

GENOME-WIDE STUDIES ON POST-TRANSCRIPTIONAL REGULATION

by
Anna McGeachy

A dissertation submitted to Johns Hopkins University in conformity with the requirements
for the degree of Doctor of Philosophy

Baltimore, MD
April 2016

Abstract

Regulation of gene expression is essential for cellular survival and complex responses or phenotypes. Improvements in sequencing technology enable investigation of RNA transcription as a means to understand complex biological systems. However, there is increasing evidence that the correlation between RNA abundance and protein abundance is weak, as the range of associated protein abundances observed for transcripts with similar abundance spans two orders of magnitude. This discrepancy in abundances can be partially explained by changes in transcript stability or translational efficiency, collectively known as post-transcriptional regulation. Single gene studies have demonstrated that post-transcriptional regulation by trans- or cis-acting factors is essential for cell survival and differentiation. Post-transcriptional regulation can be driven by trans-acting RNA binding proteins (RBP) or cis-acting RNA-encoded elements. Genome-wide RNA association studies have identified an extensive number of RBPs; however, the regulatory roles of these RBPs are largely unclear. We provide a method for generating and functionally characterizing post-transcriptional regulators in genome-wide expression libraries. Briefly, we generate genome-wide expression libraries utilizing large-volume turbidostats. We then characterize the post-transcriptional regulation activity of each expression fragment through tethering to a reporter RNA encoding a fluorescent protein. Our method recovers known post-transcriptional regulators as well as novel regulators. We also performed a genome-wide survey for cis-acting RNA elements known as upstream open reading frames (uORFs). uORFs are thought to negatively regulate transcript translation by titrating ribosomes away from downstream coding sequence (CDS). Under translation-limited conditions, uORFs are thought to be bypassed in favor of the CDS. We identified several previously uncharacterized uORFs as important regulators of the cellular stress response. In addition, we demonstrate that

uORFs are occupied in stress, arguing against the current model for uORF regulation. Collectively, our data provides a genome-wide view of post-transcriptional mechanisms by both trans- and cis-acting factors.

Preface and Acknowledgements

The following dissertation is in large part the sole, unpublished work of the thesis author, Anna McGeachy, with the following exceptions:

- Chapter 4 includes a publication co-written with the thesis advisor, Nicholas Ingolia. The chapter was originally published online in The EMBO Journal on February 19, 2016. In line with The EMBO Journal and Wiley Publishing's permission policies, authors are allowed to reprint in whole or part the text in a new publication by the authors as long as appropriate citation is given ([http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1460-2075/homepage/Permissions.html](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1460-2075/homepage/Permissions.html)).
- Chapter 5 includes a publication performed in collaboration with Carmela Sidrauski and was co-written with Carmela Sidrauski, Nicholas Ingolia, and Peter Walter. The publication was originally published online in eLife on February 9, 2016. In line with eLife's open access policies, the article is governed by the Creative Commons license (<http://creativecommons.org/licenses/by/4.0/>). As such, there are no limitations on reprinting the work in whole or in part as long as proper attribution is given. For the purpose of this thesis, we (Anna McGeachy and Nicholas Ingolia, two of the four authors) have edited the manuscript to focus on the contribution of the thesis writer, Anna McGeachy. Additional content that was included in the article at the time of publication but was not performed in part or whole by Anna McGeachy will be added in the appendix under a section called "Results from Collaborators".

The Appendix focuses on the aforementioned "Results from Collaborators".

These were performed largely by Carmela Sidrauski with writing and statistical

consultation from the thesis writer, Anna McGeachy. They are included as a part of the publication in Chapter 5 originally, and they are reproduced in the Appendix as they are a significant component of the publication, but were not done in majority by the thesis writer, Anna McGeachy.

Table of Contents

| | |
|------------------------------|-----|
| Abstract | ii |
| Preface and Acknowledgements | iii |
| Table of Contents | v |
| List of Tables | ix |
| List of Figures | x |

Chapter 1

Introduction to post-transcriptional regulation and RNA binding proteins

| | |
|---|----|
| Preface | 1 |
| A brief history of RNA | 1 |
| History of RBPs in post-transcriptional regulation | 4 |
| Mechanistic examples of RBPs | 9 |
| Pumilio | 9 |
| Ash1 | 11 |
| Gw182 | 13 |
| Global studies of RBP classification and characterization | 16 |
| Computational prediction of RBP and RBP targets | 16 |
| High-throughput experiments defining RBPs and RBP targets | 19 |

Chapter 2

Selection and generation of diverse genome-wide expression libraries using large-scale turbidostats

| | |
|------------|----|
| Purpose | 24 |
| Motivation | 24 |

| | |
|--|----|
| Introduction | 24 |
| Results | 26 |
| Large-volume turbidostats | 26 |
| Generating a genome-wide expression library | 32 |
| Selection of a comprehensive genome-wide library | 35 |
| Discussion | 39 |

Chapter 3

Genome-wide survey for post-transcriptional regulators in *Saccharomyces cerevisiae*

| | |
|---|----|
| Purpose | 41 |
| Motivation | 41 |
| Introduction | 42 |
| Results | 43 |
| Developing a two-color, <i>in vivo</i> tether assay | 43 |
| Preparing a comprehensive library of potential regulators | 48 |
| Selecting for potential regulators | 49 |
| Examples of select identified regulators | 50 |
| Discussion | 55 |

Chapter 4

Upstream open reading frames (uORFs) as important post-transcriptional regulatory domains

| | |
|---|----|
| Introduction and preface | 59 |
| Publication | |
| Starting too soon: upstream reading frames repress downstream translation | 60 |

Chapter 5

Genome-wide interrogation of the Unfolded Protein Response (UPR) and the role of ISRIB

| | |
|---|----|
| Introduction and preface | 66 |
| Publication | |
| The small molecule ISRIB reverses the effects of eIF2 α phosphorylation on translation and stress granule assembly | 71 |
| Abstract | 72 |
| Introduction | 73 |
| Results | 76 |
| Ribosome profiling of ER stress in mammalian cells | 76 |
| ISRIB substantially reduced the translational effects elicited by stress and eIF2 α phosphorylation | 81 |
| Discussion | 83 |
| Translational regulation upon ISR induction | 84 |
| Materials and Methods | 88 |
| Chemicals | 88 |
| Cell culture | 88 |
| Isolation of ribosome footprints and RNA | 88 |
| Generation of sequencing libraries and data analysis | 89 |
| Acknowledgements | 90 |
| Supplement | 90 |

Appendix 1

| | |
|--|-----|
| Results from Collaborators | |
| Introduction and preface | 99 |
| Publication | |
| The small molecule ISRIB reverses the effects of eIF2 α phosphorylation on translation and stress granule assembly | 100 |
| Results | 101 |
| ISRIB prevents formation of stress granules exclusively triggered by eIF2 α phosphorylation | 101 |
| ISRIB triggers rapid disassembly of stress granules and restores translation | 103 |
| Discussion | 108 |
| Materials and Methods | 110 |
| Immunofluorescence | 110 |
| Live cell microscopy | 110 |
| Protein analysis | 111 |
| [³⁵ S]-methionine incorporation | 111 |
| Supplement | 111 |
| Bibliography | 115 |
| Acknowledgements | 150 |
| Curriculum Vitae | 151 |

List of Tables

Table 2-1: Selecting protein coding sequences requires bulk selection

List of Figures

Chapter 2

Figure 2-1: Detecting cell density in a culture

Figure 2-2: Turbidostat growth chambers

Figure 2-3: Growth measurements using IR detection

Figure 2-4: Size selecting genomic DNA fragments

Figure 2-5: Generation a genome-wide protein-expression library

Figure 2-6: Generations of growth required to select for protein-expression library

Figure 2-7: Genomic coverage and framing in pre- and post-recombination libraries

Chapter 3

Figure 3-1: RNA binding proteins are often unexpected from existing annotation

Figure 3-2: Dual fluorescent reporter tethering system for measuring post-transcriptional regulation in vivo

Figure 3-3: Testing positive and negative regulation in dual fluorescent tethering systems

Figure 3-4: Abundance of fragments in potential activator or repressor libraries

Figure 3-5: Two potential repressors from the post-transcriptional regulator screen

Figure 3-6: Two potential activators from the post-transcriptional regulator screen

Figure 3-7: Summary of the screen

Figure 3-8: Future directions

Chapter 4

Figure 4-1: Ribosome profiling detects uORF translation and downstream repression

Chapter 5

Figure 5-0: The unfolded protein response (UPR) drives adaptation through three pathways, one specific to translation

Figure 5-1: Translational regulation upon ER stress in mammalian cells

Figure supplement 5-1: Ribosome and mRNA densities in the 5' UTR of ATF4 and SLC35A4

Figure supplement 5-2: Translation regulation of mTOR targets upon ER-stress

Figure supplement 5-3: Correlation plots for duplicate ribosome profiling experiments

Figure supplement 5-4: Mean mRNA abundances of all genes mapped

Appendix: Results from Collaborators

Figure A1-2: ISRIB blocks stress granule formation induced by eIF2 α phosphorylation

Figure A1-3: ISRIB addition rapidly dissolves pre-formed stress granules in live cells restoring translation

Video A1-1: ISRIB triggers stress granule disassembly

Video A1-2: ISRIB does not trigger disassembly of P-bodies

Figure A1-3 supplement A1-1: ISRIB does response and inactive analog in stress granule assay

Figure A1-3 supplement A1-2: Representative SDS-PAGE gel of [³⁵S]-methionine pulse as described in Figure 3D

Chapter 1:

Introduction to post-transcriptional regulation and RNA binding proteins

Preface

Regulation of the subset of transcripts and proteins produced at a given time is essential for many processes, including stress response, development, and cellular homeostasis. The active unit of gene expression is typically the protein produced from translating a transcript; however many genome-wide studies use abundances determined from RNA sequencing as a proxy for protein expression. Several studies have demonstrated a weak correlation between RNA and protein abundances (Ingolia et al. 2009, Schwanhäusser et al. 2011, Li et al. 2014, Jovanovic et al. 2015, Cheng et al. 2015). One way in which this discrepancy can arise is through post-transcriptional regulation. Post-transcriptional regulation encompasses the processes regulating the lifespan, localization, and translational efficiency of an RNA molecule. One method by which post-transcriptional regulation can be exerted is through the activity of RNA binding proteins (RBPs). There are several well-studied examples that describe the RBPs as essential regulators of stress, development, and additional pathways. To understand the field of post-transcriptional regulation, we first describe the history and study of RNA. Following this, we describe the post-transcriptional regulation of single genes. Finally, we discuss the current work and methodologies for post-transcriptional regulation. Together, the introduction reviews the background and future directions for understanding post-transcriptional regulation.

1. A brief history of RNA

Although it is now clear that RNA and its regulation is biologically essential, this was not always appreciated. RNA was described as early as Miescher's 1871 studies on "nuclein" and biochemically separated as a non-DNA nucleic acid in 1893 as "yeast nucleic acid" (Darnell 2011). However, both DNA and RNA were under-studied for several decades in favor of proteins. While DNA was established as the genetic material responsible for conveying information between generations (Avery et al. 1944; Watson and Crick, 1953) after seminal studies in the 1940s and early 1950s, RNA studies remained relatively scarce.

The study of RNA in the 1940s and 1950s focused on cytoplasmic RNA and its role in the synthesis of protein. Jean Brachet proposed this connection in 1941 (Darnell 2011). In work that continued from 1943 and into the early 1950s, Albert Claude used the new technology of ultracentrifugation to separate RNA and protein containing "microsomes" from cellular lysate. He further characterized microsomes using electron microscopy (EM; Rheinberger 1995, Darnell 2011). In the mid-1950s, Paul Zamecnik's group specifically demonstrated that the combination of RNA and protein from microsomal fractions was capable of protein synthesis (Littlefield et al. 1955) in what we now recognize as the ribosome. Ribosomes are perhaps one of the earliest examples of functional ribonucleoprotein complexes (RNPs); however, the interplay between ribosomal RNAs (rRNAs) and ribosomal proteins is still being actively researched.

It was clear in the late 1950s that RNA was necessary for translation in the context of the ribosome, and advances in the 1960s demonstrated that the intermediate polymer we now know as messenger RNA (mRNA) acted as a template for ribosome-mediated translation. The Brenner-Jacob-Meselson experiments published in 1961 provided evidence that newly synthesized bacteriophage RNA associated with pre-existing

bacterial ribosomes to generate bacteriophage protein product (Brenner et al. 1961). These intermediate RNA products were shown to be transcribed from defined genomic loci using molecular hybridization (Hayashi et al. 1963). With these experiments and several others, it became clear that RNA served as an intermediate messenger for conversion of DNA-encoded information to protein (reviewed briefly in Penman et al. 1963 and extensively in Darnell 2011).

The hybridization of messenger RNA (mRNA) to DNA provided additional information about the life cycle and regulation of mRNAs. At this point, rRNA was known to be processed from a larger precursor into a smaller, functional sequences (Scherrer et al. 1963). A similar size discrepancy was seen in cytoplasmic versus nuclear non-ribosomal RNA (Tonegawa et al 1973). Molecular hybridization of adenovirus mRNA to DNA resulted in several loops of unmatched DNA, suggesting that RNA processing occurred in a manner that connected non-contiguous transcribed sequences together in mature cytoplasmic RNA molecules by the process now known as splicing (Berget et al. 1977; Chow et al. 1977) . It also provides one of the first examples of RNPs regulating RNA biology. The studies of RNPs in splicing began with a clinical observation of antigen-reactive RNP species from the sera of systemic lupus erythematosus (SLE) patients in 1971 by Mattioli and Reichlin. These were further characterized by Lerner and Steitz in 1979 who found that this population consisted of multiple RNP species. Lerner et al. and Rogers and Wall hypothesized a link between these RNPs and splicing in 1980. Direct evidence for their necessity in pre-mRNA processing was provided in 1983 by Padgett et al. and throughout the following decade (for examples, see Frendeway and Keller, 1985; Bringmann and Lührmann, 1986; review in Dreyfuss et al. 1988). These early studies of splicing provide some of the first insights into protein-containing complexes as important players in co-transcriptional regulation of RNA fates. Subsequent advancements have

characterized the steps and proteins required for mRNA biogenesis: transcription start site (TSS) selection, alternative splicing, polyadenylation signal sequence selection, and more. In the years following, studies began to illuminate that the complexities of mRNA regulation extend from nuclear co-transcriptional events into cytoplasmic post-transcriptional regulation.

2. History of RBPs in post-transcriptional regulation

One of the first studies of post-transcriptional regulation focused on the essential process of localization and incorporation of membrane-associated proteins. Blobel and Dobberstein (1975) observed targeting of the transcript encoding murine immunoglobulin to the ER on actively translating ribosomes. They hypothesized that mRNA transcripts for membrane-targeted proteins contained a cis-acting feature in the emerging peptide that could drive ER localization; however, recognition and localization of the actively translating transcript required the action of an adaptor. Walter and Blobel (1982) identified a particle responsible for targeting composed of 6 previously identified proteins and a 7S RNA; they termed this RNP the signal recognition particle (SRP). The SRP identifies a series of hydrophobic residues in the emerging nascent peptide chain (NC; reviewed in Walter and Lingappa 1986; Akopian et al. 2013). Upon signal peptide recognition, the SRP interacts with the ribosome and mRNA to pause translation (Walter et al. 1981). The transcript and ribosome NC (RNC) are directed by SRP-driven interactions to the ER SRP receptor; here, the ribosome resumes translation concurrent with translocation of the peptide into the membrane, and the SRP is released and recycled for subsequent targeting events (reviewed in Walter and Lingappa 1986, Akopian et al. 2013). Thus, the SRP can be thought of as an RBP complex that recognizes multiple mRNA targets to regulate translation localization.

Concurrent with studies of SRP, regulated mRNA localization and translation were recognized to be important for *Drosophila melanogaster* embryogenesis. The early *Drosophila* embryo is a syncytium for the first 13 rounds of nuclear division; during this time, the zygotic genome is largely transcriptionally silent (Zalokar, 1976). Early transplantation assays established that embryonic patterning decisions are established prior to cellularization and activation of zygotic nuclei (Kauffman 1980; discussed in Mahowald and Hardy, 1985). As such, many of the early patterning decisions are dependent on regulation of maternally deposited mRNA (for examples, see Bull 1966, Lehmann and Nüsslein-Volhard 1986, Schupbach and Wieschaus 1986). Direct connection of phenotypes to regulated mRNA localization and translation remained elusive until the application of *in situ* hybridization for mRNA (Gall and Pardue, 1969; Hafen et al. 1983) and antibody-affinity protein identification (Coons et al. 1941; Klämbt and Schmidt, 1986).

Regulation of the *bicoid* mRNA serves as a historical paradigm for understanding spatial patterning in the *Drosophila* embryo and in cellular biology in general. *Bicoid* is essential for establishing the anterior pole; mutants of *bicoid* lack anterior head and thorax structures (Frohnhofer and Nüsslein-Volhard, 1986). Transplantation of wild-type *bicoid* cytoplasm into *bicoid* mutants is capable of rescuing abdominal structures in a location and dose dependent manner, suggesting a concentration-dependent gradient emanating from an anchored site (reviewed in Nüsslein-Volhard et al. 1987). Use of *in situ* hybridization showed establishment of localized *bicoid* mRNA at the anterior pole of the oocyte during oogenesis and nurse cell contribution (Berleth et al 1988; Johnston et al. 1989). In contrast, Bicoid protein is not detectable by antibodies until after fertilization; moreover, the protein extends much further down the anterior-posterior (A-P) axis than

the corresponding transcript (Dreiver and Nüsslein-Volhard, 1988). The temporal and spatial control of *bicoid* mRNA localization and translation requires the interplay of a number of proteins. Exupurantia and Exu-like bridge the 3' untranslated region (UTR) of *bicoid* for transport along microtubules into the anterior pole from nurse cells (MacDonald et al. 1995; Cha et al. 2001). Anchoring at the pole is provided by a number of proteins, including Staufén, ESCRT-II, and others (reviewed in Lasko 2012). Some yet unknown protein drives poly(A) lengthening and translational activation upon fertilization (Sallés et al. 1994) with stabilization of the transcript potentially driven by Pumilio (Gamberi et al. 2002). Transcript stability is also tightly regulated in the *Drosophila* embryo (reviewed in Temme et al 2014) and a key point for RBP-mediated post-transcriptional regulation across eukaryotes.

Transcript stability is globally regulated through RBP-mediated post-transcriptional action. In 1963 Penman et al. established that the abundance of eukaryotic mRNAs decreases after observing a decline in cytoplasmic mRNA signal in the absence of new transcription. Pulse-chase experiments using radioactive nucleotides during the 1970s and 1980s demonstrated that mRNAs exhibit a range of decay rates (reviewed briefly in Darnell 2011 and extensively in Belasco and Brawerman 1993). These experiments revealed two important molecular details about the regulation of transcript stability. First, poly(A) tail length changes over time (Gorski et al. 1975) and transcript stability correlates with poly(A) length (Nudel et al 1976). Second, poly(A) binding protein (PABP) affects the stability of transcripts by protecting the poly(A) tail from nuclease attack (reviewed in Bernstein and Ross 1989, Belasco and Brawerman 1993). Although these early studies focused predominantly on *globin* mRNA, we know now that regulating stability is important for a variety of biological functions including but not limited to responding to stress (reviewed in Khabar 2014, Schoenburg and Maquat 2012),

inhibiting host translation during viral infection (Glaunsinger and Ganem 2006), and differentiation of diverse tissue types (reviewed in Belasco and Brawerman 1993, Temme et al. 2014).

As a part of these varied regulation pathways, disruption of the Pabp-poly(A) tail interaction shortens the poly(A) tail and initiates decapping and exonucleolytic decay (reviewed in Belasco and Brawerman 1993, Beelman and Parker 1995). One of the RBPs recruited as a result of poly(A) shortening and Pabp disruption is Ccr4-Not, a multi-protein complex conserved in eukaryotes (reviewed in Collart and Timmers 2004, Coller and Parker 2004, Doidge et al. 2012, Temme et al. 2014). When recruited to a transcript with a pre-shortened poly(A) tail, the complex acts at both the 5' and 3' end of the transcript to promote decay. First, Ccr4-Not promotes 3' to 5' exonucleolytic digestion by completing poly(A) tail removal through its Ccr4 deadenylases (reviewed in Shirai et al. 2014, Collart and Timmers 2004). Second, 5' to 3' decay is driven by Ccr4-Not scaffolding the recruitment of decapping factors and subsequent decay machinery (reviewed in Collart and Timmers 2004, Shirai et al. 2014). In knockout studies, Ccr4-Not and its subunits have been shown to be important in regulating both general and specific subsets of transcripts accounting for approximately 85% of the yeast genome (Azzouz et al. 2009).

One specific example of Ccr4-Not recruitment and decay is as a component in the micro-RNA (miRNA) driven degradation process (Fabian et al. 2011). Post-transcriptional regulation by small RNAs (smRNA) was first described in 1993 and has been an active area of study since. The Ruvkun (Wightman et al. 1993) and Ambros (Lee et al. 1993) labs simultaneously described repression of the *lin-14* transcript dependent on RNA-RNA interaction with the small RNA *lin-4* (reviewed in Darnell 2011;

He and Hannon 2004). Double stranded smRNA homology-dependent regulation was established as a genome-wide and evolutionary conserved mode of regulation through two papers from the Ruvkun lab in 2000 (Pasquinelli et al. 2000; Reinhart et al. 2000). The mechanism and associated RBPs for endogenous dsRNA processing and target repression were discovered in the years immediately following (reviewed in He and Hannon 2004). Briefly, cells transcribe long endogenous imperfect RNA hairpins, primary microRNAs, that undergo a number of nuclear processing steps to produce single-stranded micro-RNAs approximately 22 nucleotides in length (reviewed in Ha and Kim, 2014). As a part of the miRNA production process, the miRNA is loaded into a member of the Argonaute (Ago) family of RBPs. Ago proteins cooperate with additional factors (e.g., Gw182) to form what is called the RNA-induced silencing complex (RISC; Eulalio et al. 2008). The RISC is localized to mRNA transcripts bearing at least partial complementarity to the loaded miRNA. Once bound to a target mRNA, the RISC proteins can recruit additional proteins such as Ccr4-Not (Fabian et al. 2011; review of associated proteins in Triteschler et al. 2010, Ho and Mardsen 2014) to first drive translational repression (Bazzini et al. 2012, Djuranovic et al. 2012) followed by transcript degradation (Guo et al. 2010). Importantly, because miRNAs can target many mRNAs concurrently, understanding the full scope of miRNA-mediated regulation requires moving from single gene studies into genome-wide experiments (highlighted through the genome-wide in silico predictions as early as Pasquinelli et al. 2000, reviewed in He and Hannon 2004, and demonstrated in Giraldez et al. 2006).

Much of what we understand about post-transcriptional regulation comes from these specific historical examples: mRNA localization (SRP and *Drosophila* embryogenesis), translational repression (SRP, *Drosophila* embryogenesis, and miRNA), and regulation of transcript stability (*Drosophila* embryogenesis and miRNA). These early examples

provided the paradigms by which we define subsequent RBPs. We ask mechanistic questions about what transcript the protein binds, how specificity is achieved, and what role the RBP plays in the regulation of transcript localization, translation, and stability.

3. Mechanistic examples of RBPs

The mechanism of RBP binding, specificity, and target regulation has been well defined for a number of RBPs. In this section, we focus on a few examples that define characteristics we can use to develop a framework to understand new RBPs and their role in biology.

Pumilio

As described in the previous section, restricting the expression of mRNAs in the early *Drosophila* syncytium is essential for embryogenesis. The anterior organizer, Bicoid, is described above; the posterior organizer is Nanos (Lehmann and Nüsslein-Volhard 1986, Lehmann and Nüsslein-Volhard 1991). Along the anterior-posterior axis, the localization of *nanos* is the inverse of *bicoid*: posterior localization of mRNA and a posterior-to-anterior gradient of protein (Wang et al. 1994). After the establishment of the initial anterior-posterior gradients, there is a second stage of pattern formation dependent on several genes, including the RBP Pumilio (Lehmann and Nüsslein-Volhard 1987b; reviewed in Nüsslein-Volhard et al. 1987). Mutants for *pumilio* are missing abdominal segments but maintain a normal posterior; moreover, the *pumilio* product was not capable of generating abdominal segments, suggesting that it mitigates its regulation via other signaling molecules (Lehmann and Nüsslein-Volhard 1987b). In subsequent studies, Pumilio has been implicated in varying protein complexes for

diverse responses of many different mRNAs (Parisi and Lin 1999, Wickens et al. 2002, Gerber et al. 2006).

The effect of Pumilio is perhaps best studied through its interactions with another of the *Drosophila* patterning genes, *hunchback*. Similar to many other *Drosophila* morphogens, *hunchback* was first identified by the morphological and segment abnormalities caused by its absence; specifically, Hunchback is important for formation of the abdomen as mutants lack gnathal structures and thoracic bands (Lehmann and Nüsslein-Volhard 1986). Hunchback protein is localized in a gradient from anterior to posterior similar to *bcd*; however, in contrast to *bcd*, *hunchback* mRNA is evenly distributed across the early embryo (Tautz 1988, Tautz and Pfeifle 1989). In the absence of Nanos, the Hunchback gradient extends further into the posterior (Tautz 1988; Wharton and Struhl 1991).

Pumilio acts as a third factor in the regulation of *hunchback* translation, as demonstrated in loss of localized Hunchback expression in *pumilio* mutants (Lehmann and Nüsslein-Volhard 1987a, Lehmann and Nüsslein-Volhard 1991, reviewed in MacDonald 1992, reviewed in Barker et al. 1992).

Pumilio regulation of *hunchback* mRNA is dependent on the localization of Pumilio to the transcript. However, early studies of the Pumilio protein sequence did not identify any known RBDs (Macdonald 1992; Murata and Wharton 1995). It was shown, however, that the *hunchback* mRNA has two conserved sequences in the 3' UTR. These sequences were originally named Nanos responsive elements (NREs) as they were identified as essential for Nanos-mediated repression of *hunchback* mRNA in the embryonic posterior (Wharton and Struhl 1991). In line with the recognition that the loss of Hunchback localization seen in Nanos was dependent on Pumilio (Barker et al. 1992), it was shown that NREs were actually bound by Pumilio rather than Nanos (Murata and Wharton,

1995). With binding sites for Pumilio identified in *hunchback* mRNA, it was possible to define the protein region of Pumilio necessary for recruitment (Zamore et al. 1997). Importantly, recognition of this RBD founded an entire family of RBPs with similar RBDs (Zamore et al. 1997). Pumilio bound to *hunchback* mRNA then recruits Nanos (Murata and Wharton 1995). The Nanos- and Pumilio-bound hb mRNA is deadenylated (Wreden et al 1997) and separately translationally repressed (Chagnovich and Lehmann, 2001; Sonoda and Wharton, 2001). Translational repression is the result of *brain tumor* protein (Brat) binding to the *hunchback* mRNA-Pumilio-Nanos complex (Sonoda and Wharton, 2001). Brat recruits the cap-binding competitor d4EHP, which in turn represses translation by preventing binding of the canonical ribosome recruitment factor eukaryotic initiation factor 4E (eIF4E; Cho et al. 2006). While these mechanisms were uncovered by the study of *hunchback* mRNA regulation, additional studies have identified many other Pumilio targets in the *Drosophila* transcriptome (Gerber et al. 2006). This gives Pumilio two roles across the genome: first, it offers specificity for transcript selection, and second, it acts as a scaffold for protein-protein (Brat-d4EHP) interactions that can enact post-transcriptional regulation.

Pumilio is the establishing member of a much larger family of PUF (Pumilio and F box) proteins that are characterized by a regular eight helix repeat RNA binding domain (Zamore et al. 1997; reviewed in Filipovska et al 2001). PUF proteins are conserved across eukaryotes and participate in a wide range of post-transcriptional regulatory pathways, including glucose stress-response through Puf3 in yeast (Lee and Tu, 2015), anti-viral defense in mammalian cells (reviewed in Schwerk et al, 2015), and more (see Quenault et al. 2010 for additional review). Moreover, Pumilio and the many other members of the PUF family exemplify two characteristics of RBP-mediated post-transcriptional control: first, driving localized mRNA translation or repression, and

second, separating the activities of transcript specificity and post-transcriptional control into separate proteins or domains.

Ash1

Although the unicellular budding yeast *Saccharomyces cerevisiae* is far smaller than the syncytial *Drosophila* embryo, localized translational control is essential for the yeast life cycle. Haploid wild-type *S. cerevisiae* can switch mating type (from a to α and vice-versa) through HO-mediated (Homothallic switching endonuclease) rearrangement of the mating type locus (reviewed in Herskowitz, 1988). However, this mating-type switch only occurs in cells that have budded at least once (mother cells), but not in newly generated daughter cells. Screens for cells in which the daughter cells were capable of switching mating type prior to their first cellular division recovered the gene *ash1* (asymmetric synthesis of HO; Sil and Herskowitz 1996; Bobola et al. 1996). Its protein product, Ash1, is located predominantly in the nucleus of the daughter cell where it represses transcription of HO (Sil et al. 1996; Bobola et al. 1996). She1-5 (Swi5-dependent HO expression) proteins are important for the asymmetric distribution of Ash1p (Jansen et al. 1996).

As discussed previously about Bicoid in *Drosophila*, asymmetric protein distribution can result from regulated mRNA localization. Several labs simultaneously reported that *ash1* mRNA is asymmetrically distributed to the distal tip of the developing daughter cell (Long et al. 1997; Takizawa et al 1997). Moreover, this localization is dependent on the *ash1* mRNA 3' UTR and She1–5 proteins (Long et al. 1997; Takizawa et al 1997; Bertrand et al. 1998). Similar to Pumilio regulation of *hunchback* mRNA 3' NREs, specificity for *ash1* mRNA is established through an RBP interacting with the CDS and structural elements

in the 3' UTR (Gonzalez et al. 1999; Chartrand et al. 1999). The *ash1* mRNA-specifying RBP is She2 (Long et al. 2000; Böhl et al. 2000) which interacts with several proteins. In the nucleus, She2 cooperatively loads the RBPs Loc1 (Long et al. 2001; Shahbadian et al. 2014) and Puf6 (Gu et al. 2004; Shahbadian et al. 2014). In the cytoplasm, She2 interacts with proteins required for transport (She3 and She1/Myo4) and membrane tethering of the *ash1* transcript (She5p/Bni1; Takizawa and Vale, 2000; Long et al. 2000; Böhl et al. 2000).

The asymmetric distribution of Ash1 depends on the localization of *ash1* mRNA (Chartrand et al 2002; Irie et al. 2002). To restrict the region of Ash1 expression, *ash1* mRNA remains translationally repressed until it is delivered to the distal tip of the daughter cell (Chartrand et al 2002; Irie et al. 2002). Translational repression is mediated through at least two proteins -- Khd1 (Irie et al. 2002) and Puf6 (Gu et al. 2004). Puf6 represses translation through interaction with eIF5B (eukaryotic initiation factor 5B) to prevent 60S subunit recruitment (Deng et al. 2008). Khd1 binds eIF4G1 (eukaryotic initiation factor 4G1) and drives translational repression through an unknown mechanism (Paquin et al. 2007). Both Puf6 and Khd1 are phosphorylated by bud tip-localized kinases. This modification decreases their affinity for RNA (Deng et al. 2008; Paquin et al. 2007). Once Khd1 and Puf6 are released from the *ash1* transcript, translational repression is abrogated, and translation of the *ash1* mRNA can occur at the distal bud tip.

The study of *ash1* mRNA localization and translational repression provides an important case study for modes of RNA binding of proteins (She2, Puf6, and Khd1) in both the 3' UTR and the body of the gene. In addition, the complex network of protein-protein interactions (She2-Myo4, She2-She3, Puf6-eIF5B, Khd1-eIF4G1) highlights the

substantial interplay between multiple pathways to drive regulation of developmentally important genes.

Gw182

In addition to transcript specificity being provided by a protein, transcripts can be targeted for post-transcriptional regulation through Watson-Crick base-pairing with small RNA (smRNA) molecules. One of the extensively studied smRNA pathways is microRNA (miRNA) mediated regulation. These miRNAs are 20-22 nucleotide RNAs that are derived from longer, endogenous hairpins (He and Hannon 2004; Kim 2005, Winter et al. 2009; Ha and Kim 2014) and delivered to the Argonaute (Ago) protein and its associated RNA induced silencing complex (RISC). The RISC complex is directed to specific targets through a combination of perfect pairing (seed sequence, 6–8 nucleotides) and imperfect pairing between the target transcript and loaded miRNA (Schirle et al. 2014). Regulation by miRNAs has been implicated in a number of developmental pathways (*miR-430* in zebrafish, Giraldez et al. 2006; *lin-4* in *C. elegans*, reviewed in He and Hannon 2004) as well as in the establishment of human disease (reviewed in Li and Kowdley 2012). While their wide-ranging consequences are a result of miRNA complementary base-pairing, the mechanism for localization to and repression of the target transcript is remarkably well conserved.

A central component of RISC is the Gw182 protein. Like the early discoveries on snRNPs, Gw182 was originally discovered through sera antibodies of patients with autoimmune disease (Eystathiou et al. 2002). Gw182 has a functional RNA binding domain responsible for binding a large number of mRNA molecules (Eystathiou et al. 2002). Importantly, Gw182 and its associated mRNAs often occurred in cytoplasmic

punctae that did not colocalize with any known organelle markers, earning these structures the name GW bodies (GWB, Eystathioy et al. 2002). Proteins associated with mRNA decay were identified in the GWB, consistent with the known composition of mRNA degradation punctae, processing bodies (P-bodies), in yeast (reviewed in Stinton et al. 2004). Gw182 was identified as a part of the RISC through Argonaute 2 (Ago2) colocalization studies (Sen and Blau 2005) as well as being required for miRNA-mediated decay (Rehwinkel et al. 2005). Studies in 2006 (Behm-Ansmant et al.) and 2008 (Eulalio et al.) established Gw182 as an essential component for proper miRNA-RISC and Ago function. The inclusion of Gw182 in RISC for miRNA silencing opens questions about how Gw182 directs post-transcriptional regulation. Given the role of RBDs in directing post-transcriptional regulation (e.g. Pumilio and She2p), we can start to answer these questions by approaching the structural characteristics of Gw182 protein.

Gw182 has several characterized domains, including the eponymous GW repeat-rich domains at the N-terminus and silencing domain (Eulalio et al. 2009b). As mentioned previously, Gw182 interacts with Ago as well as components of the mRNA destabilizing machinery (Rehwinkel et al. 2005; Behm-Ansmant et al. 2006; Eulalio et al. 2008). Behm-Ansmant et al. (2006) used tethering assays to establish that the N-terminal GW domains are necessary for interacting with Ago for productive miRNA-mediated repression (also reviewed and named Ago-binding domain in Eulalio et al. 2009b). Briefly, tethering assays utilize small RNA hairpins usually placed in the 3' UTR of a reporter gene. Small viral peptides display high-affinity interactions for these hairpins and can be used to drive artificial recruitment of a protein domain of interest to assay its effect on the reporter gene (reviewed in Baron-Benhamou et al. 2004; Collier and Wickens 2007; Keryer-Bibens et al. 2012). Tethering assays also functionally

characterized the RRM (Eulalio et al. 2009a) and DUF (Lazzaretti et al. 2009; Zipprich et al. 2009) domains for their interaction with various mRNA decay complexes (reviewed and named the silencing domain in Eulalio et al. 2009b). Specifically, this silencing domain directly interacts with Pabp to compete for cap-binding and drive decapping (Zekri et al. 2009, reviewed in Triteschler et al. 2010) as well as directly recruiting Ccr4-Not to promote deadenylation and downstream degradation (Chekulaeva et al. 2011; Fabian et al. 2011). Through the Ago-binding domain, Gw182 is capable of assisting in the RNA-RNA mediated recruitment of miRNA-loaded Ago. Through the silencing domain, Gw182 acts to drive both translational repression and transcript degradation. The activities of these domains are separable, as shown by tethering assays. The ability to separate domains and retain their function is a key feature in many post-transcriptional regulators, and one that allows us to study individual protein domains for their post-transcriptional activity. Moreover, Gw182 demonstrates the paradigm of one or more RBPs acting as RNA binders with activation or repression provided by the recruitment of additional protein factors.

Pumilio, *ash1* mRNA, and Gw182 are three well-studied examples of RBP-mediated regulation. They provide paradigms for understanding how recruitment of co-regulators can regulate transcript lifespan as well as how the activities of transcript binding and regulator recruitment can be separated and studied individually. These single gene studies have provided invaluable information for our understanding of the involvement of RBPs in post-transcriptional regulation.

4. Global studies of RBP classification and characterization

Advances in technology, including microarrays, high-throughput sequencing, and improvements in mass spectrometry, have made it possible to extend single-gene studies to genome-wide approaches. Systematic approaches to understanding RBPs aim to answer three questions: what proteins bind RNA, what RNAs do these proteins bind, and what role do these RBPs play in RNA regulation?

Computational prediction of RBP and RBP targets

Many of the computerized databases of RBPs have their basis in literature reviews. Cook et al. (2010) used the information from a literature review to establish the RNA Binding Protein Database (RBPDB). Other internet-accessible collections of aggregated data include sources like the Gene Ontology Consortium (Gene Ontology Consortium 2015) and organism-specific sites like the *Saccharomyces* Genome Database (SGD; Cherry et al. 2011). Tsvetanova et al. (2010) and Gertsberger et al. (2014) have created manually curated lists of RBPs from genome databases for *S. cerevisiae* and mammalian cells respectively. While these methods are not inherently high-throughput, they are often used as benchmarks and guidance for RBP screens.

Several studies have aimed to expand on predicted RBPs and target mRNAs using algorithmic predictions on sequence and structural information (reviewed in Si et al. 2015). As discussed with Pumilio, there are known families of RBPs, typically defined by similarities in protein sequence and structure. Some examples of this include the aforementioned PUF family, proteins bearing K-homology (KH) domains, and the small RNA associated Piwi-Argonaut-Zwille (PAZ) domain proteins, and many others (reviewed in Lunde et al. 2007; Glisovic et al. 2008; Cook et al. 2010; Cléry and Allain 2015). By observing conserved residues and motifs from some members of a family, it is

possible to computationally predict additional family members using genomic or proteomic sequence. Han et al. (2004) use the primary sequences of known RBPs to train support vector machines (SVM) to identify RBPs as well as specific domains (e.g. KH domain). SVMs are one form of machine learning (ML); additional ML approaches applied to RBP prediction include Bayesian networks (Choi and Han 2011) or random forests (Ma et al. 2015). In general, each iteration of ML algorithm works to incorporate additional metrics to predict RBPs based on protein sequence, protein structure, and evolutionary conservation (see Shao et al. 2009, Kumar et al. 2011, Ma et al. 2015 for additional examples; review provided by Si et al. 2015). While these studies extended machine learning to the study of RNA-protein interactions, they are inherently limited by existing datasets that define an RNA binding protein. Candidate sequences identified as false positives (predicted RBP not annotated as an RBP) may in fact show RNA binding activity, but we lack data in support of this function. A historical example comes from Pumilio, which was originally not predicted to have RNA binding activity based on what was known about RNA binding protein motifs at the time (MacDonald 1992); however, it is now the founder of an expansive family of RBPs (Filipovska et al 2001). Computational studies are limited because they require information from previous studies to set expectations and predictions. This limitation is especially important given that many of the proteins identified in RNA-interaction studies lack known RBDs (Tsvetanova et al. 2010; Beckmann et al. 2015).

Similar to the work done for predicting RBPs, several computational models exist to predict the targets and binding specificity of RBPs. Sequence-level binding preferences can be predicted using an algorithm originally designed for DNA and protein motif identification: Multiple EM (Expectation Maximization) for Motif Elicitation (MEME) (Bailey and Elkan 1995; Bailey et al. 2006). Briefly, larger sequences containing the

RBPs smaller preferred binding site are experimentally derived through cross linking immunoprecipitation (CLIP, Ule et al. 2003) or systematic evolution of ligands by exponential enrichment (SELEX, Ellington and Szostak 1990; Tuerk and Gold 1990). The MEME algorithm derives matching subsequences or motifs from a collection of sequences in a mathematically rigorous way without prior knowledge about the expected subsequence. However, in contrast to DNA, RNA has more opportunity and flexibility to develop secondary structure. Because of this, RNA specificity can occur through structure-level information as well as sequence-level specificity. MEME in RNAs Including secondary Structure (MEMERIS, Hiller et al. 2006) incorporates the EM on maximum likelihood (ML) of MEME to include an additional layer of calculation that describes the probability of a nucleotide to be base paired or unpaired (PU, Hiller et al. 2006). Hiller et al. (2006) can identify motifs in both simulation and biological datasets. Several improvements have been made on MEME and MEMERIS in recent years (reviewed by Li et al. 2013). Two examples are RNAContext which incorporates RNA looping information (Kazan et al. 2010) and GraphProt which preserves complex, multi-nucleotide secondary structures using graph-based encoding (Maticzka et al. 2014). As with identifying RBPs, there are web-based interfaces for some of the existing computational methods such as RBPmotif (Kazan and Morris 2013). These modeling softwares are limited by built in assumptions; for example, RNA folding predictions may not match RNA structure in a biological context. Furthermore, the models are limited by the experimental biases of the input datasets: SELEX has the potential to capture high-affinity but low biological-relevance sequences; CLIP can also capture transient, off-target interactions. As such, any computationally defined RBP binding sequence should be biochemically verified using traditional methods or the recent high-throughput flow-cell methods from Buenrostro et al. (2014).

High-throughput experiments defining RBPs and RBP targets

The previous sections discussed the approaches for studying RBPs including developmental phenotypes, single gene IP, pulse-chase assays, and genetic knockouts. The information provided from these studies is invaluable, but it focuses on single actors in what is a complex system of interactions. The advent of microarrays, high throughput sequencing, and improvements in mass spectrometry make it possible to start answering questions genome-wide.

A direct approach to defining RBPs is to see what RNAs and proteins are physically interacting in the cell. In one of the earliest studies, Tsvetanova et al. (2010) used two complementary methods in *Saccharomyces cerevisiae*: 1. mass spectrometry (MS) on proteins that copurified with oligo-dT selected mRNA and 2. total RNA binding to protein microarrays. Using both techniques, they recovered numerous proteins (68 by IP, 12 by microarray) that were not previously characterized to bind RNA (Tsvetanova et al. 2010). Moreover, the majority of these novel RBPs lacked characterized RBDs or known roles in RNA biology; two biological groups recovered in their screen were secretory pathway proteins and metabolic enzymes (Tsvetanova et al. 2010). The method was adapted to use photoactivatable-ribonucleoside crosslinking (PAR-CL) in two contemporaneous papers in mammalian cell lines (HeLa, Castello et al. 2012; HEK, Baltz et al. 2012). Both Castello et al. (2012) and Baltz et al. (2012) identified around 800 RBPs. As in the Tsvetanova study, many of these RBPs were not annotated to bind RNA previously, did not have computationally predicted RNA binding domains, and included non-RNA associated proteins such as metabolic enzymes (Castello et al. 2012, Baltz et al. 2012). A flurry of similar studies followed using different cell types (Mitchell et al. 2012 in yeast; Kwon et al. 2013 in embryonic stem cells; Beckmann et al. 2015 in human liver cells and yeast). This cohort of papers identified many previously uncharacterized RBPs, many

without known RBDs or roles in RNA biology. Beckmann et al. 2015 coined the term enigmaRBPs for these unexpected RBPs. These studies have opened the realm of RBPs beyond proteins that can be strictly defined by computational studies or known functional annotation. Next, we aim to understand what mRNAs are targeted by RBPs and the functional role of bound RBPs in affecting post-transcriptional regulation.

There are several approaches to identify targets for a known RBP. While it is possible to assay the RNAs bound by specific RBPs without crosslinking (see Tenebaum et al. 2000 for example), the development of crosslinking immunoprecipitation paired to RNA sequencing (CLIP-seq; reviewed extensively in Darnell 2010) chemically preserves protein-RNA interactions and enriches for the protein of interest. Briefly, cells are treated with ultraviolet (UV) light to form zero-length covalent links between an RNA and directly bound proteins. Immunoprecipitation is used to enrich for the protein of interest. Following immunoprecipitation, the protein can be degraded to liberate the RNA fragments for sequencing library preparation. Early efforts utilizing low-throughput Sanger sequencing identified just over 300 targets of the splicing regulator RBP Nova in neuronal cells (CLIP, Ule et al. 2003). A similar approach was applied in mammalian cell culture to purify transcripts that bear iron response element (IRE) landing sites for the RBP regulators iron response proteins 1 and 2 (Irp1, Irp2); purified transcripts were then identified using microarrays, allowing for the identification of more target transcripts than the earlier low-throughput sequencing approach (Sanchez et al. 2006). Cross-linking and analysis of cDNAs (CRAC) was used to identify RBP targets through both Sanger and high-throughput sequencing in analysis of rRNA processing (Granneman et al. 2009) and several stages of yeast RNA processing (Tuck and Tollervey, 2013). The original CLIP Nova study was later repeated using high-throughput sequencing, increasing the number of known Nova targets to nearly 2,500 (HITS-CLIP, Licatalosi et al. 2008). A

similar study of the RBPs in yeast P-bodies found a surprising lack of specificity of several RBPs (eg Pat1, Lsm1, Dhh1; Mitchell et al 2012). CLIP can also be enhanced by the metabolic incorporation of photoactivatable ribonucleosides, which improves crosslinking efficiency (PAR-CLIP, Hafner et al. 2010). Nucleotide-level information of protein binding can be obtained using reverse transcription up to the crosslinking site that serves as a physical impediment to the polymerase (iCLIP, König et al. 2010). An alternative to directly immunoprecipitating the protein of interest is to use a GFP-fusion and fluorescently labeled RNA approach (Strein et al. 2014). To identify what RNAs are bound by all potential RBPs rather than just those by a specific protein, protein-bound RNA can be isolated using sucrose sedimentation; this approach has determined that >70% of the yeast genome is bound by an RBP at any given time (Freeberg et al. 2013). Given the pervasive binding of transcripts by Pab, the 70% measured by Freeberg et al. 2013 very likely underestimates RBP binding. With numerous assays to define the specific targets of RBPs in a high-throughput way, we next aim to answer how these RBPs are regulating their target transcripts.

Functional characterization of RBPs has been done on single genes using tethering assays (e.g. PABP, Coller et al. 1998; Gw182, Fabian et al. 2011). However, there is not a published method for the high-throughput analysis of RBP function. From RNA-association studies, we know that there are a large number of unexpected or enigmRBPs (Tsvetanova et al. 2010, Beckmann et al. 2015); these studies provide lists of regulators, but they do not provide functional information for what these RBPs are doing to bound transcripts. In addition, we predict that there is extensive opportunity for post-transcriptional regulation by the measured discrepancies in RNA-protein abundances (Ingolia et al. 2009, Schwanhausser et al. 2011, Li et al. 2014). To mechanistically connect potential post-transcriptional regulation with proposed post-

transcriptional regulators, we need to expand the functional annotation of these regulators at a genome-level. To address this need, we have developed both physical technologies and molecular biology tools to create a system for functionally characterizing RBPs and post-transcriptional regulators in a way that is both comprehensive and genome-wide.

Chapter 2:

Selection and generation of diverse genome-wide expression libraries using large-scale turbidostats

Purpose

To perform functional characterization of proteins at a genome-wide scale, we devised a method and designed custom large volume turbidostats for investigating protein expression libraries that represent all potential proteins multiple times.

Motivation

One difficulty in generating genome-scale expression libraries is selecting in-frame fragments without stop codons from the pool of genomic sequences. We overcome this challenge by selecting for expression of an in-frame marker downstream of the tether protein. We perform this selection on approximately 100 million genomic fragments in a custom, large volume turbidostat. In contrast to previously described turbidostats, the larger volume growth chambers accommodate selection of diverse expression libraries. Furthermore, our custom turbidostats provide an affordable alternative to comparable commercially available bioreactors. We show that our system yields tens of thousands of distinct, in-frame fragments in a cost-effective manner.

Introduction

Recent technological advances enable systems-level studies. Rather than single gene studies, entire transcriptomes or proteomes can be examined simultaneously. While it is possible to measure the transcription or translation of a gene in the cell, we need more

methods for functionally characterizing proteins at a genome-wide level. One approach is knockout libraries that aim to systematically disrupt each gene in a genome individually; there are several gene knockout libraries available for *Saccharomyces cerevisiae* (*S. cerevisiae*). While these libraries are useful resources, they cannot be used to study genes that are essential for cell survival. In addition, the role of a single gene can be confounded due to secondary effects from total gene knockout.

Considerable effort is being put forth by the Boone lab generating pairwise knockouts to establish genetic interaction (GI) maps (reviewed in Costanzo et al. 2011). Briefly, GI maps look for genetic epistasis, alleviation of one candidate gene phenotype, or synthetic negative phenotypes between pairs of genes. Combinations of genes that are affected positively or negatively in a paired knockout can point to shared biological pathways or molecular mechanisms. As such, they can be informative to establish the order and mechanism of complex biological pathways. Higher eukaryotes have classically been less tractable for generating single or double knockouts; as a proxy for genetic disruption, transcription or translation of a gene can be lessened or knocked down through the use of RNA interference or transcriptional interference. Bassik et al. (2013) have demonstrated a method for pairwise GI maps in mammalian systems through the use of high coverage short hairpin RNA (shRNA) libraries. However, GI maps still require picking candidate genes to knock out based on prior knowledge.

We developed a method for creating a genome-wide protein expression library, and applied it to the study of post-transcriptional regulators. There are greater than 600 known RNA binding proteins (RBPs) defined in the literature in the budding yeast *S. cerevisiae* alone. This number is likely an underestimate given the difficulty of predicting RBPs based on protein sequence or functional annotation alone (Tsvetanova et al. 2010; Beckmann et al. 2015). To overcome the challenge of predicting which proteins or

domains are capable of binding RNA, we aim to clone and screen every protein and protein domain for post-transcriptional activity. If we conservatively assume that every gene in the yeast genome (~6000; Müller et al. 2002; Byrne and Wolfe 2005) has at least two domains, preparing a genome-wide library requires cloning 12,000 constructs and prior knowledge of domain boundaries. Instead of candidate-driven massively parallel cloning, we use randomly fragmented genomic DNA to generate a library of fragments largely unconstrained by prior expectations. These libraries are selected for coding sequence using custom-built large volume turbidostats.

Results

Large-volume turbidostats

The ability to perform comprehensive screens is limited by the ability to make large and diverse libraries of coding sequence fragments. The coverage of a genome-scale expression library is bounded by the mathematical constraints in developing the library. We use yeast genomic DNA as a source of DNA sequence with equal representation of every gene in the genome. The yeast genome is twelve megabases and double stranded, and as such it has approximately 2.4×10^7 potential base pairs representing a start point of a fragment (Table 2-1). An estimated 72.9% of the yeast genome or 1.75×10^7 base pairs are contained within exons (Alexander et al. 2010). Because coding sequences occur with a three base pair periodicity, only one third will be in frame at the start, one third in frame at the end. Moreover, only half of these in-frame fragments will be on the correct coding strand. Therefore, of the 1.75×10^7 , only 9.72×10^5 base pairs will be in a fragment that is both exonic and in the correct frame (Table 2-1). This population represents only 4.5% of the total genomic fragments possible. Enriching for this limited

population requires a method of bulk selection. To avoid the bottlenecks and technical difficulties associated with enriching for these fragments by plating yeast transformed with the library, we opted to use liquid selection of a large culture. To maintain an optimal density of yeast and exponential growth, we developed large volume turbidostats. Turbidostats are continuous growth culture chambers that have been used for a number of microbial studies (Bryson and Szybalski 1952, Sorgeloos et al. 1972, Takahashi et al. 2015); however, the volumes traditionally used in these experiments are insufficient for the size of our library. We adapted the optics design from the Toprak et al. (2013) morbidostat to continuously culture, apply selection, and monitor growth of large volume yeast cultures.

| Class | |
|---|----------|
| Base pairs in yeast genome | 2.40E+07 |
| Exonic base pairs in yeast genome | 1.75E+07 |
| In frame, coding exonic base pairs | 9.72E+05 |
| Fraction of total fragments that are coding | 4.05E-02 |

Table 2-1: Selecting protein coding sequences requires bulk selection.

Calculations are as follows: The yeast genome is 12 megabases and double stranded, therefore 24 million base pairs. 72.9% of that is estimated to be exonic (Alexander et al. 2010), where 72.9% of 24 million is 17.5 million. Only 1/18th of the coding base pairs will produce in frame fragments ($\frac{1}{3}$ in frame at the start, $\frac{1}{3}$ in frame at the end, $\frac{1}{2}$ on the coding strand), so 972 thousand coding fragments are possible. 972 thousand fragments are only .045 of the total 24 million potential fragments.

A turbidostat links culture turbidity to an inflow of media in order to maintain a constant cell density. To detect culture turbidity, we use an infrared (IR) light emitting diode (LED) to emit IR light into the culture vessel. Microbes in the culture scatter the IR, and this scatter can be detected using a light semi-conductor (Figure 2-1A; similar to Toprak et al. 2013). Measuring microbial density through measurement of scattered light is conceptually similar to traditional methods of culture density measurements using absorbance of 600 nanometer (nm) light through a cuvette to linearly estimate cell density (Figure 2-1B). To avoid interference with ambient light sources, we use an IR LED for the emitter and a semiconductor inset into a polylactic acid (PLA) 3D printed band coated in a visible and IR absorbing paint (Ultra Flat Black, Krylon). The band also maintains a constant geometry between the emitter and detector (Figure 2-1C). Comparing voltage-IR measurements to traditional spectrophotometer measurements (optical density, OD), we can detect linear changes that correlate IR to OD, and in turn cell density, using our detector (Figure 2-1D).

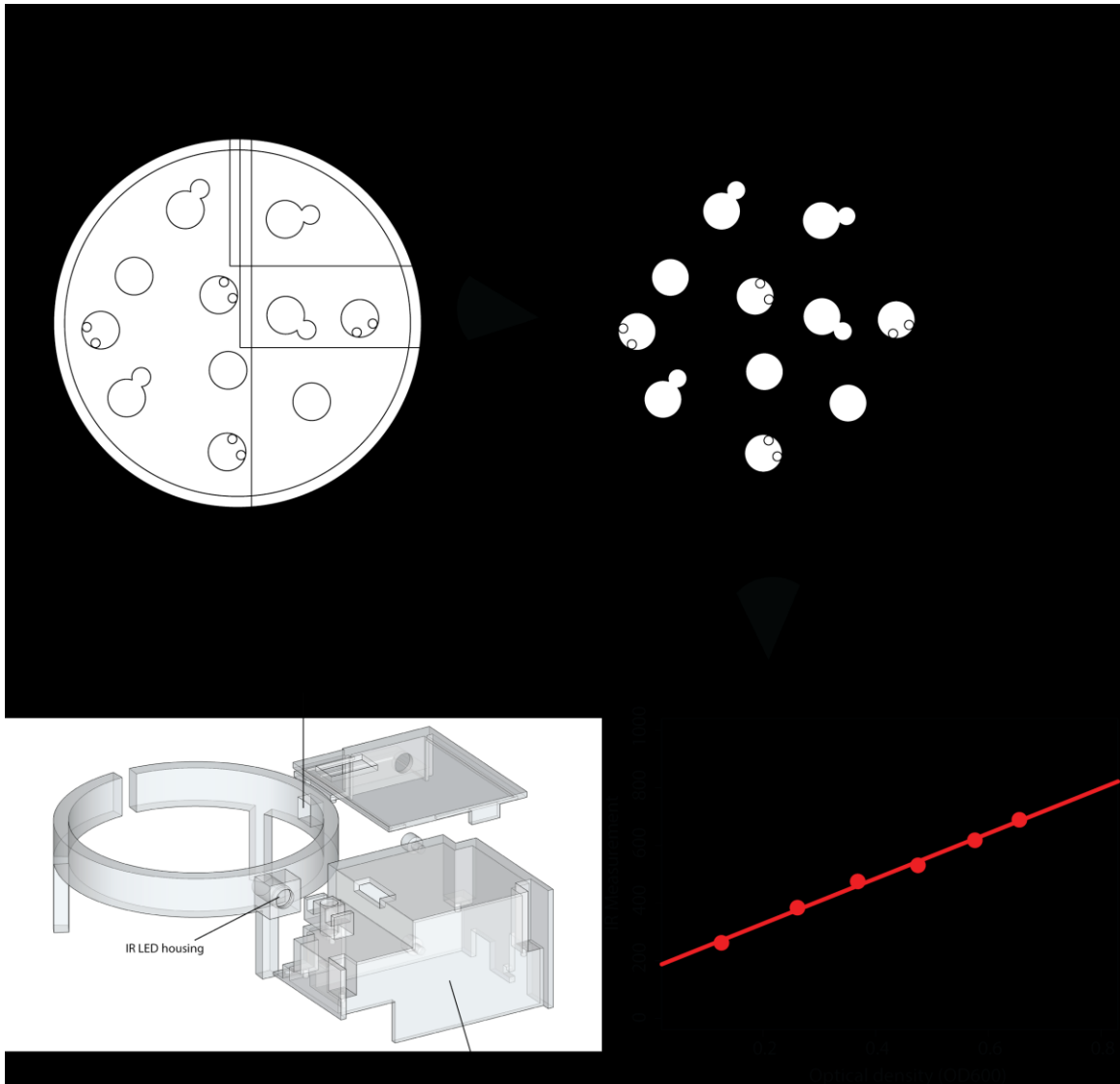


Figure 2-1: Detecting cell density in a culture

(A) We use the approach described by Toprak et al. (2013) in which measurement is done in the growth vessel rather than requiring subsampling with a traditional spectrophotometer. Light is emitted into the culture vessel from an IR LED, and the level of reflection can be empirically correlated to the traditional OD or number of cells. **(B)** Traditional methods use a standardized 1 cm² cuvette. 600 nanometer (nm) light is passed through the cuvette holding 1 milliliter (mL) of culture, and the absorbance is measured by a detector directly across from the emitter. This can then be empirically correlated to number of cells in the culture. **(C)** Toprak et al used an angle of 135° between emitter and detector. We experimentally determined that a 90° angle gave better results (data not shown). To maintain the angle and exclude external light, we mounted the emitter and detector in a custom 3D printed band made of polylactic acid (PLA) and coated in paint that

blocks all visible and IR light (Ultra Flat Black, Krylon). To eliminate electronic noise, an Arduino mini is mounted directly onto the band to receive and process signal from the IR semiconductor. **(D)** There is a linear association between IR measurement and optical density, demonstrated here for four separate bands, demonstrating that we can use reflected IR to accurately measure cell density.

The goal of the turbidostat is to maintain a constant cell density by pumping in new media when the density exceeds a user-defined threshold. The culture density is continually monitored through IR measurements that are collected and parsed by an Arduino mini processor every second (Figure 2-1C). If the measured turbidity is higher than the target (calibrated with each experiment to be approximately $OD_{600}=0.6$), the Arduino will power on a peristaltic pump connected to a media source (Figure 2-2A) to add new media until the desired turbidity is achieved (Figure 2-2B). The growth chamber maintains a positive pressure through pumping in filtered air (Figure 2-2C). This positive pressure as well as an internal line set just above the desired volume push out volume as new media is pumped in (Figure 2D).

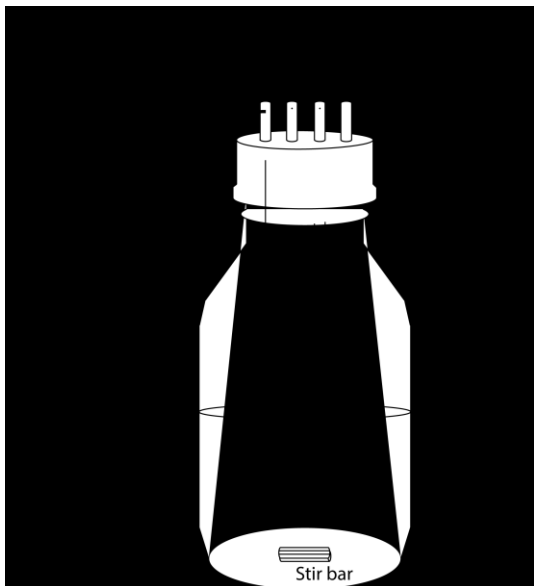


Figure 2-2: Turbidostat growth chambers

The growth vessel of the turbidostat is a standard 500mL glass bottle (VWR) with a 4-port lid (VICI) with custom poly ether etherketone (PEEK) lines. The culture volumes are set at approximately 300mL. Media is supplied through the action of a density-dependent Arduino controlled peristaltic pump **(A)** into the growth chambers first PEEK line **(B)**. The chamber maintains aeration and positive pressure through pumping in

hydrated and filtered air **(C)**. Additional aeration is provided through the action of a stir bar and stir plate. This addition of media raises the level of media in the vessel, and through positive pressure and a line

placed just above the desired volume (**D**), an equivalent amount of culture is removed. A fourth line attaches to a syringe for inoculation of the chamber with yeast as well as allowing for sampling or collecting a portion of the vessel at any time (**E**).

In addition to making decisions about the current cell density and pumping, the Arduino program provides a record of the volume of media pumped over time. By measuring the rate of media replacement, it is possible to track the growth rate of the culture (Figure 2-3). We can then compare the measured rate of growth to expected values to determine if the culture is maintaining exponential growth. This information is useful in determining the status of the culture with regard to stress and cellular homeostasis.

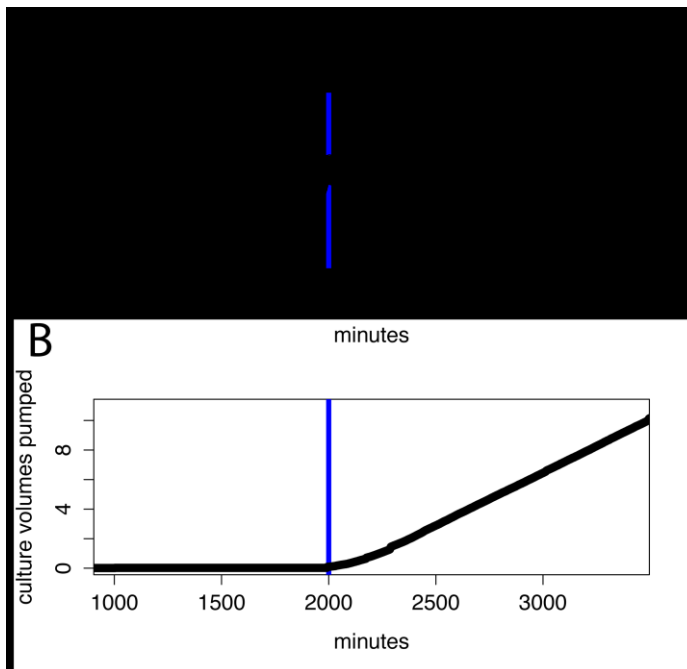


Figure 2-3: Growth measurement using IR detection

The turbidostat growth chamber is fitted with an optics ring as described in figure 1. Target IR (**A**) is calibrated based on the linear relationship between IR and OD and an initial manual OD measurement. This graph shows yeast transformation following adjustment (at 1000 minutes), through the target approaching target density (OD \approx 0.6) at approximately 2000 minutes. A blue line marks this point when the culture

reaches target density and pumping activates. Knowing the volume of the culture, we can calculate the number of volumes pumped over time (**B**). This allows us to calculate the doubling time through linear regression of pumping; here, the yeast are doubling every 96 minutes, in line with expected growth rates in synthetic media.

Generating a genome-wide expression library

Several recent studies have highlighted that many RBPs have neither annotated RBDs or RNA-associated function (Tsvetanova et al 2010; Beckmann et al. 2015). While these studies provide a list of potential RBPs and post-transcriptional regulators, they don't ascribe functional effects to the RNA-protein interaction. Current techniques to determine the function of RBPs are done on single genes. While it is technically possible to create and characterize the >600 identified RBPs in yeast alone, we chose to create a comprehensive genome-wide library.

An ideal protein expression library will carry in equal measure all potential coding sequences represented in the genome. As we've developed the growth chambers for yeast growth, we want a source for all potential coding sequences in yeast. The yeast genome is very information-rich; estimates put the protein coding content of the yeast genome at 72.9%, (Alexander et al. 2010). Because of this, we can use yeast genomic DNA (Figure 2-4A) as a source of DNA representing each gene in equal abundance rather than relying on normalized cDNA. The average yeast gene is 1.6kb (Milo et al. 2010), and we hypothesize there are 2-3 domains per gene. To try to capture one domain per library candidate, we fragment genomic DNA (Figure 2-4B) and select for the expected size of a domain (Figure 2-4C) with some expected reduction in size during library prep (Mills et al. unpublished).

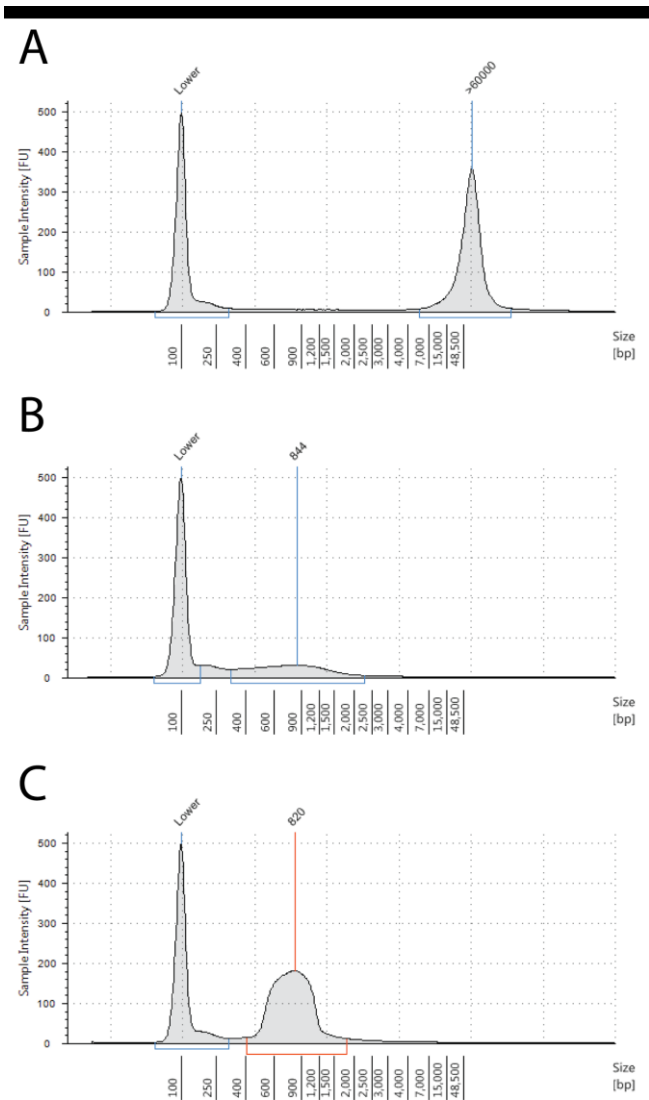


Figure 2-4: Size selecting genomic DNA fragments

(A) Commercially available yeast genomic DNA (EMD Millipore catalog #69240-3) is present in high molecular weight species. **(B)**

This genomic DNA is fragmented via sonication using a Covaris S220 at 5% duty factor, 175W peak incidence power, 200 cycles per burst, and 25 seconds (TruSeq DNA PCR-Free Library Prep Reference, Illumina 2016).

(C) Given the average yeast gene of 1.6kb and 2-4 domains per gene, we estimate the average domain size to be between 300-800bp. To enrich for fragments in this size range, fragment genomic DNA is size selected between 500-1200bp using a Pippin Prep™ 1.5% agarose cassette. All samples shown above are run on an Agilent Bioanalyzer Genomic DNA cassette to assay effectiveness of sonication and size selection.

After generating DNA fragments that are size-selected for presumptive protein-domain encoding potential, we required a method for expressing and selecting for in-frame fragments. The sonicated fragments are prepared for incorporation into an expression vector using either the NEBNext® or Illumina library prep kits with standard Illumina-based primers (Figure 2-5A). These Illumina primer sites are used as homology arms for

in vivo recombination into a linearized yeast expression vector. Importantly, the expression vector encodes for a selectable auxotrophic marker. This marker is expressed downstream of the incorporated fragment through the poly-cistronic viral sequence 2A (Donnelly et al. 2001; Kim et al. 2011). Briefly, 2A fails to form a peptide bond between the upstream and downstream sequences, allowing for the expression of two distinct protein products from a single reading frame (Donnelly et al. 2001). The auxotrophic marker will only be successfully expressed if the upstream fragment is in frame at the start, in the end, and on the coding strand (Figure 2-5B). As mentioned previously, correct framing occurs in only 4.5% of all total genomic base pairs. As such, using a poly-cistronic transcript bearing a selective marker allows us to select for only these desired coding fragments via permissive growth.

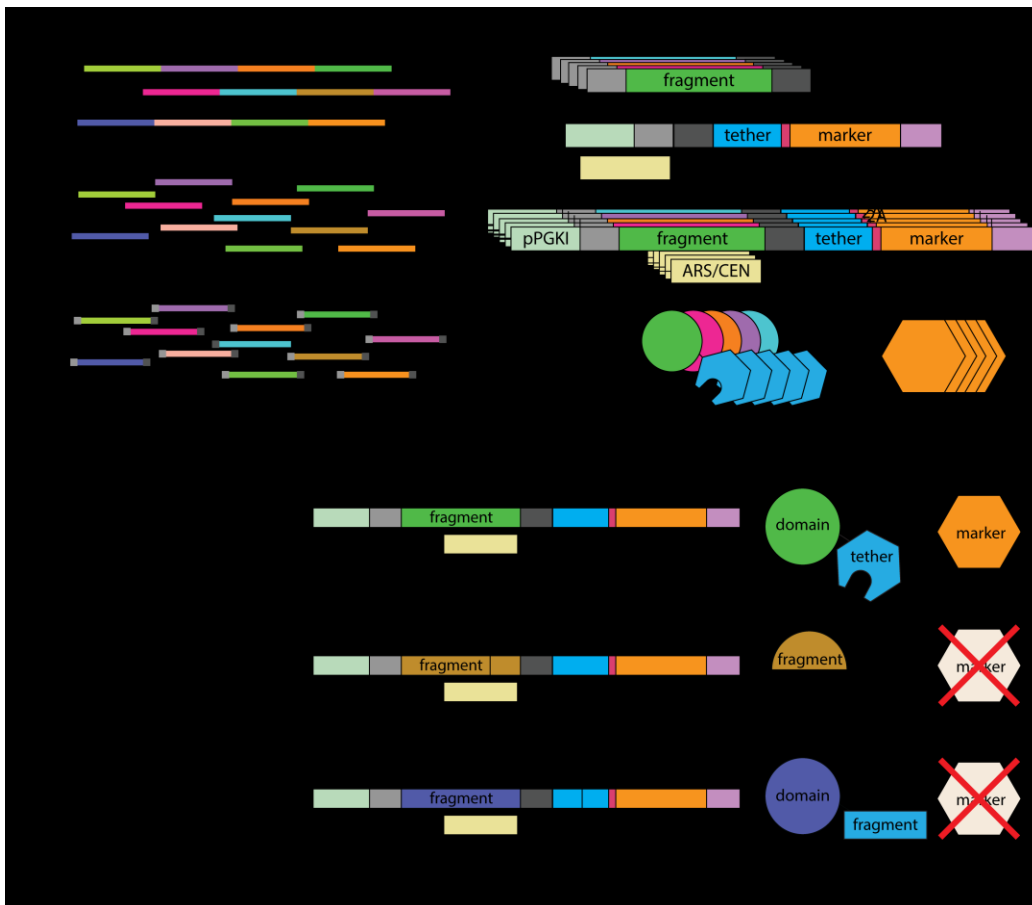


Figure 2-5: Generating a genome-wide protein-expression library

To screen proteins at a genome-wide level, we developed a method for generating a genome-wide protein expression library. **(A)** Yeast genomic DNA is information-rich, and therefore used as a normalized cDNA proxy. (i.) Different genomic regions are represented by different colors. (ii.) The genomic DNA is fragmented (as in Figure 4) and (iii.) given Illumina adaptors (represented by light and dark grey boxes). (iv.) These fragments can then be recombined into a yeast expression vector carrying essential yeast features such as a promoter (PGKI promoter, pPGKI) and ARS/CEN. The fragment recombination region is followed by a tether for studying post-transcriptional regulation. Downstream is a viral sequence, 2A, that prevents peptide bond formation between the upstream and downstream sequence. (v.) The peptide-skipping of 2A produces two distinct protein products: the upstream domain-tether and the downstream selective marker. **(B)** The combination of an upstream fragment recombination site, downstream 2A, and further downstream selective marker selects only for the 4.5% of the total potential genomic fragments that encode a protein domain. (i) Fragments that are on the coding strand and enter and leave in frame will produce a functional domain and downstream marker. (ii) Fragments of the coding strand, but that are out of frame at the start are likely to be interrupted by an in-fragment premature termination codon (PTC), resulting in no selective marker expression. (iii) Fragments of coding strand, but that are out of frame at the end will contain an in-tether PTC, resulting in no selective marker expression.

Selection of a comprehensive genome-wide library

The actual selection of the expression library occurs in turbidostat-grown yeast. We transform yeast with the aforementioned linearized expression vector and a recombination-compatible library of potential coding fragments. In early development tests we established that for these transformations, $1.6 \times 10^{-3}\%$ of the total population is both transformed and expresses an in-frame library fragment and downstream selective marker. Explicitly, for every 6×10^8 yeast put into the transformation reaction, only 1×10^4 result in successful recombination of in-frame, coding strand library fragments (Figure 2-6A). Because this population of 1×10^4 yeast can read through to and express the downstream selective marker, they are capable of growing in selective media. In contrast, the remaining yeast that are either not transformed, not properly recombined,

or are carrying recombinant expression vectors with out of frame or non-coding sequence cannot grow under the selective conditions (Figure 2-6A). To allow for the minor population of transformed yeast to overcome the non-selective population before pumping occurs (at OD=0.4–0.6), we inoculate large volume turbidostats at an OD₆₀₀~0.2. Given the expected portion of the population carrying our selection-positive expression library, we can computationally predict the rate at which yeast carrying expressing fragments from the library will outpace their non-expressing counterparts. Because the turbidostats maintain a constant volume of culture media, we can represent the relative abundance of the selective or non-selective populations as a percentage of total cell count (Figure 2-6B). We collect yeast when we predict that approximately 99.9% of the total population is represented by successful in-frame and coding library fragments.

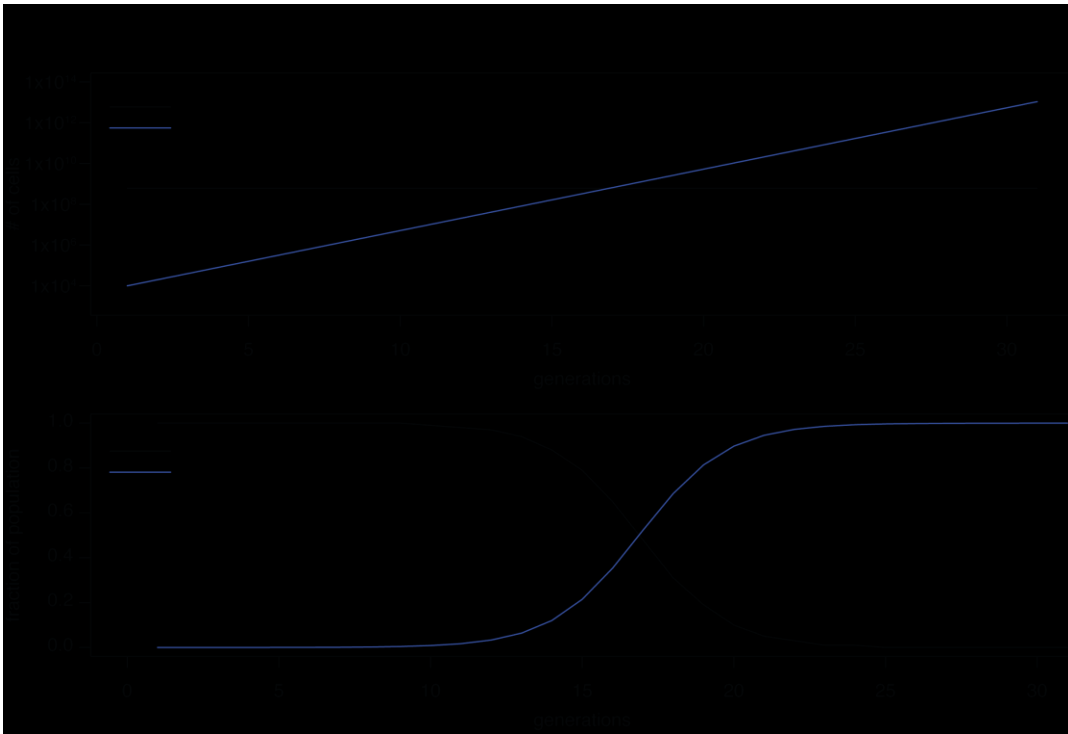


Figure 2-6: Generations of growth required to select for protein-expression library

(A) Assuming we transform 6×10^8 cells, we positively transform, recombine, and have proper strand and framing for 1×10^4 cells. Because this latter population is capable of surviving under selection, it grows while the non-marker population remains constant. Between 15 and 20 generations, the selective population overtakes the non-selective cells. **(B)** Because the turbidostats maintain a constant volume, we can calculate the percentage of population represented by protein-expression library containing cells. Around 23 generations, the population reaches ~99.9% selective library cells. In addition, because the population of selective cells began as such a small percentage of the original population, we know that at this point each individual recombination event is represented by many individual daughter cells.

We next assayed how much of the protein-coding potential of the genome we are capable of recovering. To determine the starting population of potential fragments, we sequenced a library of pre-recombination, pre-selection library fragments. To assay the extent of genome coverage in our post-selection libraries, we prepared and sequenced a library from the yeast collected at point where 99.9% of the population is expected to express transformed protein. We observe good genomic coverage in both the pre-recombination and post-selection libraries (Figure 2-7A). As expected through fragmentation, this pre-recombination and pre-selection library shows even distribution across all potential reading frame combinations (Figure 2-7B). In contrast, yeast that had been subjected to recombination and auxotrophic selection showed a dramatic (seven-fold) enrichment for in-frame and coding sequences (Figure 2-7C). In this post-selection library, each gene in the genome is represented by a distinct in-frame fragment approximately 21 times across all expected size ranges. Together, this demonstrates the validity of our approach in generating and selecting for genome-wide protein expression libraries.

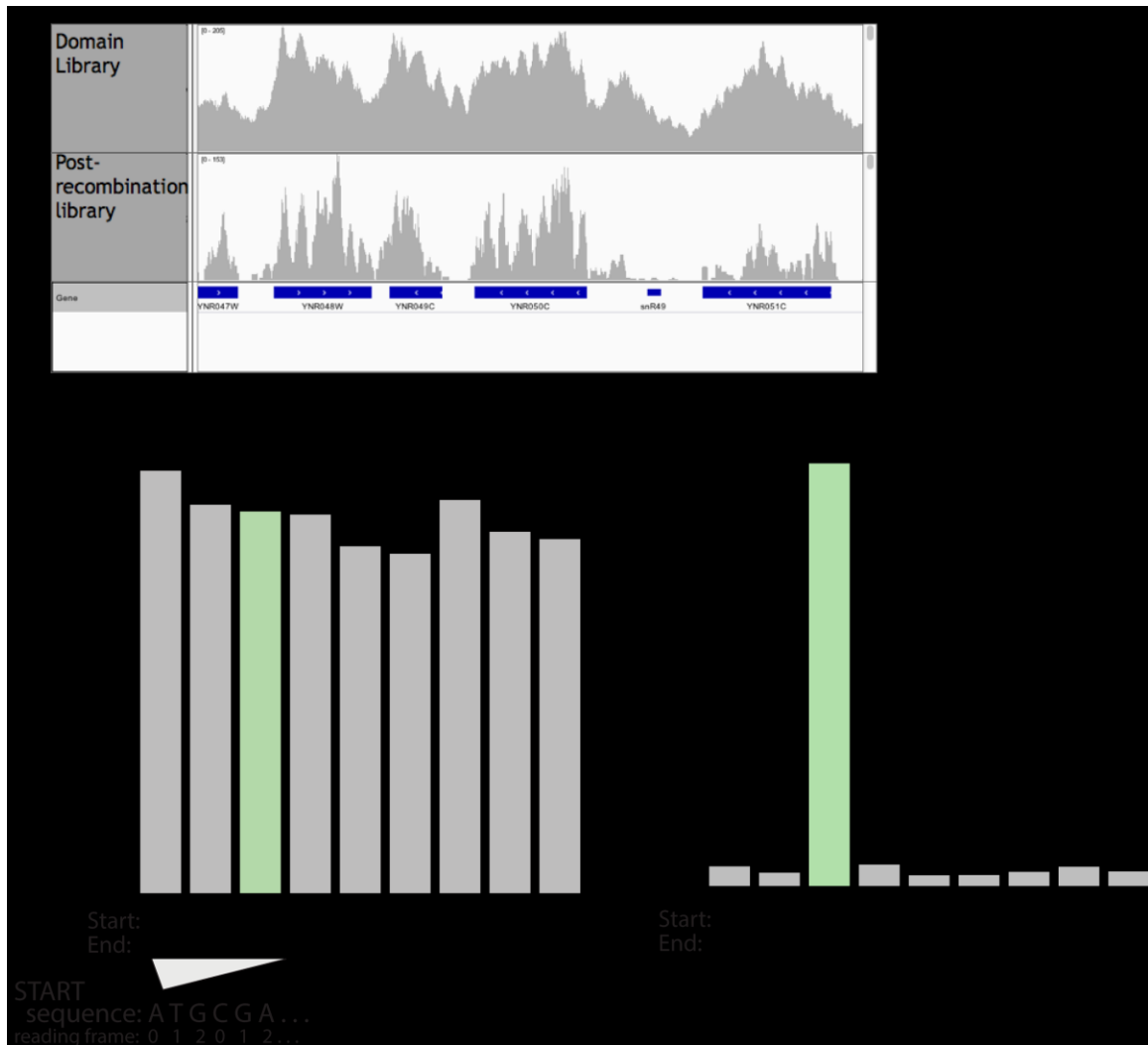


Figure 2-7: Genomic coverage and framing in pre- and post-selection libraries

Libraries were generated as described in figures 4 and 5. **(A)** An example Integrated Genome Viewer (IGV) generated plot for read counts in pre- and post-selection libraries. In both libraries, there are many copies aligning to each position (up to 200 at some positions). In the post-selection library, there is an enrichment for fragments aligning exclusively to coding sequences. Given the constraints on selection for coding, in-frame fragments, this is to be expected. **(B)** In a pre-selection library, genomic fragments map equally to any of nine frame combinations. The frame that would result in protein-expressing and downstream selective marker expression is highlighted in green and represents only 11% of the total population. **(C)** Yeast were transformed, recombined, and grown under selection for marker expression until the culture was predominantly library expressing cells (as in figure 6). DNA was isolated from these cells and sequencing libraries prepared. Following selection, the coding sequence, in-frame fragments represent 76% of the

population, a nearly seven-fold increase from pre-selection. This predominantly protein-coding library can then be used to screen for protein activities in a genome-wide level approach.

Discussion

We have incorporated two novel techniques in this experiment. First, we built custom large-volume turbidostats capable of growing yeast at a user-specified density. Second, we used these large-volume turbidostats in the selection of an in-frame fusion library; specifically, we use polycistronic expression of auxotrophic marker to generate a genome-wide protein expression library. We later used these turbidostat selected genomic libraries to assay for post-transcriptional regulators via fluorescence activated cell sorting (FACS). However, the ability to generate genome-wide expression libraries is widely applicable to a number of diverse approaches. The large volume turbidostats can be assembled with readily available components and limited technical expertise. They therefore provide an easily customizable and affordable alternative to commercially available bioreactors or existing small-volume turbidostats. In addition, because the media source of the turbidostats can be easily changed during an experiment, these turbidostats can be used for varying selective pressure or growth conditions (such as increasing drug concentration or reducing nutrient concentrations).

The method we describe for generating a genome-wide library protein expression library provides a mechanism for functional characterization of proteins and protein domains on a previously intractable level. We use this for investigating post-transcriptional regulatory potential across the genome. However, this method could easily be adapted for selection of directed evolution libraries, mutational or deletion scanning, functional studies of protein-protein interactions, and more. In combination with the large-volume turbidostats,

there is nearly unlimited potential for screening genome-wide libraries of protein for specific activities of interest.

Chapter 3:

Genome-wide survey for post-transcriptional regulators in *Saccharomyces cerevisiae*

Purpose

This chapter will focus on a genome-wide survey for active post-transcriptional regulators in yeast utilizing the protein expression libraries generated in chapter 2.

Motivation

Post-transcriptional regulation plays an important role in modulating gene expression in situations of stress, development, and even non-stressed growth. Transcript translation and stability can change in response to binding of regulatory RNA binding proteins (RBPs). Proteomic studies have identified over 500 RBPs in *Saccharomyces cerevisiae* (Tsvetanova et al. 2010; Beckmann et al. 2015); however, the functional impact of these proteins on the transcripts they bind remains unclear. Tethered function assays have been used to elucidate the post-transcriptional regulatory effect of proteins and protein domains. We adapt this tethered function assay into an approach for an unbiased survey of regulatory activity in proteins of the budding yeast *S. cerevisiae*. Our system is capable of interrogating tens of thousands of unique protein fragments genome-wide. We recover known post-transcriptional regulators and reveal new activities in uncharacterized proteins. Our approach offers a comprehensive view of the mechanisms underlying post-transcriptional control in yeast and functional annotation of cryptic RNA binding proteins.

Introduction

Regulating gene expression is essential for biological processes, ranging from development and differentiation to stress response and cellular homeostasis. Many studies rely on RNA abundance measurements from RNA-seq to estimate gene expression. However, in most cases, the protein resulting from the expressed transcript is the effector for biological action. Several recent studies have demonstrated that the correlation between RNA and protein abundances is weak; the abundance of proteins encoded by mRNAs with similar abundance can differ in abundance by up to 100-fold (Ingolia et al. 2009, Schwanhausser et al. 2011, Li et al. 2014). The differences between RNA and protein levels can be regulated through changes in transcript stability, transcript localization, and translational efficiency, collectively termed post-transcriptional regulation. We know of several specific examples where post-transcriptional regulation is essential for development (*bicoid* mRNA, Nanos, and *hunchback* mRNA in *Drosophila* embryogenesis), stress response (Atf4 and Gcn4 in mammals and yeast, respectively), and cellular homeostasis (Pabp across eukaryotes). Given the importance of post-transcriptional regulation, there have been relatively few studies that focus on the global impact of post-transcriptional regulation.

RNA binding proteins (RBPs) are an attractive candidate for regulating transcript fates. RBPs are often defined either by the presence of a conserved RNA binding domain (RBD) or by empirical detection of proteins that associate with transcripts. However, in a literature review by Tsvetanova et al. (2010), many of the proteins defined as RBPs lack canonical RBDs (Figure 3-1A). Moreover, proteins experimentally determined as associating with RNA often lack functional characterization relating to RNA biology or transcript regulation (Figure 3-1B). This raises the question: what are the functions of

these unexpected RBPs? We address this question by measuring the effect of protein fragments tethered to a reporter transcript in a genome-wide library of protein fragments. Using this method, we have characterized over ten thousand protein fragments for their functional role in post-transcriptional activity.

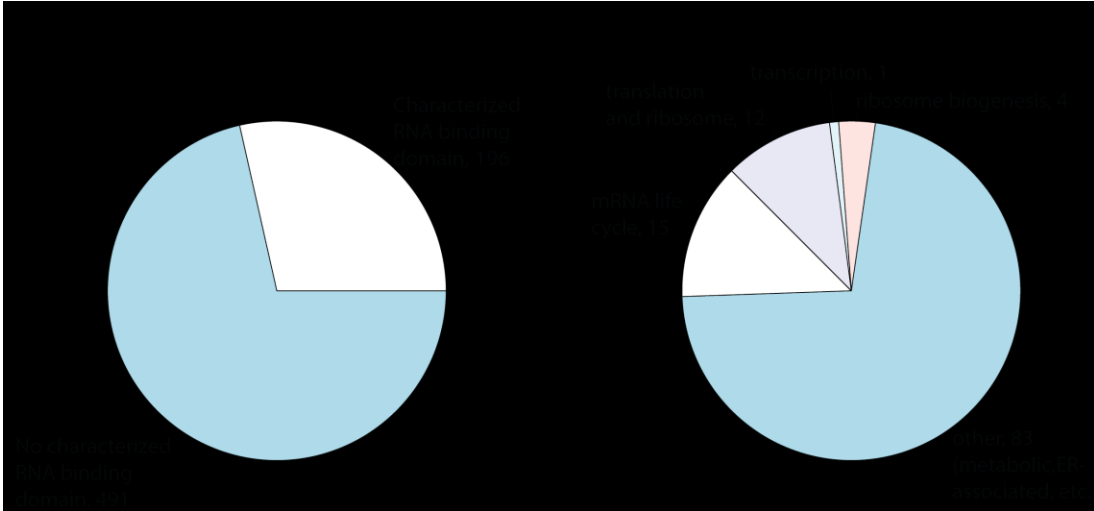


Figure 3-1: RNA binding proteins are often unexpected from existing annotation

Data collected and plotted from Tsvetanova et al. 2010. **(A)** In a literature review of annotated RBPs, just over 25% of the total proteins had a canonical RBD (e.g. RRM, PUF, ZH, etc). In contrast, the majority lacked known RBDs; this makes it difficult to computationally predict an RBD using analysis of primary sequence and protein homology. **(B)** Tsvetanova et al. 2010 assayed the proteins bound to RNA by attaching total cell RNA to beads, flowing cellular lysate over the beads, and assaying the proteins that bound by mass spectrometry. Of the proteins identified, again just over 25% of them have annotations that match known RNA functions. In contrast, the remainder had metabolic, endoplasmic reticulum, vesicular, or other ontologies. The lack of RNA-associated functions further complicates the prediction of RNA binding proteins based on existing annotations. Moreover, it raises the question of how these unexpected RBPs binding and regulating transcripts.

Results

Developing a two-color, in vivo tether assay

Post-transcriptional regulation can be separated into two steps: recruitment of the RBP to a transcript and the exerting post-transcription control on the target transcript. Studies on the post-transcriptional regulator Gw182 demonstrate that these activities can be functionally independent of each other; the modularity of recruitment and repression domains is a common but not strictly essential feature of many RBPs (reviewed in Lunde et al. 2007). Briefly, Gw182 is recruited to a target transcript through the interaction of its N-terminal domain with a microRNA-loaded Argonaute 1 or 2 (Rehwinkel et al. 2005, Chekulaeva et al. 2009, Chekulaeva et al. 2011). Gw182 exerts its regulatory effect through the C-terminal region by interfering with mRNA circularization and by recruiting decay factors (reviewed in Triteschler et al. 2010). The mRNA decay and translational repression by Gw182 can be recapitulated through artificial recruitment of either the C-terminal or Silencing Domain alone (Chekulaeva et al. 2009, Zekri et al. 2009). Artificial recruitment can be accomplished with a set of mRNA encoded hairpins in a reporter mRNA and a small viral peptide that has high affinity for these hairpins and is fused to the protein of interest (Coller and Wickens, 2007). This combination of hairpins and small peptide has been used to study a number of post-transcriptional regulators, including but not limited to Pabp (Coller et al. 1998, Gray et al. 2000, Tsuboi and Inada 2010), Pop2 (Finoux and Séraphin, 2006), and exon junction complex proteins (EJC, Nott et al. 2004).

Tethering is usually performed *ex vivo* in cell lysate or *in vitro*; we adapted the tethering system to work in living cells with sensitive, single-cell detection using a dual fluorescence reporter system (Figure 3-2A). We express two distinct fluorescent reporter transcripts, eYfp (yellow) and mCherry (red), each bearing one of two orthogonal tethering cassettes, boxb or pp7 (Figure 3-2B). Functionally, the corresponding viral

peptide will bind only its target hairpins, not the orthogonal set. This peptide-hairpin specificity allows us to target a putative regulator-tether fusion to only one of the two transcripts. As a result, only the tethered fluorescent reporter will read out the effect of post-transcriptional activity. The other fluorescent reporter will act as an internal control, allowing for precise, single cell measurements. If the tethered fragment is a post-transcriptional repressor or co-repressor, there will be a decrease in abundance of the target reporter protein without a corresponding change in the non-target reporter protein abundance (Figure 3-2C). Conversely, a post-transcriptional activator or co-activator will increase the level of the target but not the non-target (Figure 3-2E). Conceptually, we have designed this system to allow for swapping the target fluorescent reporter or tethering system to account for reporter or tether specific effects.

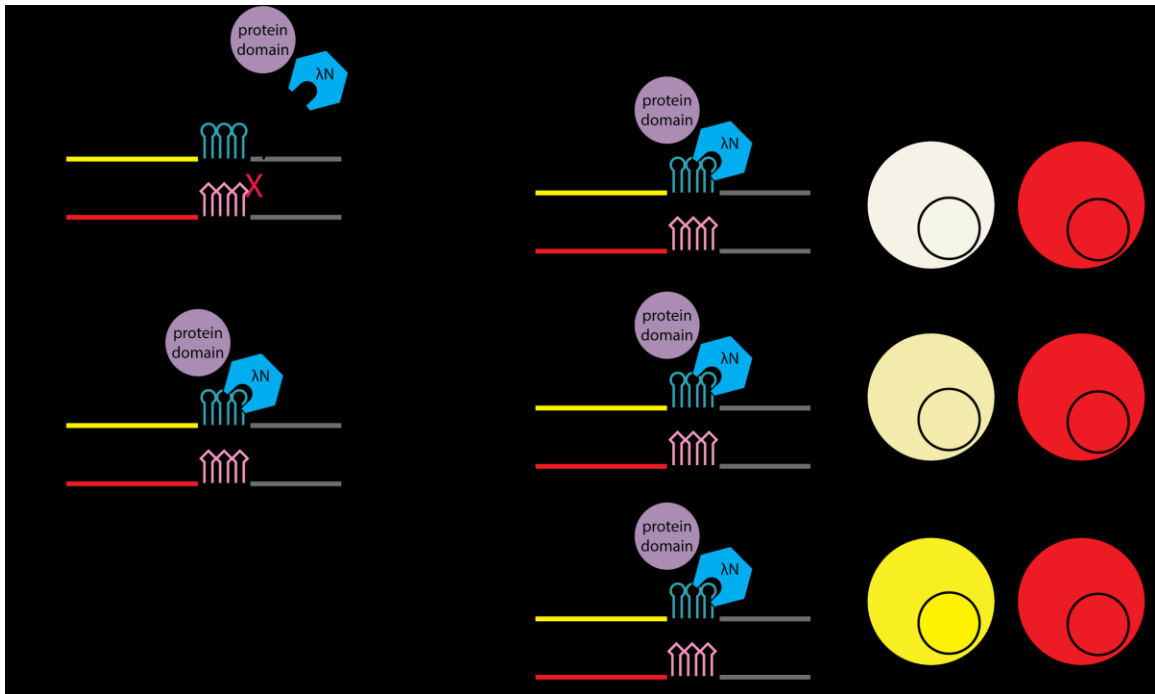


Figure 3-2: Dual fluorescent reporter tethering system for measuring post-transcriptional regulation in vivo

Design and expectation of dual reporter system. **(A)** To perform tethering assays in living cells with sensitive, single-cell precision, we use two fluorescent reporters (eYfp and mCherry) bearing orthogonal tethering cassettes containing 3–5 hairpin repeats (boxb and pp7). These constructs were transformed into the URA3 locus of S288C α or α cells. The pairs of reporters were mated to create a red and yellow positive strain expressed biallelically. **(B)** The use of orthogonal tethering systems means that the corresponding viral peptide, such as λ N, will only be recruited to its partner hairpin sequence and not the other hairpin set. **(C)** If the tethered protein domain is a post-transcriptional repressor or co-repressor, the total protein output of the target transcript will be decreased whereas the non-target transcript will remain unchanged. This ratio is determined relative to a non-regulator or tether-only ratio of the fluorescent proteins. **(D)** If the protein domain is not a post-transcriptional regulator, there will be no change in the ratio of fluorescent proteins. **(E)** If the protein domain is a post-transcriptional activator or co-activator, there will be an increase in the target protein output with no change in the non-target. By measuring the extent of deviation of the repressor or activator yellow to red ratios relative to a non-regulator, we can observe both the direction and extent of regulatory effect. The system is designed such that we can swap the tether used or the transcript targeted to account for artifacts of experimental design.

After building the dual-fluorescent tether system, we wanted to verify that the system was capable of reporting both positive and negative regulation. To test this, we cloned a number of known post-transcriptional regulators and measured the resulting ratios of target to non-target fluorescence using flow cytometry (Figure 3-3A). To confirm that post-transcriptional regulation could occur using our 3' UTR encoded hairpins and the fusion tether, we tested both a decapping (5' UTR) regulator, Dhh1 (reviewed in Collier and Parker 2004), as well as a deadenylation (3' UTR) protein, Pop2 (Daugeron et al. 2001, reviewed in Schoenberg and Maquat 2012). Both negative regulators were capable of driving repression using boxb- λ N tethering to either mCherry (Figure 3-3B) or Yfp (Figure 3-3C). There are very few translational activators defined in the literature; however, Collier et al. (1998) show that in yeast, tethering the poly(A) binding protein 1 (Pab1) to the unstable mating factor alpha (*mfa2*) transcript was sufficient to stabilize the

transcript, thus acting as a post-transcriptional activator. Tethering Pab1 to mCherry results in a distinct increase in the mCherry/Yfp ratio (Figure 3-3B). Tethering to Yfp yields a smaller but reproducible increase (Figure 3-3C). Although both fluorescent reporters are transcriptionally expressed from the same promoter and genomic location, there is less dynamic range for up-regulation using Yfp. The difference in dynamic range highlights the importance of the ability to swap tethers and reporters to account for regulation that could be tether- or reporter-specific. Overall, we are capable of detecting both positive and negative regulation at both the 5' and 3' UTR of RNA despite the hairpins being encoded in the 3' UTR. The results from these known regulators validate our system and suggest that we can use it to assay previously uncharacterized post-transcriptional regulators.

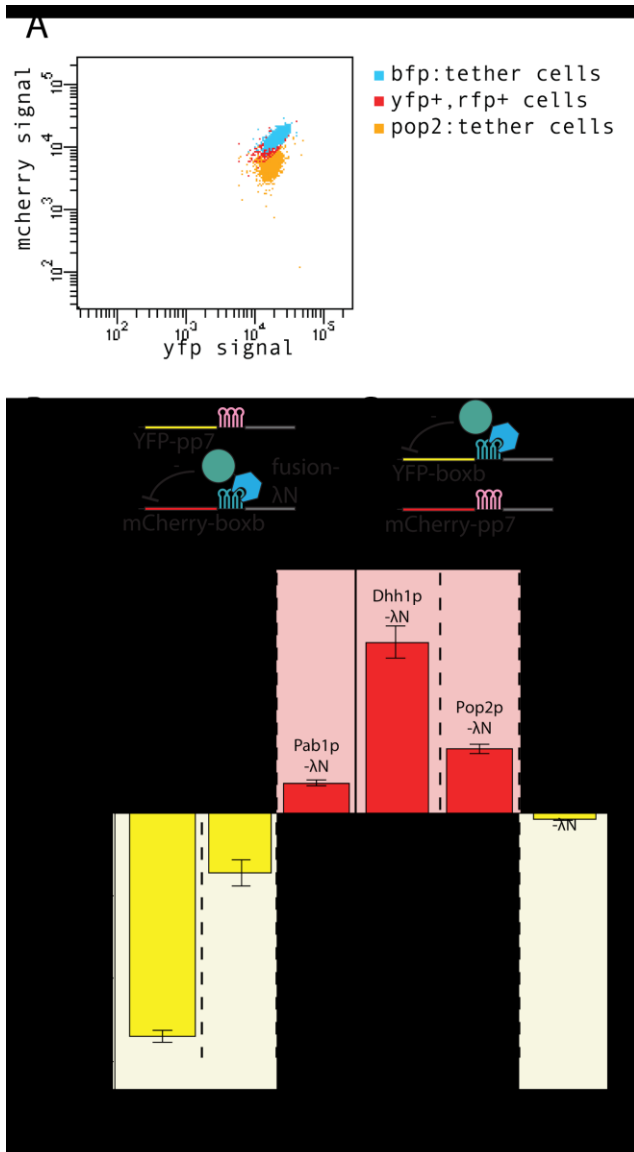


Figure 3-3: Testing positive and negative regulation in dual fluorescent tethering system

After constructing the dual fluorescent strains, we assayed for post-transcriptional regulation using known post-transcriptional regulators. In figures, the protein tethers are referred to by their yeast nomenclature wherein a p is appended to the end of the name to signify protein. **(A)** Fluorescence levels are measured using flow cytometry. Each dot represents a single cell. Expressed fluorescent proteins show a distribution of relative fluorescent signal. The BFP-tether cells (blue) overlap with the no tether population (red). When the nuclease domain of the post-transcriptional repressor Pop2 (Daugeron et al. 2001) is recruited to mCherry transcripts, a corresponding decrease in the mCherry signal but not Yfp signal is observed.

(B) This signal can be summarized calculating

the median fluorescence of 10,000 cells and taking a ratio between mCherry and Yfp. Using base 2 logarithm transformation, the ratio is visualized above or below the y-axis for increased mCherry and Yfp expression, respectively. Dhh1 and Pop21 are known post-transcriptional repressors. Pab1 is a known post-transcriptional activator. The expected direction of change (a decrease in ratio for Dhh1 and Pop2 and an increase for Pab1 when tethering to mCherry as shown at the top) is highlighted by the shaded underlay. Where possible we tethered regions of proteins previously verified for post-transcriptional significance through tethering or other assays. As expected if the system is functioning, repression of mCherry can be observed by tethering by both the 5' repressor Dhh1 (full length, from Aditya Radhakrishnan, unpublished) or the 3' repressor Pop2 (nuclease domain, as defined by Thore et al. 2003) despite tethering being directed to

the 3' UTR. Activation is also detectable by an upward shift as the result of tethering Pab1 (first two RRM, as defined by Collier et al. 1998, Gray et al. 2000). **(C)** Using the same tether, we can switch the transcript hairpins to swap the targeted transcript. Dhh1 and Pop2 show similar levels of repression; however, the increase by Pab1p is diminished. Although the fluorescent reporters are expressed from the same promoter and are allelic to each other, Yfp is more stable. As a result, there is less dynamic range to detect positive regulation of Yfp translation. This result highlights the importance of our system's ability to swap tether target or tether system to account for reporter or tether specific effects.

Preparing a comprehensive library of potential regulators

Many of the RBPs identified in mRNA interaction screens lack either known RBDs or RNA-associated functions. Because of this difficulty in predicting RBPs, we chose to assay the entire genome for post-transcriptional regulators rather than using a literature defined list of targets. In chapter 2, we describe the generation and selection of genome-wide protein expression libraries using custom large-volume turbidostats. Briefly, information-rich yeast genomic DNA is fragmented via sonication and is subcloned via in vivo recombination into an expression vector. This expression vector contains a selective marker downstream of the fragment-tether fusion and a 2A sequence. The 2A sequence is a viral sequence that allows for polycistronic translation by preventing peptide bond formation between the upstream and downstream translation products (Donnelly et al. 2001; Kim et al. 2011). Only cells carrying productively translated fragment-tether fusions will translate the downstream selective marker, allowing for enrichment of rare in-frame recombinants. Using our custom-built turbidostats, we can apply selection to generate populations of cells carrying fragment-tether fusions that represent every gene in the genome multiple times.

Selecting for potential regulators

Genome-wide libraries of fragment tethers are generated in cells carrying the fluorescent-hairpin reporter system. Following marker selection and library generation, subpopulations are sorted by fluorescent activated cell sorting (FACS) to detect either an increase (Figure 3-2E) or decrease (Figure 3-2C) in the ratio of reporter to non-reporter fluorescence. Potential activator or repressor fusions are identified by DNA extraction, PCR recovery, and sequencing library generation from the sorted populations. The resulting libraries are submitted for high-throughput sequencing. A fragment is defined by a pair of distinct start and stop sites in the genome. Statistically speaking, each distinct fragment is likely to be the result of a single recombination event. Accordingly, the abundance of each fragment in the sorted library is likely indicative of the strength of that fragment as a regulator. Plotting the abundance of each in either the potential activator or potential repressor libraries (Figure 3-4) reveals that our FACS selection acts to deplete fragments of one type from the opposing library (e.g. activators from repressor library), with the majority of the fragments occurring in only one library or another. Fragments that occur with some significant frequency of abundance in both libraries can be classified by their relative enrichment in one library. With thousands of counts per fragment and functional depletion of fragments with opposing action, we demonstrate we can successfully generate libraries of potential activators or regulators.

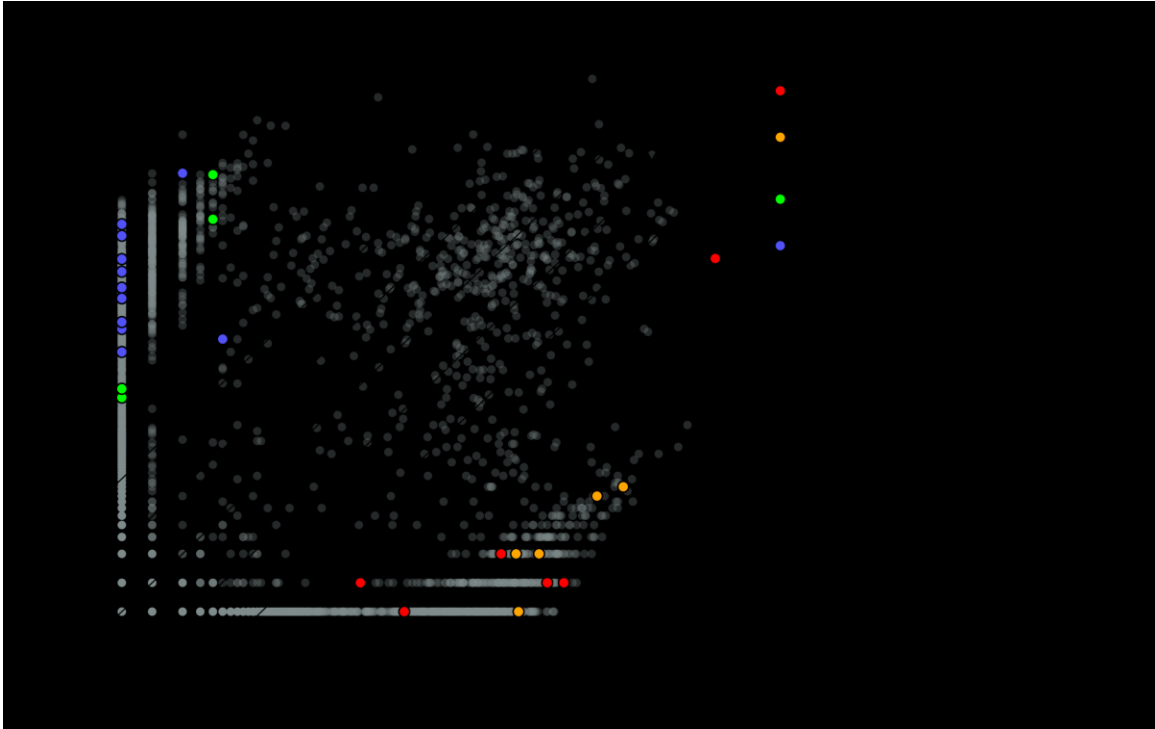


Figure 3-4: Abundance of fragments in potential activator or repressor libraries

A distinct fragment is defined as a pair of start and stop sites in the genome. Plotted are the abundances of all in frame fragments from either the potential activator or potential repressor libraries. Each dot represents a distinct fragment, and each dot is plotted with limited opacity. As a result, the areas with many distinct fragments at that level are much darker than regions where there is limited overplotting. These darkest regions occur along the axes, demonstrating that most fragments are functionally depleted in one library versus the other. Diagonal lines on the graph indicate 10- or 100-fold enriched in one library versus the other as written. Highlighted are fragments corresponding to a few genes of interest (*pat1*, red; *ymr295c*, orange; *cdc19*, green; *ypr204w*, blue) that will be discussed in the following section.

Examples of select identified regulators

From our catalog of over ten thousand potential regulatory fragments, we selected those genes showing the highest representation or enrichment in one library for validation. We picked two highly enriched genes out of the repressor library: *pat1*, which had the most distinct fragments (Figure 3-5A), and *ymr295c*, which had the largest fold enrichment in

one library (Figure 3-5B). Multiple distinct fragments were recovered with many unique sequencing reads for both proteins. Pat1 is a post-transcriptional repressor that links deadenylation and decapping through scaffolding interactions between deadenylation and decapping proteins (Haas et al. 2010, Ozgur et al. 2010, Standart and Marnef 2012, Sharif et al. 2013). Pat1 has several protein domains: N-terminal domain (NTD), Proline-rich domain (P-rich), Mid domain, and the IV and V domains alternatively called the C-terminal domain (Standart and Marnef 2012, Figure 5A). In tethering assays, the N-terminal domains and Pro-rich domains are necessary and sufficient to drive transcript decay and reporter repression (Haas et al. 2010, Ozgur et al. 2010). The other domains (Mid, IV, V) are necessary for rescuing full repression in vivo and may act to stimulate and localize Pat1, but are not necessary for inducing repression via tethering (Haas et al. 2010). The fragments we recover from our library all localize in the NTD and P-rich domains, consistent with their direct repression of transcripts (Figure 3-5A). The repressor with the most enriched fragments, Ymr295c, is a protein of unknown function that associates with the ribosome (Fleischer et al. 2006) and localizes to the bud and cell periphery (Tkach et al. 2012). The majority of the fragments occur within the N-terminus of the protein, with one fragment in the C-terminal region (Figure 3-5B). To verify the fragments we recovered from the screen, we cloned the sequence represented by the overlap of all fragments from Pat1 or the full-length Ymr295c into the λ N tether (Figure 3-5A, 3-5B). As expected, Pat1 has a strong repressive effect regardless of the reporter used (Figure 3-5C). In contrast to our expectations, full-length Ymr295c has a slight activating effect (Figure 3-5C). Since Ymr295c's fragments occur either in multiple locations in the N-terminal domain or in the C-terminal domain alone, we may be recovering activity in the screen that is lost in the full-length protein. Alternatively, Ymr295c may be acting as a post-transcriptional regulator dependent on additional cellular signals, such as Puf3 which acts as either an activator or repressor dependent

on the metabolic context (Lee and Tu 2015). To address if only a portion of Ymr295c alone can act as a repressor, we are cloning fragments of *ymr295c* that relate to either just the N- or C-terminus or single fragments. Overall, however, our system is capable of recovering post-transcriptional activity of a known and well-characterized regulator, Pat1, with domain-level information.



Figure 3-5: Two potential repressors from the post-transcriptional regulator screen

We verified two repressors identified in our screen. **(A)** Pat1 had the most total distinct fragments mapping to it. Prior work defined the domain structure of Pat1 (light blue). Specifically, the N-terminal domain (NTD) and Proline-rich (P-rich) domain are necessary and sufficient for stimulating mRNA decay, whereas the Mid, IV, and V domains act as platforms for interacting with other decay proteins *in vivo* (Haas et al. 2010, Ozgur et al. 2010). Our library hits (red) all align to these NTD and P-rich domains, consistent with the previous literature regarding their activity. We cloned a union of these fragments for verification of the screen (teal). **(B)** Ymr295c fragments exhibited the highest degree of enrichment between libraries. Ymr295c is a protein of unknown function, and as such, lacks domain annotations. The majority of our fragments (orange) map to the N-terminus, with one fragment in the C-terminus. As the protein is short, we cloned the entirety of the protein for verification. **(C)** Verification of the library hits was done through cloning either a union of library hits (Frag-Pat1) or the entire protein (Ymr295c) upstream of a λN tether. Pat1-λN shows a robust repression of both reporter-hairpin fusions. Full-length Ymr295c surprisingly acts as an activator, highlighting the possibility that our screen could pick up non-biological activity (as in, not within the context of the full protein)

activity. To address the possibility that sub-regions of Ymr295c could act as repressors, we are cloning smaller portions of Ymr295c.

We applied the same analysis for the potential repressor library to the potential activator library with two highly enriched genes: *ypr204w*, which had the most distinct fragments (Figure 3-6A), and *cdc19*, which had the highest degree of enrichment between libraries (Figure 3-6B). The CDC19 protein is the major pyruvate kinase for yeast (Ciriacy and Breitenbach 1979; Burke et al. 1983). The domain structure and regulation of Cdc19 activity is heavily studied (Jurica et al. 1998, Fenton and Blair 2002, Xu et al. 2012). However, prior to a very recent study from Beckman et al. (2016), Cdc19 was not annotated as an RBP. As Cdc19 lacks either known RBDs or RNA function, Beckman et al. (2016) term it an enigma RBP (enigmRBP). We recover domains in our positive regulatory library that overlap with the B and the C domains (Figure 6A), both of which are unstructured in the absence of fructose-1,6-bisphosphate (FBP; Jurica et al. 1998). The positive regulator with the most aligned fragments, Ypr204w, is a putative helicase encoded in the telomeric Y' element. Two early studies on this protein family predicted a potential helicase (Figure 6B, grey) in the N-terminus (Yamada et al. 1998; Shiratori et al. 1999). However, none of our library fragments aligned to this region. Active regulatory fragments align to a highly repetitive region. The disorder prediction software, GlobPlot (Figure 3-6B, Linding et al. 2003) confirmed that our library hits all occurred within regions of low-complexity and predicted disorder (Figure 6B, light blue and goldenrod respectively), consistent with the importance of disordered regions in regulating RNA biology (reviewed in Calabretta and Richard, 2015). Moreover, when sections of protein corresponding to the library hits were cloned into a tether construct, they yielded small but repeatable increases in target fluorescence (Figure 3-6C). Together, these confirmed

activator hits demonstrate that our system is capable of detecting positive regulators of post-transcriptional regulation.

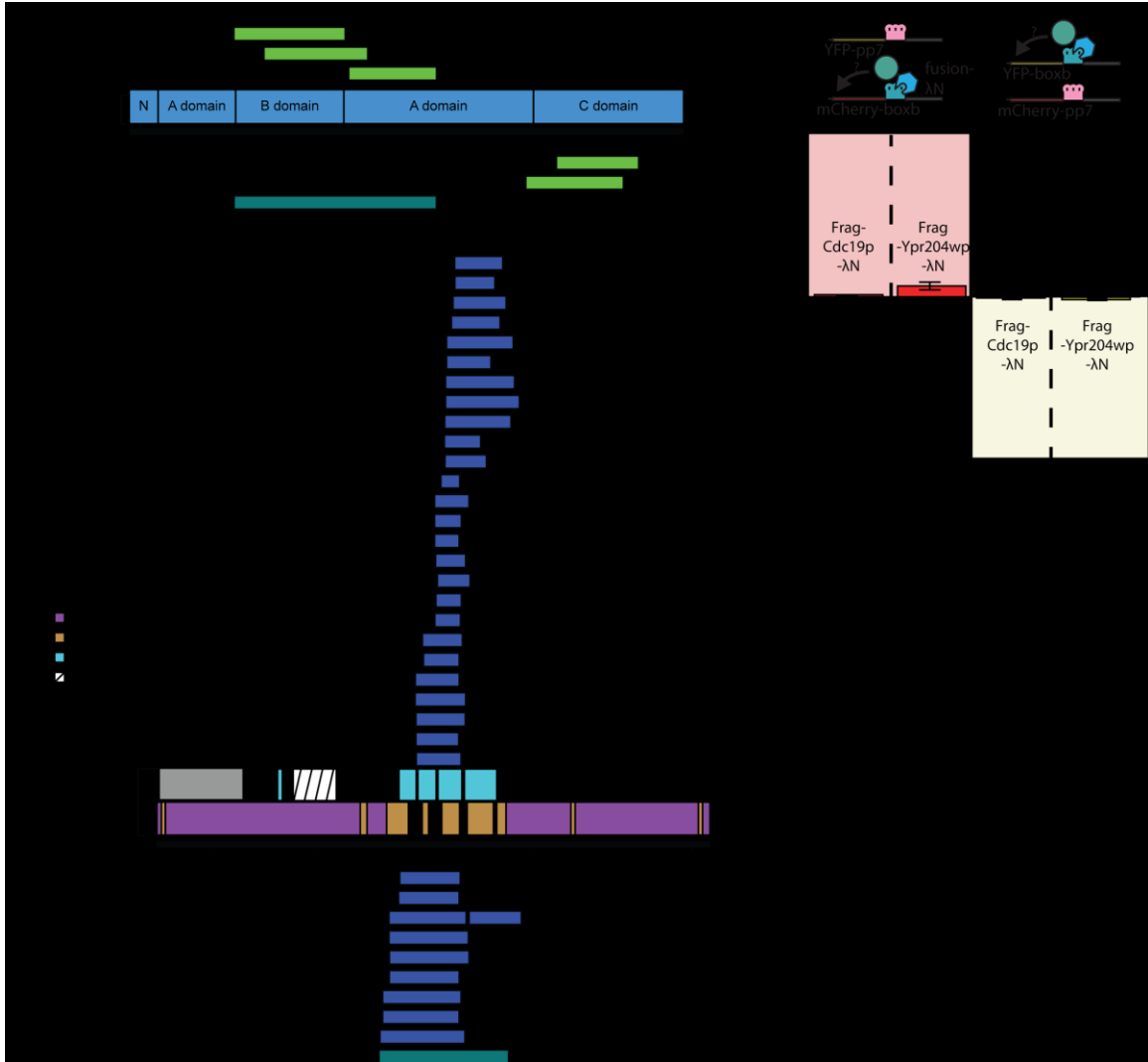


Figure 3-6: Two potential activators from the post-transcriptional regulator screen

We selected two hits from our post-transcriptional regulator screen for verification. **(A)** The gene *cdc19* had the most enriched fragments (light green). Cdc19 is the major yeast pyruvate kinase, and as such has well-annotated domains (light blue). Cdc19 has a split catalytic barrel (A domains), a barrel cap (B domain), and a regulator C-terminus (C domain; Jurica et al. 1998). Our fragments align with the B and C domains, both of which are predicted to be unfolded prior to the binding of the allosteric regulator, fructose-1,6-biphosphate (Jurica et al. 1998). **(B)** YPR204W had the most hits of any protein in the activator library (dark blue). Ypr204wp is predicted to have a helicase domain (grey; Yamada et al. 1998, Shiratori et al. 1999).

However, none of our fragments aligned to this region. We used the disorder prediction software GlobPlot to analyze this sequence (Linding et al. 2003). The fragments we recover from our screen (dark blue) overlap with the regions predicted by GlobPlot to have low complexity (light blue) and disorder (goldenrod), as with CDC19. **(C)** We subcloned a region of CDC19p (A, teal) and YPR204W (B, teal) and tested their ability to act as positive regulators. They both give a small, but reproducible increase in target fluorescence. As with the initial tether system verification studies, Yfp has a smaller dynamic range for detecting increased fluorescence.

Discussion

Post-transcriptional regulation is essential for gene expression in a diverse number of biological processes. In recent years, technological improvements in mass spectrometry, RNA sequencing, and the advent of ribosome profiling have highlighted that post-transcriptional regulation is of genome-wide importance rather than a unique feature of a few discrete genes. We anticipate that many post-transcriptional events occur through the interaction of transcripts with RNA binding proteins (RBPs). Several studies have cataloged the ensemble of proteins bound to the transcriptome (Hogan et al. 2008, Tsvetanova et al. 2010, Mitchell et al. 2012, Baltz et al. 2012, Kwon et al. 2013; Beckmann et al. 2015). RBPs recovered from these studies often lack RBDs or annotated RNA-associated functions. To expand our understanding of functional roles of RBPs, we developed a methodology for functionally characterizing post-transcriptional regulators on a genome-wide level. This methodology incorporates our approach for generating genome-wide protein expression libraries with an *in vivo* tethering system to characterize post-transcriptional regulators through changes in target to non-target fluorescence (Figure 3-7).



Figure 3-7: Summary of screen

We sought to probe post-transcriptional activity on a genome level. (i.) To accomplish this, we fragmented information-rich yeast genomic DNA. (ii.) This population was recombined into expression vectors in dual fluorescent tethering yeast. (iii.) Populations of in-frame fragments were selected in custom large-volume turbidostats using marker-based selection (as described in chapter 2). (iv.) Resulting populations of in-frame tether fusions have the potential to act as non-regulators, repressors, or activators. (v.) Using fluorescence activated cell sorting (FACS), we can separate these three different populations. These sorted cells are used to generate libraries of fragments that correspond to potential activators or repressors. Using a single color and tether combination (mCherry-boxb, fragment- λ N), we can recover tens of thousands of potential activators and repressors. We verified three out of four initial hits from this screen, and we are actively developing a statistical framework for identifying more functional post-transcriptional regulators.

Our design allows us to easily swap the targeted reporter or the tether, accounting for reporter- or tether-specific variation. From a single run of this screen using one set of reporters and a single tether, we identified over ten thousand fragments that can be classified as activators or repressors based on their enrichment in the resulting libraries. In this study, we describe four proteins identified in our screen, two activators and two repressors. For three out of the four hits, we demonstrate the validity of the screen in generating and selecting for bona fide post-transcriptional regulators. We are currently developing a statistical framework for parsing fragment- and protein-level significance of the remaining several thousand regulatory fragments as either activators or repressors. Using the information from this statistical analysis, we can confidently characterize

proteins for their functional roles, extending our understanding from identifying proteins to functionally describing them.

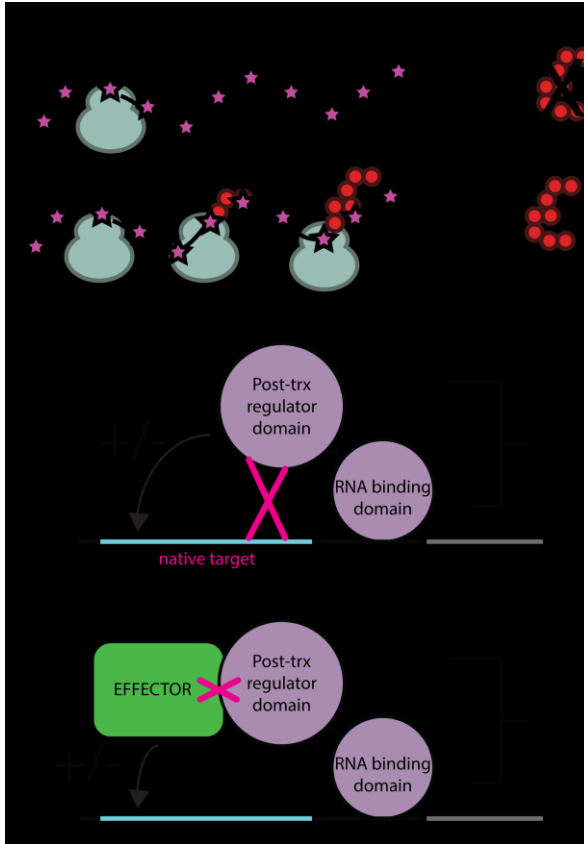


Figure 3-8: Future directions

Our screen provides information about the regulatory outcome (positive or negative) that a potential regulator exerts on a reporter transcript. It does not currently define the level of regulation, the native targets, or the molecular mechanism for regulation. We are developing several tools to address these questions. (A) We are optimizing flow cytometry fluorescence in situ hybridization (FLOW-FISH) to simultaneously assay reporter transcript abundance and protein abundance to distinguish protein stability from translational efficiency. (B) Simultaneously, we're developing tandem affinity purification (TAP) tagging systems to perform native target identification using high-throughput sequencing cross-linking immunoprecipitation (HITS-CLIP). (C) We can apply this same purification system to assay associated effector proteins using cross-linking immunoprecipitation mass spectrometry (CLIP-MS).

In the future, we plan to follow up on our current validated hits (Pat1, Cdc19, and Ypr204w) and other statistically defined regulators using molecular techniques to define both native targets and the mechanism for regulation. Post-transcriptional regulation can occur at either the level of modulating stability or translational efficiency. To differentiate these two levels, we are optimizing a protocol for flow cytometry fluorescent in situ hybridization (FLOW-FISH) to simultaneously measure abundance of target and non-target transcripts relative to their protein abundances (Figure 3-8A). In addition, we want to know the biological role of these reporters: what transcripts do they target, and how do they enact their regulation? To identify native transcript targets, we are developing a tandem affinity purification tagging (TAP-tagging) system using high-throughput sequencing cross-linking immunoprecipitation (HITS-CLIP; Figure 3-8B). This same TAP-tagging system can be used to identify co-acting protein partners by using cross-linking immunoprecipitation mass spectrometry (CLIP-MS; Figure 3-8C). These approaches will provide systems-level information about the role and mechanism of proteins identified through our screen. In addition, the systems described here are easily adapted for use in other biological systems such as cultured mammalian cells. The combination of our screen with downstream identification of targets and co-actors will advance our understanding of the extent and mechanism of post-transcriptional regulation.

Chapter 4:

Upstream open reading frames (uORFs) as important post-transcriptional regulatory elements

Introduction and preface

Prior to this chapter, we focused on post-transcriptional regulation as mediated by trans-acting RBPs. However, there are several well studied examples where cis-acting features play a role in determining the post-transcriptional regulation of a transcript. Of particular interest are the upstream open reading frames (uORFs). A uORF is defined as an in-frame start and stop codon that occur within the 5' untranslated region (UTR) or transcript leader (TL; Arribere & Gilbert, 2013) of a transcript. These ORFs can occur strictly within the TL or can overlap with the coding sequence (CDS) of the transcript itself. Until the last decade, uORFs were thought to reside on only a handful of transcripts; however, many uORFs have been identified since. The following piece was written by the thesis writer, Anna McGeachy, as a review of the current understanding of uORFs for The EMBO Journal in tandem with the publication of a paper (Johnstone et al. 2016). The article highlights the nature of uORFs, our expectations of their translation, and how these cis-acting factors are thought to regulate post-transcriptional fates.

The review was originally published online in The EMBO Journal on February 19, 2016. In line with The EMBO Journal and Wiley Publishing's permission policies, authors are allowed to reprint in whole or part the text in a new publication by the authors as long as appropriate citation is given ([http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1460-](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1460-)

2075/homepage/Permissions.html). For the purpose of this thesis, we the authors (Anna McGeachy and Nicholas Ingolia) have reproduced the manuscript in its original form as it provides a useful review by the thesis writer, Anna McGeachy, of the state of understanding uORF regulation. The only change made to the manuscript is to rename the figures to match the scheme (chapter-figure number) of the thesis. An online version of the original text can be found via EMBO's archiving ([embj.201693946](https://doi.org/10.15252/embj.201693946)) or through the article DOI ([10.15252/embj.201693946](https://doi.org/10.15252/embj.201693946)).

Publication

Starting too soon: upstream reading frames repress downstream translation

Upstream open reading frames (uORFs) are known to regulate a few specific transcripts, and recent computational and experimental studies have suggested candidate uORF regulation across the genome. In this issue, Johnstone et al (2016) use ribosome profiling to identify translated uORFs and measure their effects on downstream translation. Furthermore, they show that regulatory uORFs are conserved across species and subject to selective constraint. Recognizing the potential of uORFs in regulating translation expands our understanding of the dynamic regulation of gene expression.

Differences in the translation level of distinct mRNAs play an important role in controlling the production of the encoded protein. Changes in translation drive posttranscriptional gene expression programs that play critical roles in diverse processes ranging from cellular stress responses to memory formation. Despite the importance of differential

translation, we have a limited understanding of the cis-acting mRNA features that determine the stability or translation state of an mRNA (Brar et al, 2012; Arribere & Gilbert, 2013; Calvo et al, 2009). Johnstone et al (2016) now provide evidence for a global impact of short upstream open reading frames (uORFs) on translation.

In eukaryotes, the small subunit of the ribosome typically scans from the 5' end of an mRNA until it recognizes a start codon. The transcript leader sequence (TLS), also known as the 5' untranslated region, scanned by the ribosome can thus modulate translation initiation in order to control protein synthesis. In particular, the presence of upstream start codons in the transcript leader can recruit scanning ribosomes to an alternate reading frame, reducing the fraction that reach the start codon for the major protein (Johnstone et al, 2016; Arribere & Gilbert, 2013). Regions that show no overlap with the CDS are termed upstream open reading frames (uORFs), while those that start prior to CDS AUG but finish within the CDS are called overlapping open reading frames (oORFs; Johnstone et al, 2016). Calvo et al (2009) provided direct evidence that the presence of a uORF represses downstream CDS translation in the context of reporters bearing endogenous, uORF-containing TLSs. Furthermore, uORFs can specify dynamic regulation. In the stress-specific transcription factors ATF4 (in mammals) and GCN4 (in yeast), upstream initiation sites repress protein production under normal cellular conditions (Sonenberg & Hinnebusch, 2009). Under stress, global translation levels decrease; however, ATF4 and GCN4 synthesis is paradoxically upregulated in an upstream initiation site-dependent manner.

While it is possible to computationally identify upstream TLS AUG codons, these potential initiation sites may not actually be engaged by ribosomes. The advent of ribosome profiling now allows direct interrogation of ribosome occupancy in TLSs (Figure 4-1A) and has identified translated uORFs (Ingolia et al, 2009; Arribere & Gilbert, 2013) that in some cases produce peptides detectable by mass spectrometry (Bazzini et al, 2014; Saghatelian & Couso, 2015). Ribosome profiling provides information about ribosome occupancy at single nucleotide precision, and several metrics now exist to calculate the likelihood of a ribosome profiling signal corresponding to active translation of a given reading frame (Ingolia et al, 2014; Chew et al, 2013). Bazzini et al (2014) utilized the triplet periodicity from such nucleotide level profiles to classify translation status on a genome wide level across zebrafish developmental stages, using a metric termed ORF score (Figure 5-1B). Johnstone et al (2016) now ask how empirically verified uORF translation acts across the transcriptome.

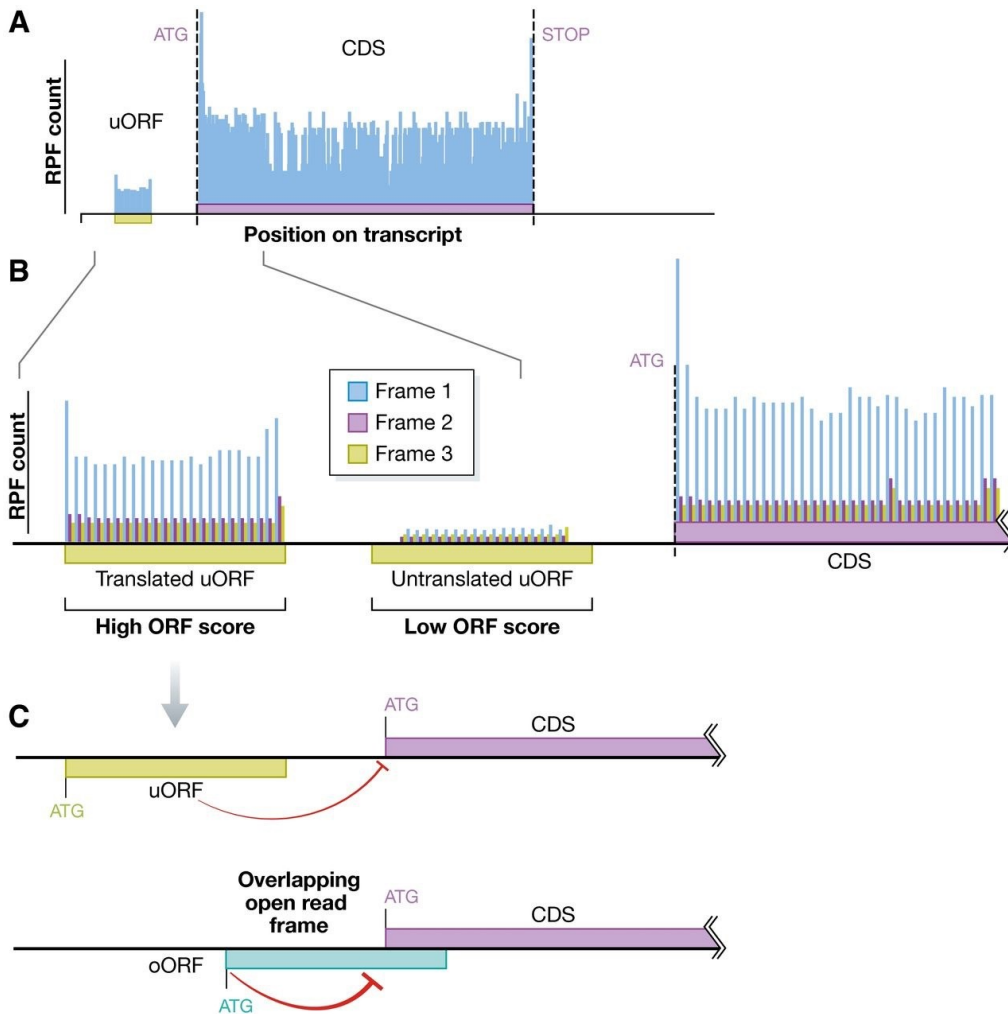


Figure 4- 1: Ribosome profiling detects uORF translation and downstream repression

(A) Ribosome profiling maps the location of ribosomes across a transcript using deep sequencing of ribosome protected fragments (RPFs) (Ingolia et al, 2009). In addition to the major protein coding sequence (CDS), RPFs can be seen in upstream open reading frames (uORF). **(B)** RPFs offer nucleotide level resolution that can map to any of the three potential reading frames. Bazzini et al (2014) utilized this triplet periodicity to generate an ORFscore. This ORFscore can be used to classify actively translated regions. In this issue, Johnstone et al (2016) apply ORFscore classification to separate computationally defined uORFs by their translation status. **(C)** Johnstone et al (2016) find that confidently translated uORFs correlate with repression of the downstream CDS translation. Moreover, overlapping open reading frames (oORFs) act as stronger repressors of CDS translation.

Using ORF score, Johnstone et al (2016) identify over a thousand zebrafish transcripts with confidently translated uORFs, including several developmentally important genes. Previous studies suggest uORFs generally repress the downstream CDS (Calvo et al, 2009; Arribere & Gilbert, 2013), and using matched RNA sequencing, Johnstone et al (2016) calculate a score of translation efficiency (TE) for the major transcript product as well. They show that on a genomewide scale, the presence of a translated uORF correlates with decreased downstream CDS translation and transcript stability relative to non-uORF-containing transcripts (Figure 5-1C). Moreover, the repression of the downstream CDS is greater in the presence of overlap with the upstream region (oORFs) or stronger uORF translation.

uORFs provide functionally important repression, mediated by the titration of initiating ribosomes away from downstream genes. Johnstone et al (2016) demonstrate conservation of the presence of uORFs and selective constraints on sequence features that confer strong translation initiation. Furthermore, these sequence features positively correlate with the strength of downstream translation repression across zebrafish genes. In contrast, the peptide sequences of regulatory uORFs do not show selective constraint as seen in truly functional micropeptides (Saghatelian & Couso, 2015). This suggests that the functional importance of uORFs is dependent on features that drive uORF translation rather than the specific peptide produced.

The next challenge for the field will be understanding the mechanism by which these TLS reading frames can provide dynamic regulation. In the case of ATF4 and GCN4, translation of the CDS is only accomplished under cases of cellular stress, and loss of

this stress-induced translation is linked to diabetes and neurodegenerative disorders (Sidrauski et al, 2015; Sonenberg & Hinnebusch, 2009). This regulation depends on phosphorylation of translation initiation factor eIF2 α , which is thought to promote the bypass of repressive uORFs.

Because eIF2 α is required for essentially all translation initiation events, this acts as a global method of regulation. Dynamic regulation of specific transcripts can result from the interaction between repressive uORFs and sequence-specific RNA-binding proteins, as seen in *Drosophila* SXL2 control of msl2 translation (Medenbach et al, 2011). Direct detection and measurement of uORF translation by ribosome profiling, as demonstrated by Johnstone et al (2016), promises further insights into the dynamic process of uORF-mediated translational control.

Chapter 5:

Genome-wide interrogation of the Unfolded Protein Response (UPR) and the role of Integrative Stress Response Inhibitor (ISRIB)

Introduction and preface

A key component of cellular homeostasis is the management of the translation state of the cell. Accumulation of unfolded or unmodified proteins in the lumen of the endoplasmic reticulum is an indication of cellular stress. Detection of ER-protein stress triggers a pathway known as the unfolded protein response (UPR). Disregulation of the UPR is responsible for several human diseases, including neurodegenerative disorders (Sonenberg & Hinnebusch, 2009; Pavitt and Ron 2012, Scheper and Hoozemans, 2015), diabetes (Harding et al. 2001, Scheuner et al. 2005, Wek et al. 2006; Walter and Ron 2011), and cancer (Walter and Ron 2011; Vandewynckel et al. 2013). As such, understanding how the UPR is activated and how it controls stress responsive gene expression programs will impact many aspects of human health. The Walter lab identified a small molecule, integrated stress response inhibitor (ISRIB), that alleviated UPR-driven phenotypes (Siduraski et al. 2013). In collaboration with the Walter lab, the thesis writer and advisor, Anna McGeachy and Nicholas Ingolia respectively, performed a global analysis of the UPR.

The UPR consists of three pathways (Figure 5-0A), each signaled by a distinct ER-membrane embedded protein: activating transcription factor 6 (ATF6), inositol-requiring enzyme 1 (Ire1), and Pkr-like ER-kinase (Perk). Two of these factors, Atf6 and Ire1, are associated with transcriptional pathways. Briefly Atf6, is a transcription factor that is

sequestered to the ER until detection of ER-stress (Haze et al. 1999). Upon activation of the UPR, Atf6 is trafficked to the Golgi and subsequently cleaved by S1P and S2P; the N-terminal fragment is then free to translocate into the nucleus and enact transcriptional control (Figure 5-0A-i; Haze et al. 1999; Ye et al. 2000; Walter and Ron 2011). Ire1 has an inter-luminal domain for the detection of ER stress, upon which Ire1 oligomerizes to activate intrinsic RNase and kinase activities (Chen and Brandizzi 2013). While Ire1 can act as a general cellular RNase to degrade ER-associated mRNAs, Ire1 specifically drives the removal of a latent intron in *hac1* (fungal) or *xbp1* (metazoan) mRNA to promote cytosolic splicing of the mature transcript (Figure 5-0A-ii; Yoshida et al. 2001; Walter and Ron 2011; Chen and Brandizzi 2013). Xbp1/Hac1 then act as a transcription factor to induce expression of genes to mitigate and adapt to stress (Figure 5-0A-ii; Walter and Ron 2011; Chen and Brandizzi 2013).

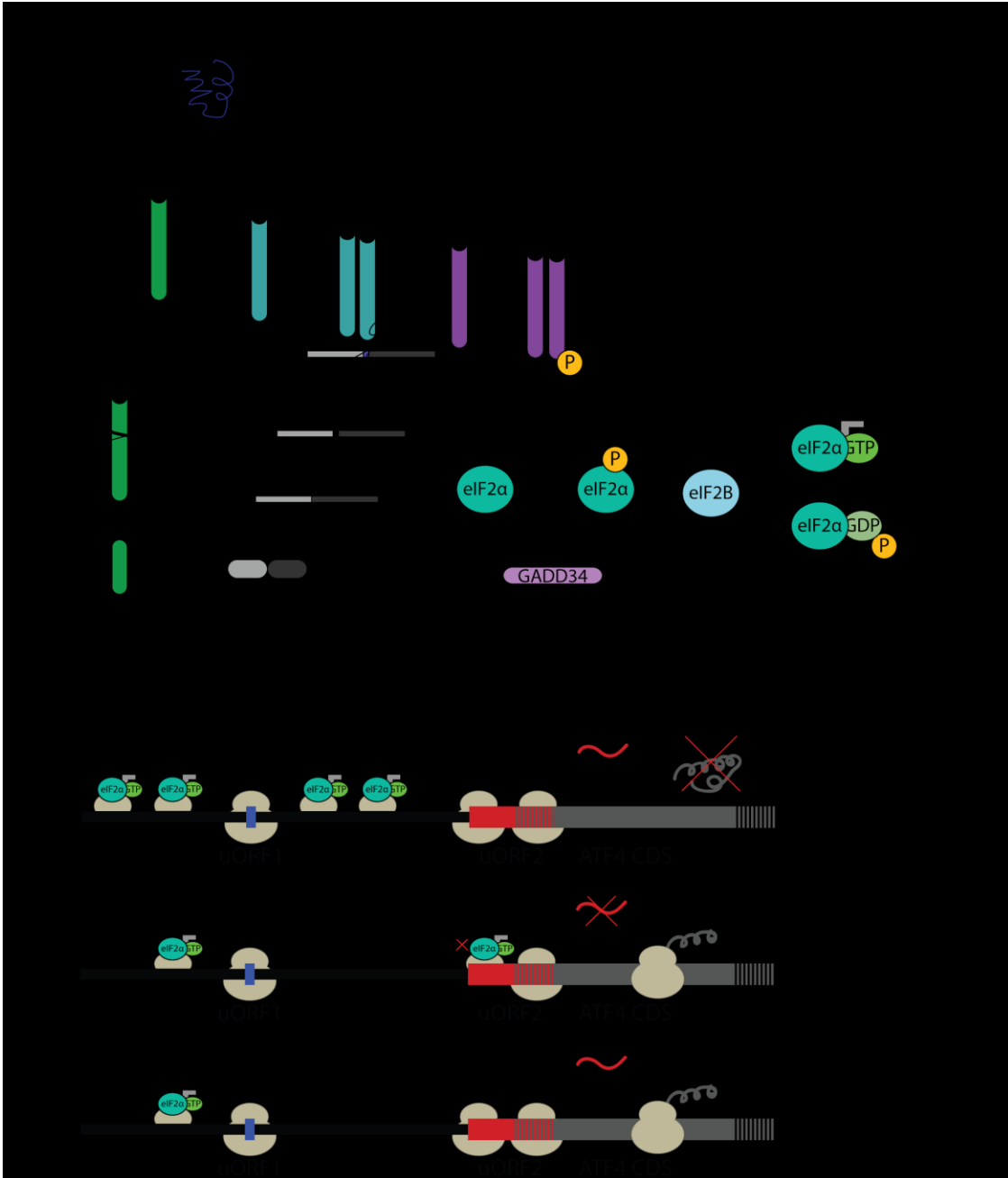


Figure 5-0: The unfolded protein response (UPR) drives adaptation through three pathways, one specific to translation

(A) The UPR has three sensor and downstream pathways: Atf6 (A-i), Ire1 (A-ii), and Perk (A-iii). (i.) Upon detection of ER stress, Atf6 is translocated to the Golgi and cleaved to liberate the N-terminal transcription factor portion. (ii.) Ire1 remains ER localized and forms higher-order oligomers. These oligomers have RNase activity that drive the removal of a latent intron in *xbp1* mRNA. The remaining exons are spliced and translated to produce active Xbp1 transcription factor protein. (iii.) Perk activates through auto-dimerization

and auto-phosphorylation. The activated Perk's major substrate is eIF2 α . The ternary complex of eIF2 α -GTP-Met-tRNA is essential for each round of translation initiation. Following initiation, the tRNA is released and GTP is hydrolyzed. Typically, eIF2 α is recharged for subsequent rounds of initiation through the action of eIF2B. Phosphorylated-eIF2 α cannot be recycled. As such, this drives a global block of translation; however, there are specific transcripts that experience paradoxical upregulation. **(B)** The model for this paradoxical regulation is dependent on the presence of upstream open reading frames (uORFs) in the specifically upregulated transcripts, here shown for Atf4. (i.) Under normal cellular conditions, abundant ternary eIF2 α complex allows for re-initiation at the uORFs, but not the major transcript coding sequence (CDS). (ii.) Under the current model, when eIF2 α becomes limited, the delay in re-initiation permits translation initiation at the CDS, but not the downstream uORFs. (iii.) In contrast to the standing model, our ribosome profiling data shows evidence of translation at all uORFs as well as the CDS. While the mechanism for permissive re-initiation at both uORF and CDS is unclear, it provides evidence against the current model of strictly eIF2 α -limited driven regulation.

The third UPR pathway, Perk (Figure 6-0A-iii, is implicated in translational regulation. Similar to IRE1, Perk has an inter-luminal domain for the detection of ER-stress (Ron and Walter 2007). Upon detection of stress, Perk becomes activated through self-association and -phosphorylation (Figure 5-0A-iii). The primary substrate for the activated kinase domain is eukaryotic initiation factor 2 subunit alpha (eIF2 α); however, there is evidence for additional substrates (Cullinan et al. 2003; Sidrauski et al. 2015, below). The substrate eIF2 α is a key factor in mediating cellular translation response to stress as it is the convergence point for the four-pronged integrative stress response (ISR, Figure 5-0A; Harding et al. 2003, Wek et al. 2006, Roux and Topisirovic 2012, Chen 2014, Castilho et al. 2014). The importance of eIF2 α in mediating cellular stress responses derives from the central role of eIF2 α in translation initiation. eIF2 α associates with charged methionine-tRNA (Figure 5-0A-iii, grey) and GTP (Figure 5-0A-iii, bright green) and as such is essential for each round of translation initiation. Start codon selection drives hydrolysis of GTP and release of eIF2 α (Kapp and Lorsch, 2004;

Sonenberg and Hinnebusch, 2009; Jackson et al. 2010). The complex is recharged and reloaded for subsequent rounds of translation initiation through interactions with eukaryotic initiation factor 2 subunit B (eIF2B, Figure 5-0A-iii, light blue). Phosphorylation of eIF2 α at serine 51 by PERK or any of the ISR kinases blocks recharging by eIF2B. As a result, the pool of initiation-competent eIF2 α is drastically limited in stressed cells, blocking translation initiation at a global level (Kapp and Lorsch, 2004; Sonenberg and Hinnebusch, 2009; Jackson et al. 2010). However, under these competent eIF2 α -limited conditions, translation of several transcripts is paradoxically upregulated, including ATF4 (Harding et al. 2003), ATF5, and CHOP (Figure 5-0A-iii; Jackson et al. 2010). Our analysis of the global UPR response recovered the expected targets as well as several novel translational upregulation targets (Siduraski et al. 2015, below).

The paradoxical upregulation of these specific transcripts is thought to be driven by an upstream open reading frame (uORF; Figure 5-0B). uORFs are defined as an in-frame start and stop codon contained in the 5' untranslated region (UTR) or transcript leader (TL; Arribere and Gilbert, 2013) and have been the topic of several recent global studies (Brar et al. 2012, Bazzini et al. 2014, Johnstone et al. 2016). Under normal cellular conditions, the translation-competent eIF2 α preinitiation complex (PIC) is thought to be abundant across TLs (Figure 5-0B-i). Given this abundance, the ribosome is thought to start translation at the first optimal start codon (e.g. uORF1) and be capable of re-initiating at downstream ORFs (Figure 5-0B-i). When eIF2 α is limited through phosphorylation-blockage of recycling, the current model predicts that the limited abundance allows for delayed scanning and reinitiation at the major transcript coding sequence (CDS) only (Figure 5-0B-ii; Wek et al. 2006; Sonenberg and Hinnebusch 2009; Jackson et al. 2010; Pavitt and Ron 2012). However, based on the ribosome profiling data we collected under UPR conditions, initiation occurs at all uORFs as well

as the major CDS (Figure 5-0B-iii; Sidrauski et al. 2015, below). While the mechanism for permissive translation at the CDS is unclear, our data provides evidence against the current model and motivation for further downstream studies.

The following was originally published online in eLife on February 9, 2016. In line with eLife's open access policies, the article is governed by the Creative Commons license (<http://creativecommons.org/licenses/by/4.0/>). As such, there are no limitations on reprinting the work in whole or in part as long as proper attribution is given. For the purpose of this thesis, we (Anna McGeachy and Nicholas Ingolia, two of the four authors)

have edited the manuscript to focus on the contribution of the thesis writer, Anna McGeachy. Additional content that was included in the article at the time of publication but was not performed in part or whole by Anna McGeachy will be added in the appendix under a section called "Results from Collaborators". In addition, the manuscript has been edited to rename the figures to match the scheme (chapter-figure number) of the thesis. An online version of the original text can be found via eLife's archiving (eLife 2015;4:e05033) or through the article DOI (<http://dx.doi.org/10.7554/eLife.05033>).

Publication

The small molecule ISRIB reverses the effects of eIF2 α phosphorylation on translation and stress granule assembly

Carmela Sidrauski^{1*}, Anna M McGeachy², Nicholas T Ingolia², Peter Walter^{1*}

¹ Department of Biochemistry and Biophysics, Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, United States; ² Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, United States

**For correspondence: carmelas@me.com (CS); Peter. Walter@ucsf.edu (PW)*

Competing interests: The authors declare that no competing interests exist.

Received: 07 October 2014

Accepted: 04 February 2015

Published: 26 February 2015

Reviewing editor: David Ron, University of Cambridge, United Kingdom

Cite as eLife 2015;4:e05033

Abstract

Previously, we identified ISRIB as a potent inhibitor of the integrated stress response (ISR) and showed that ISRIB makes cells resistant to the effects of eIF2 α phosphorylation and enhances long-term memory in rodents (Sidrauski et al., 2013). Here, we show by genome-wide in vivo ribosome profiling that translation of a restricted subset of mRNAs is induced upon ISR activation. ISRIB substantially reversed the translational effects elicited by phosphorylation of eIF2 α and induced no major changes in translation or mRNA levels in unstressed cells. eIF2 α phosphorylation-induced stress granule (SG) formation was blocked by ISRIB. Strikingly, ISRIB addition to stressed cells with pre-formed SGs induced their rapid disassembly, liberating mRNAs into the actively translating pool. Restoration of mRNA translation and modulation of SG dynamics may

be an effective treatment of neurodegenerative diseases characterized by eIF2 α phosphorylation, SG formation, and cognitive loss.

Introduction

Diverse cellular conditions activate an integrated stress response (ISR) that rapidly reduces overall protein synthesis while sustaining or enhancing translation of specific transcripts whose products support adaptive stress responses. The ISR is mediated by diverse stress-sensing kinases that converge on a common target, serine 51 in eukaryotic translation initiation factor alpha (eIF2 α) eliciting both global and gene-specific translational effects (Harding et al., 2003; Wek et al., 2006). Mammalian genomes encode four eIF2 α kinases that drive this response: PKR-like endoplasmic reticulum (ER) kinase (PERK) is activated by the accumulation of unfolded polypeptides in the lumen of the ER, general control non-derepressible 2 (GCN2) kinase by amino acid starvation and UV light, protein kinase RNA-activated (PKR) by viral infection, and heme-regulated eIF2 α kinase (HRI) by heme deficiency and redox stress. The eIF2 α kinase PERK is also part of the unfolded protein response (UPR). This intracellular stress signaling network is comprised of three ER-localized transmembrane sensors, IRE1, ATF6, and PERK, which initiate unique signaling cascades upon sensing an increase in unfolded proteins in the ER lumen (Walter and Ron, 2011; Pavitt and Ron, 2012).

The common mediator of the ISR, eIF2 α , is a subunit of an essential translation initiation factor conserved throughout eukaryotes and archaea. The heterotrimeric eIF2 complex (composed of subunits α , β and γ) brings initiator methionyl tRNA (Met-tRNA_i) to translation initiation complexes and mediates start codon recognition. It binds GTP along

with Met-tRNA_i to form a ternary complex (eIF2- GTP-Met-tRNA_i) that assembles, along with the 40S ribosomal subunit and several other initiation factors, into the 43S pre-initiation complex (PIC). The 43S PIC is recruited to the 5' methylguanine cap of an mRNA and scans the 5'UTR for an AUG initiation codon (Hinnebusch and Lorsch, 2012). Start site codon recognition triggers GTP hydrolysis and phosphate release, which is followed by release of eIF2 from the 40S subunit, allowing binding of the 60S ribosomal subunit to join. After these events, the elongation phase of protein synthesis ensues. To engage in a new round of initiation, the newly released eIF2 complex has to be re-loaded with GTP, a reaction catalyzed by its dedicated guanine nucleotide exchange factor (GEF), the heteropentameric eukaryotic initiation factor 2B (eIF2B). Phosphorylation of eIF2 α does not directly affect its function in the PIC, but rather inhibits eIF2B, thereby depleting ternary complex and reducing translation initiation (Krishnamoorthy et al., 2001). eIF2B complex is limiting in cells, present in lower abundance than eIF2; a small amount of phospho-eIF2 α therefore acts as a competitive inhibitor with dramatic effects on eIF2B activity. When eIF2B is inhibited and ternary complex is unavailable, the rate of translation initiation decreases.

Unimpaired elongation in the face of reduced initiation allows translating ribosomes to run off of their mRNAs, generating naked mRNAs that can then bind to RNA-binding proteins (RBPs) and form messenger ribonucleoproteins, which can further assemble into stress granules (SGs). These cytoplasmic, non-membrane bounded organelles contain translationally stalled and silent mRNAs, 40S ribosomal subunits and their associated pre-initiation factors and RBPs; these RBPs facilitate the nucleation and reversible aggregation of SGs through reversible, low-affinity protein–protein interactions mediated by their low complexity domains (Buchan and Parker, 2009; Kedersha and Anderson, 2009; Kato et al., 2012).

Paradoxically, under conditions of reduced ternary complex formation and protein synthesis, a group of mRNAs is translationally up-regulated. These mRNAs contain short upstream open reading frames (uORFs) in their 5' UTRs, which are required for their ISR-responsive translational control (Hinnebusch, 2005; Jackson et al., 2010). These target transcripts include mammalian ATF4 (a cAMP response element binding transcription factor) and CHOP (a pro-apoptotic transcription factor) (Harding et al., 2000; Vattam and Wek, 2004; Palam et al., 2011). ATF4 regulates the transcription of many genes involved in metabolism and nutrient uptake and thus is a major regulator of the transcriptional changes that ensue upon eIF2 α phosphorylation and ISR induction (Harding et al., 2003). Although activation of this cellular program can initially mitigate the stress and confer cytoprotection, persistent and severe stress and its associated reduction in protein synthesis and CHOP activation lead to apoptosis (Tabas and Ron, 2011; Lu et al., 2014).

In animals, the ISR has been implicated in diverse processes ranging from the regulation of insulin production to learning and memory. These effects were studied first using genetics by generating knockout mice lacking individual eIF2 α kinases as well as a knock-in of the non-phosphorylatable allele eIF2 α S51A (Eif2s1S51A). Homozygous loss of eIF2 α phosphorylation leads to perinatal death but heterozygous eIF2 α +/S51A animals, which have reduced levels of eIF2 α phosphorylation, grow into healthy adults showing phenotypes that demonstrate the importance of translation initiation in establishment of long-term memories (Scheuner et al., 2001). Behavioral tests demonstrated that PKR $-/-$, GCN2 $-/-$ and eIF2 α +/S51A animals display enhanced memory consolidation in learning paradigms of light training (Costa-Mattioli et al., 2005, 2007; Zhu et al., 2011). Pharmacological modulation of eIF2 α phosphorylation

represented an important advance, allowing easier discrimination between developmental and acute effects of ISR reduction and circumventing the lethal phenotype of homozygous eIF2 α S51A/S51A. Recent work identified small molecules that modulate the ISR pathway at distinct steps: (1) kinase inhibitors that target PERK or PKR (Jammi et al., 2003; Atkins et al., 2013); (2) an activator of HRI (Chen et al., 2011); (3) salubrinal, an inhibitor of eIF2 α phosphatases (Boyce et al., 2005); and (4) ISRIB (Sidrauski et al., 2013). By a yet unknown mechanism, ISRIB blunts the effects of eIF2 α phosphorylation in cells and thus represents the first bona fide ISR inhibitor acting downstream of all eIF2 α kinases.

Here, we show that ISRIB reverses comprehensively and specifically the effects of eIF2 α phosphorylation. By profiling the genome-wide translational program downstream of the ISR, we present the application of ribosome profiling to the ISR in mammalian cells, which allowed us to identify and quantify the translational changes that take place upon its induction and ISRIB treatment. Moreover, live cell imaging revealed that ISRIB addition can trigger a remarkably fast dissolution of phospho-eIF2 α -dependent SGs in stressed cells, restoring translation.

Results

Ribosome profiling of ER stress in mammalian cells

We used ribosome profiling to characterize translational changes induced by ER stress. Deep sequencing of ribosome-protected mRNA fragments provides global, quantitative measurements of translation and reveals the precise location of ribosomes on each mRNA (Ingolia et al., 2009, 2011). We triggered the UPR in HEK293T cells by treating

them with tunicamycin (Tm), a toxin that blocks N-linked glycosylation of ER-resident proteins. We chose to analyze an early time point (1 hr) in order to focus on translational changes preceding the extensive transcriptional induction that takes place upon activation of the three branches of the UPR (for a time course of UPR induction, see Figure 5-3—figure supplement 5-1 in 10.7554/eLife.00498). After 1 hr of Tm or mock treatment, we added cycloheximide (CHX) to arrest translating ribosomes, lysed the cells, and digested the extract with nuclease to degrade mRNAs not protected by ribosomes. In parallel, we isolated total mRNAs to monitor any changes in mRNA levels. Ribosome profiling data revealed a discrete subset of mRNAs that were translationally up- or down-regulated more than twofold after UPR induction (Figure 5-1A, above or below box) as seen by changes in abundance of ribosome-protected fragments (RPF) ('Ribo-Seq', y-axis) without corresponding changes in mRNA levels ('mRNA-Seq', x-axis). Data points representing statistically significant changes in expression between Tm-treated and untreated ('UT') samples are highlighted in black.

Consistent with the well-established presence of regulatory uORFs in their 5'-UTRs, this genome-wide analysis identified four previously extensively studied mRNAs that displayed significant translational upregulation: ATF4, ATF5, CHOP, and GADD34, (Figure 5-1A, colored pink). The mRNAs encoding the closely paralogous transcription factors ATF4 and ATF5 are known translational targets of the ISR (Lu et al., 2004; Vatter and Wek, 2004; Zhou et al., 2008). They contain two uORFs (the second one overlapping with the coding sequence [CDS]) that govern their enhanced translational efficiency. The mRNAs encoding the pro-apoptotic transcription factor CHOP and the regulatory subunit of the eIF2 α phosphatase GADD34 were also significantly upregulated at the translational level. Although both CHOP and GADD34 are also known transcriptional targets of ATF4, we did not detect significant induction of their mRNAs at

this early time point (Figure 5-1A, lack of displacement along x-axis) indicating that at the time point chosen our analysis exclusively reports on translational effects. CHOP and GADD34 mRNAs also contain uORFs that allow for translational regulation upon eIF2 α phosphorylation (Lee et al., 2009; Palam et al., 2011).

We identified a total of 78 mRNAs whose translation changed significantly and substantially (more than twofold) upon ER stress in HEK293T cells (listed in Figure 5-1—source data 5-2A). GO term analysis revealed the involvement of these genes in diverse functions and several encode for proteins with entirely unknown functions. Besides the four known ISR translational targets described above, six mRNAs in the list contain previously mapped uORFs as validated by ribosome profiling in the presence of a translation initiation inhibitor to mark initiation sites (Figure 5-1A, colored green and Figure 5-1—source data 5-2B) (Lee et al., 2012). Whereas 5% of the non-significantly changed genes in the Tm sample contain previously identified AUG-initiated uORFs, 14% of genes in the list of ISR-translational targets contain uORFs, indicating a significant enrichment ($p < 0.003$, chi-squared test with Yates correction).

A seventh and novel uORF-containing translational target of the ISR encodes SLC35A4, a putative nucleotide-sugar transporter (Song, 2013). It was recently shown that the longest uORF of SLC35A4 is indeed translated because peptides corresponding to the encoded polypeptide were found in a whole proteome mass spectrometry study (Kim et al., 2014). Analysis of RPFs in the uORFs of the SLC35A4 and ATF4 mRNAs revealed significant ribosome density, further confirming that these regulatory uORFs are normally translated (Figure 5-1—figure supplement 5-1). Due to the reduced mRNA expression levels of ATF5, CHOP, and GADD34 in the absence of stress or at early time-points of UPR activation, we did not analyze the RPFs or mRNA reads at specific locations along

these genes, as the read numbers were low.

Interestingly, there was a slight reduction in translation of mRNAs encoding ribosomal proteins and translation elongation factors (Figure 5-1—figure supplement 5-2, panel A). The translation of this functionally related class of ~100 abundant mRNAs, which have a 5' terminal oligopyrimidine (5' TOP) motif, is controlled by the activity of the mTOR kinase (Meyuh, 2000; Tang et al., 2001; Hsieh et al., 2012; Thoreen et al., 2012). The concerted changes that we observed in their translation upon UPR activation suggest that ER stress and eIF2 α phosphorylation affects 5' TOP translation in HEK293T cells.

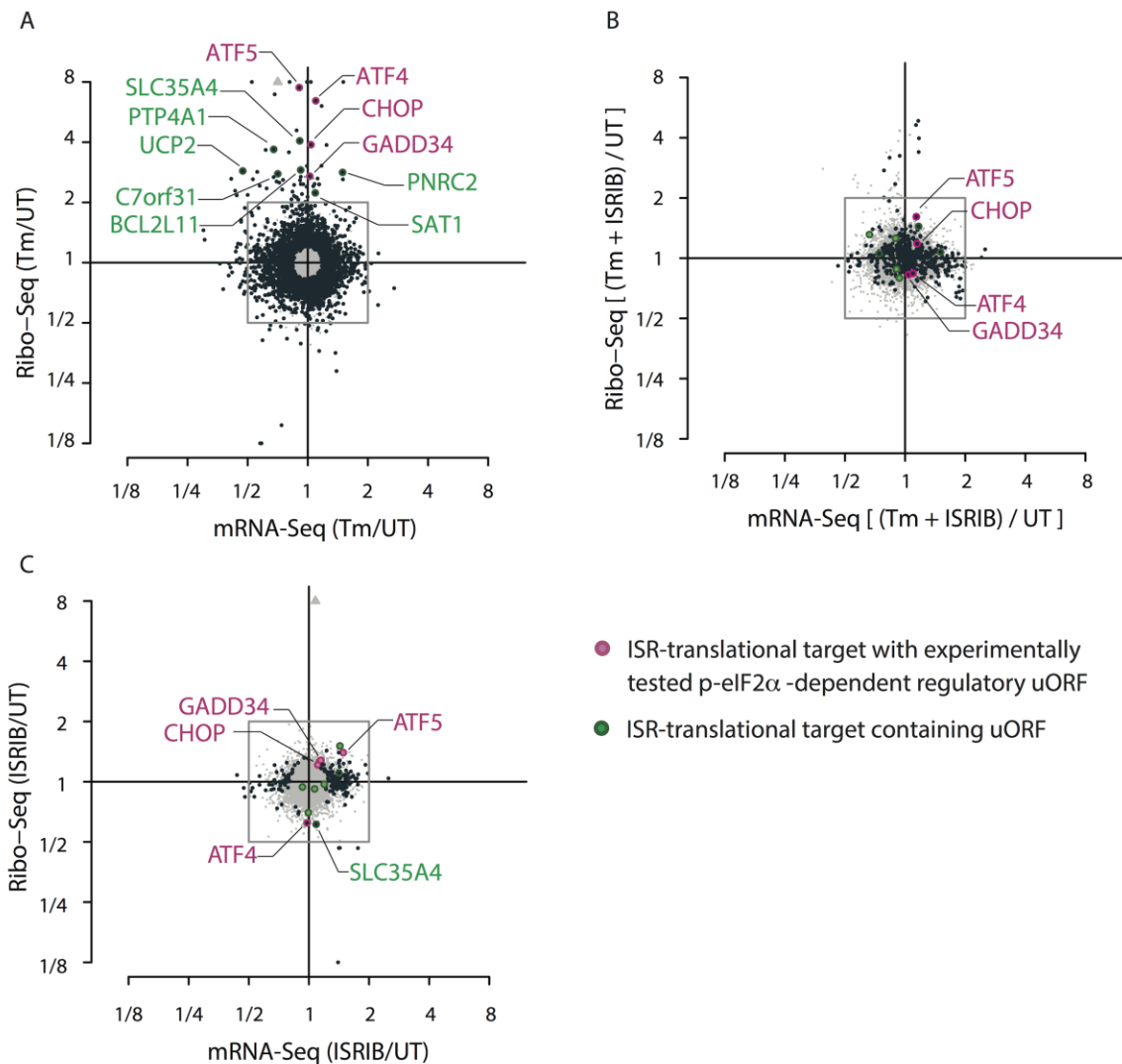


Figure 5-1. Translational regulation upon ER stress in mammalian cells.

(A) Translational and mRNA changes in HEK293T cells upon ER stress. HEK293T cells were treated with or without 1 µg/ml of Tm for 1 hr. The y-axis represents fold changes in ribosome-protected fragments (Ribo-Seq) between Tm-treated and control samples. The x-axis represents fold changes in mRNA levels (mRNA-Seq) between Tm-treated and control samples. Data points reflecting significant changes (FDR-corrected p-value < 0.1) between Tm treated and untreated ('UT') samples are shown in black and non-significant changes are shown in light grey. Note that genes with significant changes (black circles) are numerous in Tm-treated cells and thus the cloud of genes with no significant changes (grey circles) is mostly hidden in the background. Genes with substantially enhanced RFPs and uORFs that are known to be phospho-eIF2α-dependently regulated are labeled pink. ISR-translational targets that contain previously identified uORFs are labeled in green. Triangles denote genes that fall beyond the axis range. The genes inside the grey box are those that change less than twofold in RPF or mRNA reads. Figure 5-1—source data 5-2A contains a list of all genes that change more than twofold in RPFs during Tm induction (FDR-corrected p-value < 0.1, corresponding to black circles above and below the box). **(B)** Translational and mRNA changes in cells co-treated with Tm and ISRIB. HEK293T cells were treated with or without 1 µg/ml of Tm and 200 nM ISRIB for 1 hr. The y-axis represents fold changes in ribosome-protected fragments (Ribo-Seq) between Tm + ISRIB-treated and control samples. The x-axis represents fold changes in mRNA levels (mRNA-Seq) between Tm + ISRIB-treated and control samples. Genes that significantly change when ISRIB co-administration modulates the effects of Tm treatment are shown in black (FDR-corrected p-value < 0.1). Figure 5-1—source data 5-2C contains a list of all genes that change more than twofold in RPFs during Tm and ISRIB treatment (FDR-corrected p-value < 0.1). The identity of the ISR-translational targets that contain previously identified uORFs (labeled in green) was not included in this panel as they all collapsed to the center of the plot. **(C)** Translational and mRNA changes in ISRIB-treated cells. HEK293T cells were treated with or without 200 nM ISRIB for 1 hr. The y-axis represents fold changes in ribosome-protected fragments (Ribo-Seq) between ISRIB-treated and control samples. The x-axis represents fold changes in mRNA levels (mRNA-Seq) between ISRIB-treated and control samples. Data points reflecting significant changes (FDR-corrected p-value < 0.1) between ISRIB-treated and untreated ('UT') samples are shown in black and non-significant changes are shown in light grey. Figure 5-1—source data 5-2D contains a list of all genes that change more than twofold in RPFs during ISRIB treatment (FDR-corrected p-value < 0.1, corresponding to black circles outside of the box). ATF4 and SLC35A4 (labeled in this panel) showed reduced translational efficiency upon addition of ISRIB. Two biological replicates were analyzed per condition. Number of reads

aligned to the genome and ORFs for all samples are Figure 5-1. found in Figure 5-1—source data 5-2E. Correlation plots for the replicates for each condition are found in Figure 5-1—figure supplement 5-3. mRNA abundance for all ORFs mapped are found in Figure 5-1—figure supplement 5-4. Read counts for all conditions and each individual transcript are found in Figure 5-1—source data 5-1. The Ribo-seq and mRNA-seq data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE65778.

DOI: 10.7554/eLife.05033.002

The following source data and figure supplements are available for figure 1 (placed at the end of this chapter):

Source data 5-1. Read counts for all conditions and each individual transcript.

DOI: 10.7554/eLife.05033.003

Source data 5-2. Source data for Figure 5-1.

DOI: 10.7554/eLife.05033.004

Figure supplement 5-1. Ribosome and mRNA densities in the 5'UTR of ATF4 and SLC35A4.

DOI: 10.7554/eLife.05033.005

Figure supplement 5-2. Translational regulation of mTOR targets upon ER-stress.

DOI: 10.7554/eLife.05033.006

Figure supplement 5-3. Correlation plots for duplicate ribosome profiling experiments.

DOI: 10.7554/eLife.05033.007

Figure supplement 5-4. Mean mRNA abundance of all genes mapped.

DOI: 10.7554/eLife.05033.008

ISRIB substantially reduced the translational effects elicited by stress and eIF2 α phosphorylation

To study the translational effects of the small molecule ISRIB at a genome-wide level, we analyzed changes in RPFs and mRNA levels after addition of the drug to both ER-stressed and unstressed cells. As seen in Figure 5-1B, ISRIB comprehensively blocked

the translational changes that take place upon ER-stress. A large number of genes with a significant change in expression upon stress collapsed to the center of the plot with ISRIB and Tm co-treatment (Figure 6-1B, highlighted in black). Importantly, ISRIB abolished the induction of the known phospho-eIF2 α -dependent translational targets (Figure 5-1B, colored pink) and the seven ISR-translational targets with previously identified uORFs (Figure 5-1B, colored green). The mRNAs that remained translationally induced in the presence of ISRIB are listed in Figure 5-1—source data 5-2C. In addition, ISRIB reversed the reduction in translation of mTOR target mRNAs upon ER stress (Figure 5-1—figure supplement 5-2, panel B).

Importantly, ISRIB treatment alone did not have general effects on translation in non-stressed cells, as revealed by the lack of substantial changes in RPFs in most cellular mRNAs, nor did it cause any significant changes in mRNA levels (Figure 5-1C) and mTOR target expression (Figure 5-1—figure supplement 5-2, panel C). In the absence of ER stress, ISRIB-treated cells behaved like untreated cells with the exception of a reduction in the basal level of translation of ATF4 and SLC35A4 mRNAs and a few additional mRNAs (Figure 5-1C and Figure 5-1—source data 5-2C). Taken together, these data strongly support the notion that ISRIB does not have global effects on translation, transcription, or mRNA stability in non-stressed cells and underscores its remarkable ability to counteract selectively the translational changes elicited by eIF2 α phosphorylation in stressed cells.

1.

2. *[Two sections originally featured in the published version have been removed here:*

ISRIB prevents formation of stress granules exclusively triggered by eIF2 α phosphorylation

ISRIB triggers rapid disassembly of stress granules and restores translation

While they contribute to the overall story, the work down within was not performed by the thesis writer. These sections can be found in the appendix in a section titled “Results from Collaborators”.]

Discussion

ISRIB is the first reported antagonist of the ISR that blocks signaling downstream of all eIF2 α kinases. It was shown to have good pharmacokinetic properties and brain penetration, making it a useful tool to study the systemic effects of acute inhibition of the pathway. We showed that ISRIB administration enhances long-term memory in rodents (Sidrauski et al., 2013). More recently, we showed by electrical recordings in brain slices that by preventing AMPAR down-regulation in the post-synaptic neuron, ISRIB blocks mGluR-mediated long-term depression (LTD), an effect that is dependent on eIF2 α phosphorylation (Di Prisco et al., 2014). Comprehensive analyses of the cellular effects and kinetics of action of ISRIB are critical for interpretation of its in vivo effects and assessment of its therapeutic potential.

Our translational and transcriptional profiling confirmed that ISRIB treatment of ER-stressed cells substantially and comprehensively blocks the translational effects of eIF2 α phosphorylation. ISRIB blocked SG formation that was triggered by eIF2 α phosphorylation but did not abolish their assembly upon eIF4A inhibition; eIF4A inhibitors do not cause eIF2 α phosphorylation and can induce SGs in eIF2 α S51A/S51A cells (Mazroui et al., 2006; Mokaš et al., 2009). These data further support the notion that ISRIB solely inhibits cellular events that are a consequence of eIF2 α

phosphorylation. In agreement with these observations, we previously showed by polyribosome sedimentation analysis that ISRIB does not reverse bulk translational down-regulation triggered by inhibition of CAP-dependent initiation (Sidrauski et al., 2013). Moreover, ISRIB treatment alone did not induce overall changes in translation or mRNA levels. Taken together these data demonstrate that ISRIB is a pharmacological agent that acutely and specifically blocks the ISR and is thus an invaluable tool for in vivo studies.

Translational regulation upon ISR induction

The method of ribosome profiling can monitor in vivo translation comprehensively and with nucleotide resolution (Ingolia et al., 2009). We used this method to monitor translation of all cellular mRNAs upon ISR activation. We found that a limited set of mRNAs is preferentially translated in a substantial manner upon a reduction in ternary complex assembly. Although previous large-scale analyses have revealed that almost 45% of all 5' UTRs have at least one upstream uORF (Calvo et al., 2009; Ingolia et al., 2011), our data revealed that only a few of these mRNAs contain uORFs with regulatory properties that significantly enhance translation of their downstream coding sequences upon eIF2 α phosphorylation. The canonical ISR translational targets, ATF4, ATF5, CHOP, and GADD34 mRNAs were significantly induced upon 1 hr treatment with the ER-stressor tunicamycin. The stress-induced, uORF-mediated regulation of GCN4 translation in yeast established the paradigm for this mode of regulation (Dever et al., 1995; Grant et al., 1995). As in mammalian cells, GCN2 is activated in amino acid-starved yeast by the accumulation of uncharged tRNAs, catalyzing eIF2 α phosphorylation. The transcript encoding GCN4, a bZIP transcription factor with

homology to mammalian ATF4, has four uORFs that modulate translation of its coding sequence upon stress. GCN4 induction is thought to occur via a re-scanning mechanism that allows 40S ribosomal subunits to remain mRNA-bound after completing the translation of short reading frames and subsequently reinitiate in the downstream coding sequence after reloading with ternary complex (Hinnebusch, 2005). The select mRNAs that are translationally upregulated in mammalian cells have uORFs that vary in number, length, and distance from the coding sequences. As was observed for GCN4, the uORF2 of ATF4 mRNA showed ribosome density, supporting the notion that it is translated under normal growth conditions (Ingo-lia et al., 2009). Whether the same mechanism of rescanning is utilized by all these mRNAs is not known but, like ATF4, their regulation depends on their uORFs (Vattem and Wek, 2004; Zhou et al., 2008; Lee et al., 2009; Palam et al., 2011).

SLC35A4 is a novel translational target of the ISR. Ribosome profiling of HEK293T cells upon arsenite treatment, a potent inducer of eIF2 α phosphorylation, also revealed the increased synthesis of SLC35A4 (Andreev et al., 2015). It belongs to a large family of nucleotide sugar transporters (NSTs) that are highly conserved transmembrane antiporters localized to the ER or Golgi apparatus (Song, 2013). The role of SLC35A4 in cells is unknown but it may function as the elusive ER-localized UDP-glucose transporter. This hypothesis is particularly attractive in the context of our data because unfolded ER proteins, which trigger the ISR, are continuously de- and re-glucosylated on their N-glycans using UDP-glucose as the glucose donor. Proteins with monoglucosylated N-glycans bind calnexin or calreticulin which promote protein folding. Translational induction of SLC35A4 may thus quickly enhance UDP-glucose transport into the ER lumen upon the accumulation of unfolded proteins in order to promote this

pro-folding pathway.

Ribosome profiling upon activation of the UPR uncovered additional mRNAs induced upon eIF2 α phosphorylation (our data, Reid et al., 2014; Andreev et al., 2015). Several of these mRNAs encode for proteins with entirely unknown functions and the remaining targets are involved in a wide range of cellular processes. Whether these ISR-induced translational targets are similarly regulated by the presence of uORFs in their 5' UTRs remains to be determined with the construction of synthetic translational reporters. ISRIB blocked their differential translation, suggesting that these changes were due to phospho-eIF2 α . There may be additional transcripts that are synthesized later during ISR activation, downstream of the early transcription factor targets such as ATF4, as well as tissue-specific mRNAs that are controlled by phospho-eIF2 α . For example, OPHN1 is a neuron-specific mRNA containing uORFs that is translationally upregulated after mGluR engagement and eIF2 α phosphorylation and induces LTD. By blocking the effects of phospho-eIF2 α in cells, ISRIB also blocks mGluR-dependent LTD (Di Prisco et al., 2014). Ribosome profiling of glutamatergic neurons upon ISR induction may reveal additional transcripts whose translational control contributes to the molecular events underlying memory.

The ribosome profiling data presented here revealed that eIF2 α phosphorylation modestly, yet significantly, decreased translation of a large number of ribosomal proteins and elongation factors. Although the decrease in translation of ribosomal proteins and elongation factors upon eIF2 α phosphorylation is small in magnitude, its effects on bulk protein synthesis in the cell are significant as these represent a large number of highly expressed proteins. Translation of these mRNAs was previously shown to be under control of mTOR kinase, which regulates mRNA cap-binding factor eIF4E via

phosphorylation of inhibitory eIF4E-binding proteins, thereby adjusting protein synthesis in cells in response to the cell's energy and nutrient status (Ma and Blenis, 2009). In this way, mTOR preferentially regulates translation of a group of mRNAs characterized by 5' TOP motifs (Hsieh et al., 2012; Thoreen et al., 2012). Upon mTOR inhibition, translation of 5' TOP mRNAs is reduced and the expression of factors required for protein synthesis is diminished.

The observed effect that eIF2 α phosphorylation preferentially decreased translation of 5' TOP mRNAs could, in principle, be due to inhibition of mTOR in response to ER stress. However, ISRIB reversed the translational changes, indicating that they are likely to be downstream consequences of eIF2 α phosphorylation. Thus, if these translational changes do reflect altered mTOR activity, then the change in mTOR signaling must result from reduced translation mediated by eIF2 α phosphorylation. Alternatively, eIF2 α phosphorylation may lead to silencing of these mRNAs by recruiting them into SGs. The RNA binding proteins TIA-1 and TIAR, which are prominently SG-associated, were previously shown to bind to TOP mRNAs, leading to their translational downregulation upon amino acid starvation. This effect required both mTOR inhibition and GCN2 activation, the latter resulting in eIF2 phosphorylation (Damgaard and Lykke-Andersen, 2011). SGs also have been shown to recruit signaling molecules including upstream negative regulators of mTORC1 (raptor and DYRK3) and mTORC1 itself, and thus, SG formation may reduce their presence in the cytosol and impede translation of 5' TOP mRNAs (Thedieck et al., 2013; Wippich et al., 2013).

[One section originally featured in the published version have been removed here:

Stress granule dynamics and ISRIB

While they contribute to the overall story, the work down within was not performed by the thesis writer. These sections can be found in the appendix in a section titled “Results from Collaborators”.]

Materials and methods

Chemicals

Tunicamycin was obtained from Calbiochem EMB Bioscience. Thapsigargin, cycloheximide and sodium arsenite were obtained from Sigma–Aldrich. Hippuristanol and pateamine A were a kind gift from Jerry Pelletier. GSK797800 (PERK inhibitor) was obtained from TRC Inc. ISRIB (Sidrauski et al., 2013) and an inactive analog (754125) (Di Prisco et al., 2014) were synthesized in-house.

Cell culture

HEK293T, U2OS, and U2OS GFP-G3BP/Dcp1-RFP cells were maintained at 37 °C, 5% CO₂ in DMEM media supplemented with 10% FBS, L-glutamine and antibiotics (penicillin and streptomycin). U2OS cells stably expressing G3BP-GFP/Dcp1-RFP cells were a kind gift from Nancy Kedersha (Kedersha et al., 2008).

Isolation of ribosome footprints and RNA

HEK293T cells were treated with or without 1 µg/ml of tunicamycin, tunicamycin and ISRIB (200 nM), or ISRIB for 1 hr. Cycloheximide (CHX) (100 µg/ml) was added for 2 min, cells were washed with ice cold PBS (with 100 µg/ml of CHX) and lysed in 20 mM Tris pH = 7.4 (RT), 200 mM NaCl, 15 mM MgCl₂, 1 mM DTT, 8% glycerol, 100 µg/ml

CHX, 1% Triton and protease inhibitors (Roche complete EDTA-free). A syringe (25G5/8) was used to triturate cells, the lysate was clarified at 12,000 rpm for 10 min and half of the lysate was used for RNA extraction (Trizol, Invitrogen, Carlsbad, CA) and the other half was digested with RNase I (Ambion). The amount of RNase I and time of incubation was optimized for each sample based on the collapse of polyribosomes to the monosome peak as analyzed by analytical polyribosome gradients. The reaction was quenched with SUPERaseIn (Ambion, Life Technologies) and the digested lysate was then loaded on an 800 μ l sucrose cushion (1.7 g of sucrose was dissolved in 3.9 ml of lysis buffer without Triton) and centrifuged in a TLA100.2 rotor at 70,000 rpm for 4 hr. The pellet was resuspended in 10 mM Tris pH = 7 (RT), and RNA was extracted (phenol/chloroform).

Generation of sequencing libraries and data analysis

Sequencing libraries were generated as described in Ingolia et al., 2012. For data analysis, we used DESeq as described by Anders and Huber (2010). P-adj values (p-values) were calculated using the R command 'p.adjust' for multiple comparisons and the BH method (Benjamini and Hochberg, 1995) to correct for false discovery rate. The data in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE65778.

1.

2. *[Four sections originally featured in the published version have been removed here:*

3.

Immunofluorescence

4.

Live cell microscopy

Protein analysis

[35S]-methionine incorporation

While they contribute to the overall story, the work down within was not performed by the thesis writer. These sections can be found in the appendix in a section titled “Results from Collaborators”.]

Acknowledgements

We thank Margaret Elvekrog, Voytek Okreglak, Shelley Starck, and Jirka Peschek for editing the manuscript and the members of the Walter lab for helpful discussions.

Author contributions

CS, NTI, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; AMMG, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; PW, Analysis and interpretation of data, Drafting or revising the article

Supplement

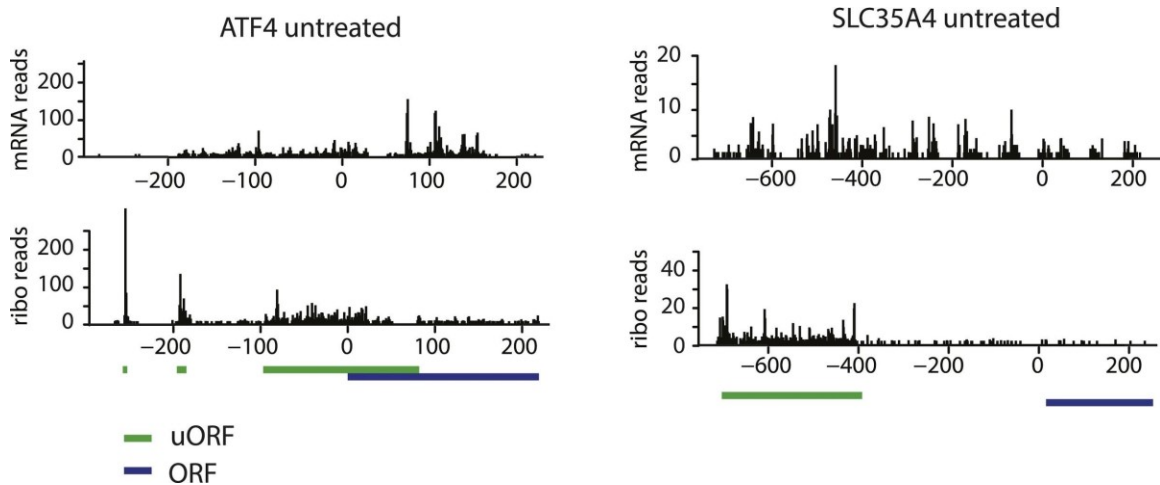


Figure supplement 5-1. Ribosome and mRNA densities in the 5'UTR of ATF4 and SLC35A4.

DOI: 10.7554/eLife.05033.005

mRNA reads (y-axis) are represented along the sequence of each gene (x-axis) in the upper panel.

Ribosome footprint (ribo) reads (y-axis) are represented along the sequence of each gene (x-axis) in the lower panel. The known and predicted uORFs are indicated along the sequence in green. The ORF is indicated along the sequence in blue.

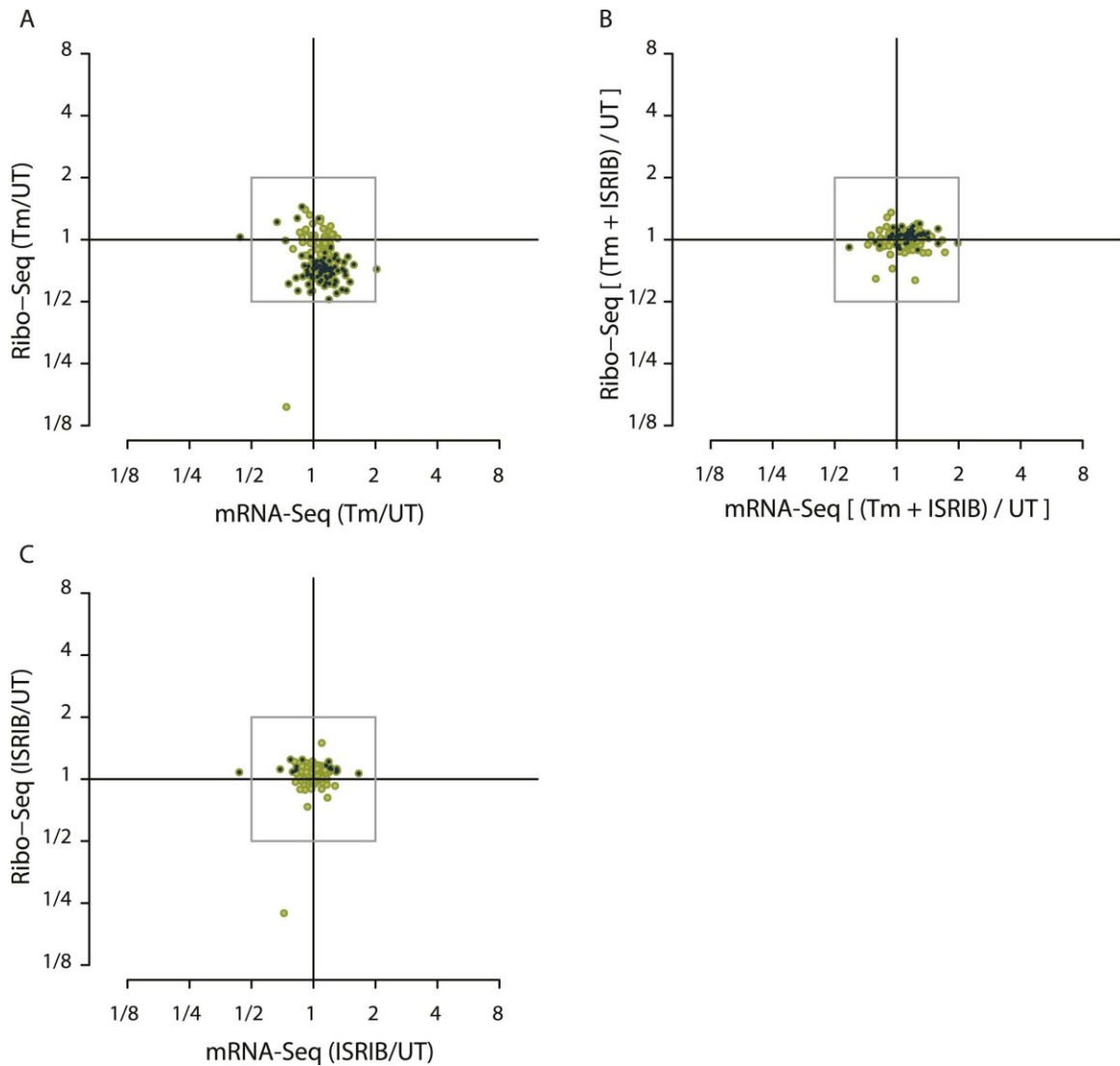


Figure supplement 5-2. Translational regulation of mTOR targets upon ER-stress.

DOI: 10.7554/eLife.05033.006

(A) Cells were treated with or without 1 μ g/ml of Tm for 1 hr. The y-axis represents fold changes in ribosome-protected fragments (Ribo-Seq) between Tm-treated and control samples. The x-axis represents

fold changes in mRNA levels (mRNA-Seq) between Tm-treated and control samples. Only mTOR translational targets are plotted (colored light green). Significant changes in mTOR genes (FDR-corrected p-value < 0.1) between Tm-treated and untreated (UT) are highlighted in black. (B) Cells were treated with or without 1 µg/ml of Tm and 200 nM ISRIB for 1 hr. The y-axis represents fold changes in ribosome-protected fragments (Ribo-Seq) between Tm + ISRIB-treated and control samples. The x-axis represents fold changes in mRNA levels (mRNA-Seq) between Tm + ISRIB-treated and control samples. Only mTOR translational targets are plotted (colored light green). Genes that significantly change when ISRIB co-administration modulates the effects of Tm treatment are shown in black (FDR-corrected p-value < 0.1). (C) Cells were treated with or without 200 nM ISRIB for 1 hr. The y-axis represents fold changes in ribosome-protected fragments (Ribo-Seq) between ISRIB-treated and control samples. The x-axis represents fold changes in mRNA levels (mRNA-Seq) between ISRIB-treated and control samples. Only mTOR translational targets are plotted (colored light green). Significant changes in mTOR genes (FDR-corrected p-value < 0.1) between ISRIB-treated and untreated (UT) are highlighted in black.

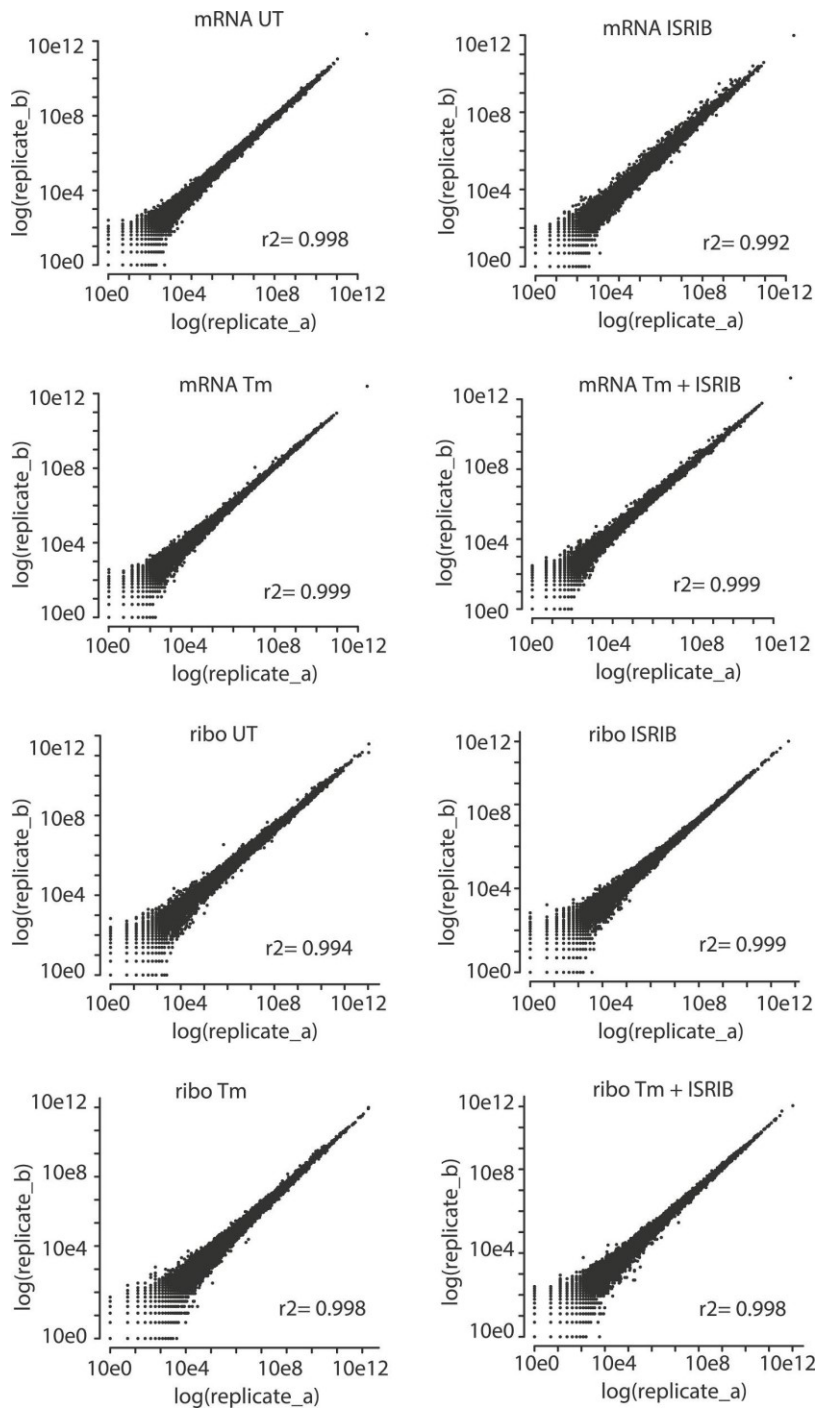


Figure supplement 5-3. Correlation plots for duplicate ribosome profiling experiments.

DOI: 10.7554/eLife.05033.007

The detected ribosome (ribo) or mRNA density is plotted for each gene in each experimental condition (untreated [UT], tunicamycin [Tm], tunicamycin + ISRIB [Tm + ISRIB] and ISRIB). Correlation coefficients (r²) between replicates (A and B) in each condition are indicated in the lower right for each panel.

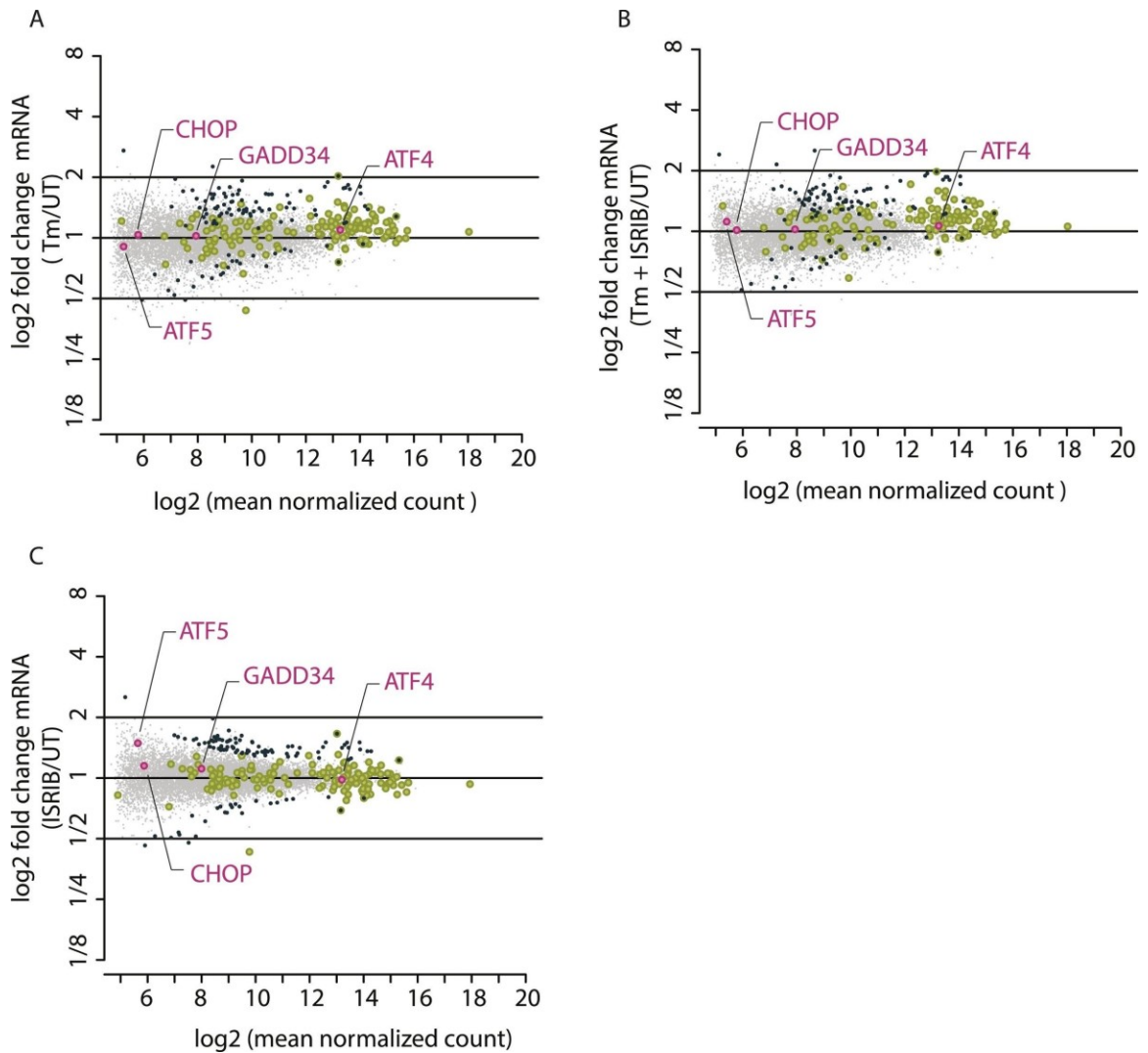


Figure supplement 5-4. Mean mRNA abundance of all genes mapped.

DOI: 10.7554/eLife.05033.008

The x-axis represents log₂ (mean normalized count) for each mRNA mapped and the y-axis represents log₂ fold changes in mRNA abundance in the different experimental conditions: Tm (panel A), Tm + ISRIB (panel B), and ISRIB (panel C). Previously known phospho-eIF2 α -dependent ISR translational targets are highlighted in pink.

Chapter 6:

Conclusion

Introduction

We have expanded our understanding of post-transcriptional regulation by identifying post-transcriptional activators and repressors out of a genome-scale protein expression library. Our work complements the RBP literature with functional description for putative regulators of translation and mRNA stability. We have employed ribosome profiling to capture the full range of ISR-mediated genes, thereby identifying additional uORFs and providing nucleotide-level resolution data on uORF occupancy. We demonstrate the impact of uORFs on downstream CDS regulation and regulatory potential for cis-acting features. Taken together, we have provided new insights into the cis- and trans-acting factors of underlying post-transcriptional regulation at a genomic scale.

Results

One major contribution of this thesis is a validated approach for genome-scale functional characterization of post-transcriptional regulators. In order to achieve this goal, we developed a system and equipment for creating and screening a genome-wide protein expression library. Building these genome-wide libraries provides a comprehensive and less biased approach to capturing potential regulators, which is important because sequence and functional annotations alone are not sufficient to predict RBPs. We then developed an assay to identify trans-acting regulators using a protein-RNA tethering system with fluorescent reporters. We combined this assay with our genome-wide

expression libraries to identify tens of thousands of unique potential regulatory fragments as well as verify the approach on a handful of candidate genes. Our catalogue of regulators extends our understanding of post-transcriptional regulation beyond the well-studied examples discussed in the introduction to the genome-level studies that are increasingly essential for understanding complex molecular and cellular biology.

We reviewed the role of cis-acting post-transcriptional regulation by uORFs on a genome-level. We then used ribosome profiling to characterize the global response to ER stress utilizing RNA sequencing and ribosome profiling in collaboration with the Walter lab. In this study, we detected the expected upregulation in translation of known transcripts as well as provide empirical evidence for additional upregulated transcripts all bearing uORFs. Our ribosome profiling data also provides nucleotide resolution information about ribosome occupancy and translation of these uORFs. Importantly, we demonstrate that the uORFs are occupied under stress conditions, providing evidence to revise the model of regulation in response to eIF2 α phosphorylation. The addition of novel uORF regulated transcripts as well as insights into the mechanism of this regulation expands the understanding of cis-acting regulators on post-transcriptional regulation to a genome-wide level with information that will answer long-held questions about this paradoxical upregulation.

Future directions

Our studies on the cis-acting regulatory uORFs raise a number of questions. We review and empirically identify that uORFs are more pervasive than originally anticipated. Studies like Johnstone et al. (2016) start to answer the function of uORFs genome-wide. However, the studies here are correlative rather than causative. Moreover, our studies in the context of the UPR suggest that eIF2 α -limited re-initiation alone is insufficient to

explain CDS upregulation under stress. By incorporating ribosome profiling data studies at known and newly identified uORFs, we can start to ask questions about the mechanism for CDS-specific upregulation while maintaining uORF expression. Looking at this question genome-wide may provide information that follows the pattern of ribosome trafficking, especially in association with other ribosome profiling studies that probe ribosome translation status using drugs or knockouts (Lareau et al. 2014, Guydosh and Green 2014).

We also raise many questions by extending our understanding of trans-acting post-transcriptional regulators. We verified a handful of candidates; however, our data indicate that many more await identification and validation. Moreover, we can use HITS-CLIP and CLIP-MS to address what transcripts are being targeted as well as the mechanistic details of how candidates post-transcriptionally regulate their targets. While we aim to answer some of these questions for a select number of candidates, the screening technique is applicable to additional biological systems. In particular, we can ask how and why unexpected hits such as the glycolysis enzyme Cdc19p bind RNA. We propose that by characterizing the regulatory effect, target RNAs, and co-activators, we can understand a system of metabolic regulation that can extend to other metabolic RBPs. Moreover, as many of these enigmRBPs are conserved from yeast to humans (Beckmann et al. 2016), our results should reflect broadly applicable biological principles.

Conclusion

Genome-wide studies have revealed that post-transcriptional regulation is both pervasive and essential. To understand the mechanisms of post-transcriptional

regulation, we need to both catalog and functionally characterize the cis- and trans-acting factors that contribute to the complex networks that underlie transcript stability and translational efficiency. Our comprehensive analyses on the cis-acting uORFs and trans-acting post-transcriptional regulators advances our study of post-transcriptional control by enlivening the parts list provided by interactome studies to contain functional information.

Appendix 1:

Results from Collaborators

Introduction and preface

Previously in this thesis, the paper “The small molecule ISRIB reverses the effects of eIF2 α phosphorylation on translation and stress granule assembly” was edited to highlight the contributions of the thesis writer and article second author, Anna McGeachy. In line with departmental policy, any results not generated in part or in whole by the thesis writer, but included in a thesis reproduced manuscript, must be presented separately in the appendix under the heading “Results from Collaborators”. The sections previously excluded include:

Results

1. *ISRIB prevents formation of stress granules exclusively triggered by eIF2 α phosphorylation*
2. *ISRIB triggers rapid disassembly of stress granules and restores translation*

1.

Discussion

Stress granule dynamics and ISRIB

1.

2.

3. Materials and Methods

Immunofluorescence

Live cell microscopy

Protein analysis

[35S]-methionine incorporation

In these parts, the thesis writer, Anna McGeachy, provided manuscript editing and
4. statistical consult. However, the experiments were performed by the first author,
Carmela Sidrauski. Since these results add to the article, they will be reproduced below.
These sections were originally published online in eLife on February 9, 2016. In line with
eLife's open access policies, the article is governed by the Creative Commons license
(<http://creativecommons.org/licenses/by/4.0/>). As such, there are no limitations on
reprinting the work in whole or in part as long as proper attribution is given.

**The small molecule ISRIB reverses the effects of eIF2 α phosphorylation on
translation and stress granule assembly**

Carmela Sidrauski^{1*}, Anna M McGeachy², Nicholas T Ingolia², Peter Walter^{1*}

¹ Department of Biochemistry and Biophysics, Howard Hughes Medical Institute, University of California,
San Francis, San Francisco, United States; ² Department of Molecular and Cell Biology, University of
California, Berkeley, Berkeley, United States

**For correspondence: carmelas@me.com (CS); Peter.Walter@ucsf.edu (PW)*

Competing interests: The authors declare that no competing interests exist.

Received: 07 October 2014

Accepted: 04 February 2015

Published: 26 February 2015

Reviewing editor: David Ron, University of Cambridge, United Kingdom

Results

ISRIB prevents formation of stress granules exclusively triggered by eIF2 α phosphorylation

Phosphorylation of eIF2 α and reduction of ternary complex formation are tightly linked to the formation of stress granules (SGs) (Kedersha et al., 2002). ISRIB renders cells insensitive to the effects of eIF2 α phosphorylation, thus leading to the prediction that it prevents SG formation as well. We tested this hypothesis by inducing SG formation using thapsigargin (Tg), a potent ER stressor that inhibits the ER calcium pump and was recently shown by ribosome profiling to yield analogous translational effects to tunicamycin (Reid et al., 2014). Microscopic detection of SGs required a stronger induction of ER stress than commonly achieved with Tm, making Tg the preferred inducer. We monitored SGs by performing immunofluorescence on eIF3a, a translation initiation factor that is recruited into SGs. As expected, we found that ISRIB significantly reduced their assembly upon co-treatment with Tg (Figure A1-2A,B). In addition, ISRIB prevented SG formation induced by arsenite (Ars), another widely used inducer of eIF2 α phosphorylation via activation of HRI. As expected, both treatments induced eIF2 α phosphorylation but only Tg induced the ER-resident kinase, PERK, as seen by its shift in mobility that is due to its extensive auto-phosphorylation (Figure A1-2C). Both the PERK mobility shift and eIF2 α phosphorylation elicited by Tg treatment were blocked by a PERK inhibitor (GSK707800; Axten et al., 2012) but not by ISRIB. Like ISRIB, and as expected by the block in eIF2 α phosphorylation, the GSK PERK inhibitor prevented SG induction upon Tg addition (Figure A1-2A).

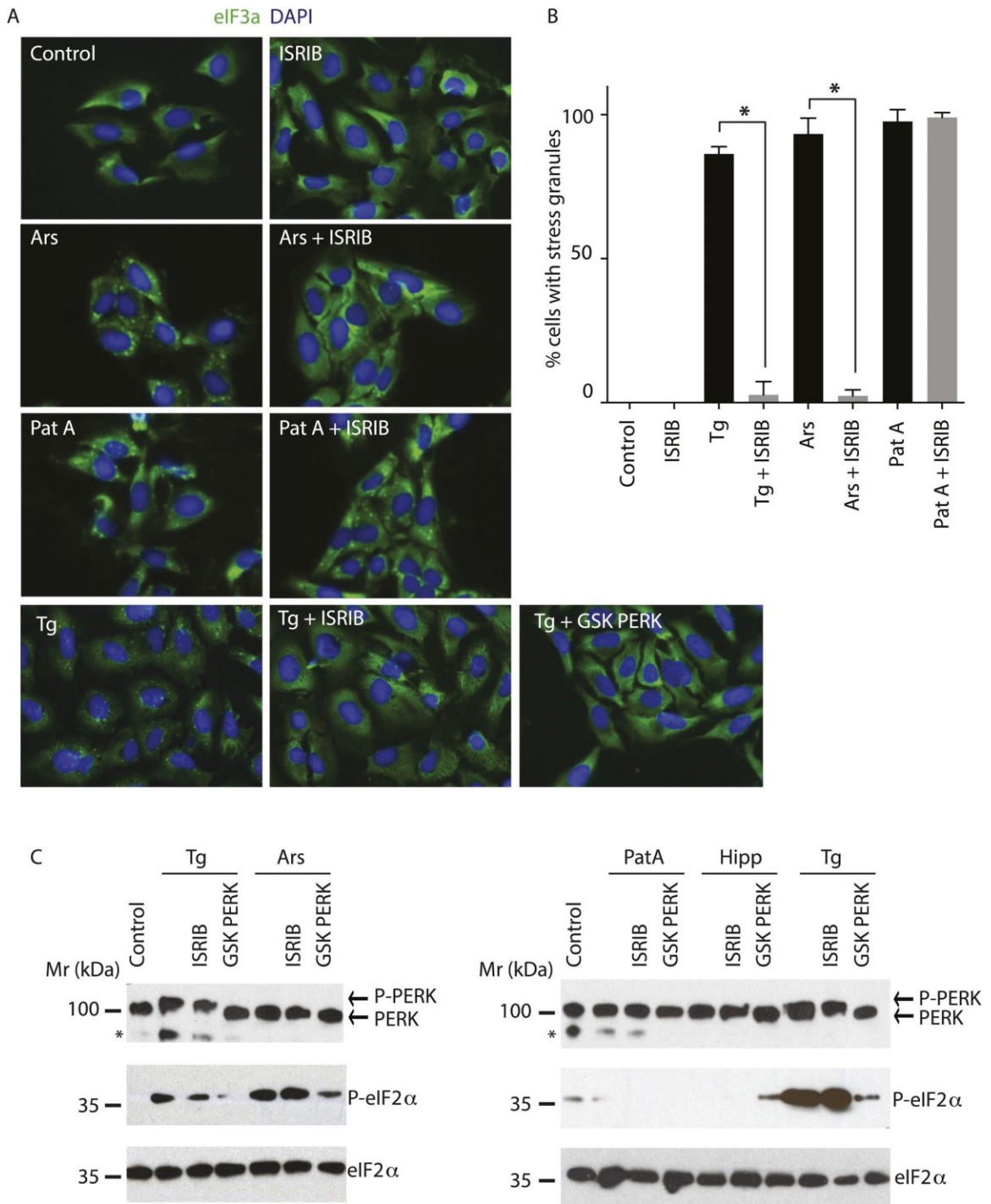


Figure A1-2. ISRIB blocks stress granule formation induced by eIF2 α phosphorylation.

(A) Immunofluorescence analysis (eIF3a) of U2OS cells treated with 200 nM Tg for 1 hr, 250 μ M Ars for 30 min, or 100 nM Pat A for 30 min in the presence or absence of 200 nM ISRIB or 1 μ M GSK797800 PERK inhibitor. A secondary Alexa Dye 488 anti-rabbit antibody was used to visualize eIF3a and DAPI was used to visualize nuclei. Representative images of at least two biological replicates are shown. **(B)** Quantitation of

the percentage of cells containing stress granules in the different conditions described in A. Images were collected from at least two independent experiments and the number of cells with SGs or no SGs counted. The total number of cells counted for each condition was (sum of all replicates): Control (N = 81), ISRIB (N = 94), Tg (N = 122), Tg + ISRIB (N = 71), Ars (N = 85), Ars + ISRIB (N = 84), Pat A (N = 47) and Pat A + ISRIB (N = 64). No cells had SGs in Tg + PERK inh (N = 71). p-values are derived from a Student's t-test, *p < 0.05. (C) Immunoblot analysis of PERK, phospho eIF2 α , and total eIF2 α in U2OS cells treated as in A. Hippuristanol (Hipp) was used at 300 nM for 30 min. The right blot was overexposed to confirm the absence of induction of eIF2 α phosphorylation upon Pat A and Hipp treatment. A representative blot of three independent experiments is shown. The asterisk (*) represents a background band or degradation product. DOI: 10.7554/eLife.05033.009

SG formation can also be induced in the absence of eIF2 α phosphorylation by inhibiting the eIF4A helicase, which is part of the cap-binding eIF4F complex (Mazroui et al., 2006). Pateamine A (Pat A) binds to and inhibits this enzyme and blocks scanning of the PIC and translation initiation (Dang et al., 2006). In agreement, Pat A-induced SG formation but it did not cause eIF2 α phosphorylation (Figure A1-2A,C). In contrast to phospho-eIF2 α -induced SGs, Pat-A-induced SGs were not reduced by ISRIB (Figure A1-2A,B). Thus, ISRIB blocks phospho-eIF2 α -dependent SG induction selectively.

ISRIB triggers rapid disassembly of stress granules and restores translation

To visualize SG formation in living cells and to assess the effects of ISRIB on pre-formed SGs, we took advantage of a stable cell line expressing G3BP fused to GFP (Kedersha et al., 2008). In contrast to cell lines that overexpress SG-associated RNA binding proteins like G3BP, in this single clone-derived cell line, low expression of the fusion protein preserves stress-dependent regulation of SG assembly. We confirmed that in

this cell line ISRIB significantly reduces SG formation driven by stresses that cause eIF2 α phosphorylation (Tg and Ars) but not by phospho-eIF2 α -independent induction through eIF4A inhibition (Pat A and hippuristanol [Hipp]) (Figure A1-3A,B) (Cencic et al., 2012). To match the strength of the stresses used in these experiments and minimize the toxic effects of these agents, we used the shortest incubation time and the lowest concentration of each stressor that resulted in SG formation in the majority of cells. ISRIB has an EC₅₀ of 5 nM as previously measured using an uORFs- ATF4-driven luciferase reporter in HEK293T cells (Sidrauski et al., 2013). In close agreement with the high potency measured in the reporter assay, ISRIB significantly reduced SG formation even at concentrations as low as 2 nM in U2OS cells and as expected, an inactive analog, ISRIBinact, did not reduce their formation (Figure A1-3—figure supplement A1-1).

Treatment of cells with CHX disassembles SGs in the presence of ongoing stress (Kedersha et al., 2000; Mollet et al., 2008). This observation as well as other pharmacological and microscopy data revealed that SGs are highly dynamic structures with mRNAs quickly shuttling in and out. When these mRNAs leave SGs, translation is reinitiated; CHX then immobilizes elongating ribosomes and prevents mRNA re-entry into SGs. Because polyribosome disassembly is blocked by CHX yet required for SG assembly, CHX treatment dissolves pre-formed SGs. As seen in Figure 3C, a 10-min treatment with CHX following Tg induction of SGs (40 min) was sufficient to observe disassembly. Like CHX, ISRIB addition disassembled SGs within 10 min, even in the prolonged presence of the stressor Tg (Figure A1-3C). Whereas ISRIB restored translation of mRNAs that are liberated from SGs, as seen by the quick recovery in [35S]-methionine incorporation, CHX further reduced protein synthesis (Figure A1-3D

and Figure A1-3—figure supplement A1-2). These experiments demonstrate that ISRIB triggers disassembly of pre-formed SGs by loading dissociating mRNAs with actively translating ribosomes.

We next looked at the kinetics of SG disassembly upon ISRIB addition. Strikingly, after only 5 min of ISRIB treatment, Tg-induced SGs were no longer observed in cells (Figure A1-3E and Video A1-1). We also investigated the impact of ISRIB on P-bodies, a molecularly distinct class of RNA aggregates that serve as centers of mRNA decay (Kedersha and Anderson, 2009). The mRNA decay factor Dcp1 serves as a marker for these structures, and we visualized them in living cells using the fusion protein Dcp1-RFP. We saw that P-bodies were constitutively present in a percentage of the cells and were not affected by ISRIB treatment or by the stressors used to induce SGs over the time-course experiments explored here (Figure A1-3E red arrows, Video A1-2 and data not shown).

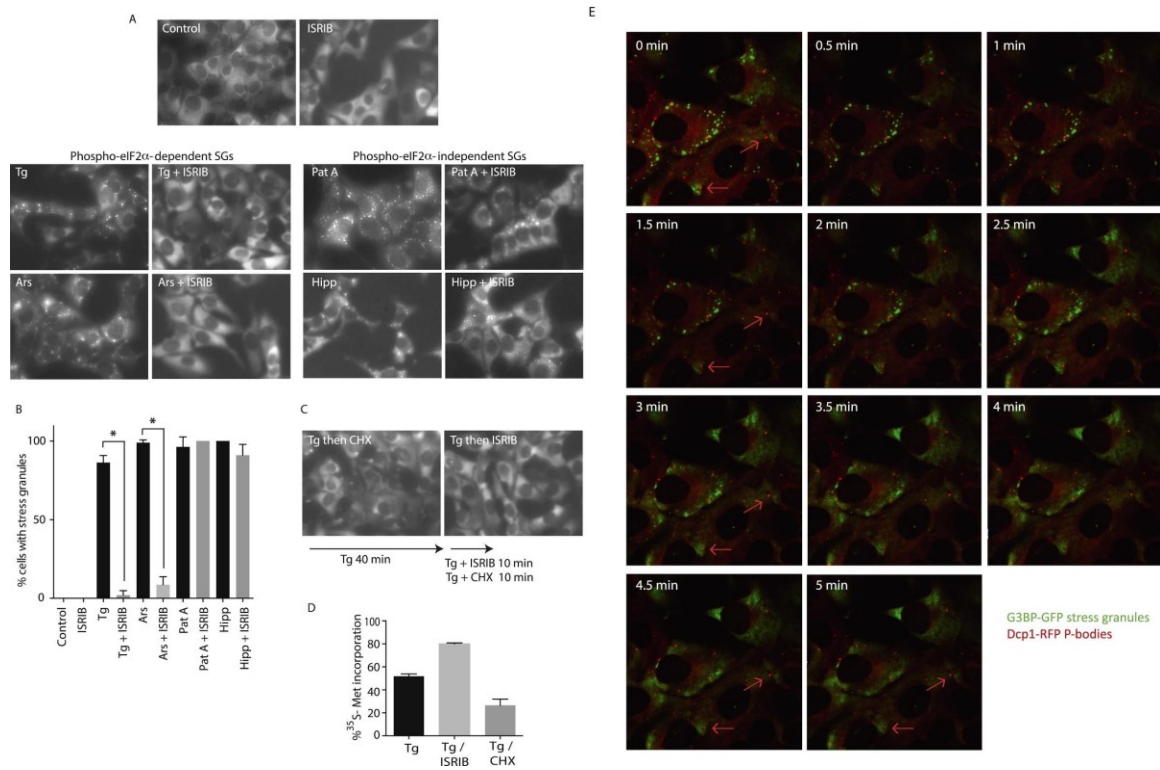


Figure A1-3. ISIRIB addition rapidly dissolves pre-formed stress granules in live cells restoring translation.

(A) Live cell imaging of stress granules in U2OS cells stably expressing G3BP-GFP (SG marker). Cells were treated with 200 nM Tg for 40 min, 250 μM Ars for 30 min, 100 nM Pat A for 30 min, or 300 nM Hipp in the presence or absence of 200 nM ISIRIB. Cells were imaged using an epifluorescence microscope. Representative images of at least two biological replicates are shown. **(B)** Quantitation of the percentage of cells containing stress granules in the different conditions described in A. Images were collected from at least two independent experiments and the number of cells with SGs or no SGs counted. The number of cells analyzed for each condition was (sum of replicates): Control (N = 98), ISIRIB (N = 81), Tg (N = 101), Tg + ISIRIB (N = 84), Ars (N = 80), Ars + ISIRIB (N = 55), Pat A (N = 58), Pat A + ISIRIB (N = 50), Hipp (N = 41) and Hipp + ISIRIB (N = 52). p-values are derived from a Student's t-test, *p < 0.05. **(C)** Stress granules were pre-formed with Tg for 40 min (as in Figure 3A) and then CHX (50 $\mu\text{g}/\text{ml}$) or ISIRIB (200 nM) was added to the well, incubated for 10 min and images were collected. Representative images of at least two biological replicates are shown. **(D)** ISIRIB quickly restores mRNA translation upon disassembly of stress granules. Cells were treated as in C with 200 nM Tg for 40 min and then DMSO, CHX (50 $\mu\text{g}/\text{ml}$), or ISIRIB (200 nM) was added at the same time as [^{35}S]-methionine. Cells were lysed after 15 min, protein was run in an SDS-PAGE gel and radioactivity was measured in each lane (N = 2, mean \pm SD). **(E)** ISIRIB quickly dissolves

stress granules but does not affect P-bodies. Live cell imaging of U2OS cells stably expressing G3BP-GFP (SG marker) and Dcp1-RFP (P-body marker). Cells were treated with 200 nM Tg for 45 min followed by addition of 200 nM ISRIB at t = 0 min to the well and then imaged using spinning disk confocal microscopy. Images were collected every 30 s. The red arrows point to two representative P-bodies. Representative images of at least three biological replicates are shown.

DOI: 10.7554/eLife.05033.010

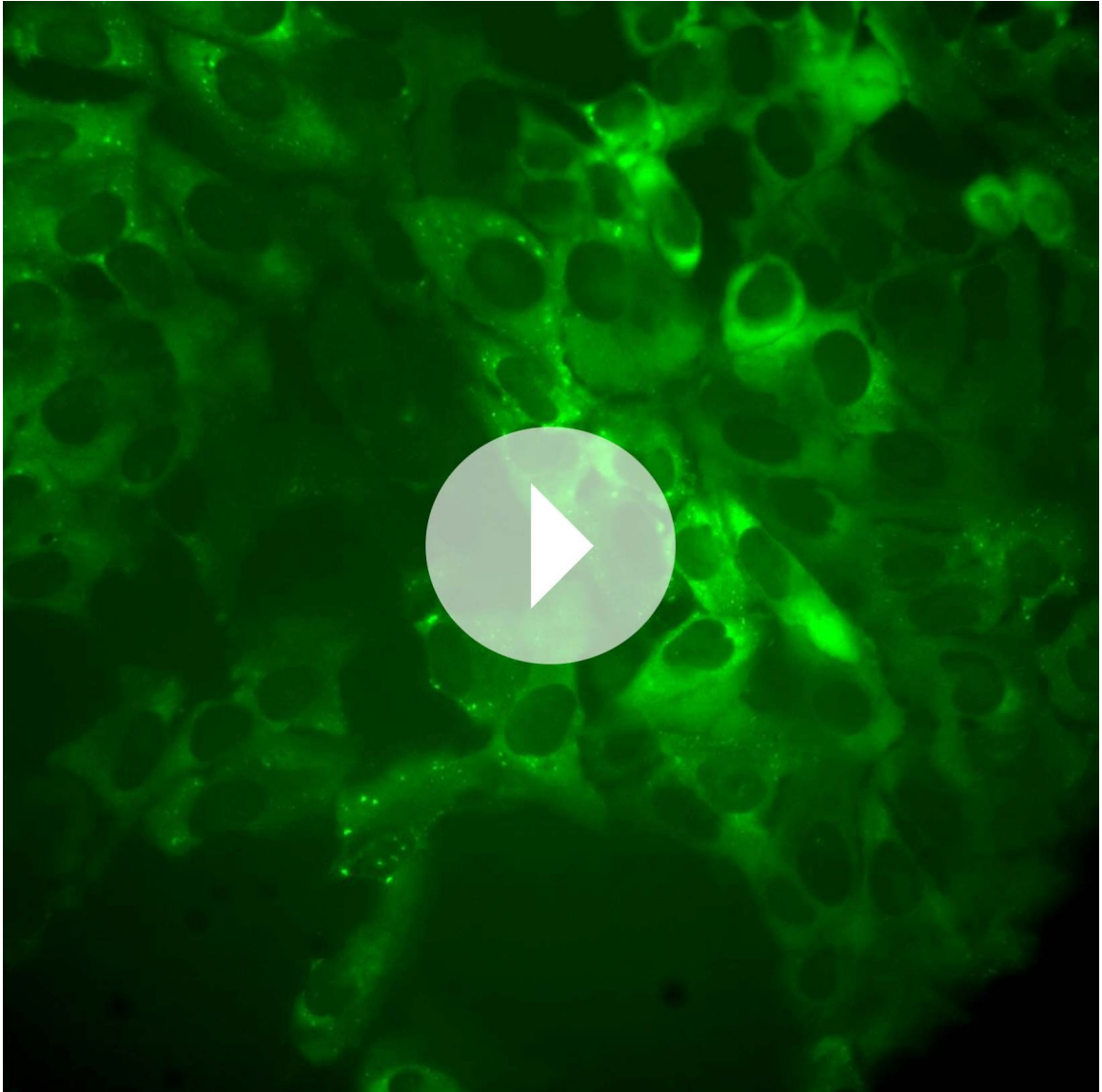
The following figure supplements are available for figure A1-3:

Figure supplement A1-1. ISRIB dose response and inactive analog in stress granule assay.

DOI: 10.7554/eLife.05033.011

Figure supplement A1-2. Representative SDS-PAGE gel of [35S]-methionine pulse as described in Figure 3D.

DOI: 10.7554/eLife.05033.012

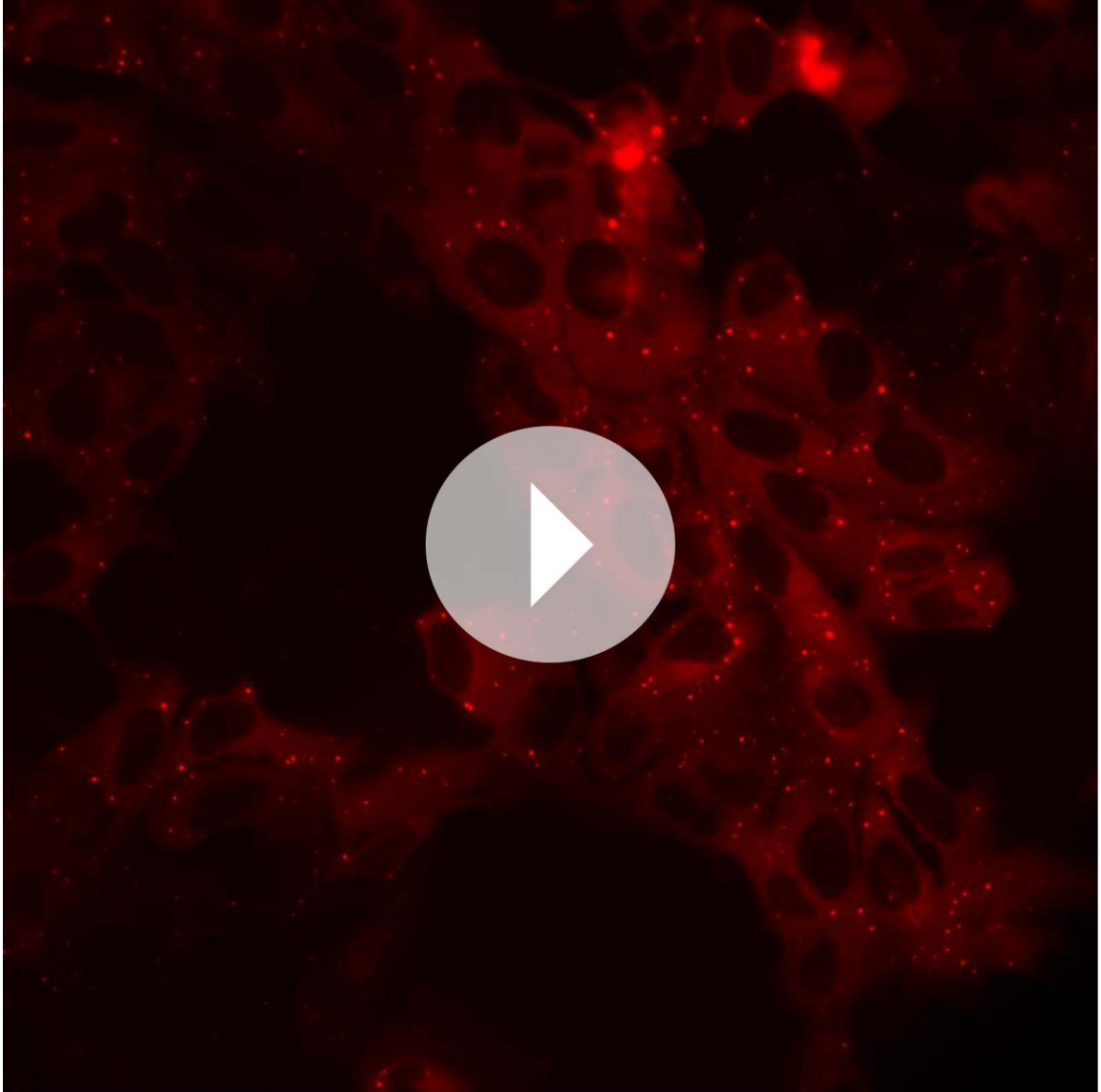


Video A1-1.

ISRIB triggers stress granule disassembly.

U2OS cells stably expressing G3BP-GFP (SG marker) and Dcp1-RFP (P-body marker) were treated with 200 nM Tg for 40 min and then 200 nM ISRIB was added at $t = 0$ min to the well and imaged using an epifluorescence microscope. Images of G3BP-GFP (SGs) were collected every 30 s.

DOI: <http://dx.doi.org/10.7554/eLife.05033.013>



Video A1-2.

ISRIB does not trigger disassembly of P-bodies.

Images of Dcp1-RFP (P-bodies) corresponding to the same field of cells as in Video 1 were collected every 30 s.

DOI: <http://dx.doi.org/10.7554/eLife.05033.014>

Discussion

Stress granule dynamics and ISRIB

The dynamic nature of SGs allowed us to monitor the action of ISRIB upon its addition to live cells in real time. Strikingly, addition of ISRIB to stressed cells with pre-formed SGs lead to their quick dissolution (less than 5 min), liberating mRNAs back into the translational pool. A pulse of [35S]-methionine confirmed the fast recovery in protein synthesis even in the presence of stress. Although the molecular target of ISRIB remains unknown, its quick action suggests a direct effect on translation initiation. Phospho-eIF2 α resistance has been observed both in yeast and in mammalian cells. In yeast, mutations in eIF2B (the GEF for eIF2) and eIF5 (the 48S PIC-associated GTPase-activating protein for eIF2) have been reported to make cells insensitive to this phosphorylation event (Vazquez de Aldana and Hinnebusch, 1994; Pavitt et al., 1997, 1998). In mammalian cells, TLR4 engagement in macrophages leads to increased eIF2B activity by removal of an inhibitory phosphorylation and insensitivity to ISR activation (Woo et al., 2012). Thus, ISRIB may directly or indirectly enhance the activity of eIF2B, eIF5, or other initiation factors, thus quickly reversing the cellular effects of phosphorylated eIF2 α .

SGs contain a large number of RBPs that harbor low complexity sequence domains that nucleate through transient, low affinity interactions (Kato et al., 2012). These RBPs usually contain several RNA-binding domains and can associate with more than one mRNA; this multi-valency further favors the coalescence of RNA-protein granules. A conspicuous feature of some degenerative diseases is the cytoplasmic or nuclear aggregation of RBPs, driven in some cases by pathogenic mutations. TDP-43 and FUS mutations are found in amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD) (Li et al., 2013), and mutations in hnRNPA1 and hnRNPA2/B1 have also been found in ALS (Kim et al., 2013). Recent reports have also described the presence of RNA and RBPs in aggregates that form in prion disease, tauopathies, and

Alzheimer's (Vanderweyde et al., 2012; Ash et al., 2014). The impact of these cytosolic aggregates on SG dynamics is not known, though they may hamper the ability of SGs to properly dissolve, thereby contributing to sustained translational attenuation and neurodegeneration. By quickly disassembling SGs even in the presence of stress, ISRIB may provide a useful therapeutic intervention in these diseases by antagonizing the cellular effects of pathogenic RNA-protein assemblies.

Materials and Methods

Immunofluorescence

U2OS cells were seeded on 4-well chamber slides (Lab-Tek) 18 hr prior to processing for immunofluorescence. Cells (80% confluent) were fixed with ice-cold methanol. The cells were then rinsed with PBS (Sigma) and blocked for 1 hr at room temperature in 0.5% BSA in PBS. The cells were then incubated overnight at 4 °C with an anti-eIF3A rabbit antibody (#3411; Cell Signaling Technology) at a 1:1000 dilution in blocking buffer. The next morning the slides were washed three times (5 min each time) with PBS and then incubated for 1 hr at room temperature in a 1:1000 dilution (in 0.5% BSA in PBS) of secondary anti-rabbit antibody labeled with Alexa Dye 488 (Molecular Probes). The slides were washed three additional times with PBS. The slides were then mounted with antifade reagent with DAPI (Life Technologies P-36931). Lastly, the slides were imaged using a Zeiss Axiovert 200M epifluorescence microscope.

Live cell microscopy

U2OS G3BP-GFP/Dcp1-RFP cells were plated in 8-well Lab-Tek chamber slides and switched to imaging media (lacking phenol red) upon addition of different stress inducers. Cells were either imaged using a Zeiss Axiovert 200M epifluorescence microscope or in a heated chamber using a spinning confocal epifluorescence

microscope (Eclipse Ti-Nikon) and an Andor iXon3 camera.

Protein analysis

Cells were washed with PBS and lysed in SDS-PAGE loading buffer (1% SDS, 62.5 mM Tris-HCl pH 6.8, 10% glycerol). Lysates were sonicated and loaded on Any-kD SDS-PAGE gels (BioRad). Proteins were transferred onto nitrocellulose and probed with primary antibodies diluted in Tris-buffered saline supplemented with 0.1% Tween 20 and 5% BSA. The following antibodies were used: PERK (D11A8) (1:1000), eIF2 α (#9722; Cell Signaling technology) (1:1000), phospho-eIF2 α (Ser51) (44728G; Invitrogen). An HRP-conjugated secondary antibody (Amersham) was employed to detect immune-reactive bands using enhanced chemiluminescence (SuperSignal, Thermo Scientific).

[35S]-methionine incorporation

U2OS GFP-G3BP/mRFP-DCP1a cells were seeded on 12-well plates, allowed to recover overnight and treated with 100 nM Tg for 40 min. ISRIB (200 nM) or CHX (50 μ g/ml) was added at the same time as 50 μ Ci of [35S]-methionine (Perkin Elmer) and incubated for 15 min. Cells were lysed by addition of SDS-PAGE loading buffer. Lysates were sonicated and equal amounts were loaded on SDS-PAGE gels (BioRad). The gel was dried and radioactive methionine incorporation was detected by exposure to a phosphor-screen and visualized with a Typhoon 9400 Variable Mode Imager (GE Healthcare).

Supplement

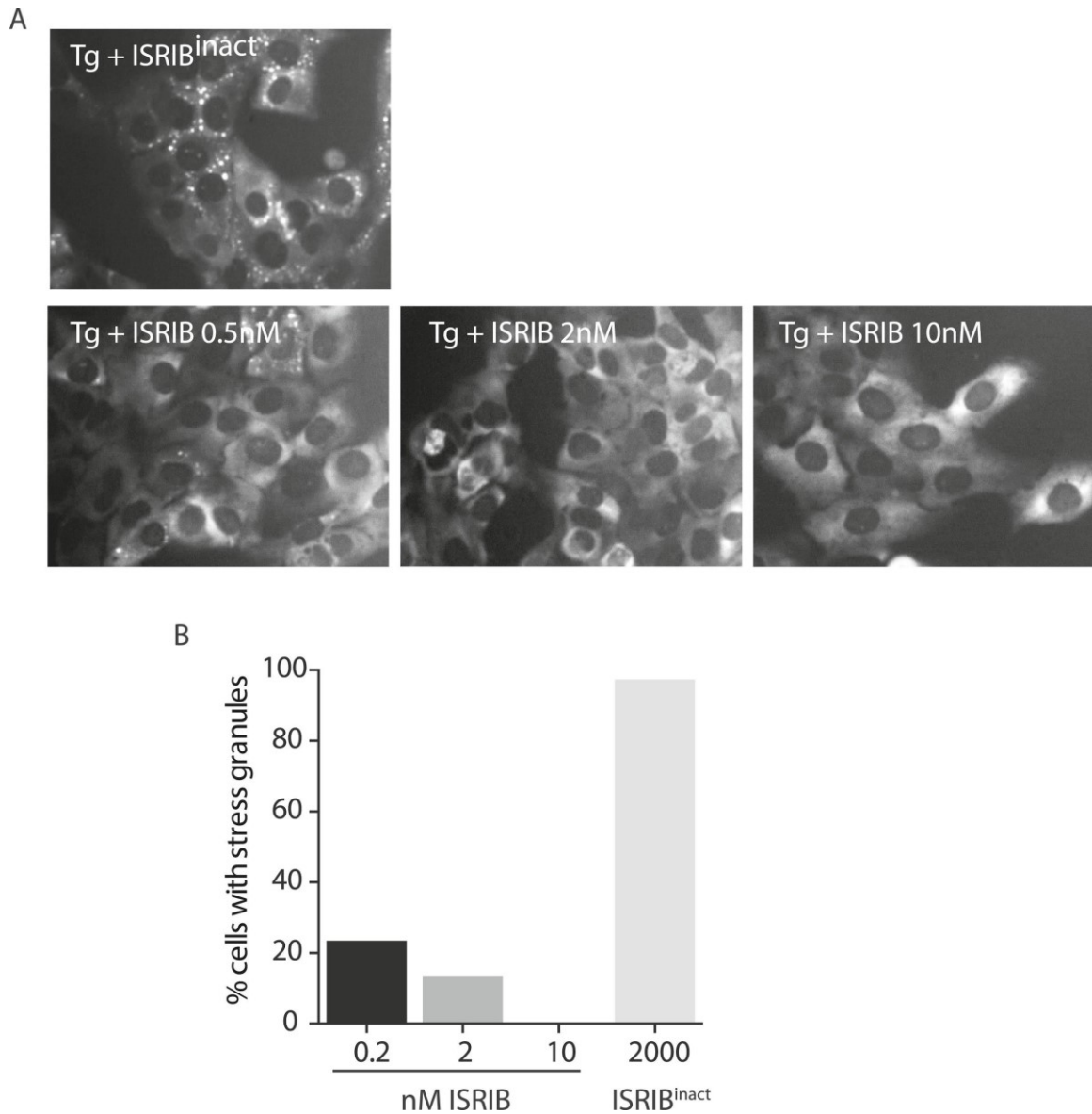


Figure A1-3—figure supplement A1-1.

ISRIB dose response and inactive analog in stress granule assay.

(A) Live cell imaging of SGs in U2OS cells stably expressing G3BP-GFP. Cells were treated with 200 nM Tg and different doses of ISRIB (as indicated) or 2 μ M of an inactive analog of ISRIB (ISRIB^{inact}).

Representative images of at least two biological replicates are shown. (B) Quantitation of the percentage of cells containing stress granules in the different conditions. The number of cells analyzed for each condition were: ISRIB^{inact} (N = 37), 0.5 nM ISRIB (N = 81), 2 nM ISRIB (N = 91), 10 nM ISRIB (N = 43).

DOI: <http://dx.doi.org/10.7554/eLife.05033.011>

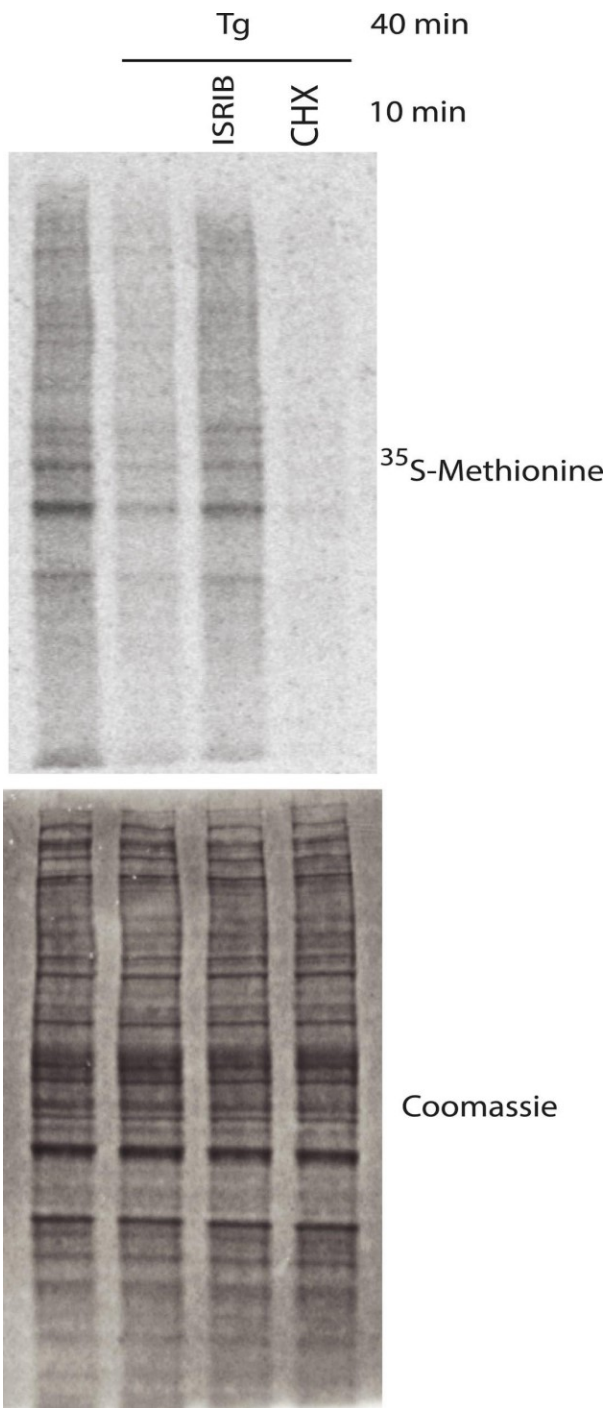


Figure A1-3—figure supplement A1-2.

Representative SDS-PAGE gel of [³⁵S]-methionine pulse as described in Figure 3D.

Top panel is an autoradiogram and bottom panel is total protein of the same gel as shown by Coomassie staining.

DOI: <http://dx.doi.org/10.7554/eLife.05033.012>

Bibliography

- Akopian D, Shen K, Zhang X, Shan S-O (2013) Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annu Rev Biochem* **82**: 693–721.
- Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB (2010) Annotating non-coding regions of the genome. *Nature Publishing Group* **11**: 559–571.
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology* **11**: R106.
- Andreev DE, O'Connor PBF, Fahey C, Kenny EM, Terenin IM, Dmitriev SE, Cormican P, Morris DW, Shatsky IN, Baranov PV (2015) Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *eLife* **4**: e03971–21.
- Arribere JA, Gilbert WV (2013) Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Research* **23**: 977–987.
- Ash PEA, Vanderweyde TE, Youmans KL, Apicco DJ, Wolozin B (2014) Pathological stress granules in Alzheimer's disease. *Brain Research* **1584**: 52–58.
- Atkins C, Liu Q, Minthorn E, Zhang SY, Figueroa DJ, Moss K, Stanley TB, Sanders B, Goetz A, Gaul N, et al. (2013) Characterization of a Novel PERK Kinase Inhibitor with Antitumor and Antiangiogenic Activity. *Cancer Research* **73**: 1993–2002.
- Avery OT, MacLeod CM, McCarty M (1944) STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM

PNEUMOCOCCUS TYPE III. *The Journal of Experimental Medicine* **79**: 137–158.

Axten JM, Medina JR, Feng Y, Shu A, Romeril SP, Grant SW, Li WHH, Heerding DA, Minthorn E, Mencken T, et al. (2012) Discovery of 7-Methyl-5-(1-[[3-(trifluoromethyl)phenyl]acetyl]-2,3-dihydro-1H-indol-5-yl)-7H-pyrrolo[2,3-d]pyrimidin-4-amine (GSK2606414), a Potent and Selective First-in-Class Inhibitor of Protein Kinase R (PKR)-like Endoplasmic Reticulum Kinase (PERK). *J Med Chem* **55**: 7193–7207.

Azzouz N, Panasenko OO, Deluen C, Hsieh J, Theiler G, Collart MA (2009) Specific roles for the Ccr4-Not complex subunits in expression of the genome. *RNA* **15**: 377–383.

Bailey TL, Elkan C (1995) Unsupervised Learning of Multiple Motifs in Biopolymers Using Expectation Maximization. *Machine Learning* **21**: 51–80.

Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* **34**: W369–W373.

Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, et al. (2012) The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Molecular Cell* **46**: 674–690.

Barker DD, Wang C, Moore J, Dickinson LK, Lehmann R (1992) Pumilio is essential for function but not distribution of the *Drosophila* abdominal determinant Nanos. *Genes & Development* **6**: 2312–2326.

Bassik MC, Kampmann M, Lebbink RJ, Wang S, Hein MY, Poser I, Weibezahn J, Horlbeck MA, Chen S, Mann M, et al. (2013) A Systematic Mammalian Genetic Interaction Map Reveals Pathways Underlying Ricin Susceptibility. *CELL* **152**: 909–922.

- Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, et al. (2014) Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *The EMBO Journal* **33**: 937–938.
- Bazzini AA, Lee MT, Giraldez AJ (2012) Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**: 233–237.
- Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume A-M, Schwarzl T, Curk TZ, Foehr S, Huber W, et al. (2015) The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nature Communications* **6**: 1–9.
- Beelman CA, Parker R (1995) Degradation of mRNA in Eukaryotes. *CELL* **81**: 179–183.
- Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E (2006) mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes & Development* **20**: 1885–1898.
- Belasco JG, Brawerman G (1993) *Control of Messenger RNA Stability*. Woodhead Publishing Limited.
- Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B Methodological* **57**: 289–300.
- Berget SM, Moore C, Sharp PA (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* **74**: 3171–3175.
- Berleth T, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, Noll M, Nüsslein-Volhard C (1988) The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *The EMBO Journal* **7**: 1749–1756.

- Bernstein P, Ross J (1989) Poly(A), poly(A) binding protein and the regulation of mRNA stability. *Trends in Biochemical Sciences* **14**: 373–377.
- Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM (1998) Localization of ASH1 mRNA particles in living yeast. *Molecular Cell* **2**: 437–445.
- Blobel G, Dobberstein B (1975) Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *The Journal of Cell Biology* **67**: 835–851.
- Bobola N, Jansen RP, Shin TH, Nasmyth K (1996) Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells. *CELL* **84**: 699–709.
- Böhl F, Kruse C, Frank A, Ferring D, Jansen RP (2000) She2p, a novel RNA-binding protein tethers ASH1 mRNA to the Myo4p myosin motor via She3p. *The EMBO Journal* **19**: 5514–5524.
- Boyce M, Bryant KF, Jousse C, Long K, Harding HP, Scheuner D, Kaufman RJ, Ma D, Coen DM, Ron D, et al. (2005) A selective inhibitor of eIF2alpha dephosphorylation protects cells from ER stress. *Science* **307**: 935–939.
- Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS (2012) High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science* **335**: 552–557.
- Brenner S, Jacob F, Meselson M (1961) An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. **190**: 576–581.
- Bringmann P, Lührmann R (1986) Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cels. 1–8.
- Bryson V, Szybalski W (1952) Microbial Selection. *Science* **116**: 45–51.

- Buchan JR, Parker R (2009) Eukaryotic Stress Granules: The Ins and Outs of Translation. *Molecular Cell* **36**: 932–941.
- Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, Snyder MP, Greenleaf WJ (2014) Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nature Biotechnology* **32**: 562–568.
- Bull AL (1966) Bicaudal, a genetic factor which affects the polarity of the embryo in *Drosophila melanogaster*. *Journal of Experimental Zoology* **161**: 221–241.
- Burke RL, Tekamp-Olson P, Najarian R (1983) The Isolation, Characterization, and Sequence of the Pyruvate Kinase Gene of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **258**: 2193–2201.
- Byrne KP, Wolfe KH (2005) The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research* **15**: 1456–1461.
- Calabretta S, Richard S (2015) Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends in Biochemical Sciences* **40**: 662–672.
- Calvo SE, Pagliarini DJ, Mootha VK (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci USA* **106**: 7507–7512.
- Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. (2012) Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *CELL* **149**: 1393–1406.
- Castilho BA, Shanmugam R, Silva RC, Ramesh R, Himme BM, Sattlegger E (2014) Keeping the eIF2 alpha kinase Gcn2 in check. *BBA - Molecular Cell Research* **1843**: 1948–1968.

- Cencic R, Galicia-Vzquez G, Pelletier J (2012) *Inhibitors of Translation Targeting Eukaryotic Translation Initiation Factor 4A*. Elsevier Inc.
- Cha B-J, Koppetsch BS, Theurkauf WE (2001) In Vivo Analysis of *Drosophila* bicoid mRNA Localization Reveals a Novel Microtubule-Dependent Axis Specification Pathway. *CELL* **106**: 35–46.
- Chagnovich D, Lehmann R (2001) Poly(A)-independent regulation of maternal hunchback translation in the *Drosophila* embryo. *PNAS* **98**: 11359–11354.
- Chartrand P, Meng XH, Huttelmaier S, Donato D, Singer RH (2002) Asymmetric sorting of ash1p in yeast results from inhibition of translation by localization elements in the mRNA. *Molecular Cell* **10**: 1319–1330.
- Chartrand P, Meng XH, Singer RH, Long RM (1999) Structural elements required for the localization of ASH1 mRNA and of a green fluorescent protein reporter particle in vivo. *Current Biology* **9**: 333–336.
- Chekulaeva M, Filipowicz W, Parker R (2009) Multiple independent domains of dGW182 function in miRNA-mediated repression in *Drosophila*. *RNA* **15**: 794–803.
- Chekulaeva M, Mathys H, Zipprich JT, Attig J, Colic M, Parker R, Filipowicz W (2011) miRNA repression involves GW182-mediated recruitment of CCR4–NOT through conserved W-containing motifs. *Nature Structural & Molecular Biology* **18**: 1218–1226.
- Chen J-J (2014) Translational control by heme-regulated eIF2 α kinase during erythropoiesis. *Current Opinion in Hematology* **21**: 172–178.
- Chen T, Ozel D, Qiao Y, Harbinski F, Chen L, Denoyelle SEV, He X, Zvereva N, Supko JG, Chorev M, et al. (2011) Chemical genetics identify eIF2 α kinase heme-regulated inhibitor as an anticancer target. *Nature Chemical Biology* **7**: 608–614.

- Chen Y, Brandizzi F (2013) IRE1: ER stress sensor and cell fate executor. *Trends in Cell Biology* **23**: 547–555.
- Cheng Z, Teo G, Krueger S, Rock TM, Koh HW, Choi H, Vogel C (2016) Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. *Mol Syst Biol* **12**: 855–855.
- Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, et al. (2011) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research* **40**: D700–D705.
- Chew GL, Pauli A, Rinn JL, Regev A, Schier AF, Valen E (2013) Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development* **140**: 2828–2834.
- Cho PF, Gamberi C, Cho-Park YA, Cho-Park IB, Lasko P (2006) Cap-Dependent Translational Inhibition Establishes Two Opposing Morphogen Gradients in *Drosophila* Embryos. *Current Biology* **16**: 2035–2041.
- Choi S, Han K (2011) Prediction of RNA-binding amino acids from protein and RNA sequences. *BMC Bioinformatics* **12 Suppl 13**: S7.
- Chow LT, Gelinis RE, Broker TR, Roberts RJ (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *CELL* **12**: 1–8.
- Ciriacy M, Breitenbach I (1979) Physiological effects of seven different blocks in glycolysis in *Saccharomyces cerevisiae*. *Journal of Bacteriology* **139**: 152–160.
- Collart MA, Timmers HTM (2004) The Eukaryotic Ccr4-Not Complex: A Regulatory Platform Integrating mRNA Metabolism with Cellular Signaling Pathways. 1–34.

- Coller J, Gray NK, Wickens MP (1998) mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. *Genes & Development* **12**: 3226–3235.
- Coller J, Parker R (2004) Eukaryotic mRNA Decapping. *Annu Rev Biochem* **73**: 861–890.
- Coller J, Wickens M (2007) Tethered Function Assays: An Adaptable Approach to Study RNA Regulatory Proteins. In, *Translation Initiation: Extract Systems and Molecular Genetics* pp 299–321. Elsevier.
- Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR (2011) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Research* **39**: D301–D308.
- Coons AH, Creech HJ, Jones RN (1941) Immunological Properties of an Antibody Containing a Fluorescent Group. *Proc Soc Exp Biol Med* **47**: 200–202.
- Costa-Mattioli M, Gobert D, Harding H, Herdy B, Azzi M, Bruno M, Bidinosti M, Ben Mamou C, Marcinkiewicz E, Yoshida M, et al. (2005) Translational control of hippocampal synaptic plasticity and memory by the eIF2 α kinase GCN2. *Nature* **436**: 1166–1173.
- Costa-Mattioli M, Gobert D, Stern E, Gamache K, Colina R, Cuello C, Sossin W, Kaufman R, Pelletier J, Rosenblum K, et al. (2007) eIF2 α Phosphorylation Bidirectionally Regulates the Switch from Short- to Long-Term Synaptic Plasticity and Memory. *CELL* **129**: 195–206.
- Costanzo M, Baryshnikova A, Myers CL, Andrews B, Boone C (2011) Charting the genetic interaction map of a cell. *Current Opinion in Biotechnology* **22**: 66–74.
- Cullinan SB, Zhang D, Hannink M, Arvisais E, Kaufman RJ, Diehl JA (2003) Nrf2 Is a Direct PERK Substrate and Effector of PERK-Dependent Cell Survival. *Molecular and Cellular Biology* **23**: 7198–7209.

- Damgaard CK, Lykke-Andersen J (2011) Translational coregulation of 5'TOP mRNAs by TIA-1 and TIAR. *Genes & Development* **25**: 2057–2068.
- Dang Y, Kedersha N, Low W-K, Romo D, Gorospe M, Kaufman R, Anderson P, Liu JO (2006) Eukaryotic initiation factor 2alpha-independent pathway of stress granule induction by the natural product pateamine A. *Journal of Biological Chemistry* **281**: 32870–32878.
- Darnell JE (2012) *RNA: Life's Indispensable Molecule*.
- Darnell RB (2010) HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *WIREs RNA* **1**: 266–286.
- Daugeron MC, Mauxion F, Séraphin B (2001) The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. *Nucleic Acids Research* **29**: 2448–2455.
- Deng Y, Singer RH, Gu W (2008) Translation of ASH1 mRNA is repressed by Puf6p-Fun12p/eIF5B interaction and released by CK2 phosphorylation. *Genes & Development* **22**: 1037–1050.
- Dever TE, Yang W, Aström S, Byström AS, Hinnebusch AG (1995) Modulation of tRNA(iMet), eIF-2, and eIF-2B expression shows that GCN4 translation is inversely coupled to the level of eIF-2.GTP.Met-tRNA(iMet) ternary complexes. *Molecular and Cellular Biology* **15**: 6351–6363.
- Di Prisco GV, Huang W, Buffington SA, Hsu C-C, Bonnen PE, Placzek AN, Sidrauski C, Krnjević K, Kaufman RJ, Walter P, et al. (2014) Translational control of mGluR-dependent long-term depression and object-place learning by eIF2α. *Nature Publishing Group* **17**: 1073–1082.
- Djuranovic S, Nahvi A, Green R (2012) miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336**: 237–240.

- Doidge R, Mittal S, Aslam A, Winkler GS (2012) Deadenylation of cytoplasmic mRNA by the mammalian Ccr4-Not complex. *Biochim Soc Trans* **40**: 896–901.
- Donnelly ML, Hughes LE, Luke G, Mendoza H, Dam ten E, Gani D, Ryan MD (2001) The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J Gen Virol* **82**: 1027–1041.
- Dreyfuss G, Philipson L, Mattaj IW (1988) Ribonucleoprotein particles in cellular processes. *The Journal of Cell Biology* **106**: 1419–1425.
- Driever W, Nüsslein-Volhard C (1988) A Gradient of bicoid Protein in *Drosophila* Embryos. *CELL* **54**: 83–93.
- Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**: 818–822.
- Eulalio A, Eulalio A, Triteschler F, Triteschler F, Izaurralde E, Izaurralde E (2009) The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. *RNA* **15**: 1433–1442.
- Eulalio A, Helms S, Fritsch C, Fauser M, Izaurralde E (2009) A C-terminal silencing domain in GW182 is essential for miRNA function. *RNA* **15**: 1067–1077.
- Eulalio A, Huntzinger E, Izaurralde E (2008) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol* **15**: 346–353.
- Eulalio A, Triteschler F, Büttner R, Weichenrieder O, Izaurralde E, Truffault V (2009) The RRM domain in GW182 proteins contributes to miRNA-mediated gene silencing. *Nucleic Acids Research* **37**: 2974–2983.
- Eystathiou T, Chan EKL, Tenenbaum SA, Keene JD, Griffith K, Fritzler MJ (2002) A Phosphorylated Cytoplasmic Autoantigen, GW182, Associates with a

- Unique Population of Human mRNAs within Novel Cytoplasmic Speckles. *Mol Biol Cell* **13**: 1338–1351.
- Fabian MR, Cieplak MK, Frank F, Morita M, Green J, Srikumar T, Nagar B, Yamamoto T, Raught B, Duchaine TF, 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.
- Fenton AW, Blair JB (2002) Kinetic and Allosteric Consequences of Mutations in the Subunit and Domain Interfaces and the Allosteric Site of Yeast Pyruvate Kinase. *Archives of Biochemistry and Biophysics* **397**: 28–39.
- Filipovska A, Razif MFM, Nygård KKA, Rackham O (2011) A universal code for RNA recognition by PUF proteins. *Nature Chemical Biology* **7**: 425–424.
- Finoux A-L, Séraphin B (2006) In vivo targeting of the yeast Pop2 deadenylase subunit to reporter transcripts induces their rapid degradation and generates new decay intermediates. *Journal of Biological Chemistry* **281**: 25940–25947.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806–811.
- Fleischer TC, Weaver CM, McAfee KJ, Jennings JL, Link AJ (2006) Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes & Development* **20**: 1294–1307.
- Freeberg MA, Han T, Moresco JJ, Kong A, Yang Y-C, Lu ZJ, Yates JR, Kim JK (2013) Pervasive and dynamic protein binding sites of the mRNA transcriptome in *Saccharomyces cerevisiae*. *Genome Biology* **14**: R13.
- Frendewey D, Keller W (1985) Stepwise Assembly of a Pre-mRNA Splicing Complex Requires U-snRNPs and Specific Intron Sequences. *CELL* 1–13.

- Frohnhofer HG, Nüsslein-Volhard C (1986) Organization of anterior pattern in the *Drosophila* embryo by the maternal gene bicoid. *Nature* **324**: 120–125.
- Gall JG, Lou Pardue M (1969) Formation and Detection of RNA-DNA Hybrid Molecules in Cytological Preparations. **63**: 378–383.
- Gamberi C, Peterson DS, He L, Gottlieb E (2002) An anterior function for the *Drosophila* posterior determinant Pumilio. *Development* **129**: 2699–2710.
- Gerber AP, Luschnig S, Krasnow MA, Brown PO, Herschlag D (2006) Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **103**: 4487–4492.
- Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. *Nat Rev Genet* **15**: 829–845.
- Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF (2006) Zebrafish MiR-430 Promotes Deadenylation and Clearance of Maternal mRNAs. **312**: 75–79.
- Glaunsinger BA, Ganem DE (2006) Messenger RNA Turnover and its Regulation in Herpesviral Infection. In pp 337–394. Elsevier.
- Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters* **582**: 1977–1986.
- Gonzalez I, Buonomo SB, Nasmyth K, Ahsen von U (1999) ASH1 mRNA localization in yeast involves multiple secondary structural elements and Ash1 protein translation. *Curr Biol* **9**: 337–340.
- Gorski J, Morrison MR, Merkel CG, LINGREL JB (1975) Poly(A) size class distribution in globin mRNAs as a function of time. *Nature* **253**: 749–751.
- Granneman S, Kudla G, Petfalski E, Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci USA* **106**: 9613–9618.

- Grant CM, Miller PF, Hinnebusch AG (1995) Sequences 5' of the first upstream open reading frame in GCN4 mRNA are required for efficient translational reinitiation. *Nucleic Acids Research* **23**: 3980–3988.
- Gray NK, Collier JM, Dickson KS, Wickens M (2000) Multiple portions of poly(A)-binding protein stimulate translation in vivo. *The EMBO Journal* **19**: 4723–4733.
- Gu W, Deng Y, Zenklusen D, Singer RH (2004) A new yeast PUF family protein, Puf6p, represses ASH1 mRNA translation and is required for its localization. *Genes & Development* **18**: 1452–1465.
- Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**: 835–840.
- Guydosh NR, Green R (2014) Dom34 Rescues Ribosomes in 3' Untranslated Regions. *CELL* **156**: 950–962.
- Ha M, Kim VN (2014) Regulation of microRNA biogenesis. *Nature Publishing Group* **15**: 509–524.
- Haas G, Braun JE, Igreja C, Tritschler F, Nishihara T, Izaurralde E (2010) HPat provides a link between deadenylation and decapping in metazoa. *The Journal of Cell Biology* **189**: 289–302.
- Hafen E, Levine M, Garber RL, Gehring WJ (1983) An improved in situ hybridization method for the detection of cellular RNAs in *Drosophila* tissue sections and its application for localizing transcripts of the homeotic Antennapedia gene complex. *The EMBO Journal* **2**: 617–623.
- Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp A-C, Munschauer M, et al. (2010) Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *CELL* **141**: 129–141.

- Han LY, Cai CZ, Lo SL, Chung MCM, Chen YZ (2004) Prediction of RNA-binding proteins from primary sequence by a support vector machine approach. *RNA* **10**: 355–368.
- Harding HP, Novoa I, Zhang Y, Zeng H, Wek R, Schapira M, Ron D (2000) Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Molecular Cell* **6**: 1099–1108.
- Harding HP, Zeng H, Zhang Y, Jungries R, Chung P, Plesken H, Sabatini DD, Ron D (2001) Diabetes Mellitus and Exocrine Pancreatic Dysfunction in Perk^{-/-} Mice Reveals a Role for Translational Control in Secretory Cell Survival. *Molecular Cell* **7**: 1153–1163.
- Harding HP, Zhang Y, Zeng H, Novoa I, Lu PD, Calton M, Sadri N, Yun C, Popko B, Paules R, et al. (2003) An Integrated Stress Response Regulates Amino Acid Metabolism and Resistance to Oxidative Stress. *Molecular Cell* **11**: 619–633.
- HAYASHI M, HAYASHI MN, SPIEGELMAN S (1963) RESTRICTION OF IN VIVO GENETIC TRANSCRIPTION TO ONE OF THE COMPLEMENTARY STRANDS OF DNA. *Proc Natl Acad Sci USA* **50**: 664–672.
- Haze K, Yoshida H, Yanagi H, Yura T, Mori K (1999) Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress. *Mol Biol Cell* **10**: 3787–3799.
- He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* **5**: 522–531.
- Herskowitz I (1988) Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiological Reviews* 536–553.

- Hiller M, Pudimat R, Busch A, Backofen R (2006) Using RNA secondary structures to guide sequence motif finding towards single-stranded regions. *Nucleic Acids Research* **34**: e117–e117.
- Hinnebusch AG (2005) Translational regulation of GCN4 and the general amino acid control of yeast. *Annual Review of Microbiology* **59**: 407–450.
- Hinnebusch AG, Lorsch JR (2012) The Mechanism of Eukaryotic Translation Initiation: New Insights and Challenges. *Cold Spring Harbor Perspectives in Biology* **4**: a011544–a011544.
- Ho JJD, Marsden PA (2013) Competition and collaboration between RNA-binding proteins and microRNAs. *WIREs RNA* **5**: 69–86.
- Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO (2008) Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System. *Plos Biol* **6**: e255.
- Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A, Shi EY, Stumpf CR, Christensen C, Bonham MJ, et al. (2012) The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**: 55–61.
- Hurtaud C, Gelly C, Bouillaud F, Lévi-Meyrueis C (2006) Translation control of UCP2 synthesis by the upstream open reading frame. *Cell Mol Life Sci* **63**: 1780–1789.
- Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7**: 1534–1550.
- Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJS, Jackson SE, Wills MR, Weissman JS (2014) Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *CellReports* **8**: 1365–1379.

- Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218–223.
- Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *CELL* **147**: 789–802.
- Irie K, Tadauchi T, Takizawa PA, Vale RD, Matsumoto K, Herskowitz I (2002) The Khd1 protein, which has three KH RNA-binding motifs, is required for proper localization of ASH1 mRNA in yeast. *The EMBO Journal* **21**: 1158–1167.
- Jackson RJ, Hellen CUT, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**: 1–15.
- Jammi NV, Whitby LR, Beal PA (2003) Small molecule inhibitors of the RNA-dependent protein kinase. *Biochem Biophys Res Commun* **308**: 50–57.
- Jansen RP, Dowzer C, Michaelis C, Galova M, Nasmyth K (1996) Mother cell-specific HO expression in budding yeast depends on the unconventional myosin myo4p and other cytoplasmic proteins. *CELL* **84**: 687–697.
- Johnstone TG, Bazzini AA, Giraldez AJ (2016) Upstream ORFs are prevalent translational repressors in vertebrates. *The EMBO Journal* 1–18.
- Jonhston DS, Driever W, Berleth T, Richstein S, Nüsslein-Volhard C (1989) Multiple steps in the localization of bicoid RNA to the anterior pole of the *Drosophila* oocyte. *Development* **107**: 13–19.
- Jovanovic M, Rooney MS, Mertins P, Przybylski D, Chevrier N, Satija R, Rodriguez EH, Fields AP, Schwartz S, Raychowdhury R, et al. (2015) Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**: 1259038.

- Jurica MS, Mesecar A, Heath PJ, Shi W, Nowak T, Stoddard BL (1998) The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure* **6**: 195–210.
- Kapp LD, Lorsch JR (2004) THE MOLECULAR MECHANICS OF EUKARYOTIC TRANSLATION. *Annu Rev Biochem* **73**: 657–704.
- Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, et al. (2012) Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *CELL* **149**: 753–767.
- Kauffman SA (1980) Heterotopic Transplantation in the Syncytial Blastoderm of the *Drosophila*: Evidence for Anterior and Posterior Nuclear Commitments. *Wilhelm Roux Archives of Developmental Biology* **189**: 135–145.
- Kazan H, Morris Q (2013) RBPmotif: a web server for the discovery of sequence and structure preferences of RNA-binding proteins. *Nucleic Acids Research* **41**: W180–W186.
- Kazan H, Ray D, Chan ET, Hughes TR, Morris Q (2010) RNAcontext: A New Method for Learning the Sequence and Structure Binding Preferences of RNA-Binding Proteins. *PLoS Comput Biol* **6**: e1000832–10.
- Kedersha N, Anderson P (2009) *Chapter 4 - Regulation of Translation by Stress Granules and Processing Bodies*. Elsevier Inc.
- Kedersha N, Chen S, Gilks N, Li W, Miller IJ, Stahl J, Anderson P (2002) Evidence that ternary complex (eIF2-GTP-tRNA(i)(Met))-deficient preinitiation complexes are core constituents of mammalian stress granules. *Mol Biol Cell* **13**: 195–210.

- Kedersha N, Cho MR, Li W, Yacono PW, Chen S, Gilks N, Golan DE, Anderson P (2000) Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. *The Journal of Cell Biology* **151**: 1257–1268.
- Kedersha N, Tisdale S, Hickman T, Anderson P (2008) *Chapter 26 - Real-Time and Quantitative Imaging of Mammalian Stress Granules and Processing Bodies*. Elsevier Inc.
- Keryer-Bibens C, Barreau C, Osborne HB (2012) Tethering of proteins to RNAs by bacteriophage proteins. *Biology of the Cell* **100**: 125–138.
- Khabar KSA (2014) Post-Transcriptional Control of Cytokine Gene Expression in Health and Disease. *Journal of Interferon & Cytokine Research* **34**: 215–219.
- Kim HJ, Kim NC, Wang Y-D, Scarborough EA, Moore J, Diaz Z, MacLea KS, Freibaum B, Li S, Molliex A, et al. (2013) Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**: 467–473.
- Kim JH, Lee S-R, Li L-H, Park H-J, Park J-H, Lee KY, Kim M-K, Shin BA, Choi S-Y (2011) High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. *PLoS ONE* **6**: e18556–e18558.
- Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, et al. (2014) A draft map of the human proteome. *Nature* **509**: 575–581.
- Kim VN (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* **6**: 376–385.
- Klämbt C, Schmidt O (1986) Developmental expression and tissue distribution of the lethal (2) giant larvae protein of *Drosophila melanogaster*. *The EMBO Journal* **5**: 2955–2961.

- König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**: 909–915.
- Krishnamoorthy T, Pavitt GD, Zhang F, Dever TE, Hinnebusch AG (2001) Tight Binding of the Phosphorylated Subunit of Initiation Factor 2 (eIF2) to the Regulatory Subunits of Guanine Nucleotide Exchange Factor eIF2B Is Required for Inhibition of Translation Initiation. *Molecular and Cellular Biology* **21**: 5018–5030.
- Kumar M, Gromiha MM, Raghava GPS (2011) SVM based prediction of RNA-binding proteins using binding residues and evolutionary information. *J Mol Recognit* **24**: 303–313.
- Kwon SC, Yi H, Eichelbaum K, Föhr S, Fischer B, You KT, Castello A, Krijgsveld J, Hentze MW, Kim VN (2013) The RNA-binding protein repertoire of embryonic stem cells. *Nature Structural & Molecular Biology* **20**: 1122–1130.
- Lareau LF, Hite DH, Hogan GJ, Brown PO (2014) Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife* **3**: 190–16.
- Lasko P (2012) mRNA Localization and Translational Control in *Drosophila* Oogenesis. *Cold Spring Harbor Perspectives in Biology* **4**: a012294–a012294.
- Lazzaretti D, Tournier I, Izaurralde E (2009) The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts independently of Argonaute proteins. *RNA* **15**: 1059–1066.
- Lee C-D, Tu BP (2015) Glucose-Regulated Phosphorylation of the PUF Protein Puf3 Regulates the Translational Fate of Its Bound mRNAs and Association with RNA Granules. *CellReports* **11**: 1–26.

- Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*. **75**: 843–854.
- Lee S, Liu B, Lee S, Huang S-X, Shen B, Qian S-B (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci USA* **109**: E2424–E2432.
- Lee Y-Y, Cevallos RC, Jan E (2009) An upstream open reading frame regulates translation of GADD34 during cellular stresses that induce eIF2alpha phosphorylation. *Journal of Biological Chemistry* **284**: 6661–6673.
- Lehmann R, Nüsslein-Volhard C (1986) Abdominal Segmentation, Pole Cell Formation, and Embryonic Polarity Require the Localized Activity of *oskar*, a Maternal Gene in *Drosophila*. *CELL* **47**: 141–152.
- Lehmann R, Nüsslein-Volhard C (1987a) *hunchback*, A Gene Required for Segmentation of an Anterior and Posterior Region of the *Drosophila* Embryo. *Developmental Biology* **119**: 402–417.
- Lehmann R, Nüsslein-Volhard C (1987b) Involvement of the *pumilio* gene in the transport of an abdominal signal in *Drosophila* embryo. *Nature* **329**: 167–170.
- Lehmann R, Nüsslein-Volhard C (1991) The maternal gene *nanos* has a central role in posterior pattern formation in the *Drosophila* embryo. *Development* **112**: 679–691.
- Lerner MR, Steitz JA (1979) Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci USA* **76**: 5495–5499.
- Lerner MR, Steitz JA, Boyle JA, Mount SM, Wolin SL (1980) Are snRNPs involved in splicing? **283**: 220–224.

- Li JJ, Bickel PJ, Biggin MD (2014) System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2**: e270–26.
- Li X, Kazan H, Lipshitz HD, Morris QD (2013) Finding the target sites of RNA-binding proteins. *WIREs RNA* **5**: 111–130.
- Li Y, Kowdley KV (2012) MicroRNAs in Common Human Diseases. *Genomics, Proteomics & Bioinformatics* **10**: 246–253.
- Li YR, King OD, Shorter J, Gitler AD (2013) Stress granules as crucibles of ALS pathogenesis. *The Journal of Cell Biology* **201**: 361–372.
- Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**: 464–469.
- Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Research* **31**: 3701–3708.
- Littlefield JW, Keller EB, Gross J, Zamecnik PC (1955) Studies on Cytoplasmic Ribonucleoprotein Particles from the Liver of the Rat. 1–15.
- Long RM, Gu W, Lorimer E, Singer RH, Chartrand P (2000) She2p is a novel RNA-binding protein that recruits the Myo4p-She3p complex to ASH1 mRNA. *The EMBO Journal* **19**: 6592–6601.
- Long RM, Gu W, Meng X, Gonsalvez G, Singer RH, Chartrand P (2001) An exclusively nuclear RNA-binding protein affects asymmetric localization of ASH1 mRNA and Ash1p in yeast. *The Journal of Cell Biology* **153**: 307–318.
- Long RM, Singer RH, Meng X, Gonzalez I, Nasmyth K, Jansen RP (1997) Mating type switching in yeast controlled by asymmetric localization of ASH1 mRNA. *Science* **277**: 383–387.

- Lu M, Lawrence DA, Marsters S, Acosta-Alvear D, Kimmig P, Mendez AS, Paton AW, Paton JC, Walter P, Ashkenazi A (2014) Opposing unfolded-protein-response signals converge on death receptor 5 to control apoptosis. *Science* **345**: 98–101.
- Lu PD, Harding HP, Ron D (2004) Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *The Journal of Cell Biology* **167**: 27–33.
- Lunde BM, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* **8**: 479–490.
- Ma X, Guo J, Sun X (2015) Sequence-Based Prediction of RNA-Binding Proteins Using Random Forest with Minimum Redundancy Maximum Relevance Feature Selection. *BioMed Research International* **2015**: 1–10.
- Ma XM, Blenis J (2009) Molecular mechanisms of mTOR-mediated translational control. *Nat Rev Mol Cell Biol* **10**: 307–318.
- MacDonald PM (1992) The *Drosophila* pumilio gene: an unusually long transcription unit and an unusual protein. *Development* **114**: 221–232.
- MacDonald PM, Leask A, Kerr K (1995) exl protein specifically binds BLE1, a bicoid mRNA localization element, and is required for one phase of its activity. *Proc Natl Acad Sci USA* **92**: 10787–10791.
- Mahowald AP, Hardy PA (1985) Genetics of *Drosophila* Embryogenesis. *Annual Review of Genetics* **19**: 149–177.
- Mandal S, Mandal A, Johansson HE, Orjalo AV, Park MH (2013) Depletion of cellular polyamines, spermidine and spermine, causes a total arrest in translation and growth in mammalian cells. *Proc Natl Acad Sci USA* **110**: 2169–2174.
- Maticzka D, Lange SJ, Costa F, Backofen R (2014) GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biology* **15**: R17–R18.

- Mattioli M, Reichlin M (1971) Characterization of a Soluble Nuclear Ribonucleoprotein Antigen Reactive with Sle Sera. 1–11.
- Mazroui R, Sukarieh R, Bordeleau M-E, Kaufman RJ, Northcote P, Tanaka J, Gallouzi I, Pelletier J (2006) Inhibition of ribosome recruitment induces stress granule formation independently of eukaryotic initiation factor 2alpha phosphorylation. *Mol Biol Cell* **17**: 4212–4219.
- McGeachy AM, Ingolia NT (2016) Starting too soon: upstream reading frames repress downstream translation. *The EMBO Journal*.
- Medenbach J, Seiler M, Hentze MW (2011) Translational Control via Protein-Regulated Upstream Open Reading Frames. *CELL* **145**: 902–913.
- Meyuhas O (2000) Synthesis of the translational apparatus is regulated at the translational level. *Eur J Biochem* **267**: 6321–6330.
- Milo R, Jorgensen P, Moran U, Weber G, Springer M (2010) BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Research* **38**: D750–D753.
- Mitchell SF, Jain S, She M, Parker R (2012) Global analysis of yeast mRNPs. *Nat Struct Mol Biol* **20**: 127–133.
- Mokas S, Mills JR, Garreau C, Fournier M-J, Robert F, Arya P, Kaufman RJ, Pelletier J, Mazroui R (2009) Uncoupling stress granule assembly and translation initiation inhibition. *Mol Biol Cell* **20**: 2673–2683.
- Mollet S, Cougot N, Wilczynska A, Dautry F, Kress M, Bertrand E, Weil D (2008) Translationally repressed mRNA transiently cycles through stress granules during stress. *Mol Biol Cell* **19**: 4469–4479.
- Müller A, MacCallum RM, Sternberg MJE (2002) Structural Characterization of the Human Proteome. *Genome Research* **12**: 1625–1641.

- Murata Y, Wharton RP (1995) Binding Pumilio to Maternal hunchback mRNA Is Required for Posterior Patterning in *Drosophila* Embryos. 1–10.
- Nott A, Le Hir H, Moore MJ (2004) Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes & Development* **18**: 210–222.
- Nudel U, Soreq H, Littauer UZ (1976) Globin mRNA species containing poly(A) segments of different lengths. Their functional stability in *Xenopus* oocytes. *Eur J Biochem* **64**: 115–121.
- Nüsslein-Volhard C, Frohnhofer HG, Lehmann R (1987) Determination of Anteroposterior Polarity in *Drosophila*. 1–8.
- Ozgur S, Chekulaeva M, Stoecklin G (2010) Human Pat1b Connects Deadenylation with mRNA Decapping and Controls the Assembly of Processing Bodies. *Molecular and Cellular Biology* **30**: 4308–4323.
- Padgett RA, Mount SM, Steitz JA, Sharp PA (1983) Splicing of Messenger RNA Precursors Is Inhibited by Antisera to Small Nuclear Ribonucleoprotein. 1–7.
- Palam LR, Baird TD, Wek RC (2011) Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. *J Biol Chem* **286**: 10939–10949.
- Paquin N, Ménade M, Poirier G, Donato D, Drouet E, Chartrand P (2007) Local activation of yeast ASH1 mRNA translation through phosphorylation of Khd1p by the casein kinase Yck1p. *Molecular Cell* **26**: 795–809.
- Parisi M, Lin H (1999) The *Drosophila* pumilio Gene Encodes Two Functional Protein Isoforms That Play Multiple Roles in Germline Development, Gonadogenesis, Oogenesis, and Embryogenesis. *Genetics* **153**: 235–250.
- Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degan B, Müller P, et al. (2000) Conservation of the sequence

- and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**: 86–89.
- Pavitt GD, Ramaiah KV, Kimball SR, Hinnebusch AG (1998) eIF2 independently binds two distinct eIF2B subcomplexes that catalyze and regulate guanine-nucleotide exchange. *Genes & Development* **12**: 514–526.
- Pavitt GD, Ron D (2012) New Insights into Translational Regulation in the Endoplasmic Reticulum Unfolded Protein Response. *Cold Spring Harbor Perspectives in Biology* **4**: a012278–a012278.
- Pavitt GD, Yang W, Hinnebusch AG (1997) Homologous segments in three subunits of the guanine nucleotide exchange factor eIF2B mediate translational regulation by phosphorylation of eIF2. *Molecular and Cellular Biology* **17**: 1298–1313.
- Penman S, Scherrer K, Becker Y, Darnell JE (1963) Polyribosomes in normal and poliovirus-infected HeLa cells and their relationship to messenger-RNA. *Proc Natl Acad Sci USA* **49**: 654–662.
- Quenault T, Lithgow T, Traven A (2011) PUF proteins: repression, activation and mRNA localization. *Trends in Cell Biology* **21**: 104–112.
- Rehwinkel J, Behm-Ansmant I, Gatfield D, Izaurralde E (2005) A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* **11**: 1640–1647.
- Reid DW, Chen Q, Tay ASL, Shenolikar S, Nicchitta CV (2014) The Unfolded Protein Response Triggers Selective mRNA Release from the Endoplasmic Reticulum. *CELL* **158**: 1362–1374.
- Reinhart BJ, Slack F, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901–906.

- Rheinberger H-J (1995) From Microsomes to Ribosomes: "Strategies" of 'Representation'. *Journal of the History of Biology* **28**: 49–89.
- Rogers J, Wall R (1980) A mechanism for RNA splicing. *PNAS* **77**: 1877–1879.
- Ron D, Walter P (2007) Signal integration in the endoplasmic reticulum unfolded protein response. *Nat Rev Mol Cell Biol* **8**: 519–529.
- Roux PP, Topisirovic I (2012) Regulation of mRNA Translation by Signaling Pathways. *Cold Spring Harbor Perspectives in Biology* **4**: a012252–a012252.
- Saghatelian A, Couso JP (2015) Discovery and characterization of smORF-encoded bioactive polypeptides. *Nature Chemical Biology* **11**: 909–916.
- Sanchez M, Galy B, Hentze MW, Muckenthaler MU (2007) Identification of target mRNAs of regulatory RNA-binding proteins using mRNP immunopurification and microarrays. *Nat Protoc* **2**: 2033–2042.
- Scheper W, Hoozemans JJM (2015) The unfolded protein response in neurodegenerative diseases: a neuropathological perspective. *Acta Neuropathologica* **130**: 315–331.
- SCHERRER K, LATHAM H, DARNELL JE (1963) Demonstration of an unstable RNA and of a precursor to ribosomal RNA in HeLa cells. *Proc Natl Acad Sci USA* **49**: 240–248.
- Scheuner D, Mierde DV, Song B, Flamez D, Creemers JWM, Tsukamoto K, Ribick M, Schuit FC, Kaufman RJ (2005) Control of mRNA translation preserves endoplasmic reticulum function in beta cells and maintains glucose homeostasis. *Nat Med* **11**: 757–764.
- Scheuner D, Song B, McEwen E, Liu C, Laybutt R, Gillespie P, Saunders T, Bonner-Weir S, Kaufman RJ (2001) Translational control is required for the unfolded protein response and in vivo glucose homeostasis. *Molecular Cell* **7**: 1165–1176.

- Schirle NT, Sheu-Gruttadauria J, MacRae IJ (2014) Structural basis for microRNA targeting. *Science* **346**: 608–613.
- Schoenberg DR, Maquat LE (2012) Regulation of cytoplasmic mRNA decay. *Nat Rev Genet* 1–15.
- Schupbach T, Wieschaus E (1986) Germline Autonomy of Maternal-Effect Mutations Altering the Embryonic Body Pattern of *Drosophila*. *Developmental Biology* **113**: 443–448.
- Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2012) Global quantification of mammalian gene expression control. *Nature* **473**: 337–342.
- Schwerk J, Jarret AP, Joslyn RC, Savan R (2015) Landscape of post-transcriptional gene regulation during hepatitis C virus infection. *Current Opinion in Virology* **12**: 75–84.
- Shahbadian K, Jeronimo C, Forget A, Robert F, Chartrand P (2014) Co-transcriptional recruitment of Puf6 by She2 couples translational repression to mRNA localization. *Nucleic Acids Research* **42**: 8692–8704.
- Shao X, Tian Y, Wu L, Wang Y, Jing L, Deng N (2009) Predicting DNA- and RNA-binding proteins from sequences with kernel methods. *Journal of Theoretical Biology* **258**: 289–293.
- Sharif H, Ozgur S, Sharma K, Basquin C, Urlaub H, Conti E (2013) Structural analysis of the yeast Dhh1-Pat1 complex reveals how Dhh1 engages Pat1, Edc3 and RNA in mutually exclusive interactions. *Nucleic Acids Research* **41**: 8377–8390.
- Shirai Y-T, Suzuki T, Morita M, Takahashi A, Yamamoto T (2014) Multifunctional roles of the mammalian CCR4--NOT complex in physiological phenomena. *Frontiers in Genetics* **5**: 1–11.

- Shiratori A, Shibata T, Arisawa M, Hanaoka F, Murakami Y, Eki T (1999) Systematic identification, classification, and characterization of the open reading frames which encode novel helicase-related proteins in *Saccharomyces cerevisiae* by gene disruption and Northern analysis. *Yeast* **15**: 219–253.
- Si J, Cui J, Cheng J, Wu R (2015) Computational Prediction of RNA-Binding Proteins and Binding Sites. *IJMS* **16**: 26303–26317.
- Sidrauski C, Acosta-Alvear D, Khoutorsky A, Vedantham P, Hearn BR, Li H, Gamache K, Gallagher CM, Ang KK-H, Wilson C, et al. (2013) Pharmacological brake-release of mRNA translation enhances cognitive memory. *eLife* **2**: 7193–22.
- Sidrauski C, McGeachy AM, Ingolia NT, Walter P (2015) The small molecule ISRIB reverses the effects of eIF2 α phosphorylation on translation and stress granule assembly. *eLife* **4**:
- Sil A, Herskowitz I (1996) Identification of asymmetrically localized determinant, Ash1p, required for lineage-specific transcription of the yeast HO gene. *CELL* **84**: 711–722.
- Sonenberg N, Hinnebusch AG (2009) Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *CELL* **136**: 731–745.
- Song Z (2013) Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. *Molecular Aspects of Medicine* **34**: 590–600.
- Sonoda J, Wharton RP (2001) *Drosophila* Brain Tumor is a translational repressor. 1–12.
- Sorgeloos P, Van Outryve E, Persoone G, Cattoir-Reynaerts A (1976) New type of turbidostat with intermittent determination of cell density outside the culture vessel. *Applied and Environmental Microbiology* **31**: 327–331.
- Standart N, Marnef A (2012) Pat1 proteins: regulation mRNAs from birth to death? *BioMolecular Concepts* **3**: 1–13.

- Stinton LM, Eystathioy T, Selak S, Chan EKL, Fritzler MJ (2004) Autoantibodies to protein transport and messenger RNA processing pathways: endosomes, lysosomes, Golgi complex, proteasomes, assemblyosomes, exosomes, and GW bodies. *Clinical Immunology* **110**: 30–44.
- Strein C, Alleaume A-M, Rothbauer U, Hentze MW, Castello A (2014) A versatile assay for RNA-binding proteins in living cells. *RNA* **20**: 721–731.
- Tabas I, Ron D (2011) Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress. *Nature Cell Biology* **13**: 184–190.
- Takahashi CN, Miller AW, Ekness F, Dunham MJ, Klavins E (2015) A Low Cost, Customizable Turbidostat for Use in Synthetic Circuit Characterization. *ACS Synth Biol* **4**: 32–38.
- Takizawa PA, Sil A, Swedlow JR, Herskowitz I, Vale RD (1997) Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast. 1–4.
- Takizawa PA, Vale RD (2000) The myosin motor, Myo4p, binds Ash1 mRNA via the adapter protein, She3p. *Proc Natl Acad Sci USA* **97**: 5273–5278.
- Tang H, Hornstein E, Stolovich M, Levy G, Livingstone M, Templeton D, Avruch J, Meyuhas O (2001) Amino Acid-Induced Translation of TOP mRNAs Is Fully Dependent on Phosphatidylinositol 3-Kinase-Mediated Signaling, Is Partially Inhibited by Rapamycin, and Is Independent of S6K1 and rpS6 Phosphorylation. *Molecular and Cellular Biology* **21**: 8671–8683.
- Tautz D (1988) Regulation of the *Drosophila* segmentation gene hunchback by two maternal morphogenetic centers. *Nature* **332**: 1–4.
- Tautz D, Pfeifle C (1989) A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene hunchback. *Chromosoma* **98**: 81–85.

- Temme C, Simonelig M, Wahler E (2014) Deadenylation of mRNA by the CCR4–NOT complex in *Drosophila*: molecular and developmental aspects. **5**: 1–11.
- Tenebaum SA, Carson CC, Lager PJ, Keene JD (2000) Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *PNAS* **97**: 14085–14090.
- The Gene Ontology Consortium (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Research* **43**: D1049–D1056.
- Thedieck K, Holzwarth B, Prentzell MT, Boehlke C, Kläsener K, Ruf S, Sonntag AG, Maerz L, Grellscheid S-N, Kremmer E, et al. (2013) Inhibition of mTORC1 by Astrin and Stress Granules Prevents Apoptosis in Cancer Cells. *CELL* **154**: 859–874.
- Thore S, Mauxion F, Séraphin B, Suck D (2003) X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. *EMBO reports* **4**: 1150–1155.
- Thoreen CC, Chantranupong L, Keys HR, Wang T, Gray NS, Sabatini DM (2012) A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **486**: 109–113.
- Tkach JM, Yimit A, Lee AY, Riffle M, Costanzo M, Jaschob D, Hendry JA, Ou J, Moffat J, Boone C, et al. (2012) Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nature Cell Biology* **14**: 966–976.
- Tonegawa S, Baldi I (1973) Electrophoretically homogeneous myeloma light chain mRNA and its translation in vitro. *Biochem Biophys Res Commun* **51**: 81–87.
- Toprak E, Veres A, Yildiz S, Pedraza JM, Chait R, Paulsson J, Kishony R (2013) Building a morbidostat: an automated continuous- culture device for

- studying bacterial drug resistance under dynamically sustained drug inhibition. *Nat Protoc* **8**: 555–567.
- Tritschler F, Huntzinger E, Izaurralde E (2010) Role of GW182 proteins and PABPC1 in the miRNA pathway: a sense of déjà vu. 1–6.
- Tsuboi T, Inada T (2010) Tethering of Poly(A)-binding Protein Interferes with Non-translated mRNA Decay from the 5' End in Yeast. *Journal of Biological Chemistry* **285**: 33589–33601.
- Tsvetanova NG, Klass DM, Salzman J, Brown PO (2010) Proteome-Wide Search Reveals Unexpected RNA-Binding Proteins in *Saccharomyces cerevisiae*. *PLoS ONE* **5**: e12671.
- Tuck AC, Tollervey D (2013) A Transcriptome-wide Atlas of RNP Composition Reveals Diverse Classes of mRNAs and lncRNAs. *CELL* **154**: 996–1009.
- Tuerk C, Gold L (1990) Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* **249**: 505–510.
- Ule J (2003) CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science* **302**: 1212–1215.
- Vanderweyde T, Yu H, Varnum M, Liu-Yesucevitz L, Citro A, Ikezu T, Duff K, Wolozin B (2012) Contrasting Pathology of the Stress Granule Proteins TIA-1 and G3BP in Tauopathies. *Journal of Neuroscience* **32**: 8270–8283.
- Vandewynckel Y-P, Laukens D, Geerts A, Bogaerts E, Paridaens A, Verhelst X, Janssens S, Heindryckx F, Van Vlierberghe H (2013) The paradox of the unfolded protein response in cancer. *Anticancer Res* **33**: 4683–4694.
- Vattem KM, Wek RC (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci USA* **101**: 11269–11274.

- Vazquez de Aldana CR, Hinnebusch AG (1994) Mutations in the GCD7 subunit of yeast guanine nucleotide exchange factor eIF-2B overcome the inhibitory effects of phosphorylated eIF-2 on translation initiation. *Molecular and Cellular Biology* **14**: 3208–3222.
- Walter P, Blobel G (1982) Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* **299**: 691–698.
- Walter P, Ibrahimi I, Blobel G (1981) Translocation of Proteins Across the Endoplasmic Reticulum. *The Journal of Cell Biology* **91**: 545–550.
- Walter P, Lingappa VR (1986) Mechanism of Protein Translocation Across the Endoplasmic Reticulum Membrane. *Annual Review of Cell Biology* **2**: 499–516.
- Walter P, Ron D (2011) The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **334**: 1081–1086.
- Wang C, Dickinson LK, Lehmann R (1994) Genetics of nanos localization in *Drosophila*. *Developmental Dynamics* **199**: 103–115.
- WATSON JD, CRICK FHC (1953) Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**: 737–738.
- Wek RC, Jiang H-Y, Anthony TG (2006) Coping with stress: eIF2 kinases and translational control. *Biochem Soc Trans* **34**: 7–11.
- Wharton RP, Struhl G (1991) RNA Regulatory Elements Mediate Control of *Drosophila* Body Pattern by the Posterior Morphogen nanos. *CELL* **67**: 955–967.
- Wickens MP, Bernstein DS, Kimble J, Parker R (2002) A PUF family portrait: 3' UTR regulation as a way of life. 1–8.

- Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional Regulation of the Heterochronic Gene *lin-14* by *lin-4* Mediates Temporal Pattern Formation in *C. elegans*. *CELL* **75**: 855–862.
- Winter J, Jung S, Keller S, Gregory RI, Diederichs S (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology* **11**: 228–234.
- Wippich F, Bodenmiller B, Trajkovska MG, Wanka S, Aebersold R, Pelkmans L (2013) Dual Specificity Kinase DYRK3 Couples Stress Granule Condensation/Dissolution to mTORC1 Signaling. *CELL* **152**: 791–805.
- Woo CW, Kutzler L, Kimball SR, Tabas I (2012) Toll-like receptor activation suppresses ER stress factor CHOP and translation inhibition through activation of eIF2B. *Nature Cell Biology* **14**: 192–200.
- Wreden C, Verrotti AC, Schisa JA, Lieberfarb MA, Strickland S (1997) Nanos and pumilio establish embryonic polarity in *Drosophila* by promoting posterior deadenylation of hunchback mRNA. 1–9.
- Xu Y-F, Zhao X, Glass DS, Absalan F, Perlman DH, Broach JR, Rabinowitz JD (2012) Regulation of Yeast Pyruvate Kinase by Ultrasensitive Allostery Independent of Phosphorylation. *Molecular Cell* **48**: 52–62.
- Yamada M, Hayatsu N, Matsuura A, Ishikawa F (1998) Y'-Help1, a DNA helicase encoded by the yeast subtelomeric Y' element, is induced in survivors defective for telomerase. *Journal of Biological Chemistry* **273**: 33360–33366.
- Ye J, Rawson RB, Komuro R, Chen X, Davé UP, Prywes R, Brown MS, Goldstein JL (2000) ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Molecular Cell* **6**: 1355–1364.

- Yoshida H, Matsui T, Yamamoto A, Okada T, Mori K (2001) XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *CELL* **107**: 881–891.
- Yu B (2005) Methylation as a Crucial Step in Plant microRNA Biogenesis. *Science* **307**: 932–935.
- Yu L, Kelly U, Ebright JN, Malek G, Saloupis P, Rickman DW, McKay BS, Arshavsky VY, Bowes Rickman C (2007) Oxidative stress-induced expression and modulation of Phosphatase of Regenerating Liver-1 (PRL-1) in mammalian retina. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1773**: 1473–1482.
- Zalokar M (1976) Autoradiographic study of protein and RNA formation during early development of *Drosophila* eggs. *Developmental Biology* **49**: 425–437.
- Zamore PD, Williamson JR, Lehmann R (1997) The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *RNA* **3**: 1421–1433.
- Zekri L, Huntzinger E, Heimstadt S, Izaurralde E (2009) The Silencing Domain of GW182 Interacts with PABPC1 To Promote Translational Repression and Degradation of MicroRNA Targets and Is Required for Target Release. *Molecular and Cellular Biology* **29**: 6220–6231.
- Zhou D, Palam LR, Jiang L, Narasimhan J, Staschke KA, Wek RC (2008) Phosphorylation of eIF2 directs ATF5 translational control in response to diverse stress conditions. *Journal of Biological Chemistry* **283**: 7064–7073.
- Zhu PJ, Huang W, Kalikulov D, Yoo JW, Placzek AN, Stoica L, Zhou H, Bell JC, Friedlander MJ, Krnjević K, et al. (2011) Suppression of PKR Promotes Network Excitability and Enhanced Cognition by Interferon- γ -Mediated Disinhibition. *CELL* **147**: 1384–1396.

Zipprich JT, Bhattacharyya S, Mathys H, Filipowicz W (2009) Importance of the C-terminal domain of the human GW182 protein TNRC6C for translational repression. *RNA* **15**: 781–793.

Acknowledgements

I, the thesis writer, Anna McGeachy, would like to thank several individuals for their assistance in the production of the enclosed work. In particular, Zuriyah Meacham and Nicholas Ingolia provided wet lab assistance and mentorship respectively for the content in Chapters 2 and 3. I would like to thank Carmela Sidrauski and Peter Walter for including me in the collaboration over the UPR and ISRIB presented in Chapter 5. I would also like to thank my official thesis readers, Nicholas Ingolia and Joseph Gall, for their input and editing. My thanks also goes to my unofficial reader and editor, Brian Farley, for his input and guidance. Of course, no thesis is written in isolation, and I thank my significant other, Derek Reeve, for his patience, willingness to move multiple times, and unending support.

Curriculum Vitae

Anna McGeachy

Born: Phoenix, AZ

February 25, 1988

Barker Hall Rm 416

224.213.1825

University of California Berkeley

anna.mcgeachy@gmail.com

Barker Hall

Berkeley, CA 94720

Positions Graduate Student August 2010 – present

Johns Hopkins University,

Program in Cell, Molecular, Developmental Biology, and Biophysics

Carnegie Institution, Dept. of Embryology

Visiting Student Researcher

November 2013 - present

University of California Berkeley, Dept. of Molecular and Cell Biology

Education **Johns Hopkins University** Ph.D., Biology, June 2016 (expected)

University of Chicago B.A., Biology: Genetics, June 2010

Research **Johns Hopkins University-UC Berkeley** May 2010 – present

Experience Graduate student researcher, Dept. of Cell, Mol., Dev., and Biophys.

Advisor: *Nicholas Ingolia*

Genome-wide survey for post-transcriptional regulators. Discovery of novel mechanisms controlling translation and RNA decay by adapting tethered function assay to a high-throughput, unbiased format in *S. cerevisiae* using flow cytometry and high-throughput sequencing.

University of Chicago

Winter 2009 – August 2010

Undergraduate research assistant, Dept. of Ecology and Evolution

Advisor: *Martin Kreitman*

Investigate hidden genetic variance in a complex, robust developmental phenotype using artificial perturbation of naturally occurring variable egg size in wild-caught *D. melanogaster* within a segmentation patterning (specifically, giant and eve)

University of Chicago

Fall 2007 – Fall 2008

Undergraduate research assistant, Ben May Cancer Institute

Advisor: *Akira Imamoto*

Studies on DiGeorges Syndrome (DGS) in C57BL6 *M. musculus*; establishment of an ERK2 knockout line and maintenance of collaborator-provided neo-cassette ERK1 lines for investigation of interaction with lab genes of interest TBX1, Crk, Crkl

Oklahoma Medical Research Foundation

Summer 2007, 2008

Undergraduate student researcher, Arthritis and Immunology

Advisor: *Joshua Ojwang*

Assisted in sample prep, blood processing and DNA extractions for studies within the Arthritis and Immunology Department; contributed in the progression of SLEGEN and large-scale SNP association studies

Oklahoma Medical Research Foundation

Summer 2006

Sir Alexander Fleming Scholar, Arthritis and Immunology

Advisor: *John B. Harley*

Investigating genetic contributions to susceptibility in Systemic Lupus Erythematosus (SLE), focusing on eukaryotic initiation factor 3 subunit M (eIF3M) and isolated SNPs on 11p13

Oklahoma City Community College

Spring 2006

Assistant Technician, Biotechnology Program

Advisors: *Charlotte Mulvihill and Don Bell*

Research, Development and Production for Biotechnology programs area high schools

Fellowships and DuPont Teaching Award for Excellence in Undergraduate Education

(2012)

Honors

DuPont Teaching Award for Excellence in Undergraduate Education

(2011)

Sir Alexander Fleming Scholar (2006)

Betty S. and David C. Hess Scholarship (2008-2010)

Publications

Sidrauski, C., **McGeachy, A. M.**, Ingolia, N. T., & Walter, P. (2015).

The small molecule ISRIB reverses the effects of eIF2 α

phosphorylation on translation and stress granule assembly. *ELife*, 4.

doi:10.7554/eLife.05033

Ingolia NT, Brar GA, Rouskin S, **McGeachy AM**, Weissman JS.

Genome-Wide Annotation and Quantitation of Translation by

Ribosome Profiling. *Current Protocols in Molecular Biology*, 4.18.

1-4.18. 19

Ingolia NT, Brar GA, Rouskin S, **McGeachy AM**, Weissman JS. The

ribosome profiling strategy for monitoring translation in vivo by deep

sequencing of ribosome-protected mRNA fragments. *Nat Protoc*.

2012

Jul 26;7(8):1534-50.

Invited Talks

Genome-wide survey for post-transcriptional regulators in

Saccharomyces cerevisiae

Bay Area RNA Club 2015. 2015 July 9.

Bulk selection for diverse yeast expression libraries using large

volume turbidostats

UC Berkeley Microbial Student Symposium. 2015 April 25.

Analysis of sequencing data using Generalized Linear Models (GLMs)

Carnegie Institution of Science Nitty Gritty Workflow meeting.

2013 Oct 10.

Posters

Comprehensive profiling of the eIF2 α -mediated unfolded protein response.

EMBO Conference on Protein Synthesis and Translational Control

Heidelberg, Germany

References

Nicholas Ingolia

Assistant Professor, Molecular and Cell Biology,

University of California, Berkeley

Joe Gall

Staff Member (PI), Carnegie Institution, Department of Embryology

Adjunct Professor, Biology, Johns Hopkins University

Rachel Green

Professor, Molecular Biology and Genetics,

Johns Hopkins Medical Institute

Investigator, Howard Hughes Medical Institute

Liana Lareau

Fellow (PI), California Institute for Quantitative Biosciences,
University of California, Berkeley

Peter Walters and Carmela Sidrauski

Principal Investigator, University of California, San Francisco (Peter)

Investigator, Howard Hughes Medical Institute (Peter)

Scientist, Calico Labs (Carmela)