Regulation of gene expression by RNA binding proteins and microRNAs

Kyle Cottrell
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds
Part of the Biochemistry Commons, Cell Biology Commons, and the Molecular Biology Commons

Recommended Citation
https://openscholarship.wustl.edu/art_sci_etds/1192

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Dissertation Examination Committee:
Sergej Djuranovic, Chair
Kathleen Hall
Tim Schedl
Sheila Stewart
Andrew Yoo

Regulation of gene expression by RNA binding proteins and microRNAs
by
Kyle Aaron Cottrell

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2017
St. Louis, Missouri
# Table of Contents

List of Figures........................................................................................................................................ vii
List of Tables ........................................................................................................................................... ix
Acknowledgments .................................................................................................................................... x
Abstract................................................................................................................................................ xii

Chapter 1: Introduction ............................................................................................................................ 1
  1.1 The messenger RNA ................................................................. 1
  1.2 Translation ............................................................................ 3
  1.3 Decay .................................................................................. 5
  1.4 Post-transcriptional Regulation .............................................. 5
    1.4.1 microRNAs ................................................................. 5
    1.4.2 RNA binding proteins .................................................. 8
  1.5 Conclusions .......................................................................... 11
  1.6 References .......................................................................... 13

Chapter 2: Urb-RIP – an adaptable and efficient approach for immunoprecipitation of RNAs and associated RNAs/proteins .................................................................................................................. 22
  Preface .................................................................................... 22
  2.1 Abstract ............................................................................. 23
  2.2 Introduction ........................................................................ 23
  2.3 Materials and Methods ......................................................... 26
    2.3.1 Construction of 2HA-Urb and SLII-tagged RNA constructs .................................................. 26
    2.3.2 Cell culture and transfection ......................................... 27
    2.3.4 Crosslinking and immunoprecipitation of SLII-tagged RNA (Urb-RIP) ............................ 28
    2.3.5 RNA analysis by RT-qPCR ........................................... 30
    2.3.6 Protein analysis by western blotting ........................................ 30
  2.4 Results .............................................................................. 31
    2.4.1 General description of Urb-RIP ..................................... 31
    2.4.2 Urb-RIP allows for enrichment of mRNAs ...................... 33
    2.4.3 Urb-RIP enriches for tagged mRNA and a bound RNA binding protein .......................... 35
5.3 References ........................................................................................................................................... 191
List of Figures

Figure 2.1: Schematic illustration of the Urb-RIP protocol and potential applications................. 31
Figure 2.2: Urb-RIP enriches for mCherry-mRNA and bound PABP ........................................ 34
Figure 2.3: Urb-RIP shows Argonaute and miRNA binding to miRNA-targeted messages in human cells .................................................................................................................. 37
Figure 2.4: Urb-RIP confirms binding of PABP to BC200.............................................................. 38
Supplemental Figure 2.1: Optimization of pulldown protocol ...................................................... 52
Supplemental Figure 2.2: RNA integrity after UV irradiation ...................................................... 53
Supplemental Figure 2.3: Optimization of blocking and preclearing .......................................... 54
Supplemental Figure 2.4: Analysis of U1-snRNA binding during Urb-RIP .................................... 55
Supplemental Figure 2.5: Reproducibility of Urb-RIP ................................................................. 56
Supplemental Figure 2.6: Analysis of RNA Pulldown by Bioanalyzer ......................................... 57
Figure 3.1: 3’UTR influences translatability and miRNA-mediated repression ......................... 67
Figure 3.2: Codon optimality and 5’UTR elements modulate miRNA-mediated repression ...... 70
Figure 3.3: Translation efficiency is a determinant of the magnitude of miRNA-mediated translational repression but not RNA degradation of endogenous miRNA targets ............. 73
Figure 3.4: Model - Translation efficiency and mRNA elements influence the magnitude of miRNA-mediated repression ........................................................................................................ 79
Supplemental Figure 3.1: RNA decay and 3’UTR length of Renilla Luciferase Reporters ...... 97
Supplemental Figure 3.2: Bantam antagomir abrogates repression of 3’UTR reporters .......... 98
Supplemental Figure 3.3: RNA abundance correlates with non-targeted reporter protein expression ........................................................................................................................................ 99
Supplemental Figure 3.4: Reporter transcripts terminated by the histone H3 stem loop show correlation between repression and non-targeted reporter expression ........................................... 100
Supplemental Figure 3.5: Reporters with moderate codon optimality show robust repression . 101
Supplemental Figure 3.6: 5’UTR structure and uORF influence non-targeted and targeted reporters differentially .................................................................................................................. 102
Supplemental Figure 3.7: Statistical analysis of the interaction between TE and translational repression ...................................................................................................................................... 103
Supplemental Figure 3.8: TE influences the magnitude of repression by miR-155 independent of the seed pairing ............................................................................................................. 104
Supplemental Figure 3.9: TE influences the magnitude of repression by miR-1 ................................ 105
Supplemental Figure 3.10: TE along with mRNA abundance can predict fold change of RPF. 106
Supplemental Figure 3.11: Correlation between TE and RPF fold change for miR-155 targets in LPS activated B-cells ........................................................................................................... 107
Supplemental Figure 3.12: Comparison of distributions of TE multiplied by RPF Fold Change across groups and time points ............................................................................................... 108
List of Tables

<table>
<thead>
<tr>
<th>Table Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplemental Table 2.1: Cloning primers and oligonucleotides</td>
<td>58</td>
</tr>
<tr>
<td>Supplemental Table 2.2: qPCR and Reverse Transcription Priemrs</td>
<td>59</td>
</tr>
<tr>
<td>Supplemental Table 2.3: Pulldown of Non-target RNAs in mCherry Pulldown</td>
<td>60</td>
</tr>
<tr>
<td>Supplemental Table 2.4: Enrichment of mCherry mRNA by Urb-RIP</td>
<td>61</td>
</tr>
<tr>
<td>Supplemental Table 2.5: Pulldown of Non-target RNAs in BC200 Pulldown</td>
<td>62</td>
</tr>
<tr>
<td>Supplemental Table 3.1: Primers and oligonucleotides</td>
<td>115</td>
</tr>
<tr>
<td>Supplemental Table 3.2: Expression of miRNAs in S2 cells</td>
<td>117</td>
</tr>
<tr>
<td>Supplemental Table 3: Primers for miRNA qRT-PCR</td>
<td>118</td>
</tr>
<tr>
<td>Supplemental Table 4.1: Primers and Oligonucleotides</td>
<td>179</td>
</tr>
<tr>
<td>Supplemental Table 4.2: Binding sites for RBP and miRNA</td>
<td>181</td>
</tr>
</tbody>
</table>
Acknowledgments

Science is a community effort. Each scientist relies on the work of others as a foundation for their own projects. Within our labs we rely on our lab-mates for support and our mentors for guidance. We rely on the input of our research community to strengthen our work.

I would like to thank my lab-mates: Jessey Erath and Laura Arthur, not just for their support of my research but for at times providing a distraction from the day-to-day stresses of the laboratory. I would like to thank the undergraduate students (Denise Rogers) and rotation students who have contributed to my projects (Yansel Nunez and Kellan Weston). I would like to thank my lab manager Slavica Pavlovic-Djurunic not just for keeping the lab running and stocked, but for being there to talk about life outside of lab. I would like to thank my PI, Sergej Djuranovic, for guiding my research when necessary but also for giving me the freedom to control the direction of my projects. Outside of the Djuranovic lab I would like to thank the many members of the Washington University community that I have interacted with through the years. Whenever I have needed a reagent or an expert opinion on some technique someone was always willing to help. I would like to thank the members of my thesis committee for providing constructive feedback during each of my updates. I want to thank Barak Cohen and the members of his lab, in particular Hemangi Chaudhari. Barak and Hemangi were instrumental in one of the projects described here. Of course, none of this research would have been possible without our funding sources: GM007067 and GM112824.

Finally, I want to thank my wife, Monica. Her support has been essential throughout my time as a graduate student.

Kyle Cottrell

Washington University in St. Louis

December 2017
Dedicated to Lyla, my daughter
ABSTRACT OF THE DISSERTATION

Regulation of gene expression by RNA binding proteins and microRNAs

by

Kyle Aaron Cottrell

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Cell Biology

Washington University in St. Louis, 2017

Dr. Sergej Djuranovic, Chair

Regulation of gene expression is essential to life. Post-transcriptional regulation of gene expression is a complex process with many inputs that lead to changes in localization, translation and stability of mRNAs. The translation and stability of many mRNAs is regulated by cis-elements, such as mRNA-structure or codon optimality; and by trans-acting factors such as RBPs and miRNAs. Here I report on the complex interactions between RBPs, miRNAs and characteristics of their target mRNAs in respect to effects on translation and RNA stability.

Using a reporter based approach we studied modulation of microRNA-mediated repression by various mRNA characteristics. We observed the influence of codon optimality, 5’UTR structure, uORFs and translation efficiency on the magnitude of miRNA-mediated repression. To study functional interactions between RBPs and miRNAs, we developed a new method: PTRE-seq. This method utilizes a massively parallel reporter library to study the individual and combined effects of RBPs and miRNAs on translation and RNA stability. Using PTRE-seq we observed epistatic interactions between AU-rich elements and miRNA binding sites. In addition to PTRE-seq, we developed a novel method for immunoprecipitation of mRNAs that will facilitate the identification of miRNAs and RBPs bound to mRNAs of interest.
Chapter 1: Introduction

1.1 The messenger RNA

Messenger RNAs (mRNA) carry genetic information required for protein synthesis. The transcription of each mRNA is tightly controlled by transcription factors and chromatin state to maintain the appropriate amount of gene product. In addition to transcriptional control of gene expression, mRNAs are also regulated post-transcriptionally and the protein product is regulated post-translationally. The structure of the mRNA can be broken down into three parts: the 5’-untranslated region (5’UTR, sometimes referred to as the transcript leader ¹), the coding sequence, and the 3’UTR. The coding sequence possesses the sequence information that will be translated into a polypeptide via the translation machinery. Coding sequences are made up of an open reading frame, a set of triplet codons beginning with the start codon (AUG) and a stop codon (UAG, UGA or UAA in most organisms). The 5’ and 3’UTRs of the mRNA possess regulatory elements that control the localization, translation and degradation of the mRNA.

Eukaryotic mRNAs have evolved key structural and sequence elements that facilitate the localization, translation and degradation of the mRNA. All mRNAs are transcribed by RNA polymerase II (RNAP II). A hallmark of RNAP II transcripts is the presence of a 5’ cap. The cap consists of a modified nucleotide, 7-methyl guanosine, which is affixed to the 5’ of the mRNA through a 5’ – 5’ triphosphate linkage. The cap serves an important role in protecting the mRNA from degradation by exonucleases but also as a binding site for proteins that regulate splicing, nuclear export and translation ². Decapping of the mRNA is required for 5’ – 3’ exonucleatic degradation of the mRNA ³.
Like the 5’ end of the mRNA, the 3’ end is also protected. This protection comes in the form of a long tract of adenosine nucleotides, known as the poly(A)-tail. This tail is added to mRNAs co-transcriptionally by a collection of proteins, including cleavage and polyadenylation specificity factor (CPSF), that detect the polyadenylation signal and promote cleavage and the processive addition of adenosines to the 3’ end of the mRNA by poly(A)-polymerase (PAP). PAP adds 250 adenosines to the 3’ end of the mRNA before it stops due to a loss of interaction with CPSF. The poly(A)-tail serves several purposes: protects the mRNA from degradation, facilitates nuclear export and promotes translation. The functions of the poly(A)-tail are carried out largely through poly(A)-binding proteins (PABP).

Along with processing of the 5’ and 3’ ends of the mRNA during transcription, in many higher eukaryotes a significant amount of information must be removed from the pre-mRNA. Pre-mRNAs contain both introns and exons. Introns generally do not contain coding information and are thus removed via a process known as mRNA splicing. Splicing is carried out by a large complex of RNAs and proteins known as the spliceosome. For some mRNAs alternative splicing of some introns or exons can lead to transcript variants. These transcript variants could contain different coding sequences, or UTRs. As such, transcript variants often contain different regulatory elements in the coding sequence or UTRs that may affect localization, translation or degradation of the mRNA. Beyond removing introns, splicing plays an important role in facilitating mRNA export and surveillance of abnormal mRNAs. Both of these processes are carried out by the exon-junction complex (EJC). The EJC is a large protein complex that remains bound to the mRNA at the site of an exon-exon junction following splicing. Some proteins within the EJC facilitate export of the mRNA from the nucleus. The EJC is also used as a
reference point by the RNA surveillance pathway known as non-sense mediated decay to determine if a stop-codon is premature \(^\text{10}\).

### 1.2 Translation

Eukaryotic translation can be broken into three phases: initiation, elongation and termination. The initiation phase begins with the recruitment of the pre-initiation complex (PIC) to the mRNA \(^\text{11}\). The PIC is made up of the small subunit of the ribosome; the ternary complex which contains eIF2, GTP and the initiator tRNA; and the initiation factors eIF1, eIF1A, eIF3 and eIF5. The PIC is recruited to the mRNA by the eIF4F complex. The eIF4F complex is made up of the cap binding protein eIF4E, the helicase eIF4A and the scaffold protein eIF4G. Loading of the eIF4F complex is facilitated by the protein eIF4B which activates the helicase activity of eIF4A. The helicase activity of eIF4A opens secondary structure near the cap to allow for binding of the eIF4F complex and recruitment of the PIC. Once the PIC is recruited, it scans the 5’UTR in search of a start codon. Scanning is ATP dependent and thought to be facilitated by the helicase eIF4A \(^\text{12}\) but other helicases likely contribute \(^\text{13}\). Once a start codon is found through base-pairing with the initiator tRNA, the GTP bound to eIF2 is hydrolyzed which results in a conformation change of the complex. The large subunit is then recruited by eIF5B-GTP and many of the initiation factors are released. Hydrolysis of GTP causes release of eIF5B and eIF1A, leaving the completed ribosome with the initiator tRNA. Sequences within the 5’UTR can reduce the translation efficiency of the mRNA. For instance, stable secondary structures reduce translation \(^\text{14}\). The presence of start codons upstream of the main ORF start codon can reduce initiation of the main ORF \(^\text{15}\). Besides uORFs and mRNA structure, several mRNA trans-acting factors have been shown to inhibit translation initiation in order to control gene expression \(^\text{16}\).
Following initiation, the ribosome along with the elongation factors and charged tRNAs faithfully incorporate amino acids into the growing polypeptide. The ribosome has three tRNA binding sites: the aminoacyl-tRNA site (A), the peptidyl-tRNA site (P) and the tRNA exit site (E). After initiation the initiator tRNA is located within the P site. Aminoacyl-tRNAs in complex with eEF1A and GTP sample the codon within the A site through codon-anticodon base pairing. When the cognate codon is recognized, GTP is hydrolyzed and eEF1A-GDP is dissociated. The peptidyl transferase center of the ribosome catalyzes peptide bond formation. The elongation factor eEF2 then translocates the peptidyl tRNA to the P site through GTP hydrolysis. The rate of elongation is dependent on the codon usage within the open reading frame. Recently several studies have revealed a correlation between efficient codon usage and mRNA stability. Besides codon usage, the protein FRMP regulates elongation in order to control gene expression.

For most eukaryotes there are three stop codons: UAG, UGA and UAA. The stop codon is recognized not by a tRNA but by a protein, eukaryotic release factor 1 (eRF1). eRF1 recognizes all three stop codons. Together with the GTPase eRF2, eRF1 promotes the release of the nascent polypeptide in a GTP dependent manner. Following release of the nascent peptide the ribosomal subunits dissociate.

mRNAs can be translated repeatedly until they are degraded. Typically, mRNAs are translated by more than one ribosome at a time, producing what is referred to as a polysome. To facilitate this process, some mRNAs form a “closed-loop structure”. The closed loop forms through interactions of PABP, bound at the 3’ end of the mRNA, and eIF4G, bound to the cap binding protein at the 5’ end of the mRNA. This structure is thought to facilitate the local
recycling of ribosomes \(^{26}\). Some mRNA \textit{trans}-acting factors are thought to target this structure to reduce translation efficiency \(^{27,28}\).

### 1.3 Decay

All mRNAs are eventually degraded. Most mRNAs are degraded in a deadenylation dependent manner \(^3\). In this process the poly(A)-tail is removed by one of three deadenylase complexes, CCR4-NOT, Pan2/Pan3 or PARN. The deadenylated mRNA is then susceptible to degradation in the 3’-5’ direction by a large protein complex known as the exosome. The exosome is made up of eleven subunits that degrade the mRNA into single RNA nucleotides. Following deadenylation, the mRNA is also decapped and degraded in a 5’-3’ direction. Decapping is carried out by the decapping protein Dcp2 and its accessory protein Dcp1. The decapped mRNA is then a substrate for 5’-3’ degradation by the RNA exonuclease Xrn1.

Messenger RNA degradation, like transcription, is tightly regulated to control gene expression. Degradation of many mRNAs is initiated by \textit{trans}-acting factors that bind the mRNA and recruit the deadenylase and/or decapping factors. MicroRNAs and RNA binding proteins have been shown to promote mRNA degradation in this fashion \(^{29-33}\).

### 1.4 Post-transcriptional Regulation

#### 1.4.1 microRNAs

MicroRNAs are short, 21-23 nt, noncoding RNAs that were discovered in the early 1990s \(^{34}\). They are known to post-transcriptionally regulate mRNAs. MicroRNAs are endogenous, many reside within their own genomic locus while others are within introns of other genes. The human genome encodes \(>2500\) miRNAs \(^{35}\). The primary miRNA transcript, pri-miRNA, contains a stem-loop structure flanked on either side by additional RNA sequence. The stem-loop
structure is recognized by Drosha which cleaves the RNA sequences on either side of the stem-loop producing the precursor miRNA, pre-miRNA. The pre-miRNA structure which contains the mature miRNA and its guide strand is exported from the nucleus by Exportin 5. In the cytoplasm, Dicer cleaves the loop from the pre-miRNA and loads the mature miRNA into one of the Argonaute proteins. In Drosophila, miRNAs are loaded into Ago1. Argonaute interacts with another protein, GW182, to form the RNA induced silencing complex, or the miRNA induced silencing complex (miRISC).

Base-pairing between the miRNA and target sequences within mRNAs recruits the miRISC to the mRNA. miRNAs tend to pair with sites in the 3’UTR of their target mRNAs. The 5’ 6-8 nt of the miRNA form the seed region. The extent of base-pairing within the seed region influences the efficiency of miRNA-mediated repression. For most effective miRNA targets there is base-pairing outside of the seed region as well. Rarely, miRNAs can bind with perfect complementarity to the target mRNA. When this perfect base-pairing occurs, Argonaute endonucleolytically cleaves the mRNA. Many mRNAs are targeted by multiple miRNAs and/or have multiple binding sites for a given miRNA.

MiRISC binding to a target mRNA leads to translational repression and mRNA degradation. Kinetic studies of miRNA-mediated repression in Drosophila, Zebrafish and human cultured cells showed translational repression to precede mRNA degradation. Conversely, it has been recently observed that the two events may be coupled; with mRNA degradation occurring co-translationally.

The exact mechanism of translational repression by the miRISC is still debated. It is generally thought that the miRISC inhibits translation at the initiation step; after cap-binding but before recruitment of the large subunit. Two helicases that have roles in translation have been
implicated in miRNA-mediated repression: eIF4A and DDX6. As a component of the eIF4F complex, eIF4A is involved in recruitment of the PIC to the mRNA and is thought to facilitate scanning of the 5’UTR. Knockdown and inhibition of eIF4A2 in human cells reduced miRNA-mediated repression ⁴⁶. Recently it has been observed that miRISC binding to an mRNA causes dissociation of eIF4A1 and eIF4A2 (or eIF4A in Drosophila) from the mRNA ⁴⁷, ⁴⁸. Following dissociation of eIF4As the rest of the eIF4F complex also dissociates. It has been proposed that the miRISC is inhibiting scanning by targeting eIF4A ⁴⁹. In fact, messages with certain internal ribosome entry sites (IRES) that do not require scanning to initiate translation are refractory to miRNA-mediated repression ⁴⁶, ⁴⁹. Furthermore, messages with unstructured 5’UTRs are also refractory to miRNA-mediated repression ⁴⁶. While all of these data suggest the miRISC causes translational repression by targeting eIF4A and inhibiting scanning, a recent report challenged this paradigm by showing cells lacking eIF4A2 have functional miRNA-mediated repression ⁵⁰. It is possible that in the case of a complete knockout, as opposed to a transient knockdown or inhibition, that the cell has compensated for the activity of eIF4A2 in some way. To further complicate this story, another RNA helicase DDX6 has also been implicated in miRNA-mediated repression.

DDX6 is a DEAD-box helicase that promotes translational repression and decapping of mRNAs ⁵¹. Knockdown of DDX6 in human cells reduced miRNA-mediated repression, but not siRNA-mediated RNA silencing ⁵². DDX6 is known to interact with miRISC through CNOT1, a component of the CCR4-NOT deadenylase complex that interacts with GW182 ⁵³. Disruption of the interaction of DDX6 with CNOT1 reduced miRNA-mediated repression. As DDX6 has roles in both mRNA decay and translational repression it is thought that it may act as an effector for miRNA-mediated translational repression and decay.
MicroRNA targets often have reduced RNA stability as well as reduced translation. The GW182 component of the miRISC acts as scaffold that recruits deadenylase factors to the targeted mRNA \(^{31, 54}\). The CNOT1 subunit of the CCR4-NOT deadenylase complex interacts with two motifs within GW182. These interactions recruit the deadenylase complex to miRNA-targets and facilitate deadenylation. Further interactions with GW182 and the PAN2-PAN3 deadenylase complex, at a different interaction motif, recruits yet another deadenylase complex to the miRNA-targeted mRNA \(^{54}\). Beyond recruitment of deadenylase complexes to the miRISC bound mRNA, an interaction motif within GW182 also binds PABP \(^{55}\). This interaction is thought to facilitate deadenylation in a spatial manner. As well as deadenylation, many miRNA-targets are decapped and subsequently degraded in a 5’-3’ manner. The decapping complex of DCP1 and DCP2 is required for miRNA-mediated mRNA degradation. The miRISC promotes association of the decapping accessory proteins DCP1, HPat1 and me31b (Drosophila homologue of DDX6) \(^{56}\).

While we now know much more regarding the mechanism of miRNA-mediated repression, there remain questions: What is the true mechanism of translational repression, inhibition of scanning by dissociation of eIF4A or recruitment of DDX6, or both? Why are many miRNA-targets well repressed while others are not? Is the mechanism of miRNA-mediated repression the same for all targets?

**1.4.2 RNA binding proteins**

RNA binding proteins (RBPs) serve many roles in RNA biology: splicing, RNA editing, polyadenylation, localization, deadenylation, translation activation and repression, etc. The human genome encodes >1000 RBPs \(^{57-61}\). In HeLa cells alone there are >800 mRNA interacting
RBP s 57. Many of those RBPs have unknown functions. A handful of those RBPs have more defined roles in regulating the stability and translation of mRNAs. One such RBP is Pumilio.

Pumilio is a member of the Puf family of RBPs that are conserved from yeast to humans 62. Pumilio along with an accessory protein Nanos binds primarily to the 3’UTR of mRNAs. Pumilio plays an important role in early development of Drosophila, where it represses the translation of the morphogen Hunchback 63-65. The Pumilio recognition element (PRE), sometimes referred to as a Nanos recognition element, is an unstructured RNA sequence of UGUANAGA. This binding site is highly conserved 66. Much like miRNA-mediated repression, Pumilio binding promotes translation repression and degradation of its targets 27, 32, 64, 66-68. Pumilio recruits the CCR4-NOT deadenylase complex to target mRNAs to promote deadenylation and decay 27, 32. Translational repression by Pumilio is thought to occur by displacement of PABP which disrupts closed-loop complex formation 27. It has also been proposed that Pumilio promotes translation repression by interfering with the cap binding protein eIF4E 68.

The Pumilio accessory protein Nanos is tightly regulated during Drosophila development. Nanos expression is confined to the posterior portion of the embryo, in the pole-bodies that will give rise to the germline. Nanos is regulated by the RBP Smaug. Smaug is conserved from yeast (Vts1) to humans (SAMD4A and SAMD4B) and regulates both translation and RNA-decay 69-71. Smaug binds to Smaug recognition elements (SREs) in target mRNAs. The Nanos 3’UTR has two SREs. SREs are stem-loop structures with a 4-8 nt loop that contains CNGG immediately following the 5’ loop-closing nucleotide 71, 72. Smaug recruits several effector proteins that promote translational repression and decay of its target mRNAs. Smaug is thought to inhibit translation through recruitment of the cap binding protein Cup which competes
with eIF4E to inhibit cap-dependent translation. It has also been proposed that Smaug inhibits translation and promotes RNA-decay through recruitment of Ago1 to the mRNA. Smaug also recruits deadenylase complexes to target mRNAs to promote deadenylation and subsequent decay. The actions of Smaug are antagonized by the RNA-binding protein Oskar. Oskar binding to the Nanos 3’UTR causes dissociation of Smaug and stabilizes the Nanos mRNA. The complex messenger ribonuclear protein (mRNP) that forms on the Nanos mRNA through the binding of multiple RBPs and accessory proteins to the Nanos 3’UTR leads to tight regulation of its translation and stability.

While the RBPs Pumilio and Smaug have well defined binding sites, another class of RBPs have a less discrete binding site, AU-rich elements (ARE). AREs are categorized into three classes based on the sequence content of the AU-rich element. For instance, class I AREs contain several copies of an AUUUA motif within a U-rich region. There are 20 RBPs known to interact with AREs. These proteins are involved in splicing, translation and decay. A well-studied family of ARE-binding proteins (ARE-BP) is the homologues of Drosophila embryonic lethal abnormal vision (elav). There are four elav homologues in mammals. The most common and ubiquitously expressed mammalian homologue of elav is HuR (ELAVL1). HuR has been shown to stabilize bound mRNAs and also has a role in splicing. Interestingly, HuR has been shown to promote and antagonize the action of the miRISC. Other ARE-BP have been shown to destabilize mRNA targets (TTP) or to have dual roles in stabilizing and destabilizing mRNA targets (Auf1).

While we have an idea of the mechanism of action of some RBPs, like Smaug and Pumilio, for many others the function and the mechanism are unknown. For the remaining RBPs,
what are their functions and mechanisms? How do RBPs interact functionally with one another and with other mRNA trans-factors such as the miRISC?

1.5 Conclusions

Control of gene expression is essential to life. While much control is exerted at the level of transcription, many genes are regulated post-transcriptionally. Messenger RNAs can be regulated in many ways: alternative splicing, localization, translation and stability. Trans-acting factors such as RBPs and non-coding RNAs like miRNAs are important post-transcriptional regulators. With over 2500 miRNAs and >1000 RBPs encoded in the human genome there is likely to be considerable overlap in the targets of each trans-acting factor. Furthermore, the immense variability in mRNA structure, sequence elements and codon usage provide further complexity to post-transcriptional regulation.

Understanding how various trans-acting factors functionally, and potentially physically, interact with each other is essential to our understanding of post-transcriptional regulation. Furthermore, understanding how mRNA features such as 5’UTR structure, uORFs, codon optimality, etc. affect the regulatory capacity of trans-acting factors is also of importance.

Here I report on our efforts to elucidate mechanisms of post-transcriptional regulation. In Chapter 1, I describe a novel method for RNA immunoprecipitation (RIP). This method, Urb-RIP uses the RNA recognition motif of the “resurrected” snRNA-binding protein Urb to enrich transcripts containing a stem-loop tag. This method can be used to identify proteins and RNAs that interact with an RNA of interest. In Chapter 2, I describe the results of our systematic analysis of modulation of miRNA-mediated repression by mRNA characteristics. In this study we used a reporter system to assay the effects of various mRNA elements and characteristics on miRNA-mediated repression. We observed modulation of miRNA-mediated repression by
5’UTR structure, uORFs, codon optimality, 3’UTR sequence and translation efficiency. These results provide insight into the wide variability of miRNA-mediated repression observed in the literature. In Chapter 4, I describe a novel approach to studying the mechanism and interactions of miRNAs and RBPs. This method, post-transcriptional regulatory element sequencing (PTRE-seq) employs a massively parallel reporter system to study post-transcriptional regulation by miRNAs and RBPs. By studying the effects of RBPs and miRNAs, individually or in combination, on RNA stability and translation we were able to identify interactions between these *trans*-factors and differences in their mechanisms of action. Finally, in Chapter 5, I describe preliminary data and future directions for the study of post-transcriptional regulation by miRNAs and RBPs.
1.6 References


Chapter 2: Urb-RIP – an adaptable and efficient approach for immunoprecipitation of RNAs and associated RNAs/proteins

Preface

The following work was completed by myself and Sergej Djuranovic. S.D. conceived and supervised the project. K.A.C. performed all experiments, data analysis and wrote the manuscript with contributions from S.D.

This chapter is published in its entirety [Cottrell KA, Djuranovic S. 2016. Urb-RIP - An Adaptable and Efficient Approach for Immunoprecipitation of RNAs and Associated RNAs/Proteins. PLoS One 11: e0167877.] and is available at https://doi.org/10.1371/journal.pone.0167877. This article is distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

We thank the Kathleen Hall (Washington University Medical School) for a plasmid containing Urb-RRM1 and for advice. We also thank Denise Rogers and Yansel Nunez for assistance in cloning. This work was supported by an NIH training grant T32 GM007067 to KAC and NIH Grant RO1 GM112824 to SD.
2.1 Abstract

Post-transcriptional regulation of gene expression is an important process that is mediated by interactions between mRNAs and RNA binding proteins (RBP), non-coding RNAs (ncRNA) or ribonucleoproteins (RNP). Key to the study of post-transcriptional regulation of mRNAs and the function of ncRNAs such as long non-coding RNAs (IncRNAs) is an understanding of what factors are interacting with these transcripts. While several techniques exist for the enrichment of a transcript whether it is an mRNA or an ncRNA, many of these techniques are cumbersome or limited in their application. Here we present a novel method for the immunoprecipitation of mRNAs and ncRNAs, Urb–RNA immunoprecipitation (Urb-RIP). This method employs the RRM1 domain of the “resurrected” snRNA-binding protein Urb to enrich messages containing a stem-loop tag. Unlike techniques which employ the MS2 protein, which require large repeats of the MS2 binding element, Urb-RIP requires only one stem-loop. This method routinely provides over ~100-fold enrichment of tagged messages. Using this technique we have shown enrichment of tagged mRNAs and IncRNAs as well as miRNAs and RNA-binding proteins bound to those messages. We have confirmed, using Urb-RIP, interaction between RNA PolIII transcribed IncRNA BC200 and polyA binding protein.

2.2 Introduction

Regulation of gene expression at the post-transcriptional level is a complex process involving many trans factors such as RNA-binding proteins (RBPs), non-coding RNA (ncRNA), and ribonucleoproteins (RNPs) [1, 2]. In order to fully understand this process for any given RNA it is essential that we know what factors are bound to the transcript. This knowledge will prove useful in designing therapies that target trans factors or the RNA itself. While there exist many molecular techniques for purification of RNAs of interest [3-18] and in silico tools for
identification of RNA binding protein (RBP) or ncRNA binding sites on your RNA of interest [19-24], many of these tools have limitations in their applicability, efficiency or false positive rate. Current techniques for RNA purification fall into one of three classes: RBP-mediated [3-10], aptamer and oligonucleotide-mediated [11-18], or direct purification of biotinylated RNA [25, 26]. While each of these techniques has been used successfully, they often require unique experimental designs that make them potentially less adaptable and more time consuming.

Pulldown of RNA of interest using aptamers or oligonucleotides relies on base pairing of a biotinylated-oligonucleotide to an RNA of interest or binding of a structure inserted into an RNA of interest to compound, usually a metabolite. While both techniques have been used successfully they both suffer from the same difficulty, RNA structure. Folding of the RNA can disrupt the formation of the aptamer or occlude the binding site of an oligonucleotide. For pulldown with oligonucleotides this can be abrogated by tiling across the entire RNA with multiple oligonucleotides; however this increases the chance of pairing with other RNAs beside the RNA of interest.

Direct purification of a biotinylated RNA involves in vitro synthesis of an RNA of interest and tagging with biotin, this is usually done with a biotinylated cap or 5’ nucleotide. The biotinylated RNA is then incubated with a cell lysate and subsequently precipitated with a streptavidin matrix. The biggest drawback of this technique is that the RNA is introduced to a lysate as opposed to being transcribed within the cell as normal. It is well appreciated that numerous proteins bind to RNAs concurrent with transcription or splicing. These interactions may not occur when an in vitro transcribed RNA is incubated with a cell lysate, potentially leading to false negative results.
Likely the most common technique for affinity based RNA-purification is RNA-immunoprecipitation (RIP) using the bacteriophage MS2-coat protein. This approach uses an epitope tagged MS2-protein to enrich an RNA of interest containing the MS2 hairpin. While this technique has been widely and successfully used [3-8] it is not without its pitfalls. The main pitfall being a lack of efficiency; RNA of interest are routinely tagged with multiple MS2-hairpins often up to two dozen [3-8]. The addition of a large number of MS2-hairpins adds a significant amount of mass to the RNA of interest and can result in relatively poor enrichment, less than one order of magnitude [3].

Here we report a new method for targeted RNA pull-down that is both efficient and highly adaptable. Our approach, which we have named Urb – RNA immunoprecipitation (Urb-RIP), utilizes the RNA recognition motif 1 (RRM1) domain of the “resurrected” snRNA-binding protein Urb to enrich transcripts containing a stem-loop tag. The RRM1 domain of Urb binds stem-loop II (SLII) of the U1-snRNA and SLIV of the U2-snRNA with high affinity [27]. Urb-RIP uses a single SLII-tag to allow binding of Urb-RRM1 to an RNA of interest. Prior to cell lysis we employ crosslinking by UV irradiation to produce RNA-protein crosslinks between Urb and the tagged RNA or other proteins bound to the RNA much like CLIP techniques [28, 29]. Following immunoprecipitation it is possible to specifically elute RNA or protein bound to the RNA of interest. We have validated Urb-RIP using transcripts generated by RNA polymerase II and III. Pull-down of mRNAs was highly efficient and provided enrichment of RNA binding proteins bound to the message. Using Urb-RIP pull-down of a miRNA-targeted reporter we have enriched for the miRNA which targets the tagged mRNA as well as Argonaute protein, part of the miRNA-induced silencing complex (miRISC). Finally, we confirmed the binding of polyA-binding protein (PABP) to the RNA PolIII transcribed IncRNA BC200 using Urb-RIP.
2.3 Materials and Methods

2.3.1 Construction of 2HA-Urb and SLII-tagged RNA constructs

All primers used for cloning can be found in Supplemental Table 1. The RRM1 domain of Urb was a kind gift from the laboratory of Kathleen Hall. Two rounds of PCR were performed with first 2HA-RRM1-URB forward-1 and then 2HA-RRM1-URB forward-2 each with 2HA-RRM1-URB reverse to add 2x HA tags followed by a TEV protease site to URB-RRM1. 2HA-TEV-URB-RRM1 was then cloned into pENTR-D-TOPO (Invitrogen). Site-directed mutagenesis was performed using the 2HA-URB NarI mutagenesis primers to introduce a NarI restriction site in between the coding region for the TEV protease and N-terminus of URB-RRM1. The pENTR-2HA-TEV-URB-RRM1 plasmid containing the inserted NarI site was then digested with NarI and ligation was performed to insert a single FLAG tag using the FLAG oligonucleotides 1 and 2. LR-Clonase II (Invitrogen) was used to transfer the pENTR-2HA-TEV-FLAG-URB-RRM1 insert to the destination plasmid pT-RexDEST31 (Invitrogen), making pT-Rex-2HA-TEV-FLAG-URB-RRM1 (referred to above in the text as pT-Rex-2HA-URB). The pENTR-2HA-TEV-FLAG-URB-RRM1 plasmid was also recombined with pCDNA5-Frt-TO (Invitrogen) to make a pCDNA5-2HA-URB.

To facilitate SLII-tagging of mRNAs a destination vector was constructed that would place a SLII-tag in the 3’UTR. The SLII-tag was inserted into the 3’UTR of pcDNA-DEST40 to make pcDNA-DEST40-SLII. pcDNA-DEST40 was digested with SacII and ligation was performed to insert the SLII-tag using the SLII Tag SacII oligonucleotides 1 and 2.

LR-Clonase II (Invitrogen) was used to transfer the pENTR-mCherry insert to the destination plasmid pcDNA-DEST40 or pcDNA-DEST40-SLII, making pcDNA-mCh and pcDNA-mCh-SLII.
The transcript sequence for the lncRNA-BC200 was amplified from HEK 293 genomic DNA purified using DNeasy Blood & Tissue Kit (Qiagen) by the BC200 forward and BC200 reverse (with PolIII terminator) primers. The PCR product was digested with SpeI and XbaI and ligated into pSM2 vector (Addgene) digested with the same restriction enzymes. The SLII-tag with XbaI/SpeI overhangs, SLII Tag SpeI oligonucleotides 1 and 2, was ligated into the pSM2-BC200 plasmid digested with SpeI. The inserts from pSM2-BC200 and pSM2-SLII-BC200 were recombined into pcDNA-DEST40ΔCMV using LR-Clonase II (Invitrogen) to make pcDNAΔ-BC200 and pcDNAΔ-SLII-BC200. pcDNA-DEST40ΔCMV was constructed by digest to remove the CMV promoter with SpeI (NEB) and SacI (NEB) followed by ligation with the CMV deletion oligonucleotides 1 and 2.

A region of pAWH-Rluc-let-7-A114-N40-HhR [30] containing eight let-7 binding sites was amplified using the pAWH let-7 sites forward and reverse primers. This PCR product was gel purified and phosphorylated using T4-PNK (NEB). This product was then ligated into the pcDNA-mCh and pcDNA-mCh-SLII plasmids that had been digested with Pmel and dephosphorylated with Antarctic phosphatase (NEB).

To make EGFP-FLAG-AGO2, EGFP was PCR amplified using the EGFP forward and reverse-overlap primers and Ago2 was amplified using the Ago2 forward-overlap and reverse primers. The PCR products were stitched together using overlap PCR and cloned into pENTR-D-TOPO. LR-Clonase II (Invitrogen) was then used to transfer the EGFP-FLAG-Ago2 cassette into pcDNA-DEST40 making pcDNA-EGFP-FLAG-Ago2.

2.3.2 Cell culture and transfection

T-REx™-293 cells and Flp-In T-REx™-293 (Invitrogen) were grown in DMEM (Gibco) supplemented with 10% heat-inactivated FBS (Gibco), 1x Penicilin streptomycin and glutamine
(Gibco) and 1x MEM Non-Essential Amino Acids (Gibco). T-REx™-293 were kept under
selection with 5 µg/mL blasticidin. Transfection was performed using Xtreme Gene 9 (Roche)
per manufacturer’s recommendations. The plasmid was transfected at a ratio of 1 µg per 2 µL of
transfection reagent. For a 10 cm dish 8 µg of plasmid was used. Two stable cell lines were made
for expression of 2HA-Urb. The pT-REx™-2HA-Urb plasmid was transfected into T-REx™-
293 cells and the cells were selected with 0.5 mg/mL Geneticin (Invitrogen) to produce a stable
cell line. This cell line (T-REx™-293-2HA-Urb) was maintained in 5 µg/mL blasticidin and 0.5
mg/mL geneticin. The pCDNA5-2HA-Urb plasmid was co-transfected with the Flippase
expressing plasmid pOG-44 (Invitrogen) into Flp-In T-REx™-293 and the cells were selected
with 0.1 mg/mL hygromycin to produce a stable cell line. This cell line (Flp-In T-REx™-293-
2HA-Urb) was maintained in 5 µg/mL blasticidin and 0.1 mg/mL hygromycin. These cell lines
were used interchangeably with minimal differences in Urb-RIP efficiency (data not shown).

2.3.4 Crosslinking and immunoprecipitation of SLII-tagged RNA (Urb-RIP)

A detailed protocol for the Urb-RIP method and the recipes for all buffers listed below
can be found in the Supplemental Methods (S1_text), this protocol has been adapted from
previous CLIP protocols [28, 29]. Prior to performing Urb-RIP one 10 cm plate of T-REx-293 -
2HA-Urb or Flp-In T-REx-293-2HA-Urb cells was transfected with a tagged RNA of interest or
an untagged control as described above. Four hours later the media was removed and 2HA-Urb
expression was induced by addition of fresh media containing 2 µg/mL doxycycline. The next
day the cells were transferred into a single 15 cm dish in media containing doxycycline as
before. The following day the cells were washed briefly with cold PBS prior to UV-irradiation at
400 mJ/cm² using a Stratalinker 1800. The cells were suspended in cold PBS by pipetting and
transferred to a conical tube. The cells were pelleted, resuspended in PBS and transferred to a 1.7
mL microfuge tube before pelleting a second time. The pelleted cells were lysed in 1% NP-40 Lysis Buffer containing 1x Complete Protease Inhibitor (Roche) and 0.5 units/µL RNase inhibitor (RNasin, Promega or RnaseOUT, Invitrogen). Lysis occurred over 20 minutes on ice and was followed by centrifugation at 15,000 g for 20 minutes at 4 °C to clear the insoluble fraction. The protein concentration of the lysate was quantified by DC Protein Assay (Bio-Rad). At least 1 mg of total protein was loaded onto anti-HA magnetic beads (Pierce) that had been previously blocked for 1 hr in 4% BSA with 0.5 µg/µL yeast tRNA. An aliquot of the lysate (5% of the amount used for IP) was kept for western analysis and RNA isolation. The lysate was incubated on the anti-HA beads for 1 hr with rotation at 4 °C. Following binding the beads were washed twice with Low Salt Wash Buffer and twice more with High Salt Wash Buffer. The beads were resuspended in water and transferred to two fresh microfuge tubes for elution. Elution of protein was carried out by suspending the beads in reducing sample buffer (XT-Sample Buffer, Bio-Rad, with XT-sample reducing agent, Bio-Rad) and heating at 95 °C for 7 minutes. Elution of RNA was performed by resuspending the beads in 200 µL of Proteinase K Buffer containing 32 units of proteinase K (NEB) and incubation for 20 minutes at 37 °C. After 20 minutes an equal volume of Proteinase K Urea Buffer was added and the samples were incubated another 20 minutes at 37 °C. Following incubation the RNA was extracted using low pH phenol:chloroform and precipitated by ethanol precipitation with glycogen added as a carrier. The precipitated RNA was washed with 70% ethanol, dried and resuspended in water. The RNA was DNase treated (Turbo DNase, Ambion) Isolation of RNA from the input sample was performed in the same manner as elution of the beads.
2.3.5 RNA analysis by RT-qPCR

Total cDNA synthesis was performed using iScript Supermix per manufacturer protocol (Bio-Rad). For miRNA reverse transcription 6 pmol of the let-7 reverse transcription primer [31] was added to the reaction containing 1X iScript Supermix and cDNA was synthesized per manufacturer protocol. Quantitative PCR was performed using iQ SYBR Green Supermix (Bio-Rad) on the CFX96 Real-Time system with Bio-Rad CFX Manager 3.0 software, with a standard 3 step PCR cycle with initial denaturation at 95 °C for 3 min denaturation at 95 °C 10 s, annealing at 55 °C for 10 s, extension at 72 °C for 30 s. Cycle threshold (Ct) values were normalized to GAPDH, except where indicated. All qPCR and reverse transcription primers can be found in Supplemental Table 2.

2.3.6 Protein analysis by western blotting

Protein samples were resolved on a 4-12% Bis-Tris precast gel (Bio-Rad). The resolved proteins were transferred using Trans-Blot SD (Bio-Rad) onto Immuno-Blot PVDF (Bio-Rad). The membrane was blocked in 5% milk in 1x PBS with 1% Tween (PBST) for a minimum of 1 hour. The following primary antibodies were used in western analysis at the given dilution: HA-HRP, 1:2000 (Santa Cruz, sc-7392); PABP, 1:1000 (Abcam, ab21060); FLAG-HRP, 1:5000 (Sigma, F1804); Beta-Actin-HRP, 1:2000 (BioLegend, 643807); GAPDH-HRP, 1:2000 (BioLegend, 649203); Nop56, 1:2000, (Bethyl, A302-721A-T); Anti-mouse IgG HRP, 1:10,000 (Cell Signaling, 7076S); Anti-Rabbit IgG HRP (Cell Signaling, 7074S). All antibodies were diluted in 5% milk in PBST and incubated with the membrane for 2 hours at room temperature or overnight at 4 °C. The membrane washed with PBST prior to incubation with secondary antibody against mouse or rabbit coupled to horse-radish peroxidase (HRP) (Cell Signaling). The secondary was diluted 1:10,000 in 5% milk in PBST and allowed to incubate for 1 hour at room
temperature. The membrane was washed with PBST and HRP activity was detected using SuperSignal West Pico or Dura (Thermo Scientific). The membrane was imaged by Bio-Rad Molecular Imager ChemiDoc XRS System with Image Lab software (Bio-Rad).

2.4 Results

2.4.1 General description of Urb-RIP

We turned to the recently “resurrected” urbilatariian homologue of the SNF/U1A/U2B family of proteins, Urb, to create Urb-RIP as a tool for pull-down of RNA of interest, Figure 1.

![Urb-RIP Protocol](image)

**Figure 2.1: Schematic illustration of the Urb-RIP protocol and potential applications.**

The first step of the Urb-RIP protocol is to tag an RNA of interest with the SLII-tag, illustrated in the figure. The tagged RNA and in parallel an untagged control are coexpressed with 2HA-Urb in a cell line of interest. After a period of time the cells are UV irradiated to produce RNA-protein crosslinks. The cells are then lysed and immunoprecipitation is performed using blocked anti-HA magnetic beads. The RNA or protein is then eluted and analyzed by an appropriate method. This approach should be applicable to mRNAs and lncRNAs and amendable to a wide range of methods for eluate analysis.

We have tagged the RRM1 domain of Urb with two N-terminal hemmaglutanin A (HA) tags followed by a TEV protease cleavage site and a single FLAG tag, Figure 2a, 2HA-Urb. The
RRM1 domain of Urb binds the U1-snRNA SLII and U2-snRNA SLIV with high affinity, $1.2 \times 10^{-9}$ M and $1.5 \times 10^{-8}$ M respectively [27]. This high affinity binding allows Urb-RIP to be performed with an RNA of interest bearing only a single SLII-tag. Urb is suspected to have a structure very similar to the SNF and U1A/U2B proteins for which it is a hypothetical ancestor [27]. It is expected that Urb needs a loop structure for binding, much like U1A [32]. In fact, a nearly identical variant of the Urb RRM1 domain that we are using here, Urb-V, was found to bind SLII of the U1-snRNA greater than 300 fold more efficiently than a linear RNA containing the loop sequence from SLII [33]. Hence, we were careful to maintain the SLII structure in our construct by adding unstructured ‘CAA’ repeats on either side. On the 5’ side of the stem-loop there is seven ‘CAA’ repeats while there are three on the 3’ side, Figure 1. We have also incorporated a restriction enzyme site on one side of the SLII-tag to aid in subsequent cloning. As such the total length of the engineered pull-down sequence is 60 nucleotides. The Urb-RIP procedure consists of five main steps, Figure 1. First, a SLII-tagged-RNA of interest is co-expressed with 2HA-Urb. After a period of time the cells are UV irradiated to induce RNA-protein crosslinks. Following UV irradiation cell lysis and immunoprecipitation with magnetic anti-HA matrix is performed followed by thorough washing. Prior to immunoprecipitation the anti-HA matrix is blocked with bovine serum albumin (BSA) and yeast tRNA to reduce non-specific binding. Proteinase K is used to elute the tagged RNA by degrading URB, as well as any other proteins in the IP product. The proteinase K is removed by phenol:chloroform extraction of the eluate and the RNA is precipitated using standard ethanol precipitation. The proteinase K treatment is an important step in the procedure much like in HITS-CLIP as it degrades proteins covalently bound to the RNA that could interfere with reverse transcriptase during cDNA synthesis [28]. Protein is eluted from the anti-HA matrix by adding reducing sample buffer and
boiling. Finally, the eluted RNA or protein can be analyzed by a method of choice, qRT-PCR, western blot, northern blot, RNA-seq, proteomics, etc.

2.4.2 Urb-RIP allows for enrichment of mRNAs

We first sought to validate Urb-RIP using a mCherry reporter containing a single SLII-tag, Figure 2a. This reporter was transfected into a stable T-REx™-293 cell line for the inducible expression of 2HA-Urb, we will refer to this cell line as 293-2HA-Urb. As a control, a parallel transfection was performed with a mCherry reporter lacking the SLII-tag. These reporters were used to optimize the pull-down conditions, as well as the amount of UV-irradiation and the procedure for blocking anti-HA-beads, Figures S1, S3 and S4. The pull-down efficiency for a mCherry mRNA construct with a single SLII-tag was high for non-UV irradiated conditions (approx. 100 fold over the untagged control mRNA) and could further be improved by UV-induced RNA-protein crosslinking. We found that cross-linking with 400 mJ/cm² of UV prior to lysis provided enrichment over non-irradiated samples (approximately 2 fold) for the SLII-tagged mCherry RNA, Figure S1. Importantly we did not observe overt cleavage of RNA following UV irradiation as determined by standard analysis of rRNA integrity by denaturing agarose gel, Figure S2. Blocking of the beads with yeast tRNA and BSA increased enrichment while pre-clearing of the lysate with protein A/G matrix improved specificity of the tagged mRNA pull-down but did not improve overall enrichment, Figure S3. Following optimization of Urb-RIP we were able to readily obtain enrichment of our SLII-tagged mCherry reporter of ~350 fold, Figure 2b. Importantly the addition of the SLII-tag to the 3’UTR of mCherry only modestly reduced expression, Figure 2C. We found that one of the pitfalls of the Urb-RIP approach is the ability of the 2HA-Urb protein to interact with SLII of the endogenous U1-snRNA. However, we
observed that blocking of matrix prior to immunoprecipitation reduced enrichment of U1-snRNA, Figure S4.

Figure 2.2: Urb-RIP enriches for mCherry-mRNA and bound PABP

a Schematic describing 2HA-Urb construct used for Urb-RIP and reporter, mCherry (mCh)-mRNA tagged with SLII, used to validate and optimize Urb-RIP. b Enrichment of mCh-mRNA by Urb-RIP as determined by qPCR. The cell line 293-2HA-Urb was transfected with a plasmid expressing mCherry-mRNA untagged or tagged with SLII. Two days after transfection the cells were UV-irradiated at 400 mJ/cm² and subsequently lysed. Immunoprecipitation was performed using the Urb-RIP protocol and RNA was eluted with proteinase K treatment. qRT-PCR was performed using mCh and GAPDH primers. c Comparison of relative expression amounts of mCherry with (+SLII) and without (-SLII) tagging. qPCR results show a modest reduction in mCherry expression upon insertion of SLII. Relative levels of mCherry are normalized to GAPDH. d western blot shows enrichment of PABP following Urb-RIP of mCh. Half of the immunoprecipitate from above was eluted with sample buffer and analyzed by western blot with antibody against proteins listed. Samples labeled input represent 5% of the total sample used for Urb-RIP. e quantification of western blot in d. For panels b and c mean ± SD of three independent experiments are shown. For panel e mean ±SD of three independent experiments are shown. * p<0.05, ** p<0.01
2.4.3 Urb-RIP enriches for tagged mRNA and a bound RNA binding protein

To validate Urb-RIPs ability to identify RBP bound to RNA of interest we assessed enrichment of PABP bound to a tagged mRNA. We used the mCherry reporters described above, Figure 2. These reporters were transfected into 293-2HA-Urb. Analysis of Urb-RIP using these constructs showed enrichment of mCherry-mRNA and PABP, Figure 2B, D and E. The presence of background PABP binding was not surprising as the Urb-RIP product often contains traces of non-tagged mRNAs, Supplemental Table 3. However, the tagged-RNA is efficiently immunoprecipitated and much more abundant in the Urb-RIP pull-downs than the untagged; for example the tagged RNA can be detected by qPCR approximately 8 cycles before the untagged RNA, Supplemental Table 3. While in the Urb-RIP input the untagged and tagged RNA are detected with less than a cycle difference. GAPDH can routinely be detected with a threshold cycle in the mid-thirties but is sometimes undetectable, Supplemental Table 3 and Supplemental Table 4. The same is true for Actin mRNA. Even highly abundant RNAs such as 7SK are much less abundant in the IP eluate than our tagged mRNA, approximately 2% of the abundance of the tagged mRNA, Supplemental Table S3. To control for potential binding of 2HA-Urb to RNAs containing a sequence similar to SLII, i.e. the loop from SLII, we used qPCR to detect binding to the TIMM50 mRNA. While TIMM50 contains a sequence identical to the loop of our SLII-tag we did not observe any enrichment upon IP, Supplemental Table 3. Importantly we did not observe binding of abundant proteins such as beta-actin or GAPDH. Also, we did not observe binding of the RBP Nop56, which binds the box C/D snoRNAs and is involved in ribosome biogenesis [34, 35].
2.4.4 Urb-RIP enriches for miRNAs and Ago2

A common desire in the field of post-transcriptional regulation of gene expression is the ability to identify miRNAs and miRNA-induced silencing complexes (miRISC) that target an RNA of interest. While there exists several in silico tools to identify potential binding sites for a miRNA, it is often the case that their predictions produce false positives [36, 37]. An alternative approach has been to utilize HITS-CLIP or PAR-CLIP which can identify targets of a given miRNA [28, 38]. In order to identify which miRNAs are capable of binding an RNA of interest the most direct approach would be to enrich for the RNA through a pull-down. With this in mind we sought to test Urb-RIPs ability to identify a miRNA bound to an RNA of interest. We used a let-7 reporter construct which contains multiple binding sites for let-7 in the 3’UTR of the reporter gene [30]. The insertion of eight let-7 sites has been shown previously to reduce expression of such reporters [30] and we have observed repression of our mCherry-let-7 reporter by reduction in mCherry fluorescence (data not shown). We inserted the SLII-tag between the let-7 sites and the polyadenylation signal, Figure 3A. This construct along with a control lacking the SLII-tag and the mCherry constructs described in Figure 2 were transfected into the 293-2HA-Urb cell line used above and Urb-RIP was performed two days later. Analysis of the immunoprecipitated RNAs revealed enrichment of let-7 bound to the mCherry-let-7 reporter, Figure 3B and C. We next transfected the mCherry-let-7 reporter along with GFP-FLAG-Ago2 and performed Urb-RIP as before. In parallel we assayed the GFP-FLAG-Ago2 binding to the control reporter lacking the SLII-tag. Analysis of the Urb-RIP product by western blot revealed modest enrichment of GFP-FLAG-Ago2 in the sample containing the SLII-tagged mCherry-let-7 reporter, Figure 3D and E. As seen with PABP in Figure 2 there was GFP-FLAG-Ago2 in the
control sample. This can be attributed to overexpression of tagged-Ago2 construct and the presence of trace amounts of various RNA species in the control pull-down as described above.

Figure 2.3: Urb-RIP shows Argonaute and miRNA binding to miRNA-targeted messages in human cells

- **a** Schematic describing the *let*-7 reporters used to validate Urb-RIPs ability to identify interacting miRNA.
- **b** Enrichment of *let*-7 and mCh-mRNA by Urb-RIP as determined by qPCR. The cell line 293-2HA-Urb was transfected with plasmids expressing the constructs described in **a** as well as a plasmid for expression of GFP-FLAG-Ago2. Two days after transfection the cells were UV-irradiated at 400 mJ/cm² and subsequently lysed. Immunoprecipitation was performed using the Urb-RIP protocol and RNA was eluted with proteinase K treatment. qRT-PCR was performed using mCh, *let*-7 and GAPDH primers.
- **c** enrichment of *let*-7 normalized to mCh abundance in the immunoprecipitate.
- **d** western blot shows enrichment of GFP-FLAG-Ago2 following Urb-RIP. The mCh-*let*-7-SLII reporters from above were co-transfected with GFP-FLAG-Ago2. Two days after transfection Urb-RIP was performed. The eluted protein as well as input was analyzed by western blot with antibody against FLAG (GFP-FLAG-Ago2 and 2HA-FLAG-Urb). Samples labeled input represent 5% of the total sample used for Urb-RIP.
- **e** quantification of western blot in **d**, normalized to 2HA-Urb, relative to mCh-*let*-7-SLII.

2.4.5 Urb-RIP confirms binding of PABP to the RNAPIII IncRNA BC200

In order to show the adaptability of Urb-RIP for other RNA transcripts we used our method to identify factors bound to the IncRNA BC200. BC200 is a well described IncRNA
transcribed by RNA PolIII [39-43]. The BC200 transcript contains a large A-rich element which was shown in vitro to interact with PABP [40-43]. We tagged the 5’ end of BC200 with SLII and expressed it using the U6 promoter in a 2HA-Urb stable cell line as described in Figure 4A. The use of a single SLII-tag allows us to add a relatively short sequence to the natural BC200 transcript, Figure 4A, in comparison with tagging BC200 with MS2 hairpins which would approximately double or triple the length of the BC200 transcript if tagged with 12 or 24 MS2 hairpins as is common.

Figure 2.4: Urb-RIP confirms binding of PABP to BC200
a Schematic describing the BC200 construct used for pull-down by Urb-RIP. b Enrichment of BC200+SLII by Urb-RIP as determined by qPCR. A stable HEK-293 cell line for the inducible expression of 2HA-Urb was transfected with plasmids expressing the constructs described in a. Two days after transfection the cells were UV-irradiated at 400 mJ/cm² and subsequently lysed. Immunoprecipitation was performed using the Urb-RIP protocol and RNA was eluted with proteinase K treatment. qRT-PCR was performed using BC200 and GAPDH primers. c Comparison of relative expression amounts of BC200 with (+SLII) and without (-SLII) tagging. qPCR results show no change in BC200 expression upon insertion of SLII. Relative levels of BC200 are normalized to GAPDH. d Western blot shows enrichment of PABP following Urb-RIP of BC200+SLII. e Half of the immunoprecipitate from above was eluted with sample buffer and analyzed by western blot with antibody against proteins listed. Samples labeled input represent 5% of the total sample used for Urb-RIP.
Analysis of western blot in d, normalized to 2HA-Urb, relative to mCh-SLII. For panels b and c mean ± SD of two independent experiments are shown. For panel e mean ±SD of three independent experiments are shown. * p<0.05.

We transfected our SLII-tagged BC200 construct as well as untagged control in parallel into the 293-2HA-Urb cell-line used previously. Urb-RIP with SLII-tagged and control BC200 expressing cells was performed two days after transfection. Analysis of the pull-down efficiency showed substantial enrichment of BC200 IncRNA, which was readily more than 2000 fold, Figure 4B and Figure S5. The addition of the SLII-tag had no effect on BC200 expression, Figure 4C. Analysis of the immunoprecipitated RNA by bioanalyzer showed a prominent peak for BC200 in the BC200+SLII pulldown, this peak was absent from the control pulldown, Figure S6. Furthermore, analysis of non-target RNAs by qPCR showed little binding during the pull-down of BC200, Supplemental Table 5, consistent with the results of mCherry pull-down, Supplemental Table 3. We could also confirm binding of PABP to BC200 by western blot analysis of immunoprecipitated tagged-BC200, Figure 4D and E. As such we could show that Urb-RIP method can be used equally well for untranslated IncRNA that may act in post-transcriptional control of gene expression.

2.5 Discussion

We have presented Urb-RIP, an adaptable and efficient approach to affinity purify specific RNAs and to identify interacting RNAs, RBPs or RNPs. Our method uses a novel affinity tag for RNA affinity purification. We utilize the RRM domain of a recently “resurrected” snRNA-binding protein, Urb [27]. Urb-RIP takes advantage of the high affinity binding of the RRM1 domain of Urb to SLII of the U1-snRNA to affinity purify an RNA of interest using a single stem-loop tag. By epitope tagging the RRM1 domain of Urb making 2HA-Urb we can affinity purify any RNA of interest containing the SLII-tag with anti-HA matrix in a single
purification step with higher efficiency than most of the current methods. In order to improve the effectiveness of this technique we have incorporated UV induced crosslinking. Much like CLIP or HITS-CLIP, the UV crosslinking used in Urb-RIP helps to stabilize RNA-protein complexes [29, 44]. While UV crosslinking is notoriously inefficient, between 1-5% [44], we have opted to include it in this protocol in order to help maintain interactions with more transient or weakly interacting RBPs. It is possible to perform Urb-RIP without UV crosslinking however we have found that it is slightly less efficient than with crosslinking, Figure S1. We have shown that Urb-RIP can provide enrichment of RBPs and miRNAs bound to immunoprecipitated RNAs.

This method provides many advantages over the commonly used MS2 system for RNA purification. Urb-RIP requires only one SLII-tag there by limiting the mass added to the RNA of interest. The single tag also makes cloning much easier as the tag can be synthesized using a single DNA oligonucleotide and its complement and simply ligated into a plasmid of interest or added to the template sequence of an RNA of interest through PCR. We have not observed any aggregation of yellow fluorescent protein tagged 2HA-Urb or negative effects on cellular homeostasis upon continuous expression of 2HA-Urb in stable cell lines (data not shown). This gives Urb-RIP an advantage over other RNA pull down methods. Aggregation of the MS2 protein is a common problem and requires tight control of expression in order to be mitigated [4, 45]. While mutations in MS2 coat proteins may reduce the oligomerization pattern of the protein [8, 46], requirement of the multiple binding loops still increases possibility for aggregation and reduction in the immunoprecipitation of active RNP complexes on targeted RNA transcripts.

Urb-RIP proved capable of enrichment of tagged mRNAs and lncRNAs from cell lysates as well as for their trans regulators: ncRNAs, RBPs and RNPs, Figures 2, 3 and 4. Urb-RIP method proved capable of enriching for a miRNA and Argonaute, miRISC component, bound to
an RNA of interest (Figure 3A and B). We used a reporter for the miRNA let-7 to confirm the ability of Urb-RIP to identify interacting miRNAs by qPCR. As such the ability of Urb-RIP to identify miRNAs bound to RNA of interest could be a very valuable tool to many researchers.

Using our method we were able to confirm binding of PABP to the PolIII in vivo transcribed human IncRNA BC200. We showed that human BC200 IncRNA can be efficiently immunoprecipititated using our Urb-RIP method, Figure 4. Further western blot analysis of immunoprecipitated material bound to tagged-BC200 showed subtle and reproducible enrichment of PABP. The interaction of PABP to an internal tract of adenosines in human BC200 and mouse BC1 IncRNAs has been shown previously in vitro by either electrophoretic mobility shift assay (EMSA) or by immunoprecipitation of PABP bound RNAs from cells transfected with in vitro transcribed BC200 [40-43]. By using Urb-RIP method we were able to show, for the first time, PABP and BC200 interaction by pulling-down the IncRNA.

In addition, our Urb-RIP method has been recently coupled with mass-spectrometry to identify RBPs bound to the H/ACA snoRNA ACA11. These analyses confirmed previous results [25] and revealed novel potential interactors of ACA11 snoRNA (N Mahanaj, S Liu and M Tomasson, manuscript in preparation).

An important consideration when tagging an RNA with the SLII-tag is the location the tag is to be inserted. Here we have inserted the tag into the 3’UTR of mRNAs and the 5’ end of the IncRNA BC200. For IncRNAs we suspect the tag could be placed at either end of transcript. It would likely be best to avoid the middle as to not perturb the structure of the RNA. For mRNAs the tag should be placed in the 3’UTR. Inserting the tag into the 5’UTR or coding sequence will likely lead to displacement of 2HA-Urb from the message by the translation
machinery, or stalling of the ribosome or pre-initiation complex. In all cases the ideal location of the tag may need to be empirically determined.

One of the pitfalls of the Urb-RIP approach is the ability of the Urb-RRM1 domain to interact with endogenous U1-snRNA. However, this binding has not prevented us from identifying RBPs and RNAs bound to Urb-RIP purified mRNAs and lncRNAs. It may be possible to mitigate this issue by pre-clearing the lysate with an antibody against a U1-snRNP factor. Additionally, mutants of Urb RRM1 domain, which show similar or higher affinity to SLII hairpin [27], can be used for further improvement of the method. An additional pitfall of our method or any other RBP-mediated RNA-pulldown, such as pulldown with MS2, is that the RNA of interest is exogenous and in many cases overexpressed. This along with overexpression of the RBP used for pulldown, be it MS2 or 2HA-Urb, should be considered when designing the experiment.

Taken together our results show that Urb-RIP provides an adaptable and efficient approach for pull down of RNA of interest and their interacting proteins and ncRNAs. We predict Urb-RIP will work efficiently in most cell lines and can be coupled with many techniques for the analysis of interacting proteins and ncRNAs. Urb-RIP has the potential to become a useful tool in the study of post-transcriptional regulation of mRNA and the function of lncRNAs.
2.6 References


2.6 Supplemental Information

2.6.1 Detailed Protocol

Before you begin:

- Make PK/7 M urea buffer, recipe follows protocol.

Cross Linking:

1. Remove media from cells and wash with 10 mL ice-cold PBS.
2. Add 2 mL of ice-cold PBS to the plate to keep the cells moist. Keep plate on ice until crosslinking.
3. Crosslink in Stratalinker 1800 at 400 mJ/cm².
4. Add 8 mL of ice-cold PBS to the plate and collect cells by pipetting or scraping. Transfer suspension to a 15 mL conical tube.
5. Spin cells at 1000 g for 3 minutes at 4 °C.
6. Remove supernatant and resuspend in 1 mL of cold PBS and transfer to a 1.7 mL tube.
7. Spin cells as before.
8. Remove supernatant and estimate cell volume.

Lysis and Sample preparation.

9. Lyse cells with 3-4 volumes of 1% NP-40 lysis buffer + PI with 0.5 units/µL RNase inhibitor
   a. Lyse on ice for 20 minutes
   b. Spin at maximum speed for 20 minutes at 4 °C.
   c. Transfer supernatant to a fresh tube
10. Quantitate total protein in lysate
    a. Calculate volume of lysate needed for each sample to IP at least ~1000 µg of total protein, reserve 5-10% for input RNA and Western.
       i. Bring lysate for IP to 300 µL with lysis buffer, add RNase inhibitor to 0.5 units/µL
       ii. For input RNA add 200 µL proteinase K buffer with proteinase K (pre-incubated). Follow procedure for RNA elution below.
       iii. For western input control add sample buffer to 1x
           1. Boil for 7 minutes and store at -20 °C.
Blocking Beads

1. For each IP you will need one tube of blocked anti-HA beads
2. Add 50 uL of beads to a 1.5 mL tube add 150 uL of TBS-T, vortex
3. Separate beads with magnet and remove supernatant
4. Add 1 mL of TBS-T, mix by inversion for 1 minute, collect beads with magnetic stand and remove supernatant
5. Add 300 uL of blocking buffer (lysis buffer w/PI, 4% BSA) add 15 uL of yeast tRNA (10 mg/mL) block for 1 hr with rotation at 4 C.
6. Separate beads and remove the supernatant.
7. Wash with 300 uL of TBS-T three times, leave in last wash at 4 C until ready for binding.

Binding

1. Remove TBS-T from blocked anti-HA beads
2. Add lysate, allow to bind at 4 °C for 1 hr with rotation

Washing

1. Separate beads, remove and save supernatant
   a. Transfer 5% of supernatant (15 µL) and add sample buffer to 1x
   b. Boil for 7 minutes.
2. Wash twice with low salt wash buffer, 500 µL/wash, vortex for 10s at ~1000rpm
3. Wash twice with high-salt wash buffer, 500 µL/wash, vortex for 10s at ~1000rpm
4. Separate beads, remove the last wash and add 500 µL of pure water
5. Mix and split beads into two tubes, one for protein elution and one for RNA elution.

Elute for protein or RNA

- For protein add 1x sample buffer
  a. Boil beads for 7 minutes and transfer the supernatant to a new tube.
- For RNA add 200 µL of proteinase K buffer + proteinase K (160 µL of proteinase K buffer and 40 µL Proteinase K (NEB)). **Note**: make a mastermix of the proteinase K buffer and incubate with proteinase K for 20 minutes prior to elution to kill RNase.
  a. Incubate 20 minutes at 37 °C, 1000 rpm.
  b. Add 200 µL of PK/Urea buffer.
  c. Incubate as above.
  d. Add 400 µL of acid-phenol/chloroform, vortex and let sit for 5 minutes
e. Spin at maximum speed in cold centrifuge for 15 minutes
f. Take aqueous phase and add 1 µL of glycogen, 1/10\textsuperscript{th} volume of 3 M NaOAc (pH 5.5) and 2.5 volumes of 100% ethanol.
g. Precipitate overnight at -20 °C.
h. Spin at maximum speed in cold centrifuge for 30 minutes
i. Wash pellet in 1 mL of 70% ethanol
j. Let dry for 5 minutes at RT
k. Resuspend in 10-20 µL of water.
Supplemental Figure 2.1: Optimization of pulldown protocol.
A stable HEK-293 cell line for the inducible expression of 2HA-Urb was transfected with a plasmid expressing mCherry-mRNA untagged or tagged with SLII. Two days after transfection the cells were UV-irradiated at doses shown and subsequently lysed. Immunoprecipitation was performed using the Urb-RIP protocol and RNA was eluted with proteinase K treatment. qRT-PCR was performed using mChery and GAPDH primers. Pulldown efficiency was quantified by qRT-PCR analysis of enrichment of mCherry+SLII relative to mCherry, the abundance of both messages was normalized to GAPDH.
Supplemental Figure 2.2: RNA integrity after UV irradiation

RNA was isolated from control or UV irradiated (400 mJ/cm²) 293-2HA-Urb cells using the Qiagen RNeasy Kit per manufacturer’s protocol. Two micrograms of RNA was mixed with 3 µL of 10x MOPS buffer, 6 µL of formaldehyde and formamide to 30 µL prior to denaturation at 80 °C for 15 minutes. The RNA was cooled on ice and 2x RNA Loading Dye was added (10mM EDTA, 50% glycerol v/v, 0.25% bromophenol blue and xylene cyanol) along with ethidium bromide. The samples were loaded on a 1.2% denaturing agarose gel, resolved and the gel was imaged. 

M: 0.1-2 kB RNA ladder
1: No UV
2: 400 mJ/cm² of UV
Supplemental Figure 2.3: Optimization of blocking and preclearing.
A stable HEK-293 cell line for the inducible expression of 2HA-Urb was transfected with a plasmid expressing mCherry-mRNA untagged or tagged with SLII. Two days after transfection the cells were UV-irradiated at 400 mJ/cm² and subsequently lysed. The lysate was loaded onto untreated beads or beads blocked with 300 µL of 4% BSA, 0.5 µg/mL yeast tRNA in TBST. For one sample the lysate was cleared by incubation with Protein A/G beads for 1 hour prior to loading on the blocked beads. Following binding the beads were processed following the Urb-RIP protocol and RNA was eluted with proteinase K treatment. qRT-PCR was performed using mCherry and GAPDH primers. Pulldown efficiency was quantified by qRT-PCR analysis of enrichment of mCherry+SLII relative to mCherry, the abundance of both messages was normalized to GAPDH.
Supplemental Figure 2.4: Analysis of U1-snRNA binding during Urb-RIP.
A stable HEK-293 cell line for the inducible expression of 2HA-Urb was transfected with a plasmid expressing mCherry-mRNA untagged or tagged with SLII. Two days after transfection the cells were UV-irradiated at 400 mJ/cm² and subsequently lysed. The lysate was loaded onto untreated beads or beads blocked with 300 µL of 4% BSA, 0.5 µg/mL yeast tRNA in TBST. For one sample the lysate was cleared by incubation with Protein A/G beads for 1 hour prior to loading on the blocked beads. Following binding the beads were processed following the Urb-RIP protocol and RNA was eluted with proteinase K treatment. qRT-PCR was performed using mCherry, U1-snRNA and GAPDH primers. a Enrichment of U1-snRNA relative to the input abundance was determined by qRT-PCR, normalized to GAPDH. b Abundance of U1-snRNA in the immunoprecipitate relative to mCherry was determined by qRT-PCR, normalized to GAPDH.
Supplemental Figure 2.5: Reproducibility of Urb-RIP
Enrichment of BC200+SLII by Urb-RIP as determined by qPCR. A stable HEK-293 cell line for the inducible expression of 2HA-Urb was transfected with plasmids expressing the constructs described in Figure 4a. Two days after transfection the cells were UV-irradiated at 400 mJ/cm² and subsequently lysed. Immunoprecipitation was performed using the Urb-RIP protocol and RNA was eluted with proteinase K treatment. qRT-PCR was performed using BC200 and GAPDH primers. a and b Enrichment of BC200 by qRT-PCR, normalized to GAPDH, from two independent experiments. c and d Western blot analysis of PABP and 2HA-Urb abundance in the Urb-RIP product and input. Samples labeled input represent 5% of the total sample used for Urb-RIP.
Supplemental Figure 2.6: Analysis of RNA Pulldown by Bioanalyzer
A stable HEK-293 cell line for the inducible expression of 2HA-Urb was transfected with plasmids expressing the constructs described in Figure 4a. Two days after transfection the cells were UV-irradiated at 400 mJ/cm² and subsequently lysed. Immunoprecipitation was performed using the Urb-RIP protocol and RNA was eluted with proteinase K treatment. The eluted RNA as well as RNA from the isolated from the Urb-RIP input was analyzed by Agilent 2100 Bioanalyzer. The analysis for the input samples a and b are shown as well as the IP eluate c and d. The analysis of the IP eluate shows a strong peak for BC200 in the pull-down of BC200+SLII, d, this peak was absent in the control pulldown, c. There is a peak for the U1 and U2-snRNA in both IP eluates.
### 2.6.2 Supplementary Tables

#### Supplemental Table 2.1: Cloning primers and oligonucleotides

<table>
<thead>
<tr>
<th>Primer/Oligo Name</th>
<th>Sequence (5'-&gt;3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>2HA-RRM1-URB Forward 1</td>
<td>ATGTATCCGTATATGTGTGCGGGATTATGCGGGCGCGTGATCCG TAGTATGTGCGGGATTATGCGGGAAAAACCTGTATTTTCAGGGC GACATCCGGCCGGAACACACG</td>
</tr>
<tr>
<td>2HA-RRM1-URB Forward 2</td>
<td>CACCATGTATCCGTATGTATGTGACC</td>
</tr>
<tr>
<td>2HA-RRM1-URB Reverse</td>
<td>CATTAGCGTTTACAAAAGGTGC</td>
</tr>
<tr>
<td>2HA-URB NarI Mut. Forward</td>
<td>GGCGAAAAAACCTGTATTTTCAGGGCGGCCGACATCGGCCGCGAAGC</td>
</tr>
<tr>
<td>2HA-URB NarI Mut. Reverse</td>
<td>GGGTCGGGCGGATGTCGGCGGCCCTTGAATAACAGGTGTTCCG</td>
</tr>
<tr>
<td>Flag Oligo 1</td>
<td>CGATTACAAGGAGCGATGACGATAAGGG</td>
</tr>
<tr>
<td>Flag Oligo 2</td>
<td>CGCCCTATATCGCTATCGCTTTTTAAT</td>
</tr>
<tr>
<td>SLII Tag SacII Oligo 1</td>
<td>GGCAACAAACAAACAAACAAACAAACAAACAGGAGACCATTGCACTCCG GTTTCGCCGCAACAGATATCCGCG</td>
</tr>
<tr>
<td>SLII Tag SacII Oligo 2</td>
<td>GGGATATCTTTTGTGTGGAAGCCGGAGGTCAATGTCTCCTT GTTGTGTGTGTGTGTGTGCGC</td>
</tr>
<tr>
<td>BC200 F</td>
<td>CTAGACTTAGAATAAGGCGGGCGGCCGCTGTCAC</td>
</tr>
<tr>
<td>BC200 R (with terminator)</td>
<td>CTAGTCATGAAAAAAGGGGGGGGTTGGTGG</td>
</tr>
<tr>
<td>SLII Tag SpeI Oligo 1</td>
<td>AATTCCACACAAACACAAACAAACAAACAGGAGACCATTGCACTCCG CCGGTTCGACCAACAGATATCCG</td>
</tr>
<tr>
<td>SLII Tag SpeI Oligo 2</td>
<td>TTAAGCTATAGAACAACACACCTTTGGCGCTACGTTACCAGA GGAACAACAAACAACACACACC</td>
</tr>
<tr>
<td>CMV Deletion Oligo 1</td>
<td>CCGAGGATATCCGAGA</td>
</tr>
<tr>
<td>CMV Deletion Oligo 2</td>
<td>CTAGTCTCGGATATCCCTCGGAGCT</td>
</tr>
<tr>
<td>pAWH let-7 sites F</td>
<td>GCAGTAATTCTAGGGCGATCGC</td>
</tr>
<tr>
<td>pAWH let-7 sites R</td>
<td>CGCTGGCGCGCTGCGAGA</td>
</tr>
<tr>
<td>EGFP Forward</td>
<td>CACCATGGGCGCAGTCACAAAGGAGATCGAGGGCG</td>
</tr>
<tr>
<td>EGFP Reverse-overlap</td>
<td>GCGGGGCGCCTCCCGAGTAGGATCCCCGCGAGCTTGCGCGTTCACGAGATCGAGAGTGTCTTAC AGCTCGTCC</td>
</tr>
<tr>
<td>Ago2 Forward-overlap</td>
<td>GCTGTAACAGGCGATCGGCTGGCGGAGGCTACTCGGCGAGCGCCGCCGCGAAGTCGGACGACC</td>
</tr>
<tr>
<td>Ago2 Reverse</td>
<td>GTAGCGCGCGCTCAAGCAAGTACATGGGTGCGACAGAGTGCTTGG</td>
</tr>
</tbody>
</table>
### Supplemental Table 2.2: qPCR and Reverse Transcription Priemrs

<table>
<thead>
<tr>
<th>Primer/Oligo Name</th>
<th>Sequence (5'→3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>mCherry qRT Forward</td>
<td>CAAGGGCGAGGAGGATAACAT</td>
</tr>
<tr>
<td>mCherry qRT Reverse</td>
<td>ACATGAACCTGAAGGGACAGG</td>
</tr>
<tr>
<td>U1-snRNA qRT Forward</td>
<td>CTTACCTGCGAGGGGAGATAAC</td>
</tr>
<tr>
<td>U1-snRNA qRT Reverse</td>
<td>TCCGGAGTGCAATGGATAAG</td>
</tr>
<tr>
<td>GAPDH qRT Forward</td>
<td>GAGTCAACGGATTTGGTCGT</td>
</tr>
<tr>
<td>GAPDH qRT Reverse</td>
<td>GACAAGCTTCCTCGTTCAG</td>
</tr>
<tr>
<td>TIMM50 qRT Forward [1]</td>
<td>GCGTTGGTGTTGCGAGGTA</td>
</tr>
<tr>
<td>TIMM50 qRT Reverse [1]</td>
<td>ACGGGAGGCGGGGAAAGG</td>
</tr>
<tr>
<td>Actin qRT Forward</td>
<td>AGAAAATCTGGCACCACACC</td>
</tr>
<tr>
<td>Actin qRT Reverse</td>
<td>AGAGGCGTACAGGGGATAGCA</td>
</tr>
<tr>
<td>7SK qRT Forward [2]</td>
<td>CCCCTGTAGAATCCCTCAAC</td>
</tr>
<tr>
<td>7SK qRT Reverse [2]</td>
<td>CACATGCAGCGCTCTTTT</td>
</tr>
<tr>
<td>BC200 qRT Forward</td>
<td>CTGGGCAATATAGCGAGACC</td>
</tr>
<tr>
<td>BC200 qRT Reverse</td>
<td>GGTTGTTGCTTGGAGGGAAG</td>
</tr>
<tr>
<td>let-7 Reverse Transcription [3]</td>
<td>CGCATATCGCGTCATTACAGAAACTACAA</td>
</tr>
<tr>
<td>let-7 qRT Forward [3]</td>
<td>TCGCATATCGCGTCATTACAGA</td>
</tr>
<tr>
<td>let-7 qRT Reverse [3]</td>
<td>GCGGAGTTGAGGTAGTGGTTG</td>
</tr>
</tbody>
</table>
## Supplemental Table 2.3: Pulldown of Non-target RNAs in mCherry Pulldown

<table>
<thead>
<tr>
<th></th>
<th>Input (mCh – SLII)</th>
<th>Input (mCh + SLII)</th>
<th>IP (mCH – SLII)</th>
<th>IP (mCh + SLII)</th>
<th>% Input (mCh - IP)</th>
<th>% Input (mCh + IP)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>mCherry</strong></td>
<td>23.05</td>
<td>23.66</td>
<td>35.13</td>
<td>26.93</td>
<td>0.0231</td>
<td>10.3913</td>
</tr>
<tr>
<td><strong>GAPDH</strong></td>
<td>27.47</td>
<td>28.41</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Actin</strong></td>
<td>27.72</td>
<td>28.06</td>
<td>41.01</td>
<td>43.19</td>
<td>0.0100</td>
<td>0.0028</td>
</tr>
<tr>
<td><strong>7SK</strong></td>
<td>20.85</td>
<td>21.41</td>
<td>32.33</td>
<td>32.46</td>
<td>0.0349</td>
<td>0.0472</td>
</tr>
<tr>
<td><strong>TIMM50</strong></td>
<td>32.00</td>
<td>32.37</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>18s rRNA</strong></td>
<td>11.72</td>
<td>12.45</td>
<td>27.79</td>
<td>25.53</td>
<td>0.0015</td>
<td>0.0115</td>
</tr>
<tr>
<td><strong>U1 snRNA</strong></td>
<td>18.79</td>
<td>19.15</td>
<td>19.09</td>
<td>18.74</td>
<td>81.4040</td>
<td>132.5872</td>
</tr>
<tr>
<td><strong>U2 snRNA</strong></td>
<td>21.01</td>
<td>21.28</td>
<td>24.21</td>
<td>24.09</td>
<td>10.8633</td>
<td>14.2748</td>
</tr>
</tbody>
</table>

N/A: Not detected within 45 cycles
### Supplemental Table 2.4: Enrichment of mCherry mRNA by Urb-RIP

<table>
<thead>
<tr>
<th>qPCR Target</th>
<th>mCherry</th>
<th>mCherry</th>
<th>GAPDH</th>
<th>GAPDH</th>
<th>mCherry</th>
<th>mCherry</th>
<th>GAPDH</th>
<th>GAPDH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfected</td>
<td>mCh-SLII</td>
<td>mCh+SLII</td>
<td>mCh-SLII</td>
<td>mCh+SLII</td>
<td>mCh-SLII</td>
<td>mCh+SLII</td>
<td>mCh-SLII</td>
<td>mCh+SLII</td>
</tr>
<tr>
<td>Trial 1</td>
<td>25.04</td>
<td>25.02</td>
<td>29.29</td>
<td>29.29</td>
<td>28.44</td>
<td>21.01</td>
<td>31.32</td>
<td>32.33</td>
</tr>
<tr>
<td>Trial 2</td>
<td>19.61</td>
<td>20.32</td>
<td>19.61</td>
<td>20.32</td>
<td>25.86</td>
<td>20.15</td>
<td>32.18</td>
<td>34.78</td>
</tr>
<tr>
<td>Trial 3</td>
<td>23.74</td>
<td>25.36</td>
<td>28.99</td>
<td>29.62</td>
<td>29.24</td>
<td>21.44</td>
<td>33.38</td>
<td>33.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>mCh-GAPDH</th>
<th>mCh-GAPDH</th>
<th>Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1</td>
<td>-2.88</td>
<td>-11.32</td>
</tr>
<tr>
<td>Trial 2</td>
<td>-6.32</td>
<td>-14.63</td>
</tr>
<tr>
<td>Trial 3</td>
<td>-4.14</td>
<td>-11.88</td>
</tr>
</tbody>
</table>
### Supplemental Table S5: Pulldown of Non-target RNAs in BC200 Pulldown

<table>
<thead>
<tr>
<th></th>
<th>Average Ct</th>
<th></th>
<th></th>
<th></th>
<th>% Input (BC200 - IP)</th>
<th>% Input (BC200 + IP)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Input (BC200 - SLII)</td>
<td>Input (BC200 + SLII)</td>
<td>IP (BC200 - SLII)</td>
<td>IP (BC200 + SLII)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC200</td>
<td>14.28</td>
<td>14.08</td>
<td>26.72</td>
<td>17.48</td>
<td>0.018</td>
<td>9.433</td>
</tr>
<tr>
<td>GAPDH</td>
<td>22.73</td>
<td>22.51</td>
<td>36.12</td>
<td>40.57</td>
<td>0.009</td>
<td>0.000</td>
</tr>
<tr>
<td>Actin</td>
<td>23.42</td>
<td>23.41</td>
<td>38.69</td>
<td>35.26</td>
<td>0.003</td>
<td>0.027</td>
</tr>
<tr>
<td>5SK</td>
<td>22.34</td>
<td>22.24</td>
<td>30.21</td>
<td>30.46</td>
<td>0.428</td>
<td>0.335</td>
</tr>
<tr>
<td>TIMM50</td>
<td>26.20</td>
<td>26.44</td>
<td>35.31</td>
<td>34.51</td>
<td>0.180</td>
<td>0.372</td>
</tr>
<tr>
<td>18s rRNA</td>
<td>11.66</td>
<td>11.16</td>
<td>26.32</td>
<td>25.62</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>U1 snRNA</td>
<td>18.90</td>
<td>18.83</td>
<td>17.38</td>
<td>18.58</td>
<td>285.56</td>
<td>118.88</td>
</tr>
<tr>
<td>U2 snRNA</td>
<td>20.15</td>
<td>19.95</td>
<td>20.44</td>
<td>22.67</td>
<td>81.61</td>
<td>15.18</td>
</tr>
</tbody>
</table>

N/A: Not detected within 45 cycles
Chapter 3: Translation efficiency is a determinant of the magnitude of miRNA-mediated repression

Preface

The following work was completed by myself, Pawel Szczesny and Sergej Djuranovic. S.D. initiated and directed this research. K.A.C did the experiments. K.A.C. and P.S. did the computational analysis. K.A.C wrote the manuscript with contributions from P.S. and S.D.

This chapter is published in its entirety [Cottrell KA, Szczesny P, Djuranovic S. Translation efficiency is a major determinant of the magnitude of miRNA-mediated repression. Sci Rep 7:14884] and is available at http://rdcu.be/yImu. This article is distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

We thank Kathleen Hall, Tim Schedl and Andrew Yoo for comments. Funding from NIH (GM007067 to KAC and GM112824 to SD).
3.1 Abstract

MicroRNAs are well known regulators of mRNA stability and translation. However, the magnitude of both translational repression and mRNA decay induced by miRNA binding varies greatly between miRNA targets. This can be the result of cis and trans factors that affect miRNA binding or action. We set out to address this issue by studying how various mRNA characteristics affect miRNA-mediated repression. Using a dual luciferase reporter system, we systematically analyzed the ability of selected mRNA elements to modulate miRNA-mediated repression. We found that changing the 3’UTR of a miRNA-targeted reporter modulates translational repression by affecting the translation efficiency. This 3’UTR dependent modulation can be further altered by changing the codon-optimality or 5’UTR of the luciferase reporter. We observed maximal repression with intermediate codon optimality and weak repression with very high or low codon optimality. Analysis of ribosome profiling and RNA-seq data for endogenous miRNA targets revealed translation efficiency as a key determinant of the magnitude of miRNA-mediated translational repression. Messages with high translation efficiency were more robustly repressed. Together our results reveal modulation of miRNA-mediated repression by characteristics and features of the 5’UTR, CDS and 3’UTR.
3.2 Introduction

MicroRNAs are short, endogenous non-coding RNAs that along with associated Argonaute proteins form the miRNA-induced silencing complex (miRISC) which acts by inhibiting translation and causing mRNA decay \(^1\textsuperscript{-4}\). The magnitude of translational repression and mRNA decay for each miRNA target can vary greatly \(^5\textsuperscript{-11}\). The variation in repression for some targets can be explained by poor miRNA binding \(^12\), or RNA binding proteins (RBPs) modulating repression \(^13\textsuperscript{-15}\). It is well appreciated that alternative transcription start site selection, splicing and polyadenylation can lead to transcript variants that differ by their 5’untranslated region (UTR), coding sequence (CDS) and/or 3’UTR \(^16\textsuperscript{-18}\). In some cases, these transcript isoforms have altered repression by miRNAs \(^9,19\). We hypothesized that mRNA elements such as the 5’UTR, CDS and 3’UTR could modulate miRNA-mediated repression; to address this we systematically analyzed the effects of various mRNA elements on the magnitude of miRNA-mediated repression.

The 5’UTR, CDS and 3’UTR are important regulatory regions of the mRNA. Structure within the 5’UTR has been shown to affect mRNA translation by impeding the initiation process \(^20,21\). The presence of upstream translation start sites and upstream open reading frames (uORF) has been shown to repress translation \(^22\textsuperscript{-26}\). Along with these repressive elements the 5’UTR is also home to binding sites for RBPs that can act on mRNA translation and stability \(^27\). Like the 5’UTR the 3’UTR is an important regulatory region. The 3’UTR typically contains binding sites for many RBPs and miRNAs. The RBPs that bind to the 3’UTR can influence the translation, stability and localization of the mRNA \(^28,29\). Sandwiched between the 5’ and 3’UTR is the CDS. The CDS is a series of mRNA codons that are translated into a protein product. In recent years, it has become apparent that the stability and translation of many mRNAs is regulated by their
unique codon usage \(^{30-34}\). Messages with more optimal codons have a faster translation elongation rate and tend to be more stable \(^{30-34}\). Together, the UTRs and the CDS regulate the stability and translation of the mRNA.

Our systematic analysis showed that the magnitude of miRNA-mediated repression is dependent on the translational efficiency of the non-targeted reporter; a characteristic which can be modulated by changing the 3’UTR, codon optimality of the CDS, and 5’ UTR. Additional analysis of whole genome mRNA-seq and ribosome profiling data revealed that translation efficiency of the target mRNA is also a determinant of the magnitude of miRNA-mediated repression. Our data indicate that variation in the magnitude of miRNA-mediated translational repression observed in previous reporter and global studies \(^6, 8, 11, 35, 36\) can be, in part, explained by the variation in translation efficiency of the targeted message or influenced by the composition of the 5’UTR, CDS and 3’UTR.

### 3.3 Results

#### 3.3.1 The 3’UTR modulates translatability and miRNA-mediated repression

Using a previously defined reporter system targeted by the miRNA bantam in *Drosophila* S2 cells \(^{11}\) (Fig. 1A), we assessed the ability of the 3’UTR, CDS and 5’UTR to modulate miRNA-mediated repression. The reporter system includes a targeted (T) *Renilla* luciferase reporter that contains six target sites for the miRNA bantam in the 3’UTR and a non-targeted (NT) reporter containing reversed bantam sites, both of which are tightly controlled by the metallothionein promoter \(^{37}\). The 3’UTR has been implicated in regulation of translation and mRNA stability \(^{28, 38-40}\). In order to assess how different 3’UTRs modulate repression of our reporter system we inserted the 3’UTR of several different genes from *Drosophila melanogaster* downstream of the miRNA target sites or reversed sites in our reporters (Fig. 1B).
Figure 3.1: 3’UTR influences translatability and miRNA-mediated repression.

A Schematic describing the reporters used for this study. The 3’UTR of each of the genes in B were cloned downstream of the miRNA target sites yielding a T and NT reporter for each 3’UTR. Mtn designates the metallothionein promoter. B Table describes some attributes of the 3’UTRs used in this study. *The length of the 3’UTR as reported by Flybase (http://flybase.org/) are actual lengths of 3’UTRs based on 3’ RACE (Fig. S1C). ^Predicted number of miRNA binding sites within each 3’UTR as transcribed (3’ RACE). The value in the parenthesis represents the number of binding sites for miRNAs previously found to be expressed in Drosophila S2 cells. C The 3’UTR of the reporter greatly affected repression (ratio of NT/T for luciferase activity). Dual-luciferase assay was used to determine repression. Normalization was carried out using firefly luciferase activity. D NT reporter expression and normalized ratio NT/T for luciferase activity show a statistically significant correlation. E Expression of the NT reporter varies greatly while T reporter expression shows little variation. F Correlation between several characteristics of our 3’UTR reporters and repression. All data are depicted as mean ± SD.

The 3’UTRs of these genes were chosen because they represent a wide range of cellular functions and have widely varying lengths (Fig. 1B). None of the selected 3’UTRs contained a predicted bantam binding site. We used a dual luciferase assay to determine repression of each reporter, as done previously. Strikingly we observed a wide variation in the magnitude of
translational repression, with the reporter containing the GAPDH 3’UTR being repressed over 100-fold while the Cad87 3’UTR reporter was repressed less than 5-fold (Fig. 1C). There was very little variation in the magnitude of mRNA-degradation and no significant correlation between mRNA-degradation and translational repression (Fig. S1A and S1B). The expression of the NT reporter for each 3’UTR, however, correlated strongly with the observed translational repression, $r = 0.904$ and $r_s = 0.964$ (Fig. 1D). We define the variation in NT reporter expression as ‘translatability’. The NT reporter expression varied over two orders of magnitude. While repression by bantam was abrogated upon co-transfection with a bantam antagonir the 3’UTR dependent effect on translatability was still observed with reporters lacking the bantam target sites or the control sequence (Fig. S2). Additionally, there was a significant correlation between mRNA expression of our NT 3’UTR reporters and translatability (Fig. S3A and S3B). As such, our results indicate that both mRNA expression and the translation rate of our reporter constructs are altered by the 3’UTR. Differences in the protein expression levels (Fig. 1E), however, exceeded observed differences at the mRNA level (Fig. S1A). Interestingly, we saw less than two-fold variation in the protein expression of the targeted reporter for each 3’UTR (Fig. 1E). Together these data show that the magnitude of miRNA-mediated translational repression is dependent on the translatability of the target mRNA.

We did not observe correlation between repression or translatability and several other characteristics of the 3’UTRs (Fig. 1F). We performed 3’ rapid amplification of cDNA ends (3’RACE) to determine the length of each 3’UTR as expressed. While there was not a statistically significant correlation between 3’UTR length and translational repression for all our reporters, (Fig. 1F), we did observe a negative correlation for several 3’UTR reporters (Fig. S1, D-E). It is well known that the presence and length of the poly(A)-tail can influence translation
and multiple models have indicated that the progressive shortening of the poly(A)-tail is one of the mechanisms through which miRNAs exert translational repression on their targets. Using a commercially available poly(A)-tail length assay, we determined the poly(A)-tail length of both the NT and T reporters. Consistent with previous work in *Drosophila* S2 cells we did not observe a shortening of the poly(A)-tail in our miRNA-target reporters (Fig S3C-E). To test whether the presence of a poly(A)-tail is needed for the observed correlation between translatability and repression of our reporters we replaced the poly(A)-tail with the histone H3 stem loop (H3-SL). The switch from the poly(A)-tail to the H3-SL created an additional set of variations in the magnitude of miRNA-mediated repression (Fig. S4A). We have again observed high variability of NT reporter expression compared to T reporter at the protein level (Fig. S4 B and C). We again did not observe any significant effects on mRNA ratio between NT and T reporters (Fig. S4D). While there was a clear difference between the translatability of reporters terminated by a poly(A)-tail or the H3-SL (Fig. S4B), the presence or absence of poly(A)-tail did not influence the correlation between translatability and miRNA-mediated repression (Fig. 1D and S4E). This result indicates that the observed correlation of the magnitude of miRNA-mediated repression and translatability is independent of the poly(A)-tail.

### 3.3.1 miRNA-mediated repression is modulated by changes in codon optimality and 5’UTR

Having observed a wide variation in the translatability of our reporters simply by changing the 3’UTR, we wanted to explore how changes in other parts of the mRNA could affect miRNA-mediated repression. We changed the codon optimality of *Renilla* luciferase in our GAPDH, CrebA and Cad87 reporters. The original coding sequence (CDS) of *Renilla* had an optimality of 0.387 by the tRNA adaption index (tAI). This value is close to the median tAI of all *D. melanogaster* genes (Fig. 2A).
Figure 3.2: Codon optimality and 5'UTR elements modulate miRNA-mediated repression.
The coding sequence of Renilla luciferase in the 3'UTR reporters for GAPDH, CrebA and Cad87 was modified to increase or decrease the codon optimality. A The tAI (tRNA Adaptation Index) of the original Renilla luciferase along with the modified versions is described. The inset histogram describes the frequency of tAI across all D. melanogaster genes. B Repression as determined by dual-luciferase assay was robust for reporters containing Renilla luciferase with moderate codon optimality, 0.387 and 0.494 tAI. *** p<0.001 by ANOVA. Stem-loop (SL) structures were inserted into the 5'UTR of the reporters containing the GAPDH and Cad87 3'UTR. C Schematic describing the 5'UTR inserts used in panel D. D Repression was reduced for the GAPDH reporter containing the 15 bp SL inserted in the 5'UTR, while repression of the Cad87 reporter was unaffected by changes to the 5'UTR structure. ** p<0.01 by t-test. All data are depicted as mean ± SD.

To sample a range of different tAI, values we created reporters with tAI of 0.602, 0.494 and 0.298 (Fig. 2A). Reducing the codon optimality of Renilla luciferase within our reporters reduced expression of the NT and T reporters by nearly three orders of magnitude (Fig. S5B and C). Interestingly, the repression of the GAPDH, CrebA and Cad87 3’UTR reporters was affected differently by changes to codon optimality (Fig. 2B). The GAPDH and CrebA reporters had peak repression when using the 0.387 tAI CDS, while the Cad87 reporter had peak repression using the 0.494 tAI CDS. All reporters showed reduced repression with the highest and lowest codon optimalities, 0.298 and 0.602. Consistent with our previous results, we found that the repression
and expression of the reporters were affected differently by changes to the codon optimality depending on which 3’UTR was present (Fig. S5). This result again highlights the interaction between the 3’UTR and translatability. Together, the 3’UTR and, codon optimality determine the magnitude of miRNA-mediated repression.

To further examine how mRNA elements may influence miRNA-mediated repression we assayed effects of the 5’UTR on miRNA-mediated repression. Recent evidence supports a model in which the miRISC inhibits translation via targeting the helicase eIF4A during 5’UTR scanning in search for a start codon. It is also known that introduction of secondary structure into 5’UTR affects initiation rate and total protein output. Therefore, we inserted stem-loop structures into the 5’UTR of our GAPDH and Cad87 3’UTR reporters, which showed maximal and minimal miRNA-mediated repression respectively (Fig. 1C and 2C). As seen in previous studies, insertion of the stem-loop structures in the 5’UTR greatly reduced reporter expression measured by luciferase activity (Fig. S6, A and B). Interestingly, we found that the addition of stem-loop structures had no effect on repression of the Cad87 3’UTR reporter, which was minimal with the control insert (Fig. 2D). For our GAPDH 3’UTR reporter, which showed maximal repression with the control insert, we observed reduced repression upon insertion of a 15 bp stem-loop (Fig. 2D). In addition to testing the effect of specific 5’UTR elements, we also made reporters with 5’UTRs from Drosophila mRNAs. In particular, we paired the 3’UTR of the reporters described in Figure 1 with their cognate 5’UTR. We again observed wide variation in the magnitude of repression, consistent with similar previous studies (Fig. S6C). To test the effect of short upstream ORFs on miRNA-mediated repression we inserted a short sequence coding for hemagglutinin-A epitope 55 nt upstream of the Renilla luciferase start site. The introduction of this short uORF in the GAPDH 3’UTR reporter increased repression by two fold.
(Fig. S6E). These results should be taken with caution, however, since translation of uORFs usually leads to activation of mRNA surveillance mechanisms. These events usually result in efficient and targeted mRNA decay of mRNAs with translated uORFs. Therefore, it is not surprising that we observed great reduction in protein output of the downstream encoded luciferase reporter (Fig. S6F). Our data on 5’UTR structure and codon optimality in the context of different 3’UTRs indicate a connection between translatability and magnitude of miRNA-mediated translational repression.

### 3.3.2 Translation efficiency is a determinant of miRNA-mediated repression

While reporter studies are valuable for understanding the mechanism of miRNA-mediated translational repression and mRNA deadenylation and degradation, they might be limited since they study a relatively small number of targeted messages in controllable in vivo or in vitro conditions. MicroRNAs in living cells act on hundreds of endogenous genes which have more varied mRNA sequences than reporters. Moreover, cellular physiology is under constant change due to the complex level of transcriptional, translational and post-translational control, which are influenced by developmental, environmental and other physiological cues. In order to test the generality of our reporter studies, we turned to whole genome analysis of miRNA targets in HeLa cells. The most striking observation made using our reporters is the strong correlation seen between translatability and repression. We compared fold change of ribosome protected fragments (RPF) for miR-155 targets following mock transfection or miR-155 transfection with translation efficiency of the target message in the absence of the miRNA (mock transfection). Translation efficiency (TE) is the ratio of RPF and RNA abundance determined by ribosome profiling and RNA-seq. We consider TE as a good proxy for
translatability. Comparison of TE and fold change of RPF revealed a strong interaction (Fig. 3A, C and E, Fig. S7).

Figure 3.3: Translation efficiency is a determinant of the magnitude of miRNA-mediated translational repression but not RNA degradation of endogenous miRNA targets
Cumulative distributions of fold change of RPF (ribosome protected fragments), A and C or fold change of RNA, B for all miR-155 predicted targets (http://targetscan.org/) in data from Guo et al., 2010. Fold change is calculated as the log2 normalized RPF or RNAseq reads for miR-155 transfected divided by mock transfected. TE for each transcript in the absence of miR-155 (mock transfection) was calculated by normalized RPF divided by normalized RNAseq reads. All miR-155 targets are binned by TE, above or below the median (“High” or “Low”), A and B or by TE quartiles, C. D and F, Correspondence between RPF fold change and TE, D, or RNA fold change and TE, F. E miR-155 targets were binned by the number of conserved and poorly conserved binding sites for miR-155 as well as TE. ** p<0.01, *** p<0.001 by Kolmogorov–Smirnov test.

We binned all miR-155 targets by TE, either above or below the median TE or by quartiles. Transcripts with high TE were more repressed than those with low TE (log2 of RPF median fold
change: High\textsubscript{TE} = -0.486, Low\textsubscript{TE} = -0.044, All miR-155 targets = -0.346). Separating miR-155 targets by TE quartile produced similar results (High = -0.588, Med.High = -0.419, Med. Low = -0.185, Low = 0.093). Further stratifying the miR-155 targets into messages containing a 6-mer, 7-mer-a1, 7-mer-m8 and 8-mer binding sites revealed that this interaction is not dependent on the type of miRNA-target pairing that is present (Fig. S8). These results were consistent with those of miR-1 transfection (Fig. S9). Interestingly, TE had no influence on miRNA-mediated mRNA decay (Fig. 3, B and F). Since a correlation between translational repression (Fold Change of RPF) and mRNA-decay (Fold Change of RNAseq) had been shown previously \(^6, 53\), the observation that TE influences translational repression but not mRNA-decay is surprising. This result suggests that the correlation between translational repression and mRNA-decay seen previously might be also dependent on TE. This was confirmed by the nonlinear model fitting where the Pearson correlation coefficient between measured and predicted RPF fold change was much improved (0.44 vs 0.74) when TE was available as a variable in addition to the measurements of mRNA levels (Fig. S10). UTR-related variables (like length or MFE) did not improve the basic model.

Using the same approach, we analyzed recent data of miR-155 induced response in lipopolysaccharide (LPS) stimulated B-cells (Fig. S11). miR-155 is induced upon LPS stimulation in primary macrophages, dendritic cells, B and T cells \(^54-56\). Compared to exogenous expression of miR-155 in Hela cells (Guo et al., 2010), increased levels of miR-155 during LPS response is required for both translational and transcriptional activation and differentiation of B and T cells to cells characterized by production of IgM and switched antigen-specific antibodies \(^57, 58\). Results from our analysis of this environmentally induced miRNA response further support our earlier observations and the correlation between translation efficiency and the magnitude of
miRNA-mediated repression. Transcripts with high TE were more repressed than those with low TE (Fig. S11). This trend was observed at multiple time points and was specific for miR-155 targets (Fig. S12). Additional analysis of gene ontology groups (Fig. S13) identified similarly enriched functions for groups of genes selected by higher-than-median TE values or, independently, the significant level of repression (RPF FC values below -0.25), which independently implies the correlation between these two variables.

Having previously shown interactions between mRNA characteristics such as codon optimality and 5’UTR structure with miRNA-mediated repression we sought to study these interactions globally using ribosome profiling data. Interestingly, we did not observe any influence of 3’UTR length, transcript length, tAI, 5’UTR structure or 3’UTR structure on miRNA-mediated repression (Fig. S9, 14 and 15). We also did not find correlation between miRNA-mediated mRNA degradation or translational repression with global measurements of mRNA half-lives \(^{59}\) (Fig. S16). However, TE, especially in combination with mRNA degradation, was predictive of the magnitude of miRNA-mediated repression.

### 3.4 Discussion

By using a systematic approach, we have revealed several mRNA elements capable of modulating miRNA-mediated repression. Our observations suggest efficient translational repression by the miRISC depends on the translation efficiency of the target.

Using luciferase reporters in *Drosophila* S2 cells we observed a strong interaction between the 3’UTR and the magnitude of repression. This result was largely driven by 3’UTR dependent differences in the translatability of the reporters. Translatability is likely the output of different mRNA characteristics such as *cis* and *trans* factors that modulate translation rate and mRNA stability. 3’UTR characteristics such as GC content and structure could not explain this
variation. While there was not a significant correlation between 3′UTR length and translatability there was a trend for a few of the 3′UTRs tested. This observation is supported by analysis of several whole-genome studies of miRNA-mediated repression which showed messages containing shorter 3′UTRs are more repressed than messages with long 3′UTRs \(^{60}\). Beyond structural features of the 3′UTR, the presence of RBPs or miRNAs likely influences the translatability of some of the 3′UTRs tested. Several RBPs have been shown to modulate miRNA-mediated repression \(^{13-15}\). We cannot exclude that additional binding of RBPs or other miRNAs to the assayed 3′UTRs may also affect the translation rate but we assume that these effects are preserved in both targeted and non-targeted reporter. In order to thoroughly address the possibility of RBPs modulating the miRNA-mediated repression of our 3′UTR reporters we need a more thorough understanding of which RBPs are bound to those 3′UTRs and how those RBPs functionally interact with the miRISC. Beyond RBPs many of the miRNAs that are predicted to target the 3′UTRs are either not expressed or expressed at a very low level (Table S2).

Upon changing the codon optimality of our miRNA-targeted reporters we observed variation in repression. Reporters with very high or low codon optimality were poorly repressed compared to reporters with intermediate optimality. This was true for all reporters but there were differences in the expression and repression of the reporters that were 3′UTR dependent. This observation suggests some interplay between the 3′UTR and codon optimality, which is consistent with recent report that the stability of maternally deposited mRNAs in zebrafish is regulated by the combined effect of codon optimality and 3′UTR length \(^{61}\). Furthermore, the variability of miRNA-mediated repression caused by changes in codon optimality indicates again that translatability has an influence on the magnitude of miRNA-mediated repression.
Paradoxically, the reporters with the highest codon optimality and highest expression were poorly repressed. A possible explanation for this finding can be that the miRISC is most effective at inhibiting the translation of efficiently translated mRNAs. While codon optimality is thought to influence the rate of translation elongation, the overall rate of translation includes the rate of initiation and termination. Our results suggest that when one of these rates are changed, but not the others, the efficiency of miRNA-mediated repression is altered. An intriguing possible explanation for the effects of the various 3’UTRs on miRNA-mediated repression is that each 3’UTR is affecting either the initiation or elongation rate. This could help to explain the interplay between codon optimality and the 3’UTR, in one potential scenario the 3’UTR is increasing or decreasing the initiation rate which could enhance or repress the effects of changing the codon optimality. For instance, the overall translation rate of a message with very slow initiation may be less affected by increasing codon optimality. This balance between initiation rate and elongation rate would be reflected as a change in TE. Messages with more balanced translation would have higher TE, and as we have shown messages with higher TE are more repressed by the miRISC.

Our analysis of previously published ribosome profiling data revealed TE to be a determinant of miRNA-mediated repression. This observation was true for several miRNAs across multiple cell lines. This finding was consistent with our 3’UTR reporter study where we observed a correlation between the translatability of the reporter and its miRNA-mediated repression. In the context of what is known about miRNA-mediated repression these findings make sense. Since the miRISC inhibits translation, messages that are translated well should show the most repression. These findings and those made using reporters help to explain the wide variation seen in the magnitude of translational repression using various reporters and in whole-
genome studies of miRNA function. We were unable to find any correlation between miRNA-mediated repression and various mRNA characteristics within the ribosome profiling data. We suspect that since each transcript possesses many varied features (tAI, CDS length, transcript length, 3’UTR length, 5’UTR length, 5’UTR structure, binding sites for miRNAs, RBPs, the presence of uORFs, etc.) that the interactions of any one of this features and miRNA-mediated repression are subtle due to this complexity. Perhaps with more knowledge of the interactions of these elements with each other and the miRISC a more sophisticated model could be built to predict the magnitude of miRNA-mediated repression.

Our data also help to further define the mechanism of miRNA-mediated repression. When considering our results with kinetic analyses of miRNA-mediated repression, which have previously shown translational repression preceding mRNA-decay \(^5, 7, 10, 11, 36\); a model for miRISC function can be generated in which translational repression precedes mRNA decay, and while the magnitude of translational repression is dependent on TE the magnitude of mRNA-decay is not (Fig. 4). Recently it has been observed that miRNA targeted mRNAs can be degraded co-translationally \(^51\). This observation directly links the translation status of a miRNA target with its decay. We suspect that the magnitude of mRNA-decay is dependent on the susceptibility of the message to deadenylation and decay which may vary from cell-to-cell based on the abundance of decapping/deadenylation factors and from message-to-message based on the presence of cis and trans elements that affect this process. This model therefore allows for a scenario in which an mRNA may serve as an effective target for translational repression because of its TE but not for mRNA-decay or vice versa. Our model fits well with the hypothesis that miRNAs serve dual functions: to induce robust changes in gene-expression during development
and other biological processes or small changes in gene-expression to balance stochastic gene-transcription \(^1\).

![Diagram](image)

**Figure 3.4: Model - Translation efficiency and mRNA elements influence the magnitude of miRNA-mediated repression.**

Multiple mRNA elements along with translation efficiency influence the magnitude of miRNA-mediated repression. MiRNA targets with relatively high TE will be more robustly repressed than targets with relatively low TE. Some mRNA elements may directly influence the magnitude of miRNA-mediated repression while others may have an indirect effect by changing TE.

Finally, our analysis of endogenous miRNA targets highlights the difficulty of studying the effects of mRNA elements and characteristics such as translation elongation and initiation rates on miRNA-mediated repression at the whole genome level. Each message possesses so many variables that the effects of any one variable on miRNA-mediated repression are masked. Additionally, biological processes are under complex control at different molecular levels. An example of this can be seen during activation of immune cells where changes in the transcriptome and proteome results from epigenetic, transcriptional and post-transcriptional regulation, controlled in part by miRNAs \(^{62-64}\). Due to the complexity of these changes, and interactions between key factors at each regulatory level, it will be hard to tease apart the direct influence of one specific factor, even for post-transcriptional regulation alone \(^{62}\). Therefore, one
approach to study the detailed mechanics of miRNA-mediated repression is to use reporters in appropriate cells types. More comprehensive analysis of the effects of cis and trans elements of the mRNA on miRNA-mediated repression will be essential for pinpointing the mechanism of miRNA-mediated repression and refining models of effective miRNA target prediction.

3.5 Materials and Methods

3.5.1 Construction of Reporters

All primers used for cloning can be found in the Supplemental Table 1. Renilla luciferase along with either six bantam sites or six flipped bantam sites were PCR amplified from pMT-DEST48-HID and pMT-DEST48 FLP using the Renilla forward and HID/FLP overlap reverse primers. This PCR product was used in overlap PCR to construct the 3’UTR reporters used in the study. The 3’UTRs were amplified by the primers designated in Supplemental Table 1 (for example: GAPDH forward overlap and reverse). The forward primer for each 3’UTR contained a 25-26 nt sequence complimentary to the HID/FLP overlap reverse primer used above. The PCR product for each 3’UTR and the PCR product containing Renilla luciferase and the targeted/non-targeted bantam sites were used in overlap PCR with a Renilla forward primer and a reverse primer specific for each 3’UTR. This product was then cloned into pENTR/D-TOPO (Invitrogen) per the manufacturer’s protocol. For the Rpl32 3’UTR reporter a reverse primer containing the Rpl32 3’UTR (Rpl32 reverse) was used to add the Rpl32 3’UTR to the PCR product containing Renilla luciferase and the bantam sites. The constructs were confirmed by Sanger sequencing and subsequently cloned into the pMT-DEST48-p(A)sΔ plasmid using LR-Clonase (Invitrogen). These constructs were confirmed by Sanger sequencing. The pMT-DEST48-p(A)sΔ plasmid was made by site directed mutagenesis to remove the SV40 p(A) signal
from pMT-DEST48 (Invitrogen) using the SV40 p(A)s mutagenesis forward and reverse primers.

To make 3’UTR reporters terminated by the H3 stem-loop we first constructed pMT-DEST48-H3. The pMT-DEST-48-p(A)sΔ plasmid was digested with Pmel and subsequently ligated with oligonucleotides H3-SL oligonucleotide 1 and 2. The 3’UTR for GAPDH, Hsp70, Alpha-tubulin, beta-tubulin and CrebA were amplified with the forward primer used for the initial cloning of the 3’UTR and a reverse primer located upstream of the native p(A) signal (for example: GAPDH-p(A) reverse). Overlap PCR was performed as described above to fuse Renilla luciferase and the bantam sites with the 3’UTR and this product was subsequently cloned into pENTR/D-TOPO (Invitrogen) per the manufacturer’s protocol. The constructs were confirmed by Sanger sequencing and subsequently cloned into the pMT-DEST-48-H3 plasmid using LR-Clonase (Invitrogen).

To make 3’UTR reporters with codon modified Renilla coding sequences we digested the expression plasmid for each 3’UTR reporter (for example: pMT-pAs-GAPDH) with NcoI and KpnI to remove Renilla luciferase. The digest was resolved on an agarose gel and the appropriate band was excised and purified. This product was then ligated with coding sequence for Renilla luciferase with a tAI of 0.602, 0.494 or 0.298. The Renilla luciferase coding sequence was synthesized by Invitrogen (coding sequence shown in Supplementary Information) and was digested with NcoI and KpnI prior to ligation. The resulting constructs were confirmed by Sanger sequencing.

To make insertions into the 5’UTR of our 3’UTR reporters we digested the desired 3’UTR reporter with SacII. The digest product was ligated with oligos containing the 5’UTR
insert (control and stem-loops of 9, 12, 15 and 18 bps as well as the uORF control or uORF). The resulting constructs were confirmed by Sanger sequencing.

To insert the cognate 5’UTR for each 3’UTR reporter described above we digested the vector containing the 3’UTR reporter (pMT-DEST48) with MscI and NcoI. This digest removed the 5’UTR present in the vector. The 5’UTRs to be inserted were PCR amplified from S2 cell cDNA with primers described in Table S1. The forward primer for each contained 20 nt corresponding to the transcription start site and flanking bases (5’ CCAATGTGCATCAGTTGTGG 3’) that were removed from the vector by the digest. The PCR product was digested with NcoI and the product was ligated into the digested plasmids described above.

The 3’UTR reporters were made without bantam target sites or control sequences by overlap PCR using primers designed to amplify the 3’UTR and primers designed to amplify Renilla luciferase described in Table S1. This product was then cloned into pENTR/D-TOPO (Invitrogen) per the manufacturer’s protocol. The constructs were confirmed by Sanger sequencing and subsequently cloned into the pMT-DEST48-p(A)sΔ plasmid using LR-Clonase (Invitrogen).

3.5.2 Transfection and Luciferase Assay

For most experiments 100 ng of the 3’UTR reporter as well as 100 ng each of pMT-firefly-luciferase, pAC-bantam and 200 ng of pMT-bantam were transfected into one well of a six well dish containing drosophila S2 cells (Invitrogen). The transfection was performed using Effectene (Qiagen) per the manufacturer’s protocol. Four hours after transfection the media was removed and replaced with media containing 500 µg/mL CuSO₄ to induce expression of the 3’UTR reporter as well as firefly luciferase and bantam. Two hours post induction the media was
removed and the cells were briefly washed with media containing 50 µg/mL bathocuproine disulfonate (BCS). Following the wash 2 mL of media containing 50 µg/mL BCS was added to each well and the cells were resuspended and split between two wells in separate 12-well plates. The cells were then allowed to incubate for 16 hours. After 16 hours one of the 12-well plates was harvested for measurement of luciferase activity while the other was used to isolate RNA, see below. For the luciferase assay the culture media was removed from the cells and 250-400 µL of 1x Passive Lysis Buffer (Promega) was added. The cells were lysed for 15 minutes while rocking at room temperature. The lysate was cleared by centrifugation at 21,000 g for 1 minute. An aliquot of the lysate was then used to measure firefly and Renilla luciferase activity using the Dual Glo Luciferase System (Promega) and the Glomax plate reader (Promega) per the manufacturer’s instructions. All luciferase assays were performed in triplicate. Renilla luciferase activity was normalized to firefly luciferase.

For co-transfection of antagonomers, the transfection was carried out in a 12-well format. Cells were transfected with 50 ng of the 3’UTR reporter as well as 50 ng each of pMT-firefly-luciferase, pAC-bantam and 100 ng of pMT-bantam or for antagonim treated cells 50 ng of the 3’UTR reporter as well as 50 ng pMT-firefly-luciferase, 200 nM batnam antagonim (IDT) and 150 ng of pMT-CFP to maintain the DNA concentration. The cells were induced and harvested as described above.

3.5.3 RNA Analysis

RNA was extracted from S2 cells using either Ribosol (Amresco) or SIGMA RNA mini-prep per manufacturer’s protocol. Isolated RNA was DNase treated with Turbo DNase (Ambion) prior to cDNA synthesis. For cDNA synthesis 5x iScript Supermix (Bio-Rad) was used per manufacturer’s protocol. Quantitative PCR was performed with primers targeting Renilla
luciferase or firefly luciferase, Supplemental Table 1. For 3’RACE: cDNA was synthesized using 3’RACE RT primer and 5x iScript Select (Bio-Rad) per manufacturer’s protocol. First round PCR was performed with Renilla-tail forward and 3’RACE External Amp primers. Second round PCR was performed with the overlap-forward primer for each 3’UTR being amplified (for example: GAPDH forward overlap) and 3’RACE amplification primer. The PCR products were resolved on an agarose gel. Prominent bands were excised and sequenced by Sanger sequencing. For qPCR of mature miRNAs we followed the protocol described previously 65. The primers used for this analysis are described in Table S3.

For analysis of p(A)-tail length we used the Poly(A) Tail-Length Assay Kit from Thermo-Fisher. The assay was performed per the manufacturer’s protocol. The primers used are described in Table S1: Hsp70a R, GAPDH R2 and HID/FLP F.

3.5.4 Cell Culture

Drosophila Schneider 2 (S2) cells were maintained in High Five Serum Free Media (Invitrogen) supplemented with 1 x penicillin, streptomycin and glutamine (PSG) (Gibco) and 20 mM glutamine (Gibco).

3.5.5 Analysis of Ribosome Profiling and RNA-seq

The accession number for ribosome profiling and RNA-seq data used in this study is GSE22004. Fold change of RPF and RNA-seq was calculated as described in Guo et al., 2010 6. Translation efficiency (TE) was calculated using RPF and RNA-seq rpkM from mock transfection, TE = (rpkM_{RPF}/rpkM_{RNA}). We obtained transcript, CDS and 3’UTR length for human genes from Ensembl using BioMart 66,67. mRNA half-lives were obtained from 5’-bromo-uridine (BrU) immunoprecipitation chase-deep sequencing analysis of HeLa mRNAs 59. miR-
155 or miR-1 targets were predicted using TargetScan\textsuperscript{60, 68, 69}. The tRNA adaptation index for each gene was calculated using CodonR (https://github.com/dbgoodman/ecre_cds_analysis/tree/master/codonR). For this analysis the CDS of all human or \textit{Drosophila} genes was obtained from the UCSC Table Browser\textsuperscript{70} and the tRNA gene table for human or \textit{Drosophila} was obtained from the GtRNAdb\textsuperscript{71}. Analysis of ribosome profiling and RNA-seq data was performed in R 3.2.4\textsuperscript{55} using packages ggplot2\textsuperscript{56} and extrafont\textsuperscript{57}. Scripts in R used for analysis are available at the Github repository under MIT license (https://github.com/freesci/translationefficiency). Gene ontology terms enrichment assessed with FunRich\textsuperscript{72}.

3.5.6 Model fitting

All genes with complete information (mRNA and RPF levels) from miR-155 repression experiment were further analyzed for the relationships between fold change, TE and other variables. In addition to statistics collected above, we have calculated MFE of both UTRs using Vienna package\textsuperscript{73} and normalized against sequence length using the approach described by Trotta\textsuperscript{74}. These variables were later imported into Eureqa software from Nutonian that dynamically fits a variety of equations into the data. Several experiments were done using different approaches to scoring function, from absolute error (the software default) to $R^2$ coefficient of determination which was chosen for the final plots.
3.6 References


60. Saito, T. & Sætrom, P. Target gene expression levels and competition between transfected and endogenous microRNAs are strong confounding factors in microRNA high-throughput experiments. *Silence* 3, 3 (2012).


3.7 Supplemental Information

3.7.1 Renilla Luciferase Coding Sequences

>Ren_luc_0.298

ATGGGGGACTTCTAAAGTATATGACCCCTGAACAAAGAAAAAGAATGATAACTGGGCGCTCAATGGTGGG
CAAGGTGTAAACAATGAAGTATAGCTACTGACTCTTTTAAATATATATATGACTCTGAAAGCATGCAGA
AATGCAGTAATATTTTTACATGGGAAATGCAGCATCTCTTATCTATGGGACATGTAGTTCTCTCATATA
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG

>Ren_luc_0.387

ATGGGGGACTTCTAAAGTATATGACCCCTGAACAAAGAAAAAGAATGATAACTGGGCGCTCAATGGTGGG
CAAGGTGTAAACAATGAAGTATAGCTACTGACTCTTTTAAATATATATATGACTCTGAAAGCATGCAGA
AATGCAGTAATATTTTTACATGGGAAATGCAGCATCTCTTATCTATGGGACATGTAGTTCTCTCATATA
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG
GAACCTGTAGGGCAGGTGTATAATTCTACTTTAATAGGGATGGGAAATGCAGTAAATTTTTTACATGGG

TGGAAACCATGTTGCCATCAAAAAATCATGAGAAAGTTAGAACCAGAAGAAATTTGCAGCATATCTTGAA
CCATTCAAAGAGAAAGGTGAAAGTTCCGTCCCTGCTCAACATTATCATAGTGGCTTATGGAAATCCCGTGAAT
AGGTGGTAAACCTGACGTTGTAACAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGATT
ACAAAAATGTTTATTTGAAACTCGACCAGAAGATTCTTTCTAAATTTTCGCAAGAAGATGCAACCTGTGA
GAAAAATATATCAAACTCGTTCTGGAGCGAGTTCTCTCAAAYATGAACAAT

>Ren_luc_0.494
ATGGGCACCTCAAGGTGTTACGATCCAGACGATGAAAGCTGGAACCTGAGGAGTTTGCCGCCTATCTTGAA
AGGTGGTAAACCTGACGTTGTAACAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGATT
ACAAAAATGTTTATTTGAAACTCGACCAGAAGATTCTTTCTAAATTTTCGCAAGAAGATGCAACCTGTGA
GAAAAATATATCAAACTCGTTCTGGAGCGAGTTCTCTCAAAYATGAACAAT

>Ren_luc_0.602
ATGGGCACCTCAAGGTGTTACGATCCAGACGATGAAAGCTGGAACCTGAGGAGTTTGCCGCCTATCTTGAA
AGGTGGTAAACCTGACGTTGTAACAAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGATT
ACAAAAATGTTTATTTGAAACTCGACCAGAAGATTCTTTCTAAATTTTCGCAAGAAGATGCAACCTGTGA
GAAAAATATATCAAACTCGTTCTGGAGCGAGTTCTCTCAAAYATGAACAAT
CTCCTACCGCCTGCTGGACCACTACAAGTACCTGACCACCTGGTCTCGAGCTGCTGAACCTGCCCAAGA
AGATCATCTTCGTGGCCACGACTGGGGCGCCTGCCTGGCCTCCACTACTCCTACGAGCACCAGGAC
ACATCGAGGAGGAACATCGCCCTTGATCAAGTCGAGGAGGAGGACATCGCCCTGATCAAGTCGAGGAGGGCGAGAAGATGGTGCTGGAGAACAACTT
CTTCGTGGAGACCATGCTGCCCCCTCCAAGAGATCATCGAAGCTGGAGCCCGAGGAGTTCGCCGCCTACC
TGGAGCCCTTTCAAGGAGAAGGGCGAGGTGCGCCGCCCAACCCCTGAATGTGGCCACGCAGATCCACCT
GGTGAAAGGCGGGCAAGCCCCGACGTCGTGCAGATCGGCAACTACACGCCTACCTGCGCGCCTCC
GACGACCTGGCAAGATGGTTCATCAGTGCGACCCGCTCCTCTTCTCCAACGCACTCAGTGAGGAGGC
CAAGAAGTTCCCAACACCAGTTCTGGAAGGTTGGAAGGGCGCTCGACCTTCTCCAGGAGGAGAGGCCCCCG
ACAGAATGGGGAAGATCAGTCTCGTGAGCAGCCTGCTGAAGAACGACGAGTAA
Supplemental Figure 3.1: RNA-decay and 3'UTR length of Renilla Luciferase Reporters

A The 3'UTR of the reporter minimally affected mRNA degradation. Quantitative RT-PCR was performed to determine fold RNA degradation (Normalized Ratio NT/T for mRNA). B There is no correlation between mRNA degradation and fold repression (normalized ratio of NT/T for luciferase activity). C Agarose gel of 3’RACE products for 3’UTR reporters used in this study. Stars indicate bands that were excised and sequenced. D Correspondence between the length of each 3’UTR (assigned by the most prominent band for each 3’UTR in C) as determined by sequencing of C and repression (normalized ratio of NT/T for luciferase activity) or NT Expression E. All data are depicted as mean ± SD.
Supplemental Figure 3.2: Bantam antagonmir abrogates repression of 3’UTR reporters

A Normalized ratio of NT/T expression as determined by luciferase assay. Co-transfection with bantam antagonmir with a final concentration of 200 nM caused derepression of each 3’UTR reporter tested. B The normalized NT expression for the reporters in panel A is shown. There is no significant increase in expression upon co-transfection with bantam antagonmir. C The bantam T/NT sites from the reporters in Figure 1A were removed and the 3’UTR was directly fused to the Renilla luciferase coding sequence. The expression of these new 3’UTR reporters is shown in panel C. All data are depicted as mean ± SD.
Supplemental Figure 3.3: RNA abundance correlates with non-targeted reporter protein expression

A Normalized mRNA abundances for each 3’UTR reporter measured in Figure 1 of the main text. B Correlation between mRNA abundances and NT reporter luciferase activity. All data are depicted as mean ± SD. C Poly-A tail length analysis for GAPDH and Hsp70a 3’UTR reporters. The p(A)-tail length was assayed using the GI-tailing approach. GSP control refers to PCR amplified using a GSP that binds at the end of the 3’UTR and a forward primer that binds at the end of the bantam target/non-target sites. For p(A)-tail PCR a p(C) primer was used. Panels D and E show the band intensity along the y-axis of the gel for each lane.
Supplemental Figure 3.4: Reporter transcripts terminated by the histone H3 stem loop show correlation between repression and non-targeted reporter expression

A The 3'UTR reporters from Figure 1 were cloned with the histone H3 stem loop (H3-SL) which results in non-poly(A)-tail and PABP driven translation through stem-loop binding protein (SLBP). Dual-luciferase assay was used to determine repression of each reporter. Normalization was carried out using firefly luciferase activity. B Expression levels of the NT reporters with poly(A)-tail and H3-SL. C Expression levels of the T reporters with poly(A)-tail and H3-SL. D Reporters terminated by the H3-SL show minimal miRNA-mediated mRNA degradation as measured by the ratio of NT/T Renilla mRNA by qRT-PCR. E Reporters with poly(A)-tail or the H3-SL show correlation between translatability (NT expression) and repression (normalized ratio of NT/T for luciferase activity). All data are depicted as mean ± SD.
Supplemental Figure 3.5: Reporters with moderate codon optimality show robust repression
A Normalized ratio NT/T for luciferase activity from Figure 3B is shown as relative repression, with all values for each 3’UTR set relative to the maximal repression observed for each 3’UTR reporter. B The NT reporter expression of all 3’UTR reporters is shown. NT reporter Renilla luciferase activity was normalized to firefly luciferase activity and set relative to expression of the 0.602 tAI reporter. C The T reporter expression of all 3’UTR reporters is shown. T reporter Renilla luciferase activity was normalized to firefly activity and set relative to expression of the 0.602 tAI reporter. All data are depicted as mean ± SD.
Supplemental Figure 3.6: 5’UTR structure and uORF influence non-targeted and targeted reporters differentially

A The NT reporter expression of both 3’UTR reporters is shown. NT reporter Renilla luciferase activity was normalized to firefly activity and set relative to expression of the SL control reporter. B The T reporter expression of both 3’UTR reporters is shown. T reporter Renilla luciferase activity was normalized to firefly luciferase activity and set relative to expression of the SL control reporter. C The cognate 5’UTR of each 3’UTR reporter used in Figure 1A was cloned in place of the 5’UTR in the vector pMT-DEST48. The repression of each reporter is shown. D Schematic describing the 5’UTR inserts used in panels E and F. E Insertion of an uORF in the 5’UTR of the GAPDH 3’UTR reporter increased repression. F The expression of the NT and T reporter for GAPDH was reduced upon insertion of the uORF. The expression of the T reporter was reduced more than the NT reporter. All data are depicted as mean ± SD.
Supplemental Figure 3.7: Statistical analysis of the interaction between TE and translational repression

Cumulative distributions of fold change of RPFs, A and B for all miR-155 predicted targets (http://targetscan.org/) in data from Guo et al., 2010. Fold change is calculated as the log$_2$ normalized RPF reads for miR-155 transfected divided by mock transfected. TE for each transcript in the absence of miR-155 (mock transfection) was calculated by normalized RPF divided by normalized RNAseq reads. All miR-155 targets are binned by TE, above or below the median (“High” or “Low”), A, or by TE quartiles, B. The tables below each plot describe the p-value for Kolmogorov-Smirnov tests of each population.
Supplemental Figure 3.8: TE influences the magnitude of repression by miR-155 independent of the seed pairing
Cumulative distributions of fold change of RPFs, for miR-155 predicted targets containing a conserved or poorly conserved 8-mer, 7-mer-a1, 7-mer-m8, or 6-mer seed pairing with miR-155 (http://targetscan.org/) in data from Guo et al., 2010. Fold change is calculated as the log₂ normalized RPF reads for miR-155 transfected divided by mock transfected. TE for each transcript in the absence of miR-155 (mock transfection) was calculated by normalized RPF divided by normalized RNAseq reads. All miR-155 targets are binned by TE, above or below the median (“High” or “Low”).
Supplemental Figure 3.9: TE influences the magnitude of repression by miR-1
Cumulative distributions of fold change of RPF or RNAseq for all miR-1 predicted targets (http://targetscan.org/) in data from Guo et al., 2010. Fold change is calculated as the log2 normalized RPF or RNAseq reads for miR-1 transfected divided by mock transfected. TE for each transcript in the absence of miR-1 (mock transfection) was calculated by normalized RPF divided by normalized RNAseq reads. All miR-1 targets are binned by TE, above or below the median (“High” or “Low”), A or by TE quartiles, B. C. miR-1 targets are binned by the number of conserved and poorly conserved binding sites for miR-1 as well as TE. ** p<0.01 by Kolmogorov–Smirnov test. D miR-1 targets are binned by 3’UTR length, above the median “Long”, or below the median “Short”. E miR-1 targets are binned by 3’UTR length into quartiles, “Long”, “Med.Long”, “Med.Short”, and “Short”. F All miR-1 targets are binned by transcript length, above the median “Long”, or below the median “Short”. G miR-1 targets are binned by transcript length into quartiles, “Long”, “Med.Long”, “Med.Short”, and “Short”. H miR-1 targets are binned by tAI, above the median “High”, or below the median “Low”. I miR-1 targets are binned by tAI into quartiles, “High”, “Med.High”, “Med.Low”, and “Low”.
Supplemental Figure 3.10: TE along with mRNA abundance can predict fold change of RPF

Plots showing correlation between measured FC and predicted based on the different equations derived from the data for miR155 repressed sample. For the prediction we have chosen not the best equation but the one that was among the best and among the simplest. These typically were not scoring worse than 0.05 in R-goodness-of-fit units compared to the complex versions. A Correlation between Fold Change RPF and Fold Change of mRNA from corresponding RNAseq experiments. B Prediction of Fold Change RPF using only mRNA levels and the corresponding equation derived by Eureqa. C Best prediction of Fold Change RPF using more variables than mRNA levels and the corresponding equation derived by Eureqa. Only length of 3’ UTR was selected as the variable improving the model. However, the selected equation does not make sense mathematically, as FC is a logarithm. D Prediction of FC using mRNA levels and TE and the corresponding equation derived by Eureqa.
Supplemental Figure 3.11: Correlation between TE and RPF fold change for miR-155 targets in LPS activated B-cells

Correspondence between RPF fold change and TE, A and D, or RNA fold change and TE, B and E, at 2 hr, A and D, or 4 hr, B and E, post activation of B-cells with LPS in data from Eichorn et al., 2014. Fold change is calculated as the log₂ normalized RPF or RNAseq reads for WT B-cells divided by miR-155 knockout B-cells. TE for each transcript in the absence of miR-155 (knockout) was calculated by normalized RPF divided by normalized RNAseq reads. Cumulative distributions of fold change of RPF, C and F, for all miR-155 predicted targets in B-cells at 2 hr post activation, C, or 4 hr post activation, F. All miR-155 targets are binned by TE quartiles.
Supplemental Figure 3.12: Comparison of distributions of TE multiplied by RPF Fold Change across groups and time points

Across time points TE*RPF FC gradually shifts toward left for targets and stays the same or shifts towards right for non-targeted genes. Statistics calculated using Kolmogorov-Smirnov test. A. 2hrs, p-value that TE*RPF_FC is smaller for targets than for control: 0.002756 B. 4hrs, p-value that TE*RPF_FC is smaller for targets than for control: 8.037e-05 C. 8 hrs, p-value that TE*RPF_FC is smaller for targets than for control: 0.003012 D. 48 hrs, p-value that TE*RPF_FC is smaller for targets than for control: < 2.2e-16.
B

TE low (238 genes)

FC > 0.25 (57 genes)

TE high (236 genes)

FC < -0.25 (126 genes)
Supplemental Figure 3.13: Gene ontology analysis for 4 groups: TE high, TE low and RPF FC below -0.25 and RPF FC > 0.25.
Eight most statistically significant terms are provided for each group. Timepoints: A 2hrs, B 8hrs. Highly repressed genes (FC values below -0.25 or TE high) have a similar functional profiles. Due to small sets used in case of fold-change-selected groups, uncorrected p-values are reported in all cases.
Supplemental Figure 3.14: 3’UTR length, transcript length and tAI do not influence fold change of RPF
Cumulative distributions of fold change of RPF for all miR-155 predicted targets (targetscan.org) in data from Guo et al., 2010. Fold change is calculated as the log2 normalized RPF for miR-155 transfected divided by mock transfected. A We binned all miR-155 targets by 3’UTR length, above the median “Long”, or below the median “Short”. B We binned all miR-155 targets by 3’UTR length into quartiles, “Long”, “Med.Long”, “Med.Short”, and “Short”. C We binned all miR-155 targets by transcript length, above the median “Long”, or below the median “Short”. D We binned all miR-155 targets by transcript length into quartiles, “Long”, “Med.Long”, “Med.Short”, and “Short”. E We binned all miR-155 targets by tAI, above the median “High”, or below the median “Low”. F We binned all miR-155 targets by tAI into quartiles, “High”, “Med.High”, “Med.Low”, and “Low”. 
Supplemental Figure 3.15: 5' UTR and 3' UTR structure do not correlate with fold change of RPF

Lack of correlation between Fold Change RPF and normalized MFEs of 5' UTR, A, and 3' UTR, B, of miR-155 targeted genes. Normalization of MFE was done according to Trotta method (Trotta, 2014) which removes all dependence of sequence length from the final MFE values.
Supplemental Figure 3.16: mRNA half-life does not correlate with Fold Change of RPF, RNA or TE
Lack of correlation between miR-155 induced Fold Change of RNA, A, RPF, B, and TE, C, and mRNA half-life in untransfected HeLa, Tani et al., 2012.
### Supplemental Table 3.1: Primers and oligonucleotides

<table>
<thead>
<tr>
<th>Primer/Oligo Name</th>
<th>Sequence (5'-&gt;3')</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Renilla luc. forward</strong></td>
<td>CACCATGGGCACTTCAAGTAATTATG</td>
</tr>
<tr>
<td><strong>HID/FLP reverse</strong></td>
<td>GTATCTTATCATGTCGTCGACGCG</td>
</tr>
<tr>
<td><strong>SV40 p(A)s mut. forward</strong></td>
<td>CCATTATAAGCTGCAATTAACACACAAAG</td>
</tr>
<tr>
<td><strong>SV40 p(A)s mut. reverse</strong></td>
<td>GCAAATGTGTGGTTAATGACGAT</td>
</tr>
<tr>
<td><strong>GAPDH overlap forward</strong></td>
<td>GTTCGAGCAAGATGTAAGCTAGCATGCAAAAG</td>
</tr>
<tr>
<td><strong>GAPDH reverse</strong></td>
<td>AGATGCGGAA</td>
</tr>
<tr>
<td><strong>α-tubulin overlap forward</strong></td>
<td>CACGTTCGAGCAGACATGATAAGATACATTCGAATTTG</td>
</tr>
<tr>
<td><strong>α-tubulin reverse</strong></td>
<td>CTTATTTCTGTTTCTGAATTG</td>
</tr>
<tr>
<td><strong>β-tubulin overlap forward</strong></td>
<td>CACGTTCGAGCAGACATGATAAGATACATTCGAATTTG</td>
</tr>
<tr>
<td><strong>β-tubulin reverse</strong></td>
<td>CTTATTTCTGTTTCTGAATTG</td>
</tr>
<tr>
<td><strong>Rpl32 reverse w/overlap</strong></td>
<td>CACGTTCGAGCAGACATGATAAGATACATTCGAATTTG</td>
</tr>
<tr>
<td><strong>Hsp70a overlap forward</strong></td>
<td>GTTCGAGCAAGATGTAAGCTAGCATGCAAAAG</td>
</tr>
<tr>
<td><strong>Hsp70a reverse</strong></td>
<td>AGATGCGGAA</td>
</tr>
<tr>
<td><strong>CrebA overlap forward</strong></td>
<td>CACGTTCGAGCAGACATGATAAGATACATTCGAATTTG</td>
</tr>
<tr>
<td><strong>CrebA reverse</strong></td>
<td>CTTATTTCTGTTTCTGAATTG</td>
</tr>
<tr>
<td><strong>Cad87 overlap forward</strong></td>
<td>CACGTTCGAGCAGACATGATAAGATACATTCGAATTTG</td>
</tr>
<tr>
<td><strong>Cad87 reverse</strong></td>
<td>CTTATTTCTGTTTCTGAATTG</td>
</tr>
<tr>
<td><strong>H3-SL oligo 1</strong></td>
<td>AAATAATCGGTTCCTTTTCAGGACCACAAAG</td>
</tr>
<tr>
<td><strong>H3-SL oligo 2</strong></td>
<td>AAATAATCGGTTCCTTTTCAGGACCACAAAG</td>
</tr>
<tr>
<td><strong>GAPDH –p(A)s reverse</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>α-tubulin –p(A)s reverse</strong></td>
<td>CTTATTTCTGTTTCTGAATTG</td>
</tr>
<tr>
<td><strong>β-tubulin –p(A)s reverse</strong></td>
<td>CTTATTTCTGTTTCTGAATTG</td>
</tr>
<tr>
<td><strong>Hsp70a –p(A)s reverse</strong></td>
<td>CTTATTTCTGTTTCTGAATTG</td>
</tr>
<tr>
<td><strong>CrebA –p(A)s reverse</strong></td>
<td>CTTATTTCTGTTTCTGAATTG</td>
</tr>
<tr>
<td><strong>Control 5'UTR insert oligo 1</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>Control 5'UTR insert oligo 2</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>9-SL 5'UTR insert oligo 1</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>9-SL 5'UTR insert oligo 2</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>12-SL 5'UTR insert oligo 1</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>12-SL 5'UTR insert oligo 2</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>15-SL 5'UTR insert oligo 1</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>15-SL 5'UTR insert oligo 2</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>uORF Control insert oligo 1</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>uORF Control insert oligo 2</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td><strong>uORF Control insert oligo 2</strong></td>
<td>CAACAACAAATAATATGTGACTTGC</td>
</tr>
<tr>
<td>HA uORF insert oligo 2</td>
<td>GTTACGCCTAATCCGGCAGTCGTACGGATACCGCCATTGTGC</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>3'RACE RT</td>
<td>GGCCTAGCTGTTACTGGGCAAGGGTCAGCTGACTAGTACTTTTTTT</td>
</tr>
<tr>
<td>3'RACE External Amp</td>
<td>GGCCTAGCTGTTACTGGGCAAGGGTCAGCTGACTAGTACTTTTTTT</td>
</tr>
<tr>
<td>3'RACE Amplification</td>
<td>CACACGGTGACTAGTAC</td>
</tr>
<tr>
<td>Renilla-tail Forward</td>
<td>GTGAAAAATATATCAAATCGTTCGTTGG</td>
</tr>
<tr>
<td>Renilla luc. qRT forward</td>
<td>GCCATGGGCAGTCAGG</td>
</tr>
<tr>
<td>Renilla luc. qRT reverse</td>
<td>GCAACCAATCTAGCCG</td>
</tr>
<tr>
<td>Renilla 0.602 luc. qRT F</td>
<td>CCTACGAGCGACCAAGCAAAG</td>
</tr>
<tr>
<td>Renilla 0.602 luc. qRT R</td>
<td>CGATGTCCTCCTCGATGTGC</td>
</tr>
<tr>
<td>Renilla 0.494 luc. qRT F</td>
<td>GTTTTTAAGCTTGCCAAAGAG</td>
</tr>
<tr>
<td>Renilla 0.494 luc. qRT R</td>
<td>CCACGACTCAATCACTGCAAC</td>
</tr>
<tr>
<td>Renilla 0.298 luc. qRT forward</td>
<td>TGCAGCATCATCTATCTATAGGAG</td>
</tr>
<tr>
<td>Renilla 0.298 luc. qRT reverse</td>
<td>TCCCCAGTATTACCTGACTTCCC</td>
</tr>
<tr>
<td>Firefly luc. qRT forward</td>
<td>CCAGGGATTTCAATGATG</td>
</tr>
<tr>
<td>Firefly luc. qRT reverse</td>
<td>AAATCTCAGCAGATGTTCT</td>
</tr>
<tr>
<td>GAPDH 5'UTR F</td>
<td>CCAATGGAATCACTGATTTG</td>
</tr>
<tr>
<td>GAPDH 5'UTR R</td>
<td>GTGCCCATGTTGTTGCTTCCTT</td>
</tr>
<tr>
<td>α-tubulin 5'UTR F</td>
<td>CCAATGGAATCACTGATTTG</td>
</tr>
<tr>
<td>α-tubulin 5'UTR R</td>
<td>GTGCCCATGTTGTTGCTTCCTT</td>
</tr>
<tr>
<td>β-tubulin 5'UTR F</td>
<td>CCAATGGAATCACTGATTTG</td>
</tr>
<tr>
<td>β-tubulin 5'UTR R</td>
<td>GTGCCCATGTTGTTGCTTCCTT</td>
</tr>
<tr>
<td>Cad87 5'UTR F</td>
<td>CCAATGGAATCACTGATTTG</td>
</tr>
<tr>
<td>Cad87 5'UTR R</td>
<td>GTGCCCATGTTGTTGCTTCCTT</td>
</tr>
<tr>
<td>Hsp70a 5'UTR F</td>
<td>CCAATGGAATCACTGATTTG</td>
</tr>
<tr>
<td>Hsp70a 5'UTR R</td>
<td>GTGCCCATGTTGTTGCTTCCTT</td>
</tr>
<tr>
<td>Rpl32 5'UTR F</td>
<td>CCAATGGAATCACTGATTTG</td>
</tr>
<tr>
<td>Rpl32 5'UTR R</td>
<td>GTGCCCATGTTGTTGCTTCCTT</td>
</tr>
<tr>
<td>Renilla overlap R</td>
<td>GTATCTTTATCATGTCGTTCGTTATTGTTTCGTTGAGAACAC</td>
</tr>
<tr>
<td>HID/FLP F</td>
<td>CACCCGCTTTGAGCAGACATGATAAGATAC</td>
</tr>
<tr>
<td>GAPDH R2</td>
<td>ATTACAGTAACAGGGCGATACTTTATTC</td>
</tr>
</tbody>
</table>
### Supplemental Table 3.2: Expression of miRNAs in S2 cells

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Predicted 3’UTR Target</th>
<th>Relative Expression RNAseq&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Relative Expression qPCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>bantam</td>
<td>N/A</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>miR-1</td>
<td>CrebA</td>
<td>0.00</td>
<td>ND</td>
</tr>
<tr>
<td>miR-7</td>
<td>Cad87</td>
<td>2.11</td>
<td>1.12</td>
</tr>
<tr>
<td>miR-8</td>
<td>Hsp70a</td>
<td>23.16</td>
<td>60.87</td>
</tr>
<tr>
<td>miR-10-3p</td>
<td>Hsp70a, Cad87</td>
<td>0.00</td>
<td>ND</td>
</tr>
<tr>
<td>miR-14</td>
<td>Hsp70a</td>
<td>335.42</td>
<td>752.97</td>
</tr>
<tr>
<td>miR-252</td>
<td>Cad87</td>
<td>75.32</td>
<td>9.76</td>
</tr>
<tr>
<td>miR-263a</td>
<td>Cad87</td>
<td>0.59</td>
<td>3.52</td>
</tr>
<tr>
<td>miR-274</td>
<td>GAPDH</td>
<td>0.08</td>
<td>0.13</td>
</tr>
<tr>
<td>miR-277</td>
<td>Cad87</td>
<td>18.68</td>
<td>20.05</td>
</tr>
<tr>
<td>miR-283</td>
<td>GAPDH</td>
<td>0.42</td>
<td>9.11</td>
</tr>
<tr>
<td>miR-304</td>
<td>Rpl32</td>
<td>0.08</td>
<td>1.30</td>
</tr>
<tr>
<td>miR-316</td>
<td>Cad87</td>
<td>0.00</td>
<td>ND</td>
</tr>
<tr>
<td>miR-956</td>
<td>Cad87</td>
<td>0.00</td>
<td>ND</td>
</tr>
<tr>
<td>miR-964</td>
<td>Cad87</td>
<td>0.00</td>
<td>ND</td>
</tr>
<tr>
<td>miR-965</td>
<td>α-tubulin</td>
<td>1.18</td>
<td>0.00</td>
</tr>
<tr>
<td>miR-981</td>
<td>Cad87</td>
<td>0.00</td>
<td>ND</td>
</tr>
<tr>
<td>miR-987</td>
<td>Cad87</td>
<td>0.00</td>
<td>ND</td>
</tr>
<tr>
<td>miR-999</td>
<td>α-tubulin, CrebA, Cad87</td>
<td>0.76</td>
<td>0.52</td>
</tr>
<tr>
<td>miR-1000</td>
<td>Cad87</td>
<td>0.00</td>
<td>ND</td>
</tr>
<tr>
<td>miR-1002</td>
<td>Cad87</td>
<td>0.00</td>
<td>ND</td>
</tr>
<tr>
<td>miR-1006</td>
<td>Cad87, Hsp70a</td>
<td>ND</td>
<td>0.07</td>
</tr>
<tr>
<td>miR-1014</td>
<td>CrebA, Hsp70a</td>
<td>ND</td>
<td>0.05</td>
</tr>
</tbody>
</table>

ND = not determined
### Table S3

**Primers for miRNA qRT-PCR**

<table>
<thead>
<tr>
<th>miRNA</th>
<th>RT Primer (5’-&gt; 3’)</th>
<th>qPCR Reverse (5’-&gt; 3’)</th>
<th>qPCR Forward (5’-&gt; 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bantam</td>
<td>ACTTAGGCGCAATGTTACAGGTTGC</td>
<td>TTTGCATTGCTGAGATCTTATTTGA</td>
<td>TGAACATTACGGTGAGCTTTAATA</td>
</tr>
<tr>
<td>mir-7</td>
<td>AGCGATCTGCTCCAGACAGCAAATC</td>
<td>TGACTCTGGAGAAGACTAGTATAT</td>
<td>TAGACATTCTGCTTCCGACAG</td>
</tr>
<tr>
<td>miR-274</td>
<td>GAAGCTCCACTTACAGTACCGTTAGT</td>
<td>GACAGAACCTTTTGAGACCACAC</td>
<td>AGGGAAGCTCCATTACAGTGC</td>
</tr>
<tr>
<td>miR-283</td>
<td>TACTGCAAAAACCTGAGTCCCGAGATACC</td>
<td>TGACTCTGATCCATCTGCAATTCT</td>
<td>AAGTACTCAAAACTTGGTAGTC</td>
</tr>
<tr>
<td>miR-304</td>
<td>CAGCAACAGAACTGGCCATCTCAGTCATC</td>
<td>TGCCACACTTAAATCTGAATTCTTCT</td>
<td>AGGACACACAGTCAAGCCATC</td>
</tr>
<tr>
<td>miR-305</td>
<td>TGCCCTCTGACAGCGCCACAGACAGGCTT</td>
<td>TGAGCTGATCTCTTGTTAATCTGCAAGA</td>
<td>CTTCCTCTTCCTGACAAGCGCCA</td>
</tr>
<tr>
<td>miR-965</td>
<td>CGATCCGTTCAGTCGTAAGGGAAAAG</td>
<td>AGTGAAGCAGTTAAGCGTATAGTTT</td>
<td>CCTGATCCGTTCCACTGCTGTA</td>
</tr>
<tr>
<td>miR-277</td>
<td>CCAGGGCAAGAATCTCTTGTGGTTACGAG</td>
<td>AGCCTGGAGTAAATGCAGACTATCT</td>
<td>GATCCGAGGCGCAATCTTTATT</td>
</tr>
<tr>
<td>miR-252</td>
<td>CTCCATTATTTGTCTTGAGTATGAGCGGA</td>
<td>CGCTTTCAAGCTAGATAGTACAGG</td>
<td>GACCTCCATTGATATGCTGACG</td>
</tr>
<tr>
<td>miR-263a</td>
<td>CGAAGTAAAGTTGATCGGCCGTGAATTCC</td>
<td>GGCCTGCAGGTAATGGCAGTAGGGAAG</td>
<td>CAGGCCAAGTTAAGTGGTGGGC</td>
</tr>
<tr>
<td>miR-1014</td>
<td>TGAGATGCCATAGGAGGGCCTGCAAATGAGA</td>
<td>AGCTGAACTAAAATTTGAATTTTCTTA</td>
<td>CGCTGAGATGCCATGGAGGTC</td>
</tr>
<tr>
<td>miR-1006</td>
<td>CTCCCTGTCCTTCCAGGAGGACTATGAAATAAG</td>
<td>AGGCTCGGATGAAATTGAATTTGCTTCT</td>
<td>GTGCTCGTATACCTGGGACG</td>
</tr>
<tr>
<td>miR-14</td>
<td>TCTTTATATCTAAGCTGATAGGTTTACCA</td>
<td>TCTTTATAGTGTTGCAATCTCCTC</td>
<td>GTATTCTGAAATTTGCTGACATCCA</td>
</tr>
<tr>
<td>miR-8</td>
<td>CTCCACCTTATATCTAAGCTGACATCTCATC</td>
<td>TCTTTACTTTTTTTTTTTTTTTTTTTT</td>
<td>TCTTCCACCTTATCTAAGCTG</td>
</tr>
</tbody>
</table>


Chapter 4: PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs

Preface

The following work was completed by me, Hemangi G. Chaudhari, Barak Cohen and Sergej Djuranovic. S.D. conceived and supervised the project and edited the manuscript. K.A.C. performed all experiments, data analysis and wrote the manuscript with contributions from the other authors. H.G.C. assisted in library design, model generation and edited the manuscript. B.A.C. helped to supervise the project, provided input throughout and edited the manuscript.

This chapter is currently under review at the journal Nature Communications.

We thank the Jason Weber and Len Maggi (Washington University in St. Louis) for assistance and equipment used in polysome profiling. We thank Marko Jovanovic and Pawel Szczesny for comments. Funding from NIH (GM007067 to KAC and GM112824 to SD). HGC and BAC supported by GM092910 and R01-HG008687
4.1 Abstract

RNA binding proteins (RBPs) and microRNAs (miRNAs) bind to sequences that are often located in the 3’ untranslated regions (UTRs) of their target mRNAs, where they regulate stability and translation efficiency. With the identification of several hundred RBPs, and several thousand miRNAs, there is an urgent need for new technologies to dissect the function of the cis-acting elements of RBPs and miRNAs. We describe post-transcriptional regulatory element sequencing (PTRE-seq), a massively-parallel method for assaying the target sequences of miRNAs and RBPs. We used PTRE-seq to dissect the sequence preferences and interactions between target sequences of the miRNA let-7 and three RBPs: SAMD4A or SAMD4B, Pumilio, and the AU-rich element binding proteins. We found that the binding sites for these effector molecules influenced different aspects of the RNA lifecycle, including RNA stability, translation efficiency, and translation initiation. In some cases, post-transcriptional control is modular, with different factors acting independently of each other, while in other cases different RNA-binding molecules show specific epistatic interactions. Deploying PTRE-seq across multiple cell lines demonstrates how the trans environment generates different effects from the same 3’UTR elements. The throughput, flexibility, and reproducibility of PTRE-seq make it a valuable new tool to study post-transcriptional regulation by 3’UTR elements.
4.2 Introduction

Cellular factors post-transcriptionally regulate mRNA by altering its modification, localization, stability, and translation\(^1\,^2\). These \textit{trans}-acting factors often bind to \textit{cis} elements within the mRNA. Two important classes of \textit{trans}-acting factors are RNA binding proteins (RBPs) and microRNAs (miRNAs).

miRNAs are short non-coding RNAs that mediate translational repression and destabilization of target mRNAs\(^3\,^{15}\). miRNAs recruit the Argonaute containing miRNA-induced silencing complex (miRISC) to specific mRNAs by base-pairing with complementary sequences within their \textit{3’UTR}\(^3,\,^6\). Mammalian cells typically express many miRNAs, with the human genome currently thought to encode 2580 miRNAs.\(^16\) Those miRNAs are predicted to target most human mRNAs\(^17\).

RBPs are a second prominent class of \textit{trans}-acting factors that affect mRNAs through processes including: splicing, adenylation/deadenylation, degradation, localization, and translation\(^2\). Recent studies have sought to identify the complete set of RBPs in mammalian cells, and based on these studies the human genome contains >1000 RBPs, most of which have unknown functions\(^18^{\text{-}}\,^{22}\). Over 800 RBPs have been identified in cultured HeLa cells alone\(^18\). One well characterized RBP is Pumilio, a member of the Puf family, which is conserved from yeast to humans\(^23\) and regulates translation and RNA-decay\(^24\,^{29}\). The RNA binding protein Smaug is also conserved from yeast (Vts1) to humans (SAMD4A and SAMD4B) and regulates both translation and RNA-decay\(^30\,^{32}\). Another well-known RBP family is the ELAVLs, homologues of the Drosophila \textit{embryonic lethal abnormal vision}, \textit{elav}\(^33,\,^34\). These proteins bind AU-rich elements within mRNAs and either stabilize or destabilize mRNAs, as well as enhance
or repress translation. While the function and mechanism of action of some RBPs has been partially elucidated, for the majority of RBPs their functions remain unknown.

Most evidence for the function of RBPs, such as Pumilio and Smaug, has come from low-throughput experiments that study their targets during embryogenesis or from reporter experiments. Given the large numbers of uncharacterized RBPs and miRNAs, we urgently need new approaches with higher throughput, which can be employed in diverse cell types and developmental stages.

Interactions between the miRISC and RBPs have been of great interest recently. With >2500 human miRNAs, that are predicted to target most mRNAs, and >1000 RBPs it is likely that many mRNAs are co-regulated by these factors. Many RBP or miRNA binding sites have been shown to occur near predicted miRNA binding sites. In many cases these binding sites are immediately adjacent or even overlap. Some RBPs cooperate with miRNAs in regulating the expression of specific genes. For example, Pumilio facilitates miRNA-mediated repression in both humans and Drosophila. HuR, a RBP that binds AU-rich elements, can also modulate miRNA-mediated repression. Understanding how mRNA trans-acting factors modulate the activity of one another is a major challenge. A tractable high-throughput approach would help unravel the interactions between different effectors of RNA regulation.

The widespread availability of high-throughput sequencing is powering the development of “omic” technologies to study miRNAs and RBPs. RNA-seq combined with ribosome profiling can reveal the effects of RBPs and miRNAs on target RNA expression and translation. While these methods provide the throughput required to study the effects of miRNAs and RBPs across the genome, they do not provide the flexibility to construct and assay large numbers of
reporters designed to dissect the effects of different combinations and affinities of RNA cis-regulatory elements.

In studies of transcriptional enhancers Massively Parallel Reporter Gene Assays (MPRAs) are useful complements to technologies that quantify the activity of endogenous genomic elements \(^{60-69}\). An analogous technology for assaying the activities of the cis-acting RNA sequences bound by RBPs and miRNAs would help unravel the network of interactions that underlies post-transcriptional regulation of mRNA. Such a system should provide the flexibility and throughput to dissect individual 3’ UTR elements, assay the effects of changes in the strength and number of cis-acting RNA elements, and detect interactions between different types of cis-acting sequences.

Recently several labs have employed plasmid or mRNA libraries to study endogenous 3’UTR elements \(^{70-74}\). These approaches generally rely on synthesizing or amplifying portions of 3’UTRs and fusing them to a reporter. While these techniques have identified 3’UTR motifs that have effects on RNA stability and protein amounts, none have been combined with polysome profiling to separate effects on RNA stability, translation efficiency, and translational initiation. In addition, naturally occurring 3’ UTRs contain many different types of elements, making it difficult to deconvolve the effects of individual sites. A synthetic approach, in which large numbers of reporters with specific combinations of elements are designed and assayed, would provide the power necessary to isolate the effects of individual binding sites, as well as the interactions between sites. Because high-throughput methods for studying synthetic elements have proven to have great utility in dissecting interactions among transcription factors, we have extended this approach to post-transcriptional regulation \(^{75-78}\).
Here we report post-transcriptional regulatory element sequencing (PTRE-seq), an approach that uses a massively parallel reporter library to study the effects of synthetic 3’UTR elements on RNA stability, translation efficiency, and translation initiation. We used PTRE-seq to study the effects of known binding sites for RBPs and miRNAs, both individually and in combination. With this approach, we determined that the binding sites for these effector molecules influenced different aspects of the RNA lifecycle, including RNA stability, translation efficiency, and translation initiation. We observed trans-acting factors acting independently or in some cases epistatically. Finally, deploying PTRE-seq across multiple cell lines revealed the influence of the trans environment on post-transcriptional regulation by specific trans-acting factors. Together these results demonstrate the throughput, flexibility and reproducibility of PTRE-seq.

4.3 Results

4.3.1 Design and application of PTRE-seq

We developed PTRE-seq to quantify the individual and combined effects of RBP and miRNA binding sites in 3’UTRs. We created 642 unique synthetic 3’UTRs composed of combinations of a let-7 binding site, the Pumilio recognition element (PRE), the Smaug recognition element (SRE), AU-rich elements (ARE), and a control sequence (‘blank’). Bioinformatics analysis indicates that nearly 300 human transcripts contain a PRE, miRNA-binding site and an ARE (Figure S1). Over 1900 transcripts contain a PRE and an ARE, 698 contain an ARE and a miRNA-binding site and 653 contain a PRE and a miRNA-binding site. Between 13-15% of each of these pairs of regulatory elements occur within 150 nt of each other, with many overlapping or immediately adjacent (Figure S1). We arranged the regulatory elements in four positions within the 3’UTR (Figure 1a), resulting in 200bp long regulatory
element section. The library included all possible combinations of the five elements within the four positions to generate the 625 unique synthetic 3’UTRs. The remaining seventeen synthetic 3’UTRs in our library contained variants of let-7 binding sites. Every unique synthetic 3’UTR is present ten times in the library, each time associated with a different co-transcribed barcode. These provide replicate measurements when the barcodes are used to quantify the relative abundance of reporter mRNAs in total or ribosome associated fractions from transfected cells. Barcoded synthetic 3’ UTR sequences were cloned downstream of a CMV promoter driven reporter gene to create a plasmid library.

We transfected HeLa cells with the library and harvested the cells after forty hours. We isolated total RNA from a portion of the cells, and the remaining cells were lysed for polysome profiling to assay translational regulation. We collected mRNAs associated with the polysome fractions (translating ribosomes) and the 40S ribosome fractions (initiating ribosomes) (Figure 1b and Figure S2).
Figure 4.1: Design and application of PTRE-seq

a Schematic of the PTRE-seq library. Each cis-regulatory element (RE) within the library is inserted into an episomal reporter as shown. CMV/TO, cytomegalovirus promoter with the 5’UTR from the vector pCDNA5/FRT/TO. EGFP, enhanced Green Fluorescent Protein. S, spacer sequence. BGH p(A)s, 3’UTR and polyadenylation signal from Bovine Growth Hormone gene. Each unique synthetic 3’UTR, made up of binding sites for the REs shown, is represented by ten barcodes.

b Representative polysome profiling trace. mRNA was isolated from 40S, and polysome fractions.

c Fold change of mRNA levels, translation efficiency, and 40S association for all reporters within the library. The reporters are arranged along the x-axis in decreasing order based on Fold Change.

Messenger RNAs associated with the polysomal fractions are considered efficiently translated $^{79}$. Since regulation of gene expression by various cis and trans-acting factors during mRNA translation often targets the translation initiation step, we separately analyzed the 40S fraction $^{80}$. The 40S fraction contains mRNAs that are bound only by the small subunit of the ribosome during the translation initiation steps. We generated cDNA from the total RNA, polysome, and 40S associated mRNA and sequenced the barcodes to determine the relative abundance of every reporter in the library, in each fraction. Counts for every barcode in cDNA were normalized by counts determined by sequencing the input plasmid library. The Pearson correlation between
replicate experiments ranged between 0.975 and 0.983 for total RNA, 0.703 and 0.787 for polysome associated RNA, and was 0.926 for 40S associated RNA (Figure S2), which allowed us to make quantitative comparisons between different synthetic 3’UTRs. To compute translation efficiency (TE), a measure of the reduction in translation beyond what is expected due to a reduction in mRNA levels, we normalized the barcode counts for each 3’UTR in the polysome fraction to its counts in total RNA. The same was done for the 40S associated RNAs to compute 40S association, which represents a proxy for the engagement of the translation initiation complex with mRNAs. In all cases, we determined the relative effect by normalizing to the control reporter, which contains four ‘blank’ sequences in the synthetic 3’UTR. For most reporters, we observed both reduced RNA expression and reduced TE, which was concomitant with an increase in 40S association (Figure 1c). Correlations can be seen between each of these metrics (Figure S2e-h). Summary statistics for PTRE-seq measurements of RNA expression and TE are shown in Figure S3. We validated our PTRE-seq findings using quantitative PCR (qPCR) and fluorescence measurements of GFP for several individual reporters from the library (Figure S4). The data we have obtained using PTRE-seq reveal the ability of this method to capture evidence for post-transcriptional regulation at different steps.

4.3.2 Linear regression and thermodynamic modeling of PTRE-seq results

The cis-elements in our library had strong effects in the data which we captured by fitting linear regression models for both RNA expression and TE to our data. For both the RNA expression model and the TE model, parameters included the identity of the element at each of the four positions and all pairwise interactions between elements at each position. The regression models captured the relationship between 3’UTR composition and relative RNA expression (five-fold cross validation, Pearson correlation 0.87-0.93) (Figure S5) and the relationship
between 3’UTR composition and TE (five-fold cross validation, Pearson correlation 0.89-0.92) (Figure S6). The model predicted well the effects of individual elements and combinations of elements on RNA expression and TE (Figure S7). Interestingly, the models accurately predicted the RNA expression and TE of reporters containing three or four different binding sites using only the individual effect of each binding sites and pairwise interactions. Models fit with higher order interaction terms failed during cross-validation. This result, combined with the observation that models with individual effects and pairwise interactions perform well, suggests that higher-order interactions have, at most, only minimal effects on RNA expression and TE.

To gain mechanistic insights, we also fit a statistical thermodynamic model to our RNA expression and TE data. Due to the position dependent nature of ARE elements (described below) we excluded synthetic 3’UTRs containing ARE elements from this analysis. This model provides a formal biophysical framework to capture saturation effects and cooperative interactions between cis-acting elements. Each 3’UTR is described as a collection of states, in which each state represents a particular configuration of bound and unbound elements on a 3’UTR. The model uses parameters that describe the free energies of interaction between RBP/miRNA-RBP/miRNA and RBP/miRNA–mRNA to compute the probability, or weight of each state. These interactions can be neighboring or non-neighboring, however, our implementation of the model does not explicitly model position of the RBPs. In each state bound factors either facilitate or inhibit the recruitment of mRNA decay machinery (or the ribosome for TE), and the weights of the different states are used to compute the probability that the mRNA decay machinery (or the ribosome for TE) is present at an mRNA. In the model, this probability is proportional to the output RNA expression or TE. Due to the position dependent nature of
ARE elements (described below) we excluded synthetic 3’UTRs containing ARE elements from this analysis.

A thermodynamic model with four independent parameters, one each for the interaction of the decay machinery with either let-7, PRE, SRE, or the “blank” site, predicted observed TE ($R = 0.92$) and RNA expression well ($R = 0.94$). The good performance of these models suggests that let-7, PRE, and SRE function mostly independently on UTRs. In most cases adding interaction terms did not improve the fit of these models to the data. This observation suggests that some of the self-interaction terms in the linear regression models (described below) are likely due to saturation of binding on UTRs with high copy numbers of cis-acting sites. The thermodynamic model naturally accounts for saturation without the need for interaction terms and describes the situation when saturation causes additional sites to have little or no effect. In two cases, the thermodynamic model for TE did improve with the addition of interaction terms, one for interaction between adjacent let-7 sites and another for interaction between adjacent PRE and let-7 sites ($R = 0.93$, Figure S8), which suggests epistatic interactions between these elements that cannot be accounted for by binding site saturation. The thermodynamic model for RNA expression also improved with the addition of five interaction terms ($R = 0.94$, Figure S9).

We sought to identify the trends in our data that underlie the strong performance of these models.

### 4.3.3 PTRE-seq reveals differences in the mechanism of post-transcriptional regulation by miRNAs and Pumilio

For each RNA element in our library there are a series of constructs that contain only that element and the control ‘blank’ sequence. This allowed us to study the individual and copy-number-dependent effect of each RNA element. For the let-7 binding site we observed a reduction in both relative RNA expression and TE (Figure 2a, c and Figure S10a, c). This suggests that not only is the abundance of the RNA reduced by the addition of let-7 sites, but
also that the remaining RNAs are translated poorly relative to the control message. Both effects were dependent on the number of let-7 binding sites in the synthetic 3’UTR and the effects appear to saturate with additional sites (Figure 2a). While our linear regression model captures well the effects of individual let-7 sites, it is easily influenced by saturation effects and thus cannot distinguish between saturation effects and true epistatic interactions (Figure 2i). To counter this we employed our thermodynamic model. The thermodynamic model requires an interaction term between let-7 binding sites that stabilizes RNA for a good fit (Figure S9). Since the thermodynamic model is robust to saturation effects, this interaction term suggests epistatic antagonism between let-7 binding sites.
Figure 4.2: PTRE-seq reveals differences in the mechanism of repression by miRNAs and Pumilio
Fold change of RNA, a, TE, c, and 40S association, e, of let-7 binding site containing reporters within the PTRE-seq library. Fold change of RNA, b, TE, d, and 40S association, f, of PRE containing reporters within the PTRE-seq library. For a-d, * P<0.05, ** P<0.01, *** P<0.001, t-test with Bonferroni correction. For panels a-f the results for all constructs containing one, two, three or four sites is shown. The data for each site in positions one-four are shown in Supplementary Figure 9. Panels g and h show
composite boxplots with fold change of RNA, TE, and translation initiation efficiency (TIE) for let-7 and PRE respectively. TIE was calculated by normalizing polysome associated RNA/40S associated RNA. The regression coefficients for linear models with parameters corresponding to let-7 alone or in combination with other let-7 sites at positions 1-4, or PREs alone or in combination with PREs at positions 1-4. In i and j, the left panels show the coefficients for RNA while the right panels show the coefficients for TE. * \( P<0.05 \), ** \( P<0.01 \), *** \( P<0.001 \), t-test. Boxplot whiskers indicate the furthest datum that is 1.5*Q1 (upper) or 1.5*Q3 (lower). For clarity, outliers have been removed from boxplots but were used for statistical analysis.

4.3.4 PTRE-seq reveals the effect of miRNA-target base-pairing on repression

The efficiency of miRNA-mediated repression depends on the number and quality of binding sites in its target. Nucleotides 2-7 of the miRNA constitute the “seed” sequence, and weak seed pairing reduces the effectiveness of miRNA-mediated repression. In addition to the 625 combinations of regulatory elements described above we included constructs in the library to study the effect of base-pairing between the miRNA and its target on repression. This included a series of constructs with one, two, or four binding sites for let-7 with either 6-mer, 7-mer-a1, 7-mer-m8 or 8-mer base pairing in the seed region, as well as binding sites for let-7 that have perfect base-pairing with the target. We observed a clear copy-number-dependent and seed-pairing dependent effect on RNA expression and TE for these reporters (Figure 3a, Figure S12). The repression at the level of RNA and TE was greatest for target sites with 8-mer or 7-mer-m8 pairing. A single copy of the perfect complement let-7 binding site was more effective at reducing RNA expression than four copies of the binding site with a mispairing bulge (Figure 3a). In addition to studying the effect of seed pairing alone we also studied the effect of endogenous let-7 binding sites. For this we made constructs containing four copies of a let-7 binding site from the 3’UTR for HMGA2, SMARCAD1, DNA2, C14orf28 and FIGNL2. While, the synthetic binding site is predicted to have the most favorable binding (Figure 3b), the sequences from two of the natural 3’UTRs (HMGA2 and FIGNL2) reduced RNA expression to a greater extent (Figure 3c). We suspect that secondary structure around the let-7 binding sites in
these reporters is contributing to let-7 binding. This can be seen by making a simple linear regression model for fold change of RNA expression with base-pairing minimal free energy (MFE) and 3’UTR secondary structure MFE as parameters. A model that includes each parameter and an interaction term gave a better fit (R=0.81) than base-pairing MFE (R=0.56) or secondary structure MFE (R=0.12) alone. Because this model was made with only a few data points it is only suggestive. This secondary structure of the 3’UTR could explain the observation that some binding sites, even with better thermodynamics, were not as well repressed.

Figure 4.3: PTRE-seq reveals the effect of the let-7 binding site on repression

a Comparison of the fold change of reporters containing synthetic let-7 binding sites with altered seed binding. Also shown are reporters containing let-7 binding sites that have perfect complement (PC)
binding to let-7. Each seed binding variant is present in either one, two or four copies. The inset describes the seed binding region of each seed-binding variant site. Table describing the natural and synthetic let-7 binding sites used in this study. MFE, minimal free energy. mirSVR, mirSVR score. Fold change of RNA and TE for reporters containing four copies each of natural or synthetic let-7 binding sites. Boxplot whiskers indicate the furthest datum that is 1.5*Q1 (upper) or 1.5*Q3 (lower). For clarity, outliers have been removed from boxplots but were used for statistical analysis.

4.3.5 Pumilio does not enhance miRISC function

Enhancement of miRNA-mediated repression by the RBP Pumilio has been observed for a handful of targeted mRNAs. In our library, the combination of a let-7 binding site and a PRE resulted in a reduction of RNA expression that was slightly less than the product of their individual effects (Figure 4a). This was true for every combination we tested. The coefficients from our linear regression model for RNA levels are positive for all combinations of PRE and let-7, while the coefficients for each alone is negative (Figure 4b). For TE we observed a modest enhancement of repression for some combinations and no effect for others (Figure 4c), this was captured by our linear regression model which showed a mix of positive and negative coefficients for the combinations of PRE and let-7 (Figure 4d). The pairwise arrangement of let-7 binding sites and PREs had no effect on repression (Figure S12). These data suggest a slight antagonism between the two elements in regard to their effects on RNA stability, and is reminiscent of the saturation we observed with additional let-7 or Pumilio binding sites. The thermodynamic model for TE includes a statistically non-significant anti-cooperative interaction between let-7 and PRE sites, while the model for RNA decay includes anti-cooperative interaction terms for a subset of let-7 and PRE binding site combinations (Figure S8 and S9). Thus, the miRISC and Pumilio function independently in most UTRs and reduced repression seen with combinations of sites is mostly because of saturation effects. Since miRNAs and Pumilio are thought to promote mRNA decay using the same pathway it isn’t surprising that when both are bound to the same message there is no enhanced degradation.
Figure 4.4: Pumilio and miRNAs function independently

The effect of a let-7, PRE or a combination of the two elements on relative expression, and c. relative TE. The median relative expression or TE is plotted across all barcodes and replicates. Red dot, the product of each individual effect, the expected result assuming independence. The regression coefficients from the linear regression model for RNA expression, b, and TE, d, for the parameters corresponding to let-7 or PREs alone or interactions between positions containing let-7 or PREs. For a and c, L = let-7 binding site, p = PRE, * P<0.05, ** P<0.01, *** P<0.001, t-test.

4.3.6 Function of AU-rich elements is dependent on their position in the 3’UTR

Several RBPs can bind AU-rich elements. ARE binding proteins, such as HuR (ELAVL1), can either stabilize or destabilize target mRNAs, and can enhance or repress translation. Other RBPs, such as tristetraprolin (TTP) and AUF1 (hnRNPD), are also ARE-
binding proteins (ARE-BP)\textsuperscript{34, 92, 93}. In our library AREs either enhanced or repressed RNA expression and TE depending on their position in the 3’UTR (Figure 5a). In the first or fourth position in our synthetic 3’UTR, ARE reduces both RNA expression and TE, while an ARE in the second position increases RNA expression and TE, and an ARE in the third position has no effect on either metric. The combination of multiple AREs altered this position dependent effect. Generally, any combination with an ARE in position four had reduced RNA and TE while all other combinations had increased RNA and TE. We observed similar effects on 40S association where any combination of ARE that reduced TE resulted in increased 40S association and \textit{vice versa} for those that increased TE (Figure 5b). These observations were captured by our linear regression model for RNA expression, which showed an ARE at position four to have a negative coefficient and an ARE at position two to have a positive coefficient (Figure 5c). Since the linear regression model cannot distinguish between saturation effects and epistatic interactions it is difficult to assign a cause. However, the results clearly show instances were AREs in specific arrangements lead to increased or decreased RNA expression and/or TE.

### 4.3.7 AU-rich elements modulate activity of miRNAs and Pumilio

The AU-rich element binding protein HuR has been shown to both activate and inhibit miRNA-mediated repression. During recovery from stress HuR relieves miRNA-mediated repression of the catalase mRNA (\textit{Cat1})\textsuperscript{47}. In contrast, HuR binding to the \textit{c-Myc} 3’UTR activates miRNA-mediated repression\textsuperscript{48}. We observed AREs in our library either enhancing or suppressing miRNA-mediated repression in a position dependent manner (Figure 5d and Figure S13). For example, a let-7 binding site at position three reduces RNA expression while an ARE at position two increases RNA expression, but the combination of ARE and let-7 reduces RNA expression more than the let-7 binding site alone. Our linear regression likely captured some of
these effects but because it cannot distinguish between saturation effects and epistatic interactions we cannot assign a cause (Figure 5e). However, it is clear that in some cases specific combinations of let-7 binding sites and AREs resulted in obvious changes to RNA expression or TE, for instance changing from a message that is stable to one that is unstable, see above example. We observed similar position dependent modulation of Pumilio activity by AREs (Figure 5f and g). Together these data show that AU-rich element binding proteins can modulate the repression by the miRISC and Pumilio in a position dependent manner, even though, as shown above, Pumilio and let-7 utilize different mechanisms to repress RNA expression and TE.

4.3.8 Post-transcriptional regulation varies across cell types

In addition to HeLa cells, we also transfected our PTRE-seq library into three other cell types: human embryonic kidney (HEK293), human neonatal dermal fibroblast (HDF), and a mouse neuroblastoma (N2A). We observed wide variation in the effect of each regulatory element tested across the four cell lines. It is possible that some of these changes could be caused by differences in transfection efficiency or transcription across the cell lines tested. The let-7 binding site caused robust reduction of RNA expression in HeLa cells, but this effect was smaller in magnitude in HEK293, HDF and N2A (Figure 6a, d and Figure S14). Neuroblastoma cells are thought to have very little expression of let-794. In contrast, we observed modest variations in the magnitude of repression by PREs across the four cell lines (Figure 6b). PREs were most effective in HeLa and least effective in N2A or HEK293 cells. Interestingly, we observed only a very modest reduction in RNA expression and no effect on TE (Figure S14) for reporters containing SREs across all cell-lines. Only when we overexpressed the Drosophila homologue of SAMD4A and SAMD4B, Smaug (mCh-Smg) did we see a substantial reduction in RNA expression (Figure 6c).
Figure 4.5: AU-rich elements modulate repression by Pumilio and miRNAs  

a The position of an ARE within the synthetic 3’UTR determines the relative TE or RNA expression.  
b The relative 40s association of ARE containing reporters.  
c Heatmap of the regression coefficients for the parameters corresponding to AREs alone. Left panel shows coefficients for RNA expression and the right panel shows coefficients for TE.  
d AREs modulate repression by miRNAs in a position dependent manner. The green box highlights an example of stimulation of miRNA-mediated RNA destabilization by an ARE.  
e The regression coefficients for the parameters corresponding to let-7 or AREs alone or interactions between positions containing let-7 or ARE.  
f AREs modulate repression by PREs in a position dependent manner.  
g The regression coefficients for the parameters corresponding to PREs or AREs alone or interactions between positions containing PRE or ARE.  

For AREs, the cell type not only altered the magnitude of the effect but could also abrogate the effect entirely (Figure 6e). For example, in HeLa the AREs could both reduce or increase RNA expression, while in HEK293 we only observed increased RNA expression by AREs. Conversely, in N2A we observed robust reductions in RNA expression by AREs but very modest increases in RNA expression. As AREs are known to be bound by multiple RBPs this finding suggests the presence of a different profile of active ARE-binding RBPs in each cell type.
Figure 4.6: The regulatory capacity of miRNAs and AU-rich elements vary across cell types
The relative expression of reporters containing let-7 binding sites, a, PREs, b, SREs, c, natural binding sites for let-7, d, or AREs, e. In panel c HeLa-mCh-Smg refers to HeLa cells that were cotransfected with the PTRE-seq library and a plasmid for expression of mCherry-Smaug. HDF, neonatal human dermal fibroblasts. HEK, human embryonic kidney. N2A, mouse neuro2A. For e, * = Blank and A =ARE. Boxplot whiskers indicate the furthest datum that is 1.5*Q1 (upper) or 1.5*Q3 (lower). For clarity, outliers have been removed from boxplots but were used for statistical analysis.
4.4 Discussion

To better understand the post-transcriptional regulation of mRNAs, we must determine how regulatory factors function both independently and in combination with each other. Towards this end we developed PTRE-seq, a powerful new high-throughput tool for interrogating the additive and combined effects of binding sites for RBPs and miRNAs on RNA stability and translation. As PTRE-seq is extended to additional RBPs and miRNAs, we will better understand the network of molecular interactions that comprise post-transcriptional regulatory systems.

Using PTRE-seq we observed decreased RNA levels and decreased association with polysomes mediated by the let-7 miRNA. By fractionating the cell lysates before analysis we determined that the reduction in polysome associated RNA was more than could be accounted for by the decrease in RNA levels alone, indicating that let-7 reduces RNA levels and reduces the efficiency with which the remaining RNA is translated. This decrease in translational efficiency also correlated with an increase in 40S association of mRNAs targeted by the miRNA let-7. These results are consistent with a proposed model in which the miRISC inhibits translation initiation at the scanning step by induced dissociation of the helicase subunit eIF4A of the eIF4F complex. The reduced rate of scanning increases the time that the 40S ribosome is bound to the message prior to identification of the start codon and recruitment of the 60S ribosome. This delay in subunit joining would increase the time mRNAs spend bound by the 40S ribosome while reducing translation efficiency. The ability of PTRE-seq to separate effects on RNA levels from effects on different steps of translation is an important advantage of this method.
Beyond identifying the mechanism of miRNA-mediated repression another major challenge in the field remains defining binding sites with gene regulatory activity of the thousands of miRNAs in the cell. Using PTRE-seq we were able to study the efficacy of target sites for let-7. Our results are consistent with previous studies: miRNA efficacy depends on thermodynamics of binding and 3’UTR structure. We observed stronger repression for messages containing more base-pairing within the seed sequence: 8-mer > 7-mer > 6-mer. This finding was consistent with previous studies of endogenous miRNA targets. Furthermore, reporters containing four copies each of endogenous let-7 binding sites showed variable repression that was not always dependent on thermodynamics of base-pairing. A simple linear regression model revealed that the secondary structure around the let-7 binding sites contributed to the magnitude of repression. This finding is consistent with a model for miRNA target prediction which incorporates the thermodynamics of miRNA binding and secondary structure near the binding site. This type of analysis could be used for other miRNAs to empirically define their binding sites with largest impact on gene regulation.

In contrast to let-7, Pumilio decreased RNA levels with only very modest effects on polysome association and no effect on 40S subunit binding. Our results are consistent with the findings that Pumilio and miRNAs inhibit translation at different steps. It is also possible that the differences we observed between let-7 binding sites and PREs could reflect differences in the kinetics of repression by the miRISC or Pumilio. Besides their individual roles in regulation of gene expression, miRNAs and Pumilio have been shown to function together. In some cases, Pumilio can activate miRNA-mediated repression of specific mRNAs. However, in our experiments the effects of let-7 and Pumilio were largely independent. Pumilio may only activate particular miRNA targets by opening certain secondary structures and
enabling miRNA binding. To test this model PTRE-seq could be performed on synthetic messages carrying miRNA binding sites and PREs in the context of varying secondary structures.

The effects of let-7 or Pumilio sites showed almost no dependency on their position in the 3’UTR. In contrast, we observed a strong positional effect for AREs. AREs are bound by several RBPs including the ELAVLs, Auf1 and TTP, and different ARE binding proteins can stabilize or destabilize mRNA targets, as well as repress or enhance translation. The dependency of ARE on position could be explained if different ARE binding proteins are binding at different positions in the 3’UTR. Although the sequence of the ARE is the same at each position, the flanking sequence context is different and RNA secondary structure may vary based on the position of the ARE (Figure S15). This altered structure might affect which ARE-BP bind to the sequence. The varied effects of AREs across cell-lines are consistent with this hypothesis. While AREs both increased and decreased RNA expression in most cell lines tested, in HEK293 we only observed increased RNA expression. As the expression of ARE-BPs is known to vary across cell and tissue types, this finding suggests that the ARE-BPs with different effects are binding to the same reporters in different cell lines. In any given cell line, the cumulative effect of multiple ARE-BPs determine the overall activity of AREs.

Our experiments revealed strong epistatic interactions between AREs and sites for let-7 and Pumilio. The AREs either enhanced or suppressed miRNA- and Pumilio-mediated repression, depending on their position in the 3'UTR. These effects were not dependent on the proximity in linear sequence space between the two binding sites. These results demonstrate that AREs can modulate repression by both miRISC and RBPs such as Pumilio. Alternatively, the presence of the PRE or miRNA binding site may modulate the effect of the ARE by changing the
secondary structure of the mRNA. These observations warrant future studies into which ARE-BP
are responsible for these epistatic interactions, and whether the effects are mediated through
mRNA secondary structure or potentially through interactions between trans-acting factors. An
intriguing possibility is that the ARE may be bound by the cytoplasmic polyadenylation element
binding protein, CPEB. The consensus binding site for CPEB is UUUUUUAU \(^{100}\), but CPEB also
binds the sequence UUUUUAU \(^{101}\), which appears once in the ARE used in our library. CPEB has
been previously shown to work with Pumilio in regulating mRNA translation \(^{102,103}\).

While we observed robust effects on the RNA expression and translation of reporters
containing let-7 binding sites, PREs and AREs, in our experiments, the SRE caused only modest
changes in RNA expression and no change in TE or 40S association. When we overexpressed
*Drosophila* Smaug in HeLa cells, we observed a reduction in the RNA levels of SRE containing
reporters. This suggests that at least in the cell lines we tested the mammalian Smaug
homologues, *SAMD4A* and *SAMD4B*, are expressed at low levels or are not efficacious.

Our results provide further evidence for the mechanisms of post-transcriptional regulation
by miRNAs, Pumilio and AREs. PTRE-seq will serve as a valuable tool studying the effects of
multiple cis-acting elements, both individually and in combination, and for unraveling their
effects on different aspects of RNA stability and translational control.

**4.5 Materials and Methods**

**4.5.1 Construction of Library**

To create the PTRE-seq library we first generated the plasmid pCDNA5/FRT/TO-EGFP-RE. EGFP was PCR amplified with EGFP-F and RE-R primers, Supplemental Table 1. This
appended a single *NheI*, *EcoRV* and *KpnI* sites downstream of EGFP. The PCR product was
ligated into pCDNA5/FRT/TO that had been previously cut with *Pmel*. 

144
A pool of 6500 unique 200-mer oligonucleotides was ordered from Agilent Technologies™. Oligonucleotides were designed to contain all combinations of either a let-7 binding site, PRE, SRE, AU-rich element or a ’blank’ control sequence. The sequence for each of these elements is described in Supplemental Table 2. Each of these unique combinations was synthesized with ten different 9 bp barcodes. This provided a total of 6250 oligonucleotides. The remaining oligonucleotides consisted of 40 additional copies of the control sequence (4 place holders, “blanks”), 50 copies of a low expression control (4x let-7 perfect complement) and a series of constructs containing natural or synthetic let-7 sites. In total, the library consisted of 642 unique ‘synthetic 3’UTRs’ each with 10 unique barcodes, except for the controls described above. Each oligo has a 5’ and 3’ priming region which are identical across all oligonucleotides. The oligonucleotides also contained a restriction enzyme sites for subsequent cloning. A generic oligonucleotide appears as follows: 5’ - GTAGCATCTGTCCGCTAGC-132nt regulatory element-ATGCATcGATATCaCTCGAGxxxxxxxxxGGTACCAGCTACTACTACG – 3’. The restriction enzymes are underlined and are from 5’ to 3’: Nhel, Nsil, EcoRV, XhoI and KpnI.

The library was PCR amplified for four cycles using Phusion High-Fidelity polymerase (NEB) and primers Lib_F and Lib_R. We cloned the amplicon into pCDNA5/FRT/TO-EGFP-RE using Nhel and KpnI. We prepared plasmid DNA from ~40,000 colonies to generate library RE_Array*. We then cloned a “spacer” sequence in between the regulatory elements and the barcode. This “spacer” is the reverse of a sequence within the BGH 3’UTR and is used for amplification of the barcodes from cDNA or plasmid. The “spacer” was ordered as a pair of oligonucleotides that were annealed to form a dsDNA oligonucleotide with overhangs compatible with DNA cleaved by Nsil and XhoI. The “spacer” was cloned into the RE_Array*
using *Nsi*I and *Xho*I. We collected plasmid DNA from ~250,000 colonies to generate the library RE_Array_1.

To clone individual reporters from the library we sequenced 94 colonies from the RE_Array* library. We chose from those clones seven reporters of interest. For the control reporter and three reporters targeted by let-7 (*7**, 7777, and 7pc-x2), we ordered oligonucleotides that were ligated into the vector pCDNA5/FRT/TO-EGFP-RE as described above. The “spacer” was ligated into the plasmids containing the reporters as described above.

### 4.5.2 Cell Culture and Transfection

HeLa (CCL-2.2, ATCC), HDFn (C0045C, Thermo Fisher), N2A (CCL-131, ATCC) and T-REx™-293 cells (R71007, Thermo Fisher) were grown in DMEM (Gibco) supplemented with 10% heat-inactivated FBS (Gibco), 1x Penicillin streptomycin and glutamine (Gibco) and 1x MEM Non-Essential Amino Acids (Gibco). Transfection was carried using the Neon Transfection System (Invitrogen) per manufacturer protocol. For each transfection, 2.5 x 10⁶ cells were electroporated with 8 µg of RE_Array_1. For transfection of mCh-Smg, we electroporated 8 µg of pCDNA-D40-mCh-Smg along with 8 µg of RE_Array_1 into HeLa cells as described above. The mCh-Smg plasmid was made by PCR amplifying the Smaug coding sequence (CDS) from *Drosophila* S2 cell cDNA using the primers described in Table S1. The Smaug CDS was fused to mCherry through overlap PCR using primers described in Table S1. This PCR product was cloned into pENTR-D-TOPO (Invitrogen) and subsequently recombined into pcDNA-D40 (Invitrogen) using LR Clonase (Invitrogen).

For transfection of the individual reporters we used Effectene (Promega). The cells were transfected in a 12-well plate with 500 ng each of the EGFP reporter and pCDNA-mCherry per the manufacturers protocol. The cells were split 24 hours later into two separate 12-well
plates and a 96-well plate. Forty hours after transfection the fluorescence was measured using a Synergy H4 plate reader (BioTek), at the same time RNA and protein was isolated from the 12-well plates.

4.5.3 RNA Isolation and Polysome Profiling

Total RNA was isolated using Qiagen RNA mini-prep per manufacturer’s protocol. For polysome profiling, cells were treated with 10 µg/ml cycloheximide for 5 minutes prior to harvesting and counting. A total of 3x10⁶ cells were lysed and the lysate was subjected to ribosome fractionation using 7% to 47% sucrose gradient (Teledyne ISCO) as described previously. RNA was isolated from 40S and polysome fractions using Ribozol (Amresco). Isolated RNA was treated with Turbo DNase (Ambion). For qPCR of rRNA from total, 40S and polysome fractions, first strand cDNA synthesis was carried out using Superscript IV reverse transcriptase (Invitrogen) with random hexamer priming. qPCR was performed with iQ™ SYBR Green master mix with the 18S and 28S rRNA primers described in Table S1.

For qPCR of the individual reporters: RNA was isolated using the Qiagen RNA mini-prep per manufacturer protocol. Isolated RNA was treated with Turbo DNase (Ambion) prior to first strand cDNA synthesis using Superscript Vilo (Invitrogen). Quantitative PCR was performed with EGFP and mCherry primers, Table S1.

4.5.4 Illumina Library Preparation

First strand cDNA synthesis for ribosome associated RNA or total RNA was carried out using Superscript IV reverse transcriptase (Invitrogen) with random hexamer priming. The barcode was amplified from cDNA or plasmid using RE_Amp_F and RE_Amp_R primers with Phusion-HF MM (NEB): 98 °C for 1 min, 22 cycles: 98 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s, and 72 °C for 5 min. The amplicon was purified using Nucleospin Gel and PCR cleanup kit
(Macherey Nagel) and subsequently digested with \textit{XhoI} and \textit{SpeI}. The digestion product was purified as before and ligated to the Illumina adapters described in Supplemental Table 1. This product was amplified using Il_Enrich\_F and Il_Enrich\_R with Phusion HF MM (NEB): 98 °C for 1 min, 21 cycles: 98 °C for 10 s, 66 °C for 30 s, 72 °C for 30 s, and 72 °C for 5 min. This product was resolved by agarose gel electrophoresis and the appropriate sized band was excised and purified using Nucleospin Gel and PCR cleanup kit (Macherey Nagel).

The Illumina library was multiplexed and run on four lanes of Illumina NextSeq machine. Barcodes counts were determined. Only barcodes with greater than >10 counts in the cDNA and plasmid pools were used for analysis.

\subsection*{4.5.5 Western Blot Analysis}

Cells that were transfected with individual reporters were lysed with Lysis Buffer (50 mM Tris pH 7.8, 150 mM NaCl, 1% NP-40). Western blot analysis was performed as described previously \textsuperscript{104}. The following primary antibodies were used in western analysis at the given dilution: GFP, 1:2000 (Clontech, 632381); \(\beta\text{-actin-HRP}, 1:2000\), (Cell Signaling, 12262); Anti-mouse IgG HRP, 1:10,000 (Cell Signaling, 7076S).

\subsection*{4.5.6 Data Analysis}

Relative RNA expression for each regulatory element was calculated as described below. In brief cDNA counts for each barcode were normalized by the plasmid counts for the same barcode. The normalized expression was set relative to the median normalized expression of the control, 4 x Blank.

\[
\text{Relative RNA Expression} = \log_2 \left( \frac{\frac{cDNA_x}{\text{plasmid}_x}}{\text{median} \left( \frac{cDNA_{control}}{\text{plasmid}_{control}} \right)} \right)
\]
Relative TE for each regulatory element was calculated as described below. In brief, polysome associated cDNA (pRNA) counts for each barcode were normalized by the plasmid counts for the same barcode. The normalized expression was set relative to the median normalized TE of the control, 4 x Blank.

\[
Relative \ TE = \log_2 \left( \frac{pRNA_x/cDNA_x}{\text{median}(pRNA_{control}/cDNA_{control})} \right)
\]

Relative 40S association for each regulatory element was calculated as described below. In brief, 40S associated cDNA (srRNA) counts for each barcode were normalized by the plasmid counts for the same barcode. The normalized expression was set relative to the median normalized 40S association of the control, 4 x Blank.

\[
Relative \ 40S \ Association = \log_2 \left( \frac{srRNA_x/cDNA_x}{\text{median}(srRNA_{control}/cDNA_{control})} \right)
\]

Relative TIE for each regulatory element was calculated as described below. In brief, cDNA from polysome associated RNA (pRNA) counts for each barcode were normalized by the cDNA from 40S associated RNA (srRNA). The normalized expression was set relative to the median normalized TIE of the control, 4 x Blank.

\[
Relative \ TIE = \log_2 \left( \frac{pRNA_x/srRNA_x}{\text{median}(pRNA_{control}/srRNA_{control})} \right)
\]

### 4.5.7 Model Fitting

For each synthetic 3’UTR, we calculated median fold change across all 10 barcodes. Median fold change values were fit to linear model with interacting terms for the let-7 binding site, PRE, SRE, AU-rich element or space-holding sequence, at four positions using the lm
function in R \(^{106, 107}\). Coefficients were obtained in reference to 'blank’ sequence at each position. For cross-validation, we randomly divided the data into five parts and used 80% of the data to train and tested on the remaining 20%. This procedure was repeated five times. The parameters for our linear regression model are shown below:

\[
\text{Relative RNA Expression} \sim P_1 + P_2 + P_3 + P_4 + \text{sum}(P_i \times P_j)
\]

\[
\text{Relative TE} \sim P_1 + P_2 + P_3 + P_4 + \text{sum}(P_i \times P_j)
\]

Where \(j = 1\) to \(4\) and \(i \neq j\), \(* = \text{Interactions*} \)

We also modeled our data using a thermodynamic framework that fits parameters that are proportional to free energies of RBP/miRNA – RBP/miRNA and RBP/miRNA–mRNA decay machinery interactions \(^{81, 82}\). We used fitting routines and custom python scripts described elsewhere \(^{77}\). All RBP/miRNA were assumed to be present at the same concentration in the cell and bind 3'UTR with same affinity. The affinity for the mRNA decay machinery to 3'UTR was set at 2 units. We fit different models with and without interactions between let-7 binding site, PRE, SRE, and the 'blank’ sequence.

4.5.8 Code Availability

Scripts used for analysis and model fitting are available at the Github repository under MIT license (https://github.com/hemangichaudhari/Cottrell_PCRE-seq_scripts).
4.6 References


4.7 Supplemental Information

4.7.1 Supplemental Figures

Supplemental Figure 4.1: Co-occurrence of regulatory elements in human 3’UTRs
The regulatory elements used in the library appear together in human transcripts. Venn-diagram showing the number of human mRNA transcripts that contain either a miRNA recognition element (MRE), PRE or ARE; and all combinations of those elements. Panels b-g show
histograms of the distance between various mRNA regulatory elements. Distances are between the start of an ARE or PRE, or the end of an MRE. The locations of AREs were obtained from the ARE Database\(^1\). The locations of MREs were obtained from TargetScan\(^2\). We identified the locations of all PREs by a search for the consensus PRE: UGUHAUA, where H is A, U or C, in all human 3’UTRs. Panels b, d and f show a broad view, +/- 4000 nt (bin-width of 100 nt) which is ~3*SD of the mean 3’UTR length in humans. Panels c, e, and g show a narrow view of +/- 150 nt (bin-width of 1 nt) which more closely resembles the size of the synthetic 3’UTRs used in the library. For MRE:ARE, 14.6% of all pairs are within 150 nt of each other. For PRE:ARE, 15.4% of all pairs are within 150 nt of each other. For MRE:PRE, 13.7% of all pairs are within 150 nt of each other.
Supplemental Figure 4.2: Replicate to replicate comparisons
The 40s fraction is enriched for the 18s rRNA of the small ribosomal subunit. a Representative polysome profiling trace. b qPCR was used to detect the abundance of the 18s and 28s rRNA in either the 40s fraction, the pooled polysome fractions or total RNA. Shown is the ratio of 18s/28s rRNA. Replicate to replicate comparisons of 40s association, c, relative RNA expression, d, and TE, g. The inset tables describe the Pearson correlation for each replicate to replicate comparison. Scatterplots of 40S association, RNA or TE, e, f and h.
Supplemental Figure 4.3: Summary statistics for PTRE-seq analysis of RNA expression and TE

Histograms show the coefficient of variance (CV) across all barcode and replicates for each reporter within the library. Panel a shows the CV for RNA expression while panel b shows the CV for RNA expression with all outliers removed. Panel c shows the CV for TE while panel d shows the CV for TE with all outliers removed. Other summary statistics for RNA expression and TE measurements are shown in the table. Outliers were defined as any value more than 1.5 *IQR above Q1 or below Q3. Panels e-j show histograms of barcode counts (e, f, g) or reporter counts (sum of all barcode counts for each reporter; h, i, j) from one replicate of plasmid, RNA and polysome associated RNA sequencing.
Supplemental Figure 4.4: Validation of PTRE-seq results by qPCR, fluorescence measurements and western blot analysis

We selected eleven reporters from our PTRE-seq library and transfected them individually into HeLa cells along with a plasmid for the expression of mCherry as a control, cells were harvested 40 hours later. Panel a shows the effect of the regulatory elements within these reporters as determined by PTRE-seq measurement of RNA expression and TE. Panel b shows a comparison between PTRE-seq measurement of RNA expression and qPCR of EGFP. For qPCR measurements of RNA expression, EGFP expression was normalized to mCherry. There is a strong correlation between PTRE-seq measurement of relative RNA expression and qRT-PCR of individual reporters, c. Panel d shows western blot analysis of EGFP and mCherry. Panel e shows a comparison between qPCR measurements of RNA expression and fluorescent measurement of EGFP protein expression. For fluorescent measurements of EGFP expression the transfected cells were split into a 96-well plate the day after transfection. 40 hours after transfection the EGFP and mCherry fluorescence was measured using a plate reader. EGFP fluorescence was normalized to mCherry. For panels a, b and d relative RNA expression, TE or fluorescence was set relative to that of the control reporter, 4x Blank.
Supplemental Figure 4.5: Fit of linear-regression model for RNA expression

Predicted relative RNA expression is plotted on the y-axis and measured relative RNA expression is plotted on the x-axis. The Pearson correlation coefficient shows a strong correlation between the predicted and measured RNA expression. b-f Five-fold cross validation of the model linear-regression model for RNA expression.
Supplemental Figure 4.6: Fit of linear-regression model for TE
Predicted relative TE expression is plotted on the y-axis and measured relative TE expression is plotted on the x-axis. The Pearson correlation coefficient shows a strong correlation between the predicted and measured TE expression. b-f Five-fold cross validation of the model linear-regression model for TE.
Supplemental Figure 4.7: Validation of the linear regression models for RNA and TE

The linear regression model described above was used to predict the relative expression and TE of reporters containing only Let-7-target site, PRE, ARE, SRE and some combinations of those sites. The model predicted well the RNA expression, a, and TE, b, of reporters containing Let-7 binding sites. The correlations between predicted and measured RNA expression and TE for other 3’UTR elements and their combinations are shown in the table.

<table>
<thead>
<tr>
<th>Regulatory Element</th>
<th>Correlation RNA Model</th>
<th>Correlation TE Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Let-7</td>
<td>0.959466645</td>
<td>0.96599286</td>
</tr>
<tr>
<td>PRE</td>
<td>0.970774702</td>
<td>0.821543573</td>
</tr>
<tr>
<td>ARE</td>
<td>0.908129791</td>
<td>0.933569444</td>
</tr>
<tr>
<td>SRE</td>
<td>0.812835373</td>
<td>0.207737487</td>
</tr>
<tr>
<td>Let-7/PRE</td>
<td>0.955130642</td>
<td>0.937768779</td>
</tr>
<tr>
<td>Let-7/ARE</td>
<td>0.927219508</td>
<td>0.926926093</td>
</tr>
<tr>
<td>ARE/PRE</td>
<td>0.939196771</td>
<td>0.91296787</td>
</tr>
</tbody>
</table>
Supplemental Figure 4.8: Thermodynamic model of contribution of PRE, let-7 and SRE to TE

a A model with six parameters explained the observed data well ($R^2 = 0.86$). Red lines represent interactions with destabilizing effect on RNA. Black lines represent interactions with stabilizing effect on RNA. Solid lines represent statistically significant interactions and dashed lines represent non-significant interactions. Parameters values are in Supplementary Table 1. b Scatter plot shows the observed (x-axis) versus predicted (y-axis) normalized RNA counts from the model shown in panel A. c A table describing the parameters used in the model and their effects on TE.
Supplemental Figure 4.9: Thermodynamic model of contribution of PRE, let-7 and SRE to RNA stability

A model with nine parameters explained the observed data well ($R^2 = 0.883$). Red lines represent interactions with destabilizing effect on RNA. Black lines represent interactions with stabilizing effect on RNA. Solid lines represent statistically significant interactions and dashed lines represent non-significant interactions. Parameters values are in Supplementary Table 1. b Scatter plot shows the observed (x-axis) versus predicted (y-axis) normalized RNA counts from the model shown in panel A. c A table describing the parameters used in the model and their effects on RNA stability.
Supplemental Figure 4.10: Detailed results of let-7 binding site containing reporters
Minimal position dependent effect of let-7 binding site, a, or PRE, b, on RNA expression or TE, c and d. Fold repression of let-7, e, or PRE, f, containing reporters shown on a linear scale to highlight saturation of the effect on repression.
Supplemental Figure 4.11: The effect of seed pairing on repression of let-7 targeted reporters
Effect of seed pairing on RNA expression and TE for reporters containing one, a, two, b, or four, c, let-7 binding sites.
Supplemental Figure 4.12: The position of the let-7 binding site or PRE has no effect on repression of RNA a, or TE, b, when both elements are present in the 3’UTR. Panels c and d represent the same data plotted in Figure 4a and b as points.
Supplemental Figure 4.13: Interactions between AREs and the miRISC or Pumilio

**a** AREs modulate repression by miRNAs in a position dependent manner. **b** AREs modulate repression by PREs in a position dependent manner. * = Blank, A = ARE, L = let-7 and p = PRE.
Supplemental Figure 4.14: Effects of regulatory elements across cell types

The effect of seed sequence is on RNA expression is shown for each cell line tested, HDF, a, N2A, b, HEK, c, and HeLa, d. Fold change of RNA and TE for reporters containing SREs, e.
Supplemental Figure 4.15: Structures of ARE containing 3’UTRs
Structure of the 3’UTR of reporters containing a single ARE in position one, a, two, b, three, c, or four, d. Structures were drawn using mfold1.
4.7.1.1 Supplementary Figure References


### Supplemental Table 4.1: Primers and Oligonucleotides

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFP-F</td>
<td>CACCATGGTGAGCAAGGGCGAGG</td>
</tr>
<tr>
<td>RE-R</td>
<td>GGTACCAGATATCTCGTAGCCACAGTCTGAGGCTATTGTACAGCTGTC CATG</td>
</tr>
<tr>
<td>Lib_F</td>
<td>GTAGCATCTGTCGAGCTGAGCTGAGG</td>
</tr>
<tr>
<td>Lib_R</td>
<td>CTGATGATGATGCAGGGGGTCACC</td>
</tr>
<tr>
<td>Spacer_F</td>
<td>TCCGACCCTTCTCATTTGTCGCTAGCTCCGACTAC</td>
</tr>
<tr>
<td>Spacer_R</td>
<td>TCGAGTATCGGAGCTGACACGGAGATCAACGGGTGGATGGCA</td>
</tr>
<tr>
<td>RE_Amp_F</td>
<td>GTTGATCTCTCGGTAGCTCC</td>
</tr>
<tr>
<td>RE_Amp_R</td>
<td>TGCGATCTAGTAAAGCAGTAGGGAGTTGGAC</td>
</tr>
<tr>
<td>Dme-Smg-Ovl-F</td>
<td>CAAGGGTGGGGCGCGTCTAATGAAGTGACTGCAACTCGGAAAGCTGAC</td>
</tr>
<tr>
<td>Dme-Smg-R</td>
<td>TTGATATCGTAAAGTGGGATGCAAACGAGGAGGGATGAGG</td>
</tr>
<tr>
<td>mCh-F</td>
<td>CACCATGGGATGAGCAGAGAAGGGATGAGGATGAGGAGG</td>
</tr>
<tr>
<td>mCh-Ovl-R</td>
<td>TGGATGAGCTCAAGCTGCAAGCTGCAAGCTGCAAGCTGCA</td>
</tr>
<tr>
<td>18S rRNA F</td>
<td>GATGCCCTTAGATGTCGAGG</td>
</tr>
<tr>
<td>18S rRNA R</td>
<td>ATGGGGTTCTAAAGGGGTGGTACC</td>
</tr>
<tr>
<td>28S rRNA F</td>
<td>AGTAAACGGCGAGTGAGG</td>
</tr>
<tr>
<td>28S rRNA R</td>
<td>GCCTCGATCGAGGAGACTTG</td>
</tr>
<tr>
<td>P1-1_F</td>
<td>AATGATACGGCGAGCATCTACACTCTCTCTCTACAGCGAGCTC</td>
</tr>
<tr>
<td>P1-1_R</td>
<td>/5’Phos/ T*CGAATCGAGCAGATCGGAAAGAAGGTGCTGTCGAGGGAAAGAGTATGCTGTCGAGGAGGACTT</td>
</tr>
<tr>
<td>P1-2_F</td>
<td>AATGATACGGCGACCCAGCATCTACACTCTCTCTCTACAGCGAGCTC</td>
</tr>
<tr>
<td>P1-2_R</td>
<td>/5’Phos/ T*CGAATAGCTCAAGATCGGAAAGAAGGTGCTGTCGAGGGAAAGAGATGAGTCTGTCGAGGAGGACTT</td>
</tr>
<tr>
<td>P1-3_F</td>
<td>AATGATACGGCGACCCAGCATCTACACTCTCTCTCTACAGCGAGCTC</td>
</tr>
<tr>
<td>P1-3_R</td>
<td>/5’Phos/ T*CGAATAGCTCAAGATCGGAAAGAAGGTGCTGTCGAGGGAAAGAGATGAGTCTGTCGAGGAGGACTT</td>
</tr>
<tr>
<td>P1-4_F</td>
<td>AATGATACGGCGACCCAGCATCTACACTCTCTCTCTACAGCGAGCTC</td>
</tr>
<tr>
<td>P1-4_R</td>
<td>/5’Phos/ T*CGATGATCCACTCATAGATCGGAAAGAAGGATGCTGTCGAGGGAAAGAGATGAGTCTGTCGAGGAGGACTT</td>
</tr>
<tr>
<td>P1-5_F</td>
<td>AATGATACGGCGACCCAGCATCTACACTCTCTCTCTACAGCGAGCTC</td>
</tr>
</tbody>
</table>
P1-5_R

/5Phos/

T*CGATACCTACTGACAGATCGGAAGAGCGTCGTGTAGGGAAGAGATGTATCTCGGTTGGTCGCCCTCATCATT

P2-1_F

/5Phos/

C*TAGAGATCGGAAGAGCGACACGTCTGAACTCCAGTCACGTTGAAGAGGAAAGAGTGTACGTGTCTCTTCGATCT

P2-1_R

CAAGCAGAAGACGGCATACGAGATCGGAAGAGCGTCGTGTAGGGAAGAGATGTATCTCGGTTGGTCGCCCTCATCATT

P2-2_F

/5Phos/

C*TAGAGATCGGAAGAGCGACACGTCTGAACTCCAGTCACGTTGAAGAGGAAAGAGTGTACGTGTCTCTTCGATCT

P2-2_R

CAAGCAGAAGACGGCATACGAGATCGGAAGAGCGTCGTGTAGGGAAGAGATGTATCTCGGTTGGTCGCCCTCATCATT

II_Enrich_F

AATGATACGGCGACCACCGAG

II_Enrich_R

CAAGCAGAAGACGGCATACGAGATCCTTGCAACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

EGFP_qF

AGGACGCAGGAGCACTACAAG

EGFP_qR

AAGTCGATGCCCTTCAGCTC

mCh_qF

CAAGGGCGAGGAGGATAACAT

mCh_qR

ACATGAACCTGAGGGAGGACAG

Control Oligo

GTAGCATACTGTCCGCTAGCATCGACCTCGACTGTGGAATCGCCCTTCTTCTATCGGGTCGCCAGGCTCCAGAGCAATCATGACCTGAGACTGGAGTCAGGAGGTAGTG

*7** Oligo

GTAGCATACTGTCCGCTAGCATCGACCTCGACTGTGGAATCGCCCTTCTTCTATCGGGTCGCCAGGCTCCAGAGCAATCATGACCTGAGACTGGAGTCAGGAGGTAGTG

7777 Oligo

GTAGCATACTGTCCGCTAGCATCGACCTCGACTGTGGAATCGCCCTTCTTCTATCGGGTCGCCAGGCTCCAGAGCAATCATGACCTGAGACTGGAGTCAGGAGGTAGTG

7pcx2 Oligo

GTAGCATACTGTCCGCTAGCATCGACCTCGACTGTGGAATCGCCCTTCTTCTATCGGGTCGCCAGGCTCCAGAGCAATCATGACCTGAGACTGGAGTCAGGAGGTAGTG
**Supplemental Table 4.2: Binding sites for RBP and miRNA**

<table>
<thead>
<tr>
<th>Regulatory Element</th>
<th>Sequence</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Let-7 binding site</td>
<td>aTCGAGACTATACAAGGATCTACCTCAGTCGca</td>
<td>Synthetic ¹</td>
</tr>
<tr>
<td>AU-rich Element (ARE)</td>
<td>UAUUUUAUUUAUUUAUUUGUUUGUUUGUUAUU</td>
<td>IL-1β 3’UTR ²</td>
</tr>
<tr>
<td>Pumilio Recognition Element (PRE)</td>
<td>gucagcuccgacuUGUAAAUAucagccucgacu</td>
<td>Consensus binding site ³</td>
</tr>
<tr>
<td>Smaug Recognition Element (SRE)</td>
<td>cacaaGCAGAGGCUCUGGCAGCUUUUGCcaaca</td>
<td>Nanos 3’UTR (Dme) ⁴</td>
</tr>
<tr>
<td>Blank</td>
<td>aucagcuccgacugcuucuguagcc</td>
<td>BGH 3’UTR, pcDNA5/FRT/TO (Invitrogen)</td>
</tr>
</tbody>
</table>

Uppercase indicates nucleotides that form the binding site, lowercase represents additional nucleotides added to maintain constant length.

### 4.7.2.1 Supplementary Table References

Chapter 5: Future Directions

5.1 Elucidating the mechanisms post-transcriptional regulation

For many RBPs and especially for miRNAs, there is evidence for multiple modes of translational repression and/or mRNA decay of their targets. Often the evidence for one mechanism is contradicted by another. In the future we would like to test some of the proposed mechanisms for miRNA-mediated or RBP-mediated post-transcriptional control using massively parallel reporter libraries. For instance, we would like to combine knockdown or inhibition of specific proteins with our PTRE-seq library described above. Because our library has binding sites for multiple trans-acting factors we will be able to study the requirement of specific proteins in regulation mediated by multiple factors at once. A downside to this type of analysis is the potential for pleiotropic effects caused by the genetic depletion or inhibition of a protein that is involved in the translation or degradation of many mRNAs. One alternative approach to studying the importance of a given protein in miRNA or RBP-mediated post-transcriptional regulation will be to use RNA immunoprecipitation. We will perform immunoprecipititation of various translation, deadenylation, or decapping factors from cells transfected with our PTRE-seq library. By doing so we can determine which factors are associated with the reporters within our library, and potentially more importantly, which factors lose association when certain 3’UTR elements are present.

While the above approaches will allow us to directly test the importance of specific proteins in miRNA or RBP-mediated regulation other approaches will be needed to study the effect of cis-elements of mRNA on post-transcriptional regulation. In Chapter 3, I described the results of our systematic assessment of mRNA characteristics on miRNA-mediated repression.
We observed modulation of miRNA-mediated repression upon changes to codon optimality, 5’UTR structure, uORFs and 3’UTR sequence. Changes to codon optimality greatly affected the repression of multiple reporters. Recent work has shown that codon usage affects mRNA stability and protein synthesis. In yeast, the DEAD-box helicase Dhh1 has been implicated in sensing codon optimality and promoting decapping and decay of mRNAs with poor codon optimality. The mammalian homologue of Dhh1 is DDX6. As discussed in Chapter 1, DDX6 has been implicated in miRNA-mediated repression. More research is needed to better understand the effects of codon optimality on miRNA-mediated repression and whether DDX6 has an important role in this system.

**Identifying factors required for post-transcriptional regulation mediated by miRNAs and RBPs**

Our PTRE-seq library allows us to explore the factors required for post-transcriptional regulation by several mRNA trans-acting factors at once. We have explored the requirement of eIF4A2 for miRNA-mediated repression using our library. The PTRE-seq library was transfected into WT and eIF4A2-ko NIH3T3 cells (a kind gift from Jerry Pelletier, McGill University). We harvested total RNA and sequenced barcodes as described in Chapter 4. We observed little to no effect of eIF4A2 loss on miRNA-mediated repression, or the RNA expression of any of our reporters, Figure 1. In the future we would like to use similar genetic and potentially pharmacological approaches to study the roles of other translation factors in post-transcriptional regulation mediated by miRNAs or RBPs. This includes inhibition or genetic depletion of other dead-box helicases: DDX6, eIF4A1 and eIF4A2.
Figure 5.1: eIF4A2 is not required for miRNA-mediated repression

a Barcode counts in RNA isolated from the WT and eIF4A2-ko cells were normalized to barcode counts for the control reporter, 4x-Blank. Points colored blue correspond to reporters that contain at least one Let-7 binding site. b Relative expression of Let-7 targeted reporters in WT and eIF4A2-ko cells.

Translation rates and post-transcriptional regulation mediated by miRNAs and RBPs

One striking observation described in Chapter 3, is the modulation of miRNA-mediated repression by codon optimality. Codon optimality is known to influence translation and RNA stability in lower organisms, yeast in particular. The effects of codon optimality in higher eukaryotes is less studied. Because our experiments in Chapter 3 were carried out using Drosophila S2 cells, we wanted to assay the effects of codon optimality on miRNA-mediated repression in human cells. We used PCR to fuse Renilla luciferase with altered codon optimality from the plasmids described in Chapter 3 to a series of four target sites for the miRNA let-7 or control sites with mutated seed binding regions. The sequences of these sites are shown in Figure 2 below. The PCR product was inserted into pENTR-D/TOPO and subsequently transferred to pcDNA-DEST40. These plasmids were then co-transfected into HeLa cells with a plasmid
encoding firefly luciferase (pcDNA-D40-FF). The cells were harvested the next day. Luciferase activity and RNA expression was determined as described in Chapter 3. We observed increased translational repression and RNA decay when codon optimality was increased. This result suggests that the effects of codon optimality on miRNA-mediated repression are conserved from *Drosophila* to humans.

In the future it will be important to repeat this experiment with additional *Renilla* coding sequences (more and less optimal). We proposed in our discussion in chapter 3 that the effects of codon optimality on miRNA-mediated repression could be caused by a mismatch in the rates of translation elongation and initiation. To test this hypothesis, it would be necessary to empirically determine the initiation and elongation rate of miRNA-targets of interest. One could then assay the effects of altering those rates on miRNA-mediated repression. It may be observed that miRNA targets with very slow initiation are poorly repressed while targets with very fast initiation are well repressed. It would be important to compare the initiation rate to the elongation rate. How is miRNA-mediated repression affected when elongation is limiting as opposed to initiation?
Figure 5.2: Codon optimality influences miRNA-mediated repression

*Renilla* luciferase reporters were constructed with a tAI of 0.29, 0.34, 0.38 and 0.44. The reporters were transfected into HeLa cells in parallel with a plasmid expressing firefly luciferase. The cells were harvested 24 hours later. The sequence of the miRNA target sites for the reporters is shown in panel a. The underlined regions indicate the let-7 binding sites with the bold bases indicating the seed-pairing region. The top sequence is the targeted sequence and the bottom is the non-target control with mutated seed (lower case bases). Increasing the tAI from 0.34 to 0.38 increased translational repression b and mRNA degradation d. Panel c shows the normalized luciferase activity (*Renilla*/Firefly) for each of the targeted reporters. There is a correlation between tAI and luciferase activity. This experiment was performed by myself and a rotation student in the Djuranovic lab, Kellan Weston.
5.2 AU-rich element binding proteins

Using PTRE-seq we observed a striking 3’UTR position dependent effect of AU-rich elements (ARE). The ARE used in our library was the same for all reporters that contained the element. However, the position of the element determined the translation efficiency and RNA expression of the mRNA. We speculated that this could be caused by differences in the structure surrounding the mRNA. In particular, we hypothesized that different structures could serve as binding sites for different ARE-binding proteins. To test this hypothesis, we performed RNA immunoprecipitation (RIP) of a well-known ARE-binding protein, HuR. We transfected HeLa cells with our PTRE-seq library described in chapter four. We performed RIP for HuR using a previously published RIP protocol. We used anti-HuR monoclonal antibody (a kind gift from Ivan Toposovic, McGill University) and Protein L magnetic beads (Fisher). We sequenced barcodes from the HuR IP and the input lysate as described in Chapter 4. Figure 3b shows the correspondence between normalized barcode counts in each sample. Several reporter mRNAs that contained AREs were enriched in the HuR IP. This finding is clearly demonstrated by the volcano-plot in Figure 3c. When we looked at reporters that contained only an ARE or the blank control sequence, we observed a position dependent effect of the ARE on HuR association. Strikingly, reporter mRNAs that had reduced relative expression were all bound by HuR. Conversely, reporter mRNAs that were not bound by HuR had increased mRNA expression.

In the future it will be important to perform a similar analysis by pulldown of other ARE-BP: TTP, Auf1, and other ELAVL family members. It will be important to confirm these results genetically by knockdown or knockout of each ARE-BP and assaying the relative expression and TE of the ARE-containing reporters within the PTRE-seq library.
Figure 5.3: Association of HuR with ARE-containing mRNAs

a Western blot showing enrichment of HuR in the immunoprecipitate of anti-HuR antibody. b Barcode counts in RNA isolated from the input or HuR-IP were normalized to barcode counts for the control reporter, 4x-Blank. Points colored blue correspond to reporters that contain at least one ARE. c Volcano blot of fold enrichment (HuR-IP counts/Input counts, relative to control reporter) and significance, q-value. The q-value was determined by an FDR correction of Wilcoxon (Mann-Whitney) p-values for the comparison of HuR-IP and Input counts for all replicate barcodes for each reporter. This experiment was performed by myself and a rotation student in the Djuranovic lab, Kellan Weston.

In parallel with the HuR immunoprecipitation described above we also performed an HuR knockdown experiment. Although one of the siRNAs used caused robust knockdown of HuR, we observed no differences in RNA expression of our reporters in the HuR knockdown cells compared to the control siRNA (si-NC). This result is surprising when compared to the results of our HuR-RIP described above. It was expected that knocking down HuR would have
resulted in a change of expression for the reporters shown to be bound by HuR. This could indicate that HuR does not regulate the expression of the reporter mRNAs in our library to which it is bound. It is also possible that the knockdown wasn’t sufficient to cause an effect on RNA expression. If HuR has very few mRNA targets in the cell-line we are using (HeLa) relative to the abundance of HuR then it is possible that we would need a more robust knockdown to see an effect on our reporter mRNAs. Complicating the use of siRNAs for this experiment is our observation that the control and HuR siRNAs caused altered expression of all the reporters in our library when compared to untransfected HeLa, Figure 4. We suspect this could be due to sequestration of RISC complexes by the siRNAs resulting in global changes in post-transcriptional regulation. To counter this complication, it may be more prudent to knockout HuR using CRISPR-Cas9.
Figure 5.4: Knockdown of HuR
HeLa cells were transfected with the PTRE-seq library described above and with a negative control siRNA or one of three siRNAs specific for HuR. Western blot analysis, a, c, and qRT-PCR, c, revealed efficient knockdown of HuR with siRNA #2. RNA was isolated from cells transfected with the negative control siRNA (siNC) and siRNA #2 (siHuR) and we analyzed RNA expression by PTRE-seq. There was very little change in RNA expression of any reporter between the two transfections, b. Panels d-f show the RNA expression of various reporters in cells transfected with siNC, siHuR or un-transfected HeLa (from chapter 4).
5.3 References


