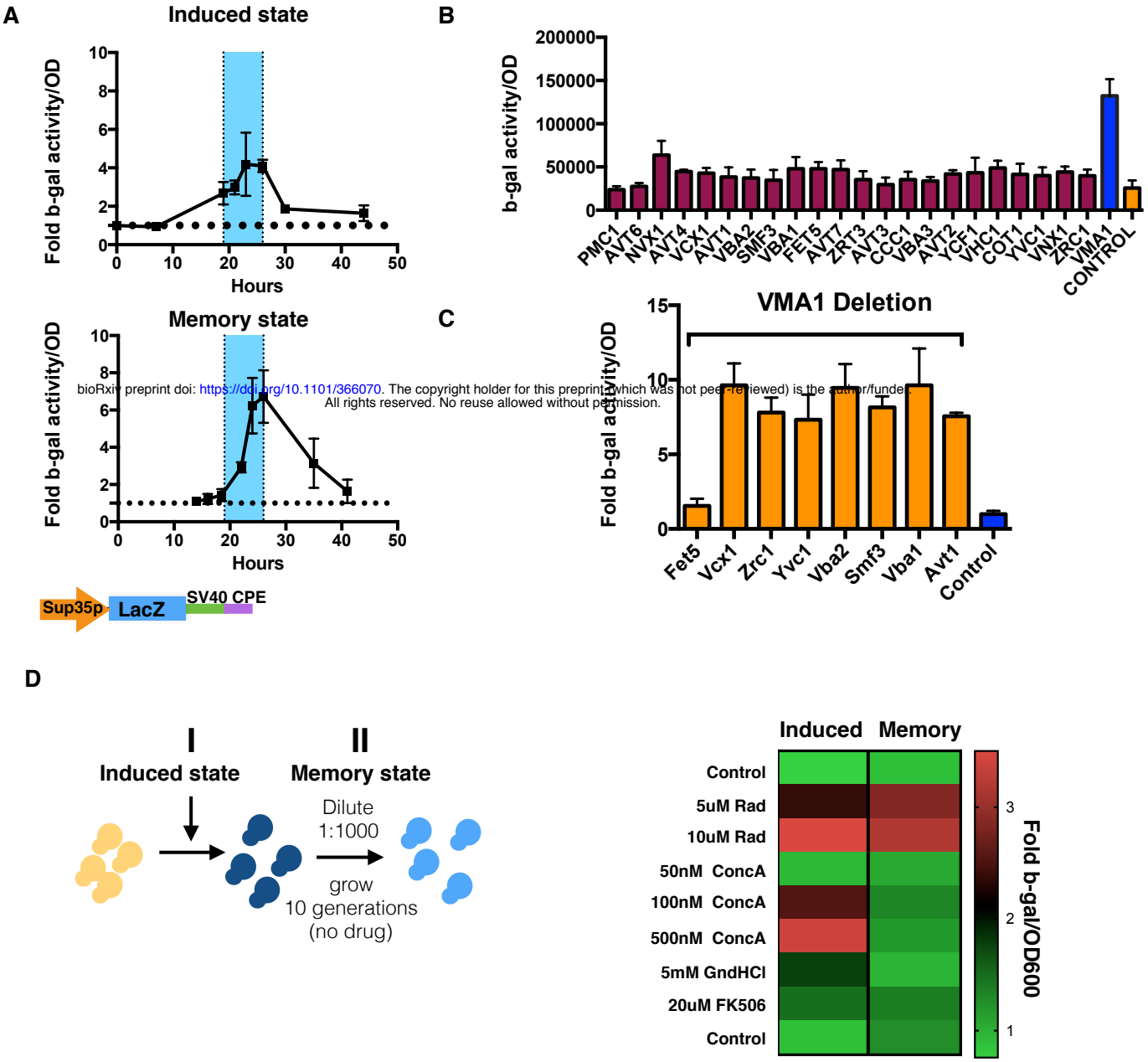


Figure 7. Perturbation of the vacuole-mitochondria axis induces the activation of translation but the effect is not persistent.

(A) The time course of fold activation of β-gal was analyzed when cells were grown in the presence or absence of the Hsp90 inhibitor radicicol (upper, induced state) and after cells were diluted 1:1000 and grown in the absence of the drug (lower memory state).

(B) β-gal activity was analyzed in strains deleted for different vacuole-associated proteins that harbored the CPEB translation reporter.

(C) The indicated vacuolar pumps were overexpressed on the background of VMA1 deletion and the activity of the CPEB translation reporter was analyzed.

(D) The induction and persistence of β-gal activity was analyzed in cells harboring the CPEB translation reporter. Cells were treated with the indicated concentration of drugs (ConcA- concanamycin A (V-ATPase inhibitor), GndHCl (Hsp104 inhibitor), FK506 (Calcineurin inhibitor). The persistent effect was analyzed after cells were diluted 1:500 and grown in the absence of the drugs for 24 hours.

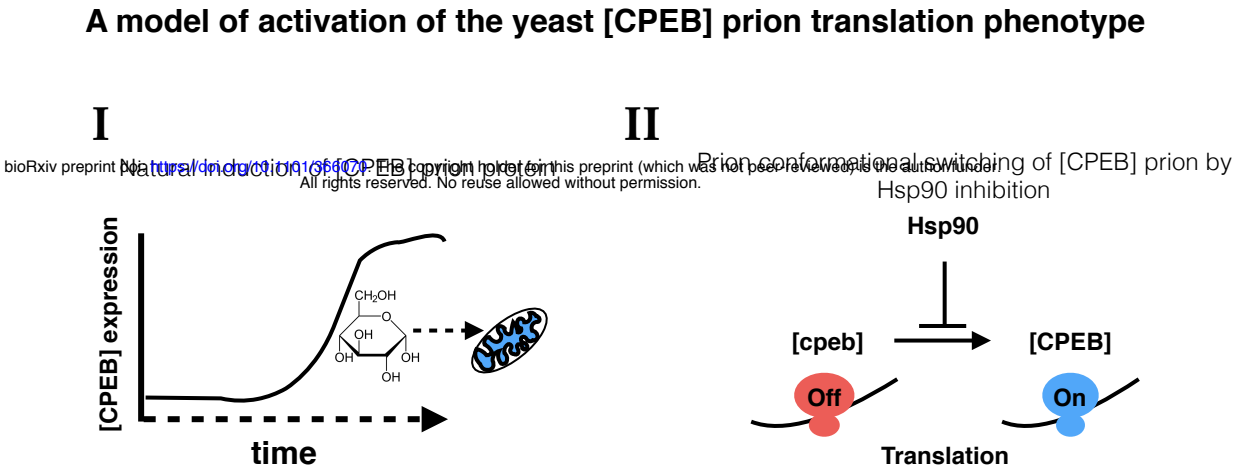


Figure 8. **A model of Hsp90 inhibition-induced translation memory.**
(i) The protein mediating the expression of the [CPEB] prion phenotype is induced during the transition from glycolysis to mitochondrial-dependent metabolism and is affected by signaling from the vacuole.
(ii) Suppression of Hsp90 induces the [CPEB] prion conformational switching. Increased prion switching with Hsp90 inhibition will enhance translation and enable epigenetic inheritance of an altered translation state.
In this model in order for the [CPEB] prion translation phenotype to be induced there is a need for the activation of the mediator prion protein (i) and the Hsp90 induced switch to enable the persistence and inheritance of the [CPEB] prion translation state (ii).

Methods

Clones and Human cell lines – To create the translation related gene library, genes with known RNA binding domains (pfam), or those annotated to have translation regulating function by Gene Ontology, Uniprot, and literature curated. The specific clones were collected from the human ORFeome [51], and those absent were cloned from cDNA by PCR. Clones were transferred into pcDNA3.1-based mammalian expression vector containing a C-terminal 3 × FLAG tag [35, 52] using Gateway recombination. Inserts were verified using restriction digestion (BsrGI, NEB) and Sanger sequencing. HEK293T cells Reporter cell lines stably expressing Renilla luciferase (Rluc)-tagged HSP90 (HSP90β) were previously described [35].

LUMIER assay – The LUMIER assay was conducted as previously described [35]. 3 × FLAG-tagged bait proteins were transfected into stable 293T cell line in 96-well format with polyethylemine (PEI). Two days after transfection, cells were washed in 1 × PBS and lysed in ice-cold HENG buffer (50 mM HEPES-KOH [pH 7.9], 150 mM NaCl, 20 mM Na_2MoO_4, 2mM EDTA, 5% glycerol, 0.5% Triton X-100). The lysate was transferred to 384-well plates coated with anti-FLAG M2 antibody (Sigma-Aldrich). Plates were incubated in cold room for 3 hr, after which plates were washed with the lysis buffer using an automated plate washer. Luminescence in each well was measured with an Envision plate reader using Gaussia FLEX luciferase kit (New England Biolabs). The normalized luminescence Z scores was used as a quantitative interaction measure.

Immunoprecipitation – HEK293T cells were transfected with 3xflag tagged bait proteins. 48 hours after transfection cells were washed in 1 × PBS and lysed in ice-cold HENG buffer. Proteins were immunoprecipitated using flag conjugated beads at 4^{o}C for 3 hours. Beads were then washed in HENG buffer to remove non-specific interactions and immunoprecipitated proteins were removed from beads by addition of loading buffer and heating 5min 95^{o}C. Samples were then loaded on Bis-Tris 4%–12% PAGE, and analyzed by western blotting on nitrocellulose membranes.

SDD-AGE- SDD-AGE was performed as detailed in [21]. A final concentration of 1% SDS was used in the loading buffer.

Read through assay- *ade1-14* read-through was assessed as detailed in [21]. Briefly, the CPEBPrD-Sup35C chimeras were individually expressed in the YRS100 strain (a derivative of YMJ584), where the endogenous Sup35 protein has been deleted, but the activity of this essential protein is covered by 416GPDSup35C. Then a plasmid shuffle was performed by plating the cells in 5-fluoroorotic acid (5-FOA). Therefore, the CPEB PrD-Sup35C was the only source of Sup35C activity in these cells.

Polysome profiling- Yeast lysates were prepared as described [53] with minor modifications. Briefly, samples were rapidly collected using filtration and ground to a fine powder using a Freezer/Mill (SPEX SamplePrep) and stored at -80 °C. Lysates were thawed and resuspended in lysis buffer (10mM Tris-HCl, pH 7.4, 5mM magnesium

chloride, 1% Triton X-100, 1% sodium deoxycholate, 100mM KCl, 0.02U/µL Superase-IN, 2mM DTT, and 100µg/mL cycloheximide), and homogenized on a Disruptor Genie (Scientific Industries) for 1 minute. The supernatant was loaded onto a 10-50% w/v linear sucrose gradient containing 20mM HEPES-KOH, pH 7.4, 5mM MgCl2, 100mM KCl, 2mM DTT, 100µg/mL cycloheximide, and 0.02U/µL Superase-IN. Gradients were centrifuged for 2hr at 36,000 r.p.m. and 4°C and then fractionated by upward displacement using a Biocomp Gradient Station with continuous absorbance monitoring at 254nm. Luciferase RNA (Promega) for qPCR normalization was added to fractions of equal volume immediately after collection. Total RNA was extracted using TRI reagent (Ambion) according to the manufacturer's instructions.

Mass spectrometry analysis-, immunoprecipitation was conducted as described above and analysis of proteins was conducted from the SDS-PAGE gel lanes for each sample were subdivided into 15 molecular weight regions. These gel bands were reduced, alkylated and digested with trypsin at 37°C overnight. The resulting peptides were extracted, concentrated and injected onto a Waters HPLC equipped with a self-packed Jupiter 3 µm C18 analytical column (0.075 mm by 10 cm, Phenomenex). Peptides were eluted using standard reverse-phase gradients. The effluent from the column was analyzed using a Thermo LTQ linear ion trap mass spectrometer (nanospray configuration) operated in a data dependent manner. The resulting fragmentation spectra were correlated against the known database using SEQUEST. Scaffold Q+S (Proteome Software) was used to provide consensus reports for the identified proteins.

Creating the new CPEB translation reporters- The original 315 LacZ CPE reporter was the original reporter described (ref), and the control SYM-CAM was also previously described (ref).

Removing the loop from original LacZ-CPE reporter- Removing the loop at the 5'UTR of the original LacZ-CPE reporter was done by digesting the plasmid with NcoI and SpeI, gel purification of the digested fragment and ligation. The plasmid was validated by sequencing to ensure that the plasmid retained the ORF sequence. This resulted in the LacZ-CPE no loop CPEB translation reporter.

Changing the Leu2 cassette to HIS3- The Leu promoter driving Leu2 gene was cut out from original reporter with BrgI and KapI digest. TEF promoter driving HIS3 gene with a TEF terminator were inserted by Gibson assembly following PCR reaction with the following primers.

His Fw Primer: gtttggccgagcggtctaagGACATGGAGGCCCAGAATAC

His Rv Primer: gaatttcatttataaagtttatCAGTATAGCGACCAGCATTC

Replacing TDH3 promoter with loop with Sup35 promoter- Sup35 promoter was used to replace the TDH3 promoter and the strong stem loop at the 5' UTR of the 315 LacZ CPE reporter. This was done by PCR of the SUP35p using the following primers:

Fw Sup35p: ctcgccatttcaaagaatacCAACCACACAAAAATCATACAAC

Rv Sup35p: tctttttggctccatggcaTGTTGCTAGTGGGCAGATATAG

The reporter was digested with HindIII and SnaBI and the Sup35 promoter was inserted using Gibson assembly. Resulting in the Sup35p Lacz-Sv40-CPE translation reporter.

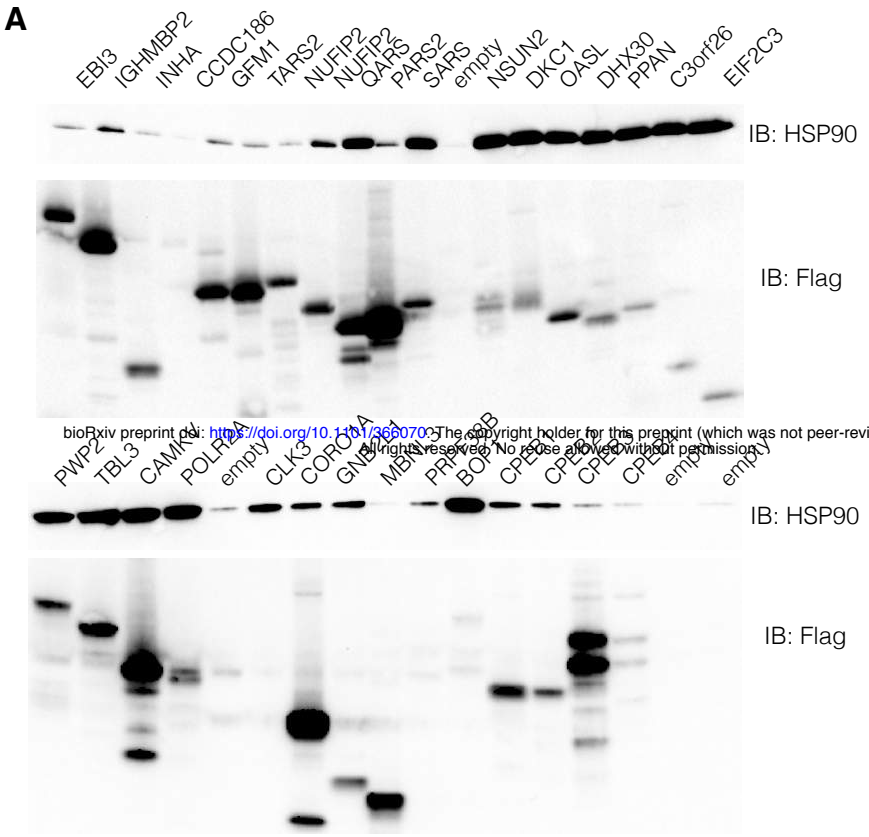
Replacing the SV40 CPE 3'UTR with an ADH1 terminator – initially a NheI digest site was introduced after the stop codons of LacZ to enable the manipulation of the 3'UTR

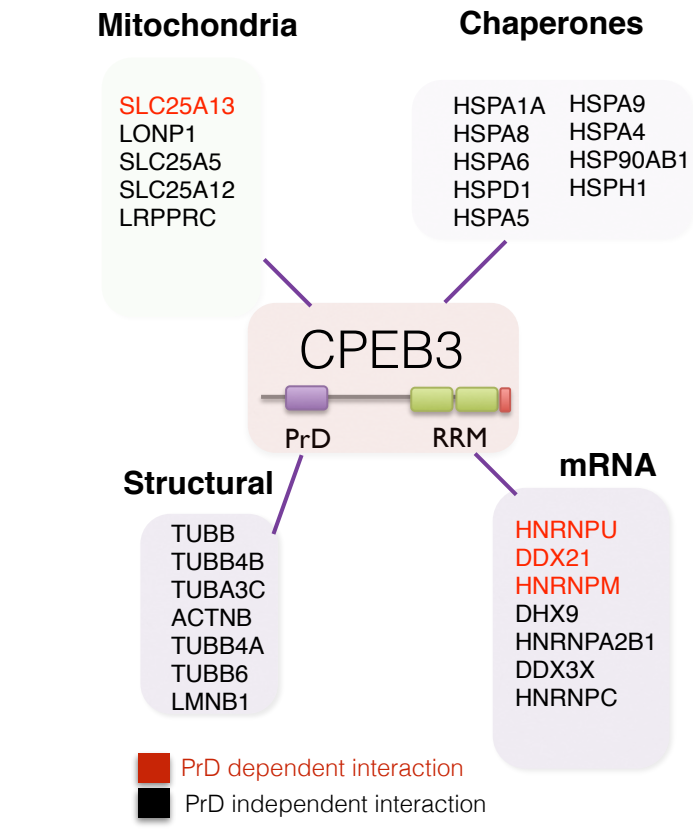
sequence. The NheI digest site was introduced by PCR of the backbone plasmid with the following primers:
Fw Backbone (NheI) (51-mer):
TGGTCTGGTGTCAAAAATAATAATAAgctagcCCGGGCAGGCCATGTCTGC
Rv backbone (54-mer):
CTCTTCTTTTTTGGCTCCATGGCAgtttaaacTGTTGCTAGTGGGCAGATATAG
The LacZ ORF was amplified with the following primers:
Fw LacZ (PmeI) (30-mer): gtttaaacTGCCATGGAGCCAAAAAAGAAG
Rv PCR lacZ (34-mer): GCTagcTTaTTaTTATTTTTGACACCAGACCAAC
The two amplified segment were assembled using Gibson assembly.
The Sup35p LacZ-SV40-CPE CPEB translation reporter with the new NheI digest site was digested with NheI SacII to remove the SV40-CPE 3'UTR and an ADH1 3'UTr was Amplified using the following primers:
Fw ADH1: gtctggtgtcaaaaataataataagctagcTAAGCAAATAGCTAAATTATATACG
Rv ADH1:
cgaattggagctccaccgcggtggcggccgCAACTGTATAAGATAGTAATAAAAATATCG
The fragments were assembled by Gibson assembly method to yield the Sup35p-LacZ-ADH1 CPEB translation reporter .

Vma1-mCherry tagging- The tagging was performed as previously described [54] using pFA6a_mCherry_HPHMX as template and the following primers :
VMA1Cterminal tagging f1: CGGCCTTGTCTGATAGTGATAAGATTACTTTGGATG
VMA1Cterminal tagging r1:
ATCAACCTGTAGGGTTCTATCGGTAGATTCAGCAAATCTTTCTTGCATAGTGCTCAA
C
VMA1Cterminal tagging f2:
GTTTAAACGAGCTCGAATTCGATATATGTAGCATTTATCTTCTGGTATATTTGTTAG
GTTTAAACGAGCTCGAATTCGATATATGTAGCATTTATCTTCTGGTATATTTGTTAG
VMA1Cterminal tagging r2: CTACCTCATAATGGATCTAAATTGCATACTAATCTCAC

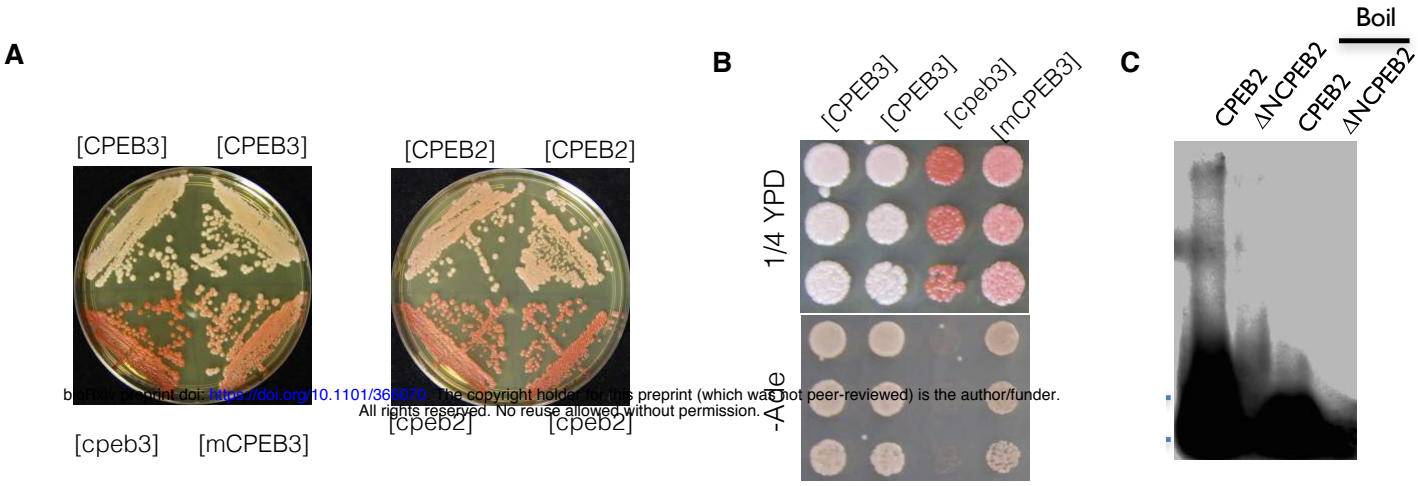
Quantitative PCR analysis- total RNA was purified via phenol/chloroform separation using phase lock tubes (five prime) followed by ethanol precipitation as previously described [55]. RT-PCR PCR was conducted with either poly-T primers or random primers using Superscript III (Thermo fisher) according to manufactures protocol. qPCR was performed with the primers for LacZ (set of 5 pairs) and the genes for normalization [56] (ALG9, TAF10, TFC1)
fLacZ1: ATC TTC CTG AGG CCG ATA CT
rLacZ1: CGG ATT GAC CGT AAT GGG ATA G
fLacZ2: CCA ACG TGA CCT ATC CCA TTA C
rLacZ2: TTC CTG TAG CCA GCT TTC ATC
fLacZ3: GTT GGA GTG ACG GCA GTT AT
rLacZ3: GCT GAT TTG TGT AGT CGG TTT ATG
fLacZ4: GCC GAA ATC CCG AAT CTC TAT C
rLacZ4: AGC AGC AGC AGA CCA TTT
fLacZ5: CAT GTT GCC ACT CGC TTT AAT
rLacZ5: GAA ACT GTT ACC CGT AGG TAG TC

ALG9
Fw: CACGGATAGTGGCTTTGGTGAACAATTAC
Rv: TATGATTATCTGGCAGCAGGAAAGAACTTGGG
TAF10
Fw: ATATTCCAGGATCAGGTCTTCCGTAGC
Rv: GTAGTCTTCTCATTCTGTTGATGTTGTTGTTG
TFC1
Fw: GCTGGCACTCATATCTTATCGTTTCACAATGG
Rv: GAACCTGCTGTCAATACCGCCTGGAG

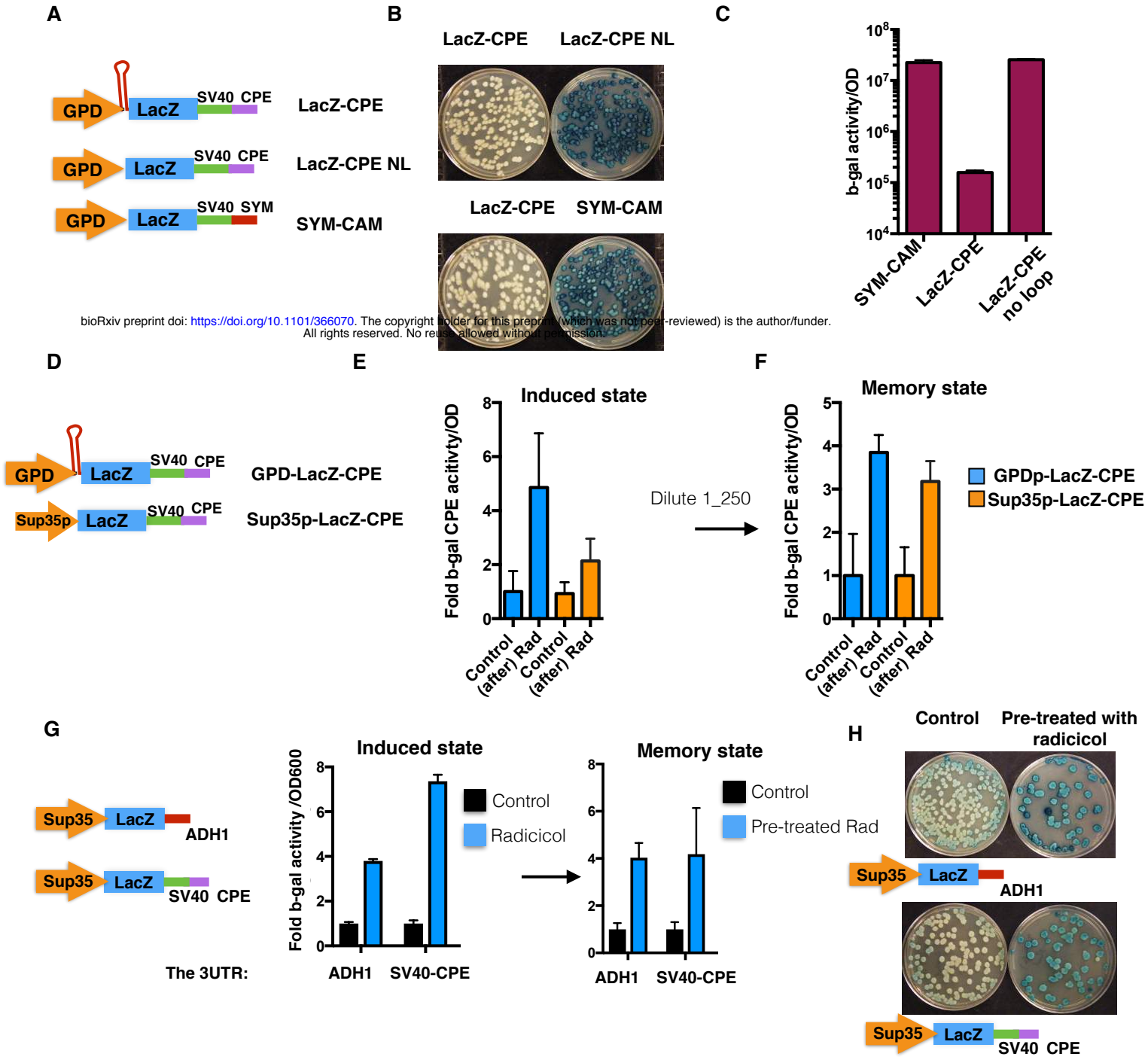




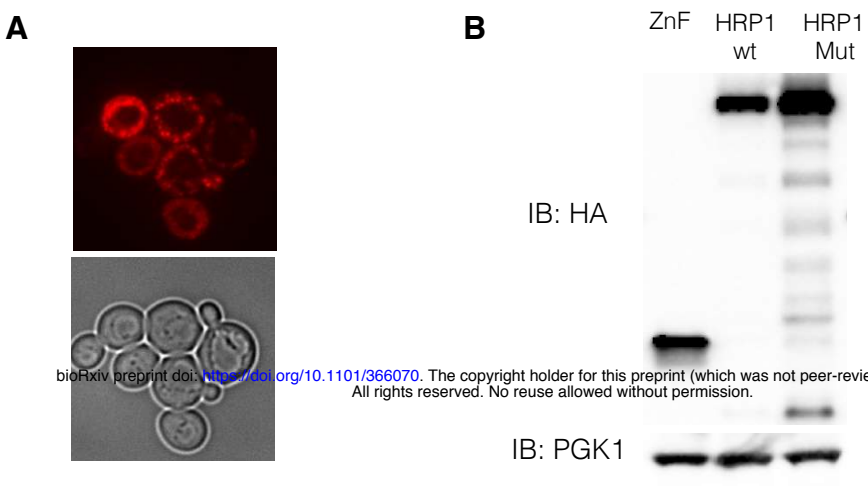
Sup figure 1:
(A) The indicated flag-tagged proteins were over expressed in 293T cells and their binding to the endogenous Hsp90 was assessed following immunoprecipitation with anti-flag antibody and immunoblotting with anti-HSP90.
(B) Schematic representation of proteins that were found to associate with CPEB3 when overexpressed in 293T cells. CPEB3-associating proteins are classified to the most prevalent categories (Chaperones, Mitochondrial, Structural or mRNA associated proteins). In red are the specific protein-protein interactions that were lost when the PrD of CPEB3 was deleted.



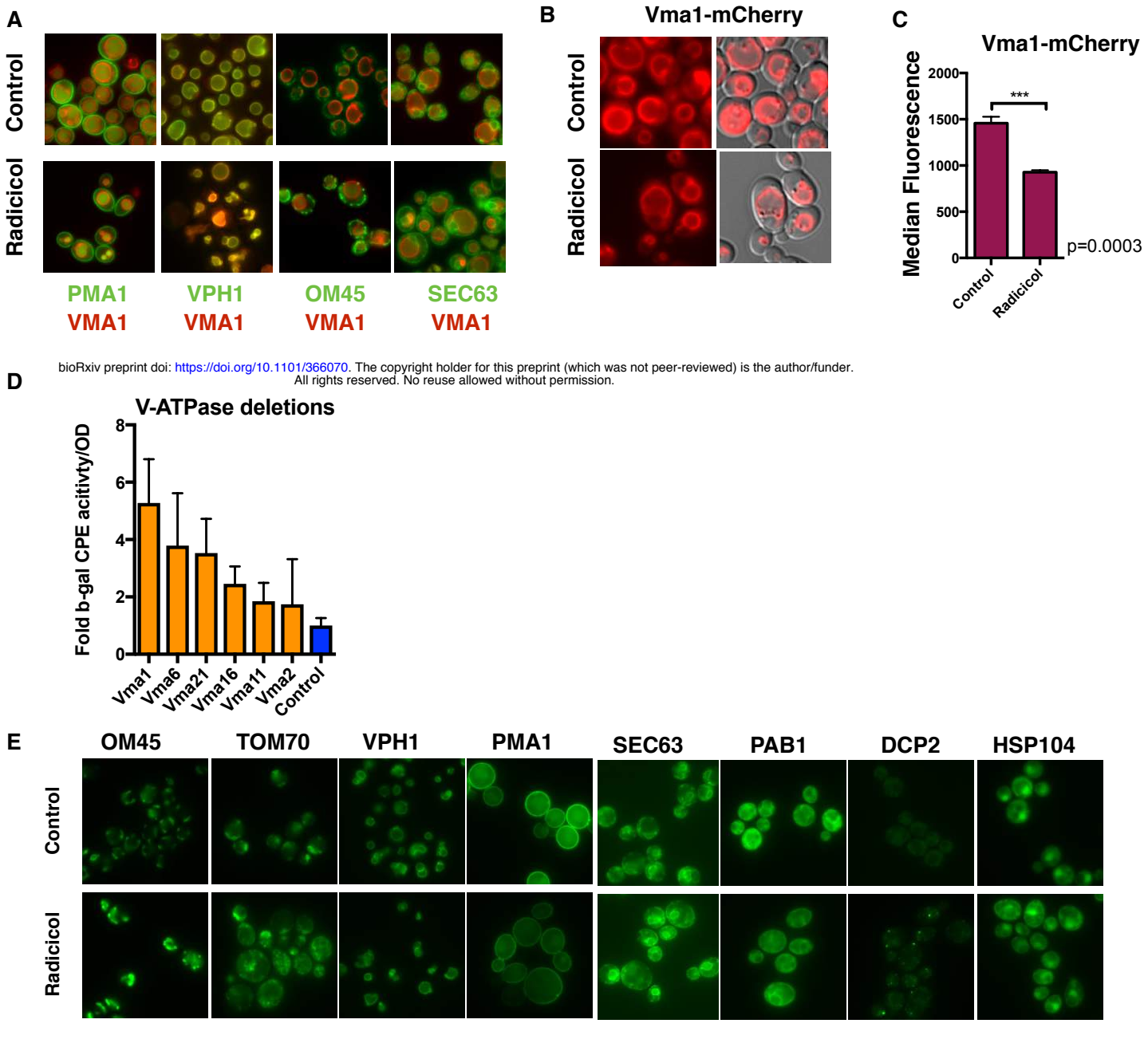
Sup figure 2.
(A) Streak out of white [CPEB2/3] and red [cpeb2/3] PrD-SUP35C colonies reveals that the epigenetic state of each strain is maintained over many generations. CPEB3 exhibited an intermediate state [mCPEB3], with altered pigmentation but lack of ability to grow on -Ade.
(B) Sup35C assay of the CPEB3 PrD. Shown is the ability of two distinct heritable conformations of the CPEB3PrD-Sup35C, white [CPEB3] and red [cpeb3], to grow on medium lacking adenine or on ¼ YPD.
(C) Cell lysates from 293T cells over expressing flag-tagged full length CPEB2 or the PrD-deleted CPEB2 were subjected to SDD-AGE analysis with and without boiling prior to loading. Proteins were detected by immunoblotting with anti-flag antibody.



Sup figure 3.
(A) Schematic of translation reporter constructs. LacZ-CPE is the original reporter used in previous studies (ref); LacZ-CPE NL is the same as LacZ-CPE but without a predicted stem loop in the 5' UTR; and LacZ SYM-CAM is the original control from previous studies (ref).
(B) Cells expressing the indicated translation reporters were grown in liquid culture, plated and then replica plated on x-gal plates to visualize the LacZ activity.
(C) The activity of b-gal in strains harboring the indicated translation reporters was measured and plotted after overnight growth.
(D) Schematic of the previously described translation reporter (ref) LacZ-CPE and the newly constructed translation reporter (Sup35p-LacZ-CPE) which lacks the stem loom at the 5' UTR and has a weaker promoter (Sup35p).
(E-F) comparison of radciciol activation (E, induced) and persistence (F, memory) (cells were grown over night and then diluted 1:250 and grown in the absence of radicicol) of translation of both the LacZ-CPE and the new Sup35p-LacZ-CPE reporter.
(G) Schematic representation of the translation reporters with alternative 3UTR sequences (left). The relative induction of b-gal activity by radicicol was measured in cells harboring the distinct translation reporters (Induced state) and the degree of persistence following dilution 1:500 and overnight growth in the absence of the inhibitor (Memory state).
(H) Cells harboring the indicated translation reporters were grown in the presence of 10uM radicicol ON and then plated on plates lacking radicicol. The plates were then replica plated onto x-gal plates and the relative b-gal activity was visualized.



Sup figure 4.
(A) Visualization of overexpression of the mkate tagged HRP1 PY521/2AA mutant after 6 hours of induction in galactose media.
(B) Western blot analysis of total protein levels in cells overexpressing the wild type HRP1, PY521/2AA HRP1 mutant and ZnF control yTRAP sensors.



Sup figure 5.
(A) Vma1 (tagged with mCherry) localization with respect to the vacuolar protein (VPH1), plasma membrane protein (PMA1), Mitochondrial protein (OM45) and ER-localized proteins (SEC63) all tagged with GFP was analyzed after overnight treatment with or without 10uM radicicol.
(B-C) VMA1 endogenously tagged with mCherry is visualized (B) and levels quantified (C) following overnight growth with or without 10uM radicicol. *** p=0.0003. unpaired t-test analysis was used.
(D) Strains with the indicated V-ATPase subunit deletions expressing the CPEB translation reporter were analyzed for their relative b-gal activity. (E) GFP-tagged OM45, TOM70, VPH1, PMA1, SEC63, PAB1, DCP2 and HSP104 were visualized following overnight treatment with or without 10uM radicicol. (F) LacZ activity in cells harboring the CPEB translation reporter grown overnight in normal (glucose) or iron-copper depleted media (-Fe-Cu) in the presence or absence of 10uM radicicol.

Timothy Jonas Eisen
teisen@wi.mit.edu | 17 Henry St. Medford, MA 02155 | (203) 671-7642

Education

Massachusetts Institute of Technology, Cambridge, MA — 2013–present
Ph. D. Biology

Brown University, Providence, RI — 2007–2011
Sc. B. Chemistry, *with honors*

Research

Laboratory of Dr. David Bartel, MIT, Whitehead Institute, HHMI — 2014–present
Graduate Student
Cytoplasmic mRNA Metabolism

Laboratory of Dr. Myriam Heiman, MIT, Broad Institute of MIT & Harvard — 2011–2013
Research Associate II
Reverse Engineering Medium Spiny Neurons

Laboratory of Dr. Gerwald Jogl, Brown Univeristy — 2010–2011
Undergraduate Research Assistant
Structural Studies of the Histone Methyltransferase SUVH5; The Interaction of LepA with the Ribosome

Laboratory of Dr. Emma Heart, Marine Biological Laboratories — 2010
NSF Graduate Research Fellow
Plasma Membrane Electron Transport in Pancreatic β cells

Teaching

MIT Lead Teaching Assistant, Undergraduate Biochemistry (7.05) — Spring 2017

MIT Teaching Assistant, Graduate Biochemistry (7.51) — Fall 2014

Brown University Teaching Assistant, Organic Chemistry (CHEM0350) — Spring 2009

Awards and Honors

Abraham J. Siegel Fellowship Award — 2019

Whitehead Institute Spirit Award for Communication and Collaboration — 2019

National Science Foundation, Graduate Research Fellowship — 2014–2019

Whitehead Institute Appreciation Awards — 2017, 2018

Brown University, Departmental Honors in Chemistry — 2011

National Science Foundation, Research Education for Undergraduates Fellowship — 2010

Presentations

Conference on Eukaryotic mRNA Processing, Cold Spring Harbor Laboratory, NY (abstract-selected speaker) — 2019

Conference on RNA Stability, Montréal, Canada (abstract-selected speaker) — 2019

Whitehead Institute Forum Seminars, Cambridge, MA (invited speaker) — 2018

Microsymposium on Small RNA Biology, Vienna, Austria (poster presentation) — 2018

Timothy Jonas Eisen
teisen@wi.mit.edu | 17 Henry St. Medford, MA 02155 | (203) 671-7642

Publications

Eisen, T.J.*, Eichhorn, S.W.*, Subtelny, A.O.*, Lin, K.S., McGeary, S.E., Gupta, S., and Bartel, D.P. (2020). The Dynamics of Cytoplasmic mRNA Metabolism. *Mol Cell* 77, 786-799 e710.

Eisen, T.J.*, Eichhorn, S.W.*, Subtelny, A.O.*, and Bartel, D.P. (2020). MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. *Mol Cell* 77, 775-785 e778.

Eisen, T.J., and Bartel, D. P. Poly(A)-Tail Length has a Significant but Modest Effect on Translational Efficiency in Neurons. (In preparation).

Tsvetkov, P., Brune, Z., **Eisen, T.J.**, Heinrich, S., Newby, G.A., Hallacli, E., Kayatekin, C., Pincus, D., and Lindquist, S. Transient Hsp90 suppression promotes a heritable change in protein translation. bioRxiv, 366070 (in revision).

Bruno, P.M., Liu, Y., Park, G.Y., Murai, J., Koch, C.E., **Eisen, T.J.**, Pritchard, J.R., Pommier, Y., Lippard, S.J., and Hemann, M.T. (2017). A subset of platinum-containing chemotherapeutic agents kills cells by inducing ribosome biogenesis stress. *Nat Med* 23, 461-471.

Kang, M.*, **Eisen, T.J.***, Eisen, E.A., Chakraborty, A.K., and Eisen, H.N. (2015). Affinity Inequality among Serum Antibodies That Originate in Lymphoid Germinal Centers. *PLoS One* 10, e0139222.

Gray, J.P., **Eisen, T.**, Cline, G.W., Smith, P.J., and Heart, E. (2011). Plasma membrane electron transport in pancreatic beta-cells is mediated in part by NQO1. *Am J Physiol Endocrinol Metab* 301, E113-121.