2015-08-04

A Gene-Centered Method For Mapping 3'UTR-RBP Interactions: A Dissertation

Alex M. Tamburino
University of Massachusetts Medical School

A GENE-CENTERED METHOD FOR MAPPING 3'UTR-RBP INTERACTIONS

A Dissertation Presented

By

Alex M. Tamburino

Submitted to the Faculty of the University of Massachusetts
Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 4, 2015

Biomedical Sciences

A GENE-CENTERED METHOD FOR MAPPING

3'UTR-RBP INTERACTIONS

A Dissertation Presented by

Alex M. Tamburino

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

Albertha J.M. Walhout, Thesis Advisor

Sean P. Ryder, Member of Committee

Allan Jacobson, Member of Committee

Scot A. Wolfe, Member of Committee

Timothy R. Hughes, Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

Job Dekker, Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school.

Anthony Carruthers, Ph.D.,

Dean of the Graduate School of Biomedical Sciences

Interdisciplinary Graduate Program

August 4, 2015

Dedication

I wish to dedicate this dissertation to my mentor, Marian. I rotated in the lab back in 2009 working on a project studying DNA-TF interactions using the yeast one-hybrid system. I didn't love the project (sorry!) but I loved the lab and I knew that I wanted to join.

I pitched a project to Marian...or more accurately several projects. The first few were duds, but she steered me through the process. Finally I pitched an idea to her that had some merit...or at least lacked glaring flaws. From there she guided me through the early stages of the process and connected me with a committee of advisors who provided the additional scientific knowledge and guidance.

All of these efforts helped get me through the many ups and downs I encountered over the years. She always gave me what I needed to suceed whether it be support, guidance, rigor, confidence or a swift kick in the ass.

For that I wish to say, Thank you Marian.

Acknowledgements

I would like to thank everyone who helped me throughout this dissertation. My mentor and thesis research advisory committee were instrumental in providing scientific guidance and knowledge from the earliest days of the project. In particular, Allan Jacobson provided critical suggestions that enabled PRIMA to get off the ground and Sean Ryder provided critical insights that helped me to overcome key obstacles and achieve success in the late stages of the project. Additionally Job Dekker helped inspire me and shape my thought process throughout all stages of the technology development.

I would also like to thank my lab mates and collaborators who contributed to and helped shape this project. In particular, Shaleen Shrestha and Ebru Kaymak contributed to the final stage experiments which helped complete this project. I also appreciated the many insightful discussions with my lab mates. Most notably, I recall many hours of deliberations John Reece-Hoyes, Lesley MacNeil and Juan Fuxman-Bass to whom I am grateful.

Abstract

Interactions between 3´ untranslated regions (UTRs) and RNA-binding proteins (RBPs) play critical roles in post-transcriptional gene regulation. Metazoan genomes encode hundreds of RBPs and thousands of 3' UTRs have been experimentally identified, yet the spectrum of interactions between 3´UTRs and RBPs remains largely unknown. Several methods are available to map these interactions, including protein-centered methods such as RBP immunoprecipitation (RIP) and cross-link immunoprecipitation (CLIP), yeast three-hybrid assays and RNAcompete. However, there is a paucity of RNA-centered approaches for assaying an RNA element of interest against multiple RBPs in a parallel, scalable manner.

Here, I present a strategy for delineating protein-RNA interaction networks using a gene centered approach. This approach includes annotating RBPs and identifying physical interactions between an RNA of interest and these RBPs using the Protein-RNA Interaction Mapping Assay (PRIMA). Few RBPs have been experimentally determined in most eukaryotic organisms. Therefore I show that existing RBP annotations can be supplemented using computational predictions of RNA binding domains (RBD) from protein sequences. A single RNA of interest can be tested using PRIMA against a library of RBPs constructed from these annotations. PRIMA utilizes the green fluorescent protein (GFP) in yeast as a reporter.

PRIMA is based on reconstitution of the interaction between the 5´ and 3´ ends of an mRNA, which increases mRNA stability and enhances translation. PRIMA recapitulates known and uncovers new interactions involving RBPs from human, *Caenorhabditis elegans* and bacteriophage with short RNA fragments and full-length 3´UTRs. The development of RBP prey libraries will enable the testing of 3´UTRs against the hundreds of RBPs, which is essential to gain broad insights into post-transcriptional gene regulation at a systems level.

Table of Contents

List of Figures and Tables

List of Symbols, Abbreviations or Nomenclature

3-AT	3-aminotriazole (3-AT)
AD	Activation domain
APA	Alternative polyadenylation
C2H2	Cysteine (2) Histidine (2) zinc finger DNA binding domain
CCCH	Cysteine (3) Histidine (1) zinc finger DNA binding domain
cDNA	Complementary DNA
ChIP	Chromatin Immunoprecipitation
CLIP	Cross-link Immunoprecipitation
DB	DNA binding domain
dsRBD	Double stranded RNA binding domain
eIF4E	Eukaryotic initiation factor 4E
eIF4G	Eukaryotic initiation factor 4G
EMSA	Electrophoretic mobility shift assay
FBE	FBF binding element
FL1	Fluorescence channel 1
FSC	Forward scatter
GFP	Green fluorescent protein
gs	gene specific
H2B	Histone 2B
HBE	Histone binding element

IP	Immunoprecipitation
IPTG	isopropyl 1-thio-β-D-galactopyranoside
KH	K homology domain
MCS	Multiple cloning site
miRNA	micro RNA
mRNA	messenger RNA
MS2	Bacteriophage MS2 coat protein
MS2BS	MS2 binding site
NRE	NOVA1 recognition element
nt	nucleotide
ONPG	ortho-Nitrophenyl-β-galactoside
ORF	Open reading frame
Pab1p	Poly(A) binding protein
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PNA	Peptide nucleic acid
PRIMA	Protein-RNA interaction mapping assay
PUF	Pumilio/FBF domain
RBD	RNA binding domain
RBP	RNA binding protein
RNAi	RNA interference

RNP	Ribonucleoprotein particle
RPM	Rotations per minute
RRM	RNA recognition motif
SAM	Sterile alpha motif
SCR	Spatial control region
SELEX	Systematic evolution of ligands by exponential enrichment
SLBP	Stem-loop binding protein
SSC	Side scatter
SXL	Sex lethal
TCR	Temporal control region
TF	Transcription factor
UTR	Untranslated region
wRBP1.0	worm RNA binding protein list 1.0
X-gal	5-bromo-4-chloro-3-indolyl-β-d-galactoside
Y3H	Yeast three-hybrid

CHAPTER I

Introduction

Proper gene expression is vital to organismal development and responses to environmental cues. Regulation of gene expression occurs at multiple levels including transcriptionally, post-transcriptionally, and post-translationally. Transcriptional and post-translational regulation of gene expression have long been a focus of genomics and systems biology, however post-transcriptional gene regulation has not received the same attention. Regulation by micro RNAs (miRNAs) is being increasingly studied, but few large scale analyses of messenger RNA (mRNA) regulation by RNA binding proteins (RBP) have been performed to date.

Many aspects of an mRNA's life cycle are regulated by RBPs including splicing, 3' end formation, nuclear-cytoplasmic export, localization, translation and stability (Moore, 2005; Glisovic *et al* 2008; Martin and Ephrussi, 2009; Mitchell and Parker, 2014). The current understanding is that mRNAs do not exist alone, but rather as ribonucleoprotein complexes (RNPs) (Keene 2007, Mitchell and Parker 2014). At the onset of transcription an mRNA receives a 7-methyl-guanosine cap at their 5' end and is subsequently bound by

eIF4E, the cap binding protein. The elongating transcript then interacts with additional RBP and RNP complexes including splicing proteins, small nuclear RNPs, exon junction proteins, and polyadenylation factors (Moore 2005). In most cases, the mRNA is polyadenylated at its 3' end and subsequently bound by the polyA binding protein, Pab1p (Mangus *et al.* 2003).

In the cytoplasm, mRNAs primarily exist in three general states: active translation, repression or degradation (Buchan and Parker 2009). The circularization of the 5' and 3' ends of the mRNA serves to promote active translation and prevent degradation of the transcript (Wells *et al.* 1998, Coller *et al.* 1998, Gray *et al.* 2000). The 5' and 3' ends are bound by eIF4E and Pab1p, respectively. Both bind the scaffold protein, eIF4G, thereby circularizing the transcript and promoting translation through enhanced ribosome binding (Wells *et al.* 1998). Further, the closed-loop scaffold provides a framework for the regulation of translation and decay by trans-acting RBPs (Szostak and Gebauer, 2013) (Figure 1.1).

RBPs can regulate the mRNA state by affecting ribosome recruitment, initiation and elongation (repression). During *Drosophila melanogaster* embryogenesis translation of many mRNAs is repressed spatially to create asymmetric protein distributions (Nakamura *et al.* 2004,

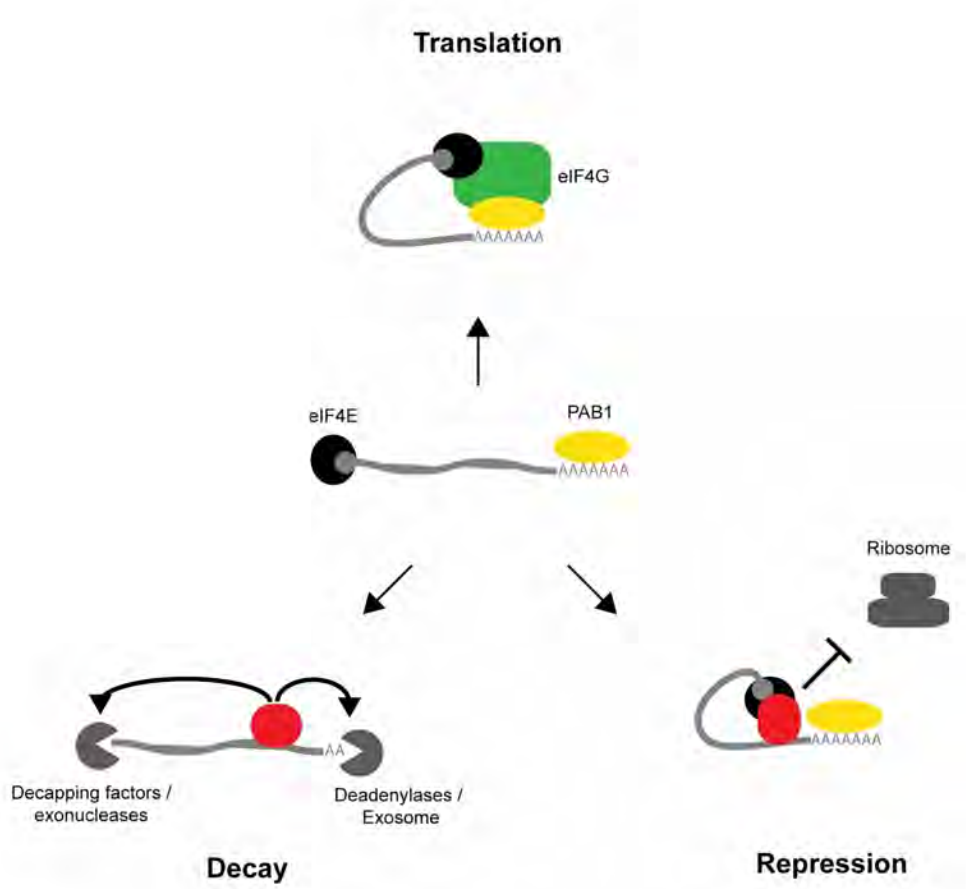


Figure 1.1: Three general states of an mRNA. In the cytoplasm, mRNAs primarily exist in three general states: active translation, repression or decay. The mRNA is bound at the 5' and 3' ends by eIF4E and Pab1p, respectively (center). Formation of the closed-loop by eIF4E results in efficient translation (top). RBPs which block this interaction result in repression of ribosomal recruitment (right). RBPs which remove the polyA tail and/or cap result in RNA decay (left).

Cho *et al.* 2005, Beckmann *et al.* 2005). For instance, *oskar* mRNA is regulated through its 3’ untranslated region (UTR) by the protein Bruno. Bruno interacts directly with eIF4E thereby disrupting the translation competent form of the closed-loop (Nakamura *et al.* 2004). The *caudal* mRNA is repressed in a related manner by Bicoid which binds the 3'UTR and recruits d4EHP. This paralog of eIF4E binds the 5’ cap but does not associate with the eIF4G scaffold protein (Cho *et al.* 2005). Lastly, the *msl-2* mRNA is inhibited by Sex-lethal (SXL) through dual mechanisms. SXL binds to the 3'UTR of *msl-2* and prevents binding of the 43S ribosomal preinitiation complex to the mRNA. Complexes which escape this first level of repression are then blocked from scanning to the downstream translation initiation codon (Beckmann *et al.* 2005).

RBPs can regulate the mRNA state by recruiting deadenylation and decapping factors that disrupt the closed-loop and destabilize the transcript (decay). Entry into mRNA decay typically begins with shortening of the poly(A) tail and removal of the 5’ cap followed by exonuclease degradation at the 5’ and 3’ ends (Garneau *et al.* 2007). In *Saccharomyces cerevisiae*, *HO* mRNA is regulated by two different RBPs, Puf4p and Puf5p (Hook *et al.* 2007, Chritton *et al.* 2010). The RBPs bind to separate elements in the *HO* 3'UTR and decrease *HO* mRNA stability by promoting deadenylation and disruption of the closed-loop (Hook *et al.* 2007). In particular, Puf5p

recruits the deadenylase Ccr4p (via Pop2p) and the decapping factor Dhh1p (via Eap1p) (Goldstrohm *et al.* 2007, Blewett and Goldstrohm 2012). Additionally, Puf5p represses translation *in vitro* independent of its effects on *HO* stability (Chritton *et al.* 2010).

C. elegans glp-1 illustrates the regulation of a single 3'UTR by many RBPs. GLP-1 is a conserved protein that is essential for germline development. *glp-1* mRNA is present throughout development, however GLP-1 protein is only expressed at distinct points (Crittenden *et al.* 1994) (Figure 1.2). The *glp-1* 3'UTR was determined sufficient for patterning of GLP-1 expression and subsequent mutation of the 3'UTR determined there were two regions which exhibited control over the temporal (TCR) and spatial (SCR) expression patterns (Evans *et al.*1994).

The regulation of *glp-1* 3'UTR elements by several RBPs was discovered through a combination of approaches. GLD-1 and POS-1 were postulated to be regulators based on correlative expression patterns and phenotypes (Marin and Evans 2003, Ogura *et al.* 2003). SPN-4 was suspected based on its physical association with POS-1 (Ogura *et al.* 2003). MEX-3, MEX-5 and OMA-1 were associated with *glp-1* through 'protein-centered' studies where the RBP was in focus (Pagano *et al.* 2007, 2009, Kaymak *et al.* 2013). In contrast, a large-scale 'gene-centered' screen focusing on the *glp-1* 3'UTR associated PUF-5

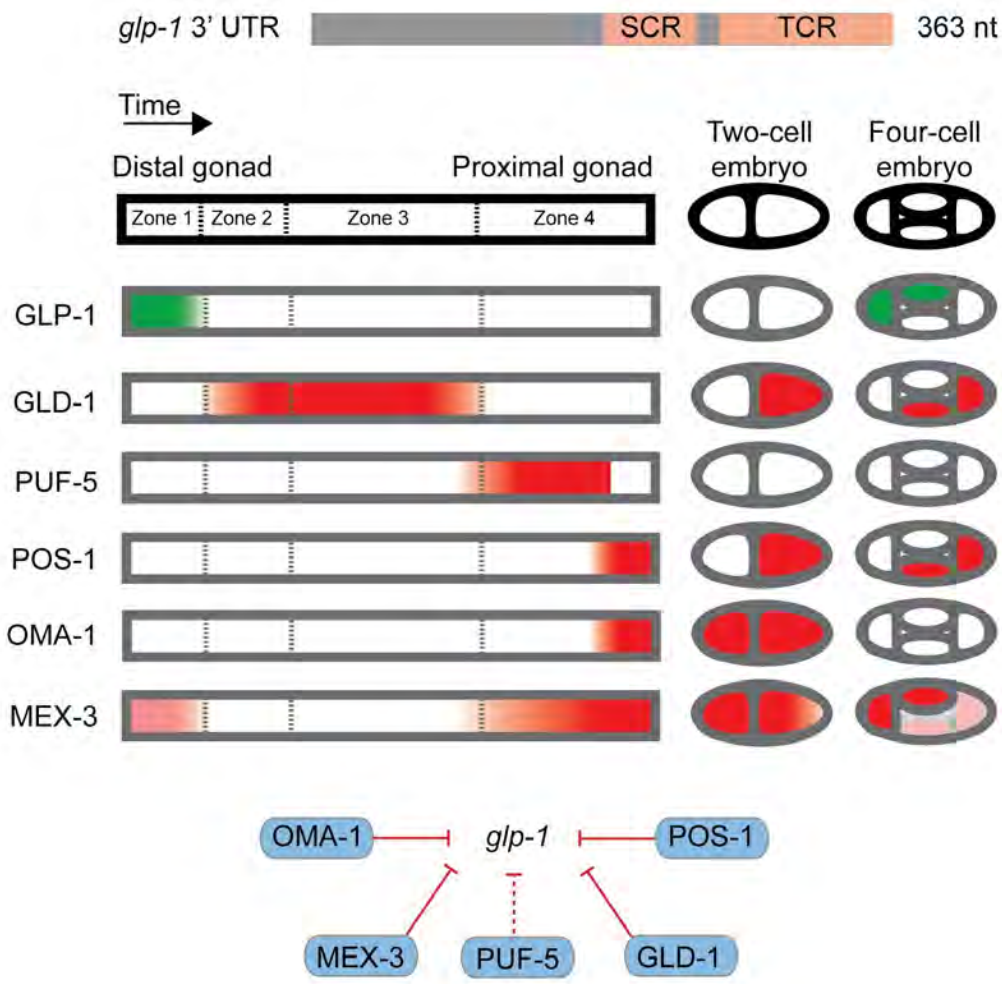


Figure 1.2: Combinatorial regulation of a 3'UTR by RBPs. The *glp-1* 3'UTR contains two *cis*-regulatory regions, SCR and TCR. Embryogenesis proceeds in an orderly fashion from the distal to proximal gonad then the two- and four-cell stages. GLP-1 protein is expressed in Zone 1 of the distal gonad and the anterior blastomeres of the four-cell stage (green). Five different RBPs known to repress *glp-1* mRNA have mostly opposite expression patterns (red).

with *glp-1*. In total, two decades of research has identified numerous regulators but few studies have taken an unbiased, comprehensive approach to identifying physical or regulatory interactions between mRNAs and RBPs. A new approach towards identifying physical interactions between a 3'UTR and many RBPs would facilitate such studies of *glp-1* and additional UTRs.

Can we assay a single 3'UTR against a library of RBPs?

Conceptually, there are two approaches for dissecting interactions with, or regulation of 3'UTRs by RBPs: protein-centered and gene-centered (Figure 1.3). Protein-centered approaches begin with an RBP of interest and seek to determine the 3'UTRs it binds; whereas, a gene-centered approach begins with the 3'UTR and seeks to determine the RBPs which bind it. Protein-centered approaches include *in vivo* interaction detection methods such as RBP immunoprecipitation (RIP) and cross-link immunoprecipitation (CLIP) or RBP specificity determination through electrophoretic mobility shift assays (EMSA), systematic evolution of ligands by exponential enrichment (SELEX), and RNAcompete (Keene *et al.* 2006, Ule *et al.* 2003, Pagano *et al.* 2011, Tuerk and Gold 1990, Ray *et*

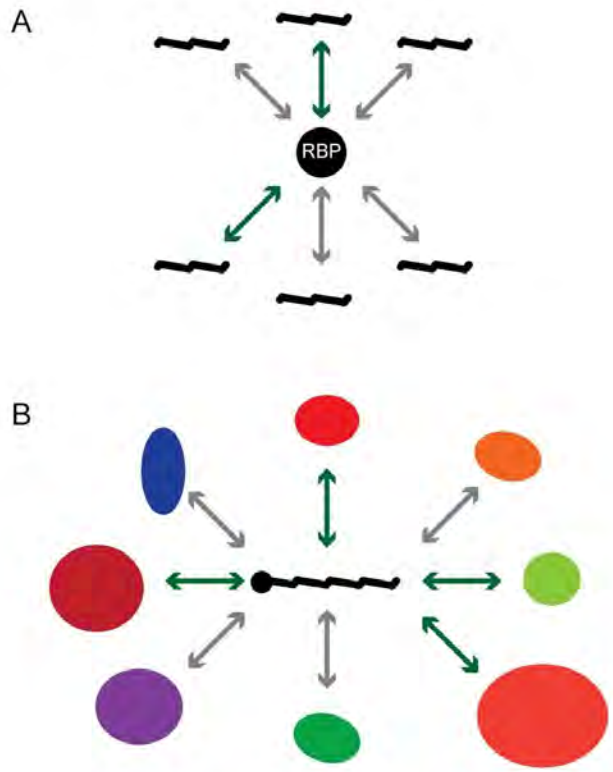


Figure 1.3: Comparison of gene- and protein-centered approaches. There are two approaches to identifying mRNA-RBP interactions: protein-centered and gene-centered. Protein-centered methods begin with a protein of interest and test for interactions amongst a set of RNAs or 3'UTRs. Gene-centered methods begin with an RNA of interest and test for interactions amongst a set of RBPs.

al. 2009). A major gene-centered approach is RNAi where the expression of a target mRNA or protein is assessed following the knockdown of individual RBPs (Kamath *et al.* 2003). A final method, yeast three-hybrid (Y3H) has been used in both manners (SenGupta *et al.* 1996).

Protein-centered techniques for identifying 3'UTR -RBP interactions

The primary *in vivo* protein-centered techniques are RIP and CLIP (Keene *et al.* 2006, Ule *et al.* 2003). In both methods a given RBP is immunoprecipitated (IP) from cell lines and the bound RNAs are identified using microarrays or RNA sequencing (Keene *et al.* 2006, Ule *et al.* 2005). The main difference is that with CLIP a cross-linking step is added before IP. The advantage of cross-linking is that more stringent washes can be utilized without losing lower affinity interactions. Both RIP and CLIP have been used to study the *C. elegans* RBPs FBF-1/2, RNP-8, and GLD-1 (Kershner *et al.* 2010, Kim *et al.* 2010, Wright *et al.* 2011, Jungkamp *et al.* 2011).

The primary *in vitro* techniques include EMSA, SELEX and RNAcompete. In all three methods a purified RBP is tested against a series of RNAs individually (EMSA) or as a library (SELEX, RNAcompete). In

EMSAs, an RNA is labeled radioactively (*e.g.* ^{32}P) or fluorescently (*e.g.* fluorescein) (Pagano *et al.* 2011). The labeled probe is incubated with purified protein and native polyacrylamide gel electrophoresis (PAGE) is performed. Binding by an RBP decreases the mobility of the probe. Quantitative measurement of the RBP-RNA interactions can be assessed by comparing binding of multiple protein concentrations against a fixed concentration of RNA (Pagano *et al.* 2011). Additionally, the RBP specificity can be quantitatively assessed by making single nucleotide mutations in the RNA and measuring the binding affinities. The specificity of multiple *C. elegans* RBPs has been determined using this approach (Pagano *et al.* 2007, 2009, Farley *et al.* 2008, 2012, Kaymak *et al.* 2013).

SELEX and RNAcompete both test a purified RBP against a library of short RNA sequences. In SELEX, the RBP is incubated with a randomized pool of purified RNAs (Tuerk and Gold 1990, Ellington and Szostak 1990). The RBP is captured then the bound RNAs are eluted and reverse transcribed into complementary strands of DNA (cDNA). The DNA is amplified via polymerase chain reaction (PCR) and transcribed back into RNA for subsequent rounds of selection. After multiple rounds, the DNA is sequenced and the enriched sequences are compiled into a representative binding motif. In RNAcompete, the library of RNAs is synthesized on an oligonucleotide microarray (Ray *et al.* 2009). The design of the microarray

helps ensure that individual RNA sequences are well represented in the library and it enables detection of bound RNAs on the same microarrays following a single round of selection. SELEX and RNAcompete have both been used to determine RNA binding specificities of several *C. elegans* RBPs (Pagano *et al.* 2009, Kaymak *et al.* 2013, Ray *et al.* 2013).

Protein-centered approaches are not ideal for research directed at understanding the regulation of a single 3'UTR. The primary reason is that there are hundreds of potential RBPs in eukaryotic organisms. To test all RBPs each technique would have to be performed hundreds of times. Additionally RIP and CLIP require suitable antibodies or epitope-tagged RBPs thus limiting broad use of these techniques. EMSA, SELEX and RNAcompete require protein purifications increasing the workload for each individual experiment. RNAcompete has been performed for hundreds of purified RBPs, however the resulting RNA binding motifs are often short and degenerate (Ray *et al.* 2013). These motifs are found in thousands of 3'UTRs, therefore additional assays are needed to test physical and regulatory interactions that are predicted to occur *in vivo* (Pagano *et al.* 2007, 2009, Kaymak *et al.* 2013).

Gene-centered techniques for identifying 3'UTR -RBP interactions

In comparison, gene-centered approaches such as RNAi enable direct testing of a single mRNA against many RBPs *in vivo*. In particular, transgenic cell lines and animal strains can be constructed wherein green fluorescent protein (GFP) is constitutively expressed from an mRNA containing a 3'UTR of interest. The expression patterns can be visualized via fluorescence microscopy. Regulatory interactions by RBPs can be determined by evaluating GFP expression following RNAi of RBPs (Pagano *et al.* 2009, Kaymak *et al.* 2013). This is a valuable approach and has the necessary throughput for testing hundreds of RBPs. Missed interactions can result due to subtle regulation, redundancy, and context dependent interactions. Also, interactions determined by RNAi are regulatory and therefore cannot be determined to occur through direct physical interaction without secondary assays.

When this project began, yeast three-hybrid (Y3H) was the best suited assay for the gene-centered detection of physical interactions between 3'UTRs and RBPs (Figure 1.4A). In Y3H, an RNA element of

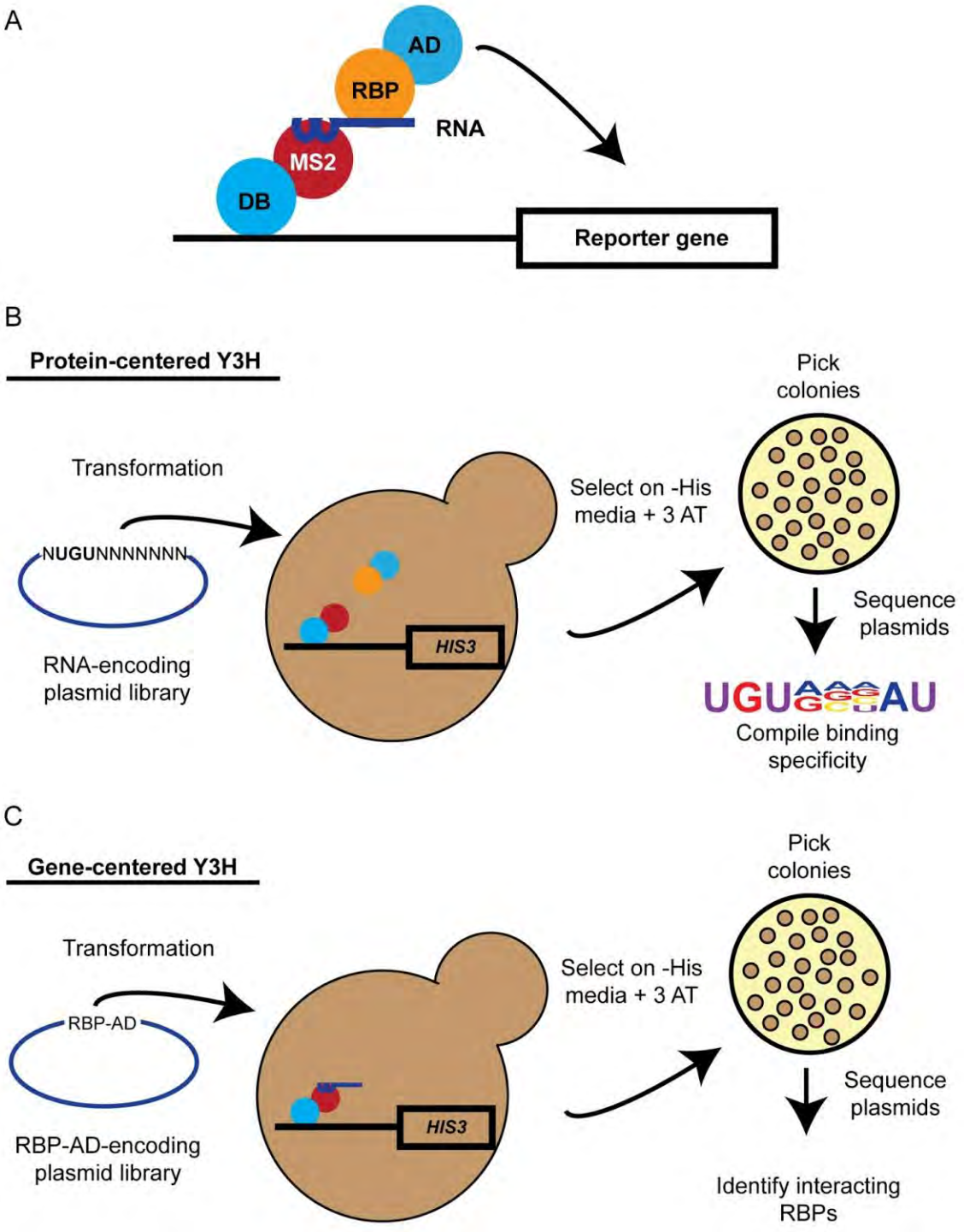

A
AD
RBP
RNA
MS2
DB
Reporter gene
B
Protein-centered Y3H
Pick colonies
Transformation
Select on -His media + 3 AT
NUGUNNNNNNN
HIS3
Sequence plasmids
RNA-encoding plasmid library
Compile binding specificity
C
Gene-centered Y3H
Pick colonies
Transformation
Select on -His media + 3 AT
RBP-AD
HIS3
Sequence plasmids
RBP-AD-encoding plasmid library
Identify interacting RBPs

Figure 1.4: Diagram of Y3H. A) Y3H tests RNA-RBP interactions using two fusion proteins and one fusion RNA. A DNA-binding domain (DB) - MS2 fusion protein binds to an RNA through two fused MS2 binding sites localizing the RNA upstream of a reporter gene. An RBP - activation domain (AD) fusion protein is coexpressed. If the RNA-RBP interaction occurs the AD is recruited upstream of the reporter thereby activating reporter expression. B) In protein-centered Y3H, a library of randomized RNA encoding plasmids are transformed into yeast containing the RBP-AD fusion and plated on selective media. If interactions occur with any transformed RNA the resulting yeast cell grows to form a visible colony. The selected colonies can be sequenced to identify the interacting RNA and a composite RBP binding site can be determined. C) In gene-centered Y3H, a library of RBP-AD encoding plasmids is transformed into yeast containing the MS2 binding site - RNA encoding plasmids. If interactions occur with any transformed RBP the resulting yeast cell grows to form a visible colony. The selected colonies can be sequenced to identify the interacting RBP.

interest is transcribed with two covalently attached high affinity sites for the bacteriophage MS2 coat protein (MS2). The RNA is constitutively expressed from an RNA polymerase III promoter and contains a RNAseP 5' leader sequence that promotes nuclear retention of the transcript. Additionally a DNA binding domain (DB) - MS2 fusion protein is co-expressed. The fusion protein serves to tether the RNA upstream of a reporter gene (*e.g. HIS3*) by binding the RNA (via MS2) and DNA binding sites (via DB) (SenGupta *et al.* 1996, Zhang *et al.* 1999). To test for an RNA-RBP interaction, the RNA 'bait' yeast is transformed with a RBP 'prey' that is fused to a TF transcription activation domain (AD). If the RBP interacts with the RNA element then the AD is recruited to the reporter gene thereby promoting reporter expression.

HIS3 and *LacZ* are frequently used as reporter genes. *HIS3* is essential for the biosynthesis of histidine and therefore yeast does not grow without the addition of exogenous histidine to the media. Additionally, low levels of 3-aminotriazole (3-AT), a competitive inhibitor of His3p, can be added to the media to increase selective pressure against low level reporter expression. Reporter activation is detected by assessing yeast growth on media lacking histidine and including 3-AT. *LacZ* encodes the beta-galactosidase protein. This enzyme converts the colorless compounds 5-bromo-4-chloro-3-indolyl-β-d-galactoside (X-gal) and

ortho-Nitrophenyl-β-galactoside (ONPG) into blue and yellow compounds, respectively. Reporter activation is detected by measuring the amount of color produced in the presence of either compound.

Y3H has been successfully utilized in both protein- and gene-centered configurations. The largest efforts using Y3H have centered on dissecting the nucleotide binding specificity of Pumilio/FBF (PUF) family members (Wickens *et al.* 2002). In a collection of works, PUF RBPs were tested in a protein-centered manner *vs.* a library of randomized Y3H RNA baits (SenGupta *et al.* 1999, Bernstein *et al.* 2005; Opperman *et al.* 2005, Stumpf *et al.* 2008, Koh *et al.* 2009) (Figure 1.4B). Bait RNAs were transformed into a yeast strain that expressed a PUF-AD prey and utilized the *HIS3* reporter. Thousands of individual colonies were screened following growth on selective media. Individual baits were sequenced to determine the interacting RNA elements and composite representations of each PUF RBP specificity was determined. As with other protein-centered assays, these specificities were used to predict target 3'UTRs which were bound by the given RBP. Physical interactions were assessed using Y3H and the *LacZ* reporter, or with orthologous assays such as EMSAs.

Most work used Y3H in the protein-centered configuration, however, it was initially utilized in a gene-centered manner (Wang *et al.* 1996, Zhang *et al.* 1997, Martin *et al.* 1997). A forward genetic screen studying *C.*

elegans sex determination identified 17 different gain-of-function mutants within a five nucleotide stretch of the *fem-3* 3'UTR (Ahringer and Kimble 1991). Additional evidence suggested that a post-transcriptional regulator was present in the worm germline, however the identity remained unknown (Ahringer *et al.* 1992). Y3H was used with the *HIS3* reporter to identify the regulatory RBPs. The wild-type *fem-3* 3'UTR element was the bait and a cDNA library of RBP-AD fusion preys was transformed into the bait strain (Figure 1.4C). Screening selected colonies identified two closely related *fem-3* binding factors, FBF-1 and FBF-2 (Zhang *et al.* 1997). Subsequent Y3H and RNAi experiments confirmed these proteins as regulatory factors that directly bound the *fem-3* 3'UTR.

Despite successful application of Y3H as a gene-centered method it has not been widely used to the study individual 3'UTRs. There are several potential, non-trivial explanations for this. First, there is a large drop in signal strength for RNA baits longer than 150 nucleotides (nt) (Zhang *et al.* 1999). The majority of metazoan 3'UTRs are longer than 150 nt (Mangone *et al.* 2010, Ulitsky *et al.* 2012, Derti *et al.* 2012) and would therefore be ineffective Y3H baits. Second, Y3H RNAs are transcribed by RNA polymerase III which is terminated by a tetrauridine signal (Zhang *et al.* 1999). 3'UTRs are particularly uridine rich and many contain tetrauridine sequences (Mangone *et al.* 2010, Jan *et al.* 2011, Ulitsky *et al.* 2012, Derti

et al. 2012). Thus testing many full-length 3'UTRs would be challenging using Y3H. Theoretically, an RNA polymerase II system without the same restrictions could be established, but this too has not been widely adapted (Zhang *et al.* 1999).

C. elegans as a model to study 3'UTR - RBP interactions

Briefly, *C. elegans* is a hermaphroditic nematode with 959 somatic cells. It is transparent, easy to culture and many techniques for genetic manipulations have been developed. In particular, transgenic strains can be generated including those which constitutively express GFP mRNAs fused to a 3'UTR of interest (Merritt *et al.* 1998). Additionally, mutant strains and RNAi clone libraries are available for the removal / downregulation of RBPs (Caenorhabditis Genome Center, Fraser *et al.* 2000, Kamath *et al.* 2003)

Development in *C. elegans* is an ordered process. There are two gonad arms which produce both sperm and oocytes (Hubbard and Greenstein, 2005). Within each gonad gametes move from the distal region proximally towards the vulva in an orderly fashion. In the distal region, a population of nuclei mitotically divide in syncytium replenishing the progenitor population (Kimble and Crittenden, 2005). As the nuclei

progress proximally they form cell walls and enter meiosis. Development continues as the oocyte moves through the loop and proximal regions. As the oocyte passes the spermatheca it is fertilized and the egg is laid at the two-cell stage of the embryo (Hubbard and Greenstein, 2005).

Transcription is absent from early regions of the germline until the 4-cell stage embryo (Seydoux *et al.* 1994, 1996, 1997). This suggested that the primary means of gene regulation in the *C. elegans* germline is post-transcriptional. Transgenic strains for 30 genes with reported germline expression patterns were generated to test this hypothesis (Merritt *et al.* 2008). The 3'UTR of each gene was cloned downstream of a GFP:Histone 2B (H2B) reporter driven by the germline ubiquitous *pie-1* promoter. Transgenic strains were generated via microparticle bombardment and randomly integrated into the genome. 17 distinct expression patterns were observed following examination of the resulting strains. Comparisons of the resulting expression patterns to those reported for the endogenous proteins showed concordance for 24/30 strains. The strains which did not match reported expression patterns were sperm-specific genes. In summary, the study demonstrated that 3'UTRs are sufficient for many germline expression patterns. Additional studies have identified many RBPs which regulate those, and other 3'UTRs. Studying *C. elegans* germline expressed

3'UTR-RBP interactions is an ideal system, however large scale analyses of *C. elegans* RBPs have not been performed.

How to assay a single 3'UTR against a library of RBPs?

The goal of this project was to develop an approach for testing a single 3'UTR against a library of *C. elegans* RBPs. The gene-centered approach had the following objectives:

1) Curate a compendium of *C. elegans* RBPs

The spectrum of *C. elegans* RBPs was not known. Cursory examinations of public databases and computational predictions identified 92 and 319 candidate RBPs indicating the total number of RBP was in the hundreds (Lublin and Evans 2007, Wang *et al.* 2009). There was a clear need for a dedicated analysis of the *C. elegans* proteome and literature similar to that performed for *C. elegans* transcription factors (TFs) (Reece-Hoyes *et al.* 2005). This database of *C. elegans* RBPs could then facilitate the construction of physical resources for gene-centered studies of 3'UTR interactions with many RBPs.

Curation of the *C. elegans* RBPs involved computational predictions of protein sequences coupled with manual annotation of RBPs with direct experimental evidence (Figure 1.5A). Known RNA binding domains were identified from literature and domain signatures were used to scan the *C. elegans* proteome for high scoring sequences. Additionally, homologs of known RBPs were identified. Lastly, RBPs were identified through a directed literature search of *C. elegans* and RNA binding proteins.

2) Develop an assay for detecting 3'UTR-RBP interactions in a gene-centered manner

The goal for this assay was to enable detection of physical interactions in a gene-centered manner. Additionally, interactions were to be detected in pair-wise and condition independent manners to remove complicating factors such as developmental or environmental states. Ideally, the method was designed in a scalable manner to enable future large scale screens. Lastly, emphases were placed on developing a cell-based system wherein the bait RNA was in a functional environment.

The assay was designed such that reporter mRNA activation by a heterologous RBP would result in high levels of signal (*i.e.* GFP fluorescence) in yeast (Figure 1.5B). The yeast based system enables the

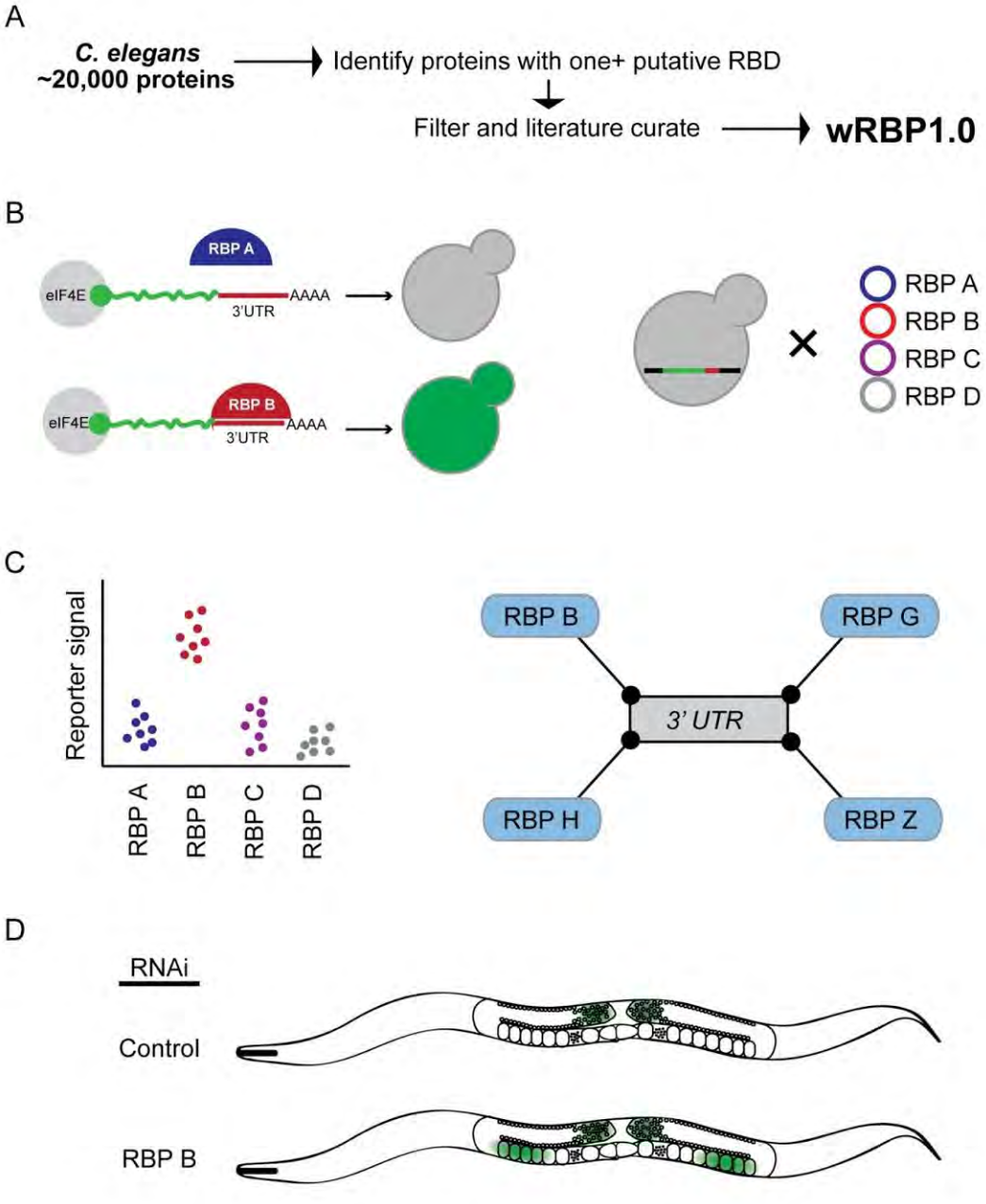

A
C. elegans
~20,000 proteins
Identify proteins with one+ putative RBD
Filter and literature curate
wRBP1.0
B
RBP A
eIF4E
AAAA
3'UTR
RBP B
eIF4E
AAAA
3'UTR
RBP A
RBP B
RBP C
RBP D
C
Reporter signal
RBP A
RBP B
RBP C
RBP D
RBP B
RBP G
3' UTR
RBP H
RBP Z
D
RNAi
Control
RBP B

Figure 1.5: Outline of a gene-centered approach to RNA-RBP interactaction mapping. A) The proteome of *C. elegans* was computationally analyzed to determine proteins sequences with one or more RNA binding domains. The computationally determined list was filtered manually to remove spurious predictions and to add literature curated RBPs that were missed. B) An assay was developed which enables the specific activation of the GFP reporter gene in yeast when an interaction occurs between an RNA and RBP. This assay is gene-centered, meaning that an individual RNA bait strain can be tested against a library of RBP preys. C) RNA-RBP interactions were tested in a pair-wise manner and those with increased signals were considered preliminary physical interactions. D) Assay defined interactions were tested *in vivo* using RNAi of the indicated RBP in a transgenic GFP : 3'UTR strain.

use of standard genetic techniques including integration of the bait-encoding DNA and subsequent transformation of the RBP prey library. This enables direct pair-wise testing of interactions in a condition independent manner. Additionally, the intact cellular environment places the interaction in a functional context. Lastly, yeast based assays can be scaled-up for genome level interaction screens (Simonis *et al.* 2009, Reece-Hoyes *et al.* 2013)

3) Test a single 3'UTR against a library of *C. elegans* RBPs

To demonstrate the effectiveness of the gene-centered approach a single 3'UTR was tested against a library of RBPs (Figure 1.5C). The ideal candidate for this is the *C. elegans glp-1* 3'UTR because there are several known RBPs which bind to and regulate it. Also, there may be additional unknown RBPs. To perform the test a subset of the curated, *C. elegans* germline expressed RBPs were tested in the assay. Comparisons were then made to known interactions and newly discovered interactions were tested using follow-up assays (Figure 1.5D).

CHAPTER II

Preface

This chapter comprises published work from the following reference:

Tamburino, A. M., Ryder, S. P., & Walhout, A. J. M. (2013). A Compendium of *Caenorhabditis elegans* RNA Binding Proteins Predicts Extensive Regulation at Multiple Levels. G3, 3, 297-304.

It is the product of my own work performed under the guidance of Marian Walhout and Sean Ryder.

A compendium of *C. elegans* RNA binding proteins predicts extensive regulation at multiple levels

Generating the right protein at the right place, the right time, and the right levels is critical during all aspects of life. Multiple levels of gene regulation coordinate the precise expression of genes throughout development and in response to environmental cues and insults. In genomics and systems biology, much attention has focused on the elucidation of regulatory networks involving transcription factors (TFs) or microRNAs (miRNAs) (Martinez and Walhout 2009; Arda and Walhout 2010). These networks include interactions in which these factors both regulate and are regulated by other molecules (Reece-Hoyes *et al.* 2011; Bartel 2009; Deplancke *et al.* 2006; Martinez *et al.* 2008; Harbison *et al.* 2004; Arda *et al.* 2010). RNA binding proteins (RBPs) are another important class of gene regulators; however, the regulatory networks in which they function remain largely uncharacterized.

Although TFs bind DNA and miRNAs interact with mRNAs, RBPs can interact with the entire spectrum of RNAs. These RNAs occur throughout the cell and can take on a vast array of functions, including serving as templates for protein synthesis (mRNA), participating as structural components of the splicing and translation machinery (rRNA, tRNA,

snRNA), and providing regulatory activity to modulate transcription, translation and chromatin structure (miRNA, siRNA, piRNA, lncRNA) (Lee and Schedl 2005, Steitz 2008; Moore and Proudfoot 2009; Carthew and Sontheimer 2009; Wahl *et al.* 2009). Physical interactions between RNA and RBPs are crucial to RNA regulation, for instance, to mediate precise mRNA 3' end formation, splicing, localization, stability, and translation. As a result of these physical interactions, RBPs can control transcript localization, levels, and translation (Shepard *et al.* 2003; Glisovic *et al.* 2008).

In contrast to RBPs, TFs are rapidly being characterized at a systems level using genome-scale methods such as chromatin immunoprecipitation (ChIP) and yeast one-hybrid assays (Walhout 2011). Among other findings these studies have demonstrated degenerate DNA binding of TFs, extensive combinatorial complexity of interactions between TFs and gene promoters, as well as both specific and promiscuous protein interactions between divergent members of the same TF family (Deplancke *et al.* 2006; Badis *et al.* 2009; Grove *et al.* 2009; Zinzen *et al.* 2009). The systems-level characterization of TFs has been greatly facilitated by high-confidence predictions of which genes in a genome encode such proteins (Reece-Hoyes *et al.* 2005; Kummerfeld and Teichmann 2006; Vaquerizas

et al. 2009). However, such compendia are not yet available for RBPs in multicellular model organisms.

Here, we present a compendium of predicted RBPs for the nematode *Caenorhabditis elegans* (wRBP1.0). We have used wRBP1.0 to begin the analysis of RBPs at a genome-wide level, using publicly available datasets. We found that RBP-encoding mRNAs have more alternative isoforms, longer 3' untranslated regions (UTRs), and more alternative polyadenylation (APA) sites than other mRNAs. In addition, RBP gene promoters interact with more TFs, RBP mRNAs are bound by more RBPs, and the 3'UTRs of RBP-encoding mRNAs are targeted by more miRNAs. Finally, RBPs are phosphorylated more frequently than other proteins. Together, our compendium and analyses provide a first step toward the characterization of RBP regulatory networks in *C. elegans* and serve as a model for the continued study of RBPs in other organisms, including humans.

RESULTS AND DISCUSSION

wRBP1.0

To curate the compendium of putative RBPs in *C. elegans*, we searched the proteome (version WS219) for each of 17 RBDs [see the section RNA binding domains (RBDs)] based on domain sequence signatures from the unified InterPro database (Quevillon *et al.* 2005; Hunter *et al.* 2009). Proteins were annotated for the presence of each domain using four separate databases (see Materials and Methods) and each protein possessing one or more RBD was included in the compendium. Low-confidence calls were removed (see Materials and Methods), and the curations were supplemented with RBPs that we identified from the literature but that were missed in the computational search. Of the total RBP set, 67% were identified by more than one method, which illustrates the robustness of our predictions (Figure 2.1A). Furthermore, the initial list contains greater than 93% of proteins that were previously curated as RNA binding (Wang *et al.* 2009), which illustrates the sensitivity of our method. It is important to note that we increased the number of putative *C. elegans* RBPs by almost threefold relative to this study (from 319 to 887). Two major reasons for this include the inclusion of additional RBDs and protein classes (*i.e.*, dsRBDs, ribosomal proteins, C2H2 zinc fingers, SAM domains) and the inclusion of additional RBPs possessing each domain (*i.e.*, 10–60% increase in KH, RRM, helicase, and CCCH zinc finger domain containing proteins). Further, 66% of the

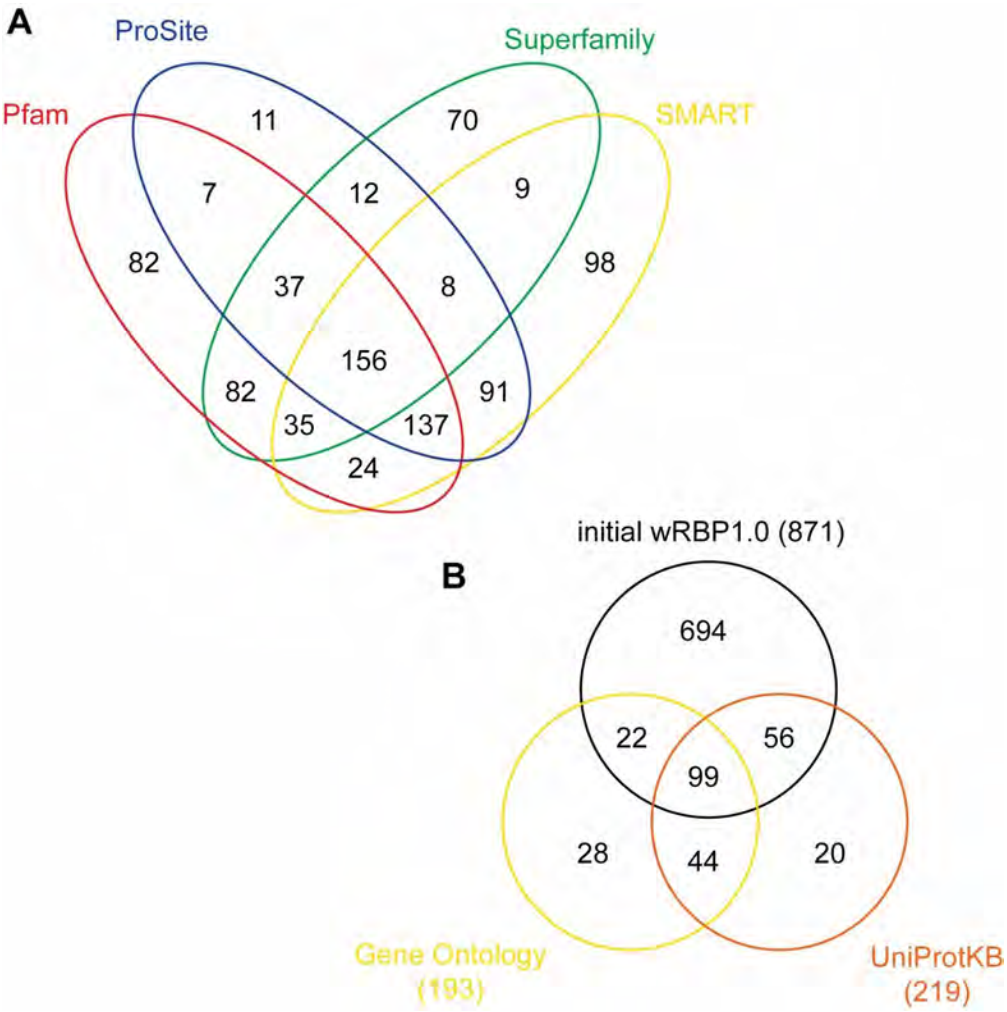


Figure 2.1: Venn diagrams of: A) Cross-validation of programs used by InteroProScan, and B) initial wRBP1.0 list together with Gene Ontology and UniProtKB listed RBPs

RBPs (177 of 269) annotated in Gene Ontology and UniProtKB databases as ‘RNA binding’ were included, again demonstrating high sensitivity (Figure 2.1B) (Gene Ontology Consortium 2000; Uniprot Consortium 2009). Next, we manually evaluated 96 RBPs that were not included in our initial list but that were annotated as RNA binding by Gene Ontology, UniProtKB or Wang *et al*. 2009. After careful consideration, we judged 16 of these to be candidate RBPs, whereas we did not have sufficient confidence to include the other 80 (data not shown). Finally, we determined that wRBP1.0 includes 220 of 230 protein listed in RBPDB (Cook *et al.* 2011) including 22 of 23 proteins with experimental evidence of RNA binding [AIN-1 is associated with the miRNA silencing complex but does not require RNA for binding (Wormbase.org)]. Altogether, this generated a final wRBP1.0 compendium of 887 genes. RBPs were then classified into Groups 1-4 based on the domains they possess (Figure 1, see below).

RNA binding domains (RBDs)

We identified a set of 17 RBDs by literature searches for proteins that bind to RNA (Figure 2.2, Table S2.1). Altogether, we identified 887 putative RBP-encoding genes (Table S2.2; see below). We divided these

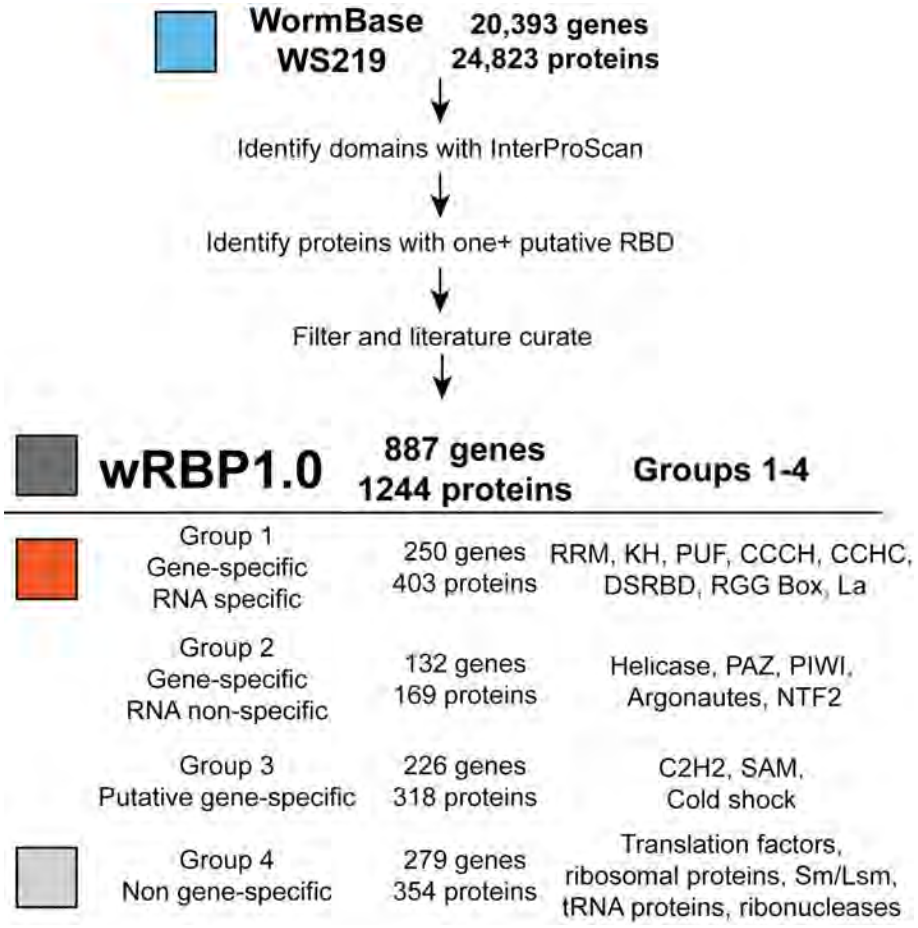


Figure 2.2: wRBP1.0. Pipeline for *C. elegans* RBP predictions. RBDs were predicted from WormBase protein annotations then filtered and literature curated. RBPs were separated into four groups according to their RBDs as indicated.

genes into four groups based on whether they are more likely to bind and regulate RNA in a gene-specific or nonspecific manner. Many RBPs contain multiple RBDs; however, only 10 of 887 genes contain domains from two or more different groups (Table S2.2). The classification of these 10 genes was first based on the presence of a sequence or structure-specific RBD.

Group 1: Gene-specific RBDs that bind RNA in a sequence-specific manner: This group contains eight RBDs that mediate binding to specific mRNAs in a sequence and/or structure-specific manner (Figure 2.2). RNA binding by these domains has been demonstrated for several individual proteins *in vitro*, and gene-specific binding has been detected for several proteins *in vivo* (Table S2.1) (Ryder *et al.* 2004; Bernstein *et al.* 2005; Opperman *et al.* 2005; Pagano *et al.* 2007; Farley *et al.* 2008; Pagano *et al.* 2009; Kershner and Kimble 2010; Wright *et al.* 2010). Direct, sequence-specific RNA binding has been shown for some *C. elegans* RBPs, but the vast majority remains untested. For instance, GLD-1 (KH domain) and FBF-1 both bind specific sequences *in vitro* (Ryder *et al.* 2004; Bernstein *et al.* 2005) and associate with specific mRNAs *in vivo* (Kershner and Kimble 2010; Wright *et al.* 2010; Jungkamp *et al.* 2011). Altogether, 250 of the 887 RBP-encoding genes are included in Group 1.

Group 2: Gene-specific RBDs that do not bind RNA in a sequence specific manner: RBDs within Group 2 bind RNA in a gene-specific manner *in vivo*. However, contrary to Group 1 RBDs, the means for this RNA binding specificity are unknown or occur in a manner that is not inherent to the RBD itself *(i.e.*, the domain contributes to RNA binding affinity rather than specificity). For instance, the argonautes ALG-1 and ALG-2 bind miRNAs through their PAZ/PIWI domains. Complementary base pairing by these miRNAs directs targeting of these proteins to specific mRNAs. Out of the 17 RBDs considered, four are placed in this group: helicase, PAZ, PIWI, and NTF2, altogether encoding 169 proteins.

Group 3: Putative gene-specific RBDs: Group 3 proteins are predicted to bind RNA in a gene- and sequence-specific manner. However, we have separated Group 3 proteins from those in Group 1 because their RBDs could be involved not only in RNA binding but also in DNA binding, or protein-protein interactions, thus making the prediction of their function ambiguous (see Table S1 for references). For instance, *Xenopus laevis* TFIIIA can bind both DNA and RNA through various combinations of its C2H2 zinc fingers (Theunissen *et al.* 1992; Lu *et al.* 2003). All proteins with the domains of group 3 are included although we expect that not all of them

will mediate RNA binding (*e.g.*, many C2H2 zinc fingers occur in TFs that bind DNA). Group 3 contains three of the 17 RBDs and 226 genes.

Group 4: Non-gene-specific RBPs, with some exceptions: The fourth group contains RBDs that typically do not bind RNA in a gene-specific manner. Many essential factors involved in general gene expression are in this group, including ribosomal proteins, transfer RNA-binding proteins, translation initiation factors, core splicing proteins and RNA degradation proteins such as ribonucleases and exosome components. Two of the 17 domains are included in this category and because many general RBPs lack clear domains, additional proteins are included based upon conservation to RBPs in other organisms. Altogether, this group contains 279 genes.

RBP-encoding genes are bound by more TFs, more RBPs, and have more splice variants

RBPs have been proposed to both fine tune gene expression as well as drive tissue and stage-specific gene expression (Blencowe 2006; Glisovic *et al.* 2008). Therefore, we hypothesized that RBPs may, as a group, be extensively regulated to mediate these functions. Here, we

tested this hypothesis using the wRBP1.0 compendium and several publicly available datasets.

Transcriptional regulation mediated by the binding of TFs to gene promoters provides a first and important level of regulation. There are 937 predicted TFs encoded by the *C. elegans* genome (Reece-Hoyes *et al.* 2005; Reece-Hoyes *et al.* 2011), and binding of 22 of these TFs (~2%) has been examined by ChIP-seq (Gerstein *et al.* 2010). Based on these data, we found that promoters of RBP genes are bound by more TFs than promoters of other genes (Figure 2.3A). Both gene-specific and general RBP promoters are bound by significantly more TFs ($p<1e-9$), indicating that transcriptional regulation is an important first step toward RBP expression. Importantly, these data were obtained using transgenic TF fusion strains. Because transgenes are often silenced in the germline (Cui and Han 2007) where many RBPs are expressed, it is possible that our analyses underestimate the enrichment. Further, this analysis was based on only 22 TFs; future studies will reveal the generality of our observation.

We next analyzed publicly available RBP-mRNA interactions. We obtained three RIP-Chip datasets for the *C. elegans* RBPs FBF-1, GLD-1 and RNP-8 (Kershner and Kimble 2010; Kim *et al.* 2010; Wright *et al.*

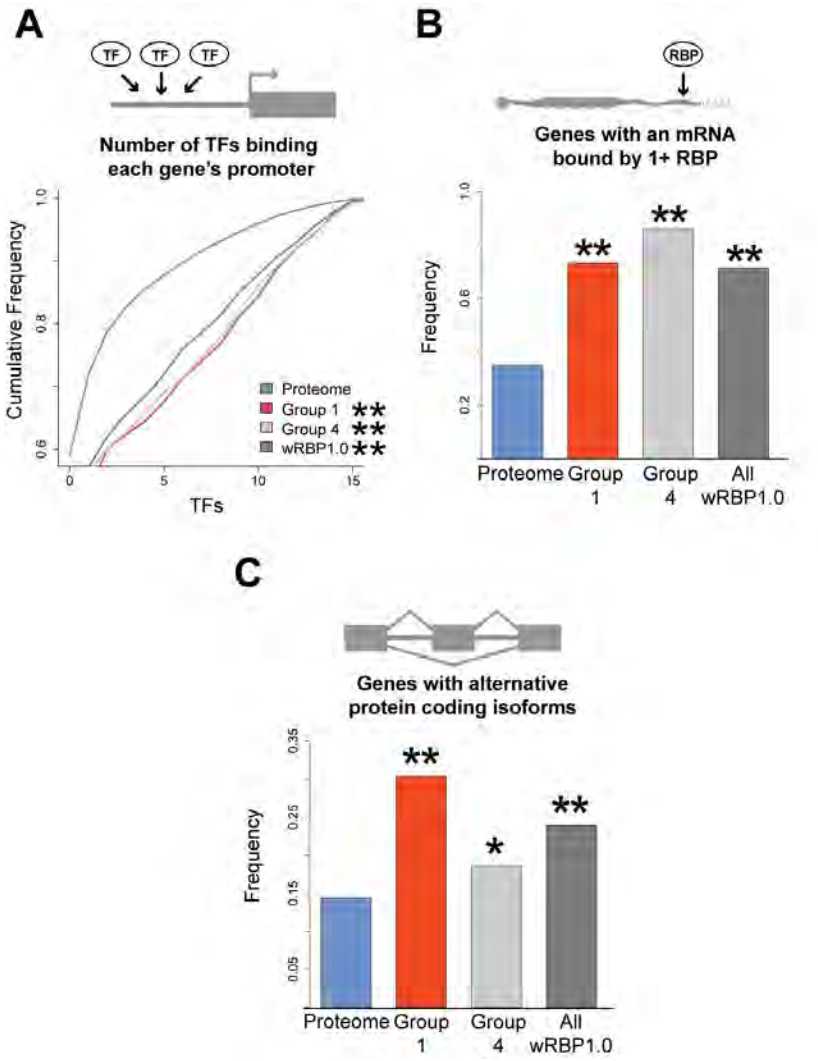


Figure 2.3: RBPs are extensively regulated by TFs and RBPs. A) More TFs bind to RBP promoters than the promoters of other genes B) RBPs bind to a higher proportion of RBP-encoding mRNAs. C) RBP genes are more frequently spliced than other genes. *p<0.05,**p<0.005, relative to proteome, hypergeometric test (frequency data), Komologorov-Smirnov test (cumulative frequency data).

2010) and found that 73% of RBP mRNAs are bound by at least one RBP, compared with only 35% of the total transcriptome (Figure 2.3B). The number of RBP mRNAs from Group 4 bound is even greater (86%). Our result is consistent with Gene Ontology enrichment analysis performed in the original studies that retrieved enrichment for 'RNA binding' and 'Nucleic acid binding' terms, respectively (Kim *et al.* 2010 and Kershner and Kimble 2010).

The binding of RBPs to mRNAs affects numerous steps of an mRNA's lifecycle, including alternative splicing (Blencowe 2006; Glisovic *et al.* 2008). To test whether *C. elegans* RBP-encoding mRNAs are more extensively spliced than other genes, we evaluated the number of protein isoforms per RBP-encoding gene by using comprehensive WormBase annotations. Approximately one-quarter of the 887 RBP-encoding genes (212; 23.9%) encode multiple isoforms, which is significantly more than the 14.4% of genes that undergo alternative splicing in the entire genome (Figure 2.3C). An even greater percentage of mRNAs encoding gene-specific RBPs in Group 1 are alternatively spliced (30.4%; Figure 2C). Through alternative splicing, the total number of RBPs increased by more than 40% (from 887 genes to 1242 proteins) and, interestingly, the number of distinct gene-specific RBPs increased by ~60% (250 genes

encoding 401 proteins). Thus, alternative splicing increases the effective number of RBPs in the *C. elegans* proteome.

RBP 3'UTRs are extensively regulated

3'UTRs affect gene expression via interactions with RBPs and miRNAs (Bartel 2009; Kuersten and Goodwin 2003). Concordantly, *C. elegans* 3'UTRs contain numerous conserved sequence elements that may interact with miRNAs or RBPs (Mangone *et al.* 2010; Jan *et al.* 2011). Using comprehensive 3'UTR annotations (www.UTRome.org), we found that RBP-encoding mRNAs have significantly longer 3'UTRs, with a median length of 156 nucleotides (nt), compared with 129 nt for the whole transcriptome (Figure 2.4A). The 3'UTRs of gene-specific RBP mRNAs (Group 1) are even longer (215 nt), whereas general RBPs have shorter 3'UTRs (Group 4; 100 nt).

Longer 3'UTRs can contain more regulatory sites, which implies that gene-specific RBPs may be more heavily regulated via their 3'UTRs, whereas general RBPs may be less extensively regulated. To test this, we first assessed the degree to which miRNAs target RBP 3'UTRs relative to all genes. In the absence of comprehensive experimental miRNA targeting data, predictions for bound target mRNAs can be made using

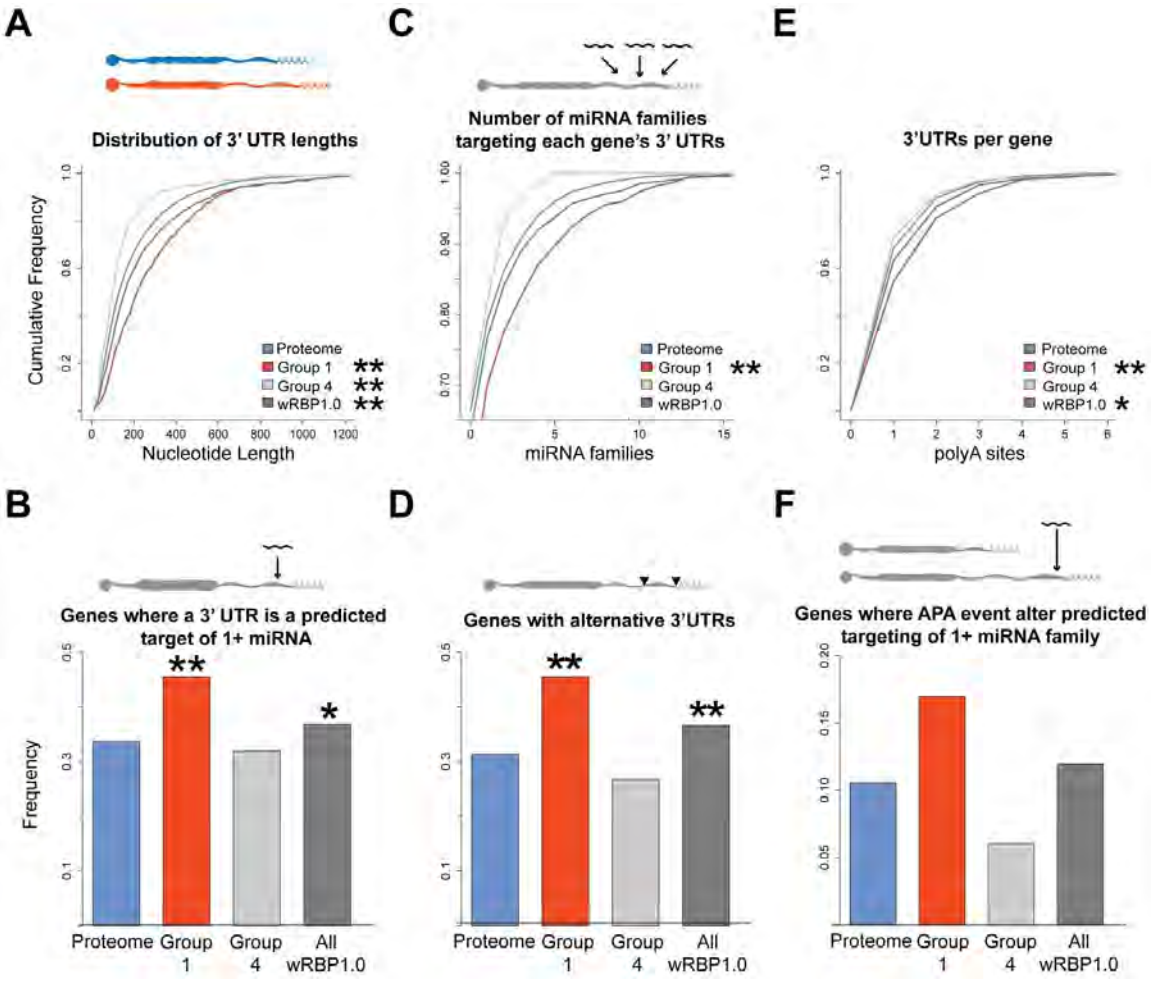

A
Distribution of 3' UTR lengths
Cumulative Frequency
Nucleotide Length
Proteome
Group 1 **
Group 4 **
wRBP1.0 **
C
Number of miRNA families targeting each gene's 3' UTRs
miRNA families
Proteome
Group 1 **
Group 4
wRBP1.0
E
3'UTRs per gene
polyA sites
Proteome
Group 1 **
Group 4
wRBP1.0 *
B
Genes where a 3' UTR is a predicted target of 1+ miRNA
Frequency
Proteome Group 1 Group 4 All wRBP1.0
D
Genes with alternative 3'UTRs
Proteome Group 1 Group 4 All wRBP1.0
F
Genes where APA event alter predicted targeting of 1+ miRNA family
Proteome Group 1 Group 4 All wRBP1.0

Figure 2.4: RBPs are extensively regulated through 3'UTRs A) RBP transcripts have longer 3'UTRs. B) RBP 3'UTRs are more heavily targeted by miRNAs. C) More miRNA families target RBP 3'UTRs. D) 3'UTR annotations show that more RBPs utilize alternative 3'UTRs, and E) that RBP genes have more alternative 3'UTRs. F) Combined miRNA target predictions and 3'UTR annotations reveal that APA affects predicted miRNA targeting. *$p<0.05$,**$p<0.005$, relative to proteome, hypergeometric test (frequency data), Komologorov-Smirnov test (cumulative frequency data).

the miRNA seed sequences (Bartel 2009). We used target predictions from TargetScan for all *C. elegans* 3'UTR sequences experimentally determined by 3P-Sequencing (3Pseq) (Jan *et al.* 2011). TargetScan predicts miRNA targets based upon stringent seed pairing as well as site number, type, context, and conservation (Bartel 2009). Comparison of RBP-encoding mRNA 3'UTRs to the 3'UTRs of all *C. elegans* mRNAs revealed that significantly more RBP 3'UTRs are predicted targets of miRNAs (Figure 2.4B). Furthermore, significantly more miRNA families target each gene-specific RBP 3'UTR compared with all 3'UTRs, indicating a potential for increased combinatorial complexity (Figure 2.4C). In contrast, general RBPs showed no significant difference in miRNA targeting compared to the total transcriptome.

It is important to note that these predictions are based on conservation of the site in multiple species and availability of the site in folded RNA. This implies that the increased number of miRNA families targeting 3'UTRs is not solely a consequence of 3'UTR length. To confirm this, we compared RBP 3'UTRs with similar length 3'UTRs from the total transcriptome by binning 3'UTRs by length (Figure 2.5). This analysis confirmed that, among the shortest 3'UTRs (*i.e.*, the first two quartiles), more miRNAs are indeed predicted to target RBP 3'UTRs, while we did not observe a difference for the longest 3'UTRs.

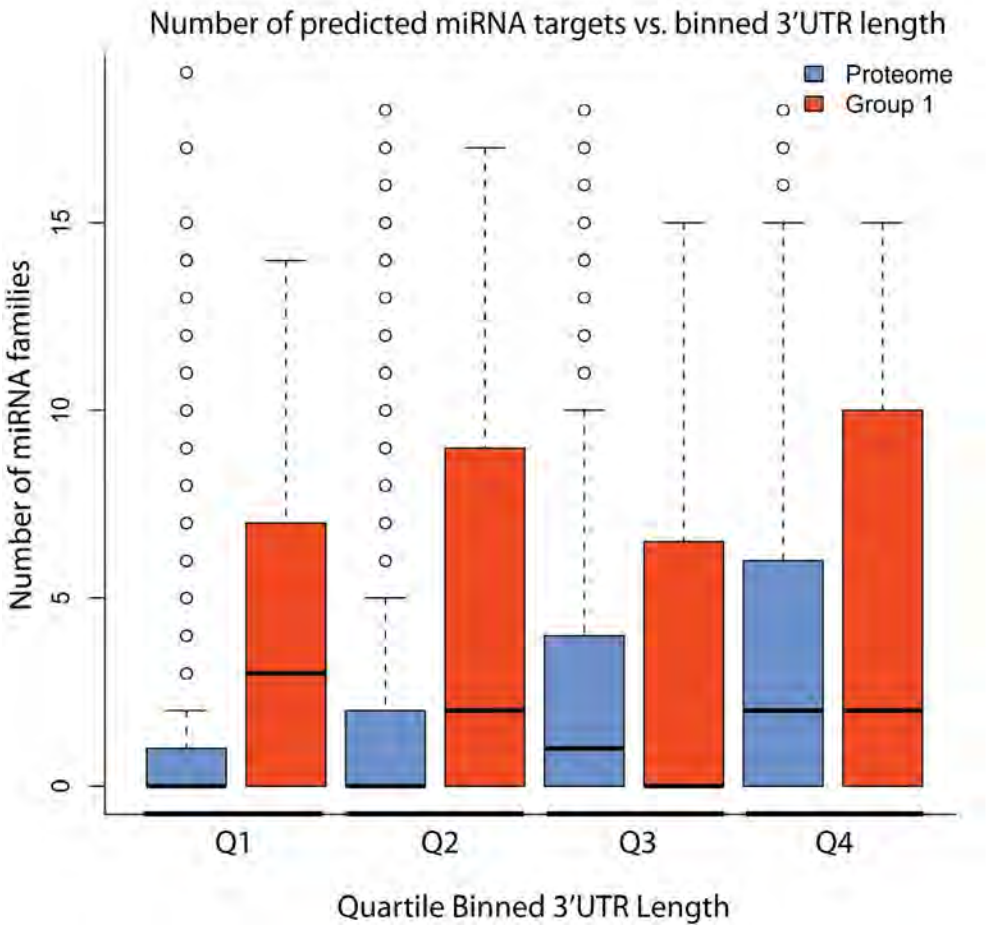


Figure 2.5: Quartile binned boxplots of miRNAs targeting RBP 3'UTRs *vs.* 3'UTRome. Quartiles were determined using the total 3'UTRome. The distribution of 3'UTR lengths are shown for each quartile.

We further evaluated miRNA targeting to RBP 3'UTRs using predictions made by mirWIP (Hammell *et al.* 2008) and argonaute ALG-1 bound 3'UTRs determined using cross-link immunoprecipitation (Figure 2.6) (Zisoulis *et al.* 2010). Both of these analyses showed that RBP 3'UTRs are indeed more frequently targeted by miRNAs, which further supports the observations made with TargetScan predictions.

Alternative 3'UTR usage provides additional unique sites of regulation for miRNAs and RBPs or, conversely, can eliminate regulatory sites for these same factors. Recently, it has been shown that shortening of 3'UTRs by alternative polyadenylation (APA) alters protein expression in proliferating cells, an effect partly attributed to the loss of miRNA binding sites (Sandberg *et al.* 2008; Mayr and Bartel 2009). Using 3'UTR annotations determined by 3P-Seq (Jan *et al.* 2011), we found that more RBPs use APA and that RBPs possess more distinct 3'UTRs than the total transcriptome (Figures 2.4D,E; results with 3'UTRome annotations were consistent, data not shown). Once again, the effect was especially pronounced for gene-specific RBPs (Group 1). We calculated the number of genes in which APA eliminates all predicted targeting sites for one or more miRNA family, thereby preventing miRNA repression and increasing gene expression. Using 3P-seq-derived 3'UTRs and TargetScan miRNA target predictions, we found that more than 15% of the gene-specific

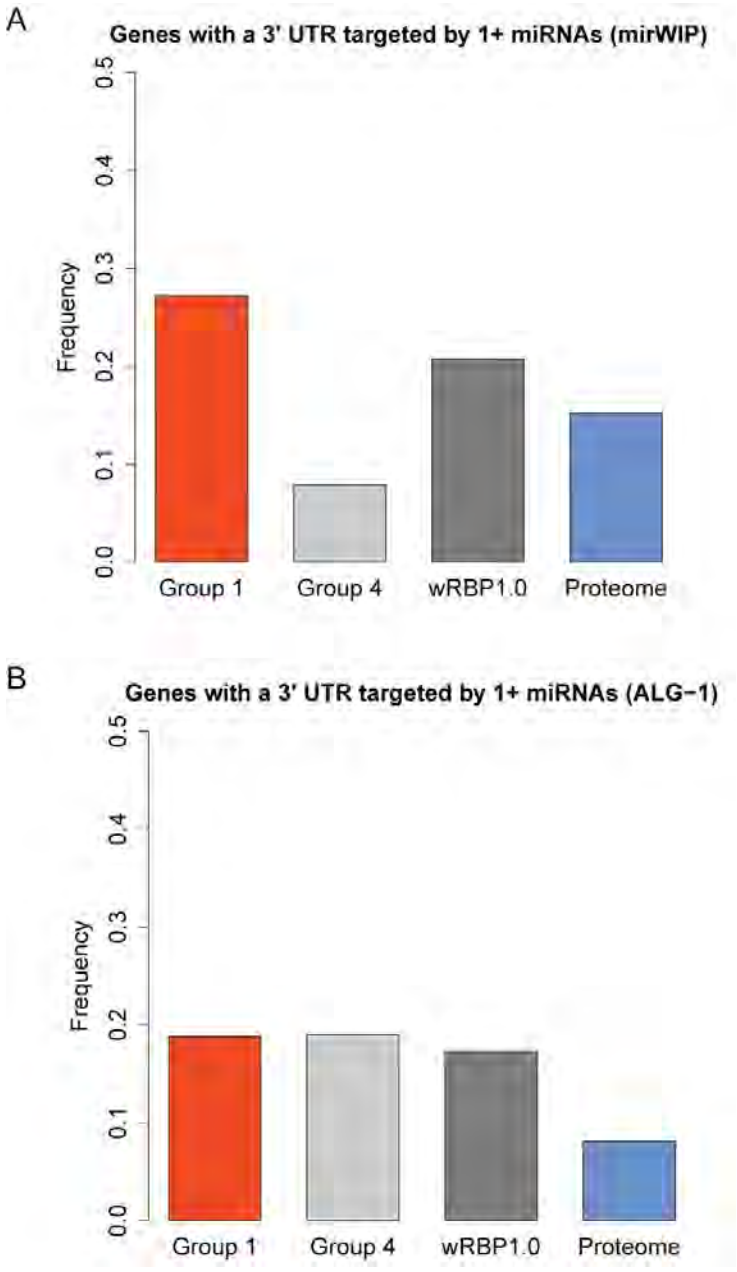


Figure 2.6: miRNA targeting. Frequency of 3'UTRs targeted by miRNAs according to A) mirWIP predictions and B) ALG-1 IP binding data

RBPs could evade potential repression by at least one miRNA family using APA, a fraction that is more than twice that of the total transcriptome (Figure 2.4F). The predicted effects of APA may also affect gene expression through the distinct binding of RBPs to alternate 3'UTRs.

RBPs are more extensively phosphorylated

Posttranslational modifications provide another mechanism to create protein diversity. In particular, phosphorylation can affect the ability of proteins to function and/or interact with binding partners (Deribe *et al.* 2010). To evaluate the degree to which RBPs are phosphorylated, we interrogated phosphoproteome data that were obtained by tandem mass spectrometry of synchronized adult worms and that identified 6,780 phosphorylation sites on 2,373 proteins (Zielinska *et al.* 2009). Because many factors can affect the ability for certain proteins to be detected in mass spectrometry, we corrected for potential biases by normalizing the frequency of detected RBPs in each group by a separate mass spectrometry study that analyzed the proteome of mixed stage worms and did not enrich for phophopeptides (Figure 2.7) (Merrihew *et al.* 2008). We found that more gene-specific RBPs are phosphorylated relative to the

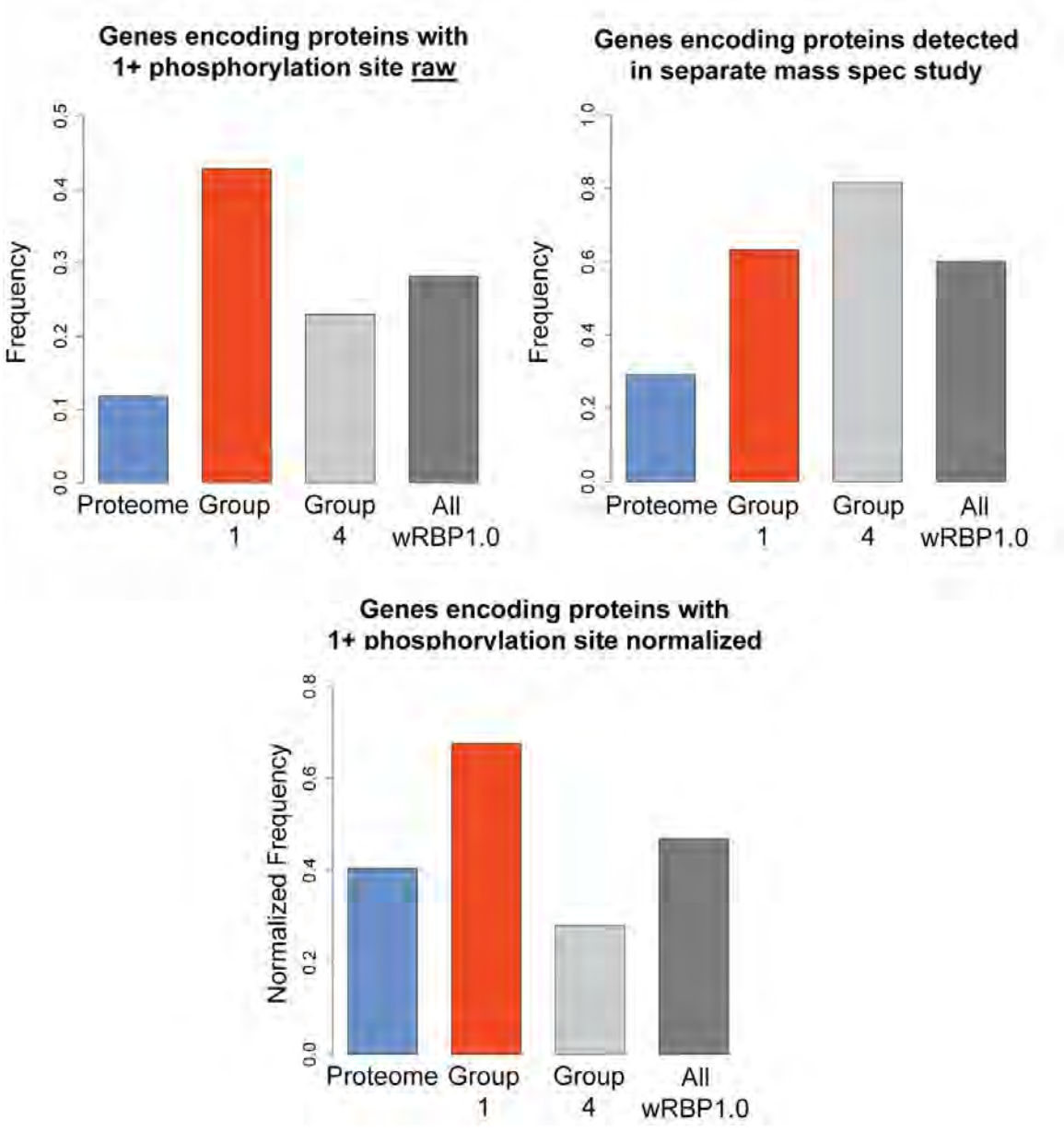


Figure 2.7: Normalization of proteomic data

entire proteome (Figure 2.8A). Furthermore, gene-specific RBPs (Group 1) have significantly more phosphorylation sites per protein than the total proteome (Figure 2.8B). In contrast, general RBPs (Group 4) are less frequently phosphorylated, although this group still contains more phosphorylation sites than entire proteome. This finding confirms the enrichment for the Gene Ontology term 'RNA binding' in the mass spectrometry dataset (Zielinska *et al.* 2009). The increased level of RBP phosphorylation further indicates that RBPs are indeed a heavily regulated class of cellular regulators.

Comparison of gene-specific RBPs (Group 1) with TFs

Group 1 RBPs are conceptually analogous to TFs in that they are predicted to bind to and regulate genes in a specific manner. Thus, we compared the characteristics of gene-specific RBPs in Group 1 to those of TFs. Although RBPs and TFs both have more isoforms than the general proteome, RBPs have significantly more isoforms compared with TFs (Figure 2.9A). This finding is interesting because RBPs often contain multiple RBDs that are differentially included in different isoforms, whereas most *C. elegans* TFs have only one DNA binding domain (Reece-Hoyes *et al.* 2005). There are more TFs bound per RBP

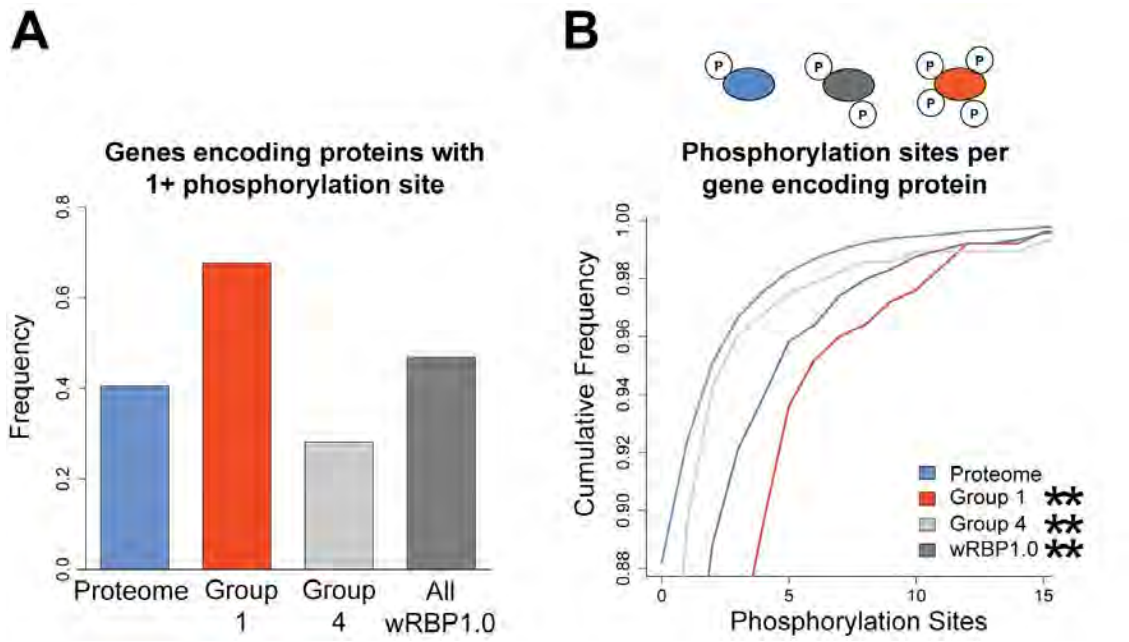


Figure 2.8: RBPs are extensively regulated post-translationally. A) More RBPs are phosphorylated. B) RBPs have more phosphorylated residues per protein. *p<0.05,**p<0.005, relative to proteome, hypergeometric test (frequency data), Komologorov-Smirnov test (cumulative frequency data).

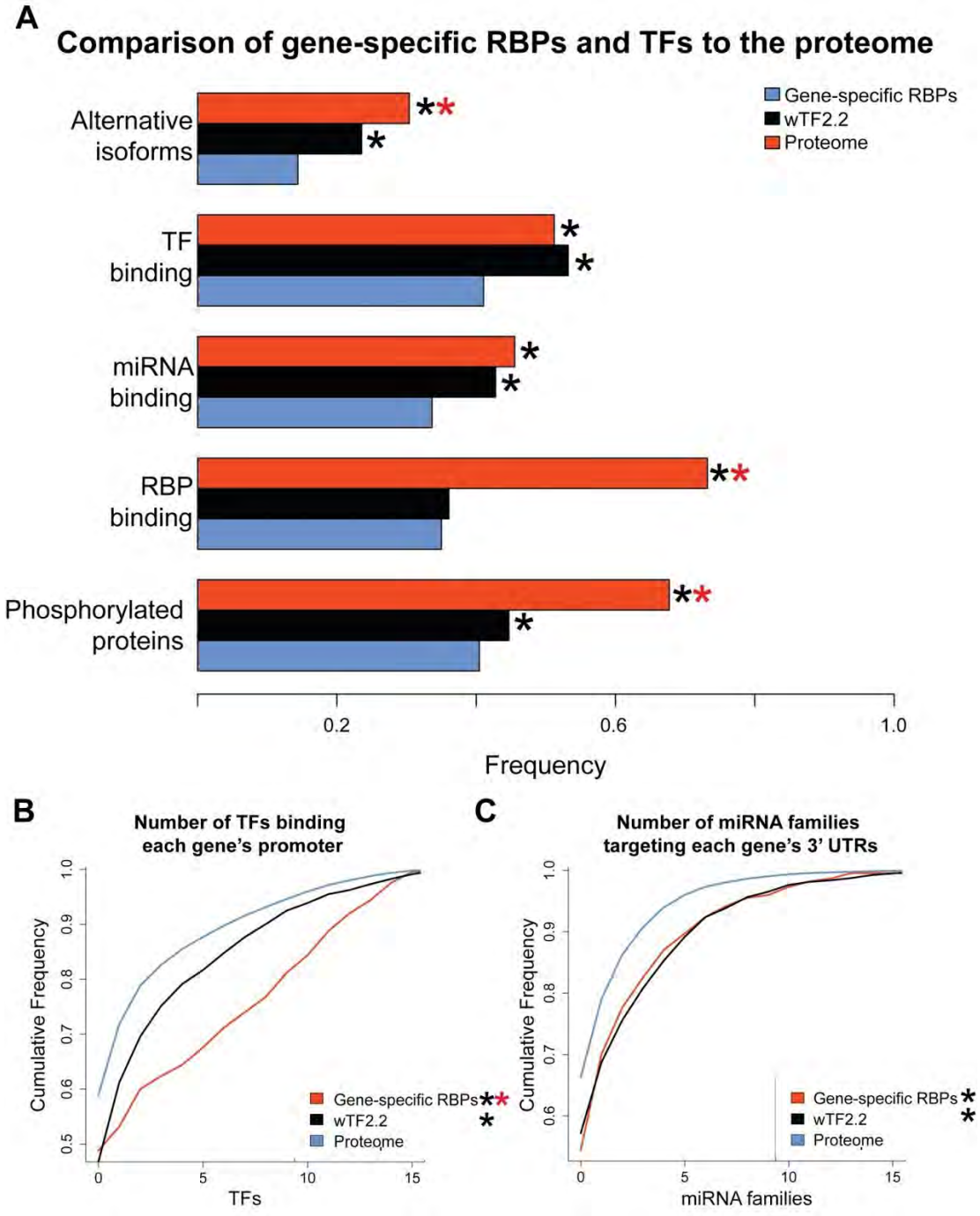

A
Comparison of gene-specific RBPs and TFs to the proteome
Gene-specific RBPs
wTF2.2
Proteome
Alternative isoforms
TF binding
miRNA binding
RBP binding
Phosphorylated proteins
0.2
0.6
1.0
Frequency
B
Number of TFs binding each gene's promoter
Cumulative Frequency
TFs
C
Number of miRNA families targeting each gene's 3' UTRs
miRNA families

Figure 2.9: Comparison of gene-specific RBPs (Group 1) to TFs. A) Comparison of alternative isoforms, TF binding, miRNA targeting, RBP binding and phosphorylation. B) Gene-specific RBPs have more TFs bound to promoters relative to TF genes. C) Gene-specific RBPs and TFs have 3'UTRs targeted by more miRNA families. *$p<0.005$, relative to proteome, **$p<0.005$ relative to wTF2.2, hypergeometric test (frequency data), Komologorov-Smirnov test (cumulative frequency data).

promoter than per TF promoter, which indicates that there may be more combinatorial complexity in the transcriptional regulation of RBP genes, or in the generation of tissue-specific gene expression patterns (Figure 2.9B). Both RBPs and TFs are predicted to have 3'UTRs that are more frequently targeted by miRNAs, and there is no difference between the numbers of distinct miRNA families that target their 3'UTRs (Figure 2.9A, C). However, there was a large, significant difference involving binding by RBPs: RBP-encoding mRNAs are more frequently bound by RBPs than TF mRNAs and mRNAs in general (Figure 2.9A). This difference could be attributed to an expression bias since RBP mRNAs are enriched in the germline (Wang *et al.* 2009) and should therefore be reevaluated after the determination of additional RBP-mRNA interaction data, including that of RBPs expressed in the soma. Finally, phosphorylation of both RBPs and TFs is significantly enriched relative to the proteome, but RBPs are more extensively phosphorylated than TFs. Taken together, both types of regulators are extensively regulated.

Conclusions

We present wRBP1.0: a comprehensive compendium of *C. elegans* RBPs. As has been demonstrated for the *C. elegans* TF compendium

(Reece-Hoyes *et al.* 2005, 2007; Grove *et al.* 2009; Reece-Hoyes *et al.* 2011) we expect that wRBP1.0 will be an invaluable resource for the creation of ORF-based clone collections, the delineation of RBP expression patterns, and RBP regulatory networks. Using wRBP1.0 and several publicly available genomic, transcriptomic and proteomic datasets, we found that RBPs are extensively regulated at each level. A question that remains is why an organism extensively regulates its RBPs. One attractive possibility is that individual RBPs mediate precise gene regulation under different developmental or environmental conditions or in distinct cells or tissues within the animal. Such diverse functionalities could potentially be greatly facilitated by a combination and layering of the different transcriptional and posttranscriptional regulatory mechanisms. Furthermore, it is likely beneficial to the animal to be able to rapidly decrease the level or activity of different RBPs, such that downstream target gene expression can change rapidly as well. Many of the regulatory trends we observed are more pronounced for gene-specific RBPs, *i.e.*, those we predict to function analogously to TFs. There are nearly four times more genes predicted to encode TFs than gene-specific RBPs in the *C. elegans* genome (937 vs. 251) (Reece-Hoyes *et al.* 2011; this study). Strikingly, however, gene-specific RBPs have more alternative isoforms and are more extensively phosphorylated than TFs. This finding could

suggest that despite fewer gene-specific RBP genes than TF genes in the *C. elegans* genome, regulatory mechanisms can increase the repertoire of RBPs, thereby diversifying their regulatory capacity. Related analyses have been performed in the unicellular eukaryote *Saccharomyces cerevisiae* (Mittal *et al.* 2009, 2011). Using a list of putative RBPs (Hogan *et al.* 2008), RBP mRNAs were shown to have shorter half-lives, greater abundance, and greater ribosome occupancy (Mittal *et al.* 2009). Additionally, it was shown that RBPs are more abundant, have longer half-lives, and decreased noise (Mittal *et al.* 2009). These trends were more pronounced for ribosomal RBPs and for RBPs with high connectivity, as defined by interaction data. Combined with complementary analyses in this study it is clear that RBPs exhibit properties distinct from the total transcriptome/proteome. It also is evident that gene-specific/low connectivity RBPs exhibit properties distinct from non gene-specific/high connectivity RBPs. Altogether, wRBP1.0 provides a starting point for the generation of RBP clone resources that can be used in system-level characterization of posttranscriptional regulatory networks, as well as a first step in the analysis of the regulation of this important class of proteins.

Materials and methods

wRBP1.0 was curated by computationally predicting RBD-containing proteins in the *C. elegans* proteome (WS219). A FASTA file containing the amino acid sequences of all protein coding isoforms in the WS219 release was downloaded from WormBase (http://wormbase.org/). This file was analyzed using a locally installed Unix version of the InterProScan software (iprscan v4.6; InterPro release 24.0; accessed September, 15, 2010) (Quevillon *et al.* 2005; Hunter *et al.* 2009) using default settings. Iprscan takes the amino acid sequence of each protein as its input and then uses several different applications to search specific databases of domain signatures. The output of iprscan is all recognizable protein domains in that protein sequence. The results were filtered to include only those domains that were identified by Pfam, SMART, Superfamily or ProSite (Punta *et al.* 2012; Letunic *et al.* 2012; Wilson *et al.* 2009; Sigrist *et al.* 2010) because these applications were most effective at detecting RBDs (data not shown). Results were then manually filtered to include only those proteins that possess one or more of 17 RBDs (Table S2.1). Of note is the RGG box, an RBD that was not included as an independent entry by any of the tools used (although it is contained within the specific Pfam domain definition FXR1P_C which encompasses two RRM domains and one RGG box). While these domains are known RBDs (Kiledjian *et al.* 1992), their

sequence and structural determinants have not been well defined. We therefore only included RGG box proteins in our list that have been implicated in the literature as RNA binding. The list was manually checked to verify the presence of known *C. elegans* RBPs and to eliminate false positive predictions, as enumerated below. Five proteins were removed from wRBP1.0 because the computationally predicted domains were much longer or shorter than known RBDs, and we were not confident in their predictions as RBDs based upon visual assessment (R12B2.5, T03G11.3, D2005.1, Y82E9BR.19, and R11H6.5). One protein was removed because it is currently annotated as a pseudogene in Wormbase (C06A1.4). Two proteins were removed due to the lack of characteristic zinc finger homology (Y60A9.3, R03D7.7). 12 RBPs were added based upon published reports that demonstrated or strongly predicted RNA binding (C18G1.4, C50E10.4, M04B2.1, R06F6.1, R144.7, T12F5.5, Y18D10A.17, Y48G8AL.6, Y53C12B.3, ZK1127.1, ZK1236.3, ZK381.4). Sixteen RBPs were added following secondary searches of genes annotated as 'RNA-binding' according to Gene Ontology, UniProtKB, or Wang *et al.* 2009 (Wang *et al.* 2009; Gene Ontology Consortium, 2000; UniProt Consortium, 2009) based upon manual inspection of all 96 RBPs using information found on Wormbase.org. Wormbase indentified several proteins with non-canonical domains including cytidine deaminases (C47D2.2,F49E8.4),

translation initation factors (T01C3.7, F53A2.6, R04A9.4, C05D9.5, Y57A10A.30) tRNA binding proteins (C41G7.1, F29C4.6, C49H3.10) and additional general factors (C12D8.11, C41G7.1, F29C4.6, C49H3.10, C11D2.7, C15C6.4, C48B6.2, F08B4.7) that were missed in our initial screen.

Genome-wide datasets were downloaded from their respective databases or publications. TF binding data was obtained from (Gerstein *et al.* 2010). RIP-Chip data for three RBPs was obtained from (Wright *et al.* 2010; Kershner *et al.* 2010; Kim *et al.* 2010). 3'UTRs were from 3'UTRome annotations, kindly provided by Marco Mangone. These annotations are reflective of two independent large scale datasets (Mangone *et al.* 2010; Jan *et al.* 2011). TargetScan miRNA target predictions were downloaded from http://www.targetscan.org/worm_52/ based on predictions that corresponded to 3'UTRs determined using 3P-Seq (Jan *et al.* 2011). mirWIP target predictions (Hammell *et al.* 2008) were kindly provided by Molly Hammell. ALG-1 targets were downloaded from the UCSC genome browser using intersection of the ALG-1 binding sites (Zisoulis *et al.* 2010) with a custom track composed of the aforementioned 3'UTRome annotations. Protein phosphorylation sites from synchronized adult worms were obtained from (Zielinska *et al.* 2009). All data were compiled into a local database. The number of alternative isoforms was defined as the

number of distinct proteins encoded by a single gene according to WormBase annotations. TF and RBP binding events as determined by ChIP-Seq and RIP-Chip were assigned to their respective genes according to the original publications. The number of miRNAs predicted to target each gene was defined as the number of unique miRNA families with one or more conserved sites predicted in any of the gene's 3'UTRs. miRNA targeting was defined to affect a gene when 1+ miRNA target site was gained/lost in an alternative 3'UTR. miRNA predictions are based on data from (Jan *et al*. 2011) and therefore the analyses of alternative 3'UTRs and their effects on miRNA targeting were based upon the same dataset. The number of post-translational modifications per protein was calculated from the number of unique residues that were phosphorylated. The number of binding events or post-translational modifications was calculated for each RNA/protein isoform and then combined non-redundantly for each gene. Hypergeometric and Komolgorov-Smirnov tests were performed using R project software (R Development Core Team, 2009).

Table S2.1 RBDs, groups and examples

Domain	**Abbrev.**	**Group**	***C.elegans* examples**	**Example protein homologs**
RNA recognition motif	RRM	1	FOX-1 RNP-8	*H. sapiens* PABP, hnRNP C
K homology (KH)	KH	1	MEX-3 GLD-1	*H. sapiens* hnRNP K
Pumilio/FBF	PUF	1	FBF-1/2	*D. melanogaster* Pumilio
CCCH zinc finger	ZF_CCCH	1	POS-1 MEX-5	*H. sapiens* TTP
CCHC zinc finger	ZF_CCHC	1	LIN-28	*H. sapiens* Lin28
Double stranded RNA binding domain	DSRBD	1	ADR-1, -2	*H. sapiens* Adar1, Adar2
RGG box	RGG	1	PGL-1	*H. sapiens* hnRNP U, FMRP
La	La	1	LARP-1 LARP-5	*H. sapiens* La
RNA Helicase	HEL	2	GLH-1	*D. melanogaster* VAS
PAZ, PIWI, Argonautes	PAZ, PIWI	2	ALG-1 CSR-1	*D. melanogaster* PIWI
Nuclear transport factor 2	NTF2	2	NXF-2	

C2H2 zinc finger	ZF_C2H2	3	MEP-1	*X. laevis* TFIIIa
Sterile alpha motif	SAM	3	BCC-1	*D. melanogaster* Smaug, *S. cerevisiae* Vts1
Cold shock	CSD	3	CEY-1 LIN-28	*H. sapiens* Unr
General factors: Translation factors, tRNA proteins, Ribosomal proteins, Ribonucleases	GF	4	IFE-4 EXO-1	*H. sapiens* eIF4G, *S. cerevisiae* Ccr4, Xrn1
Sm/Lsm	Sm	4	CAR-1	*S. cerevisiae* Lsm1, Lsm2

Table S2.2 wRBP1.0

Gene name	ORF	RBD	Group
aars-1	W02B12.6	GF	4
aars-2	F28H1.3	GF	4
acin-1	C43E11.1	RRM(1)	1
adr-1	H15N14.1	DSRBD(2)	1
adr-2	T20H4.4	DSRBD(1)	1
alg-1	F48F7.1	PAZ(1)PIWI(1)	2
alg-2	T07D3.7	PAZ(1)PIWI(1)	2
aly-1	C01F6.5	RRM(1)	1
aly-2	F23B2.6	RRM(1)	2
aly-3	M18.7	RRM(1)	1
arc-1	ZK1320.6	ZF_C2H2(1)	3
asd-1	R74.5	RRM(1)	1
asd-2	T21G5.5	KH(1)	1
bcc-1	M7.3	KH(1)SAM(1)	1
blmp-1	F25D7.3	ZF_C2H2(5)	3
btf-1	F15D4.1	HEL(2)	2
car-1	Y18D10A.17	Sm	2
cars-1	Y23H5A.7	GF	4
ccch-1	F38B7.1	ZF_CCCH(2)	1
ccch-2	F38C2.5	ZF_CCCH(2)	1
ccch-3	W05B10.2	ZF_CCCH(1)	1

ccch-5	Y57G11C.25	ZF_CCCH(2)	1
cdd-1	C47D2.2	Unknown	2
cdd-2	F49E8.4	Unknown	2
cdl-1	R06F6.1	Unknown	1
cel-1	C03D6.3	GF	4
ces-1	F43G9.11	ZF_C2H2(5)	3
cey-1	F33A8.3	CSD(1)	3
cey-2	F46F11.2	CSD(1)	3
cey-3	M01E11.5	CSD(1)	3
cey-4	Y39A1C.3	CSD(1)	3
cfim-2	D1046.1	RRM(1)	1
cgh-1	C07H6.5	HEL(2)	2
chd-3	T14G8.1	HEL(2)	2
che-1	C55B7.12	ZF_C2H2(4)	3
cid-1	K10D2.3	ZF_C2H2(1)ZF_CCHC(1)	3
cpb-1	C40H1.1	RRM(2)	1
cpb-2	C30B5.3	RRM(2)	1
cpb-3	B0414.5	RRM(2)	1
cpf-2	F56A8.6	RRM(1)	1
cps-6	C41D11.8	NUCL	4
cpsf-4	F11A10.8	ZF_CCCH(5)ZF_CCHC(2)	1
crn-1	Y47G6A.8	NUCL	4

crn-3	C14A4.4	NUCL	4
crn-4	AH9.2	NUCL	4
crn-5	C14A4.5	NUCL	4
csb-1	F53H4.1	HEL(2)	2
csr-1	F20D12.1	PAZ(1)PIWI(1)	2
cyn-13	Y116A8C.34	RRM(1)	1
dars-1	B0464.1	tRNA	4
dars-2	F10C2.6	tRNA	4
daz-1	F56D1.7	RRM(1)	1
dcr-1	K12H4.8	DSRBD(2)HEL(2)PAZ(1)NUCL(1)	1
dct-10	Y38H6C.5	ZF_CCHC(1)	1
dct-13	Y116A8C.17	ZF_CCCH(2)	1
ddx-19	T07D4.4	HEL(2)	2
ddx-23	F01F1.7	HEL(2)	2
die-1	C18D1.1	ZF_C2H2(4)	3
din-1	F07A11.6	RRM(2)	1
dnj-10	F22B7.5	ZF_CCHC(2)	1
dnj-17	T03F6.2	ZF_C2H2(2)	3
dog-1	F33H2.1	HEL(2)	2
dpy-20	T22B3.1	ZF_C2H2(2)	3
drh-1	F15B10.2	HEL(2)	2
drh-3	D2005.5	HEL(2)	2

drr-2	T12D8.2	RRM(1)	1
drsh-1	F26E4.10	DSRBD(1)NUCL(1)	1
ears-1	ZC434.5	GF	4
edc-3	R05D11.8	Sm/Lsm	4
eea-1	T10G3.5	ZF_C2H2(1)	3
egal-1	C10G6.1	NUCL	4
egl-43	R53.3	ZF_C2H2(5)	3
egl-46	K11G9.4	ZF_C2H2(3)	3
egrh-1	C27C12.2	ZF_C2H2(3)	3
egrh-2	Y55F3AM.7	ZF_C2H2(1)	3
egrh-3	Y94H6A.11	ZF_C2H2(3)	3
ehn-3	ZK616.10	ZF_C2H2(6)	3
eif-1.A	H06H21.3	GF	4
eif-3.G	F22B5.2	RRM(1)ZF_CCHC(1)	1
emb-5	T04A8.14	GF	4
ent-3	K02E11.1	GF	4
eor-1	R11E3.6	ZF_C2H2(9)	3
ergo-1	R09A1.1	PAZ(1)PIWI(1)	2
eri-1	T07A9.5	NUCL	4
eri-7	C41D11.7	HEL(1)	2
etr-1	T01D1.2	RRM(3)	1
exc-7	F35H8.5	RRM(3)	1

exo-1	F45G2.3	NUCL	4
exos-1	Y48A6B.5	NUCL	4
exos-2	Y73B6BL.3	GF	4
exos-3	F59C6.4	GF	4
exos-4.1	B0564.1	NUCL	4
exos-4.2	Y6D11A.1	GF	4
exos-7	F31D4.1	GF	4
exos-8	F41G3.14	GF	4
exos-9	F37C12.13	NUCL	4
faah-5	Y56A3A.5	RRM(1)	1
fars-1	T08B2.9	GF	4
fars-2	Y60A3A.13	GF	4
fars-3	F22B5.9	GF	4
fbf-1	H12I13.4	PUF(8)	1
fbf-2	F21H12.5	PUF(8)	1
fib-1	T01C3.7	GF	4
fkb-8	Y18D10A.25	ZF_C2H2(1)	3
fkh-7	F26D12.1	ZF_C2H2(1)	3
fog-1	Y54E10A.4	RRM(1)	1
fox-1	T07D1.4	RRM(1)	1
fozi-1	K01B6.1	ZF_C2H2(1)	3
fust-1	C27H5.3	RRM(1)	1
gars-1	T10F2.1	GF	4

gen-1	T12A2.8	NUCL	4
gla-3	T02E1.3	ZF_CCCH(2)	1
gld-1	T23G11.3	KH(1)	1
gld-3	T07F8.3	KH(2)	1
glh-1	T21G5.3	HEL(2)ZF_CCHC(4)	1
glh-2	C55B7.1	HEL(2)ZF_CCHC(6)	1
glh-3	B0414.6	HEL(2)ZF_CCHC(2)	1
glh-4	T12F5.3	HEL(2)ZF_CCHC(5)	1
gro-1	ZC395.6	ZF_C2H2(1)	3
gut-2	T10G3.6	Sm/Lsm	4
ham-2	C07A12.1	ZF_C2H2(2)	3
hars-1	T11G6.1	GF	4
hbl-1	F13D11.2	ZF_C2H2(9)	3
hel-1	C26D10.2	HEL(2)	2
hel-308	Y55B1AL.3	HEL(2)	2
him-6	T04A11.6	HEL(2)	2
him-8	T07G12.12	ZF_C2H2(2)	3
hrp-1	F42A6.7	RRM(2)	1
hrp-2	F58D5.1	RRM(3)	1
hrpf-1	W02D3.11	RRM(3)	1
hrpf-2	Y73B6BL.33	RRM(3)	1

iars-1	R11A8.6	GF	4
iars-2	C25A1.7	GF	4
ife-1	F53A2.6	GF	4
ife-2	R04A9.4	GF	4
ife-3	B0348.6	GF	4
ife-4	C05D9.5	GF	4
ife-5	Y57A10A.30	GF	4
ifg-1	M110.4	GF	4
inf-1	F57B9.6	HEL(2)	2
ire-1	C41C4.4	NUCL	4
isw-1	F37A4.8	HEL(2)	2
kars-1	T02G5.9	tRNA	4
klf-1	F56F11.3	ZF_C2H2(3)	3
klf-2	F53F8.1	ZF_C2H2(3)	3
klf-3	F54H5.4	ZF_C2H2(3)	3
laf-1	Y71H2AM.19	HEL(2)	2
larp-1	R144.7	La	1
larp-5	T12F5.5	La	1
lars-1	R74.1	GF	4
lars-2	ZK524.3	GF	4
let-418	F26F12.7	HEL(2)	2
let-716	C16A3.3	GF	4
lin-13	C03B8.4	ZF_C2H2(22)	3

lin-26	F18A1.2	ZF_C2H2(2)	3
lin-28	F02E9.2	CSD(1)ZF_CCHC(2)	1
lin-29	W03C9.4	ZF_C2H2(5)	3
lin-36	F44B9.6	ZF_C2H2(2)	3
lin-48	F34D10.5	ZF_C2H2(4)	3
lir-1	F18A1.3	ZF_C2H2(1)	3
lir-2	F18A1.4	ZF_C2H2(2)	3
lir-3	F37H8.1	ZF_C2H2(2)	3
lsl-1	F52F12.4	ZF_C2H2(5)	3
lsm-1	F40F8.9	Sm/Lsm	4
lsm-3	Y62E10A.12	Sm/Lsm	4
lsm-4	F32A5.7	Sm/Lsm	4
lsm-5	F28F8.3	Sm/Lsm	4
lsm-6	Y71G12B.14	Sm/Lsm	4
lsm-7	ZK593.7	Sm/Lsm	4
lsm-8	Y73B6BL.32	Sm/Lsm	4
lsy-2	F49H12.1	ZF_C2H2(5)	3
lsy-27	F47H4.1	ZF_C2H2(3)	3
mars-1	F58B3.5	tRNA	4
mbl-1	K02H8.1	ZF_CCCH(2)	1
mcd-1	Y51H1A.6	ZF_C2H2(1)	3
mcm-2	Y17G7B.5	GF	4
mec-8	F46A9.6	RRM(2)	1

mel-46	T06A10.1	HEL(2)	2
mel-47	EEED8.1	RRM(1)	1
mep-1	M04B2.1	ZF_C2H2(7)	3
mex-1	W03C9.7	ZF_CCCH(2)	1
mex-3	F53G12.5	KH(2)	1
mex-5	W02A2.7	ZF_CCCH(2)	1
mex-6	AH6.5	ZF_CCCH(2)	1
mlh-1	T28A8.7	GF	4
mnm-2	C10A4.8	ZF_C2H2(3)	3
moe-3	F32A11.6	ZF_CCCH(2)	1
mog-1	K03H1.2	HEL(2)	2
mog-4	C04H5.6	HEL(2)	2
mog-5	EEED8.5	HEL(2)	2
mrps-5	E02A10.1	DSRBD(1)	1
msi-1	R10E9.1	RRM(2)	1
mtr-4	W08D2.7	HEL(2)	2
mtss-1	PAR2.1	GF	4
mut-14	C14C11.6	HEL(2)	2
mut-7	ZK1098.8	NUCL	4
mvk-1	Y42G9A.4	GF	4
nars-1	F22D6.3	tRNA	4
nars-2	Y66D12A.23	tRNA	4
ncbp-2	F26A3.2	RRM(1)	1

nos-2	ZK1127.1	ZF_CCHC(Nanos)(2)	1
nos-3	Y53C12B.3	ZF_CCHC(Nanos)(2)	1
nrd-1	D1007.7	RRM(1)	1
nrde-3	R04A9.2	PAZ(1)PIWI(1)	2
ntl-4	C49H3.5	RRM(1)	1
nuo-2	T10E9.7	KH(2)	1
nxf-1	C15H11.3	NTF2	2
nxf-2	C15H11.6	NTF2	2
nxt-1	Y71F9AM.5	NTF2	2
odd-1	B0280.4	ZF_C2H2(3)	3
odd-2	C34H3.2	ZF_C2H2(3)	3
oma-1	C09G9.6	ZF_CCCH(2)	1
oma-2	ZC513.6	ZF_CCCH(2)	1
pab-1	Y106G6H.2	RRM(4)	1
pab-2	F18H3.3	RRM(4)	1
pabp-2	C17E4.5	RRM(1)	1
pag-3	F45B8.4	ZF_C2H2(5)	3
pars-1	T20H4.3	GF	4
pars-2	T27F6.5	GF	4
pash-1	T22A3.5	DSRBD(1)	1
pbrm-1	C26C6.1	ZF_C2H2(2)	3
pes-4	Y119D3B.17	KH(2)	1

pgl-1	ZK381.4	RGG	1
pgl-3	C18G1.4	RGG	1
pie-1	Y49E10.14	ZF_CCCH(2)	1
pms-2	H12C20.2	GF	4
polq-1	W03A3.2	HEL(2)	2
pos-1	F52E1.1	ZF_CCCH(2)	1
ppw-1	C18E3.7	PAZ(1)PIWI(1)	2
ppw-2	Y110A7A.18	PAZ(1)PIWI(1)	2
pqe-1	F52C9.8	NUCL	4
pqm-1	F40F8.7	ZF_C2H2(3)	3
pqn-21	C37A2.5	ZF_C2H2(1)	3
prg-1	D2030.6	PAZ(1)PIWI(1)	2
prg-2	C01G5.2	PAZ(1)PIWI(1)	2
prp-8	C50C3.6	RRM(1)	1
psa-4	F01G4.1	HEL(2)	2
psf-1	F25B5.7	RRM(2)	1
ptb-1	D2089.4	RRM(4)	1
puf-11	Y73B6BL.38	PUF(8)	1
puf-12	ZK945.3	PUF(6)	1
puf-3	Y45F10A.2	PUF(8)	1
puf-4	M4.2	PUF(6)	1
puf-5	F54C9.8	PUF(8)	1
puf-6	F18A11.1	PUF(8)	1

puf-7	B0273.2	PUF(8)	1
puf-8	C30G12.7	PUF(8)	1
puf-9	W06B11.2	PUF(8)	1
pxl-1	C28H8.6	ZF_CCHC(1)	1
pzf-1	T05G11.1	ZF_C2H2(8)	3
qars-1	Y41E3.4	GF	4
rabs-5	Y42H9AR.3	ZF_C2H2(1)	3
rad-26	C27B7.4	HEL(2)	2
rad-54	W06D4.6	HEL(2)	2
ran-4	R05D11.3	NTF2	2
rars-1	F26F4.10	GF	4
rars-2	C29H12.1	GF	4
rbd-1	T23F6.4	RRM(6)	1
rcq-5	E03A3.2	HEL(2)	2
rde-1	K08H10.7	PAZ(1)PIWI(1)	2
rde-4	T20G5.11	DSRBD(2)	1
ref-2	C47C12.3	ZF_C2H2(4)	3
repo-1	F11A10.2	ZF_C2H2(1)	3
rha-1	T07D4.3	DSRBD(2)HEL(2)	1
rha-2	C06E1.10	HEL(2)	2
rla-0	F25H2.10	GF	4
rla-1	Y37E3.7	GF	4
rla-2	Y62E10A.1	GF	4

rnf-113	K01G5.1	ZF_CCCH(1)	1
rnp-1	ZK863.7	RRM(1)ZF_CCHC(1)	1
rnp-2	K08D10.4	RRM(2)	1
rnp-3	K08D10.3	RRM(2)	1
rnp-4	R07E5.14	RRM(1)	1
rnp-5	K02F3.11	RRM(1)	1
rnp-6	Y47G6A.20	RRM(3)	1
rnp-7	K04G7.10	RRM(1)	1
rnp-8	R119.7	RRM(1)	1
rop-1	C12D8.11	GF	4
rpb-7	Y54E10BR.6	GF	4
rpb-8	F26F4.11	GF	4
rpl-1	Y71F9AL.13	GF	4
rpl-10	F10B5.1	GF	4
rpl-11.1	T22F3.4	GF	4
rpl-11.2	F07D10.1	GF	4
rpl-12	JC8.3	GF	4
rpl-13	C32E8.2	GF	4
rpl-14	C04F12.4	GF	4
rpl-15	K11H12.2	GF	4
rpl-16	M01F1.2	GF	4
rpl-17	Y48G8AL.8	GF	4
rpl-18	Y45F10D.12	GF	4

rpl-19	C09D4.5	GF	4
rpl-2	B0250.1	GF	4
rpl-20	E04A4.8	GF	4
rpl-21	C14B9.7	GF	4
rpl-22	C27A2.2	GF	4
rpl-23	B0336.10	GF	4
rpl-24.1	D1007.12	GF	4
rpl-24.2	C03D6.8	GF	4
rpl-25.1	F55D10.2	GF	4
rpl-25.2	F52B5.6	GF	4
rpl-26	F28C6.7	GF	4
rpl-27	C53H9.1	GF	4
rpl-28	R11D1.8	GF	4
rpl-29	B0513.3	GF	4
rpl-3	F13B10.2	GF	4
rpl-30	Y106G6H.3	GF	4
rpl-31	W09C5.6	GF	4
rpl-32	T24B8.1	GF	4
rpl-33	F10E7.7	GF	4
rpl-34	C42C1.14	GF	4
rpl-35	ZK652.4	GF	4
rpl-36	F37C12.4	GF	4
rpl-37	C54C6.1	GF	4

rpl-38	C06B8.8	GF	4
rpl-39	C26F1.9	GF	4
rpl-4	B0041.4	GF	4
rpl-41	C09H10.2	GF	4
rpl-43	Y48B6A.2	GF	4
rpl-5	F54C9.5	GF	4
rpl-6	R151.3	GF	4
rpl-7	F53G12.10	GF	4
rpl-7A	Y24D9A.4	GF	4
rpl-9	R13A5.8	GF	4
rps-0	B0393.1	GF	4
rps-1	F56F3.5	GF	4
rps-10	D1007.6	GF	4
rps-11	F40F11.1	GF	4
rps-12	F54E7.2	GF	4
rps-13	C16A3.9	GF	4
rps-14	F37C12.9	GF	4
rps-15	F36A2.6	GF	4
rps-16	T01C3.6	GF	4
rps-17	T08B2.10	GF	4
rps-18	Y57G11C.16	GF	4
rps-19	T05F1.3	GF	4
rps-2	C49H3.11	GF	4

rps-20	Y105E8A.16	GF	4
rps-21	F37C12.11	GF	4
rps-22	F53A3.3	GF	4
rps-23	F28D1.7	GF	4
rps-24	T07A9.11	GF	4
rps-25	K02B2.5	GF	4
rps-26	F39B2.6	GF	4
rps-27	F56E10.4	GF	4
rps-28	Y41D4B.5	GF	4
rps-29	B0412.4	GF	4
rps-3	C23G10.3	KH(1)	1
rps-30	C26F1.4	GF	4
rps-4	Y43B11AR.4	GF	4
rps-5	T05E11.1	GF	4
rps-6	Y71A12B.1	GF	4
rps-7	ZC434.2	GF	4
rps-8	F42C5.8	GF	4
rps-9	F40F8.10	GF	4
rsp-1	W02B12.3	RRM(2)	1
rsp-2	W02B12.2	RRM(2)	1
rsp-3	Y111B2A.18	RRM(2)	1
rsp-4	EEED8.7	RRM(1)	1
rsp-5	T28D9.2	RRM(1)	1

rsp-6	C33H5.12	RRM(1)	1
rsp-7	D2089.1	RRM(2)	1
rsp-8	C18D11.4	RRM(1)	1
rtel-1	F25H2.13	HEL(2)	2
sago-1	K12B6.1	PAZ(1)PIWI(1)	2
sago-2	F56A6.1	PAZ(1)PIWI(1)	2
sap-49	C08B11.5	RRM(2)	1
sars-1	C47E12.1	GF	4
sars-2	W03B1.4	GF	4
sbds-1	W06E11.4	GF	4
scrt-1	C55C2.1	ZF_C2H2(3)	3
sdc-1	F52E10.1	ZF_C2H2(4)	3
sdz-12	F12E12.5	ZF_C2H2(7)	3
sdz-38	ZK892.7	ZF_C2H2(2)	3
sea-2	K10G6.3	ZF_C2H2(5)	3
sec-24.1	F12F6.6	ZF_C2H2(1)	3
sem-4	F15C11.1	ZF_C2H2(8)	3
set-2	C26E6.9	RRM(1)	1
sfa-1	Y116A8C.32	KH(1)ZF_CCHC(2)	1
sig-7	F39H2.2	RRM(1)	1
slr-2	Y59A8B.13	ZF_C2H2(3)	3
sma-9	T05A10.1	ZF_C2H2(8)	3
smg-2	Y48G8AL.6	HEL(1)	2

smg-4	F46B6.3	RRM(1)	1
smgl-2	Y37E11AM.1	HEL(2)	2
smn-1	C41G7.1	Unknown	1
snr-1	Y116A8C.42	Sm/Lsm	4
snr-2	W08E3.1	Sm/Lsm	4
snr-3	T28D9.10	Sm/Lsm	4
snr-4	C52E4.3	Sm/Lsm	4
snr-5	ZK652.1	Sm/Lsm	4
snr-6	Y49E10.15	Sm/Lsm	4
snr-7	Y71F9B.4	Sm/Lsm	4
somi-1	M04G12.4	ZF_C2H2(1)	3
sop-2	C50E10.4	Unknown	2
sor-1	ZK1236.3	Unknown	2
spn-4	ZC404.8	RRM(1)	1
spr-3	C07A12.5	ZF_C2H2(7)	3
spr-4	C09H6.1	ZF_C2H2(9)	3
sptf-1	F45H11.1	ZF_C2H2(1)	3
sptf-2	T22C8.5	ZF_C2H2(3)	3
sptf-3	Y40B1A.4	ZF_C2H2(3)	3
sqd-1	Y73B6BL.6	RRM(2)	1
ssl-1	Y111B2A.22	HEL(2)	2
stau-1	F55A4.5	DSRBD(4)	1
sup-12	T22B2.4	RRM(1)	1

sup-26	R10E4.2	RRM(2)	1
sup-35	Y48A6C.3	ZF_C2H2(2)	3
sut-2	Y61A9LA.8	ZF_CCCH(3)	1
syd-9	ZK867.1	ZF_C2H2(4)	3
sym-2	ZK1067.6	RRM(3)	1
taf-1	W04A8.7	ZF_CCHC(1)	1
tag-146	C27A12.3	ZF_C2H2(7)	3
tag-192	T04D1.4	HEL(1)	2
tag-214	F36F2.3	ZF_CCHC(1)	1
tag-262	Y54H5A.3	RRM(1)	1
tag-76	ZK757.3	PAZ(1)PIWI(1) NUCL(1)	2
tag-96	M01D7.4	GF	4
tars-1	C47D12.6	GF	4
tdp-1	F44G4.4	RRM(2)	1
tiar-1	C18A3.5	RRM(3)	1
tiar-2	Y46G5A.13	RRM(3)	1
tiar-3	C07A4.1	RRM(2)	1
tlp-1	T23G4.1	ZF_C2H2(1)	3
top-2	K12D12.1	GF	4
tra-1	Y47D3A.6	ZF_C2H2(5)	3
tra-4	F53B3.1	ZF_C2H2(7)	3
tut-1	F29C4.6	GF	4
uaf-1	Y92C3B.2	RRM(3)	1

uaf-2	Y116A8C.35	RRM(1)ZF_CCCH(2)	1
ubl-1	H06I04.4	GF	4
ubq-2	ZK1010.1	GF	4
ubxn-1	F23C8.4	ZF_C2H2(1)	3
unc-75	C17D12.2	RRM(3)	1
unc-98	F08C6.7	ZF_C2H2(3)	3
vars-1	ZC513.4	GF	4
vars-2	Y87G2A.5	GF	4
vbh-1	Y54E10A.9	HEL(2)	2
vps-54	T21C9.2	ZF_C2H2(1)	3
wars-1	Y80D3A.1	GF	4
wars-2	C34E10.4	GF	4
wrn-1	F18C5.2	HEL(2)	2
xnp-1	B0041.7	HEL(1)	2
xpg-1	F57B10.6	NUCL	4
xpo-3	C49H3.10	GF	4
xrn-2	Y48B6A.3	ZF_CCHC(1)	1
yars-1	K08F11.4	GF	4
zag-1	F28F9.1	ZF_C2H2(5)	3
zfp-2	F35H8.3	ZF_C2H2(7)	3
zfp-3	W05H7.4	ZF_CCCH(1)	1
zim-1	T07G12.6	ZF_C2H2(1)	3
zim-2	T07G12.10	ZF_C2H2(1)	3

zim-3	T07G12.11	ZF_C2H2(1)	3
ztf-1	F54F2.5	ZF_C2H2(2)	3
ztf-11	F52F12.6	ZF_CCHC(2)	1
ztf-12	C01B7.1	ZF_C2H2(7)	3
ztf-14	M163.2	ZF_C2H2(5)	3
ztf-15	R06C7.9	ZF_C2H2(12)	3
ztf-16	R08E3.4	ZF_C2H2(7)	3
ztf-18	T24C4.7	ZF_C2H2(2)	3
ztf-19	T27B1.2	ZF_C2H2(3)	3
ztf-2	F13G3.1	ZF_C2H2(2)	3
ztf-20	Y39B6A.46	ZF_C2H2(3)	3
ztf-22	Y48C3A.4	ZF_C2H2(4)	3
ztf-23	Y54E10BR.8	ZF_C2H2(6)	3
ztf-25	Y6G8.3	ZF_C2H2(3)	3
ztf-26	F55H12.6	ZF_C2H2(3)	3
ztf-27	T09F3.1	ZF_C2H2(2)	3
ztf-28	F36F12.8	ZF_C2H2(3)	3
ztf-29	Y66D12A.12	ZF_C2H2(3)	3
ztf-3	C53D5.4	ZF_C2H2(2)	3
ztf-30	C06E1.8	ZF_C2H2(2)	3
ztf-4	T10B11.3	RRM(1)ZF_C2H2(2)	1
ztf-6	W06H12.1	ZF_C2H2(3)	3
ztf-7	F46B6.7	ZF_C2H2(4)	3

ztf-8	ZC395.8	ZF_C2H2(3)	3
ztf-9	ZK287.6	ZF_C2H2(4)	3
	C11D2.7	GF	4
	C15C6.4	GF	4
	C48B6.2	GF	4
	F08B4.7	GF	4
	B0205.11	GF	4
	C03D6.1	PIWI(1)	2
	C08H9.3	NUCL	4
	C23G10.2	NUCL	4
	C26E6.6	GF	4
	C30C11.1	GF	4
	C37A2.7	GF	4
	F10E7.5	GF	4
	F25H5.6	GF	4
	F29B9.10	GF	4
	F43C1.6	GF	4
	K01C8.6	GF	4
	K07C5.3	ZF_C2H2(1)	3
	K10C9.3	NUCL	4
	M01F1.6	GF	4
	M28.5	GF	4
	T02G5.11	ZF_CCHC(Nanos)(2)	1

	T08B2.8	GF	4
	T21B10.1	GF	4
	T26G10.3	GF	4
	W04A8.4	NUCL	4
	W05H12.2	NUCL	4
	W09C5.1	GF	4
	Y105E8A.19	GF	4
	Y119C1B.4	GF	4
	Y37H9A.5	GF	4
	Y48A6B.3	GF	4
	Y57A10A.13	NUCL	4
	C05C8.5	NUCL	4
	C08B6.8	NUCL	4
	C16A3.4	ZF_C2H2(4)	3
	F31E3.4	NUCL	4
	M02B7.2	NUCL	4
	M03C11.2	HEL(2)	2
	R02D3.8	NUCL	4
	Y17G7B.12	NUCL	4
	Y50D7A.11	HEL(1)	2
	Y50D7A.2	HEL(2)	2
	Y56A3A.33	NUCL	4
	Y95B8A.7	ZF_C2H2(2)	3

	ZK1098.3	NUCL	4
	ZK686.4	ZF_C2H2(1)	3
	B0310.2	ZF_C2H2(2)	3
	B0511.6	HEL(2)	2
	C08F8.2	HEL(2)	2
	C16A3.1	HEL(2)	2
	C24H12.4	HEL(3)	2
	C27A2.7	ZF_C2H2(4)	3
	C28H8.3	HEL(2)	2
	C46F11.4	HEL(2)	2
	C52B9.8	HEL(2)	2
	F01G4.3	HEL(2)	2
	F07E5.5	ZF_CCHC(4)	1
	F13H6.1	ZF_C2H2(4)	3
	F27D4.4	ZF_CCCH(2)	1
	F33D11.10	HEL(2)	2
	F40F12.9	ZF_CCHC(2)	1
	F52B5.3	HEL(2)	2
	F53H1.1	HEL(2)	2
	F54E12.2	HEL(2)	2
	F54G2.2	ZF_CCHC(1)	1
	F55F8.2	HEL(2)	2
	F56D2.6	HEL(2)	2

	F57B9.3	HEL(2)	2
	F57C9.4	ZF_C2H2(5)	3
	F58E10.3	HEL(2)	2
	F58G11.2	HEL(2)	2
	F59A7.8	HEL(2)	2
	H06O01.2	HEL(2)	2
	H20J04.4	HEL(2)	2
	H27M09.1	HEL(2)ZF_CCHC(1)	1
	K02D7.2	ZF_C2H2(4)	3
	K02F3.12	HEL(2)	2
	K02F6.1	ZF_CCHC(1)	1
	K07C5.6	ZF_CCHC(1)	1
	K08D12.3	ZF_CCHC(5)	1
	K11D2.4	ZF_C2H2(4)	3
	M03C11.8	HEL(2)	2
	R05D11.4	HEL(2)	2
	T05A12.4	HEL(2)	2
	T05E8.3	HEL(2)	2
	T23H2.3	HEL(2)	2
	T26G10.1	HEL(2)	2
	Y111B2A.10	ZF_C2H2(6)	3
	Y116A8C.13	HEL(2)	2
	Y23H5B.6	HEL(2)	2

	Y37E11B.1	ZF_C2H2(5)	3
	Y38A10A.6	HEL(2)	2
	Y38H8A.5	ZF_C2H2(6)	3
	Y46G5A.4	HEL(3)	2
	Y48G8AL.10	ZF_C2H2(19)	3
	Y53G8AR.9	ZF_CCCH(3)	1
	Y54E2A.4	HEL(4)	2
	Y54G11A.3	HEL(2)	2
	Y55F3AM.14	ZF_C2H2(11)	3
	Y55F3BR.1	HEL(2)	2
	Y56A3A.18	ZF_C2H2(1)	3
	Y56A3A.28	ZF_C2H2(3)	3
	Y5F2A.4	ZF_C2H2(6)	3
	Y65B4A.6	HEL(2)	2
	Y66D12A.15	HEL(2)	2
	Y67D2.6	HEL(2)	2
	Y94H6A.5	HEL(2)	2
	ZC317.1	HEL(3)	2
	ZC328.2	ZF_C2H2(3)	3
	ZK337.2	ZF_C2H2(4)	3
	ZK512.2	HEL(2)	2
	ZK686.2	HEL(2)	2
	ZK792.5	PUF(3)	1

	B0035.12	RRM(2)	1
	B0336.3	RRM(1)ZF_CCCH(1)	1
	C06G4.1	KH(3)	1
	C07H6.4	RRM(1)	1
	C08H9.2	KH(15)	1
	C12D8.1	KH(4)	1
	C25A1.4	RRM(3)	1
	C30B5.4	RRM(1)	1
	C35D6.4	ZF_CCCH(2)	1
	C36E6.1	KH(4)	1
	C44B7.2	RRM(3)	1
	C44E4.4	RRM(2)	1
	C50B8.1	RRM(1)	1
	C50D2.5	RRM(1)	1
	C56G2.1	KH(1)	1
	D1037.1	DSRBD(1)	1
	EEED8.10	RRM(1)	1
	EEED8.12	RRM(1)	1
	EEED8.4	RRM(1)	1
	F11A10.7	RRM(2)	1
	F26B1.2	KH(3)	1
	F29C4.7	RRM(3)	1
	F32B4.4	RRM(1)	1

	F38C2.7	ZF_CCCH(2)	1
	F39E9.7	DSRBD(1)	1
	F55A4.4	DSRBD(2)	1
	F58B3.7	RRM(1)	1
	H24K24.4	RRM(1)	1
	H28G03.1	RRM(2)	1
	K07H8.10	RRM(2)	1
	K07H8.9	KH(1)	1
	K08F4.2	RRM(1)	1
	M01A10.1	KH(4)	1
	M88.5	KH(4)	1
	R05D3.8	RRM(2)	1
	R05H10.2	RRM(3)	1
	R06C1.4	RRM(1)	1
	R09B3.2	RRM(1)	1
	R09B3.3	RRM(1)	1
	R11A8.7	KH(1)	1
	T04A8.6	RRM(1)	1
	T07F10.3	RRM(2)	1
	T08B2.5	RRM(2)ZF_C2H2(1)	1
	T08B6.5	RRM(1)	1
	T11G6.8	RRM(1)ZF_CCCH(1)	1

	W04D2.6	RRM(1)	1
	Y116A8C.19	ZF_CCCH(2)	1
	Y116A8C.20	ZF_CCCH(2)	1
	Y18D10A.8	KH(1)	1
	Y37D8A.21	RRM(1)	1
	Y37E3.10	GF	4
	Y55F3AM.3	RRM(3)	1
	Y57G11A.5	RRM(1)	1
	Y57G11C.36	KH(1)	1
	Y57G11C.9	RRM(1)	1
	Y59A8B.10	KH(2)	1
	Y59E9AL.36	RRM(1)	1
	Y60A9.3	ZF_CCCH(1)	1
	Y65B4A.1	RRM(2)	1
	Y69A2AR.32	KH(1)	1
	ZK418.8	KH(2)	1
	ZK418.9	KH(4)	1
	ZK616.1	RRM(2)	1
	ZK632.2	DSRBD(1)	1
	ZK973.1	GF	4
	B0280.17	KH(1)	1
	C02B10.4	RRM(1)	1
	C23H3.3	KH(1)	1

	C41G7.3	KH(2)	1
	C49H3.4	Sm/Lsm	4
	Y106G6D.7	RRM(2)	1
	Y48G1C.9	Sm/Lsm	4
	Y53C12B.2	KH(1)	1
	Y54E5A.6	DSRBD(1)	1
	ZC190.4	SAM(1)	3
	C04F12.1	PIWI(1)	2
	M03D4.6	PIWI(1)	2
	Y71G12B.8	HEL(2)	2
	ZK218.8	PIWI(1)	2
	C14B1.7	PAZ(1)PIWI(1)	2
	C16C10.3	PAZ(1)PIWI(1)	2
	F19G12.3	PAZ(1)	2
	F47E1.3	ZF_C2H2(5)	3
	F55A12.1	PAZ(1)PIWI(1)	2
	F58G1.1	PAZ(1)PIWI(1)	2
	R06C7.1	PAZ(1)PIWI(1)	2
	T22B3.2	PAZ(1)PIWI(1)	2
	T22H9.3	PAZ(1)PIWI(1)	2
	T23D8.7	PAZ(1)PIWI(1)	2
	Y49F6A.1	PAZ(1)PIWI(1)	2
	ZK1248.7	PAZ(1)PIWI(1) NUCL(1)	2

	B0303.15	GF	4
	C33E10.6	NUCL	4
	CD4.3	GF	4
	F09G8.3	GF	4
	F13G3.11	GF	4
	F33D4.5	GF	4
	F45E12.5	GF	4
	F53F4.11	GF	4
	F55C9.3	PAZ(1)	2
	F56B3.8	GF	4
	F56D1.3	GF	4
	K07A1.15	Sm/Lsm	4
	K11H3.6	GF	4
	T03D8.2	GF	4
	T04A8.11	GF	4
	T14B4.2	GF	4
	T23B12.2	GF	4
	T23B12.3	GF	4
	W01D2.1	GF	4
	W04D2.5	GF	4
	W09D10.3	GF	4
	Y105E8A.20	GF	4
	Y26D4A.9	NUCL	4

	Y37D8A.18	GF	4
	Y37E3.8	GF	4
	Y39A1A.6	GF	4
	Y48C3A.10	GF	4
	Y54E10A.7	GF	4
	Y54G11A.17	GF	4
	Y57G11C.34	GF	4
	ZC262.8	GF	4
	ZC53.4	NUCL	4
	2L52.1	ZF_C2H2(2)	3
	B0035.1	ZF_C2H2(2)	3
	B0336.7	ZF_C2H2(1)	3
	C01F6.9	ZF_C2H2(1)	3
	C02F5.12	ZF_C2H2(2)	3
	C06E2.1	ZF_C2H2(1)	3
	C08B11.3	ZF_C2H2(2)	3
	C13F10.7	ZF_CCHC(1)	1
	C29E4.13	ZF_C2H2(2)	3
	C38D9.2	ZF_CCHC(1)	1
	C46G7.3	ZF_CCHC(2)	1
	C52E12.1	ZF_C2H2(4)	3
	D1046.2	ZF_C2H2(3)	3
	D2092.10	ZF_C2H2(2)	3

	EGAP9.4	ZF_C2H2(1)	3
	F07B7.1	ZF_CCHC(1)	1
	F10B5.3	ZF_C2H2(4)	3
	F10E7.2	ZF_C2H2(1)	3
	F15D4.5	ZF_CCHC(1)	1
	F21G4.5	ZF_C2H2(1)	3
	F27D4.6	ZF_C2H2(5)	3
	F33H1.4	ZF_C2H2(3)	3
	F37B4.10	ZF_C2H2(1)	3
	F37D6.2	ZF_C2H2(4)	3
	F41G4.5	ZF_CCHC(1)	1
	F46F11.8	ZF_C2H2(1)	3
	F48F7.5	ZF_C2H2(2)	3
	F54C4.3	ZF_C2H2(5)	3
	F55A3.2	ZF_C2H2(1)	3
	F57G12.2	ZF_C2H2(1)	3
	F58D12.5	ZF_C2H2(3)	3
	K04C1.3	ZF_C2H2(1)	3
	K05F1.5	ZF_C2H2(2)	3
	K10B3.5	ZF_C2H2(2)	3
	K12H6.12	ZF_C2H2(3)	3
	R02D3.7	ZF_C2H2(3)	3
	R05D3.3	ZF_C2H2(2)	3

	R144.3	ZF_C2H2(2)	3
	T06G6.5	ZF_C2H2(1)	3
	T08G5.7	ZF_C2H2(3)	3
	T10E9.2	ZF_C2H2(3)	3
	T13H5.4	ZF_C2H2(1)	3
	T22C8.4	ZF_C2H2(2)	3
	T23F11.4	ZF_C2H2(2)	3
	W02D7.6	ZF_C2H2(3)	3
	W04B5.2	ZF_C2H2(1)	3
	W09B7.1	ZF_CCHC(1)	1
	Y113G7B.14	HEL(1)	2
	Y14H12B.1	ZF_C2H2(1)	3
	Y34D9A.7	ZF_CCHC(1)	1
	Y43E12A.2	ZF_CCHC(1)	1
	Y43F4A.3	ZF_CCHC(1)	1
	Y48E1B.7	ZF_C2H2(1)	3
	Y48G9A.11	ZF_C2H2(2)	3
	Y52B11A.9	ZF_C2H2(1)	3
	Y53F4B.5	ZF_C2H2(1)	3
	Y57A10A.31	ZF_C2H2(1)	3
	Y71A12B.8	ZF_C2H2(2)	3
	Y79H2A.3	ZF_C2H2(4)	3
	Y82E9BR.1	ZF_C2H2(3)	3

	Y82E9BR.17	ZF_C2H2(2)	3
	ZC132.6	ZF_CCHC(1)	1
	ZC15.1	ZF_CCHC(1)	1
	ZK185.1	ZF_C2H2(1)	3
	ZK632.11	ZF_CCHC(1)	1
	ZK856.10	GF	4
	B0250.4	ZF_C2H2(7)	3
	C04F5.9	ZF_C2H2(7)	3
	C09F5.3	ZF_C2H2(8)	3
	C27A12.2	ZF_C2H2(8)	3
	C27B7.5	ZF_CCHC(1)	1
	C28G1.4	ZF_C2H2(5)	3
	C28H8.9	ZF_C2H2(1)	3
	C38D4.7	ZF_C2H2(1)	3
	C46E10.8	ZF_C2H2(2)	3
	C46E10.9	ZF_C2H2(4)	3
	C52E12.6	ZF_C2H2(5)	3
	D2030.7	ZF_C2H2(2)	3
	F21A9.2	ZF_C2H2(3)	3
	F21D5.9	ZF_C2H2(1)	3
	F26A10.2	ZF_C2H2(4)	3
	F26F4.8	ZF_C2H2(6)	3
	F39B2.1	ZF_C2H2(10)	3

	F44E2.7	ZF_C2H2(3)	3
	F49F1.8	ZF_CCHC(2)	1
	F52B5.7	ZF_C2H2(1)	3
	F53H10.2	ZF_C2H2(1)	3
	F55C5.11	ZF_C2H2(2)	3
	F56D1.1	ZF_C2H2(7)	3
	F58G1.2	ZF_C2H2(7)	3
	F59B2.14	ZF_C2H2(2)	3
	H16D19.3	ZF_C2H2(3)	3
	K09F6.11	ZF_C2H2(2)	3
	K09H9.7	ZF_C2H2(2)	3
	K11D12.12	ZF_C2H2(2)	3
	M03D4.4	ZF_C2H2(7)	3
	R07E5.5	ZF_C2H2(2)	3
	R10E4.11	ZF_C2H2(5)	3
	T07D10.3	ZF_C2H2(5)	3
	T20F7.1	ZF_C2H2(2)	3
	T20H4.2	ZF_C2H2(2)	3
	T22C8.3	ZF_C2H2(1)	3
	W04D2.4	ZF_C2H2(1)	3
	Y17G7B.22	ZF_C2H2(2)	3
	Y48G8AL.9	ZF_C2H2(2)	3
	Y53H1A.2	ZF_C2H2(2)	3

	Y54G2A.20	ZF_C2H2(2)	3
	Y55F3AM.6	ZF_CCCH(4)	1
	Y67H2A.10	ZF_C2H2(3)	3
	Y73F8A.33	ZF_C2H2(3)	3
	ZC123.3	ZF_C2H2(12)	3
	ZK546.5	ZF_C2H2(6)	3
	ZK686.5	ZF_C2H2(3)	3
	C34D10.2	ZF_CCCH(5)	1
	T26A8.4	ZF_CCCH(3)	1
	C05C8.2	KH(1)	1
	E02D9.1	KH(1)	1
	F54D1.1	KH(1)	1
	H05C05.1	RRM(1)	1
	Y79H2A.4	RRM(1)	1
	C30F12.1	ZF_CCCH(1)	1
	C33H5.17	ZF_CCCH(1)	1
	C34H4.5	ZF_C2H2(1)	3
	C35D6.3	PAZ(1)	2
	C50C3.1	ZF_CCCH(1)	1
	F10G7.9	ZF_CCCH(1)	1
	F53H4.6	HEL(1)	2
	T04H1.5	ZF_C2H2(1)	3
	Y22D7AL.16	ZF_C2H2(1)	3

	Y37E11B.5	ZF_CCCH(2)	1
	BE0003N10.1	NUCL	4
	Y37F4.6	ZF_C2H2(3)	3
	Y71A12B.2	PAZ(1)	2
	Y71A12B.3	PAZ(1)	2
	ZK856.11	GF	4
	B0361.6	GF	4
	C04E6.8	NTF2	2
	C06C3.8	GF	4
	C08G5.1	GF	4
	C26D10.4	GF	4
	C34B2.6	GF	4
	C37C3.2	GF	4
	D1007.8	GF	4
	E02H1.2	KH(1)	1
	F08F3.8	GF	4
	F09E10.7	NTF2	2
	F19B2.7	GF	4
	F29C12.4	GF	4
	F46B6.6	GF	4
	F46F5.5	GF	4
	K07A12.7	GF	4
	K10C3.5	GF	4

	R02C2.7	NTF2	2
	R05D3.1	GF	4
	R09E12.9	NTF2	2
	R12E2.12	GF	4
	R13H4.6	GF	4
	R151.7	GF	4
	T02D1.8	NTF2	2
	T05E11.3	GF	4
	T05H4.6	GF	4
	T13H5.5	GF	4
	T20F5.3	GF	4
	T23G4.2	GF	4
	W10C6.2	GF	4
	Y105C5B.5	NTF2	2
	Y119D3B.14	GF	4
	Y48B6A.13	GF	4
	Y52B11A.2	GF	4
	Y75B8A.4	GF	4
	Y82E9BR.2	RRM(1)	1
	Y92H12BR.8	GF	4
	ZK930.6	GF	4

CHAPTER III

Preface

This chapter comprises work that is currently in submission with the following title and authorship:

A gene-centered method for mapping 3'UTR-RBP interactions.

Alex M. Tamburino, Ebru Kaymak, Shaleen Shrestha, Sean P. Ryder, Albertha J.M. Walhout.

The work in this chapter was performed in majority by myself with some help from Shaleen Shrestha. All worm experiments and the images shown were performed by Ebru Kaymak in the lab of Sean Ryder.

Protein-RNA Interaction Mapping Assay (PRIMA)

The post-transcriptional regulation of gene expression is vital to organismal development and homeostasis. Post-transcriptional gene regulation affects many aspects of an mRNA, including splicing pattern, 3´ end formation, nuclear-cytoplasmic export, localization, translation and stability (Glisovic *et al.* 2008). These processes are controlled by physical interactions with different RBPs that often occur through the 3´UTR (Moore 2005, Szostak and Gebauer 2013).

In the last years, thousands of 3´UTRs have been experimentally defined in several model organisms (Ulitsky *et al.* 2012, Derti *et al.* 2012, Mangone *et al.* 2010, Jan *et al.* 2011). In addition, compendia of RBPs have been predicted or experimentally determined that frequently encompass ~5% of all protein-coding genes (Tamburino *et al.* 2013, Castello *et al.* 2012, Baltz *et al.* 2012, Gerstberger *et al.* 2014). Thus, there is a vast matrix of potential interactions between 3´UTRs and RBPs that needs to be explored.

Several assays are available for the study of interactions between RBPs and mRNAs. Protein-centered methods include RIP and CLIP that

can be used to identify the mRNAs that associate with an RBP *in vivo* (Hafner *et al.* 2010). *In vitro* methods to characterize the binding specificity of RBPs include electrophoretic mobility shift (Pagano *et al.* 2011) and RNAcompete (Ray *et al.* 2009) assays that can be used to test binding of individual RBPs to single or multiple RNA elements, respectively. These methods are relatively limited because they require suitable anti-RBP antibodies and/or purified RBPs. Other methods include yeast three-hybrid (Y3H) assays (SenGupta *et al.* 1996) that are based on transcription and can be used to study RNA-RBP interactions in the milieu of the yeast nucleus. Y3H assays are limited by the length and sequence of the RNA. These limitations present large challenges for the study of 3´UTR-RBP interactions because the majority of metazoan 3´UTRs are longer than the Y3H limit and contain many tetrauridine sequences which prevent their testing in Y3H (Zhang *et al.* 1999). However, RNA-centered methods that enable, in a single experiment, the identification of the repertoire of RBPs that can bind individual 3´UTRs or other RNA elements are lacking.

Here, we present PRIMA, an RNA-centered Protein-RNA Interaction Mapping Assay that can be used to study RNA-RBP interactions with a variety of RNA elements or 3´UTRs and different RBPs within the cytoplasm of yeast cells. PRIMA is based on the endogenous function of yeast poly(A)-binding protein (Pab1p) which binds the mRNA 3´ poly(A) tail

and interacts with the 5′ end through the scaffold protein, eIF4G, and the cap binding protein, eIF4E, thereby stabilizing the mRNA and increasing translation of the mRNA into protein (Mangus *et al.* 2003). We reasoned that we could reconstitute this interaction by using a reporter mRNA that encodes GFP and replacing its poly(A) tail with a selected RNA element (*e.g.*, a 3′UTR) of interest, and fusing a candidate interacting RBP to Pab1p (Figure 3.1A). When the RBP binds the RNA element, Pab1p interacts with the 5′ end of the reporter resulting in mRNA stabilization and production of GFP. To remove the poly(A) tail from the reporter mRNA, we added a *cis*-encoded, self-cleaving hammerhead ribozyme to the 3′ end of the mRNA, just 5′ of the poly(A) tail. As a result, ribozyme catalysis removes the poly(A) tail post-transcriptionally (Dower *et al.* 2004).

The first step for a PRIMA experiment is to generate a yeast strain harboring an mRNA comprising the GFP open reading frame, a generic *C. elegans unc-54* 3′UTR and an RNA element of interest. This RNA element is referred to as the ‘RNA bait.’ The RNA bait strain is subsequently transformed with a plasmid encoding a fusion between an RBP (the ‘prey’) and Pab1p. GFP expression is then measured in ~50,000 cells per transformant, using automated fluorescence flow cytometry

Figure 3.1 PRIMA design and experimental workflow. A) In PRIMA, RNA-RBP interactions are measured by GFP expression from a reporter mRNA or 'RNA bait.' RBP 'preys' are fused to Pab1p, which binds the translation initiation machinery when bound to the 3´ end of the mRNA. The GFP reporter mRNA including a minimal *unc-54* 3´UTR (gray) and an RNA element of interest (red) is expressed without a poly(A) tail by using a *cis*-encoded, self-cleaving hammerhead ribozyme (black). An RBP-Pab1p fusion protein (half-circles) is co-expressed with the reporter bait RNA. When the bait mRNA and RBP prey do not interact the mRNA is unstable and the GFP signals remain low. In contrast, when the RBP binds the RNA element of interest, the mRNA is stabilized and translated resulting in increased GFP levels. B) A yeast RNA bait strain is transformed with an RBP-Pab1p-encoding plasmid. Multiple plasmids can be transformed in parallel. Independent colonies are isolated and grown to log phase in liquid media. GFP expression is measured in ~50,000 cells per replicate using automated flow cytometry. C) Data filtering. The 50% most uniform cells are selected according the forward scatter (FSC, size) and side scatter (SSC, granularity) profiles. Next, 'non-zero' GFP cells are selected to ensure basal mRNA expression. The minimum fluorescence threshold (FL1>2048 *i.e.* fluorescence) was determined using GFP(-) control cell populations. FL1>1024 was used for the HBE:MS2BS RNA baits due to

their low background. Finally, the peak fluorescence was determined for each replicate (see Supplementary Methods for details). D,E) The MS2BS and HBE stem-loop RNA baits were tested with their known RBP partners, MS2 and SLBP, respectively. Flow cytometry data is plotted as density *vs.* fluorescence: positive interaction (green curve) and negative control interaction (gray curve). Dot plots show the peak fluorescence for each of the eight replicates. The bar represents the mean of eight independent replicates. (**$p<0.01$, *$p<0.05$).

(Figure 3.1B). Once collected, the data is filtered to select cells of uniform size and morphology. Next ‘non-zero’ GFP cells are selected and the peak of the GFP expression population density is calculated for each replicate (see Supplementary Methods for details, Figure 3.1C, Figure 3.2).

As a proof-of-concept we used two well-characterized RNA elements: the bacteriophage MS2 stem-loop binding site (MS2BS) that interacts with the MS2 coat protein (MS2) and the stem-loop binding element from the 3´ end of histone mRNAs (HBE) that binds the mammalian stem-loop binding protein (SLBP) (Johansson *et al.* 1998, Michel *et al.* 2000). Each binding site was tested against both RBPs to simultaneously assess PRIMA’s sensitivity and specificity. Quantification by flow cytometry showed that PRIMA could detect each test interaction with high specificity as only the cognate pairs resulted in increased GFP expression (Figure 3.1D, 1E).

We further assessed the sensitivity of PRIMA by introducing single nucleotide point mutations in the MS2BS that confer interactions with MS2 of 66 nM and 300 nM *in vitro* affinities (Johansson *et al.* 1998). As expected, the highest degree of GFP expression occurs with the original high affinity MS2BS (pM affinity). The 66 nM MS2BS also resulted in

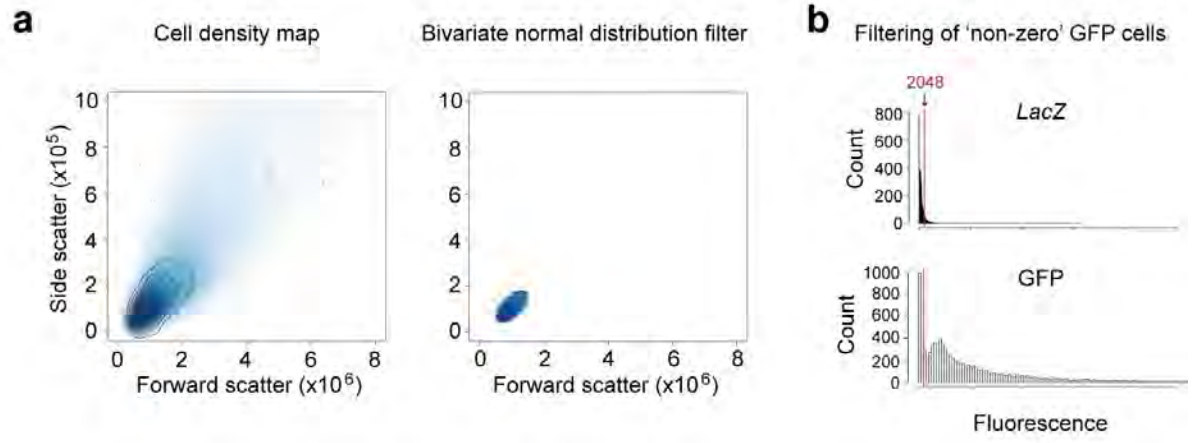
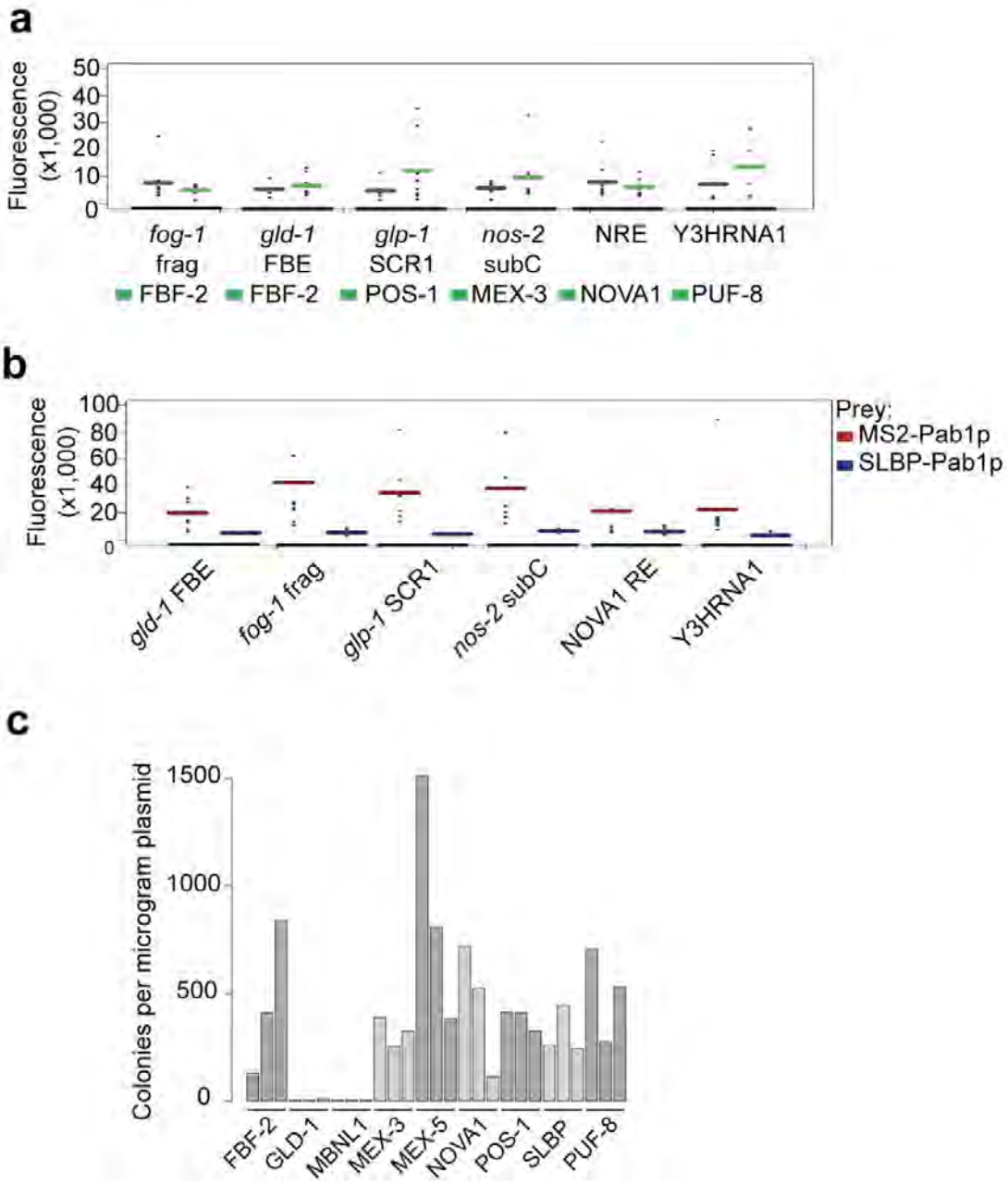


Figure 3.2 Data filtering example. A) Population density plot of forward scatter (FSC) *vs.* side scatter (SSC) values for cells collected by flow cytometry. Darker regions indicate greater numbers of cells. The two plots correspond to populations before and after the bivariate normal distribution filter is applied. B) Histograms of fluorescence intensities for *LacZ* (GFP-) and GFP with poly(A) tail (GFP+) baits. The lower fluorescence boundary for filtering 'zero' GFP cells was determined based on *LacZ* control populations.

elevated GFP levels, while the low-affinity interaction with the 300 nM MS2BS was not detected (Figure 3.1F).

Next, we tested several known interactions with RBPs from human and *C. elegans.* Initial attempts failed to induce high levels of GFP expression specifically (Figure 3.3A). There are several potential reasons for low sensitivity, including poor expression of the bait or prey, mislocalization of the RBP, for instance to the nucleus, or toxic effects of RBP expression. To address these issues, we first tested whether the RNA baits used were functional in PRIMA by introducing a high affinity MS2BS to the 3′ end. Co-expression of MS2-Pab1p increased GFP for all baits tested demonstrating that all baits are functional in PRIMA (Supplementary Figure 3.3B). Second, we tested the effects of RBP expression in *S. cerevisiae*. We obtained no or very few colonies upon transformation of GLD-1 and MBNL1-encoding plasmids, suggesting that expression of these RBPs is toxic to yeast (Supplementary Figure 3.3C). Third, we tested the functionality of the other RBPs by expressing them as an RBP-MS2-Pab1p fusion protein and introducing these constructs into the bait strain harboring a GFP reporter with a high affinity MS2BS (Supplementary Figure 3.3D). GFP was increased for all six RBP-MS2-Pab1p preys tested demonstrating that all RBPs are appropriately expressed and localized. Altogether, these results indicate


a
Fluorescence (x1,000)
50
40
30
20
10
0
fog-1 frag
gld-1 FBE
glp-1 SCR1
nos-2 subC
NRE
Y3HRNA1
FBF-2
FBF-2
POS-1
MEX-3
NOVA1
PUF-8
b
Fluorescence (x1,000)
100
80
60
40
20
0
Prey:
MS2-Pab1p
SLBP-Pab1p
gld-1 FBE
fog-1 frag
glp-1 SCR1
nos-2 subC
NOVA1 RE
Y3HRNA1
c
Colonies per microgram plasmid
1500
1000
500
0
FBF-2
GLD-1
MBNL1
MEX-3
MEX-5
NOVA1
POS-1
SLBP
PUF-8

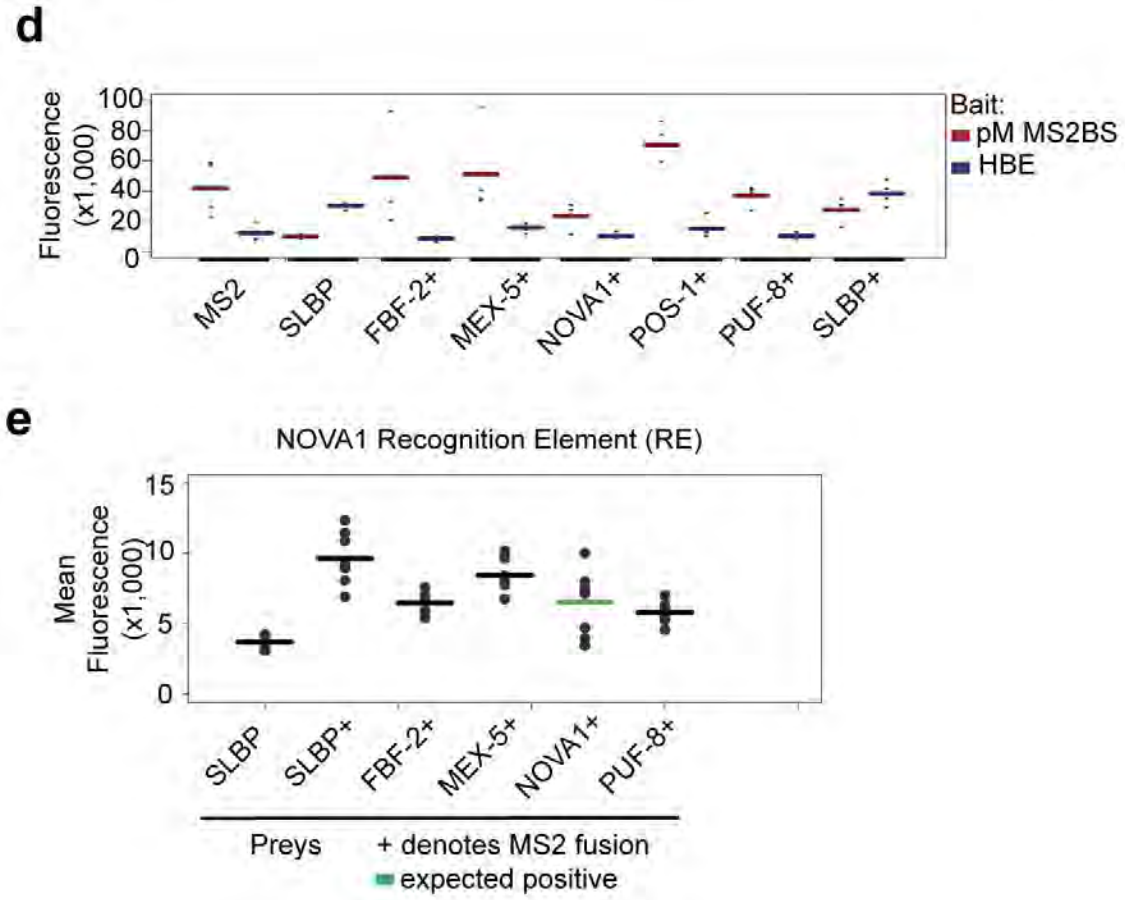

d
Fluorescence (x1,000)
100
80
60
40
20
0
Bait:
pM MS2BS
HBE
MS2
SLBP
FBF-2+
MEX-5+
NOVA1+
POS-1+
PUF-8+
SLBP+
e
NOVA1 Recognition Element (RE)
Mean Fluorescence (x1,000)
15
10
5
0
SLBP
SLBP+
FBF-2+
MEX-5+
NOVA1+
PUF-8+
Preys
+ denotes MS2 fusion
expected positive

Figure 3.3 PRIMA control experiments A) Six binding sites were tested against five RBDs. Individual replicates are shown in gray dots. Mean fluorescence levels for each RNA-RBD pair are shown as bars for SLBP-Pab1p (negative control, gray) and RBP-Pab1p (expected interaction, green). B) Test baits are functional in PRIMA. The pM MS2BS was fused to six RNA elements. PRIMA with the MS2-Pab1p prey resulted in large increases in GFP levels compared to control SLBP-Pab1p preys. C) Two prey plasmids are toxic to yeast. Equal amounts of prey plasmid DNA was transformed into the Y1H-aS2 yeast strain. Three independent replicates were performed and colonies were counted after three days of growth on Sc –Leu agar media. D) Fusion preys are functional in PRIMA. pM MS2BS and HBE baits were tested in PRIMA with RBD-MS2-Pab1p preys along with MS2-Pab1p and SLBP-Pab1p controls. The pM MS2BS bait specifically interacted with by all MS2-containing preys demonstrating their expression and localization. The HBE bait interacted with both SLBP-containing preys. E) The moderate 66 nM affinity MS2BS was fused to the NOVA1 Recognition Element (RE) binding site and tested as a bait *vs.* six preys. SLBP-Pab1p (SLBP) was a negative control and all other preys were tested as RBD-MS2-Pab1p fusions.

that, with the exception of the two toxic RBPs, all baits and preys are functional within the context of PRIMA. Therefore, we hypothesized that the cognate RBP-mRNA interaction affinities may be below the detection limits of PRIMA.

We reasoned that the sensitivity of PRIMA could be improved by including a high specificity, low affinity driver interaction adjacent to the test interaction. To test this, we introduced a moderate (66 nM) or low affinity (300 nM) MS2BS at the 3´ end of each bait RNA (Figure 3.4A). Additionally, we added an MS2 protein to the preys to create RBP-MS2-Pab1p fusion proteins. We fused the previously characterized SLBP and MS2 preys to test whether these modifications result in enhanced sensitivity. Indeed, dramatic increases in GFP production were obtained when the SLBP-MS2-Pab1p prey was tested with RNA baits that are located next to either a moderate or low-affinity MS2BS (Figure 3.4B).

Next, we re-assayed the set of known RNA-RBP interactions using the MS2 fusion strategy. The 300 nM low affinity MS2BS was fused to each RNA bait because this sequence did not show background binding in the presence of MS2-fused RBPs (Figure 3.3E). RNA-binding domains (RBD) were used in place of full-length RBPs to reduce potentials for steric hindrance. Additionally, bait constructs were integrated into the yeast genome to reduce cell-to-cell variability in bait RNA expression. Six

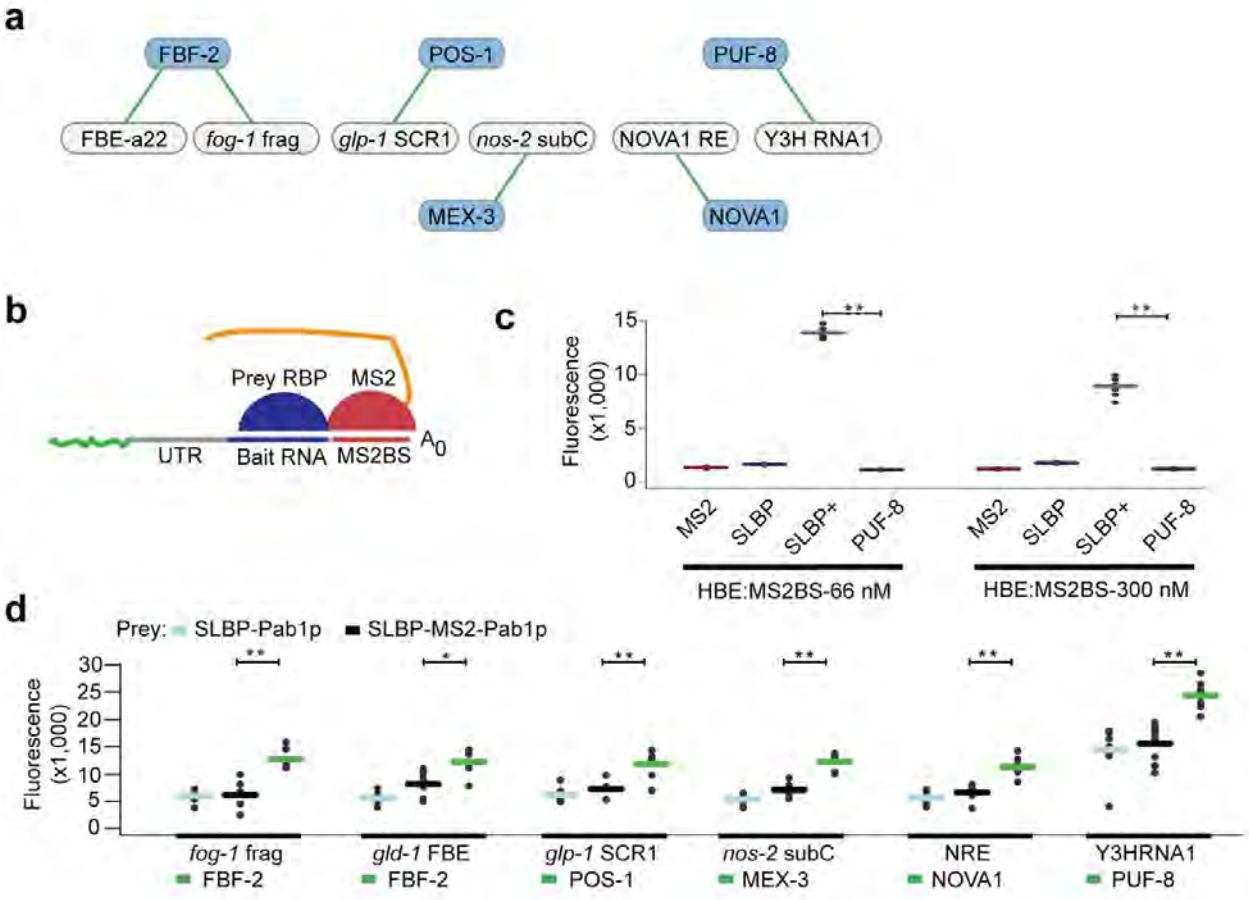

a
FBF-2
POS-1
PUF-8
FBE-a22
fog-1 frag
glp-1 SCR1
nos-2 subC
NOVA1 RE
Y3H RNA1
MEX-3
NOVA1
b
Prey RBP
MS2
UTR
Bait RNA
MS2BS
A0
c
Fluorescence (x1,000)
15
10
5
0
**
**
MS2
SLBP
SLBP+
PUF-8
MS2
SLBP
SLBP+
PUF-8
HBE:MS2BS-66 nM
HBE:MS2BS-300 nM
d
Prey: SLBP-Pab1p
SLBP-MS2-Pab1p
Fluorescence (x1,000)
30
25
20
15
10
5
0
**
*
**
**
**
**
fog-1 frag
FBF-2
gld-1 FBE
FBF-2
glp-1 SCR1
POS-1
nos-2 subC
MEX-3
NRE
NOVA1
Y3HRNA1
PUF-8

Figure 3.4 Known RNA-RBP interactions can be detected by PRIMA. A) RBDs were tested against for interactions with known binding sites using PRIMA. The binding sites were fused to the low affinity MS2BS and tested against RBD-MS2-Pab1p fusion proteins. Six binding sites were tested against five RBDs. B,C) Fusion baits containing both HBE and weak affinity MS2BS were tested against single RBP-Pab1p preys and SLBP-MS2-Pab1p (SLBP+) prey as a proof-of-concept. PUF-8-Pab1p is included as a non-binding negative control. D) Fluorescence levels for each RNA-RBD interaction. SLBP-Pab1p (SLBP) and SLBP-MS2-Pab1p (SLBP+) preys were negative controls for each bait. Bars indicate the mean fluorescence for all eight replicates. Positive interactions are shown in green (*$p<0.01$,**$p<0.001$).

different RNA baits were tested against an RBD prey (Figure 3.4A). SLBP-Pab1p and SLBP-MS2-Pab1p were included as negative controls for basal GFP expression and increases mediated by MS2 binding, respectively. Previously characterized interactions were detected for all six RNA baits (Figure 3.4D). Two sites, *gld-1* FBF binding element (FBE) and *fog-1* fragment, were bound by FBF-2 (Bernstein *et al.* 2005, Thompson *et al.* 2005). The *nos-2* subC fragment was bound by MEX-3 (Pagano *et al.* 2009, Jadhav *et al.* 2008). The NOVA-1 recognition element (RE) (Buckanovich *et al.* 1997) was bound by the human RBP and the previously characterized Y3HRNA1 fragment was bound by PUF-8 (Opperman *et al.* 2005). The *glp-1* SCR1 interaction with POS-1 was also detected in PRIMA (Farley *et al.* 2008, 2012). Overall this reference set demonstrates that PRIMA can detect different types of RNA-RBP interactions.

Next, we asked whether PRIMA can detect RNA–RBP interactions with full-length 3´UTRs as RNA baits. As baits we selected two *C. elegans* 3´UTRs, *nos-2* and *glp-1*, and tested these versus a mini-library of prey RBPs. *nos-2* and *glp-1* are expressed in the germline and early embryo of *C. elegans* (Crittenden *et al.* 1994, Jadhav *et al.* 2008). Therefore 40 RBPs were selected from previously published RBP annotations (Tamburino *et al.* 2013) that are highly expressed in the germline (Wang *et al.* 2009). The

mini-library included several well-characterized RBPs including POS-1 and PUF-5, which bind and regulate *glp-1* (Farley *et al.* 2008, Hubstenberger *et al.* 2012), and MEX-3 which binds and regulates *nos-2* (Pagano *et al.* 2009).

PRIMA detected five of six known interactions, including the interaction between *nos-2* and MEX-3 (Figure 3.5A). Further, we recapitulated the binding of POS-1 to *nos-2* (Jadhav *et al.* 2008) (Figure 3.3a). Additionally, we recapitulated three well-characterized interactions between POS-1, PUF-5 and MEX-3 with *glp-1* (Pagano *et al.* 2009, Farley *et al.* 2012, Hubstenberger *et al.* 2012, Ogura *et al.* 2003) (Figure 3.5B). The interaction between MEX-5 and *glp-1* was not observed in PRIMA (Pagano *et al.* 2007).

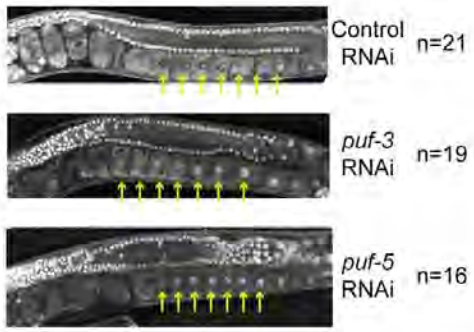
PRIMA also detected novel interactions. MEX-1, C12D8.1, GRLD-1, and DAZ-1 bound to the *nos-2* 3´UTR. MEX-1and C12D8.1 bound to *glp-1* 3´UTR (GRLD-1 and DAZ-1 were just below the threshold.) Preliminary PRIMA results with additional baits suggest that these preys bind many UTRs and are likely non-specific RBPs (data not shown). In comparison, HRP-1 and R09B3.2 bound specifically to the *nos-2* 3´UTR. Additionally, FBF-1, FBF-2 and PUF-3 bound specifically to the *glp-1* 3´UTR.

Figure 3.5 A,B) Specific interactions were detected for two full-length 3´UTRs. Two sets of 8 biological replicates were measured for each prey. The fluorescence intensity at the peak was measured for each and the two highest and two lowest samples were removed. The remaining 12 replicates and the average intensity for each prey is shown. Preys with

c

Bait	Prey	Score	p-value	Change in GFP pattern
nos-2	MEX-3	1.38	<0.001	Increase in embryos
nos-2	POS-1	1.36	<0.001	No change
nos-2	HRP-1	1.31	<0.001	No change
nos-2	R09B3.2	1.2	<0.001	No change
glp-1	FBF-2	2.04	<0.001	No change
glp-1	PUF-3	1.8	<0.001	Increase in oocytes
glp-1	PUF-5	1.44	<0.001	Increase in oocytes
glp-1	FBF-1	1.39	<0.001	No change
glp-1	POS-1	1.24	<0.003	Increase in posterior embryo

d



Pmex-5::PEST::GFP::glp-1 3'UTR

average intensity >1.20 fold compared to negative control are shown in green ($p<0.01$). Preys are labeled on the x-axis and include the fusion of MS2 to the prey (except for SLBP:Pab1p). A) The *nos-2* 3´UTR is significantly bound by eight RBP preys and B) the *glp-1* 3´UTR is bound by seven RBP preys. C) A subset of bound RNA preys were tested *in vivo* for regulation using RNAi of RBPs in transgenic *C. elegans* strains wherein PEST::GFP:H2B:3´UTR mRNA was expressed from the *mex-5* constitutive germline promoter. D) *puf-3* and *puf-5* RNAi result in increased expression

of GFP::*glp-1* 3′UTR in developing oocytes. GFP is localized to the nucleus via the H2B fusion and destabilized via the PEST domain.

We next tested the potential regulatory consequences of the 3'UTR-RBP interactions *in vivo* (Figure 3.5C, D). Single copy reporter transgene PEST::GFP::H2B mRNA containing either the *nos-2* or *glp-1* 3´UTR was expressed from the *mex-5* promoter that drives in the *C. elegans* germline. GFP expression was observed in the distal germline of both strains, while proximal expression in the oocytes is low, as previously described (Pagano *et al.* 2009, Farley *et al.* 2012). GFP levels increased in the developing oocytes of the *glp-1* 3´UTR strain following RNAi of its interactors *puf-3* and *puf-5* (Figure 3.5D). Additionally, GFP levels increased in the posterior cells of the 4-cell stage embryo of the *glp-1* 3´UTR strain following RNAi of its interactor *pos-1* as previously observed (Farley *et al.* 2012) (Figure 3.6). Furthermore, GFP levels increased in the early embryo of the *nos-2* 3´UTR strain following RNAi of its interactor *mex-3* as previously observed (Pagano *et al.* 2009) (Figure 3.6). RNAi was performed against additional PRIMA positives that did not result in a change in the pattern of GFP expression (Figure 3.5C).

In summary, we have developed and validated PRIMA, a novel protein-RNA interaction mapping assay that can be used with small RNA elements, as well as with full-length 3´UTRs. The future development of

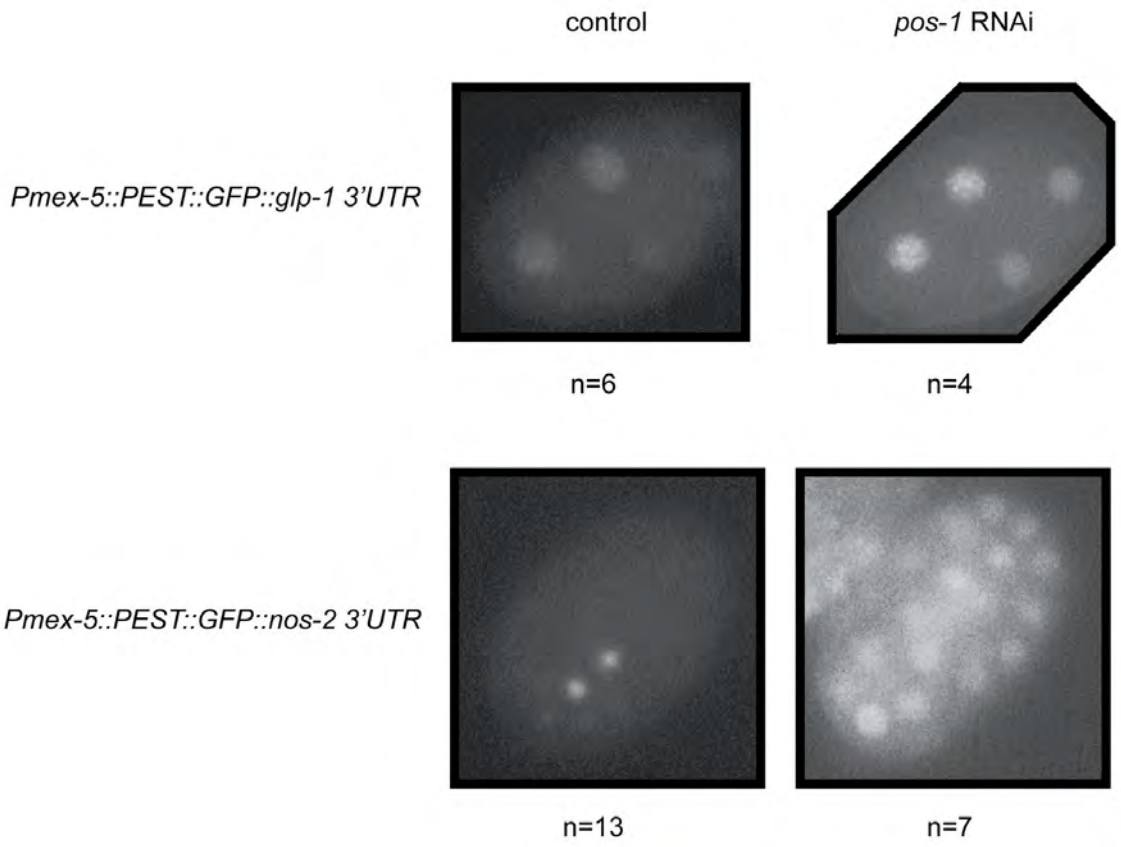


Figure 3.6 The *glp-1* and *nos-2* 3´UTRs are regulated by RBPs in the embryo. Fluorescence images of germlines from Pmex-5::GFP::H2B::*glp-1* 3´UTR and Pmex-5::GFP::H2B::*nos-2* 3´UTR strains are shown. GFP is increased in the posterior cells of the 4-cell stage embryo of GFP::H2B::*glp-1* 3´UTR strain following RNAi of *pos-1.* GFP is increased in the GFP::H2B::*nos-2* 3´UTR strain following RNAi of *mex-3*.

larger RBP prey libraries will enable the testing of full-length 3´UTRs against the hundreds of RBPs .

Methods

Cloning of RNA elements and RBPs

All DNA sequences and plasmid configurations used in this manuscript are available in Figure 3.7.

The pADH1:GFP:*unc-54*:MCS:Ribozyme plasmid expression vector was generated using sequential PCR stitching and gap repair of DNA constructs (Orr-Weaver *et al.* 1983) into the pDest22 backbone (Life Technologies). The S65T GFP sequence was amplified from pFA6:GFP (kindly provided by Paul Kaufman). The shortest *unc-54* 3´UTR isoform is included in all RNA baits. It was amplified from the 3´UTRome entry vector (Mangone *et al.* 2010). The multiple cloning site (MCS) and hammerhead ribozyme were generated synthetically (Life Technologies). Binding sites were inserted into the multiple cloning site (MCS) of the expression vector using yeast gap repair of synthetic oligos into AflII/SmaI (NEB, R0520S/R0141S) or AflII/ClaI (NEB, R0520S/R0197S) digested vectors.

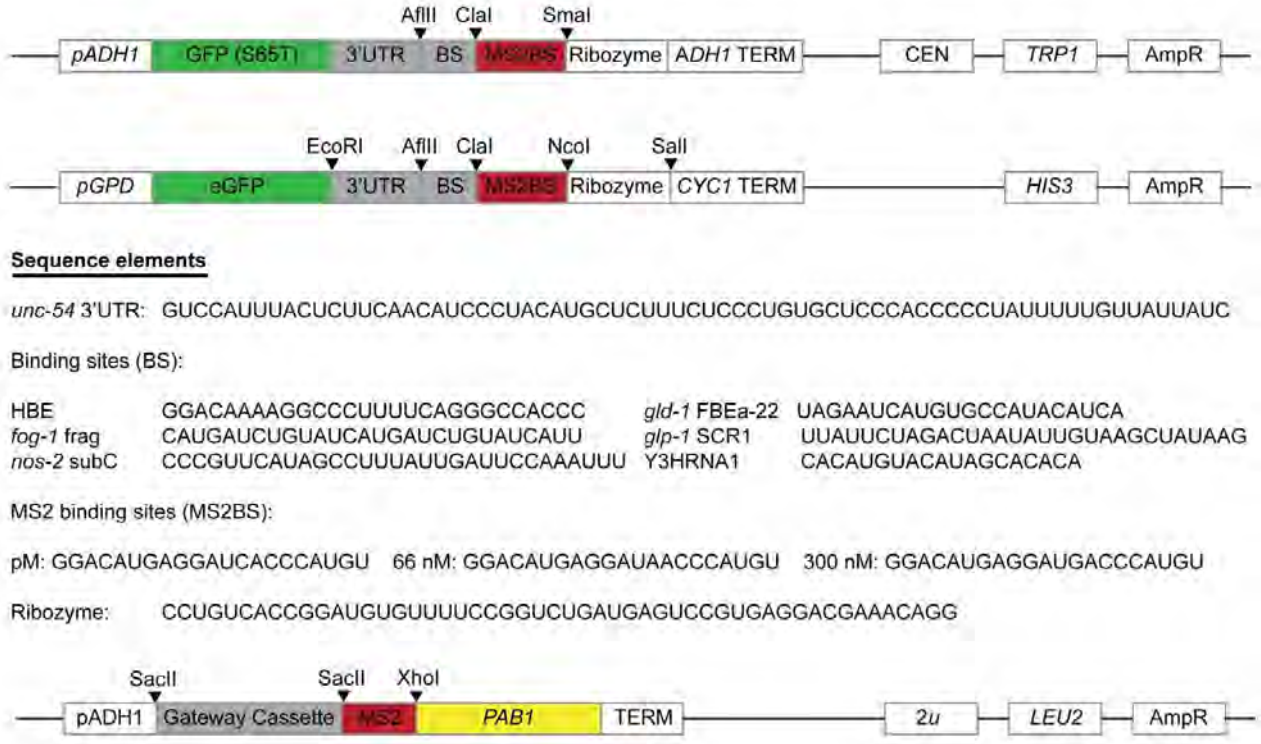


Figure 3.7 DNA sequences and plasmid configurations used in this manuscript

The pGPD:eGFP:*unc-54*:HBE:Stem-loop:Ribozyme integration expression vector was generated from pAG303GPD-EGFP-ccdB (Alberti *et al.* 2007) by inserting the 3′ end of pADH1:GFP:*unc-54*:HBE:Stem-loop:Ribozyme vector (this work) into the NotI/SalI (NEB, R3189S/R3138S) fragment. Additional RNA element constructs were generated by replacing the AflII/ClaI fragment with synthetic oligos. 3′UTR constructs were generated by replacing the EcoRI/ClaI fragment with PCR products amplified from *C. elegans* cDNA.

The pDest Pab1p vector was generated using a Gateway cassette PCR product amplified from pGBKCg (Stellberger *et al.* 2010) using Platinum HiFi Taq (Invitrogen, 11304-011) and TA cloned into pGEMT (Promega, A3600). The SacII/XhoI digested product was ligated into the SacII/XhoI site of YCplac111-MS2–Pab1p (Amrani *et al.* 2004) (kindly provided by Allan Jacobson). The pDest MS2 Pab1p vector was generated similarly using a separate SacII/SacII product ligated into the SacII site of YCplac111-MS2–Pab1p.

RBDs were determined according to the literature (Supplementary Table 1) or using InterProScan software (Jones *et al.* 2014). Domains determined using InterProScan were extended by 30 residues on both ends. Primers were designed using Primer3Plus (Untergasser *et al.* 2007) with one additional nucleotide on both ends of the RBD (to maintain frame).

Gateway B1 and B2 tails were included on the forward and reverse primers, respectively. Gateway reactions were performed as previously described (Walhout *et al.* 2000).

Yeast manipulations and assay conditions

All assays were performed using the Y1H-aS2 yeast strain (Reece-Hoyes *et al.* 2011). Plasmid expressed baits were generated by yeast transformations as previously described (Walhout and Vidal 2001) and plated on synthetic complete (Sc) -Trp agar media. Integrated baits were generated by transformation of yeast with NheI (NEB, #R3131S)-digested plasmids plated on Sc -His agar media. PRIMA assay strains were generated by yeast transformations of RNA-element harboring strains with individual prey plasmids plated on Sc -Leu, -Trp (plasmid baits) or Sc -Leu, -His (integrated baits). Individual colonies were picked and frozen at -80^{o} C in 20 % glycerol prior to performing the assay.

Assays were performed as follows: Thawed yeast strains were inoculated in 200ul appropriate Sc liquid media in 96 deep well plates and grown overnight at 30^{o} C with 200 rotations per minute (RPM) agitation. 10ul of overnight culture was diluted into 1 mL of fresh media and grown to log phase (~6.5 h). Cultures were centrifuged at 2,000 RPM for 3 min and

resuspended in 400ul of 1X Phosphate Buffered Saline (PBS). Individual cells were then measured using a BD Accuri C6 flow cytometer using the 510/15 FL1 emission filter according to manufacturer's protocols.

Data processing and quantitative scoring

The standard flow cytometry data files (FCS3.0) were exported from BD Accuri C6 software and analyzed using custom R project software (http://www.R-project.org/) and the FlowCore and FlowViz packages. Briefly, forward scatter (FSC), side scatter (SSC) and fluorescence (FL1) measurements were imported for each sample. A lower FSC cutoff of 240,000 was applied as it corresponded to cellular debris (data not shown). A uniform cell population (~50% of the population) was selected using the FSC and SSC vectors and the norm2Filter function with scale factor=1. Briefly, the norm2filter function fits a bivariate normal distribution to the dataset and selects data points according to their standard deviation from the fit.

The resulting cells were plotted as fluorescence (FL1) *vs.* cell count and the two clear peaks were observed for nearly all cell populations. The low fluorescence peak overlapped with GFP-minus (*LacZ*) control yeast, indicating that zero GFP expression was present. The high fluorescence

peak overlapped with GFP+ control yeast with poly(A) tails. We selected all ‘non-zero’ GFP cells by using a lower FL1 cutoff of 2048 which corresponded to the upper bound of GFP- control yeast. The population density was smoothed using a kernel density estimate. The peak of the density was determined for each sample. 8 replicates were tested for the initial experiments with the MS2BS, HBE, and RBP binding site baits (Figs. 1 and 2). 16 replicates (two sets of eight) were collected for each 3´UTR bait and the two highest and two lowest values were removed. The average was calculated for the remaining 12 replicates from each bait-prey pair. The average fluorescence for each test prey was compared to the average SLBP:MS2:Pab1p negative control. Test preys with >1.20 fold increase in fluorescence were considered positive provided they were statistically significant ($p<0.01$, student’s t-test).

RNAi and imaging of worm strains

Knockdowns were performed using the RNAi feeding method as described (Kamath *et al.* 2003). The RBD entry clones were cloned into the RNAi feeding vector construct L4440 using Gateway reactions and transformed into HT115(DE3) cells. The transformed colonies were grown to OD600 = 0.4 and induced with isopropyl 1-thio-β-D-galactopyranoside

(IPTG) at a final concentration of 0.4 mM for 4 hours. After induction the 50 ml cultures were concentrated 10- fold and 50 µl of the culture was added onto NGM plates containing 1mM IPTG and 100 µg/ml Ampicillin. After bleaching adult worms, in 0.5N NaOH and 2% clorox, eggs were washed once with distilled water and plated onto these plates and kept at 25° C for 2 days before imaging. HT115 strain bacteria transformed with the empty vector L4440 was used as the control RNAi.

Adult worms were placed in 0.4 mM levamisole on to 2% agarose pads before imaging. Embryo dissections were done in M9 solution and dissected eggs were mounted on 2% agarose pads. DIC and GFP fluorescence images were taken on Zeiss Axioscope 2 plus microscope (Carl Zeiss, Jena, Germany) using an oil-immersion 40X objective. Confocal images were taken under 40X magnification using Leica DMIRE2 microscope (Leica, Wetzlar, Germany) using 488 nm excitation at a 100% intensity. A single section was imaged for each worm and the line average for the scan was 16.

CHAPTER IV

Technical variations in the development of PRIMA

The PRIMA development process was designed as a series of sequential milestones (Figure 4.1A). The PRIMA concept was based upon naturally occurring processes that were learned through a directed literature search. Next, proof-of-concept experiments were designed to test the hypothetical model. Experimental processes were established including: cell growth conditions, reporter choice, detection methods and reagent (vector) generation (Figure 4.1B). Next, PRIMA technological boundaries were tested (Figure 4.1B) including the range of affinities resulting in detectable signals, the best types of yeast (mating type and strain), and methods for processing and scoring data. Chapter IV outlines crucial experiments which are not reflected in Chapter III of the thesis.

Reporter choice

Numerous reporters can be used in yeast. Selection based reporters enable yeast to grow in the absence of a particular nutrient. These include *HIS3* which is necessary for histidine biosynthesis in yeast

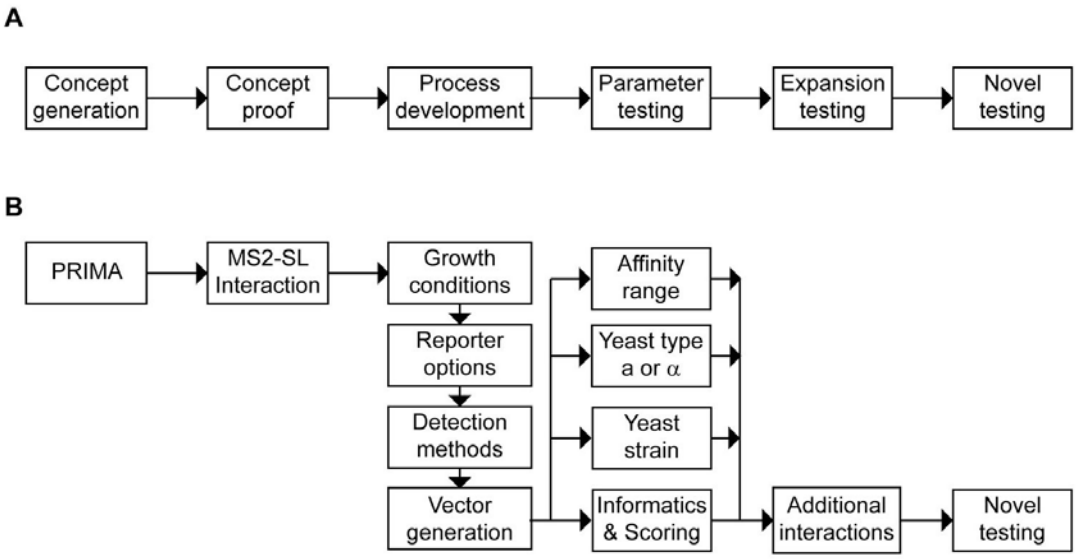


Figure 4.1 PRIMA development flow chart A) PRIMA development was designed to occur in serialized fashion from concept generation through the testing of novel interactions. B) PRIMA was developed in discrete stages during which numerous objectives and milestones were achieved.

and has been deleted from many laboratory yeast strains. When deleted, yeast cannot grow without exogenously supplied histidine. Thus *HIS3* can be reintroduced to the yeast under a new context where it reports on a condition. When the condition occurs yeast growth results and the amount of growth can be used as a qualitative or quantitative measure of reporter signal. Other nutrient based reporters include *TRP1* (tryptophan biosynthesis), *LEU2* (leucine biosynthesis), *ADE2* (adenine biosynthesis) and *URA3* (uracil biosynthesis). The main advantage of *HIS3* versus other related reporters is that low or leaky levels of reporter expression can be controlled by the addition of 3-aminotriazole (3-AT), a competitive inhibitor of the His3p enzyme. The concentration of 3-AT can be adjusted according to level of background expression.

Enzymatic reporters including *LacZ* are also frequently used (Deplancke *et al.* 2004). *LacZ* is an *Escherichia coli* gene encoding beta-galactosidase. When yeast containing the *LacZ* gene are exposed to the colorless 5-bromo-4-chloro-3-indolyl-β-d-galactoside (X-gal), a blue compound is generated in an amount proportional to the level of the reporter expression. The blue compound can be measured qualitatively by eye, or images of the yeast colonies can be analyzed for quantitative assessments (Reece-Hoyes *et al.* 2011).

Lastly, fluorescence based reporters including green fluorescent reporter (GFP) are commonly used. Yeast expressing GFP are excited by a 488 nm laser and the signal emitted at higher wavelengths (typically 509 nm +/- 10) is detected. The signal can be detected by a fluorescence microscope or measured quantitatively by fluorescence flow cytometer.

All three reporters were considered for use in PRIMA each with distinct advantages compared to the others. The main advantage of *HIS3* and *LacZ* is that readout can be conducted by eye following yeast colony growth on appropriate agar media. Additionally, many techniques for robotic handling of yeast and automated image analysis have been established enabling high-throughput data collection and analysis (Reece-Hoyes *et al.* 2011). The primary advantage of the GFP reporter is that existing methodologies enable the measurement of precise signals in single cells. Additionally, high-throughput technologies exist including automated fluorescent microscopy and flow cytometry.

The *LacZ* reporter gene was tested in PRIMA using the known high affinity interaction between the bacteriophage MS2 coat protein (MS2) and its stem-loop RNA binding site (MS2BS). *LacZ* expression was assayed using filter assays. Briefly, yeast were spotted onto nitrocellulose filters placed on synthetic yeast agar media and grown overnight at 30 C. Next, the yeast were freeze cracked by placing the filter paper in liquid nitrogen

for 15 seconds. The filter was incubated for 1 hour in buffer containing X-gal. High levels of *LacZ* expression were detected when the MS2-Pab1p prey was tested against the MS2BS bait (Figure 4.2A). Low levels of *LacZ* expression were detected when the MS2BS was absent or when a control MS2-MS2 protein was tested.

His3p reporter expression was assayed using a growth assay. Briefly, yeast were spotted on synthetic agar medium which was permissive for growth (*i.e.* supplemented with histidine) and grown overnight. Next, the yeast were transferred onto selective media (*i.e.* lacking histidine, plus 40 mM 3-AT) using a velvet cloth intermediary. Yeast were then grown at 30 C for 48 hours and imaged. High levels of growth were detected when the MS2-Pab1p prey was tested against the MS2BS bait, but not when either component was missing (Figure 4.2B).

Both reporters effectively detected the high affinity MS2BS interaction with MS2, however they were not able to clearly discriminate lower affinity interactions. The MS2-MS2BS affinity was reduced by making a single nucleotide mutation which lowered the pM affinity to 66 nM. Clear differences were observed for the pM interaction, however, no difference was observed for the 66 nM interaction (Figure 4.2C).

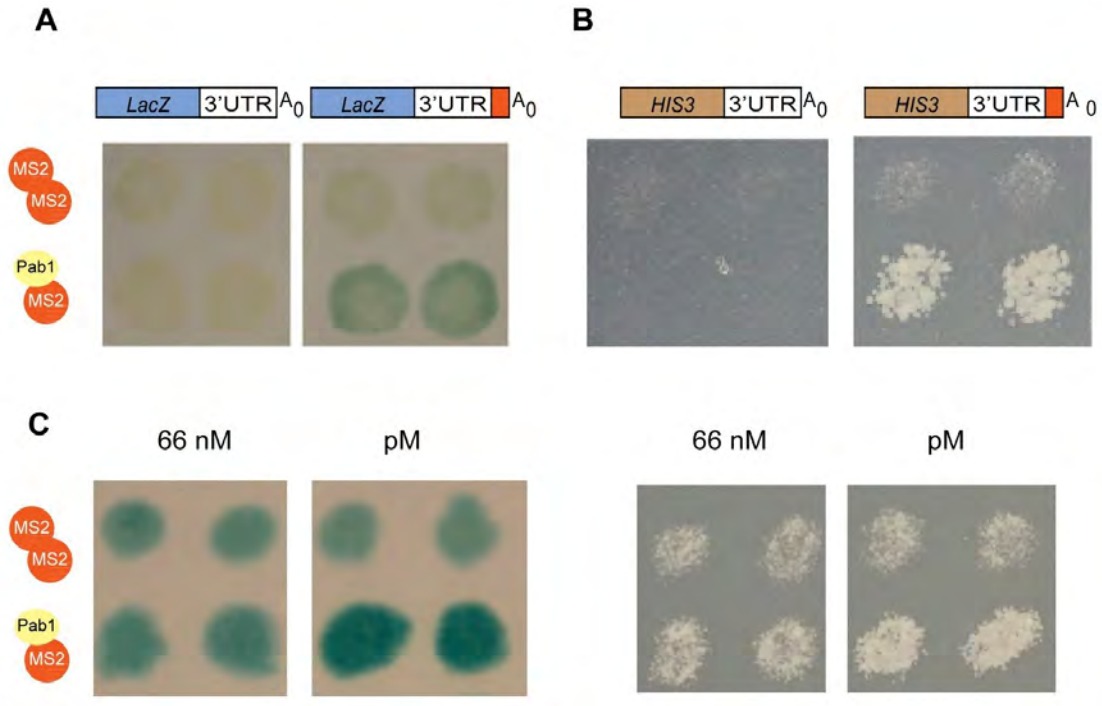


Figure 4.2 The *LacZ* and *HIS3* reporters were tested in PRIMA. A) *LacZ* mRNAs containing the minimal *unc-54* 3'UTR were tested as baits with and without the high affinity MS2BS against the MS2-MS2 negative control and the MS2-Pab1p preys. Colorless X-gal was converted to a blue compound only when the bait containing the MS2BS was tested against the MS2-Pab1p prey. B) *HIS3* mRNAs containing the minimal *unc-54* 3'UTR were tested as baits with and without the high affinity MS2BS. Yeast was able to grow only when the bait containing the MS2BS was tested against the MS2-Pab1p prey. C) Moderate affinity MS2BS (66 nM) baits failed to produced clear signals from *LacZ* or *HIS3* reporter baits.

From these experiments it was clear that PRIMA's dynamic range was low and therefore precise quantitative measurements would be necessary for the robust signal measurements. GFP reporter expression was ultimately chosen because of the ability to assay expression using multiple different quantitative platforms, which are outline below.

Detection method

There are two primary methods for the capture and evaluation of fluorescent signals: microscopy and flow cytometry. The MS2BS interaction with MS2 was evaluated using both techniques. Initial tests used fluorescence microscopy for GFP detection. Ultimately PRIMA was adapted to fluorescence cytometry for reasons that are detailed below.

Evaluation of the MS2BS interaction with MS2 using microscopy resulted in clear signal detection. Briefly, yeast colonies were grown at 30 C in synthetic liquid media overnight to saturate cultures. On the next day, cultures were diluted in fresh media to an OD600 of 0.1 and grown for 6.5 hours to reach log phase. Cultures were then centrifuged at 2,000 RPM for 3 minutes, resuspended in 1X PBS and spotted on a glass slide with coverslip. Images were taken using 630x magnification and 553 msec exposure. A reporter containing the MS2BS showed low GFP expression is

most cells when tested against an MS2-MS2 prey (Figure 4.3A). In contrast, a reporter bearing the MS2BS and an MS2-Pab1p prey resulted in an increase in GFP signal that was detectable by eye.

Next, the microscopy approach was adapted for quantitative measurements of GFP signal. Four images were taken capturing >300 cells for each sample. Cell boundaries were identified automatically using CellProfiler (Lamprecht *et al.* 2007) (Figure 4.3B). The median fluorescence intensity was calculated for each cell and the population distribution was plotted for each sample (Figure 4.3C). A two-fold increase relative to controls results when both MS2BS and MS2-Pab1p are present. Additionally, the interaction showed comparable signal to a GFP mRNA that is polyadenylated and transcribed from the same promoter.

Fluorescence cytometry provides an alternative method for quantitative measurement of GFP expression. Preparation of the cells is identical to the microscopy method. In cytometry, the samples are aspirated through a narrow nozzle and a laser light source excites the cells as they pass. Forward scattered (FSC) and side scattered (SSC) light are detected. These correspond to cell size and cell granularity, respectively. Both measurements are useful parameters for filtering outliers from the sample population. For instance, a lower FSC boundary

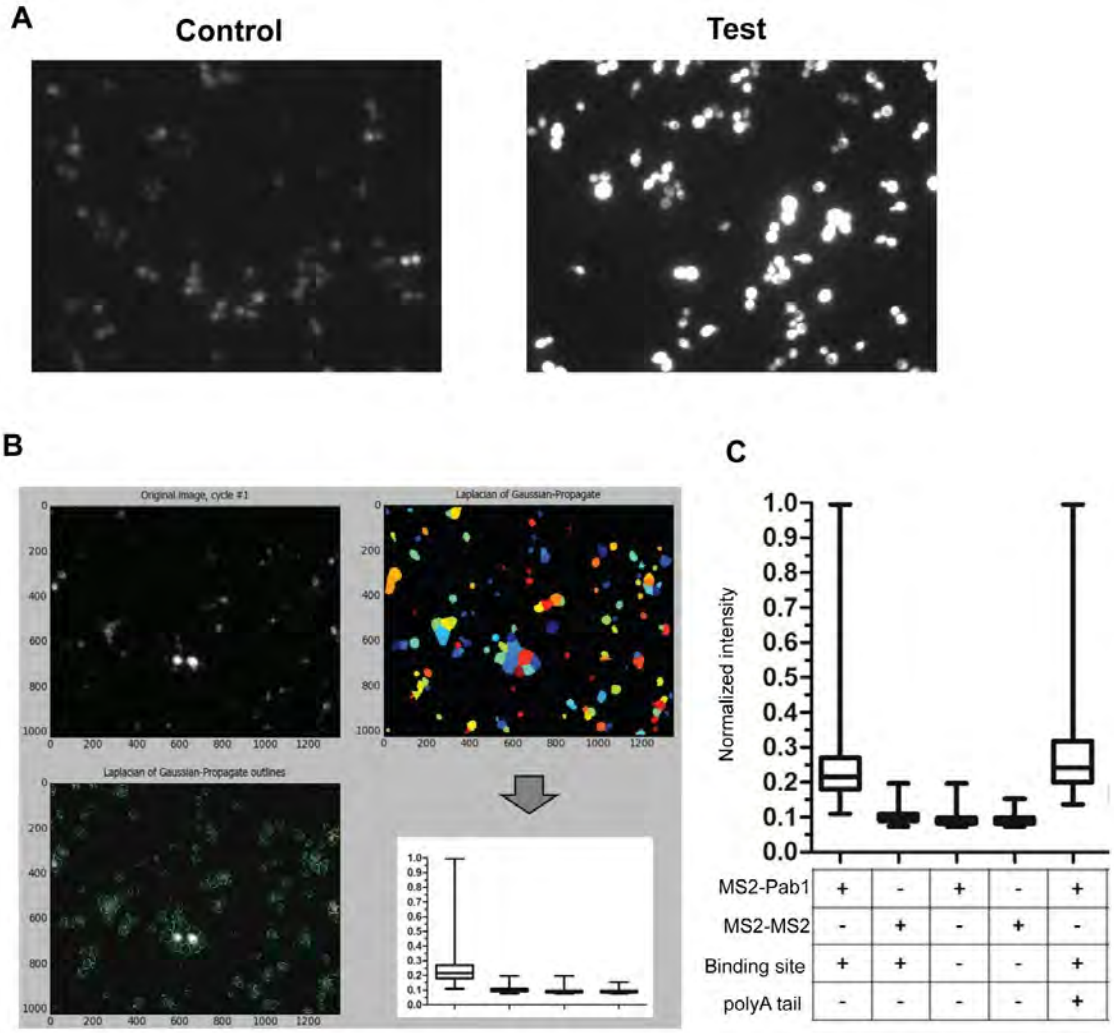

A
Control
Test
B
Original image, cycle #1
Laplacian of Gaussian-Propagate
Laplacian of Gaussian-Propagate outlines
C
Normalized intensity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
MS2-Pab1 + - + - +
MS2-MS2 - + - + -
Binding site + + - - +
polyA tail - - - - +

Figure 4.3 Detection of GFP using fluorescence microscopy A) Control yeast expressing a GFP bait containing the MS2BS showed low levels of GFP in most cells via fluorescence microscopy when tested against an MS2-MS2 prey. Test yeast expressing the same bait showed qualitative increases in GFP when tested against an MS2-Pab1p prey. B) CellProfiler was used to automatically determine yeast cell boundaries from a given fluorescence microscopy image. The median GFP expression for each cell was determined and the 256-bit pixel intensity was normalized to 1. The total population was represented as a box and whiskers plot. C) Box and whisker plots for each bait and prey combination.

can be set to remove small signals that typically correspond to debris. A fluorescence level is detected for each cell above the lower FSC boundary. The fluorescence level is measured across the entire length of the cell as it passes the detector. Therefore the quantified signal for each cell can be represented as the peak cellular intensity, or as the integrated cellular intensity. For PRIMA, the integrated cellular intensity was chosen because it captures the total GFP levels within the cell.

Flow cytometry data can be presented and evaluated in several different ways. First, fluorescence signal can be plotted against FSC wherein each point of the scatter plot corresponds to an individual cell (Figure 4.4A). This qualitative assessment shows that the MS2BS interaction with MS2-Pab1p results in more fluorescence than its interaction with MS2-MS2. Second, the distribution of cellular intensities can be evaluated (as with microscopy) (Figure 4.4B). This quantitative analysis shows that the median fluorescence increases two-fold when MS2BS is tested against MS2-Pab1p compared with MS2-MS2. Lastly, a threshold can be applied to the population (Figure 4.4C). In this approach, a threshold can be established which is exceeded by 1% of the negative control population. The same threshold is applied to the test population and the proportion of cells exceeding that threshold is measured. This

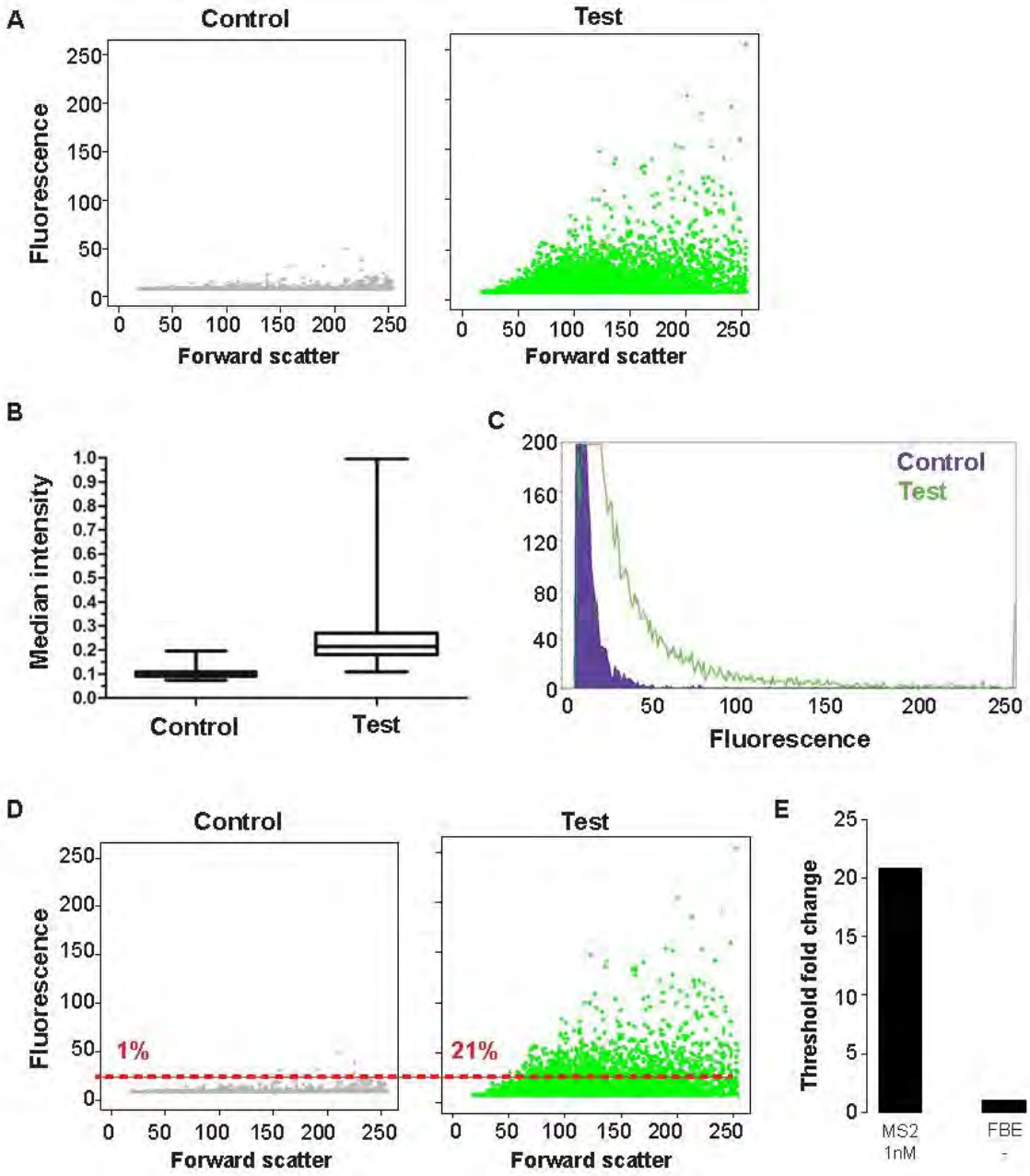

A
Control
Test
Fluorescence
Forward scatter
B
Median intensity
Control
Test
C
Control
Test
Fluorescence
D
Control
Test
1%
21%
Fluorescence
Forward scatter
E
Threshold fold change
MS2
1nM
FBE
-

Figure 4.4 Detection of GFP using flow cytometry A) Scatterplots of flow cytometry data showing the forward scatter and fluorescence measurements of each yeast cell in the population. Data is shown in 256-bit format. The control population containing the MS2BS shows low levels of GFP expression when tested against the MS2-MS2 prey. The test population shows increased levels of GFP expression in many cells when tested against the MS2-Pab1p prey. B) Box and whisker plots of the same data. C) Histogram of the same data. D) A threshold is determined in the control population at which 1% of the cells are above it. The threshold is then applied to the test population and the proportion of cells above this threshold is measured. E) Bar graph depicting the same data.

analysis shows that the MS2BS interaction with MS2-Pab1p exceeds the signal generated by MS2-MS2 by 21-fold.

Flow cytometry is the preferred method for PRIMA. The principal advantage over microscopy is that more cells can be measured in less time. A typical 63x image field contains 50-100 yeast cells. An automated microscope capable of taking two images per second would require 250 seconds to acquire images totalling 50,000 cells. In comparison, an automated flow cytometer running an average acquisition rate of 2,000 cells per second would require 25 seconds to acquire the same data. Additionally, flow cytometry does not require the added time and computation of image analysis and increased demands of image storage.

Yeast mating type and strains

Different yeast genetic backgrounds have been known to affect the interactions detected by heterologous assays such as yeast one- and yeast two-hybrid (Braun *et al.* 2009, Reece-Hoyes *et al.* 2011). This includes different forms of yeast (*i.e.* haploid or diploid) and different strains of yeast. The MS2BS interaction with MS2-Pab1p was tested in both forms of yeast and in four different genetic backgrounds to observe the effects of genetic background on PRIMA.

The MS2BS interaction with MS2-Pab1p was tested in haploid and diploid yeast. The same haploid ‘a’-type bait strains containing the MS2BS or no binding site were used for both sets of experiments. Haploid yeast were tested by transforming the bait strain with individual prey plasmids and selecting colonies which grew on synthetic yeast media lacking tryptophan and leucine. Diploid yeast strains were generated by mating ‘a’-type bait strains with ‘α’-type prey strains harboring different prey plasmids. Matings were conducted on permissive yeast media overnight followed by transfer to media lacking tryptophan and leucine for selection of diploid yeast with both plasmids.

Both yeast forms were capable of detecting the MS2BS interaction with MS2-Pab1p. There was an increase in the number of cells expressing high levels of GFP in both yeast forms (Figure 4.5A). Additionally, there was an increase in the median GFP expression and the threshold fold change for the interaction in both yeast forms (Figure 4.5B, C). Haploid yeast showed a larger increase in signal using both measures. Preparation of haploid yeast for each experiment (*i.e.* haploid transformation) is more time consuming than preparation of diploid yeast (*i.e.* mating) and it is less amenable to high throughput techniques. Overall the advantage of increased dynamic range was deemed outweigh the workload and haploid yeast were chosen for PRIMA.

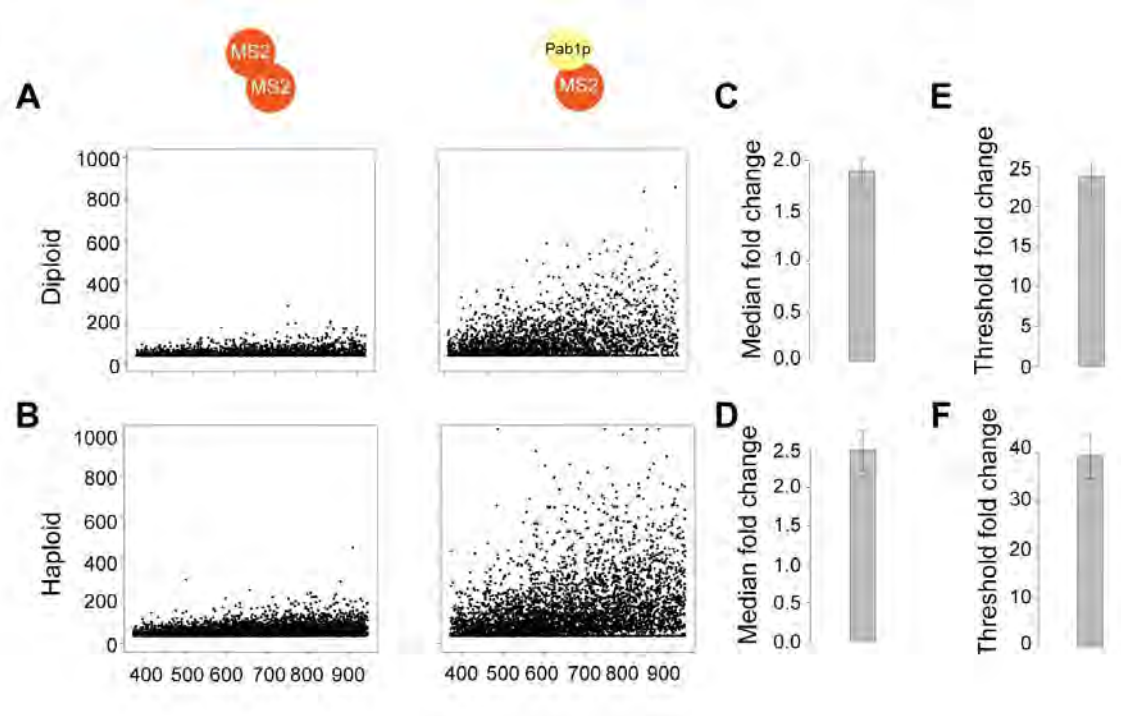


Figure 4.5 The use of diploid and haploid yeast A,B) The diploid and haploid yeast are both functional types for PRIMA. The diploid yeast have slightly reduced background in the MS2-MS2 control population relative to haploid. However, the haploid yeast have higher signal in the MS2-Pab1p test population. C,D) Median fold change for the diploid and haploid tests. E,F) Threshold fold change for the diploid and haploid tests.

The MS2BS interactions with MS2-Pab1p were tested using four different strains of yeast (Figure 4.6). Three different baits were tested including the pM MS2BS, the 66 nM MS2BS and a non-binding *fem-3* binding element (FBE). Four different yeast genetic backgrounds were tested as baits versus three different yeast genetic prey backgrounds in diploid format. This enabled a larger number of combinations to be tested. No clear differences were observed between the different bait-prey yeast combinations and therefore the original Y1HaS2 strain was retained as the PRIMA mother strain.

Conclusion

The technical iterations outlined in this chapter were critical steps in the PRIMA development process. These steps led to the choice of reporter, detection method and yeast mating types and strain.

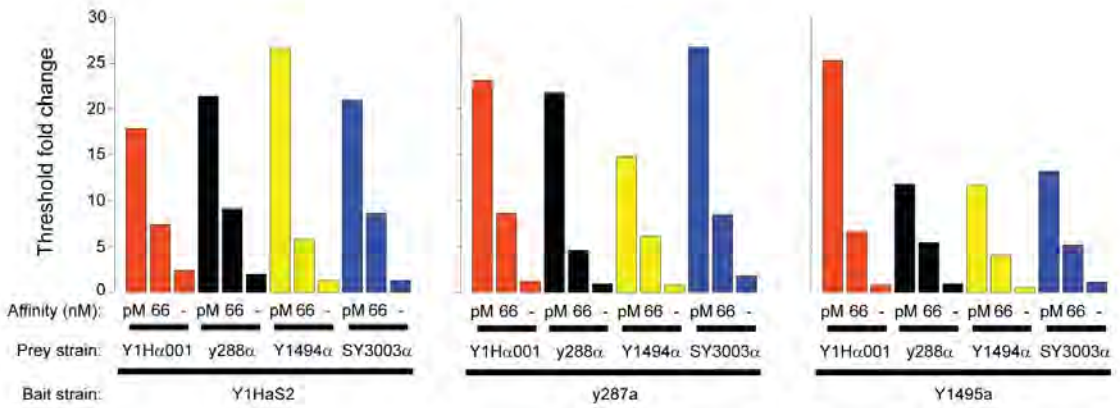


Figure 4.6 The use of different strains. Three different ‘a’ type bait yeast strains were mated to four different ‘α’ type yeast strains and tested in PRIMA. The pM, 66 nM MS2BS baits and a non-binding (FBE) bait were tested against the MS2-MS2 and MS2-Pab1p preys. The relative threshold fold change for each prey-prey pair is plotted for all 36 pair-wise combinations.

CHAPTER V

Discussion

The Protein-RNA Interaction Mapping Assay was successfully developed during this thesis project. PRIMA was designed and initially developed using known interactions between the bacteriophage MS2 coat protein (MS2) and its known binding site (MS2BS). Experiments were then expanded to include the interactions between RNA elements and several proteins from humans and *C. elegans.* PRIMA was demonstrated to recapitulate known interactions between a 3'UTR and *C. elegans* RBPs. Lastly, novel interactions were discovered including *in vivo* regulatory interactions.

There are four aspects of PRIMA which deserve additional discussion. First, it will be important to highlight where PRIMA fits in relation to current assays. Second, imminent applications of PRIMA will be profiled. Third, potential technical improvements to PRIMA will be outlined. Fourth, the long term potential to collect large amounts of data and gain new insights will be discussed.

PRIMA's impact on identifying RNA-RBP interactions

PRIMA was developed to fill a perceived void in RNA-RBP interaction technologies. PRIMA detects physical interactions in a gene-centered manner, meaning that a scientist can begin with an RNA of interest and identify the RBPs to which it binds. By demonstrating that full-length 3'UTRs can be tested in PRIMA, it is clear that PRIMA will be useful for interrogating large scale RNA-RBP gene regulatory networks.

Previously, the only existing gene-centered assay was yeast three-hybrid (Y3H). Y3H is limited in that RNAs longer than 150 nucleotides result in decreased signals. Additionally, tetrauridine signals (which are common in 3'UTRs) can terminate bait RNA production thus limiting the testable baits. In contrast, PRIMA is not limited by either threshold. This was demonstrated by successful tests of the *nos-2* and *glp-1* 3'UTRs. These UTRs are both greater than 300 nt and contain seven and nine tetrauridine stretches, respectively. Undoubtedly, PRIMA will have technical limitations, but these two critical milestones have been surpassed.

PRIMA is a heterologous yeast-based assay - it is not an *in vivo, in vitro,* or *in silico* assay. This creates many qualitative and quantitative

variables that must be considered. For instance, the expression level of a prey RBP may not reflect that of its *in vivo* cell type. Additionally, it cannot be precisely controlled (and normalized) as is the case during *in vitro* experiments. Furthermore, there are thousands of other molecules in the yeast cell. The pair-wise RNA-RBP test is therefore not entirely 1-to-1 and potential interactors include molecules which neither partner would see *in vivo.* This means that a given RNA binding site could be blocked by an unexpected yeast RBP or that a yeast RBP could alter the mRNA structure occluding or exposing binding sites. Also, the prey RBP could sequestered or altered by a yeast RNA or RBP affecting the prey's effective levels or localization.

PRIMA is not the only theoretically possible gene-centered RNA-RBP assay. It may be possible to perform gene-specific immunoprecipitation (IP) of mRNAs followed by mass spectrometry identification of the bound RBPs. Such assays have been performed following IP of total cellular mRNA through targeting of the mostly ubiquitous polyA tail (Castello *et al.* 2012, Baltz *et al.* 2012). Two analyses of specific genes were previously performed using *in vitro* synthesized RNAs and peptide nucleic acids (PNA) (Zielinski *et al.* 2006, Butter *et al.* 2009), but subsequent follow-ups have not been performed. Recently, the Xist long noncoding RNA was immunoprecipitated and corresponding bound RBPs were identified using

mass spectrometry further demonstrating the potential of this method (Chu *et al.* 2015). Major advantages of this technology include detection of interactions in an unbiased manner. However, there will likely be a sensitivity issue resulting from attempted identification of low-abundance RBPs. Additionally, IP of low abundance RNAs may be challenging. Future development of this technology will ultimately serve to complement PRIMA and advance the field of post-transcriptional regulation.

Additional improvements to PRIMA

Additional, untested improvements to PRIMA are possible. For instance, modifications to the Pab1p portion of the prey may result in improved activation and sensitivity. Truncated Pab1p constructs may improve prey expression or reduce steric hindrance between protein-protein and protein-RNA interactions. Truncated, or mutated Pab1p molecules may also reduce sequestration of the prey protein on endogenous mRNAs if they are designed to no longer bind poly(A) mRNA. Alternatively, concatenation of multiple Pab1p molecules may increase reporter activation. It is known from yeast two-hybrid experiments that the orientation of the fusion protein can affect the ability to detect interactions (Chen *et al.* 2010). Therefore the order of the prey RBP, MS2, and PAB1 should be permutated while testing

a panel of RNA-RBP interactions. Lastly, addition of a fluorescent protein marker (*e.g.* RFP) to the prey would enable real time analysis of prey expression. Flow cytometry data could be gated to include only cells with significant prey expression which may improve dynamic range. Additionally, expression data would help to identify RBP preys which are untestable in PRIMA.

Improved constructs could be designed rationally and tested individually or a they could be selected in an unbiased manner from randomized libraries. For instance, the existing Pab1p ORF in the MS2-Pab1p could be replaced with alternative ORFs derived from random PCR mutagenesis. The yeast strain expressing a GFP bait mRNA with the 66 nM MS2BS could then be transformed with the random prey library and resulting colonies would then be assayed in PRIMA using fluorescence activated cell sorting to select the yeast with the strongest GFP signals. The *PAB1* sequence from these yeast could then be sequenced to look for enriched populations with enhancing mutations. Alternatively, the *HIS3* bait with an MS2BS could be transformed with the library followed by plating of the cells on media lacking histidine and containing varying levels of 3-AT. This would allow for a rapid selection of colonies with enhancing mutations. From either population the prey plasmids can be recovered and retested using conventional PRIMA protocols.

The limits of PRIMA have not been established because only a handful of 3’ UTRs have been tested. The longest 3'UTR tested is 363 nucleotides. The majority of 3'UTRs in *C. elegans* are shorter than 363 nucleotides, however longer UTRs exist (Mangone *et al.* 2010). Furthermore, higher complexity eukaryotes have longer 3'UTRs (Ulitsky *et al.* 2012, Derti *et al.* 2012). Particular complications may arise from nonsense mediated decay of the bait mRNA as it is a length dependent process in yeast (Amrani *et al.* 2004). This issue may be circumvented using a *UPF1* deletion strain to eliminate nonsense mediated decay (Amrani *et al.* 2004). However, long sequences are more likely to contain spurious polyadenylation signals. Polyadenylated mRNAs will lead to high GFP expression through recruitment of endogenous Pab1p negating the ability of a bound prey to generate detectable signal increases. These potential concerns will need to be evaluated as additional baits are tested.

Improvements to the PRIMA pipeline may help increase throughput beyond the current capacity. The current bottleneck is the manual labor associated with picking yeast colonies following prey transformation in each bait strain. Improvements such as automated colony picking robots or revisiting a yeast mating strategy (Reece-Hoyes *et al.* 2011) would greatly increase throughput. Additionally, alternative high throughput flow cytometers may have reduced sample acquisition times compared to the BD

Accuri C6 system. Reduced acquisition times would enable an increase in the number of plates tested per machine each day.

Potential PRIMA experiments

PRIMA will enable a rapid increase in the throughput of gene-centered RNA-RBP interaction detection. There are many interesting questions to answer and many potential avenues of pursuit exist. This subsection will outline a scale-up plan and the rationale which seeks to capitalize on several worthwhile experiments en route to the delineation of larger scale gene regulatory networks.

PRIMA is a gene-centered assay and therefore it is only as valuable as its RBP prey library. The *glp-1 and nos-2* 3'UTR baits are known to be expressed in the *C. elegans* germline therefore the initial prey RBPs were chosen based on expression in this tissue. 40 of the most highly expressed germline RBPs were chosen from the list of ~250 gene-specific RBPs (Tamburino *et al.* 2013). Before testing additional PRIMA baits it would be prudent to expand this library to include the majority of gene-specific (gs) RBPs, an approximate 6-fold increase. Ultimately, additional predicted RBPs can be included in the prey library including C2H2 zinc fingers (which may be RBPs or TFs) and general RBPs such as ribosomal subunits which

may have gene-specific functions (Szostak and Gebauer 2013). Ultimately, a genome-wide prey library would enable the unbiased testing of a bait against all ORFs. However, the construction and maintenance of such a large library would be costly and time consuming. Additionally, no organism is known where >10% of it's ORFs bind RNA. Therefore, >90% of the ORFs tested in a genome-wide library would be likely be negatve for all baits tested and thus not worth the time, effort and expense involved to test them.

Concurrently, the PRIMA prey library can be expanded by searching for RBPs in an unbiased manner. Many experimentally determined RBPs contain no known RBD and therefore would not be predicted from computational motif-based predictions (Tsvetanova *et al.* 2010, Castello *et al.* 2012, Baltz *et al.* 2012). Yeast one-, two- and three-hybrids have all employed cDNA library screens to identify novel protein interactors (Deplancke *et al.* 2004, Li *et al.* 2004, Zhang *et al.* 1997). The same strategy could be adopted for PRIMA by altering the prey vector to enable cloning of cDNAs. Screens could then utilize a two step strategy. First, a bait mRNA is tested against a cDNA library using fluorescence activated cell sorting to isolate potential prey interactors. Individual preys can be determined through Sanger sequencing of individual colony plasmids or enriched preys can be determined through deep sequencing of the input and output libraries. Second, the prey plasmids can be recovered or

recloned into the Gateway compatible prey vector and tested using conventional PRIMA protocols. Alternatively, the bait mRNA can be reconfigured with a *HIS3* ORF and positive interactions can be selected using media lacking histidine and contain 3-AT.

The choice between these two methods will be based upon the number of interactions and/or resulting colonies obtained in pilot screens using known RNA-RBP interactions. Ideally, the cDNA library would be normalized to include all ORFs at an equal frequency. In practice, the library will be biased and more frequently represented ORFs may appear more often. Many colonies would need to be picked to sample the lower abundance ORFs. In contrast, deep sequencing the input and output libraries may indicate potential enrichments at all frequency levels. However, numerous sources of error may lead to erroneous hits and increase the number of retests. In contrast, the *HIS3* reporter selectivity can be increased through addition of increasing 3-AT. Potentially, a level can be reached where only true interactions are determined. The upfront work of identifying hits would likely increase, but the downstream confirmation rate would increase thereby balancing the effort.

As additional RNAs and RBPs are tested in PRIMA many novel interactions will be uncovered and a large scale comparison of PRIMA to established *in vitro* binding assays will become necessary. PRIMA

experiments to date recapitulated interactions with eight different RBPs (true positives). Several known interactions have been missed, including those between MEX-3, MEX-5, OMA-1, OMA-2 and the *glp-1* 3'UTR (false negative). Additional interactions between the *glp-1* and *nos-2* 3'UTRs and RBPs have been observed. However, these RNA - protein interactions have not been tested using an orthologous physical binding assay, such as EMSA. Furthermore, these interactions are minimal in number compared to the possible interactions that may occur between the thousands of 3'UTRs and hundreds or RBPs in any given organism. Therefore a study comparing PRIMA to an orthologous physical binding assay should be conducted with suitable sample size for comparison.

The regulatory consequences of PRIMA interactions should also be evaluated at larger scale. Interactions were determined through testing of two PRIMA 3'UTRs against 40 preys. Several known physical and regulatory interactions were recapitulated, but only one new regulatory interaction was discovered. Several possibilities exist for the discrepancy between the observed PRIMA interactions and the lack of regulatory consequences *in vivo.* Without validations of the novel physical interactions it is possible that some PRIMA interactions are false positives. In contrast, these physical interactions may be real, but their regulatory consequences may be masked through redundant interactions (*e.g.* paralogous RBPs).

Also, it is possible that germline expression of the RBP has been misannotated and should therefore be confirmed through alternative methods. In particular, the utilized annotations are based on RNA expression and thus it does not confirm protein expression in the germline (or subsection there of), as is the case for many genes.

There are many options for systematic testing of baits in PRIMA and several factors will direct this focus. Ideally, 3'UTRs will be tested using PRIMA to identify physical interactions. The positive interactions will then be assessed for regulation using *C. elegans* transgenic strains and RBP RNAi. Currently, the availability of *C. elegans* transgenic strains is limiting and the generation of additional strains is time consuming relative to the creation of PRIMA strains. Therefore, preliminary, targets may be selected according to available worm strains.

Many exciting biological questions can be investigated with a working PRIMA to RNAi pipeline, including the regulation of specific functionally related or co-expressed mRNAs. For instance, coordinated expression by RBPs within the germline is vital to development, but few examples of their own mRNA regulation have been shown. Using PRIMA, RBP mRNAs can be tested against RBPs to delineate a physical and regulatory cross-talk network. Additionally, transcription is known to be limited during oocyte maturation, but TF's must be rapidly activated during the transition to

zygotic translation at the 4-cell stage. Using PRIMA, TF mRNAs can be tested against RBPs to determine important physical events that may impact early stages of development.

Ultimately many additional 3'UTRs will be tested in PRIMA through progressive increases in the number of baits tested. Baits can be systematically chosen according to protein family (*e.g.* RBPs, TFs, Kinases, etc.), by tissue (*e.g.* germline, intestine) or both. The ultimate choice will depend upon the researcher"s biological question and the scope of the project.

PRIMA was designed to interrogate 3'UTR-mRNA interactions in a heterologous yeast based system. During development of PRIMA six short RNA binding sites were successfully tested against five RBPs. This set included one human splicing protein, NOVA1, demonstrating that additional RNA elements including splicing silencers or enhancers can be tested in PRIMA. Additionally, long noncoding RNAs may be testable in PRIMA opening up additional avenues of research.

Analyses of a large-scale PRIMA derived network

Long term, hundreds of 3'UTR baits will be tested against thousands of RBP preys. This will result in a large matrix of interactions that will contain

a wealth of data. For instance, a matrix of 200 3'UTR preys tested against 1,000 RBP preys would contain 200,000 tests of 3'UTR-RBP interactions.

A large dataset will enable the many interesting analyses. First, it will be interesting to see how many positive interactions will occur within that matrix. It is believed that many RBPs bind to a given mRNA throughout its life, therefore many RBPs will be expected to interact with each bait. It is not known what the definition of 'many' is. A given mRNA may be bound by three to five RBPs, or it may be bound by 20 to 30 or more. The initial tests with the *glp-1* and *nos-2* 3'UTRs resulted in seven and eight interactions, respectively, of the 40 RBPs tested. Extrapolating this number means that as many as 200 of 1,000 RBPs could bind to a single 3'UTR. Although it should be reiterated that the mini-library was biased towards highly expressed RBPs.

It will also be interesting to see the pattern by which RBPs bind to the set of 3'UTRs. For instance, many, or even all, 3'UTRs may be bound by the same RBP. It is possible that this RBP may regulate many 3'UTRs, as does Pab1p, or the RBP may bind with high affinity but low specificity. In contrast, there may be several RBPs which interact with one or few 3'UTRs. These RBPs may be highly specific and a means for precise regulation within the cell.

Ultimately, the bulk of the mRNAs will be bound by several RBPs and larger patterns will begin to emerge. Many RBPs will undoubtedly associate with RNAs from functionally related genes, as is the case with RNA-regulons (Keene 2007). In contrast, a given mRNA may be bound by functionally related RBPs which coordinate amongst each other to regulate the mRNA spatially and temporally. This could also indicated that the RNA is, in fact, regulating the related RBPs. Binding could result in re-localization of passenger RBPs in the cell or temporary sequestration in cellular granules. Alternatively, the mRNA may be bound by numerous, yet functionally unrelated, RBPs whose only commonality is the message. Combining the interaction data with knowledge of the RBPs' functions, expressions patterns and signalling pathways could lead to insights about the mRNA, including not only which RBPs bind, but the function, expression, pathways of its encoded protein.

Conclusion

In this thesis I have outlined a gene-centered method for mapping RNA-protein interactions. First, I have shown how computational predictions can improve the annotations of RBPs in the organism *C. elegans.* Second, I have developed PRIMA, the Protein-RNA Interaction Mapping Assay, and I

have validated it using known RNA-RBP interactions. Third, I have tested full-length 3'UTRs in PRIMA against a small library of RBPs. This test recapitulated many known interactions and discovered additional physical interactions, including one novel regulatory interaction. This work demonstrates that PRIMA works and will be of use to the study of post-transcriptional networks involving RNAs and RBPs.

Bibliography

Ahringer, J., & Kimble, J. (1991). Control of the sperm-oocyte switch in *Caenorhabditis elegans* hermaphrodites by the *fem-3* 3' untranslated region. Nature, 349, 346-348.

Ahringer, J., Rosenquist, T. A., Lawson, D. N., & Kimble, J. (1992). The *Caenorhabditis elegans* sex determining gene *fem-3* is regulated post-transcriptionally. The EMBO Journal, 11(6), 2303-2310.

Alberti, S., Gitler, A. D., & Lindquist, S. (2007). A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. Yeast, 24, 913-919.

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

Arda, H. E., Taubert, S., MacNeil, L. T., Conine, C. C., Tsuda, B., Van Gilst, M., … Walhout, A. J. M. (2010). Functional modularity of nuclear hormone

receptors in a *Caenorhabditis elegans* metabolic gene regulatory network. Molecular Systems Biology, 6(367), 1-14.

Arda, H. E., & Walhout, A. J. M. (2010). Gene-centered regulatory networks. Briefings in Functional Genomics, 9(1), 4–12.

Ascano, M., Gerstberger, S., & Tuschl, T. (2013). Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. Current Opinion in Genetics and Development, 23, 20-28.

Badis, G., Berger, M. F., Philippakis, A. a, Talukder, S., Gehrke, A. R., Jaeger, S. a, … Bulyk, M. L. (2009). Diversity and complexity in DNA recognition by transcription factors. Science, 324(5935), 1720–1723.

Baltz, A. G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., … Landthaler, M. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Molecular Cell, 46(5), 674–690.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. Cell, 136(2), 215–233.

Beckmann, K., Grskovic, M., Gebauer, F., & Hentze, M. W. (2005). A dual inhibitory mechanism restricts *msl-2* mRNA translation for dosage compensation in *Drosophila*. Cell, 122, 529-540.

Bernstein, D., Hook, B., Hajarnavis, A., Opperman, L., & Wickens, M. (2005). Binding specificity and mRNA targets of a C. elegans PUF protein, FBF-1. RNA, 11, 447-458.

Blencowe, B. J. (2006). Alternative splicing: new insights from global analyses. Cell, 126(1), 37–47.

Blewett, N. H., & Goldstrohm, A. C. (2012). A eukaryotic translation initiation factor 4E-binding protein promotes mRNA decapping and is required for PUF repression. Molecular and Cellular Biology, 32(20), 4181–4194.

Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., … Vidal, M. (2009). An experimentally derived confidence score for binary protein-protein interactions. Nature Methods, 6(1), 91-97.

Buchan, J. R., & Parker, R. (2009). Eukaryotic Stress Granules: The Ins and Outs of Translation. Molecular Cell, 36, 932-941.

Buckanovich, R. J., & Darnell, R. B. (1997). The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. Molecular and Cellular Biology, 17(6), 3194–3201.

Butter, F., Scheibe, M., Morl, M., & Mann, M. (2009). Unbiased RNA-protein interaction screen by quantitative proteomics. Proceedings of the National Academy of Sciences of the United States of America, 106(26), 10626–10631.

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell, 136(4), 642–655.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., … Hentze, M. W. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell, 149(6), 1393–1406.

Chen, Y.-C., Rajagopala, S. V., Stellberger, T., & Uetz, P. (2010). Exhaustive benchmarking of the yeast two-hybrid system. Nature Methods. 7(9). 667-668.

Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., … Chang, H. Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell, 161(2), 404–416.

Cho, P. F., Poulin, F., Cho-Park, Y. A., Cho-Park, I. B., Chicoine, J. D., Lasko, P., & Sonenberg, N. (2005). A new paradigm for translational control: Inhibition via 5’-3' mRNA tethering by Bicoid and the eIF4E cognate 4EHP. Cell, 121(3), 411–423.

Chritton, J. J., & Wickens, M. (2010). Translational repression by PUF proteins in vitro. RNA, 16, 1217-1225.

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

Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., & Hughes, T. R. (2011). RBPDB: a database of RNA-binding specificities. Nucleic Acids Research, 39(Database issue), D301–8.

Crittenden, S. L., Troemel, E. R., Evans, T. C., & Kimble, J. (1994). GLP-1 is localized to the mitotic region of the *C. elegans* germ line. Development, 120(10), 2901–2911.

Deplancke, B., Dupuy, D., Vidal, M., & Walhout, A. J. M. (2004). A Gateway-Compatible Yeast One-Hybrid System. Genome Research, 14, 2093–2101.

Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A. M., Grove, C. A., Martinez, N. J., … Walhout, A. J. M. (2006). A Gene-Centered *C. elegans* Protein-DNA Interaction Network. Cell,125, 1193-1205 .

Deribe, Y. L., Pawson, T., & Dikic, I. (2010). Post-translational modifications in signal integration. Nature Structural & Molecular Biology, 17(6), 666–72.

Derti, A., Garrett-Engele, P., MacIsaac, K. D., Stevens, R. C., Sriram, S., Chen, R., … Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. Genome Research, 22, 1173-1183 .

Dower, K., Kuperwasser, N., Merrikh, H., & Rosbash, M. (2004). A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. RNA, 10, 1888-1899 .

Ellington, A. D., & Szostak, J. (1990). Ellington and Szostak 1990. Nature, 346, 818–822.

Evans, T. C., Crittenden, S. L., Kodoyianni, V., & Kimble, J. (1994). Translational control of maternal glp-1 mRNA establishes an asymmetry in the *C. elegans* embryo. Cell, 77, 183-194.

Farley, B. M., Pagano, J. M., & Ryder, S. P. (2008). RNA target specificity of the embryonic cell fate determinant POS-1. RNA, 14, 2685, 2697.

Farley, B. M., & Ryder, S. P. (2012). POS-1 and GLD-1 repress *glp-1* translation through a conserved binding site cluster. Molecular Biology of the Cell, 23, 4473-4483.

Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., & Ahringer, J. (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. Nature, 408, 325-330.

Garneau, N. L., Wilusz, J., & Wilusz, C. J. (2007). The highways and byways of mRNA decay. Nature Reviews. Molecular Cell Biology, 8, 113-126.

Gene Ontology Consortium, T. (2000). Gene Ontology : tool for the. Gene Expression, 25(May), 25–29.

Gerstberger, S., Hafner, M., Ascano, M., & Tuschl, T. (2014). Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. Advances in Experimental Medicine and Biology, 825, 1–55.

Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., … Waterston, R. H. (2010). Integrative Analysis of the

Caenorhabditis elegans Genome by the modENCODE Project. Science, 330(6012), 1775–1787.

Glisovic, T., Bachorik, J. L., Yong, J., & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Letters, 582, 1977-1986.

Goldstrohm, A. C., Seay, D. J., Hook, B. A., & Wickens, M. (2007). PUF protein-mediated deadenylation is catalyzed by Ccr4p. Journal of Biological Chemistry, 282(1), 109-114.

Gray, N. K., Coller, J. M., Dickson, K. S., & Wickens, M. (2000). Multiple portions of poly(A)-binding protein stimulate translation in vivo. The EMBO Journal, 19(17), 4723–4733.

Grove, C. a, De Masi, F., Barrasa, M. I., Newburger, D. E., Alkema, M. J., Bulyk, M. L., & Walhout, A. J. M. (2009). A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. Cell, 138(2), 314–27.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., … Tuschl, T. (2010). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell, 141, 129-141.

Hammell, M., Long, D., Zhang, L., Lee, A., Carmack, C. S., Han, M., … Ambros, V. (2008). mirWIP : microRNA target prediction based on microRNA-containing ribonucleoprotein – enriched transcripts. Nature Methods, 5(9), 813-819.

Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., … Young, R. A. (2004). Transcriptional regulatory code of a eukaryotic genome. Nature, 99–104.

Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D., & Brown, P. O. (2008). Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biology, 6(10), e255.

Hook, B. A., Goldstrohm, A. C., Seay, D. J., & Wickens, M. (2007). Two yeast PUF proteins negatively regulate a single mRNA. Journal of Biological Chemistry, 282(21), 15430-15438.

Hubbard, E. J. A., & Greenstein, D. (2005). Introduction to the germ line. WormBook : The Online Review of *C. elegans* Biology. 1-4.

Hubstenberger, A., Cameron, C., Shtofman, R., Gutman, S., & Evans, T. C. (2012). A network of PUF proteins and Ras signaling promote mRNA repression and oogenesis in *C. elegans.* Developmental Biology, 366(2), 218-231.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., … Yeats, C. (2009). InterPro: the integrative protein signature database. Nucleic Acids Research, 37(Database issue), D211–5.

Jadhav, S., Rana, M., & Subramaniam, K. (2008). Multiple maternal proteins coordinate to restrict the translation of *C. elegans nanos-2* to primordial germ cells. Development, 135(10), 1803-1812.

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

Johansson, H. E., Dertinger, D., LeCuyer, K. A., Behlen, L. S., Greef, C. H., & Uhlenbeck, O. C. (1998). A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. Proceedings of the National Academy of Sciences of the United States of America, 95(16), 9244-9249.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics, 30(9), 1236–1240.

Jungkamp, A. C., Stoeckius, M., Mecenas, D., Grün, D., Mastrobuoni, G., Kempa, S., & Rajewsky, N. (2011). In vivo and transcriptome-wide identification of RNA binding protein target sites. Molecular Cell, 44(5), 828-840.

Kalchhauser, I., Farley, B. M., Pauli, S., Ryder, S. P., & Ciosk, R. (2011). FBF represses the Cip/Kip cell-cycle inhibitor CKI-2 to promote self-renewal of germline stem cells in *C. elegans*. The EMBO Journal, 30(18), 3823-3829.

Kamath, R. S., & Ahringer, J. (2003). Genome-wide RNAi screening in *Caenorhabditis elegans*. Methods, 30(4), 313-321.

Kaymak, E., & Ryder, S. P. (2013). RNA recognition by the *Caenorhabditis elegans* oocyte maturation determinant OMA-1. Journal of Biological Chemistry, 288(42), 30463-30472.

Keene, J. D. (2007). RNA regulons: coordination of post-transcriptional events. Nature Reviews. Genetics, 8(7), 533-543.

Keene, J. D., Komisarow, J. M., & Friedersdorf, M. B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. Nature Protocols, 1(1), 302-307.

Kershner, A. M., & Kimble, J. (2010). Genome-wide analysis of mRNA targets for *Caenorhabditis elegans* FBF, a conserved stem cell regulator. Proceedings of the National Academy of Sciences of the United States of America, 107(8), 3936-3941.

Kiledjian, M., & Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. The EMBO Journal, 11(7), 2655–64.

Kim, K. W., Wilson, T. L., & Kimble, J. (2010). GLD-2/RNP-8 cytoplasmic poly(A) polymerase is a broad-spectrum regulator of the oogenesis program. Proceedings of the National Academy of Sciences of the United States of America, 107(40), 17445-17450.

Kimble, J., & Crittenden, S. L. (2005). Germline proliferation and its control. WormBook : The Online Review of *C. elegans* Biology. 1-14.

Koh, Y. Y., Opperman, L., Stumpf, C., Mandan, A., Keles, S., & Wickens, M. (2009). A single C. elegans PUF protein binds RNA in multiple modes. RNA, 15(6), 1090-1099.

Kuersten, S., & Goodwin, E. B. (2003). The power of the 3' UTR: translational control and development. Nature Reviews. Genetics, 4(8), 626–37.

Kummerfeld, S. K., & Teichmann, S. a. (2006). DBD: a transcription factor prediction database. Nucleic Acids Research, 34(Database issue), D74–81.

Lamprecht, M. R., Sabatini, D. M., & Carpenter, A. E. (2007). CellProfiler: free, versatile software for automated biological image analysis. BioTechniques, 42(1), 71–75

Lee, M.-H., & Schedl, T. (2006). RNA-binding proteins. WormBook : The Online Review of *C. elegans* Biology, 1–13.

Letunic, I., Doerks, T., & Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Research, 40(Database issue), D302–5.

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., … Vidal, M. (2004). A map of the interactome network of the metazoan C. elegans. Science, 303(5657), 540–3.

Lu, D., Searles, M. A., & Klug, A. (2003). Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. Nature, 426(6962), 96–100.

Lublin, A. L., & Evans, T. C. (2007). The RNA-binding proteins PUF-5, PUF-6, and PUF-7 reveal multiple systems for maternal mRNA regulation during *C. elegans* oogenesis. Developmental Biology, 303(2), 635-649.

Mangone M, Manoharan AP, Thierry-Mieg D, Thierry-Mieg J, Han T, Mackowiak SD, … Kim JK. (2010). The Landscape of *C. elegans* 3'UTRs. Science, 329(5990), 432–435.

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

Marin, V. A., & Evans, T. C. (2003). Translational repression of a *C. elegans* Notch mRNA by the STAR/KH domain protein GLD-1. Development, 130(12), 2623-2632.

Martin, F., Schaller, A., Eglite, S., Schümperli, D., & Müller, B. (1997). The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. EMBO Journal, 16(4), 769-778.

Martin, K. C., & Ephrussi, A. (2009). mRNA Localization: Gene Expression in the Spatial Dimension. Cell, 136(4), 719–730.

Martinez, N. J., Ow, M. C., Barrasa, M. I., Hammell, M., Sequerra, R., Doucette-Stamm, L., … Walhout, A. J. M. (2008). A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. Genes and Development, 22(18), 2535-2549.

Martinez, N. J., & Walhout, A. J. M. (2009). The interplay between transcription factors and microRNAs in genome-scale regulatory networks. BioEssays, 435–445.

Mayr, C., & Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell, 138(4), 673–84.

Merrihew, G. E., Davis, C., Ewing, B., Williams, G., Käll, L., Frewen, B. E., … MacCoss, M. J. (2008). Use of shotgun proteomics for the identification, confirmation, and correction of *C. elegans* gene annotations. Genome Research, 18(10), 1660–9.

Merritt, C., Rasoloson, D., Ko, D., & Seydoux, G. (2008). 3′ UTRs Are the Primary Regulators of Gene Expression in the *C. elegans* Germline. Current Biology, 18(19), 1476-1482.

Michel, F., Schümperli, D., & Müller, B. (2000). Specificities of *Caenorhabditis elegans* and human hairpin binding proteins for the first nucleotide in the histone mRNA hairpin loop. RNA, 6(11), 1539-1550.

Mitchell, S. F., & Parker, R. (2014). Principles and Properties of Eukaryotic mRNPs. Molecular Cell, 54, 547-558.

Mittal, N., Roy, N., Babu, M. M., & Janga, S. C. (2009). Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. Proceedings of the National Academy of Sciences of the United States of America, 106(48), 20300–5.

Mittal, N., Scherrer, T., Gerber, A. P., & Janga, S. C. (2011). Interplay between posttranscriptional and posttranslational interactions of RNA-binding proteins. Journal of Molecular Biology, 409(3), 466–79.

Moore, M. J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. Science, 309(5740), 1514–1518.

Moore, M. J., & Proudfoot, N. J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell, 136(4), 688–700.

Nakamura, A., Sato, K., & Hanyu-Nakamura, K. (2004). *Drosophila* Cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA Translation in Oogenesis. Developmental Cell, 6(1), 69-78.

Ogura, K.-I., Kishimoto, N., Mitani, S., Gengyo-Ando, K., & Kohara, Y. (2003). Translational control of maternal *glp-1* mRNA by POS-1 and its interacting protein SPN-4 in *Caenorhabditis elegans*. Development, 130(11), 2495-2503.

Opperman, L., Hook, B., DeFino, M., Bernstein, D. S., & Wickens, M. (2005). A single spacer nucleotide determines the specificities of two mRNA regulatory proteins. Nature Structural & Molecular Biology, 12(11), 945-951.

Orr-Weaver, T. L., Szostak, J. W., & Rothstein, R. J. (1983). Genetic applications of yeast transformation with linear and gapped plasmids. Methods in Enzymology, 101, 228–245.

Pagano, J. M., Clingman, C. C., & Ryder, S. P. (2011). Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. RNA, 17(1), 14-20.

Pagano, J. M., Farley, B. M., Essien, K. I., & Ryder, S. P. (2009). RNA recognition by the embryonic cell fate determinant and germline totipotency factor MEX-3. Proceedings of the National Academy of Sciences of the United States of America, 106(48), 20252-20257.

Pagano, J. M., Farley, B. M., McCoig, L. M., & Ryder, S. P. (2007). Molecular basis of RNA recognition by the embryonic polarity determinant MEX-5. Journal of Biological Chemistry, 282(12), 8883-8894.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., … Finn, R. D. (2012). The Pfam protein families database. Nucleic Acids Research, 40(Database issue), D290–301.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., & Lopez, R. (2005). InterProScan: protein domains identifier. Nucleic Acids Research, 33(Web Server issue), W116–20.

Ray, D., Kazan, H., Chan, E. T., Peña Castillo, L., Chaudhry, S., Talukder, S., … Hughes, T. R. (2009). Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nature Biotechnology, 27(7), 667-670.

Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., … Hughes, T. R. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature, 499(7457), 172-177.

Reece-Hoyes, J. S., Deplancke, B., Shingles, J., Grove, C. A., Hope, I. A., & Walhout, A. J. M. (2005). A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biology, 6(13), R110.

Reece-Hoyes, J. S., Diallo, A., Lajoie, B., Kent, A., Shrestha, S., Kadreppa, S., … Walhout, A. J. M. (2011). Enhanced yeast one-hybrid assays for

high-throughput gene-centered regulatory network mapping. Nature Methods, 8(12), 1059-1064.

Reece-Hoyes, J. S., Pons, C., Diallo, A., Mori, A., Shrestha, S., Kadreppa, S., … Walhout, A. J. M. (2013). Extensive rewiring and complex evolutionary dynamics in a *C.elegans* multiparameter transcription factor network. Molecular Cell, 51(1), 116-127.

Ryder, S. P., Frater, L. a, Abramovitz, D. L., Goodwin, E. B., & Williamson, J. R. (2004). RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. Nature Structural & Molecular Biology, 11(1), 20–8.

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. a, & Burge, C. B. (2008). Proliferating cells express mRNAs with shortened 3’ untranslated regions and fewer microRNA target sites. Science, 320(5883), 1643–7.

Sengupta, D. J., Wickens, M., & Fields, S. (1999). Identification of RNAs that bind to a specific protein using the yeast three-hybrid system. RNA, 5(4), 596–601.

SenGupta, D. J., Zhang, B., Kraemer, B., Pochart, P., Fields, S., & Wickens, M. (1996). A three-hybrid system to detect RNA-protein interactions in vivo. Proceedings of the National Academy of Sciences of the United States of America, 93(16), 8496-8501.

Seydoux, G., & Dunn, M. A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of *Caenorhabditis elegans* and *Drosophila melanogaster*. Development, 124(11), 2191-2201.

Seydoux, G., & Fire, A. (1994). Soma-germline asymmetry in the distributions of embryonic RNAs in *Caenorhabditis elegans*. Development, 120(10), 2823-2834.

Seydoux, G., Mello, C. C., Pettitt, J., Wood, W. B., Priess, J. R., & Fire, A. (1996). Repression of gene expression in the embryonic germ lineage of *C. elegans.* Nature, 382, 713-716.

Shepard, K. a, Gerber, a P., Jambhekar, a, Takizawa, P. a, Brown, P. O., Herschlag, D., … Vale, R. D. (2003). Widespread cytoplasmic mRNA transport in yeast: identification of 22 bud-localized transcripts using DNA

microarray analysis. Proceedings of the National Academy of Sciences of the United States of America, 100(20), 11429–34.

Sigrist, C. J. A., Cerutti, L., de Castro, E., Langendijk-Genevaux, P. S., Bulliard, V., Bairoch, A., & Hulo, N. (2010). PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Research, 38(Database issue), D161–6.

Simonis, N., Rual, J.-F., Carvunis, A.-R., Tasan, M., Lemmens, I., Hirozane-Kishikawa, T., … Vidal, M. (2009). Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. Nature Methods, 6(1), 47-54.

Steitz, T. a. (2008). A structural understanding of the dynamic ribosome machine. Nature Reviews. Molecular Cell Biology, 9(3), 242–53.

Stellberger, T., Häuser, R., Baiker, A., Pothineni, V. R., Haas, J., & Uetz, P. (2010). Improving the yeast two-hybrid system with permutated fusions proteins: the Varicella Zoster Virus interactome. Proteome Science, 8, 1-9.

Stumpf, C. R., Kimble, J., & Wickens, M. (2008). A *Caenorhabditis elegans* PUF protein family with distinct RNA binding specificity. RNA, 14(8), 1550-1557.

Szostak, E., & Gebauer, F. (2013). Translational control by 3'-UTR-binding proteins. Briefings in Functional Genomics, 12(1), 58-65.

Tamburino, A. M., Ryder, S. P., & Walhout, A. J. M. (2013). A Compendium of *Caenorhabditis elegans* RNA Binding Proteins Predicts Extensive Regulation at Multiple Levels. G3, 3, 297-304.

Theunissen, O., Rudt, F., Guddat, U., Mentzel, H., & Pieler, T. (1992). RNA and DNA binding zinc fingers in *Xenopus* TFIIIA. Cell, 71(4), 679–90.

Thompson, B. E., Bernstein, D. S., Bachorik, J. L., Petcherski, A. G., Wickens, M., & Kimble, J. (2005). Dose-dependent control of proliferation and sperm specification by FOG-1/CPEB. Development, 132(15), 3471-3481.

Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science, 249(4968), 505-510.

Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., & Darnell, R. B. (2003). Ule 2003. Science, 302(5648), 1212–1215.

Ule, J., Jensen, K., Mele, A., & Darnell, R. B. (2005). CLIP: A method for identifying protein-RNA interaction sites in living cells. Methods, 37(4), 376-386.

Ulitsky, I., Shkumatava, A., Jan, C. H., Subtelny, A. O., Koppstein, D., Bell, G. W., … Bartel, D. P. (2012). Extensive alternative polyadenylation during zebrafish development. Genome Research, 22(10), 2054-2066.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research, 35(Web Server issue), W71–4.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. a, & Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. Nature Reviews. Genetics, 10(4), 252–63.

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell, 136(4), 701–18.

Walhout, A. J. M. (2011). What does biologically meaningful mean? A perspective on gene regulatory network validation. Genome Biology, 12(4), 109.

Walhout, A. J., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S., & Vidal, M. (2000). GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. Methods in Enzymology, 328, 575–592.

Walhout, A. J., & Vidal, M. (2001). High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods, 24(3), 297-306.

Wang, X., Zhao, Y., Wong, K., Ehlers, P., Kohara, Y., Jones, S. J., … Hansen, D. (2009). Identification of genes expressed in the hermaphrodite germline of *C. elegans* using SAGE. BMC Genomics, 10, 213.

Wang, Z. F., Whitfield, M. L., Ingledue, T. C., Dominski, Z., & Marzluff, W. F. (1996). The protein that binds the 3' end of histone mRNA: A novel RNA-binding protein required for histone pre-mRNA processing. Genes and Development, 10(23), 3028-3040.

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

Wickens, M., Bernstein, D. S., Kimble, J., & Parker, R. (2002). A PUF family portrait: 3'UTR regulation as a way of life. Trends in Genetics, 18(3), 150–157.

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., … Gough, J. (2009). SUPERFAMILY--sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Research, 37(Database issue), D380–6.

Wright, J. E., Gaidatzis, D., Senften, M., Farley, B. M., Westhof, E., Ryder, S. P., & Ciosk, R. (2011). A quantitative RNA code for mRNA target selection by the germline fate determinant GLD-1. The EMBO Journal, 30(3), 533-545.

Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J., & Wickens, M. P. (1997). A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. Nature, 390(6659), 477-484.

Zhang, B., Kraemer, B., Sengupta, D., Fields, S., & Wickens, M. (1999). Yeast three-hybrid system to detect and analyze interactions between RNA and protein. Methods in Enzymology, 306, 93-113.

Zielinska, D. F., Gnad, F., Jedrusik-Bode, M., Wiśniewski, J. R., & Mann, M. (2009). *Caenorhabditis elegans* has a phosphoproteome atypical for metazoans that is enriched in developmental and sex determination proteins. Journal of Proteome Research, 8(8), 4039–49.

Zielinski, J., Kilk, K., Peritz, T., Kannanayakal, T., Miyashiro, K. Y., Eiriksdottir, E., … Eberwine, J. (2006). In vivo identification of ribonucleoprotein-RNA interactions. Proceedings of the National Academy of Sciences of the United States of America, 103(5), 1557–1562.

Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M., & Furlong, E. E. M. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature, 462(7269), 65–70.

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