UNIVERSITY OF CALIFORNIA, SAN DIEGO

Beyond the seed: the identification of microRNA target sites in *Caenorhabditis elegans*

A Dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Biology

by

James P. Broughton

Committee in charge:

Professor Amy Pasquinelli, Chair
Professor Sreekanth Chalasani
Professor Jens Lykke-Andersen
Professor James Wilhelm
Professor Eugene Yeo

2016

The Dissertation of James P. Broughton is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2016

DEDICATION

For the billions of worm lives that were extinguished
so that I may learn.

And for my parents,
who always incentivized higher education
through manual labor.

EPIGRAPH

AUTONOMY

I am living without you because
of a terror, a farfetched
notion that I
can't live without you

which I must narrow down & quell,
for how can I live
worthy of you, in the
freedom of your limber engagements,

in the casual uptakes of your
sweetest compliances
if stricken in your presence
by what your absence stills:

to have you, I school myself
to let you go; how terrible
to buy that absence
before the fragrance of any presence comes:

but though I am living without
you, surely
I can't live
without you: the thought of

you hauls my heavy
body up,
floats me around,
gives my motions point, just the thought.

– *A.R. Ammons*

TABLE OF CONTENTS

Signature Page . . . iii

Dedication . . . iv

Epigraph . . . v

Table of Contents . . . vi

List of Figures . . . ix

List of Tables . . . x

Acknowledgements . . . xi

Vita . . . xiii

Abstract of the Dissertation . . . xiv

Chapter 1 Introduction . . . 1
1.1 The discovery and importance of miRNAs . . . 1
1.2 miRNA biogenesis and function in *C. elegans* . . . 3
1.3 Principles of miRNA targeting . . . 6
1.4 New tools to investigate miRNA targeting . . . 11

Chapter 2 Identifying Argonaute binding sites in *Caenorhabditis elegans* using iCLIP . . . 16
2.1 Abstract . . . 16
2.2 Introduction . . . 17
2.3 iCLIP Protocol . . . 20
2.3.1 Materials . . . 20
2.3.2 Procedure . . . 23
2.3.3 Notes . . . 34
2.4 Conclusions . . . 36
2.5 Acknowledgements . . . 37

Chapter 3 A tale of two sequences: microRNA-target chimeric reads . . . 38
3.1 Abstract . . . 38
3.2 Background . . . 39
3.2.1 Target recognition by miRNAs . . . 39
3.2.2 Challenges in the identification of miRNA targets . . . 40
3.3 Review . . . 42
3.3.1 Ligation of two RNA molecules identifies RNA-RNA interactions . . . 42

3.3.2 miRNA-target chimeras from standard CLIP-seq library preparation . . . 46
3.3.3 Bioinformatic identification of miRNA-target chimeric reads . . . 48
3.3.4 Insights from miRNA-target chimeric reads . . . 48
3.3.5 miRNA-target chimeras identify non-canonical target sites . . . 50
3.4 Conclusions . . . 52
3.5 Authors' contributions . . . 53
3.6 Acknowledgements . . . 53

Chapter 4 Pairing beyond the seed supports microRNA targeting specificity 54
4.1 Abstract . . . 54
4.2 Introduction . . . 55
4.3 Results . . . 57
4.3.1 ALG-1 iCLIP generates miRNA-target chimeras . . . 57
4.3.2 Targets identified by chimeras are misregulated in *alg-1(-)* animals . . . 63
4.3.3 Seed pairing is enriched in miRNA target sites . . . 64
4.3.4 miRNA family members bind specific sets of target sites . . . 69
4.3.5 miRNA 3' end pairing directs specific target interactions . . . 74
4.3.6 Simplified detection of endogenous miRNA target interactions by Chimera PCR (ChimP) . . . 79
4.4 Discussion . . . 82
4.4.1 Identification and validation of endogenous miRNA binding sites . . . 82
4.4.2 Features of endogenous miRNA targeting . . . 84
4.4.3 Specificity role for miRNA 3' ends . . . 86
4.5 Experimental Procedures . . . 88
4.6 Accession Numbers . . . 89
4.7 Author Contributions . . . 89
4.8 Supplemental Information . . . 89
4.8.1 Supplemental Figures . . . 89
4.8.2 Supplemental Experimental Procedures . . . 100
4.9 Acknowledgments . . . 110

Chapter 5 Aging Argonautes: Building tools to study the microRNA Argonautes in *Caenorhabditis elegans* . . . 111
5.1 Introduction . . . 111
5.2 Experimental Procedures . . . 114
5.3 Results . . . 117
5.4 Discussion . . . 126

5.5 Acknowledgements 128

Chapter 6 Conclusions 129
6.1 The importance of the miRNA 3' end in supporting targeting specificity 130
6.2 The opposing roles of the miRNA Argonautes in regulating lifespan in *C. elegans* 138

References 141

LIST OF FIGURES

Figure 1.1: Overview of miRNA biogenesis in *C. elegans* 4
Figure 1.2: Diagram showing location of the miRNA seed sequence. 7
Figure 1.3: Challenges in the identification of miRNA targets 8
Figure 1.4: Overview of steps involved in iCLIP to identify Argonaute binding sites 12

Figure 2.1: Outline of ALG-1 iCLIP method 24
Figure 2.2: ALG-1 iCLIP autoradiograph 29
Figure 2.3: Representation of denaturing PAGE used for isolating cDNA . . 32
Figure 2.4: Comparison of ALG-1 iCLIP methods and control PCR 36

Figure 3.1: Overview of CLASH and modified iPAR-CLIP methods 43
Figure 3.2: Comparison of the biochemical steps in CLIP-seq for the generation of standard CLIP-seq reads to events that can lead to the formation of miRNA-target chimeric reads in CLIP-seq or iPAR-CLIP. 47

Figure 4.1: ALG-1 iCLIP produces miRNA-target chimeric reads 59
Figure 4.2: miRNA-target chimeric reads are enriched for seed pairing . . . 67
Figure 4.3: The let-7 family of miRNAs binds divergent sets of target sites . 72
Figure 4.4: Seed pairing is not sufficient for target regulation *in vivo* 77
Figure 4.5: Chimera PCR (ChimP) enables the identification of miRNA-target chimeras by PCR 81
Figure 4.6: Features of ALG-1 iCLIP 90
Figure 4.7: The 3'UTR of *hbl-1* and *daf-12* are bound by multiple let-7 family members 92
Figure 4.8: Expression of let-7 family miRNAs in *let-7*, *miR-48*, and *miR-241* mutant animals 93
Figure 4.9: miR-58 and miR-238 family miRNAs bind non-overlapping sets of target sites 94
Figure 4.10: Comparison of seed complementarity in shared and specific sites 96
Figure 4.11: Chimera PCR (ChimP) protocol and additional examples of ChimP 97
Figure 4.12: Target sites in non-coding RNAs 99

Figure 5.1: Extended lifespan of ALG-2 119
Figure 5.2: Loss of ALG-1 expression during aging 120
Figure 5.3: Expression of ALG-2 during adulthood. 123
Figure 5.4: No-lifespan extension in ALG-2::mCherry animals. 125

LIST OF TABLES

Table 2.1: Buffers used in ALG-1 iCLIP 20
Table 2.2: Enzymes and other materials used in ALG-1 iCLIP 21
Table 2.3: Oligonucleotides used in ALG-1 iCLIP 22

Table 4.1: Oligonucleotides used for generating *lin-41(ap427)* 101
Table 4.2: Oligonucleotides used for genotyping strains 102
Table 4.3: Oligonucleotides used for qRT-PCR 106
Table 4.4: Oligonucleotides used for northern blots 107
Table 4.5: Oligonucleotides used for Chimera PCR and detection of chimeras in iCLIP libraries 109

ACKNOWLEDGEMENTS

I would like to express my profound gratitude to Dr. Amy Pasquinelli for her mentoring during my graduate career. She has provided me with endless support and expert guidance, which shaped me into a better scientist, mentor, and teacher. I feel privileged to have had the opportunity to work with her, and I will carry her enthusiasm for Northern blots forward to all my scientific endeavors. I also would like to thank my committee, Dr. Jens Lykke-Andersen, Dr. Sreekanth Chalasani, Dr. Gene Yeo, and Dr. Jim Wilhelm. They have provided direction to my research and given me unwavering support.

I also thank the friends I have made while in San Diego: Leilani, Yvonne, Suneer, Zohreh, Albert, Hannah, Eunice, Emily, Srivats, Tito, Olivia, and Lily. Without them, this dissertation would not have been possible. The support of the other members of the Pasquinelli Lab has also been invaluable. Dr. Antti Aalto, Dr. Ian Nicastro, Dr. Vanessa Mondol, Dr. Jerry Chen and William Schreiner provided wise advice in the design of experiments and shared many laughs with me along the way. In particular, I would like to thank Sarah Azoubel Lima, who joined the lab at the same time as me. Together, we jumped the hurdles and climbed the obstacles of graduate school. Her sound scientific advice and sense of humor have kept me going. Lastly, I would like to thank my partner, Isabelle. She has supported me from deciding to go to San Diego for graduate school, to late nights writing papers, to spending weekends with worms instead of at the beach.

Chapter 2, in full, is a reprint of the material as it occurs in *Methods*, Broughton, J.P. and Pasquinelli A.E., Elsevier Inc., 2013. I was the primary author.

Chapter 3, in full, is a reprint of the material as it occurs in *Genetics Selection and Evolution,* Broughton, J.P. and Pasquinelli A.E., BioMed Central, 2016. I was the primary author.

I would like to acknowledge Dr. Michael Lovci, Jessica Huang, Dr. Gene Yeo, and Dr. Amy Pasquinelli who co-authored the in press manuscript "Pairing beyond the seed supports microRNA targeting specificity" at *Molecular Cell* on which I was the primary author and was reported in Chapter 4 of this dissertation.

I would also like to thank Dr. Antti Aalto, Dr. Ian Nicastro, Dr. Jerry Chen, and Dr. Amy Pasquinelli who gave me permission to use figures from the submitted manuscript "Opposing Roles of the MicroRNA Argonautes during Aging in *Caenorhabditis elegans*" which are included in Chapter 5.

VITA

2010	Bachelor of Science, Biological Sciences and Minor in Communications, Cornell University, Ithaca, NY
2010-2011	Academic Assistant, Weill Cornell Medical College - Qatar, Doha, Qatar
2011-2016	Graduate Research Assistant, University of California, San Diego
2011-2012	BioEASI Science Writing Editor, University of California, San Diego
2012-2015	National Science Foundation - Graduate Research Program Fellow
2013	Co-chair Recruitment Committee, University of California, San Diego
2013-2014	Co-chair Biology Peer Mentoring Program, University of California, San Diego
2015-2016	Senior Graduate Teaching Mentor, University of California, San Diego
2016	Doctor of Philosophy, Biology, University of California, San Diego

PUBLICATIONS

Broughton, J.P., Lovci, M.T., Huang, J.L., Yeo, G.W., Pasquinelli, A.E. (2016). Pairing beyond the seed supports microRNA targeting specificity. *Molecular Cell*. (in press)

Broughton, J.P., Pasquinelli, A.E. (2016). A tale of two sequences: microRNA-target chimeric reads. *Genet. Sel. Evol.* 48, 31.

Broughton, J.P., Pasquinelli, A.E. (2013). Identifying Argonaute binding sites in *Caenorhabditis elegans* using iCLIP. *Methods* 63, 119-25.

Taylor, R.T., Lubick, K.J., Robertson, S.J., **Broughton, J.P.**, Bloom, M.E., Bresnahan, W.A., Best, S.M. (2011). TRIM79α, an interferon-stimulated gene product, restricts tick-borne encephalitis virus replication by degrading the viral RNA polymerase. *Cell Host Microbe* 10, 185-96.

ABSTRACT OF THE DISSERTATION

Beyond the seed: the identification of microRNA target sites in *Caenorhabditis elegans*

by

James P. Broughton

Doctor of Philosophy in Biology

University of California, San Diego, 2016

Professor Amy Pasquinelli, Chair

MicroRNAs (miRNA) are critical regulators of development, cell differentiation, and the stress response. Mature miRNAs are small (~22 nucleotides) RNA molecules that post-transcriptionally regulate their targets by acting as guides for Argonaute (AGO) proteins. Despite the importance of these small RNAs in many pathways, the rules of miRNA target site recognition remain unclear. Currently, nucleotides 2-8 of the miRNA, termed the 'seed' sequence, are known to be critical for miRNA targeting. However, it is unclear if families of miRNAs, which share the same seed sequence, functionally target the same target sites. To improve our understanding of

miRNA targeting, I generated a unique and reproducible dataset through the ligation of miRNAs to their target sites using individual nucleotide-resolution crosslinking and immunoprecipitation (iCLIP) in the nematode worm *Caenorhabditis elegans,* which is detailed in Chapter 2.

As reviewed in Chapter 3, these ligated RNAs, called miRNA-target chimeras, provide biochemical evidence of which miRNAs are bound at a specific target site. Chapter 4 details my analysis of these chimeric data. Interestingly, the majority of miRNA target sites had the potential to support 3' end interactions. In contrast to my prediction that these highly related miRNAs would bind the same sets of targets, family members from multiple miRNA families primarily targeted specific sets of transcripts. To confirm the importance of the 3' end *in vivo*, I carried out *in vivo* rescue experiments that demonstrate that seed pairing is insufficient to mediate targeting in the absence of additional complementarity to nucleotides in the 3' end of the miRNA. To confirm that 3' end interactions direct binding specificity, I developed a novel method called ChimeraPCR (ChimP) that allowed for the detection of miRNA-target chimeras without the need for analyzing sequencing datasets.

To understand the role of AGO proteins in regulating lifespan in *C. elegans*, I generated several genetic tools, which are described in Chapter 5. These tools facilitated studies of how the highly related AGO proteins, ALG-1 and ALG-2, have opposing roles in controlling lifespan through the insulin-signaling pathway.

Chapter 1

Introduction

1.1 The discovery and importance of miRNAs

MicroRNAs (miRNAs) are critical regulators of post-transcriptional gene expression. These small RNA molecules were initially discovered in the nematode worm *Caenorhabditis elegans* as transcripts that regulated larval development, but lacked the capability to code for a protein (Lee et al., 1993; Wightman et al., 1993). Although larger noncoding RNAs (ncRNAs) such as ribosomal RNAs (rRNAs) or transfer RNAs (tRNAs) were known to have essential roles in the cell, these ~22 nucleotide RNAs represented a new class of small, regulatory RNAs in eukaryotes. Since the discovery of miRNAs in *C. elegans*, thousands of miRNAs have been reported throughout the genomes of plants and animals. Currently, miRBase (release 21) contains over 28,000 entries, with over 1,800 annotated miRNAs in humans and over 300 in *C. elegans* (Kozomara and Griffiths-Jones, 2011).

MiRNA directed modulation of gene expression is essential for the regulation of development, differentiation, and the control of cellular responses (Leung and

Sharp, 2010; Reinhart et al., 2000; Sayed and Abdellatif, 2011). For example, the first two miRNAs discovered, lin-4 and let-7, control developmental timing in *C. elegans* (Reinhart et al., 2000; Wightman et al., 1993). The lin-4 miRNA regulates the transition from the first to the second larval stage in *C. elegans*, whereas the let-7 miRNA is important in regulating the transition from the last larval stage to adulthood. A complete deletion of *let-7* results in a lethal bursting phenotype due to the misregulation of the *lin-41* messenger RNA (mRNA) (Ecsedi et al., 2015; Reinhart et al., 2000).

The perfect conservation of let-7 from *C. elegans* to humans underscores the importance of these small RNA molecules in regulating cellular processes (Pasquinelli et al., 2000). Interestingly, some of the targets of let-7 are conserved from *C. elegans* to humans (Johnson et al., 2005; Kanamoto et al., 2006; Slack et al., 2000). Similar to its role in regulating development timing in *C. elegans*, in humans let-7 promotes cellular differentiation. In support of this role, let-7 is highly expressed in differentiated cells, whereas in pluripotent stem cells it is lowly expressed (Newman et al., 2008; Rybak et al., 2008; Viswanathan et al., 2008). Furthermore, the inhibition of let-7 in differentiated cells can facilitate reprograming to adopt a stem-like state (Worringer et al., 2014). Considering that many tumors feature hallmarks of self-renewing cells, it is unsurprising that let-7 acts as a tumor-suppressor and decreased let-7 expression is observed in many types of cancer (Boyerinas et al., 2010).

Although largely associated with development and cellular differentiation, miRNAs also regulate other cellular processes. For example, miR-80 and the miR-229/miR-64 cluster in *C. elegans* have been shown to promote survival during heat

stress (Nehammer et al., 2015). Other *C. elegans* miRNAs, such as miR-71 and miR-238, promote longevity, and the loss of these miRNAs leads to shortened lifespans (De Lencastre et al., 2010). The functional importance of these specific miRNAs is likely related to their direct mRNA targets. Currently, these targets are predicted based on changes in gene expression and complementarity between the miRNA and the 3' untranslated region (3'UTR) of messenger RNAs (mRNAs). However, these association-based approaches do not identify which mRNAs are actually bound and regulated by a particular miRNA, leaving the possibility that some putative targets may be misregulated due to indirect effects. Consequently, the reliable identification of miRNA target sites remains an outstanding challenge for understanding the functional role of miRNAs in many different pathways.

1.2 miRNA biogenesis and function in *C. elegans*

The majority of miRNAs in *C. elegans* and other animals are initially transcribed by RNA polymerase II (RNAP II) as long primary transcripts (Figure 1.1). Similar to other RNA polymerase II transcripts, these primary miRNAs (pri-miRNAs) are capped, polyadenylated (Cai et al., 2004), and in some cases undergo splicing (Mondol et al., 2015). A prominent stem-loop secondary structure in the pri-miRNA is cleaved by Drosha (DRSH-1), an RNase III enzyme, which forms the core of the Microprocessor complex and is bound by two Partner-of-Drosha (PASH-1) proteins (Kwon et al., 2016; Nguyen et al., 2015). DRSH-1 has been shown to bind the base of the of the stem-loop and act as molecular ruler, whereas PASH-1 likely recognizes motifs in the apical loop and enhances association of the Microprocessor

to pri-miRNAs (Nguyen et al., 2015).

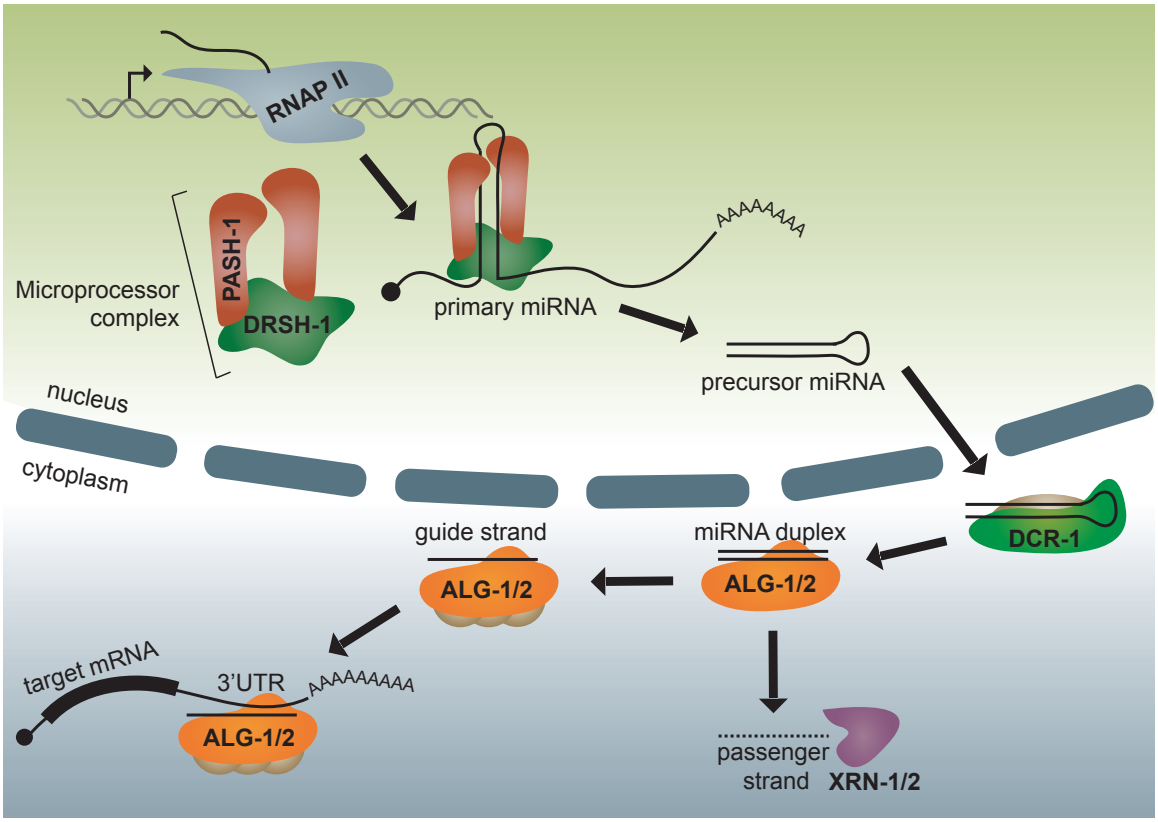


Figure 1.1: Overview of miRNA biogenesis in *C. elegans*.

The product of DRSH-1 cleavage is a ~65 nucleotide precursor miRNA (pre-miRNA). The pre-miRNA is then exported to the cytoplasm. In humans, pre-miRNA nuclear export involves Exportin-5 (Bohnsack et al., 2004; Yi et al., 2003), but in *C. elegans* there is no clear Exportin-5 homolog. However, a recent report has shown that in *exportin-5* deficient cells, miRNA biogenesis is not globally inhibited (Kim et al., 2016). This suggests that there are other factors involved in the export of pre-miRNAs from the nucleus.

Once in the cytoplasm, pre-miRNAs are recognized by another RNase III enzyme, Dicer (DCR-1), which cleaves the pre-miRNA into two ~22 nucleotide miRNAs (Ha and Kim, 2014). One strand of this miRNA duplex, known as the

passenger strand, is degraded, whereas the other strand, termed the guide strand, is loaded onto an Argonaute (AGO) protein.

AGOs are a family of RNA binding proteins found from bacteria to animals (Meister, 2013). In general, AGOs bind small RNAs, which serve as guides to direct AGO to other RNA molecules. AGO proteins contain four domains: N-terminal, PAZ, MID, and PIWI. The 3' ends of small RNAs are bound by the PAZ domain, whereas the 5' ends are located within a pocket in the MID domain (Jinek and Doudna, 2009). The PIWI domain is similar in structure to a ribonuclease-H enzyme (Song et al., 2004), and contains an active site that catalyzes the cleavage of nucleic acids (Jinek and Doudna, 2009). The activation of this 'slicing' activity of AGO generally requires perfect complementarity between the target RNA and the small guide RNA.

In *C. elegans*, AGO proteins have undergone significant expansion and specification; in mammals there are eight AGO family proteins, whereas in *C. elegans* there are 27 (Youngman and Claycomb, 2014). However, only two of these AGOs, Argonaute-Like-Gene-1 and 2 (ALG-1 and ALG-2), are specifically involved in the *C. elegans* miRNA pathway (Grishok et al., 2001). ALG-1 and ALG-2 are catalytically active, have 88% identical amino acid sequences, and are thought to be the product of a recent gene duplication event (Grishok et al., 2001; Tops et al., 2006). Despite their similarity, ALG-1 has been proposed to be the primary miRNA effector in *C. elegans* because mutations in or knockdown of *alg-1* lead to more severe developmental defects than is observed with loss of *alg-2* (Grishok et al., 2001; Tops et al., 2006; Vasquez-Rifo et al., 2012).

AGO proteins form the core of the miRNA-induced silencing complex (miRISC). The guide strand miRNA directs miRISC to target RNAs through imper-

fect base-pairing interactions (Pasquinelli, 2012). Once bound to a target RNA, miRISC partners, such as ALG-1-Interacting-protein-1 and 2 (AIN-1/AIN-2) in *C. elegans* (Zhang et al., 2007), direct the inhibition of translation or the destabilization of the target RNA. Although some groups have suggested that inhibition of translation occurs before mRNA degradation, (Bazzini et al., 2012), several comprehensive studies have emphasized that mRNA destabilization is the most common regulatory outcome in somatic tissues (Eichhorn et al., 2014; Subtelny et al., 2014).

1.3 Principles of miRNA targeting

Mature miRNAs bind target transcripts through imperfect base pairing interactions, generally in the 3'UTR of mRNAs (Pasquinelli, 2012). Functional regulation of a transcript can be achieved through as few as six base pair interactions between the miRNA and the target RNA (Bartel, 2009; Nielsen et al., 2007). The limited sequence space required for functional interactions makes the prediction of miRNA targets computationally challenging.

However, early work by many labs revealed several key features of miRNA targeting. For example, regions in 3'UTRs of regulated transcripts were found to be complementary to nucleotides on the 5' end of the mature miRNA (Brennecke et al., 2005; Grimson et al., 2007; Lee et al., 1993; Lewis et al., 2005, 2003; Moss et al., 1997; Slack et al., 2000; Wightman et al., 1993). In addition, these same sequences in the 5' end of miRNAs are conserved within and among species (Grad et al., 2003; Lee and Ambros, 2001; Lim et al., 2003). These observations led to establishment for the importance of nucleotides 2-8, termed the 'seed' sequence,

in miRNA target recognition (Figure 1.2). Subsequent experiments revealed that seed complementary was a critical feature in functional miRNA targeting (Bartel, 2009). Structural studies have provided an explanation for the importance of the seed sequence by revealing that these nucleotides are favorably positioned by the AGO to initiate contacts with the target RNA (Elkayam et al., 2012; Nakanishi et al., 2012; Schirle and MacRae, 2012).

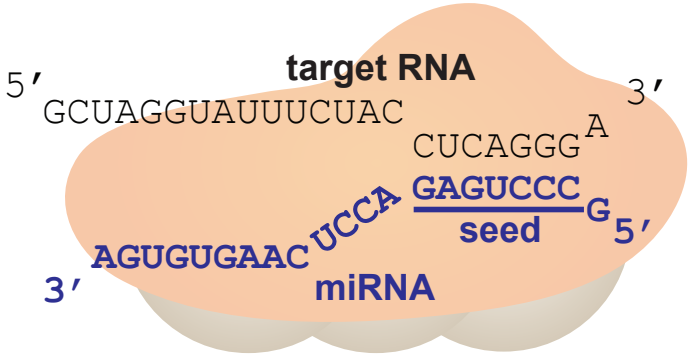


Figure 1.2: Diagram showing the location of the miRNA 'seed' sequence. Nucleotides 2-8 from the 5' end of the miRNA (shown in blue) are termed the 'seed' sequence. Perfect complementarity between the seed sequence and nucleotides in the target RNA (shown in black) is an important feature for identifying miRNA target sites. The AGO protein is shown in orange.

A variety of computational approaches have been used to predict miRNA targets. One of the first algorithms used to identify miRNA targets, called TargetScan, relied on conserved seed complementarity in 3'UTRs (Lewis et al., 2003). Alternative approaches, such as calculating the predicted minimum free energy of binding between miRNAs and regions of the 3'UTR, have also been used (Hammell et al., 2008; Krek et al., 2005; Rehmsmeier et al., 2004). More recently, algorithms have relied on the aggregation of data from multiple experiments and the calculation of target RNA features, such as secondary structure (Agarwal et al., 2015). The incorporation of secondary structure, target position, and other RNA features has

improved the accuracy of these prediction methods. Recently, the application of machine learning algorithms on experimental datasets has identified features important for miRNA targeting and helped to refine target prediction (Bandyopadhyay et al., 2015). However, computational approaches to miRNA target prediction are generally biased toward the discovery of sites that match canonical targeting motifs (perfect seed complementarity) or genic locations (3'UTRs). As a consequence, miRNA target prediction programs may miss targets that feature non-canonical base pairing interactions, occur in 5'UTRs or coding regions, and miRNA-ncRNA interactions (Figure 1.3A, B, and C).

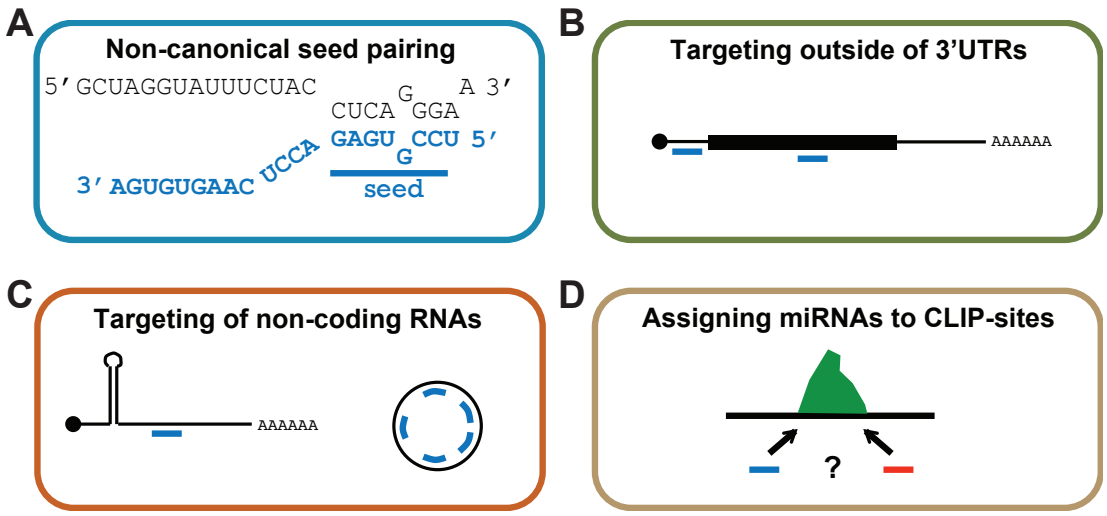


Figure 1.3: Challenges in the identification of miRNA targets. (A) Functional targets can include those with non-canonical seed pairing interactions, such as a 1 nucleotide bulge or mismatch in the seed or seed complement. (B) Some miRNA target sites have been identified outside of the 3'UTR in 5'UTRs or coding regions of mRNAs. (C) miRNAs have been shown to target non-coding RNAs such as pri-miRNAs or circRNAs. (D) CLIP-seq and PAR-CLIP studies identify binding sites for Argonaute proteins, but do not identify the specific miRNA that is responsible for a given interaction.

To determine whether predicted miRNA target sites are functional, a variety of experimental approaches have been employed. For example, forward genetic

screens allow for screening of suppressors or enhancers of phenotypes associated with loss or mutation of miRNAs (Hipfner et al., 2002; Lee et al., 1993; Reinhart et al., 2000). These screens led to the discovery of some of the earliest miRNA targets and helped to establish the basics of miRNA target recognition. The application of RNA interference (RNAi) to knockdown the expression of specific genes has also been applied to identify or confirm miRNA targets (Hunter et al., 2013; Jovanovic et al., 2010). Global methods, such as stable isotope labeling with amino acids in cell culture (SILAC) (Baek et al., 2008; Jovanovic et al., 2010; Selbach et al., 2008; Subasic et al., 2015; Vinther et al., 2006), and ribosome profiling (Bazzini et al., 2012; Guo et al., 2010) allow groups to analyze translation of miRNA targets. Methods that look at the abundance of mRNAs, such as microarrays (Grimson et al., 2007; Hristova et al., 2005) and RNA-seq (Eichhorn et al., 2014), have also been used to identify functional miRNA targets. However, similar to computational approaches, these methods are often limited by a reliance on the selection of target candidates based on previously established rules for miRNA targeting.

To address the limitations of these computational and experimental approaches, several groups turned to crosslinking immunoprecipitation and sequencing (CLIP-seq) and photoactivatable-ribonucleoside-enhanced CLIP (PAR-CLIP) to identify AGO binding sites (Chi et al., 2009; Hafner et al., 2010; Zisoulis et al., 2010). These studies provided highly detailed maps of miRNA target sites in a variety of systems. In brief, CLIP-seq and other related protocols isolate the interaction sites of RNA binding proteins by crosslinking the protein to RNA molecules using UV light, digesting unprotected RNA, and immunoprecipitating the protein-RNA complex. The bound RNA is reverse transcribed to cDNA and used to prepare sequencing libraries.

By isolating protein-RNA interactions that occur *in vivo,* CLIP-seq reveals the miRNA targeting landscape without many of the assumptions that limit other experimental approaches. The data collected from AGO CLIP-seq and PAR-CLIP experiments revealed new motifs of miRNA-target interactions (Chi et al., 2012), an abundance of coding region target sites (Chi et al., 2009; Hafner et al., 2010; Zisoulis et al., 2010), and interactions between noncoding RNAs and miRNAs (Zisoulis et al., 2012).

Despite the detailed maps of AGO binding sites generated from these studies, CLIP-based methods are not able to readily differentiate between functional and nonfunctional target sites. As a consequence, many CLIP-based experiments are complemented by the analysis of RNA or protein levels. Another limitation of CLIP-based methods is the ambiguity in which miRNA is responsible for a specific interactions (Figure 1.3D). Although many AGO binding sites contain obvious seed complementarity to a single miRNA, it is difficult to determine which miRNA is responsible for a given interaction for sites without seed complementarity or that contain seed matches to multiple miRNAs (Chi et al., 2012). The classification of AGO binding sites by seed complementarity is further complicated by the presence of miRNA families, which share the same seed sequence. In humans ~60% and in *C. elegans* ~40% of mature miRNAs belong to a miRNA family (Kozomara and Griffiths-Jones, 2011) and multiple miRNA family members can be expressed at the same time (Kato et al., 2009; Lim et al., 2003). Although miRNA families are thought to regulate overlapping sets of target RNAs due to the importance of the seed sequence in target recognition, some studies have suggested that miRNA families may have specific targets (Brennecke et al., 2005; Lin et al., 2003; Moore et al., 2015).

1.4 New tools to investigate miRNA targeting

In the past five years, several new methods and technologies have emerged which have the potential to refine our understanding of miRNA targeting. For example, the Ule lab developed individual-nucleotide resolution CLIP (iCLIP) (Konig et al., 2011; König et al., 2010) to significantly increase the recovery of unique cDNA reads and consequently the accuracy of identifying protein-RNA interactions. Similar to CLIP-seq and PAR-CLIP, the iCLIP protocol begins with crosslinking protein to RNA using UV light (Figure 1.4). RNA that is not protected by proteins is then digested with an RNase, and the resulting protein-RNA complexes are immunoprecipitated. A series of biochemical steps then prepare the isolated RNA molecules for ligation of a linker that is recognized by oligonucleotides for reverse transcription. The increased recovery of cDNAs in iCLIP is due to the inclusion of a step to circularize the cDNA after reverse transcription (Figure 1.4). This circularization step allows the capture of reverse transcription products that prematurely terminate at the crosslinking site, as well as full-length cDNAs. The circularized product is then re-linearized to allow for PCR amplification and sequencing. Comparison of CLIP-seq and iCLIP methods has shown that iCLIP can capture up to 80% more unique cDNAs (Sugimoto et al., 2012), thus making it an attractive method for identifying protein-RNA binding sites across the transcriptome.

In addition to advances in the efficiency of CLIP-like protocols, three groups have developed methods which produce miRNA-target hybrid, or chimeric, reads. These methods, such as crosslinking and sequencing of hybrids (CLASH), modified *in vivo* PAR-CLIP (iPAR-CLIP), and covalent ligation of endogenous Argonaute-bound RNAs CLIP (CLEAR-CLIP), add an intermolecular ligation step to standard

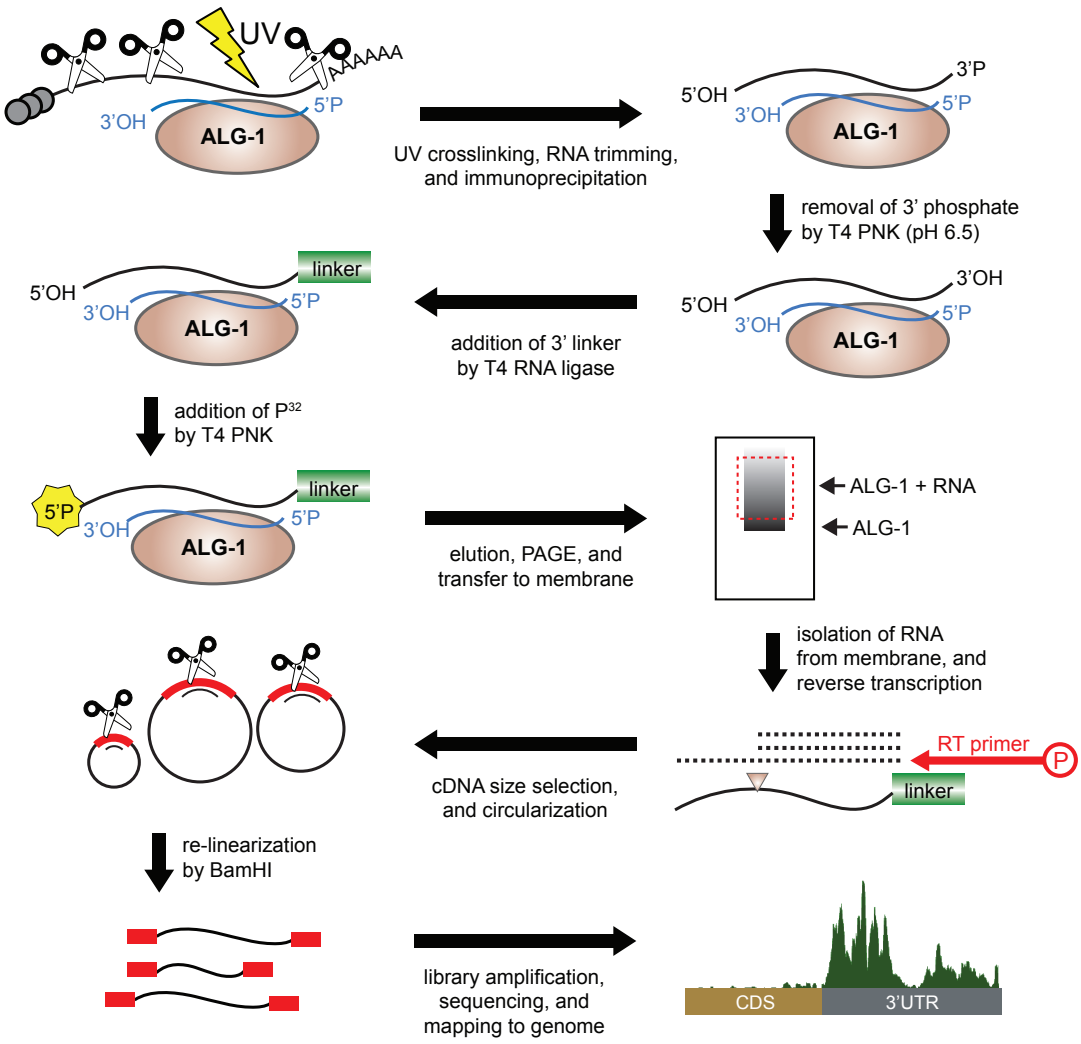


Figure 1.4: Overview of steps involved in iCLIP to identify Argonaute binding sites. See Chapter 2 for additional details.

CLIP-like methods to promote the ligation of the miRNA to its target RNA sequences (Grosswendt et al., 2014; Helwak et al., 2013; Moore et al., 2015). Chimera producing methods have the advantage of being able to unambiguously identify which miRNA is responsible for a given interaction. Although methods that produce miRNA-target chimeras offer new insights into how miRNAs find and interact with their target RNAs,

these methods are generally inefficient, with some methods generating chimeric reads that account for only 0.2% of total sequencing library reads (Grosswendt et al., 2014). These methods are also experimentally and computationally challenging for many laboratories as they take a full-week to carry out, require the use of radioactivity, and necessitate bioinformatics expertise for chimeric read identification. In addition, some research groups may be not interested in the *de novo* discovery of miRNA target sites transcriptome wide, and instead are interested in miRNA target interactions for specific target transcripts.

Many investigations into miRNA targeting have been completed using transgenic reporters, over expressing miRNAs, or knocking out multiple miRNAs (Pasquinelli, 2012; Thomas et al., 2010). However, these approaches can alter the stoichiometry between miRNAs and their targets (Mukherji et al., 2011), which may lead to artificial examples of functional regulation. Recent advances in genome editing technologies allow for the generation of double-strand breaks in DNA at pre-selected loci. These double-strand breaks can be used to generate new mutations or add tags to genes at their endogenous loci by taking advantage of DNA repair mechanisms within the cell. For example, transcription activator-like effector nucleases (TALENs) are proteins with nuclease activity that bind to pre-selected DNA sequences based on the protein's structure (Joung and Sander, 2013). Another method for genome editing utilizes the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas bacterial immune system. In its native context, the CRISPR/Cas system directs cleavage of foreign DNA, allowing it to serve as a defense against DNA viruses in bacteria (Marraffini and Sontheimer, 2010). The CRISPR/Cas system has proven to be particularly efficient and easy to work with

for making new alleles. The expression of the Cas9 RNA-binding nuclease with an engineered single guide RNA (sgRNA) has been shown to direct double-strand breaks in eukaryotic genomes (Jinek et al., 2012). Since the sequence of the sgRNA, which provides the targeting specificity to Cas9, can be easily changed, CRISPR/Cas9 genome editing is easily employed to generate new genetic tools and make custom mutations in the endogenous context. For example, miRNA target sites can be manipulated using CRISPR/Cas9 genome editing while maintaining the stoichiometry between miRNAs and their targets. Furthermore, new tools to study miRNA pathway genes can be rapidly developed.

By taking advantage of these improved methods and novel technologies, I have been able to generate the highest-resolution map of miRNA targeting in *C. elegans* to date, as detailed in Chapter 4. Specifically, I adapted and further optimized the increased resolution and accuracy of iCLIP to identify AGO binding sites in *C. elegans* (Broughton and Pasquinelli, 2013), described in Chapter 2. Furthermore, I discovered that AGO iCLIP generates miRNA-target chimeric reads at similar efficiencies to methods that are specifically designed to produce these reads. These chimeric reads are broadly discussed in Chapter 3. By analyzing over 150,000 miRNA-target chimeric reads, I found that the majority of miRNA target sites have complementarity to the miRNA 3' end. These 3' end interaction appear to affect miRNA-targeting specificity in the case of miRNA families, and these 3' end interactions are important for assigning miRNA-targeting specificity *in vivo*. In addition, I developed a novel method, called Chimera PCR (ChimP), which provides evidence of miRNA-target RNA interactions without the requirement of analyzing next-generation sequencing datasets. Finally, my work described in Chapter 5 has

contributed to research on the roles of the highly related miRNA-pathway AGOs, ALG-1 and ALG-2, in regulating the lifespan of *C. elegans*.

Chapter 2

Identifying Argonaute binding sites in *Caenorhabditis elegans* using iCLIP

2.1 Abstract

The identification of endogenous targets remains an important challenge in understanding microRNA (miRNA) function. Past approaches using *in silico* methods and reporter constructs lack biological context that may enhance or inhibit target recognition. To address these limitations, several labs have utilized crosslinking and immunoprecipitation (CLIP) of Argonaute (Ago) proteins to identify miRNA targets. Recently, the Ule Lab introduced individual-nucleotide resolution CLIP (iCLIP) to increase the sensitivity of identifying protein-RNA interaction sites. Here we adapt the iCLIP protocol for use in *Caenorhabditis elegans* to identify endogenous sites targeted by the worm Argonaute (ALG-1) primarily responsible for miRNA function.

2.2 Introduction

In many eukaryotic organisms, small RNA molecules, known as microRNAs (miRNAs), provide post-transcriptional control of gene expression through imperfect binding to target RNAs. This flexibility in targeting allows each miRNA to potentially regulate hundreds of different transcripts in a range of pathways (Pasquinelli, 2012). As a consequence of their broad regulatory potential, altered miRNA expression is implicated in many diseases (Sayed and Abdellatif, 2011).

The majority of miRNA biogenesis begins with transcription of a primary-miRNA (pri-miRNA) transcript by RNA polymerase (Finnegan and Pasquinelli, 2012). This long, poly-adenylated transcript is capable of forming an imperfectly paired stem-loop structure and is processed by Drosha and DGCR8/Pasha into a ~65 nucleotide (nt) long hairpin precursor-miRNA (pre-miRNA). Maturation of the miRNA occurs after subsequent Dicer processing of the pre-miRNA into its final ~21 nt form. Mature miRNAs are loaded onto an Argonaute (Ago) protein as part of the miRNA Induced Silencing Complex (miRISC). Classically, miRNAs are thought to help direct the miRISC to the 3'UTR of messenger RNAs (mRNAs), resulting in a decrease in gene expression through mRNA destabilization and/or translational repression (Pasquinelli, 2012). More recently, it has been shown that miRNAs can target regions throughout mRNAs (Chi et al., 2009; Hafner et al., 2010; Zisoulis et al., 2010) and non-coding RNAs (Zisoulis et al., 2012).

Over the past decade, one of the greatest challenges in the field has been identifying miRNA target sites. Predicting targets is complicated by the limited sequence information provided by miRNAs due to their small size and their ability to function even when only a partial duplex is formed (Pasquinelli, 2012). In these

cases, as few as 7 nts can provide functional targeting information to the miRISC (Bartel, 2009). *In silico* approaches are often based on common motifs, such as the "seed" sequence found in nucleotides 2-7 of the miRNA, and the propensity for miRNAs to bind to the 3'UTR of messenger RNAs. However, these programs often miss non-canonical target sites and have high rates of false positives, as the presence of a binding site does not ensure regulation. Furthermore, there is little consensus among miRNA target prediction programs. To address these problems, several labs have utilized high-throughput sequencing following ultraviolet (UV) crosslinking and immunoprecipitation (CLIP-seq or HITS-CLIP) and other similar methods to identify Argonaute/miRNA binding sites (Chi et al., 2009; Hafner et al., 2010; Kishore et al., 2011; Leung et al., 2011; Stark et al., 2012; Zisoulis et al., 2010). In HITS-CLIP, target RNAs are identified by first crosslinking with UV light, which generates a covalent bond between the protein of interest and the RNA. These protein-RNA complexes are then immunoprecipitated. By subsequently adding 5' and 3' adapters and removing the protein through Proteinase K digestion, the isolated target RNAs are reverse-transcribed, amplified, and identified through high-throughput sequencing. These studies have revealed the prevalence of non-3'UTR binding sites in target mRNAs (Chi et al., 2009; Hafner et al., 2010; Zisoulis et al., 2010), new binding motifs (Chi et al., 2012), and have the potential to improve *in silico* target prediction approaches.

Since its development, HITS-CLIP has become increasingly used for determining target sites for a range of RNA-binding proteins, particularly in the fields of splicing and small RNA research. At first, miRNA target site identification was limited to the size of the RNA fragment isolated after RNA trimming and immunoprecipitation.

However, analysis of sequencing data revealed that mutations and deletions occur during generation of the cDNA library at the crosslinking site due to the presence of an amino acid or short peptide that remains attached after Proteinase K treatment. From these data it is possible to determine RNA-protein interaction locations with increased resolution (Granneman et al., 2009; Zhang and Darnell, 2011). An adaptation on HITS-CLIP, known as photoactivatable-ribonucleoside-enhanced (PAR)-CLIP, uses photoreactive ribonucleoside analogs and UV-A light to increase crosslink efficiency and the incidence of point mutations at the crosslinking site (Hafner et al., 2010). However, PAR-CLIP requires tissues and cells to be pre-treated with the photoreactive ribonucleoside analogs before crosslinking. This may result in toxic effects (Lozzio and Wigler, 1971) that may alter endogenous conditions. Furthermore, primer extension studies and analysis of HITS-CLIP data have demonstrated that a significant portion of reverse transcription products terminate at the site of crosslinking (Sugimoto et al., 2012; Urlaub et al., 2002).

To address these problems the Ule Lab developed a new method known as individual-nucleotide resolution CLIP (iCLIP) (König et al., 2010; Konig et al., 2011). In brief, iCLIP is designed to isolate all cDNA products generated during reverse transcription, including cDNAs that truncate at the crosslinking site. Additionally, iCLIP uses barcoded primers during generation of the cDNA library to control for PCR artifacts that may occur with library amplification (Wang et al., 2010). A recent analysis of HITS-CLIP and iCLIP datasets has revealed that 80% of cDNAs truncate at the crosslink site and are absent in HITS-CLIP data (Sugimoto et al., 2012). This analysis also showed that the use of RNase I, which can cleave between any ribonucleotide pairs, in iCLIP alleviates the bias in target site identification that results

from the use of other enzymes during RNA trimming. These advantages allow iCLIP to recover additional reads and provide increased confidence in the identification of bona fide target sites.

As a complement to both HITS-CLIP and PAR-CLIP, which have been previously adapted for use in *Caenorhabditis elegans* (Jungkamp et al., 2011; Zisoulis et al., 2010), we have developed a procedure to bring the advantages of the iCLIP protocol (Konig et al., 2011) to *C. elegans* for the identification of Argonaute/miRNA regulated RNAs (Figure 2.1). For this application *C. elegans* offers several advantages: (1) Argonaute-like-gene 1 (ALG-1) is largely responsible for miRNA function, (2) crosslinking in a living animal preserves ALG-1:RNA binding context, and (3) viable *alg-1* genetic mutants exist to use as specificity controls. These advantages, along with those of iCLIP, allow us to identify miRNA:mRNA interaction sites in an entire organism with increased sensitivity and resolution.

2.3 iCLIP Protocol

2.3.1 Materials

Table 2.1: Buffers used in ALG-1 iCLIP

Buffer	Composition
M9 Buffer	22mM KH2PO4; 22mM Na2HPO4; 85mM NaCl; 1mM MgSO4
Homogenization Buffer	100mM NaCl; 25mM HEPES, pH 7.5; 250µM EDTA, pH 8.0; 0.1% (w/v) NP-40; 2mM DTT; 25U/mL rRNasin (Promega); 1 tablet / 20mL Complete Mini Protease Inhibitor (Roche)

Table 2.1: Buffers used in ALG-1 iCLIP (continued)

Buffer	Composition
RNase I Dilution Buffer	10mM Tris-HCl, pH 8.0; 100mM NaCl; 50% Glycerol
High Salt Wash Buffer	50mM Tris-HCl, pH 7.4; 1M NaCl; 1mM EDTA; 1% NP-40; 0.1% SDS; 0.5% sodium deoxycholate
Wash Buffer	20mM Tris-HCl, pH 7.4; 10mM MgCl2; 0.2% Tween-20
5X PNK Buffer, pH 6.5	350mM Tris-HCl, pH 6.5; 50mM MgCl2; 25mM DTT
4X Ligation Buffer	200 mM Tris-HCl, pH 7.4; 40mM MgCl2; 40mM DTT
MOPS-SDS Running Buffer	50mM MOPS; 50mM Tris-base; 1mM EDTA; 0.1% SDS
Bis-Tris Supplemented Transfer Buffer	25mM Bicine; 25mM Bis-Tris; 1mM EDTA; 5% Methanol; 0.01% SDS
PK Buffer	100mM Tris-HCl, pH 7.4; 50mM NaCl; 10mM EDTA
7M Urea / PK Buffer (must be made fresh)	100mM Tris-HCl, pH 7.4; 50mM NaCl; 10mM EDTA; 7M urea
Gel Extraction Buffer	300mM NaCl; 10mM Tris-HCl, pH 8.5
2X TBE-Urea Sample Buffer	90mM Tris-Base; 90mM Boric acid; 2mM EDTA; 12% Ficoll Type 400; 7M Urea; 0.01% Bromophenol Blue; 0.05% Xylene Cyanol

Table 2.2: Enzymes and other materials used in ALG-1 iCLIP

Material	Supplier
Protein G Dynabeads	Life Technologies
α-ALG-1 antibody (PA1-031)	Thermo Scientific (Pierce Antibodies)
Turbo DNase	Ambion
RNase I	Ambion
PNK Enzyme	NEB

Table 2.2: Enzymes and other materials used in ALG-1 iCLIP (continued)

Material	Supplier
rRNasin	Promega
T4 RNA Ligase I	NEB
PEG400	Sigma Aldrich
32P-γ-ATP	PerkinElmer
4x NuPAGE LDS Sample Buffer	Life Technologies
4-12% Bis-Tris NuPAGE Gel	Life Technologies
NuPAGE Antioxidant	Life Technologies
Nitrocellulose membrane (0.45µm pore)	Bio-Rad
Proteinase K (fungal)	Roche
Acid Phenol / Chloroform (5:1)	Ambion
Chloroform:IAA (25:1)	Sigma Aldrich
UreaGel System	National Diagnostics
Dark Reader Transilluminator	Clare Chemical Research
SYBR-Gold	Life Technologies
0.4µm Costar-X spin column	Corning
CircLigase II	Epicentre
BamHI	NEB
Accuprime Supermix I	Life Technologies
6% TBE gel	Life Technologies

Table 2.3: Oligonucleotides used in ALG-1 iCLIP, all oligonucleotides were obtained from IDT

Identifier	Sequence
L3 (HPLC Purified)	/5rApp/AGATCGGAAGAGCGGTTCAG/3ddc/
Rclip, XXXX = a unique four nucleotide sequence	/5phos/NNXXXXNNNAGATCGGAAGAGCG TCGTGgatcCTGAACCGC
Cut_oligo_ddc	GTTCAGGATCCACGACGCTCTTC/ddC/
P5Solexa (PAGE Purified)	AATGATACGGCGACCACCGAGATCTACA CTCTTTCCCTACACGACGCTCTTCC-GATCT
P3Solexa (PAGE Purified)	CAAGCAGAAGACGGCATACGAGATCGG TCTCGGCATTCCTGCTGAACCGCTCTTCC-GATCT
lin-41 LCS Primer 1	GAGGCAGAATGGTTGTATAA
lin-41 LCS Primer 2	ttatacaaccgttctacactca

Table 2.3: Oligonucleotides used in ALG-1 iCLIP (continued)

Identifier	Sequence
lin-41 Control Primer 1	ACATGTTTCTGGGCGATAGG
lin-41 Control Primer 2	CGTGCTGTTGGCTACTTCAA

2.3.2 Procedure

Growing worms and preparing lysates

1. Grow worms as desired so that at least 50,000 worms or enough to obtain ~1mg of total protein per sample.

2. Wash worms (3x) with 10mL M9 in 15mL tube. Resuspend in 5mL M9 after washes.

3. Rock worms for 20min on nutator at temperature used for culturing of samples. Spin down worms in clinical centrifuge for 30 seconds.

4. Plate worms on Nematode Growth Medium (NGM) plates with no food at room temperature. Be sure to avoid areas of high concentration and overlapping worms.

5. Place the NGM plates with worms into Spectrolinker XL-1000 (without the plate lids) and irradiate with UV-B at 3kJ/m2 (Energy Settings: 3000). Check worms using microscope to be sure all worms have been completely immobilized. Collect worms in M9 buffer and transfer to 2mL round bottom microcentrifuge tube. Spin worms for 30 seconds in desktop centrifuge.

6. Remove excess M9 buffer and resuspend the worms in 700µL of ice-cold homogenization buffer. Keep worms on ice.

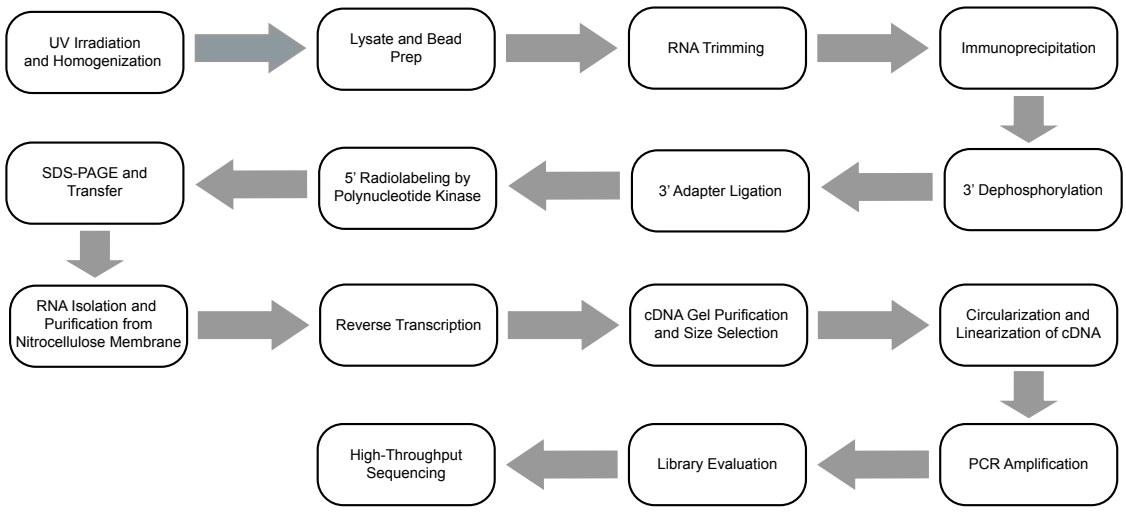


Figure 2.1: Outline of the ALG-1 iCLIP method: i) *in vivo* crosslinking stabilizes RNA-protein interactions and lysates are generated by sonication of immobilized animals; ii) lysates and beads are prepared for digestion and immunoprecipitation; iii) RNA molecules are trimmed by RNase I; iv) ALG-1 and crosslinked RNAs are immunoprecipitated using specific antibody and stringent wash conditions; v) the 2' 3' cyclic monophosphate generated by RNase I is removed by Polynucleotide Kinase in low pH (6.5) conditions; vi) ligation of adapter oligo to the 3' hydroxyl group of the RNA fragments (circularization and concatemer formation events are prevented by the presence of 3' dideoxy cytosine on the adapter); vii) the 5' end of the RNA is radiolabeled by Polynucleotide Kinase using 32P-γ-ATP; viii) samples are analyzed by electrophoresis on native SDS-PAGE under reducing conditions and transferred to a nitrocellulose membrane to isolate the band corresponding to ALG-1/RNA complexes; ix) ALG-1 is removed from the RNA by Proteinase K digestion and the RNA is purified; x) reverse transcription of isolated RNA using phosphorylated oligo; xi) cDNA is separated by size on denaturing PAGE and purified from the gel to remove sequences that are too long or consist of only adapter and oligo; xii) purified cDNA is circularized, an adapter is added that generates a specific dsDNA region in the sequence of the reverse transcription primer, and is linearized by restriction enzyme digest; xiii) PCR amplification; xiv) evaluation of library by complexity, size, and concentration; xv) high-throughput sequencing using the Illumina GAIIx system.

7. Sonicate the worms (5x) with 10 second pulses and 50 seconds resting on ice between pulses (~18W RMS output power on Dismembrator Model 100, Fisher Scientific).

8. Spin lysates at 16000xg for 15 minutes at 4°C. Transfer supernatant to new 1.5mL microcentrifuge tube.

9. Freeze lysates in dry-ice-ethanol bath. Store at -80°C.

RNA trimming and immunoprecipitation

1. Thaw lysates at 4°C on nutator for 30 minutes.

2. Quantify the protein content of lysate samples using bradford assay. Adjust lysate concentrations to be equal between all samples (~1-3mg/mL) with ice-cold homogenization buffer.

3. Save 20µL of sample for western blot analysis (input).

4. While quantifying, aliquot 50µL Protein G Dynabeads per sample. Wash beads (2x) with homogenization buffer.

5. Resuspend beads in 100µL of homogenization buffer.

6. Preclear lysates by adding 100µL 50:50 bead slurry and rocking on nutator for 1 hour at 4°C.

7. Pellet beads+lysates using magnetic rack. Transfer lysate to new chilled 1.5mL microcentrifuge tube. Keep samples on ice.

8. Add 2µL TurboDNase and 10µL RNase I dilution (1/500 RNase I dilutions are used for library preparation, while 1/50 dilutions are used to control for antibody specificity).

9. Incubate samples for 3min in Thermomixer R set to 37°C and shaking at 1100rpm. Immediately transfer samples to ice.

10. Add 7µg α-ALG-1 antibody to lysate. Rock lysate+antibody at 4°C overnight.
11. Aliquot 100µL Protein G Dynabeads per sample. Wash beads (2x) with homogenization buffer.
12. Add beads to lysate+antibody. Incubate at 4°C for 1 hour.
13. Collect the beads with magnetic rack.
14. Save 20µL of supernatant for western blot analysis (supernatant).
15. Wash (2x) the beads with 900µL ice-cold high salt wash buffer.
16. Wash (2x) the beads with 900µL ice-cold wash buffer.

Dephosphorylation of RNA 3' ends

1. Collect the beads with magnetic rack. Discard supernatant and resuspend beads in 20µL dephosphoylation mix (15µL ddH2O; 4µL 5X PNK, pH 6.5 buffer; 0.5µL PNK enzyme; 0.5µL rRNasin).
2. Incubate for 20 minutes at 37°C.
3. Add 500µL wash buffer.
4. Wash (1x) with 900µL high-salt buffer.
5. Wash (2x) with 900µL wash buffer.

Linker ligation to RNA 3' end

1. Collect the beads with magnetic rack. Discard supernatant and resuspend beads in 16µL ligation mix (9µL ddH2O; 4µL 4X ligation buffer; 1µL T4 ssRNA Ligase 1; 0.5µL rRNasin; 1.5µL 20µM L3 oligo).

2. Add 4µL PEG(400) to each reaction.
3. Incubate overnight in Thermomixer R at 16°C with intermittent shaking: 1300rpm for 15 seconds every 5min.
4. Add 500µL wash buffer.
5. Wash (2x) with 900µL high salt wash buffer.
6. Wash (2x) with 900µL wash buffer.

End labeling of RNA 5' end

1. Remove the supernatant and resuspend the beads in 80µL of hot PNK mix (67µL ddH2O; 8µL 10X PNK buffer; 1µL P32-Îş-ATP; 4µL T4 PNK enzyme).
2. Incubate for 10 minutes in Thermomixer R at 37°C with intermittent shaking: 1200rpm for 15 seconds every 4min.
3. Add 500µL wash buffer.
4. Collect beads with magnetic rack and remove supernatant. Dispose of liquid as radioactive waste.
5. Resuspend the beads in 30µL wash buffer and 10µL 4x NuPAGE LDS sample buffer. Add 2µL 1M DTT.
6. Incubate in Thermomixer R at 70°C for 10 minutes at 1200rpm.
7. Immediately place samples on magnetic rack.

SDS-PAGE and nitrocellulose transfer

1. Load the samples on a 4-12% Bis-Tris gel with 1X MOPS-SDS running buffer. Before running add 500µL NuPAGE Antioxidant to upper chamber of running buffer. Run gel at 180V.

2. Transfer the RNA-protein complexes from the gel to a pure nitrocellulose membrane (0.45µm pore size) with Bis-Tris supplemented transfer buffer at 40V for 4 hours at 4°C.

3. After the transfer, rinse the membrane in 1X PBS buffer. Wrap membrane in saran wrap and expose it to autoradiography film. Perform 1 hour, 2 hour, and overnight exposures. See Figure 2.2.

4. Using an exposed piece of film as a mask, cut out the band corresponding to the ALG-1/RNA complexes in the low RNase condition with a clean razor blade. Cut isolated membrane into smaller pieces and place into a 1.5mL microcentrifuge tube. Pieces of membrane can be stored at -80C.

RNA isolation of purification

1. Prepare 4mg/mL Proteinase K solution in 1X PK buffer and incubate this solution at 37°C for 20 minutes to deactivate any RNases that may be present.

2. Add 200µL of prepared Proteinase K solution to each tube of isolated nitrocellulose pieces and incubate for 20 minutes at 37°C at 1200rpm.

3. Add 200µL of 7M urea / 1X PK Buffer and incubate for 20 minutes at 37°C at 1200rpm.

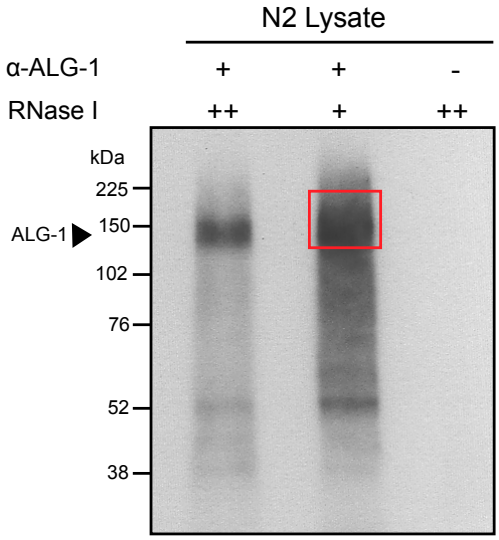


Figure 2.2: ALG-1 iCLIP autoradiograph. Autoradiograph of protein cross-linked to RNA fragments. Black arrow indicates expected migration size of ALG-1. Red square denotes region of membrane that was excised for RNA isolation and purification. RNA was partially digested by RNase I in either high (++) or low (+) concentrations. High (++) RNase treatments serve as a control for antibody specificity, as the band should be less diffuse than that of low RNase. The ALG-1 signal is absent in the no-antibody control. The bands seen below the ALG-1 signal (~52 and 38kDa) are non-specific as they are observed in the no-antibody lane in longer exposures, while the signal at ~150kDa is not.

4. Add 500µL of Acid Phenol/Chloroform. Briefly shake tubes by hand. Incubate at 37°C for 20 minutes at 37°C at 1200rpm.

5. Mix tube vigorously by hand for a few seconds and centrifuge the tubes at 16000xg for 5 minutes at room temperature.

6. Transfer aqueous layer to new 1.5mL microcentrifuge tube. Add 400µL chloroform:IAA. Mix briefly by hand and centrifuge the tubes at 16000xg for 5 minutes at room temperature.

7. Transfer the aqueous layer to a new 1.5mL microcentrifuge tube and add 40µL 3M sodium acetate, pH 5.2; 1µL 20mg/mL glycogen; 1mL 1:1

ethanol:isopropanol. Mix.

8. Precipitate overnight at -20°C.

Reverse transcription

1. Spin for 20 minutes at 16000xg at 4°C.
2. Remove the supernatant and wash the pellet with 500µL 80% ethanol.
3. Air dry pellet for exactly 8 minutes.
4. Resuspend the pellet in 7.25µL RNA/primer mix (use different Rclip primer for each experiment/replicate): 6.25µL ddH2O; 0.5µL 0.5µM Rclip primer; 0.5µL 10mM dNTP.
5. Incubate for 5 minutes at 70°C before cooling to 25°C.
6. Add 2.75µL RT mix: 2µL 5X First-Strand RT buffer; 0.5µL 0.1M DTT; 0.25µL Superscript III.
7. Incubate 5 minutes at 25°C, 20 minutes at 42°C, 40 minutes at 50°C, and 5 minutes at 80°C before cooling to 4°C.
8. Add the following: 90µL TE buffer; 10µL 3M sodium acetate, pH 5.2; 1µL 20mg/mL glycogen; 250µL 100% ethanol. Mix.
9. Precipitate overnight at -20°C.

Gel purification of cDNA

1. Pour 6% TBE-urea gel using UreaGel System (2.4mL concentrate, 6.6mL diluent, 1mL system buffer, 90µL 10% APS, 4µL TEMED).

2. Allow gel to polymerize for at least 1 hour.
3. Pre-run gel with 1X TBE buffer at 150V for 30min.
4. Spin for 20 minutes at 16000xg at 4°C.
5. Remove the supernatant and wash the pellet with 500µL 80% ethanol.
6. Air dry pellet for exactly 8 minutes.
7. Resuspend pellet in 10µL ddH2O and add 10µL 2x TBE-urea loading buffer. Heat samples to 80°C for 3 minutes directly before loading.
8. Load samples and low molecular weight marker on prepared 6% TBE-Urea gel and run at 180V.
9. Remove gel from glass plates and stain with SYBR-GOLD for 10 minutes.
10. Prepare 3 0.5mL microcentrifuge tubes for each sample by poking 3-4 holes in the bottom of each tube with 21.5 gauge needle.
11. Using dark-reader, cut three sections of gel from each lane that correspond to the sizes: 200-120nt (high), 120-85nt (medium), and 85-70nt (low). See Figure 2.3.
12. Place each gel slice into prepared 0.5mL microcentrifuge tubes.
13. Centrifuge gel pieces at 16000xg for 2 minutes to fragment the gel slices.
14. Add 400µL gel extraction buffer.
15. Rock on nutator for 2 days at 4°C.
16. Transfer supernatant and gel pieces to 0.45µm filter column and spin at 600xg for 10 seconds.

17. To flow-through add: 40µL 3M sodium acetate, pH 5.2; 1µL 20mg/mL glycogen; 1mL 100% ethanol. Mix.

18. Precipitate overnight at -20°C.

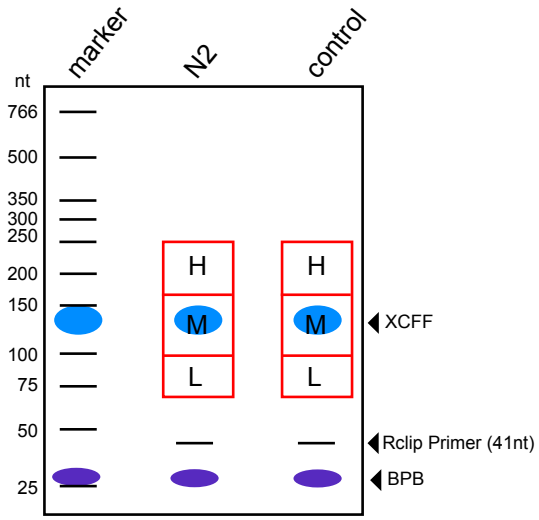


Figure 2.3: Representation of denaturing PAGE used for isolating cDNA. The gel is stained with SYBR-gold and viewed using a Dark Reader. Note that only the ladder and Rclip primer will be visible at this point, as the cDNA is not in high enough concentrations to be detected. Using a clean razor blade, the lanes corresponding to each sample are removed (red lines). The lanes are divided into high (H), medium (M), and low (L) fractions. Avoid cutting below the region of 70 nts, as the presence of adapter-Rclip primer product (52 nts) can lead to libraries that are not sufficiently complex for analysis. XCFF, Xylene cyanol FF; BPB, Bromophenol blue.

Circularization and linearization of cDNA

1. Spin for 20 minutes at 16000xg at 4°C.

2. Remove the supernatant and wash the pellet with 500µL 80% ethanol.

3. Air dry pellet for exactly 8 minutes.

4. Resuspend in 8µL ssDNA ligation mix: 6.5µL ddH2O; 0.8µL 10X CircLigase II buffer; 0.4µL 50mM MnCl2; 0.3 CircLigase II enzyme.
5. Incubate at 60°C for 1 hour.
6. Heat inactivate enzyme by incubating for 15 minutes at 80°C.
7. Add 30µL oligo annealing mix: 23µL ddH2O; 3µL 10X NEBuffer 3; 3µL 10X BSA; 1µL 10µM Cut_oligo_ddc.
8. Incubate for 1 minute at 95°C for 1 minute, then decrease the temperature every 20 seconds by 1°C until 25°C has been reached. (This can be accomplished using the Thermomixer R and changing the temperature from 95°C to 25°C, which will take approximately 25 minutes to adjust.)
9. Add 2µL BamHI and incubate for 30 minutes at 37°C.
10. Add the following: 50µL TE buffer; 10µL 3M sodium acetate, pH 5.2; 1µL 20mg/mL glycogen; 300µL 100
11. Precipitate overnight at -20°C.

PCR amplification

1. Spin for 20 minutes at 16000xg at 4°C.
2. Remove the supernatant and wash the pellet with 500µL 80% ethanol.
3. Air dry pellet for exactly 8 minutes.
4. Resuspend pellet in 19µL ddH2O and transfer to a PCR tube.
5. To the cDNA, add 1µL P5/P3 Solexa Mix (10µM each) and 20µL Accuprime Supermix I.

6. Run the following PCR Program: 94°C – 2min, [94°C – 15sec, 65°C – 30sec, 68°C – 30sec] 25-35 cycles, 68°C – 3min, 4°C hold.

7. Mix 8µL PCR product with 2µL 5xTBE loading buffer and load on a precast 6% TBE gel. Stain the gel with SYBR-GOLD or ethidium-bromide and image. See Figure 2.4A.

Quantification and sequencing

1. Quantify library concentration using QPCR and bioanalyzer.

2. Submit 15µL of the library for multiplex sequencing and store the rest at -20°C

PCR control for ALG-1 binding specificity

1. Prepare PCR reaction (1µL iCLIP library, 17.25µL ddH2O; 5µL 5X Go-Taq Buffer; 1µL 25mM MgCl2; 0.5µL 10mM dNTPs; 0.5µL lin-41 LCS Primer 1+2 Mix (10µM each); 0.25µL Go-Taq polymerase).

2. Run the following PCR program: 94°C – 2min, [94°C – 15sec, 48°C 30sec, 72°C – 30sec] 38 cycles, 72°C – 3min, 4°C hold. When using *lin-41* Control Primers 1+2, anneal at 55°C.

3. Run 10µL PCR product on 2% agarose gel and stain with ethidium-bromide. Expected product size of *lin-41* LCS 1+2 is 60 nts. See Figure 2.4B.

2.3.3 Notes

We have made several modifications to the iCLIP protocol (Konig et al., 2011). First, consistent with our previously published protocol for HITS-CLIP (Zisoulis et al., 2011), we preclear lysates before RNA trimming and immunoprecipitation. This

reduces background seen on the autoradiograph from non-specific interaction with the beads. If using another antibody, we recommend users to thoroughly evaluate immunoprecipitation conditions before proceeding with iCLIP. To further improve the autoradiograph, we prepare RNA-protein samples under reducing conditions to prevent the antibody complex from interfering with the ALG-1 signal. As can be seen in Figure 2.2, the difference between high and low RNase conditions is not as striking as seen with smaller proteins. For assessing immunoprecipitation efficiency, we recommend saving protein samples during the immunoprecipitation and after ligation of the 3' adapter for subsequent western blot analysis. To ensure complete recovery of cDNA from denaturing PAGE we have lengthened the extraction time from 2 hours at 37°C to 2 days at 4°C. In some cases, we had difficulty with ligation of the adapter used to generate dsDNA for restriction enzyme digest (Cut_oligo) to the reverse transcription primer (Rclip). This generated a distinct band (112 nts) in our evaluation of the cDNA library (Figure 2.4A), which accounted for ~90% of the PCR amplified library in some cases. To address this issue, we added a heat inactivation step after circularization and designed an adapter that has 3' dideoxy cytosine instead of 4 adenosines. These additions eliminated these non-specific products and led to cleaner libraries (Figure 2.4A). As an assessment of the quality of the library, we recommend an additional PCR step to detect the sequences known to be bound by the miRNA complex in the sample (e.g. the let-7 complementary site in the *lin-41* 3' UTR for extracts from L4 stage worms), as this provides evidence of the sensitivity and specificity of the library (Figure 2.4B). As a control, we recommend using primers that amplify an exon of *lin-41* that is not known to be bound by ALG-1.

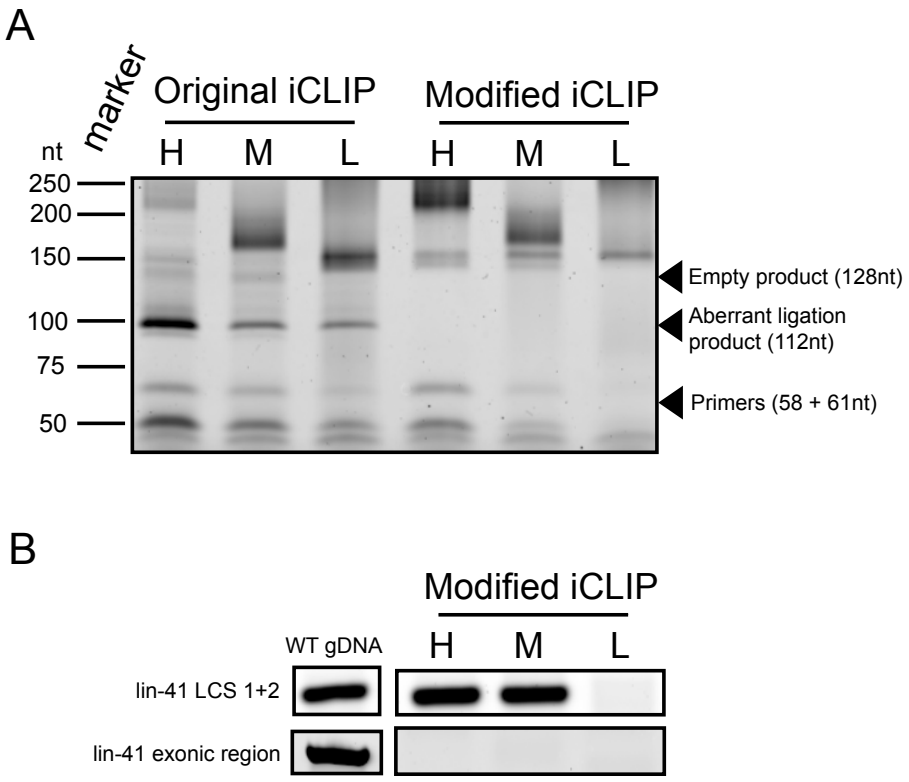


Figure 2.4: Comparison of ALG-1 iCLIP methods and control PCR (A) Comparison of cDNA libraries from the original iCLIP protocol and the modified iCLIP protocol presented here by non-denaturing PAGE. Diffuse bands corresponding to the range of sizes for each cDNA fraction removed from denaturing PAGE with 76 nts of extension by PCR primers should be observed. The expected migration size of the primer-dimer (128 nts) that contains no sequence information, the aberrant ligation product (112 nts), and the primers (58 nts and 61 nts) are indicated by black arrows. The primer-dimer product should not be present in libraries submitted for sequencing. A band corresponding to miRNA only product (~150nt) may be seen in some samples. (B) PCR controls for cDNA library specificity. The let-7 complement site (LCS) in the 3'UTR of *lin-41* is a well characterized ALG-1 interaction site and is expected to be present in iCLIP cDNA libraries. PCR of this region is expected to produce a product, while PCR of exonic regions of *lin-41*, that are not known to be bound by ALG-1, should not produce a product. Control PCRs for WT genomic DNA (gDNA) are shown.

2.4 Conclusions

The use of CLIP-based approaches to identify miRNA target sites has furthered the understanding of how these small RNAs are able to interact with their

target mRNAs (Chi et al., 2012), revealed a novel miRNA interaction with a non-coding RNA (Zisoulis et al., 2012), and provided comprehensive lists of Ago/miRNA-mediated interactions (Chi et al., 2009; Hafner et al., 2010; Kishore et al., 2011; Leung et al., 2011; Stark et al., 2012; Zisoulis et al., 2010). Here we have brought an advancement on the original CLIP methodology, iCLIP, to *C. elegans*, a powerful genetic model. The increased recovery of transcripts, as well as the ability to identify crosslink sites through mutations and the truncated cDNAs, allows iCLIP to be significantly more sensitive than other CLIP-based methods (Sugimoto et al., 2012). Improvements such as iCLIP and other future innovations to CLIP-based approaches have the potential to generate comprehensive lists of Ago/miRNA targets in a variety of specific cells, tissues, and genetic backgrounds that will increase our understanding of miRNA function.

2.5 Acknowledgements

We thank Thomas Stark and Dr. Antti Aalto for critical reading of the manuscript. We acknowledge that this protocol is an adaptation of original methods developed by the Ule lab. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1144086.

Chapter 2, in full, is a reprint of the material as it occurs in *Methods*, Broughton, J.P. and Pasquinelli A.E., Elsevier Inc., 2013. Broughton, J.P. was the primary author of this paper.

Chapter 3

A tale of two sequences: microRNA-target chimeric reads

3.1 Abstract

In animals, a functional interaction between a microRNA (miRNA) and its target RNA requires only partial base pairing. The limited number of base pair interactions required for miRNA targeting provides miRNAs with broad regulatory potential and also makes target prediction challenging. Computational approaches to target prediction have focused on identifying miRNA target sites based on known sequence features that are important for canonical targeting and may miss non-canonical targets. Current state-of-the-art experimental approaches, such as CLIP-seq (cross-linking immunoprecipitation with sequencing), PAR-CLIP (photoactivatable-ribonucleoside-enhanced CLIP), and iCLIP (individual-nucleotide resolution CLIP), require inference of which miRNA is bound at each site. Recently, the development of methods to ligate miRNAs to their target RNAs during the preparation of sequencing libraries has provided a new tool for the identification of miRNA

target sites. The chimeric, or hybrid, miRNA-target reads that are produced by these methods unambiguously identify the miRNA bound at a specific target site. The information provided by these chimeric reads has revealed extensive non-canonical interactions between miRNAs and their target mRNAs, and identified many novel interactions between miRNAs and noncoding RNAs.

3.2 Background

3.2.1 Target recognition by miRNAs

MicroRNAs (miRNAs) are an important class of regulatory molecules that function to target specific RNAs for posttranscriptional regulation (Hausser and Zavolan, 2014). Prevalent in animals and plants, miRNAs are small (~22 nucleotides), noncoding RNAs (ncRNAs) that bind to Argonaute (AGO) proteins. Once bound to Argonaute, as part of the miRNA induced silencing complex (miRISC), the miRNA guides miRISC to target RNAs. In animals, these target sites are usually located in the 3' untranslated region (3'UTR) of the mRNA, but may also reside within the coding sequence or 5'UTR. Protein production from mRNAs targeted by miRNAs is subsequently repressed due to inhibition of translation and transcript destabilization.

In animals, miRNAs interact with their targets through imperfect base pairing. The limited sequence interactions required by miRNAs to direct regulation allows a single miRNA to potentially regulate hundreds of targets in multiple pathways. Although miRNAs are flexible in their targeting ability, a large body of work has proposed a series of rules that predict canonical miRNA targeting (Bartel, 2009; Pasquinelli, 2012). Nucleotides 2-8 at the 5' end of the miRNA are known as the seed sequence and are important for miRNA target recognition. Crystal structures of miRNAs bound to Argonaute proteins have suggested that the seed sequence is

favorably positioned for initiating the interaction between miRNAs and their target RNAs (Elkayam et al., 2012; Schirle and MacRae, 2012; Schirle et al., 2014).

Perfect seed complementarity defines canonical targeting, but there are a variety of examples of imperfect or non-seed interactions (Pasquinelli, 2012). However, the extent to which miRNAs interact with their targets non-canonically and whether these targets are functional remain unclear (Agarwal et al., 2015). Additionally, recent evidence has suggested that miRNAs may have functional interactions with other ncRNAs (Hansen et al., 2013; Memczak et al., 2013; Zisoulis et al., 2012). The prevalence of these interactions is an open question.

3.2.2 Challenges in the identification of miRNA targets

The identification of miRNA target sites remains an outstanding challenge. In particular, pinpointing miRNA target sites is complicated due to the small size of miRNAs and their ability to functionally interact with their targets through imperfect base pairing. These two constraints limit the sequence information that can be used to predict targets, while also allowing a single miRNA to potentially regulate hundreds of targets.

Various research groups have developed computational approaches to predict miRNA target sites. For example, the commonly cited TargetScan algorithm was originally designed to predict target sites by looking for seed sequence complementarity and conservation in 3'UTRs (Lewis et al., 2003, 2005). However, computational prediction programs are in general limited by the current understanding of miRNA targets and may miss unexpected functional interactions, such as those between a miRNA and another noncoding RNA (Zisoulis et al., 2012). In addition, comparisons of miRNA target prediction algorithms show that there is limited overlap between the targets predicted by various programs (Min and Yoon, 2010). This suggests

that many targets identified by current miRNA target prediction algorithms are false positives. In recent years, bioinformatics approaches have improved by taking into consideration additional information, including the binding sites of Argonaute proteins and the secondary structure of the target site (Agarwal et al., 2015; Khorshid et al., 2013).

In addition to computational prediction programs, functional RNA interference (RNAi) assays have also been employed to identify miRNA targets in *Caenorhabditis elegans*. However, RNAi screens are only able to identify targets that are important for the phenotype of interest and may identify indirect targets. In *C. elegans*, the majority of single miRNA knockouts do not have an observable phenotype (Miska et al., 2007), as a consequence, the use of RNAi screens to detect targets regulated by miRNAs can be misleading. However, in other organisms, screens that inhibit miRNA targets may be more useful. For example, in *Drosophila melanogaster* >80% of miRNA mutants, 20% of which have mutations in multiple miRNAs, have phenotypes (Chen et al., 2014).

Other approaches to identify miRNA targets have focused on quantifying protein or RNA levels of candidate genes. Techniques applied to the identification of miRNA targets include stable isotope labeling by amino acids in cell culture (SILAC) (Baek et al., 2008; Selbach et al., 2008; Vinther et al., 2006) and ribosome profiling (Bazzini et al., 2012; Guo et al., 2010). These approaches can be biased by the selection of candidate targets and may reveal indirect targets (Hausser and Zavolan, 2014). Furthermore, the analysis of gene expression changes after altered miRNA levels does not identify the specific target site of the miRNA.

Recently, the identification of Argonaute binding sites through crosslinking immunoprecipitation (CLIP) based methods, such as CLIP-seq, PAR-CLIP, and iCLIP, has furthered the understanding of how miRNAs interact with their target sites

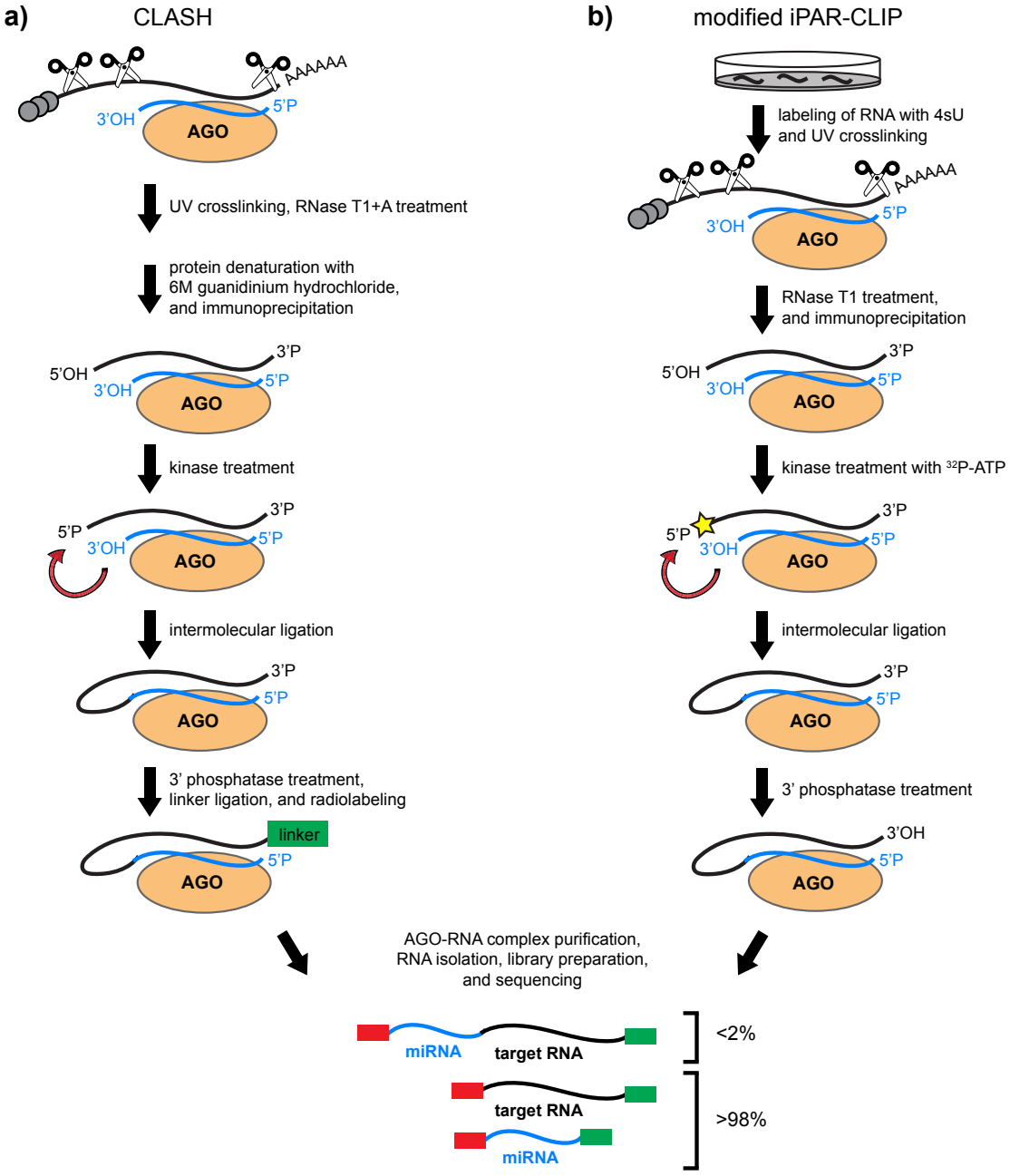
(Bosson et al., 2014; Chi et al., 2009; Hafner et al., 2010; Zisoulis et al., 2010). In general, CLIP-based methods identify protein-RNA binding sites by crosslinking proteins to interacting RNA molecules, purifying these protein-RNA complexes, and sequencing the associated RNAs. Although CLIP-based approaches define the region of an RNA that an Argonaute protein is bound to, these methods do not specifically identify the miRNA that is responsible for the identified interaction (Hausser and Zavolan, 2014). This is problematic for families of miRNAs that share the same seed sequence, for sites that contain seed complementarity to multiple miRNAs, or for sites with no obvious pairing to known miRNAs.

3.3 Review

3.3.1 Ligation of two RNA molecules identifies RNA-RNA interactions

Whereas CLIP-based methods are able to identify protein-RNA interaction sites, RNA-RNA interaction sites can be identified by crosslinking and sequencing of hybrids (CLASH) and similar approaches (Grosswendt et al., 2014; Helwak et al., 2013; Kudla et al., 2011). Akin to CLIP-seq, CLASH involves the purification and sequencing of cross-linked protein-RNA complexes. However, in CLASH additional biochemical steps promote the intermolecular ligation of RNA molecules to form a hybrid, or chimeric, read composed of two RNA molecules (Figure 3.1A).

Figure 3.1: Overview of CLASH and modified iPAR-CLIP methods. (a) CLASH and (b) modified iPAR-CLIP methods for the formation of miRNA-target chimeras. CLASH begins with trimming of unprotected RNAs in UV crosslinked lysates with RNase and denaturation of the AGO-miRNA-target RNA tertiary complex. In modified iPAR-CLIP, the sample (*C. elegans* worms, for example) must be incubated with 4-thiouridine (4sU) for RNA incorporation to enhance UV crosslinking. Both CLASH and modified iPAR-CLIP protocols phosphorylate the 5' end of the target RNA, which is then ligated to the miRNA using an exogenous RNA ligase. Subsequent 3' end phosphatase treatment prepares the RNA for linker ligation. In CLASH, the 3' linker is added during the "on-bead" biochemical steps, whereas in modified iPAR-CLIP, the 3' linker is added after RNA isolation. The majority of the reads generated from CLASH and modified iPAR-CLIP are not chimeric.


a)
CLASH
AAAAAA
3'OH
5'P
AGO
UV crosslinking, RNase T1+A treatment
protein denaturation with
6M guanidinium hydrochloride,
and immunoprecipitation
5'OH
3'P
kinase treatment
intermolecular ligation
3' phosphatase treatment,
linker ligation, and radiolabeling
linker
b)
modified iPAR-CLIP
labeling of RNA with 4sU
and UV crosslinking
RNase T1 treatment,
and immunoprecipitation
kinase treatment with 32P-ATP
3' phosphatase treatment
AGO-RNA complex purification,
RNA isolation, library preparation,
and sequencing
miRNA
target RNA
<2%
>98%

CLASH was developed after the observation that chimeric reads occurred in crosslinking and analysis of cDNAs (CRAC) data. These hybrid reads were not the product of reverse transcriptase template switching, and were likely generated as a result of the step in CRAC that ligates oligonucleotide linkers to RNA (Kudla et al., 2011). The first application of CLASH was the identification of snoRNA target sites on pre-rRNAs in yeast from C/D snoRNA associated proteins. From the sequencing library generated by CLASH for these proteins, 0.1% to 0.8% of reads were chimeric. The majority (74%) of snoRNA-pre-rRNA chimeric reads produced from this application of CLASH recovered known target sites. However, some reads identified potentially novel snoRNA-pre-rRNA sites. Other chimeric reads from Kudla et al produced rRNA-rRNA reads, which were thought to be nonspecific interactions.

The ability of CLASH to identify RNA-RNA interactions was subsequently applied to the identification of AGO1 miRNA target sites in human embryonic kidney 293 (HEK293) cells (Helwak et al., 2013). From the AGO1 CLASH data, 98% of the reads were not chimeric and contained sequence information similar to that produced by CLIP-seq. The remaining 2% of CLASH data contained chimeric reads and were composed of the mature miRNA sequence ligated to a target RNA molecule. In 69.8% of the miRNA chimeras the target RNA mapped to mRNAs. Additional RNAs found to be ligated to miRNAs included other miRNAs, rRNAs (ribosomal RNAs), tRNAs (transfer RNAs), pseudogenes, and lincRNAs (long intergenic non-coding RNAs). Through the inclusion of a control where yeast total RNA was mixed with the cell lysates before carrying out the CLASH protocol, Helwak et al demonstrated that less than 2% of CLASH chimeric reads were nonspecific. These non-mRNA targets identified by CLASH may therefore be examples of miRNA interactions with non-coding RNAs.

An alternative approach for the generation of chimeric reads has been devel-

oped through the inclusion of an intermolecular ligation step in iPAR-CLIP (Grosswendt et al., 2014). This modified version of iPAR-CLIP produced chimeric reads in *C. elegans* in a similar manner to CLASH (Figure 3.1B). From the sequencing data produced by modified iPAR-CLIP, 0.24% of reads were miRNA-target chimeras. As with CLASH, the chimeric reads appear to be highly specific with less than 2% of reads mapping to background bacterial sequences and 92% of chimeric reads mapping to mRNAs.

3.3.2 miRNA-target chimeras from standard CLIP-seq library preparation

Grosswendt et al. 2014 also found that chimeric reads were generated in iPAR-CLIP libraries that did not contain the additional intermolecular ligation step. This finding was surprising because standard iPAR-CLIP is not designed to produce the correct 5' and 3' end chemistry to allow for intermolecular ligations between miRNAs and target RNAs. However, the authors noticed that the chimeras produced by standard iPAR-CLIP tended to include a truncated miRNA sequence. They therefore concluded that the RNA trimming step in iPAR-CLIP was responsible for generating the ligated products. Specifically, RNase T1 was partially trimming the 3' end of the miRNA producing a 2'-3'-cyclic phosphate, which could then be ligated to the 5' hydroxyl of the target RNA through the action of endogenous ligases present in the lysate (Figure 3.2). The production of chimeras was less efficient in standard iPAR-CLIP than in the modified iPAR-CLIP (which included an exogenous ligase to catalyze intermolecular ligations), with only 0.16% of reads being miRNA-target chimeras. Using this information, Grosswendt et al. reanalyzed previously published CLIP-seq and PAR-CLIP data from human and mouse and found approximately 13,000 additional miRNA-target chimeras.

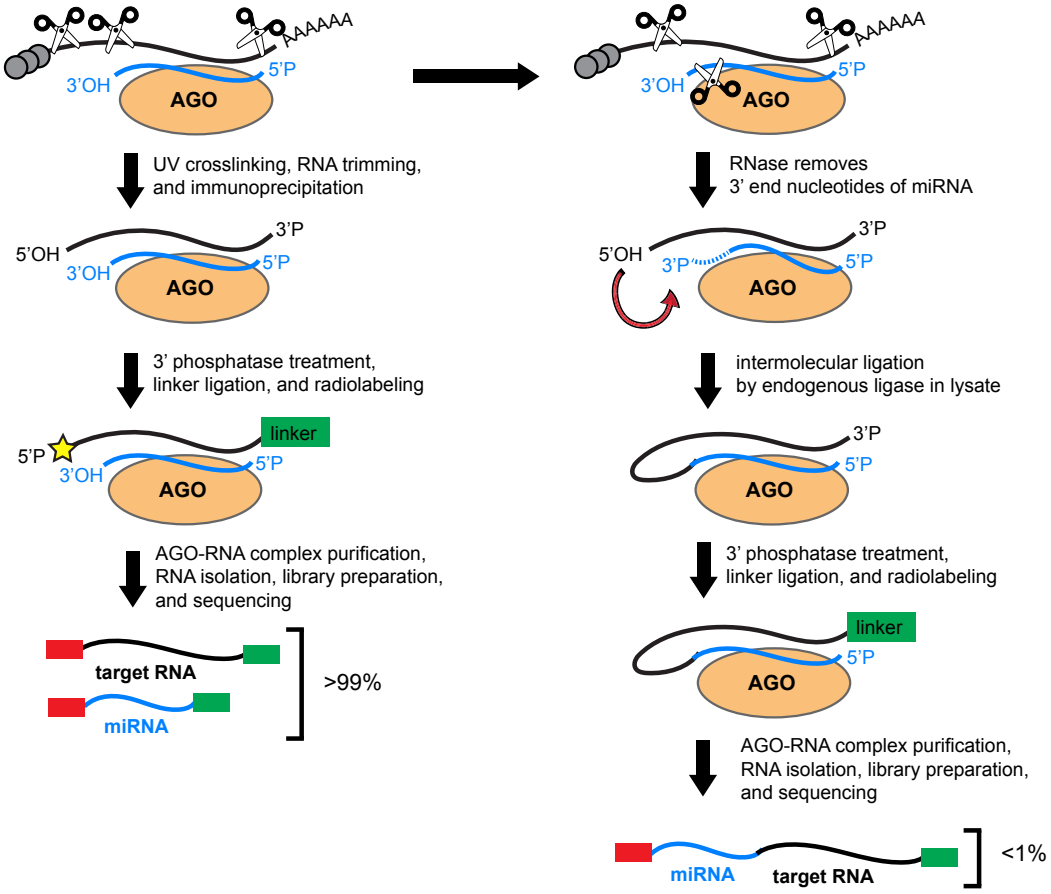


Figure 3.2: Comparison of the biochemical steps in CLIP-seq for the generation of standard CLIP-seq reads to events that can lead to the formation of miRNA-target chimeric reads in CLIP-seq or iPAR-CLIP. Standard CLIP-seq reads are generated after RNA trimming of UV crosslinked lysates and immunoprecipitation of the AGO-miRNA-target RNA tertiary complex. The 3' end of the RNA is then prepared for linker ligation and the complex is radio-labeled to facilitate the isolation of the complex. Chimeric reads may form in CLIP-seq when partial digestion of the 3' end of the miRNA by RNase during the RNA trimming step of CLIP-seq or iPAR-CLIP produces a 2'-3' cyclic phosphate or a 3' phosphate. Endogenous ligases in the lysate have been predicted to be responsible for ligation of the 3' end of the digested miRNA to the 5' phosphate of the target RNA. The subsequent steps that occur in the CLIP-seq protocol prepare the miRNA-target chimera for sequencing.

3.3.3 Bioinformatic identification of miRNA-target chimeric reads

Chimeric reads were identified similarly in CLASH and modified iPAR-CLIP. In both cases, duplicate reads and adapter sequences were removed before identifying chimeric reads. In CLASH, non-contiguous reads were identified using BLAST against transcriptome, tRNA, rRNA, and mature miRNA sequences (Helwak et al., 2013; Travis et al., 2014). Non-contiguous reads that contained a miRNA sequence were considered miRNA chimeras. Grosswendt et al searched reads from modified iPAR-CLIP for all possible 12 nucleotide sequences from mature miRNAs to identify putative miRNA-target chimeras (Grosswendt et al., 2014). The identity of the miRNA was then assigned by aligning the read to the full-length miRNA sequence. The method applied by Grosswendt et al. assured that truncated miRNAs or miRNA reads with mutations would also be recovered.

To ensure that the entire target site was identified, both Helwak et al. and Grosswendt et al. 2014 bioinformatically increased the size of the recovered target sequence. In CLASH the target sequence in the chimeric read was increased by 25 nucleotides. The reads from modified iPAR-CLIP were increased by 8 nucleotides upstream and 12 nucleotides downstream. These adjustments helped to increase the number of seed matches with the target RNA and facilitated clustering of overlapping target sequences to identify miRNA target sites (Grosswendt et al., 2014; Helwak et al., 2013).

3.3.4 Insights from miRNA-target chimeric reads

Although miRNAs are known to primarily direct Argonaute proteins to the 3'UTR of target mRNAs, many target sites identified by CLASH (42.6%) and modified

iPAR-CLIP (23.4%) are located in coding exons. Similarly, Argonaute binding sites have been identified in coding exons nearly as frequently as in 3'UTRs from CLIP-seq and PAR-CLIP datasets (Chi et al., 2009; Hafner et al., 2010; Zisoulis et al., 2010). Complementarity to miRNA seed sequences has been observed in coding exons, but the functionality of these potential target sites has remained unclear. In *C. elegans*, transcripts with coding exon Argonaute binding sites generally did not appear to be deregulated after the loss of Argonaute, whereas transcripts with 3'UTR binding sites were (Zisoulis et al., 2010). Similarly, transcripts with coding exon target sites of human Argonaute identified by PAR-CLIP in HEK293 cells were not as strongly regulated as target sites in 3'UTRs (Hafner et al., 2010). Some studies have shown that coding sequence targets function cooperatively with 3'UTR targets to enhance regulation (Fang and Rajewsky, 2011), whereas others have suggested that these target sites promote translational inhibition rather than mRNA stabilization (Hausser et al., 2013).

Since chimera-producing methods are able to identify both the miRNA and the target site, it is possible to classify the types of miRNA-target interactions that occur. Helwak et al. applied *k*-means clustering to identify five classes of miRNA-target interactions from 18,514 miRNA-mRNA chimeras. These classes included seed only, seed with supplementary (nucleotides 13-16) pairing, seed with terminal 3' end pairing, non-seed, and dispersed interactions. Targets with seed and seed with supplementary interactions were the most efficient at down-regulating targets and were the most conserved. Interestingly, 45% of miRNAs appeared to have nonrandom types of interactions with their targets with some miRNAs preferentially binding seed sites and other miRNAs having more extensive non-seed interactions. Overall, only 37% of miRNA-mRNA chimeras identified from the CLASH data contained perfect seed matches.

Grosswendt et al. 2014 also looked at the prevalence of seed interactions in modified iPAR-CLIP data and found that 43% of targets had perfect seed matches with their targets. However, when they included near-seed matches, such as 1 nucleotide mismatches and 1 nucleotide bulges, 80% of the *C. elegans* chimeras contained seed interactions. In contrast to the many non-seed interactions identified by CLASH, Grosswendt et al. observed limited evidence for 3' end interactions in their 3,627 chimeras from *C. elegans*. Similarly, the ~13,000 chimeras identified from traditional CLIP-seq and PAR-CLIP datasets also showed limited non-seed interactions.

3.3.5 miRNA-target chimeras identify non-canonical target sites

The CLASH-generated chimeras suggest that ~60% of the identified target sites were non-canonical with imperfect or non-seed interactions. To test whether non-canonical target sites for miR-92a were functional, Helwak et al. generated reporter constructs that contained miR-92a seed sites, miR-92a 3' end interaction motifs, and a combination of both seed and 3' end interaction motifs. For each of these three constructs, inhibition of miR-92a led to deregulation of the reporter. However, the construct containing just the miR-92a 3' end interaction motif was only moderately deregulated after miR-92a knockdown.

Recently, RNA expression data was independently analyzed for the regulation of miR-92a CLASH identified targets. In one dataset of miR-92a knockdown in HEK293 cells, both canonical and non-canonical miR-92a target genes were significantly deregulated (Agarwal et al., 2015). Although the non-canonical targets were deregulated, this effect was not particularly strong in comparison to the canonical targets. To further explore whether these non-canonical targets are func-

tional, expression data from the knockdown of 25 miRNAs, including miR-92a, was analyzed. In this data, the canonical miR-92a targets identified by CLASH were significantly deregulated, whereas the non-canonical miR-92a targets were not (Agarwal et al., 2015). In addition, Agarwal et al. 2015 examined the expression of non-canonical targets identified by CLASH for four miRNA families and observed that these non-canonical sites were not significantly deregulated, even if the site occurred within a 3'UTR. The slight regulation seen in the miR-92a non-canonical reporters conducted by Helwak et al., and the analysis of RNA expression in miRNA knockdowns conducted by Agarwal et al. suggest that these non-canonical target sites are either not nearly as functional as canonical seed-containing target sites or may not be functional at all.

In addition to non-canonical seed interactions, many of the chimeras identified by CLASH and modified iPAR-CLIP targeted ncRNAs, such as tRNAs, other miRNAs, and lincRNAs. Due to the low level of background ligation events with yeast RNA (CLASH) and bacterial RNA (modified iPAR-CLIP), it appears likely that many of these interactions are specific. Although the biological significance of most of these miRNA-ncRNA interactions remains to be determined, Helwak et al. demonstrated that the inhibition of a miRNA targeting a lincRNA resulted in the up-regulation of the lincRNA (Helwak et al., 2013). This indicates a functional role for some miRNA-ncRNA interactions. Competing endogenous RNAs (ceRNAs) have been proposed to sequester miRNAs from their targets (Hansen et al., 2013; Memczak et al., 2013). However, recent analysis of ceRNAs has suggested that at physiological levels many ceRNAs may not be expressed highly enough to effectively sequester miRNAs (Bosson et al., 2014). Helwak et al. propose that the prevalence of chimeras that map to rRNAs and tRNAs implies that these abundant RNAs may also have a role in sequestering miRNAs from their targets.

3.4 Conclusions

miRNA-target chimeric reads provide unambiguous determination of the identity of a miRNA that is bound at a target site, whereas previously it had to be assumed from seed complementarity or other features. In addition to correctly assigning miRNAs to their endogenous target sites, chimeras allow for detailed analysis of the types of interactions that miRNAs have with their targets. The extensive non-canonical interactions identified by CLASH may provide insights into how miRNAs choose their targets *in vivo*. While this article was in review, a new report on the analysis of miRNA-target chimeras concluded that pairing to miRNA 3' end sequences is more important than previously considered (Moore et al., 2015). In addition to patterns of hybridization with targets, analysis of chimeric sites may reveal features that explain why certain 3'UTRs are predominantly regulated by a single miRNA despite seed complementarity to other expressed miRNAs. Furthermore, chimeras allow the identification of ncRNA targets of miRNAs. These interactions with ncRNAs may be transient but still have biological importance.

Although miRNA-target chimeric reads are a unique tool to understand miRNA targeting, there are still several limitations to the current protocols. Foremost is the limited number of chimeric reads that are generated by the new methods. In CLASH and modified iPAR-CLIP, 2% and 0.24% of libraries were chimeric reads, respectively (Grosswendt et al., 2014; Helwak et al., 2013). As a consequence of the limited number of available reads, many target sites were identified by a single chimera. In the modified iPAR-CLIP data only 18.7% of target sites had more than one read (Grosswendt et al., 2014). Given this observation, it is unlikely that CLASH or modified iPAR-CLIP identify the complete set of miRNA-target interactions. Furthermore, it will be important to focus on reproducible chimeras, as isolated examples may represent sampling of targets by miRISC and not authentic targeting.

In line with these considerations, a comparison of CLASH chimeric reads to the most recent implementation of TargetScan led to the conclusion that TargetScan is better at predicting functional miRNA targets than the experimentally derived CLASH chimeras (Agarwal et al., 2015). Future work will need to focus on enriching for chimeras that represent functional targeting events to deepen our understanding of how miRISC chooses appropriate regulatory targets *in vivo*.

3.5 Authors' contributions

JPB drafted the manuscript and prepared the figures. All authors read and approved the final manuscript.

3.6 Acknowledgements

We thank members of the Pasquinelli lab for critical reading of the manuscript. Research in the Pasquinelli lab is supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1144086 (JPB) and from the NIH (GM071654) (AEP).

Chapter 3, in full, is a reprint of the material as it occurs in *Genetics Selection Evolution*, Broughton, J.P. and Pasquinelli A.E., BioMed Central, 2016. Broughton, J.P. was the primary author.

Chapter 4

Pairing beyond the seed supports microRNA targeting specificity

4.1 Abstract

MicroRNAs (miRNAs) direct Argonaute (AGO) proteins to target RNAs through imperfect base pairing. To identify endogenous miRNA target sites, we isolated AGO-bound RNAs from *Caenorhabditis elegans* by individual-nucleotide resolution crosslinking immunoprecipitation (iCLIP), which fortuitously also produced miRNA-target chimeric reads. Through the analysis of thousands of reproducible chimeras, pairing to the miRNA seed region (nucleotides 2-8) emerged as the predominant motif associated with functional interactions. Unexpectedly, we discovered that additional pairing to 3' sequences is prevalent in the majority of target sites and leads to specific targeting by members of miRNA families that share identical 5' sequences. By editing an endogenous target site, we demonstrate that 3' pairing determines targeting by specific miRNA family members and that seed pairing is not always sufficient for functional target interactions in the endogenous context.

Finally, we present a simplified method, Chimera PCR (ChimP), for the detection of specific miRNA-target interactions. Overall, our analysis of miRNA-target chimeras revealed that sequences in the 5' as well as the 3' regions of a miRNA provide the information necessary for stable and specific miRNA target interactions *in vivo*.

4.2 Introduction

MicroRNAs (miRNAs) are small RNA molecules that are bound by Argonaute (AGO) proteins. The AGO-miRNA duplex forms the core of the miRNA-induced silencing complex (miRISC), which is directed by the bound miRNA to complementary sequences in the messenger RNA (mRNA) (Pasquinelli, 2012). The miRISC co-factors then promote translational inhibition and transcript destabilization of the target RNA (Jonas and Izaurralde, 2015). Canonical miRNA-target interactions featuring complementarity to the seed sequence, nucleotides (nts) 2-8 of the miRNA, have long been recognized as critical for miRNA targeting (Bartel, 2009). Recent structural and single molecule studies have emphasized the importance of seed pairing for stable AGO binding (Chandradoss et al., 2015; Elkayam et al., 2012; Jo et al., 2015; Nakanishi et al., 2012; Salomon et al., 2015; Schirle and MacRae, 2012).

However, there are examples of functional miRNA target interactions that occur without perfect seed pairing. For example, the well-established miRNA target *lin-41* in *Caenorhabditis elegans* features two sites that are complementary to the let-7 miRNA (Slack et al., 2000; Vella et al., 2004). Neither site supports canonical seed pairing interactions; one of the sites forms a one-nucleotide bulge in the target and the other requires an unfavorable G-U pair. Imperfect seed matches have been suggested to be compensated by more extensive pairing with the 3' end of the miRNA

(Brennecke et al., 2005; Grimson et al., 2007). However, examples of conserved sites with 3' compensatory binding for weak seeds are relatively rare (Friedman et al., 2009). Although studies using 3' untranslated region (UTR) reporters have suggested that non-canonical seed sites or sites with 3' compensatory binding are functional (Helwak et al., 2013; Moore et al., 2015), a recent analysis of non-canonical target sites revealed that, even though these sites are bound by the miRNA complex, they do not appear to be broadly functional (Agarwal et al., 2015).

Many mature miRNAs can be classified based on the presence of identical seed sequences into groups called miRNA families (Lewis et al., 2003; Lim et al., 2003). Due to the shared seed sequence of miRNA family members, it is predicted that these miRNAs will regulate similar target RNAs. Phenotypic analyses of miRNA deletions in nematodes suggest that members of miRNA families have cooperative or redundant functions with each other (Alvarez-Saavedra and Horvitz, 2010). However, recent work suggests that individual family members may have independent targets, even when co-expressed (Moore et al., 2015).

Since miRNAs can regulate their targets by base pairing with as few as six nucleotides or through non-canonical interactions, the prediction of miRNA targets from sequence alone is difficult. Crosslinking immunoprecipitation and sequencing (CLIP-seq) and similar methods have been used to identify AGO binding sites on RNAs (Chi et al., 2009; Hafner et al., 2012; Zisoulis et al., 2010). However, CLIP-based approaches do not identify the miRNA that is responsible for a given interaction. Recently, methods have been developed to capture the miRNA associated with specific target sites bound by AGO (Grosswendt et al., 2014; Helwak et al., 2013; Moore et al., 2015). These methods (CLASH, modified iPAR-CLIP, CLEAR-CLIP) involve similar procedures to isolate AGO complexes, induce the ligation of miRNAs to nearby target RNA sequences, and then prepare sequencing libraries. The hybrid

reads produced by these methods are known as miRNA-target chimeras.

Here we report that individual-nucleotide resolution crosslinking immunoprecipitation (iCLIP) of endogenous AGO in *C. elegans* produces miRNA-target chimeric reads at similar efficiencies as methods designed to yield chimeras. Our analysis of thousands of reproducible miRNA-target chimeric reads unambiguously reveals the identity of the miRNA at AGO mediated target sites and points to features that promote target mRNA regulation in the endogenous context. We demonstrate the importance of interactions beyond seed pairing in specifying miRNA target sites using an endogenous *in vivo* reporter. Furthermore, we present a new method for the identification of miRNA-target chimeras that does not require the use of radioactivity or the analysis of sequencing datasets.

4.3 Results

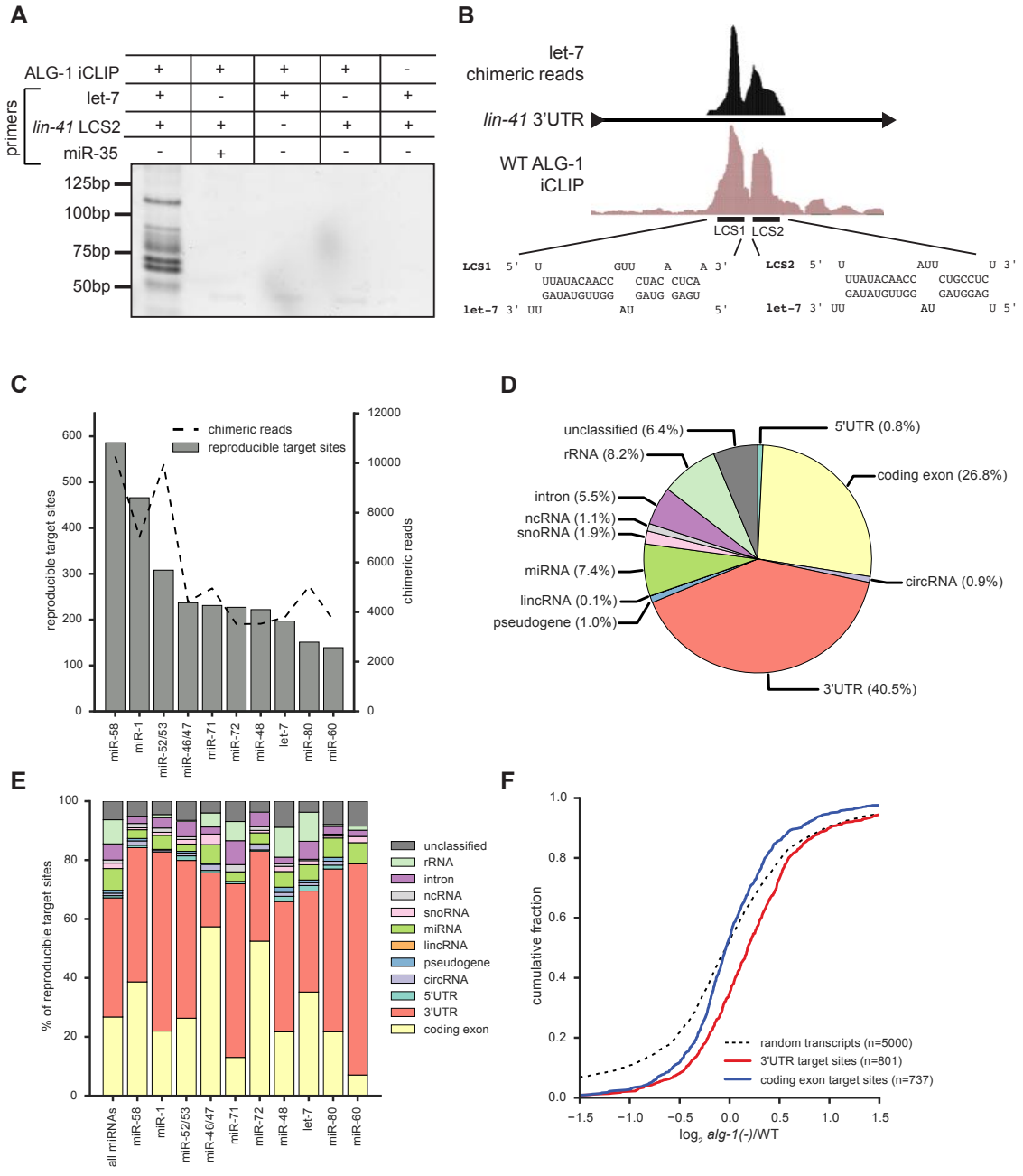
4.3.1 ALG-1 iCLIP generates miRNA-target chimeras

In *C. elegans,* the Argonaute-Like Gene 1 (ALG-1) protein is essential for normal miRNA expression and function. To generate a more complete map of ALG-1 target sites, we carried out ALG-1 iCLIP in wild-type (WT) *C. elegans* animals at the last larval stage of development, as iCLIP has been shown to recover more unique cDNAs than traditional CLIP-seq (Sugimoto et al., 2012). We analyzed ALG-1 binding sites using the CLIPper peak-finding algorithm (Lovci et al., 2013) and identified 5,006 ALG-1 binding sites that were reproducible in at least two biological replicates (Table S1). 79.9% of these binding sites occurred in 3'UTRs (Figure 4.6A).

Chimera-generating methods have provided unambiguous miRNA targeting data in a variety of organisms and cell types (Grosswendt et al., 2014; Helwak et al.,

2013; Moore et al., 2015). Interestingly, these miRNA-target chimeric reads have been reported to also occur in CLIP-seq and PAR-CLIP libraries, even without the addition of the biochemical steps intended to increase their frequency (Grosswendt et al., 2014). We tested for the presence of miRNA-target chimeras in our ALG-1 iCLIP libraries by PCR using primers for mature let-7 and the second let-7 complementary site (LCS2) in the 3'UTR of *lin-41*. PCR products were detected only for this well-established target interaction, and not for another miRNA and the same target site, or when using single primers (Figure 4.1A). This result shows that chimeric reads for authentic miRNA target sites occur in ALG-1 iCLIP and that these reads occur at a high enough frequency for PCR to detect.

Figure 4.1: ALG-1 iCLIP produces miRNA-target chimeric reads. (A) The presence of miRNA-target chimeras in ALG-1 iCLIP libraries were tested by PCR using the indicated primers. (B) let-7 chimeric reads map to LCS1 and LCS2 in the *lin-41* 3'UTR. (C) The number of target sites and chimeric reads detected for the ten miRNAs with most target sites. (D) Distribution of target sites among transcript types. (E) Genic locations of target sites for the indicated miRNAs. (F) Up-regulation in *alg-1(-)* of transcripts with target sites only in 3'UTR (red) (Mann-Whitney U test, $P<2x10-19$) but not in coding exons (blue) (Mann-Whitney U test, $P<0.35$) in comparison to randomly selected transcripts.


A
ALG-1 iCLIP
primers
let-7
lin-41 LCS2
miR-35
125bp
100bp
75bp
50bp
B
let-7
chimeric reads
lin-41 3'UTR
WT ALG-1
iCLIP
LCS1
LCS2
C
chimeric reads
reproducible target sites
reproducible target sites
chimeric reads
miR-58
miR-1
miR-52/53
miR-46/47
miR-71
miR-72
miR-48
let-7
miR-80
miR-60
D
unclassified (6.4%)
rRNA (8.2%)
intron (5.5%)
ncRNA (1.1%)
snoRNA (1.9%)
miRNA (7.4%)
lincRNA (0.1%)
pseudogene (1.0%)
5'UTR (0.8%)
coding exon (26.8%)
circRNA (0.9%)
3'UTR (40.5%)
E
% of reproducible target sites
all miRNAs
unclassified
rRNA
intron
ncRNA
snoRNA
miRNA
lincRNA
pseudogene
circRNA
5'UTR
3'UTR
coding exon
F
cumulative fraction
random transcripts (n=5000)
3'UTR target sites (n=801)
coding exon target sites (n=737)
log2 alg-1(-)/WT

To identify all of the miRNA-target chimeric reads in our ALG-1 iCLIP data, we developed a computational pipeline for their detection and mapping. We anticipated that low numbers of chimeric reads (<5000) might be recovered, as has been described for CLIP libraries that lack the intermolecular ligation step (Grosswendt et al., 2014; Moore et al., 2015). Remarkably, our analysis revealed 153,684 non-redundant chimeric reads that mapped to the *C. elegans* genome at 46,910 sites for 112 guide and 47 passenger strand miRNAs. Sites with at least two overlapping reads represented 20.6% of total sites. Non-redundant chimeric reads ranged from 1.3% to 5.1% of all reads from five independent ALG-1 iCLIP libraries. For comparison, ~2% of CLASH, ~0.2% of modified iPAR-CLIP, and ~1.5-5% of CLEAR-CLIP libraries consisted of chimeric reads (Grosswendt et al., 2014; Helwak et al., 2013; Moore et al., 2015). Initial analysis of the chimeric reads from ALG-1 iCLIP revealed that they map to known miRNA target sites. For example, let-7 chimeric reads map specifically to the two let-7 complement sites in the *lin-41* 3'UTR (Figure 4.1B). To assess the frequency of non-specific chimera formation in ALG-1 iCLIP, we mapped our chimeric reads to the *E. coli* genome. *E. coli* bacteria are the food source for *C. elegans,* and reads that map to the *E. coli* genome are commonly recovered in CLIP-based experiments (Grosswendt et al., 2014). <7% of total non-redundant miRNA-containing reads were ligated to *E. coli* sequences, indicating that nonspecific ligation events were rare. These analyses show that ALG-1 iCLIP produces miRNA-target chimeras with similar efficiency to methods specifically designed to generate chimeric reads. Moreover, these chimeric reads correctly match specific miRNAs to previously characterized miRNA target sites.

To explore the mechanism by which chimeric reads might have been generated, we examined the prevalence of full-length and 3' truncated miRNAs in our chimeric read data. Previously, it was suggested that trimming of the miRNA 3' end

by RNase treatment may allow endogenous enzymes present in the lysate to ligate the miRNA to proximal target RNA sequence (Grosswendt et al., 2014). However, the majority of miRNA-target chimeras produced by ALG-1 iCLIP were composed of full-length miRNAs. The inclusion of truncated miRNAs increased chimera identification by ~20% (Figure 4.6B), whereas >90% of the chimeras were formed by shortened miRNAs in unmodified iPAR-CLIP. Since the majority of miRNA-target chimeric reads were composed of intact miRNAs, it is likely that most iCLIP chimeras form during the biochemistry used to produce chemical moieties compatible with linker ligation steps. However, since the inclusion of 3' truncated miRNAs increased the identification of miRNA-target chimeras, it remains possible that the action of an endogenous ligase is responsible for a subset of chimeric reads. During our analysis of ALG-1 iCLIP chimeric reads, we noticed that many reads contained an untemplated nucleotide on the 5' end of the miRNA. The inclusion of the untemplated nucleotide when searching for chimeric reads increased read identification by ~30% (Figure 4.6B). This nucleotide is primarily an adenosine or thymidine and is likely added during reverse transcription (Figure 4.6C). For our computational identification of miRNA-target chimeric reads, we included both 3' truncated and 5' untemplated nucleotide addition miRNA variations.

For all subsequent analyses we considered only the 4,920 chimera-producing sites that were reproducible in at least two biological replicates (Table S2), hereafter referred to as target sites. The miRNAs with the greatest number of target sites tended to be those that were highly expressed, as determined by the number of chimeric reads (Figure 4.1C). In addition, of the 20 miRNAs with the greatest number of target sites, 80% were identified as the top 20 highest expressed miRNAs at mid L4 (Kato et al., 2009). Similarly, 83% of the guide strand miRNAs with reproducible target sites were previously shown to be associated with ALG-1 (Zisoulis

et al., 2010) (Figure 4.6D). These experiments show that ALG-1 iCLIP generates reproducible miRNA-target chimeric reads that reveal the miRNA targeting landscape in *C. elegans.*

4.3.2 Targets identified by chimeras are misregulated in *alg-1(-)* animals

AGO proteins are generally guided by miRNAs to the 3'UTR of mRNAs (Pasquinelli, 2012). The majority of target sites identified in this study occurred in mRNAs and 40.5% of all target sites mapped to 3'UTRs (Figure 4.1D). Of these 3'UTR target sites, 87% overlapped an ALG-1 binding site in at least one ALG-1 iCLIP library (Figure 4.6E). We found relatively few target sites in introns (5.5%) compared to miRNA-target chimeras from human CLASH (15%) and mouse CLEAR CLIP (36%) (Moore et al., 2015). However, both CLASH and CLEAR-CLIP considered clustered sites, whereas we considered only reproducible target sites. When we analyzed all chimeric reads (including non-reproducible) found in our ALG-1 iCLIP libraries, 16.4% of the reads mapped to introns (Figure 4.6F). The fewer miRNA intronic target sites observed when considering only reproducible sites suggests that these interactions are either unstable or occur infrequently.

Chimeras also formed with other mature miRNAs and snoRNAs, as has been previously reported (Grosswendt et al., 2014; Helwak et al., 2013; Moore et al., 2015). In addition, we observed target sites that mapped to published circRNAs (Ivanov et al., 2015; Memczak et al., 2013). Within mRNA sequences, individual miRNAs exhibited distinct binding patterns with some miRNAs having primarily 3'UTR (e.g., miR-71 and miR-60) or coding exon target sites (e.g., miR-46/47 and miR-72) (Figure 4.1E and Figure 4.6G).

In most cases, miRISC promotes transcript destabilization of bound targets

(Eichhorn et al., 2014). Hence, loss of AGO proteins or deletion of specific miRNAs results in increased target mRNA abundance. To examine if the mRNAs containing target sites identified by ALG-1 iCLIP are up-regulated in *alg-1* mutants, we performed RNA seq on WT and *alg-1(gk214)* animals, referred to hereafter as *alg-1(-)*. Compared to randomly selected genes, the mean change in expression was higher for genes with target sites in 3'UTRs but not coding exons (Figure 4.1F). Similarly in *alg-1(-)*, only transcripts with ALG-1 binding sites in 3'UTRs were up-regulated (Figure 4.6H). This observation is consistent with previous microarray analysis of *alg-1(-)* animals that showed up-regulation primarily of transcripts with ALG-1 binding in 3'UTRs (Zisoulis et al., 2010).

4.3.3 Seed pairing is enriched in miRNA target sites

The miRNA sequence can be separated into five functional domains that affect miRNA target recognition: 5' anchor (nt 1), seed sequence (nts 2-8), central region (nts 9-12), 3' supplementary region (nts 13-16), and 3' tail (nts 17-22) (Wee et al., 2012). We anticipated that complementarity to the seed sequence of the cognate miRNA would be a prominent feature in our target sites. However, it was also possible that the target sites identified by ALG-1 iCLIP would share a common sequence motif unrelated to the identity of the cognate miRNA. Using MEME motif analysis (Bailey and Elkan, 1994), the primary motif we identified was the seed complement for the cognate miRNA that was ligated to target sites of highly expressed miRNAs (Figure 4.2A). This demonstrates that the chimeras produced by ALG-1 iCLIP are dependent on the identity of the ligated miRNA. Complementarity to the 3' supplementary region of the miRNA, in addition to the seed sequence, has been suggested to moderately enhance miRNA targeting (Grimson et al., 2007). To assess the prevalence of seed and 3' supplementary pairing, we looked for

enrichment of the complementary nucleotides to these miRNA domains in and around the target sites. For both seed and 3' supplementary pairing, the presence of the complement to these sites is enriched at the target site compared to surrounding sequences (Figure 4.2B).

To examine globally how miRNAs interact with their target sites in mRNAs, we paired each miRNA with its target site using RNAhybrid (Rehmsmeier et al., 2004) and grouped similar interactions by *k*-means clustering on the predicted miRNA-target duplexes. When k=7, six groups feature seed interactions, six feature 3' interactions, whereas central and tail interactions are present in three groups (Figure 4.2C). Unlike chimeras from human miRNAs identified by CLASH, we did not detect a class of distributed interactions (Helwak et al., 2013). Our classes exhibit distinct pairing among the different miRNA functional domains. For example, Class 1 features seed-only pairing, whereas Class 3 exhibited seed pairing with 3' supplemental pairing but no central pairing. Similar to Class 3, Class 4 interactions displayed seed pairing with no central pairing but with more frequent 3' tail interactions. Class 7 interactions did not exhibit seed pairing interactions, but instead contained complementarity to interactions throughout the other miRNA domains.

Categorization of the target sites found in coding exons and 3'UTRs did not reveal enrichment for a specific class in either genic region (Figure 4.2D). However, individual miRNAs exhibited enrichment for specific classes of interactions (Figures 4.2E and 4.6I). We detected differential enrichment for each of the seven classes among the miRNAs that produce the greatest number of target sites, including differences between members of the same miRNA family, such as let-7 and miR-48. Well-established miRNA targets sites, such as those in *lin-14*, *lin-28*, *lin-41*, and *hbl-1*, generally featured Class 3 and Class 4 interactions (Figure 4.2F).

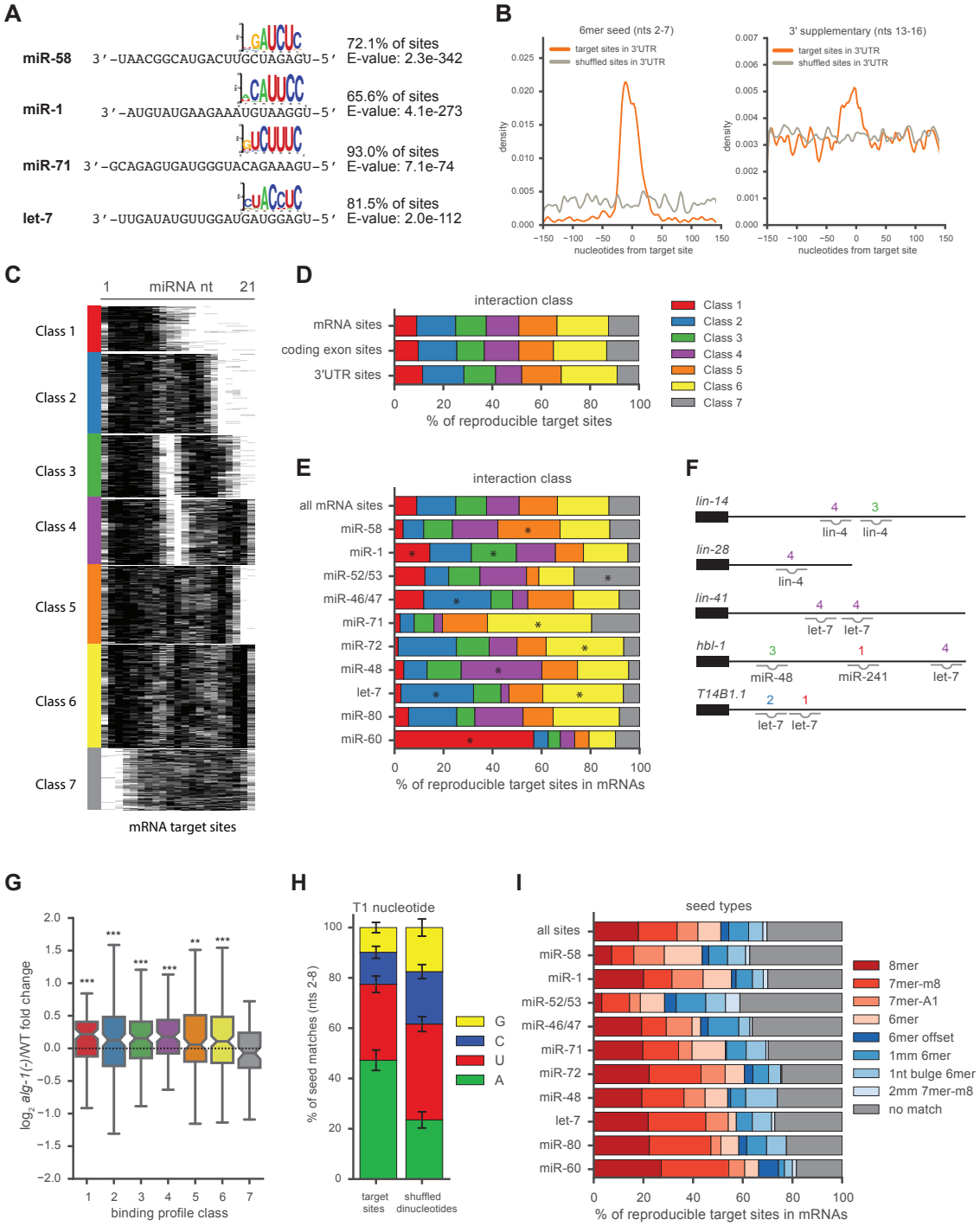
To examine if there might be functional differences among the classes of

pairing interactions, we examined the expression of transcripts with each class of interaction in their 3'UTRs in *alg-1(-)* animals compared to WT (Figure 4.2G). Among the seven classes, only the seedless Class 7 was not significantly up-regulated in comparison to randomly selected transcripts. Although it is possible that the Class 7 target sites primarily direct translational repression, it is striking that seed pairing seems to be broadly important for the regulation of target mRNA stability.

Another feature that has been associated with functional miRNA target sites is the presence of an adenosine immediately 3' of the seed complement in the target RNA, known as T1A (Lewis et al., 2005). Human AGO contains a binding pocket that recognizes T1A, which likely functions to anchor the AGO protein at the target site (Schirle et al., 2015, 2014). Adenosine at the T1 position is over-represented in our chimera derived target sites compared to sites with shuffled dinucleotides (Figure 4.2H). This analysis lends genome-wide support for previous computational and structural work pointing to the importance of the identity of the nucleotide after seed pairing in the target sequence.

Although all but one class of miRNA-target interactions exhibited general seed pairing, many sites within these classes appeared to have imperfect seed matches. We examined the complementarity of target sites to their cognate miRNAs for various classes of seed matching. Among all mRNA target sites, ~50% of interactions included 6mer (nts 2 7), 7mer m8 (nts 2 8), 7mer A1 (nts 2 7 with T1A), or a 8mer (nts 2 8 with T1A) seed interaction with the cognate miRNA, whereas ~20% of sites included a 6mer offset (nts 3-8), mismatch, or bulge seed interaction and ~30% of sites featured no match to an established seed type (Figure 4.2I). Taken together, these findings support the importance of seed pairing in miRNA function and reveal that the majority of miRNA target sites support additional 3' end complementary interactions.

Figure 4.2: miRNA-target chimeric reads are enriched for seed pairing. (A) The seed complement is the primary motif identified by MEME of targets for the indicated miRNAs. (B) Density plot showing enrichment of 6mer seed and 3' supplementary complementarity to cognate miRNAs at 3'UTR target sites. (C) miRNA-target duplex structure predictions calculated by RNAhybrid and partitioned into seven classes by *k*-means clustering. Black pixels represent pairing. (D) Distribution of classes among all mRNA, coding exon or 3'UTR target sites. (E) Distribution of classes for the indicated miRNAs. Significantly enriched classes (one-sided Fisher's exact test, $P<0.001$) are indicated (*). (F) Class interactions of established miRNA target sites. For sites with multiple miRNAs bound, the miRNA with the greatest number of chimeric reads at that site was chosen. (G) Up-regulation of transcripts in emphalg-1(-) for each interaction class (Mann-Whitney U test, *** $P<0.001$, ** $P<0.005$). (H) Adenosine at the T1 position is enriched after the seed complement for the ten miRNAs with the greatest number of target sites compared to sites with shuffled dinucleotides (Fisher's exact test, $P<0.0001$). Error bars represent $\pm$ SEM. (I) Distribution of seed complements for all mRNA target sites and the indicated miRNAs.


A
miR-58 3′-UAACGGCAUGACUUGCUAGAGU-5′ 72.1% of sites E-value: 2.3e-342
miR-1 3′-AUGUAUGAAGAAAUGUAAGGU-5′ 65.6% of sites E-value: 4.1e-273
miR-71 3′-GCAGAGUGAUGGGUACAGAAAGU-5′ 93.0% of sites E-value: 7.1e-74
let-7 3′-UUGAUAUGUUGGAUGAUGGAGU-5′ 81.5% of sites E-value: 2.0e-112
B
6mer seed (nts 2-7)
3' supplementary (nts 13-16)
target sites in 3'UTR
shuffled sites in 3'UTR
density
nucleotides from target site
C
1 miRNA nt 21
Class 1
Class 2
Class 3
Class 4
Class 5
Class 6
Class 7
mRNA target sites
D
interaction class
mRNA sites
coding exon sites
3'UTR sites
% of reproducible target sites
E
interaction class
all mRNA sites
miR-58
miR-1
miR-52/53
miR-46/47
miR-71
miR-72
miR-48
let-7
miR-80
miR-60
% of reproducible target sites in mRNAs
F
lin-14
lin-28
lin-41
hbl-1
T14B1.1
lin-4
let-7
miR-48
miR-241
G
log2 alg-1(-)/WT fold change
binding profile class
H
T1 nucleotide
% of seed matches (nts 2-8)
target sites
shuffled dinucleotides
G
C
U
A
I
seed types
all sites
% of reproducible target sites in mRNAs
8mer
7mer-m8
7mer-A1
6mer
6mer offset
1mm 6mer
1nt bulge 6mer
2mm 7mer-m8
no match

4.3.4 miRNA family members bind specific sets of target sites

Considering that seed pairing has been proposed to be not only necessary but also sufficient for miRNA targeting (Doench and Sharp, 2004; Enright et al., 2003; Lewis et al., 2005, 2003; Stark et al., 2003) and that miRNA families in *C. elegans* may act redundantly (Alvarez-Saavedra and Horvitz, 2010), we predicted that miRNA family members would bind largely overlapping sets of targets. To investigate this possibility, we examined target sites for the let-7 family of miRNAs. Three of the let-7 family of miRNAs, let-7, miR-48, and miR-241, are expressed during the last larval stage in overlapping sets of tissues (Abbott et al., 2005; Kato et al., 2009; Martinez et al., 2008). Additionally, miR-48 and miR-241 are processed from the same primary transcript. Whereas the first eight nucleotides of let-7, miR-48, and miR-241 are identical, the rest of their sequences diverge (Figure 4.3A). Surprisingly, the chimeras formed by let-7, miR-48, and miR-241 revealed that the majority of their target sites were non-overlapping (Figure 4.3B).

Target sites that were shared by multiple let-7 family members included the established let-7 family targets, *daf-12* and *hbl-1* (Abbott et al., 2005; Grosshans et al., 2005) (Figure 4.7A and B). Along the 3'UTRs of both *daf-12* and *hbl-1*, multiple let-7 family members share some target sites, but other target sites are specific or highly biased for binding to a single family member. The general binding preferences of these sites agree with early observations that some let-7 family members are predicted to pair more favorably with specific complementary sites in the *hbl-1* 3'UTR (Lin et al., 2003).

In general, transcripts bound by a single let-7 family member tended to be regulated by the miRNA pathway. We found that transcripts targeted by an individual let-7 family miRNA in their 3'UTR were significantly up-regulated in *alg-1(-)* animals (Figure 4.3C). Although the majority of let-7 target sites did not produce chimeras with

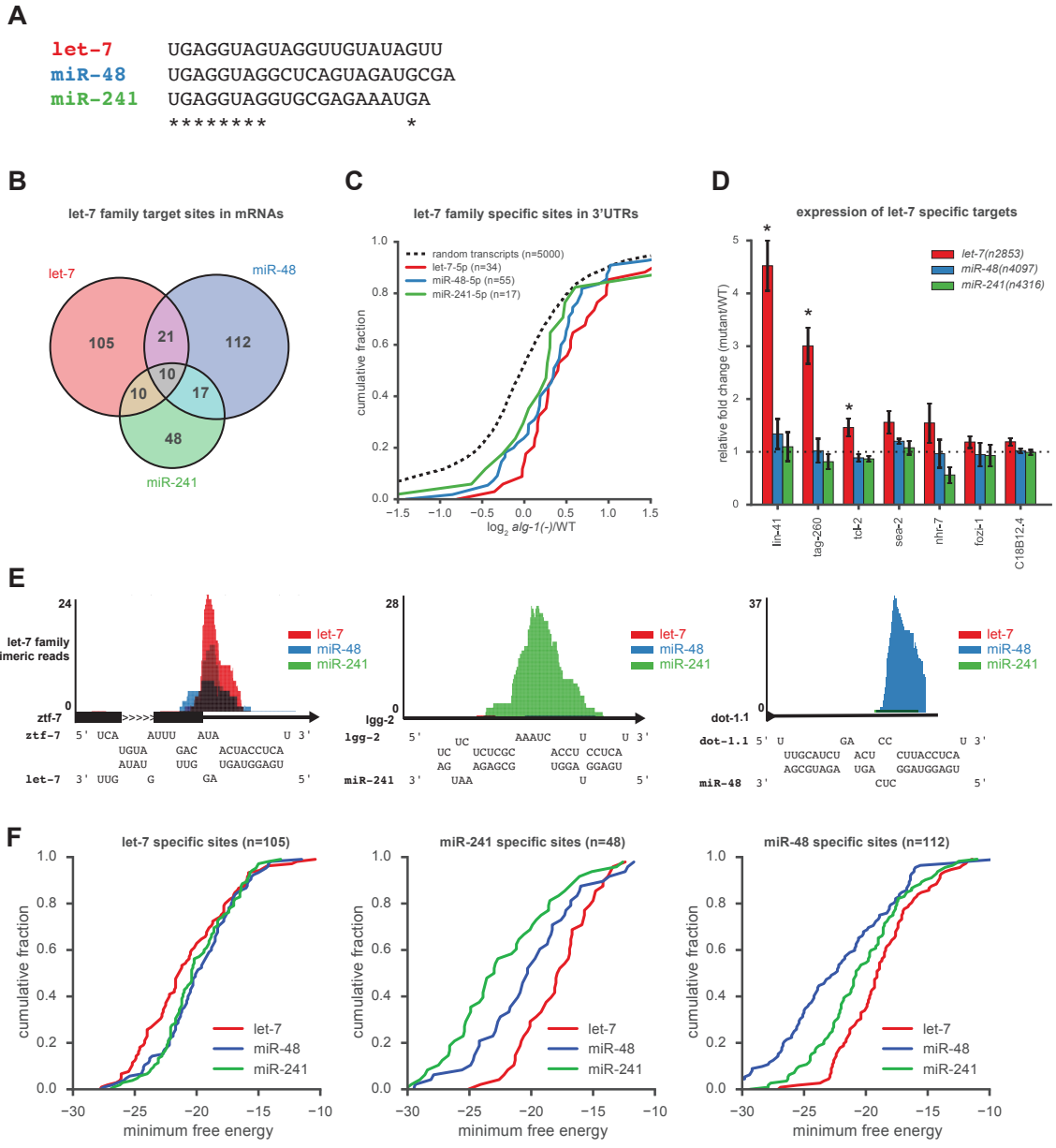
other let-7 family members (Figure 4.3B), most of the transcripts containing these let-7 specific sites produced chimeras or strong peaks representing ALG-1 binding at additional locations. While these observations suggest combinatorial regulation, we were still able to detect specific misregulation of some of these targets in animals deficient for let-7 activity; whereas these same targets were not up-regulated in *miR-48* or *miR-241* null strains (Figure 4.3D). We were unable to detect targets for miR-48 and miR-241 that appeared to be specifically misregulated, likely because both miR-48 and miR-241 are significantly up-regulated in *miR-241* and *miR-48* deletion strains, respectively (Figure 4.8).

Some let-7 target sites, such as the established site in *ztf-7* (Jovanovic et al., 2010), are shared by multiple let-7 family members, whereas others produced chimeric reads almost exclusively with a specific member (Figure 4.3E). Since nucleotides 1-8 of let-7, miR-48, and miR-241 are identical, other sequences in these miRNAs must dictate specific target interactions. With the limited number of specific sites for each let-7 family miRNA, we did not detect a miRNA region common to all let-7 family members that would be responsible for exclusive interactions. Instead, the overall binding energy is more favorable for each miRNA and its cognate sites than for other family members paired to those sites. We hybridized each let-7 family member to specific target sites using RNAhybrid to determine the minimum free energy (MFE) of the miRNA target duplex. In general, pairing was most favorable for the let-7 family member with its experimentally defined set of specific target sites (Figure 4.3F).

Specific binding by miRNA family members appears to be common for other miRNAs. The miR-58 and miR-238 families of miRNAs are expressed at mid-L4 (Kato et al., 2009) and also have divergent 3' ends (Figure 4.9A and D). For both of these families, we found primarily specific target sites with similar patterns of

favorable pairing as those observed for the let-7 family (Figure 4.9). We found that shared sites were more likely to contain perfect seed matches than specific sites (Figure 4.10A), but a strong bias for T1A in shared sites versus specific sites was not detected (Figure 4.10B). These analyses reveal that miRNA family members can exhibit divergent target interactions and that these preferences likely arise from 3' end pairing of the miRNA to its target site.

Figure 4.3: The let-7 family of miRNAs binds divergent sets of target sites. (A) Mature sequences of the three most abundant let-7 family miRNAs. (B) Overlap of target sites for let-7 family miRNAs. (C) Transcripts specifically bound by single let-7 family miRNAs in 3'UTRs are up-regulated in emphalg-1(-) compared to random transcripts (Mann-Whitney U test: let-7 targets P<1.3x10- 5, miR-48 targets P<4.4x10-5, miR-241 targets P<0.18). (D) qRT-PCR of let-7 specific target candidates in the indicated strains (* P<0.05). Error bars represent $\pm$ SEM. (E) Examples of shared and specific let-7 family targets with predicted pairing interactions. (F) RNAhybrid analysis of minimum free energy (MFE) of miRNA-target duplex for each let-7 family member to its specific mRNA target sites.


A
let-7 UGAGGUAGUAGGUUGUAUAGUU
miR-48 UGAGGUAGGCUCAGUAGAUGCGA
miR-241 UGAGGUAGGUGCGAGAAAUGA
********* *
B
let-7 family target sites in mRNAs
let-7
miR-48
miR-241
105
21
112
10
10
17
48
C
let-7 family specific sites in 3'UTRs
random transcripts (n=5000)
let-7-5p (n=34)
miR-48-5p (n=55)
miR-241-5p (n=17)
cumulative fraction
log2 alg-1(-)/WT
D
expression of let-7 specific targets
let-7(n2853)
miR-48(n4097)
miR-241(n4316)
relative fold change (mutant/WT)
lin-41
tag-260
tcl-2
sea-2
nhl-7
fozi-1
C18B12.4
E
let-7 family chimeric reads
ztf-7
lgg-2
dot-1.1
let-7
miR-48
miR-241
F
let-7 specific sites (n=105)
miR-241 specific sites (n=48)
miR-48 specific sites (n=112)
cumulative fraction
minimum free energy

4.3.5 miRNA 3' end pairing directs specific target interactions

Our detection of chimeras specific for let-7 in the *lin-41* 3'UTR is consistent with the requirement for let-7-mediated regulation of *lin-41.* Loss of *let-7* results in a lethal phenotype where animals burst through their vulvas (Reinhart et al., 2000; Slack et al., 2000), and this phenotype can be suppressed by restoring regulation of just *lin-41* (Ecsedi et al., 2015). Additionally, versions of let-7 that contain mutations in 3' end sequences do not fully rescue bursting of *let-7* mutants, pointing to the importance of the 3' end of this miRNA (Zhang et al., 2015). To further test if regulation of *lin-41* is entirely dependent on let-7, we analyzed the expression of *lin-41* in *let-7(n2853)* animals and in null mutants of *miR-48* and *miR-241*, at mid-L4 stage. Compared to WT, the levels of *lin-41* were misregulated only in *let-7(n2853)* (Figure 4.4A).

The imperfect seed pairing of LCS1 and LCS2 in *lin-41* likely necessitates strong miRNA 3' end interactions, which are much more favorable for let-7 than for miR-48 or miR-241. However, our chimera data indicated that some target sites are capable of perfect seed pairing (up to 8mer) to any of the let-7 family members yet appear to be exclusively bound by a single member (Figure 4.10A). To test the functional importance of specific miRNA targeting *in vivo*, we decided to replace LCS1 and LCS2 in the *lin-41* 3'UTR (Figure 4.4B) with a target site that was specific for another let-7 family member. We reasoned that the altered specificity of this site might switch the identity of the miRNA required to regulate *lin-41*. To minimally perturb the endogenous context, we used CRISPR/Cas9 genome editing and homologous recombination methods to introduce two copies of the miR-48 specific target site from the 3'UTR of *dot-1.1* (Figure 4.3E) at the same positions as LCS1 and LCS2 in the *lin-41* 3'UTR. The miR-48 site copied from the 3'UTR of *dot 1.1* has 7mer-A1 seed pairing with a GU-wobble at the nucleotide 8 position and

more extensive 3' end complementarity to miR-48 than to let-7 or miR-241 (Figure 4.4C). Importantly, all let-7 family members are expected to be capable of interacting with the *dot-1.1* site through seed mediated interactions. Thus, if seed pairing is sufficient for regulation, this version of *lin-41* should no longer be dependent on any one let-7 family miRNA. Animals harboring the mutant *lin-41(ap427)* allele*,* which has the miR-48 specific sites in the *lin-41* 3'UTR, display no observable phenotypes, suggesting that *lin-41* is sufficiently regulated in these animals.

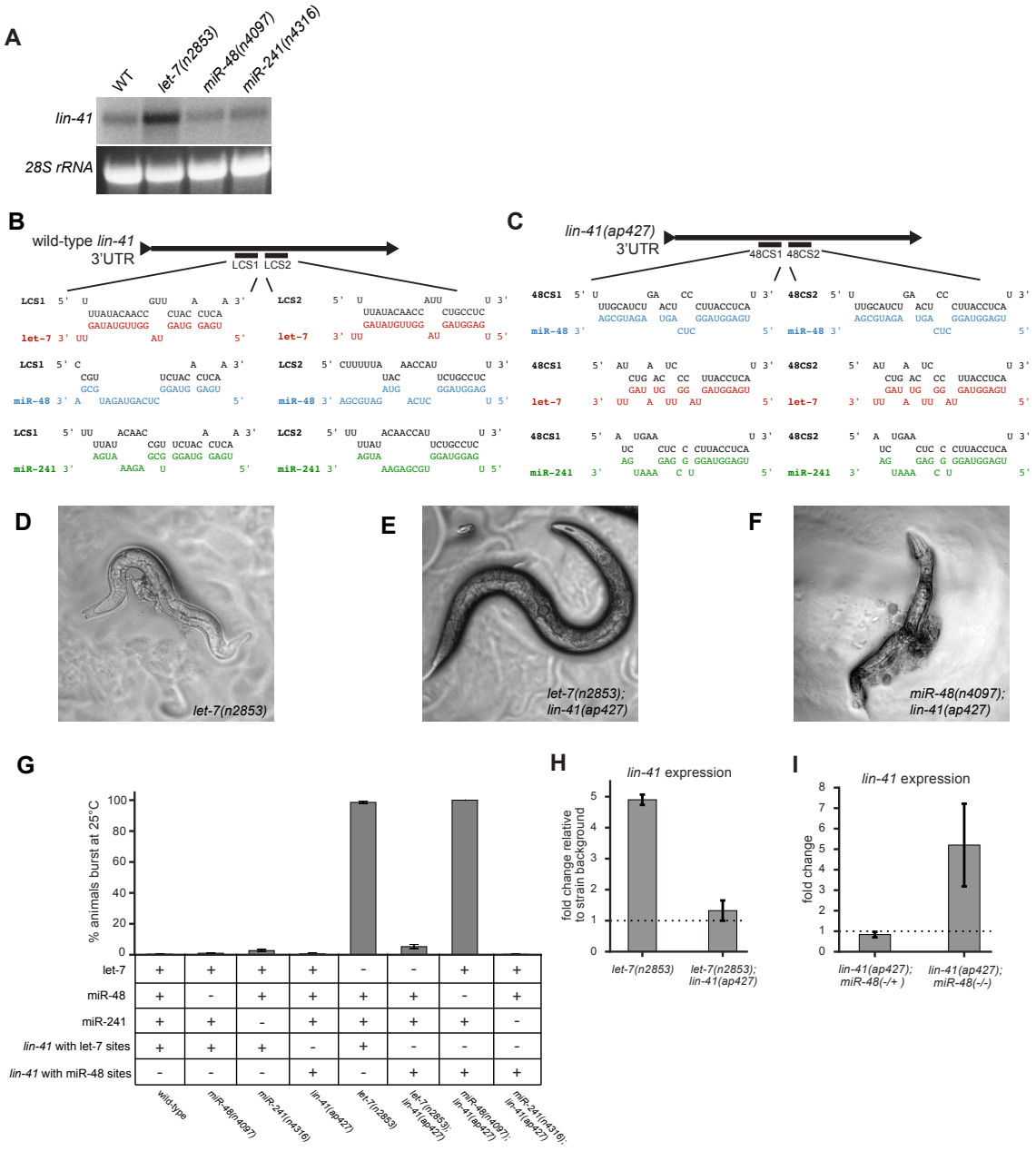
The *let-7(n2853)* mutation results in a single nucleotide change to the seed sequence of mature let-7 and decreased levels of mature let-7 (Figure 4.8C and D). At 25°C, *let-7(n2853)* animals burst through the vulva and die due to the specific misregulation of *lin-41* (Figure 4.4A and D). Remarkably, when we combined the *let-7(n2853)* mutation with *lin-41(ap427),* vulval bursting was suppressed, indicating that regulation of this edited version of *lin-41* is no longer dependent on let-7 (Figure 4.4E). We next attempted to generate a double mutant with *lin-41(ap427)* and *miR-48(n4097),* a deletion allele of the miR-48 miRNA. However, we were unable to generate a strain of homozygous double mutants because all animals burst through the vulva and died once they reached the L4 stage, regardless of culture temperature (Figure 4.4F). Finally, we generated *lin-41(ap427);miR-241(n4316)* double mutants, which had no observable phenotypes.

To quantify the dependence of *lin-41(ap427)* regulation on each let-7 family member, we conducted bursting assays at 25°C (Figure 4.4G). WT, *miR-48* null, *miR-241* null, and *lin-41(ap427)* animals did not burst at 25°C, whereas ~98% of *let-7(n2853)* animals burst and died. However, only ~7% of these let-7 mutants burst in the *lin-41(ap427)* background. As described above, a *lin-41(ap427);miR-48(n4097)* strain could not be isolated due to bursting and lethality of all double mutants, as confirmed by genotyping of the corpses. For *lin-41(ap427);miR-241* null animals,

there was no detectable bursting. The apparent miRNA specific regulation of WT *lin-41* by let-7 and *lin-41(ap427)* by miR-48 is not due to down-regulation of other let-7 family members in the miRNA mutant backgrounds (Figure 4.8).

Consistent with the phenotypes, the misregulation of *lin-41* mRNA levels in *let-7(n2853)* was prevented in the *let-7(n2853);lin-41(ap427)* strain (Figure 4.4H). Instead, expression of *lin-41* in the edited strain was found to be strongly misregulated in the absence of miR-48, as detected by single worm qRT-PCR (Figure 4.4I). Thus, the potential for pairing to miRNA 3' end sequences drives miRNA specific regulation at the molecular and phenotypic levels, which in the case of miR-48 regulation of *lin-41(ap427)* is not compensated by the presence of let-7 or miR-241. Taken together, these experiments reveal the importance of miRNA 3' end interactions in dictating target specificity among miRNA family members, even when targets share a canonical seed complement.

Figure 4.4: Seed pairing is not sufficient for target regulation *in vivo*. (A) Northern blot for *lin-41* in the indicated strains. (B-C) Diagram with binding profiles for let-7, miR-48, and miR-241 pairing to sites in WT *lin-41* and *lin-41(ap427)* where the let-7 complementary sites (LCS) have been switched miR-48 complementary sites (48CS). (D) *let-7(n2853)* animals burst through the vulva at 25°C. (E) Suppression of *let- 7(n2853)* bursting by *lin-41(ap427)*. (F) *lin-41(ap427);miR-48(n4097)* double mutants phenocopy *let-7(n2853)* vulval bursting. (G) Quantification of bursting in the presence (+) and absence (-) of the indicated gene products. Error bars represent $\pm$ SEM (H) qRT-PCR of *lin-41* in *let-7(n2853)* and *lin-41(ap427);let-7(n2853)*, normalized to WT and *lin-41(ap427)* strains, respectively. (I) Single worm qRT-PCR of *lin-41* in *lin-41(ap427);miR-48(n4097)* strains heterozygous or homozygous for the *miR-48* deletion, normalized to *lin-41(ap427);miR-48(+/+)*.


A
WT
let-7(n2853)
miR-48(n4097)
miR-241(n4316)
lin-41
28S rRNA
B
wild-type lin-41
3'UTR
LCS1 LCS2
LCS1 5' U GUU A A 3'
UUAUACAACC CUAC CUCA
GAUAUGUUGG GAUG GAGU
let-7 3' UU AU 5'
LCS2 5' U AUU U 3'
UUAUACAACC CUGCCUC
GAUAUGUUGG GAUGGAG
let-7 3' UU AU U 5'
LCS1 5' C A A 3'
CGU UCUAC CUCA
GCG GGAUG GAGU
miR-48 3' A UAGAUGACUC 5'
LCS2 5' CUUUUUA AACCAU U 3'
UAC UCUGCCUC
AUG GGAUGGAG
miR-48 3' AGCGUAG ACUC U 5'
LCS1 5' UU ACAAC A A 3'
UUAU CGU UCUAC CUCA
AGUA GCG GGAUG GAGU
miR-241 3' AAGA U 5'
LCS2 5' UU ACAACCAU U 3'
UUAU UCUGCCUC
AGUA GGAUGGAG
miR-241 3' AAGAGCGU U 5'
C
lin-41(ap427)
3'UTR
48CS1 48CS2
48CS1 5' U GA CC U 3'
UUGCAUCU ACU CUUACCUCA
AGCGUAGA UGA GGAUGGAGU
miR-48 3' CUC 5'
48CS2 5' U GA CC U 3'
UUGCAUCU ACU CUUACCUCA
AGCGUAGA UGA GGAUGGAGU
miR-48 3' CUC 5'
48CS1 5' AU A UC U 3'
CUG AC CC UUACCUCA
GAU UG GG GAUGGAGU
let-7 3' UU A UU AU 5'
48CS2 5' AU A UC U 3'
CUG AC CC UUACCUCA
GAU UG GG GAUGGAGU
let-7 3' UU A UU AU 5'
48CS1 5' A UGAA U 3'
UC CUC C CUUACCUCA
AG GAG G GGAUGGAGU
miR-241 3' UAAA C U 5'
48CS2 5' A UGAA U 3'
UC CUC C CUUACCUCA
AG GAG G GGAUGGAGU
miR-241 3' UAAA C U 5'
D
let-7(n2853)
E
let-7(n2853);
lin-41(ap427)
F
miR-48(n4097);
lin-41(ap427)
G
% animals burst at 25°C
100
80
60
40
20
0
let-7 + + + + - - + +
miR-48 + - + + + + - +
miR-241 + + - + + + + -
lin-41 with let-7 sites + + + - + - - -
lin-41 with miR-48 sites - - - + - + + +
wild-type
miR-48(n4097)
miR-241(n4316)
lin-41(ap427)
let-7(n2853)
let-7(n2853); lin-41(ap427)
miR-48(n4097); lin-41(ap427)
miR-241(n4316); lin-41(ap427)
H
lin-41 expression
fold change relative to strain background
5
4
3
2
1
0
let-7(n2853)
let-7(n2853); lin-41(ap427)
I
lin-41 expression
fold change
8
7
6
5
4
3
2
1
0
lin-41(ap427); miR-48(-/+)
lin-41(ap427); miR-48(-/-)

4.3.6 Simplified detection of endogenous miRNA target interactions by Chimera PCR (ChimP)

Although several groups have developed chimera-generating methods that can be applied in a variety of model systems, each of these protocols requires the use of radioactivity and complex sequencing data analysis. As a consequence, these methods are impractical for many research groups, which may be interested in a single miRNA or target. Considering that miRNA-target chimeras were detectable in standard iCLIP libraries using PCR (Figure 4.1A), we developed a method to facilitate the detection of miRNA-target chimeras by PCR. This method, called Chimera PCR (ChimP), allows the detection of miRNA-target chimeras without the use of radioactivity or the need for complex sequencing analyses (Figure 4.11A).

In brief, the ChimP protocol is similar to other chimera-generating methods that include an intermolecular ligation step to ligate miRNAs to their target RNAs (Grosswendt et al., 2014; Helwak et al., 2013; Moore et al., 2015). However, ChimP does not require radioactive tagging and isolation of RNA from a membrane. Instead, AGO-miRNA-target RNA tertiary complexes are treated with Proteinase K to isolate the RNAs. Libraries are then generated in a similar manner to standard iCLIP. The resulting library can then be used as the template in a PCR reaction using an oligonucleotide with the miRNA sequence as the forward primer and an oligonucleotide complementary to the target RNA as the reverse primer.

We used ChimP to confirm the miRNA-target chimeras for let-7 and *lin-41*, miR-48 and *dot-1.1*, and miR-241 and *lgg-2*, which were originally detected in our ALG-1 iCLIP reads (Figure 4.5A). To control for the possibility that the PCR products produced by ChimP were the result of amplification from a single primer, we also performed single primer controls to demonstrate that primers for both the miRNA and target sequence are required (Figure 4.5B). Furthermore, we cloned and sequenced

the PCR products for let-7 and *lin-41* and miR-48 and *dot-1.1* (Figure 4.5C). In both cases, the miRNA-target chimera contained the sequence of the two primers used for amplification separated by a small sequence that originated from the target RNA. This demonstrates that ChimP is capable of producing *bona fide* miRNA-target chimeras.

To test the fidelity of ChimP, we asked whether it could distinguish between miRNA family specific target sites. For both, *lin-41* and *dot-1.1,* we were only able to detect miRNA-target chimeras for let-7 and miR-48, respectively (Figure 4.5D and E). To further demonstrate the utility of ChimP, we also applied our method to detect miR-71 target sites to the *alg-1* and *C44F1.1* 3'UTRs that had been identified from our chimeric data (Figure 4.11B and C) and include examples of the biological reproducibility of ChimP (Figure 4.11D and E).

In some cases, we noted that primers generated products in the minus template control PCR reactions. Sequencing of these products showed that they were primer-dimers that amplified due to overlapping end complementarity. Furthermore, on occasion we found that low annealing temperatures, particularly for miRNAs with low GC content, or when library amplification primers were not fully removed from the library, led to the amplification of non-specific products, such as miRNA-rRNA reads. As a consequence, we recommend including no template control reactions as well as cloning and Sanger sequencing of fragments in pilot experiments employing ChimP.

Using ChimP, we were able to demonstrate that the let-7 specific binding of *lin-41* switches to miR-48 specific binding in the edited *lin-41(ap427)* strain (Figure 4.5D and F). These results further illustrate the utility of ChimP and show that specific binding of a miRNA target can be determined by miRNA 3' end sequences. Overall, these analyses demonstrate that ChimP is a versatile method to rapidly test for the

presence of endogenous miRNA-target interactions. With this assay, researchers can avoid laborious and computationally intensive CLIP-based chimera-producing experiments when specific miRNAs and potential target sites are of interest.

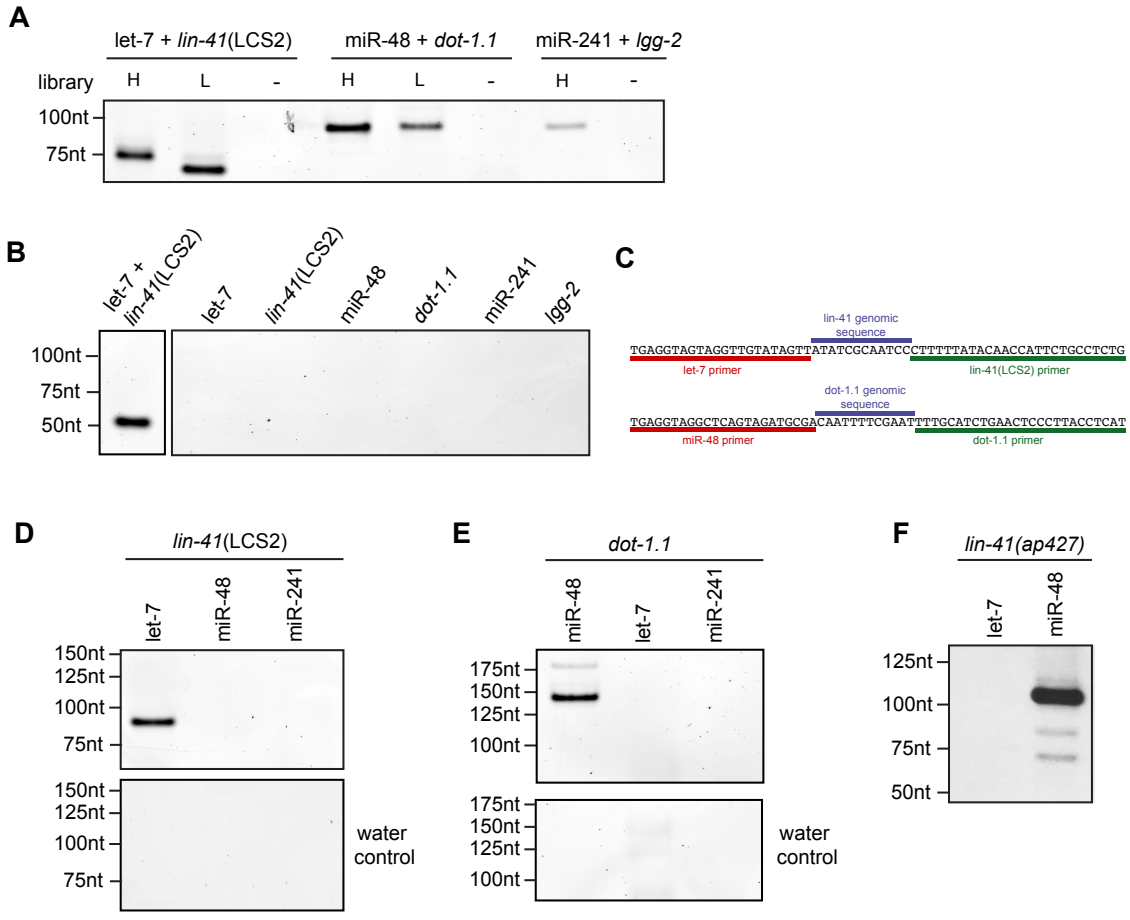


Figure 4.5: Chimera PCR (ChimP) enables the identification of miRNA-target chimeras by PCR. (A) Detection of let-7 family chimeras using ChimP. Libraries were generated using higher (H) and lower (L) molecular weight cDNAs. (B) Single primer negative controls along with a let-7 + *lin-41*(LCS2) positive control from another part of the gel. (C) Examples of sequenced ChimP products for let-7 and *lin-41*(LCS2), and miR-48 and *dot-1.1*. (D) Detection of the *lin-41*(LCS2) with let-7 but not with other let-7 family member primers by ChimP. (E) Specific detection of *dot-1.1* with a primer for miR-48. (F) The miR-48 complement sites in the *lin-41(ap427)* 3'UTR interact with miR-48 and not let-7.

4.4 Discussion

The initial aim of this study was to refine the catalog of AGO binding sites using iCLIP. Our serendipitous discovery that ALG-1 iCLIP produces miRNA-target chimeras led to the most comprehensive map of unambiguous miRNA target sites in *C. elegans* to date. Investigation of this reproducible dataset of endogenous miRNA target sites provided new insights into functional target interactions in a living animal. On a genome-wide scale, we found that miRNA-target interactions associated with regulatory outcomes generally involve seed pairing, an adenosine at the T1 position, and binding sites in mRNA 3'UTRs. Moreover, we were surprised to observe that miRNA families with divergent 3' ends target largely distinct sets of sites. We demonstrated that pairing to the miRNA 3' end not only provides specificity but can also be essential for target regulation fidelity *in vivo*. Finally, we developed ChimP to allow for the detection of chimeric reads by PCR, and anticipate that this will be a widely accessible method for interrogating potential miRNA-target site interactions.

4.4.1 Identification and validation of endogenous miRNA binding sites

The target sites identified by ALG-1 iCLIP chimeras represent endogenous miRNA-target interactions that occur at mid-L4. Presently, it is unclear why we recovered a larger fraction of chimeric reads in our iCLIP libraries (~1.3-5.1%) than did a previous directed attempt to form chimeras in *C. elegans* (~0.2%) (Grosswendt et al., 2014). Of the >150,000 miRNA-target chimeric reads generated by ALG-1 iCLIP, we conservatively used only those that were reproducible in at least two biological replicates. This requirement focused our studies on high-confidence sites

and helped to potentially eliminate off-target or transient binding events.

Although ALG-1 iCLIP produces miRNA-target chimeric reads with similar efficiencies as CLEAR-CLIP (Moore et al., 2015) and CLASH (Helwak et al., 2013), and more efficiently than modified iPAR-CLIP (Grosswendt et al., 2014), the biochemical steps were not specifically designed to promote the optimal 5'- and 3'-end chemistry that is required for chimera formation. It has been proposed that chimeric reads that form in CLIP-seq and PAR-CLIP libraries occur due to the action of an endogenous ligase present in the lysates (Grosswendt et al., 2014). In the CLIP-seq and PAR-CLIP libraries analyzed by Grosswendt et al. 2014, the majority of the miRNA sequences in chimeras were truncated at the 3'-end, likely by the RNase used to trim unprotected RNA fragments. However, full-length miRNA sequences account for the majority of miRNA-target chimeric reads in ALG-1 iCLIP data. Furthermore, our analysis of AGO-2 iCLIP data, where the 3'-linker was ligated after RNA isolation (Bosson et al., 2014), found almost no miRNA-target chimeric sequences (data not shown). This suggests that the T4 RNA ligase used to ligate the 3'-linker is responsible for the efficiency seen in ALG-1 iCLIP.

We have also demonstrated that ChimP can be used to identify miRNA-target sites for specific miRNA-target interactions of interest. An advantage of ChimP is the ability to detect chimeric events without the use of radioactivity or the bioinformatics expertise required to identify chimeras from CLIP-based methods. ChimP is sensitive enough to reproduce the specificity seen for the let-7 family miRNA targets observed in this study. Although ChimP does not identify the miRNA-target landscape across the transcriptome, it allows for the investigation of specific interactions that may be of interest to laboratories focused on certain miRNAs and targets.

4.4.2 Features of endogenous miRNA targeting

By capturing the endogenous miRISC, we were able to examine miRNA-target interactions *in vivo* without changing the stoichiometry between ALG-1, the miRNAs, and target RNAs. Perfect complementarity to at least a 6mer seed sequence was found in ~50% of target sites, and ~70% of all target sites exhibited at least partial seed complementarity. Interestingly, we did notice that the frequency of different types of seed interactions depended on the identity of the miRNA. For example, miR-60 is highly enriched for Class 1 pairing interactions, which involve complementarity to only the 5' end of the miRNA, and has greater potential for perfect pairing to nucleotides 2-8 than any of the other most abundant miRNAs (Figure 4.2E and I). This pattern could be related to the relatively low GC content of the miR-60 5' region, which might then be compensated for by strong seed pairing interactions to stabilize miRISC binding. Future studies will be important to understand the miRNA-specific positional binding preferences.

Similar to the mammalian target sites identified by CLEAR-CLIP and CLASH (Helwak et al., 2013; Moore et al., 2015), we also observed extensive predicted pairing to the 3' end of the miRNA, in addition to seed pairing. We identified seven classes of base pairing interactions between miRNAs and mRNAs, six of which featured various degrees of miRNA 3' end base pairing. Additionally, target sites were enriched for T1A, consistent with predicted features of miRNA target sites (Grimson et al., 2007; Lewis et al., 2005) and structural evidence that AGO contains a binding pocket for adenosine in the T1 position (Schirle et al., 2015, 2014).

In addition to distinct classes of pairing interactions being associated with different degrees of target regulation, the location of the target site also seems to be important for functional targeting. Similar to previous observations (Zisoulis et al., 2010), we noted that transcripts with target sites in coding exons were less

up-regulated in *alg-1(-)* animals than those with target sites in 3'UTRs. It is possible that some of these are regulated primarily at the level of translational inhibition with no detectable mRNA destabilization. Alternatively, competition between miRISC and translating ribosomes may reduce the residence time of AGO at exonic sites, thwarting effective regulation of the mRNA (Gu et al., 2009). However, not only were many target sites in coding exons reproducible but some gave rise to abundant chimeras, suggesting that these interactions are unusually stable. We noticed that some of these target sites overlap with published circRNAs for *C. elegans* (Ivanov et al., 2015; Memczak et al., 2013). Thus, it is possible that some of these stable, chimera generating target sites are derived from miRISC interactions with circRNAs and not the mRNA. In addition, circRNAs in *Drosophila* have been reported to be enriched for conserved miRNA seed complement sites, which further suggests that some circRNAs are bound by miRNAs (Westholm et al., 2014). One hypothesis is that these sites act as sponges to absorb excess miRNA load (Tay et al., 2014); however, other work has demonstrated that many circRNAs and other potential competing endogenous RNAs, due to their low expression, are not capable of functioning as miRNA sponges that titrate miRISC from mRNA targets (Bosson et al., 2014; Denzler et al., 2014). Thus, it remains to be determined what role coding-exon sites play or whether the substrates for AGO binding of this class are linear or circular RNA.

Target sites identified by ALG-1 iCLIP and other chimera-generating methods have also mapped to ncRNAs (Grosswendt et al., 2014; Helwak et al., 2013; Moore et al., 2015). Some of these interactions may represent novel functions for AGO, whereas others may arise from background ligation to highly expressed cellular RNAs. We anticipate that interactions between miRNAs and ncRNAs that are identified by chimera formation with a single miRNA are more likely to represent

functional interactions (Figure 4.12A) than sites that are bound by many unrelated miRNAs (Figure 4.12B). It remains unclear, outside of base pairing specificity, why some miRNAs seem to have preferences for particular types of ncRNA interactions. For example, miR-46/47 has twice as many snoRNA target sites, compared to all other expressed miRNAs, whereas miR-71 is devoid of snoRNA interactions (Figure 4.1E).

4.4.3 Specificity role for miRNA 3' ends

Since members of a miRNA family typically have identical 5' sequences but divergent 3' ends, they provide an ideal source for assessing contributions of the 3' supplementary region to specificity and function. Unexpectedly, specific targeting by miRNA family members seems to be much more common than anticipated and exists even in the presence of strong seed complementarity. For specific target sites, the miRNA that is bound to those sites is predicted to be more thermodynamically stable than other miRNAs of the same family. This suggests that base pairing interactions, beyond the seed sequence, are responsible for miRNA targeting specificity.

Target sites with weak seed sequence complementarity, such as those with bulged or mismatched nucleotides, are thought to be compensated by more extensive interactions with the 3' end of the miRNA (Brennecke et al., 2005; Grimson et al., 2007). As a consequence, sites with a weak seed may be more likely to be regulated by specific miRNA family members. However, single cell reporters have demonstrated that some sites with 6mer seed or 7mer-m8 seed with 3' supplementary complementarity can direct miRNA family specific regulation (Moore et al., 2015). These experiments suggested the possibility that these interactions alone could be functionally relevant.

To test whether miRNA 3' end interactions can direct functional specificity

when multiple miRNA family members are present at endogenous levels, we engineered an *in vivo* reporter based on the regulation of *lin-41*. Normally, *lin-41* is repressed solely by let-7 via two sites in its 3'UTR (Ecsedi et al., 2015; Reinhart et al., 2000; Slack et al., 2000; Vella et al., 2004). By editing those sites to become miR-48 target sites, we were able to transfer regulation of endogenous *lin-41* to miR-48. In the *miR-48* null background, animals with the edited *lin-41* 3'UTR phenocopy *let-7* null strains with completely penetrant busting and lethality (Reinhart et al., 2000). Importantly, the new miR-48 sites retain perfect seed pairing to any of the let-7 family members, yet only miR-48 appears capable of binding, as demonstrated by ChimP, and regulating this version of *lin-41.* Altogether, this work shows that miRNA family members have many distinct targets and demonstrates that targeting specificity among miRNA family members can have functional consequences *in vivo*.

Crystal structures of AGO have revealed that when bound to the miRNA alone, only nucleotides 2-6 of the miRNA seed sequence are positioned to interact with target RNAs (Elkayam et al., 2012; Nakanishi et al., 2012; Schirle and MacRae, 2012). However, when the complex includes a complement RNA, AGO undergoes a conformational change that allows seed pairing for nucleotides 2-8 of the miRNA and potentially exposes the miRNA 3' supplementary region (nts 13-16) for additional pairing interactions (Schirle et al., 2014). Our data suggest that these suspected conformational changes allow miRNAs capable of 3' end pairing interactions to outcompete miRNAs that support only seed pairing for a given site. Although *in vitro* studies have concluded that base pairing beyond the seed does not increase the affinity of AGO for a target site (Chandradoss et al., 2015; Jo et al., 2015; Salomon et al., 2015), it seems reasonable that target recognition in the more complex endogenous context could utilize additional base pairing interactions for specific and functional interactions *in vivo.*

Here we have reported the identification of thousands of examples of endogenous miRNA target sites in an intact organism. This work expands the dataset of experimentally captured miRNA-target interactions, providing a rich resource for improving target predictions and our understanding of miRNA targeting *in vivo*. For laboratories interested in select miRNA target sites, ChimP can be used to rapidly screen potential target interactions without having to analyze complex sequencing data. Our genome-wide analysis of chimeras formed by endogenous miRNA-target interactions revealed that pairing to the miRNA 3' end provides a degree of specificity not previously considered in most target prediction methods. As we have shown, the ability of a miRNA to recognize more than just seed sequences shared by all members of a family can have important consequences *in vivo*. The specificity provided by the miRNA 3' end may be especially relevant in humans where ~60% of miRNAs are part of a miRNA family (Kozomara and Griffiths-Jones, 2011). Specific miRNA family members are reportedly involved in numerous pathologies (Boyerinas et al., 2010), potentially because they have distinct targets that are not sufficiently regulated by the other family members. Overall, our results support the importance of seed pairing for functional miRNA-target interactions, but also reveal that this core motif might not always be sufficient. Instead, additional interactions with the miRNA 3' end may be necessary for specific targeting in the endogenous context.

4.5 Experimental Procedures

Nematode strains

Strains used and generation of *lin-41(ap427)* by CRISPR/Cas9 directed homologous recombination are described in the Supplemental Experimental Procedures.

ALG-1 iCLIP

ALG-1 iCLIP was performed as previously described (Broughton and Pasquinelli, 2013), using mid-L4 wild-type (N2) *C. elegans* animals grown at 25°C for 29 hours after L1 synchronization. Computational identification of ALG-1 binding sites and identification of chimeric reads is described in the Supplemental Experimental Procedures.

4.6 Accession Numbers

The sequencing data have been deposited in the SRA database under the accession numbers: SRP078361 (ALG-1 iCLIP) and SRP078368 (RNA-seq).

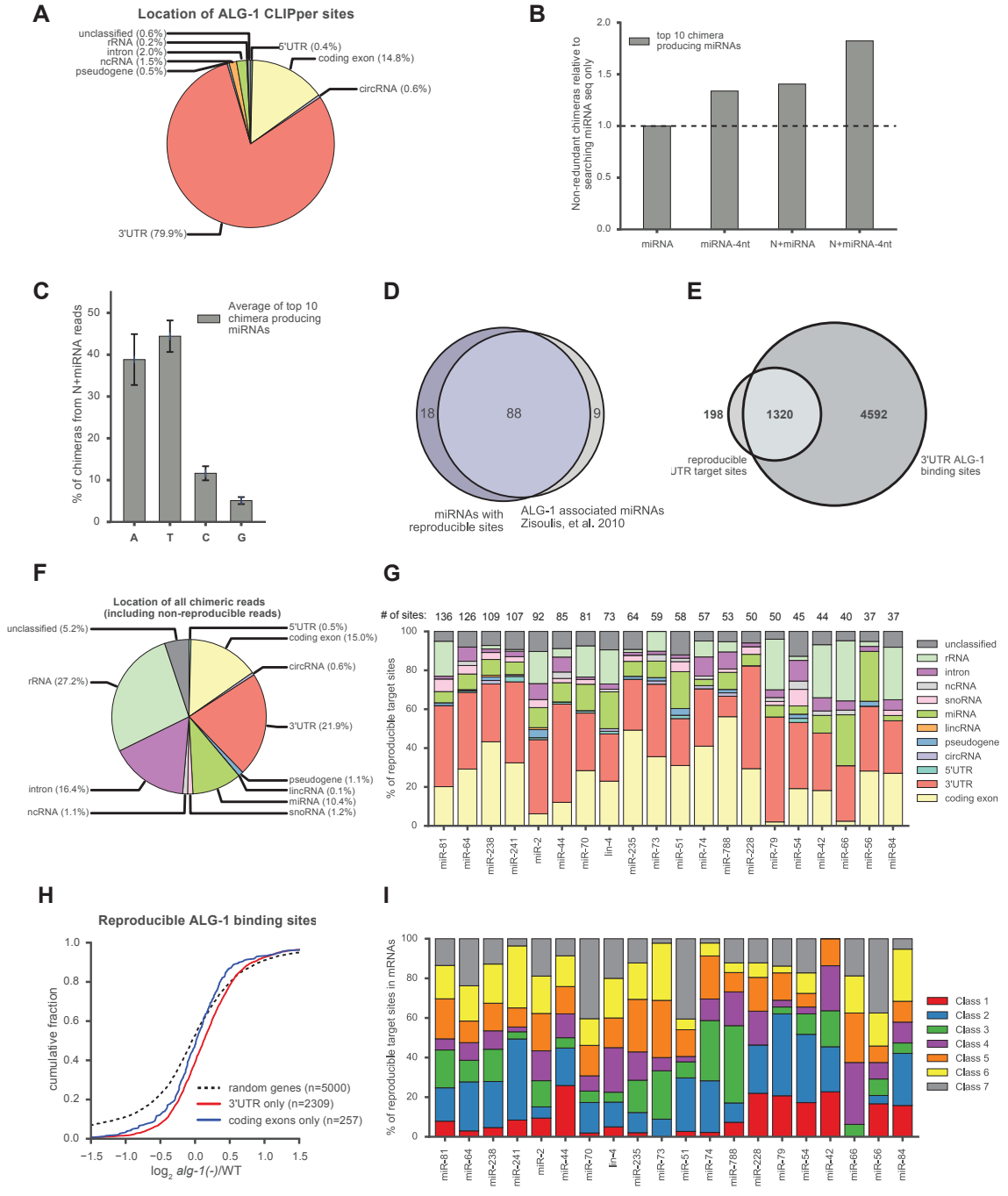
4.7 Author Contributions

J.P.B. designed and carried out all experiments with assistance from J.L.H. Computational experiments were carried out by J.P.B. and M.T.L. G.W.Y. provided advice in experimental design and analyses. A.E.P. conceived and supervised project. J.P.B. wrote the paper and all authors contributed to the final version.

4.8 Supplemental Information

4.8.1 Supplemental Figures

Figure 4.6: Features of ALG-1 iCLIP. (A) Distribution of genic locations for ALG-1 iCLIP binding sites identified by CLIPper. (B) Comparison of the number of non-redundant miRNA-target chimeric reads detected for the ten miRNAs with the most target sites using different cutadapt settings. Values are normalized to searching for the mature miRNA sequence with a tolerance of 1 nt error. (C) Identity of the untemplated nucleotide on the 5' end of the mature miRNA sequence. (D) Venn diagram showing the overlap of mature miRNAs (star strand excluded) with reproducible target sites and mature miRNAs that have been shown to be associated with ALG-1 (Zisoulis et al., 2010). (E) Venn diagram showing the overlap of 3'UTR target sites that overlap 3'UTR ALG-1 binding sites identified in at least one iCLIP library. (F) Distribution of genic locations for all chimeric reads (including those that are non-reproducible). (G) Distribution of target site genic locations for the 11-30 most target site producing miRNAs. Number of target sites for each miRNA is indicated. (H) Transcripts with ALG-1 binding sites only in the 3'UTR (red line), but not in coding exons (blue line), are up-regulated in *alg-1(-)* animals compared to random transcripts (Mann-Whitney U test, 3'UTR $P<2x10\text{-}26$, coding exon $P<0.16$). (I) Distribution of binding profile classes for the 11-30 most target site producing miRNAs.


A
Location of ALG-1 CLIPper sites
unclassified (0.6%)
rRNA (0.2%)
intron (2.0%)
ncRNA (1.5%)
pseudogene (0.5%)
5'UTR (0.4%)
coding exon (14.8%)
circRNA (0.6%)
3'UTR (79.9%)
B
top 10 chimera producing miRNAs
Non-redundant chimeras relative to searching miRNA seq only
2.0
1.5
1.0
0.5
0.0
miRNA
miRNA-4nt
N+miRNA
N+miRNA-4nt
C
Average of top 10 chimera producing miRNAs
% of chimeras from N+miRNA reads
50
40
30
20
10
0
A
T
C
G
D
18
88
9
miRNAs with reproducible sites
ALG-1 associated miRNAs Zisoulis, et al. 2010
E
198
1320
4592
reproducible UTR target sites
3'UTR ALG-1 binding sites
F
Location of all chimeric reads (including non-reproducible reads)
unclassified (5.2%)
5'UTR (0.5%)
coding exon (15.0%)
circRNA (0.6%)
rRNA (27.2%)
3'UTR (21.9%)
pseudogene (1.1%)
lincRNA (0.1%)
miRNA (10.4%)
snoRNA (1.2%)
intron (16.4%)
ncRNA (1.1%)
G
# of sites: 136 126 109 107 92 85 81 73 64 59 58 57 53 50 50 45 44 40 37 37
% of reproducible target sites
100
80
60
40
20
0
miR-81
miR-64
miR-238
miR-241
miR-2
miR-44
miR-70
lin-4
miR-235
miR-73
miR-51
miR-74
miR-788
miR-228
miR-79
miR-54
miR-42
miR-66
miR-56
miR-84
unclassified
rRNA
intron
ncRNA
snoRNA
miRNA
lincRNA
pseudogene
circRNA
5'UTR
3'UTR
coding exon
H
Reproducible ALG-1 binding sites
cumulative fraction
1.0
0.8
0.6
0.4
0.2
0.0
random genes (n=5000)
3'UTR only (n=2309)
coding exons only (n=257)
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
log2 alg-1(-)/WT
I
% of reproducible target sites in mRNAs
100
80
60
40
20
0
miR-81
miR-64
miR-238
miR-241
miR-2
miR-44
miR-70
lin-4
miR-235
miR-73
miR-51
miR-74
miR-788
miR-228
miR-79
miR-54
miR-42
miR-66
miR-56
miR-84
Class 1
Class 2
Class 3
Class 4
Class 5
Class 6
Class 7

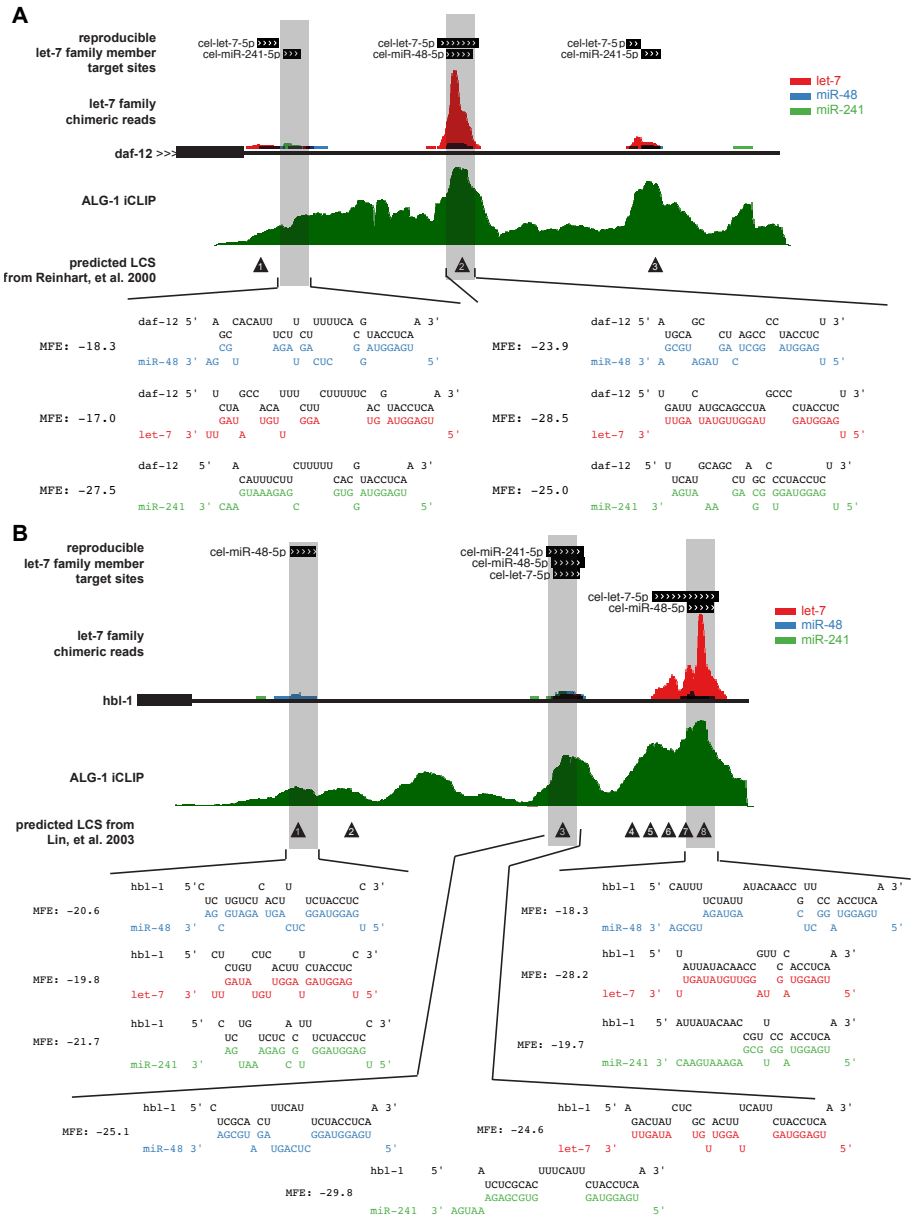


Figure 4.7: The 3'UTR of *hbl-1* and *daf-12* are bound by multiple let-7 family members. UCSC genome browser tracks for the (A) *daf-12* 3'UTR and (B) *hbl-1* 3'UTR. miRNA-target chimeric reads from five ALG-1 iCLIP libraries for let-7 family members are overlaid. ALG-1 iCLIP data from a single, representative library is shown. Reproducible target sites for all let-7 family members are shown. RNAhybrid predicted duplex structures and minimum free energy (MFE) are shown for indicated sites with let-7 family members.

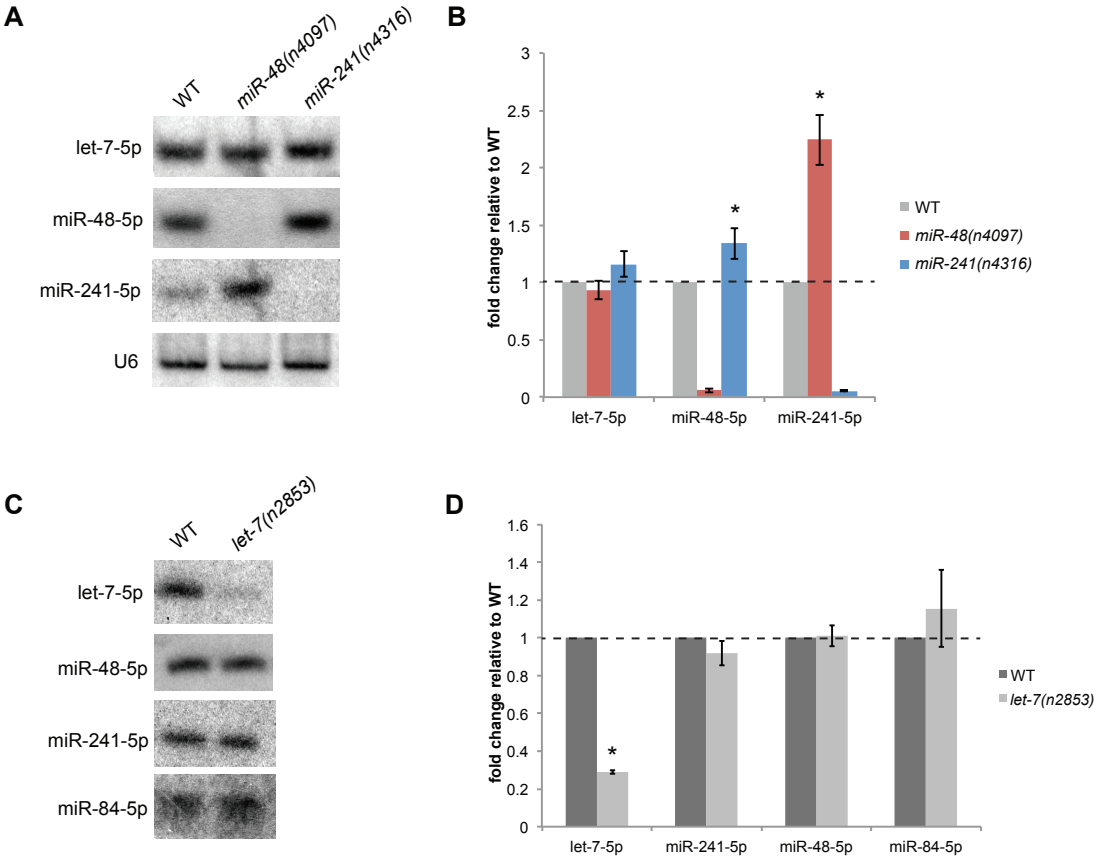


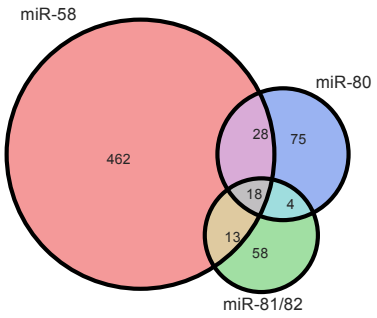
Figure 4.8: Expression of let-7 family miRNAs in *let-7*, *miR-48*, and *miR-241* mutant animals. (A) Representative northern blot of mature let-7, miR-48, and miR-241 miRNA expression in wild-type, *miR-48* and *miR-241* deletion animals. U6 snRNA is shown as a loading control. (B) Quantification of northern blots for three replicates for let-7 family expression in *miR-48* and *miR-241* deletion mutants, values are normalized to U6. Error bars represent S.E.M. Significant misregulation ($P<0.05$) is indicated by an asterisk. (C) Representative northern blot for mature let-7, miR-48, miR-241, and miR-84 miRNAs in wild-type and *let-7(n2853)* animals. (D) Quantification of northern blots for three replicates for let-7 family miRNA expression in *let-7(n2853)*. Error bars represent S.E.M. Significant misregulation ($P<0.05$) is indicated by an asterisk.

Figure 4.9: miR-58 and miR-238 family miRNAs bind non-overlapping sets of target sites. (A) Alignment of sequences for miR-58 family miRNAs. (B) Overlap of miR-58 family miRNA target sites in mRNAs. (C) miR-58 family miRNAs hybridize more favorably to their specific target sites. Minimum free energy of miRNA-target duplex calculated by RNAhybrid. (D) Alignment of sequences for miR-238 family members that are expressed at L4. (E) Overlap of miR-238 family miRNA target sites in mRNAs. (F) miR-238 family miRNAs hybridize more favorably to their specific target sites.

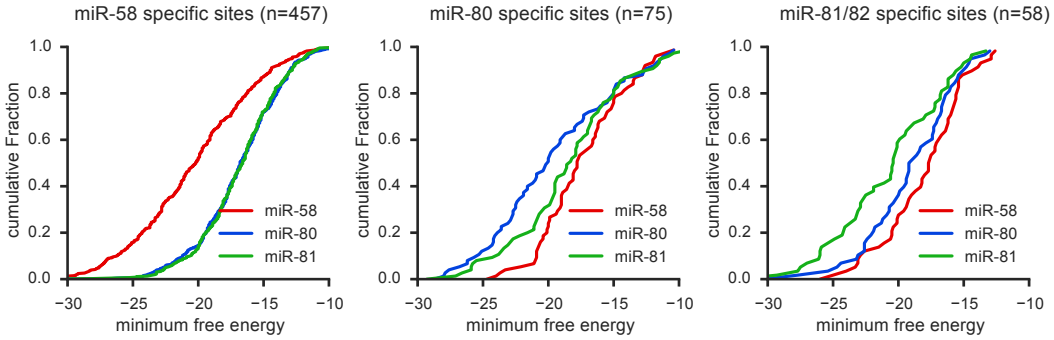
A

```
miR-58-3p       ugagaucguu--caguacggcaau
miR-80-3p       ugagaucauuaguugaaagccga-
miR-81-3p       ugagaucauc--gugaaagcuagu
miR-82-3p       ugagaucauc--gugaaagccagu
                ******* *     * * *
```

B **Overlap of miR-58 family target sites in mRNAs**



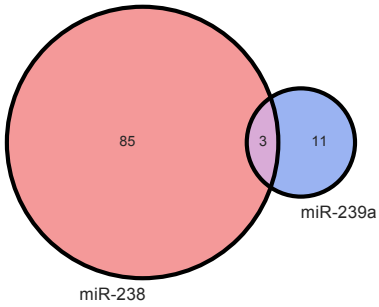
C



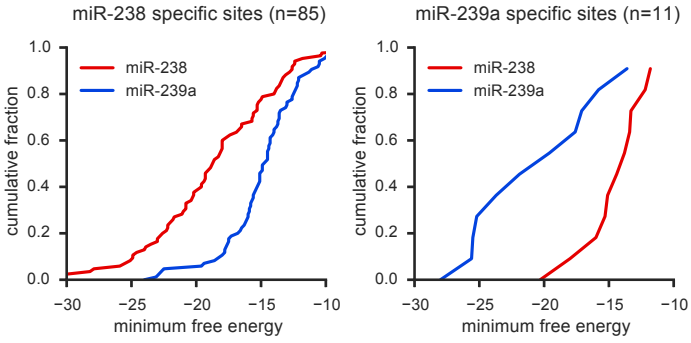
D

```
miR-238-3p      UUUGUACUCCGAUGCCAUUCAGA
miR-239a-5p     UUUGUACUACACAUAGGUACUGG
                ******** *       * * *
```

E **Overlap of miR-238 family target sites in mRNAs**



F



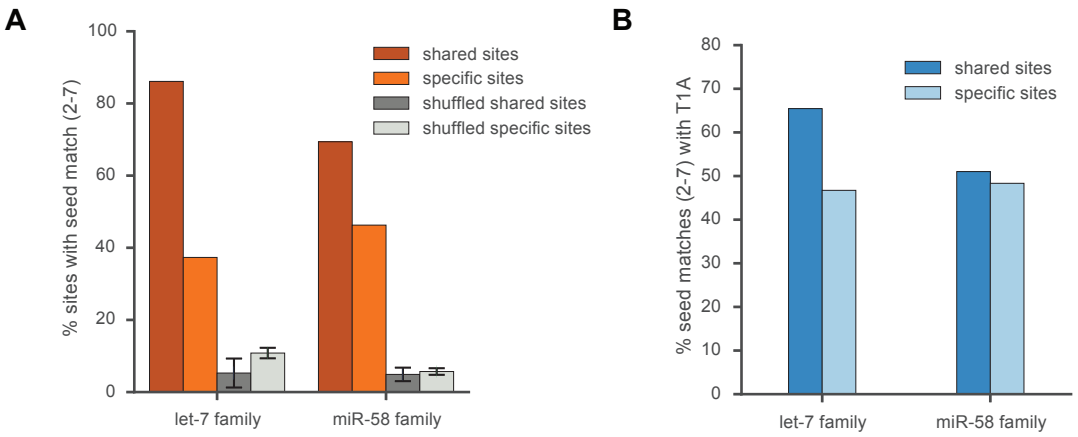
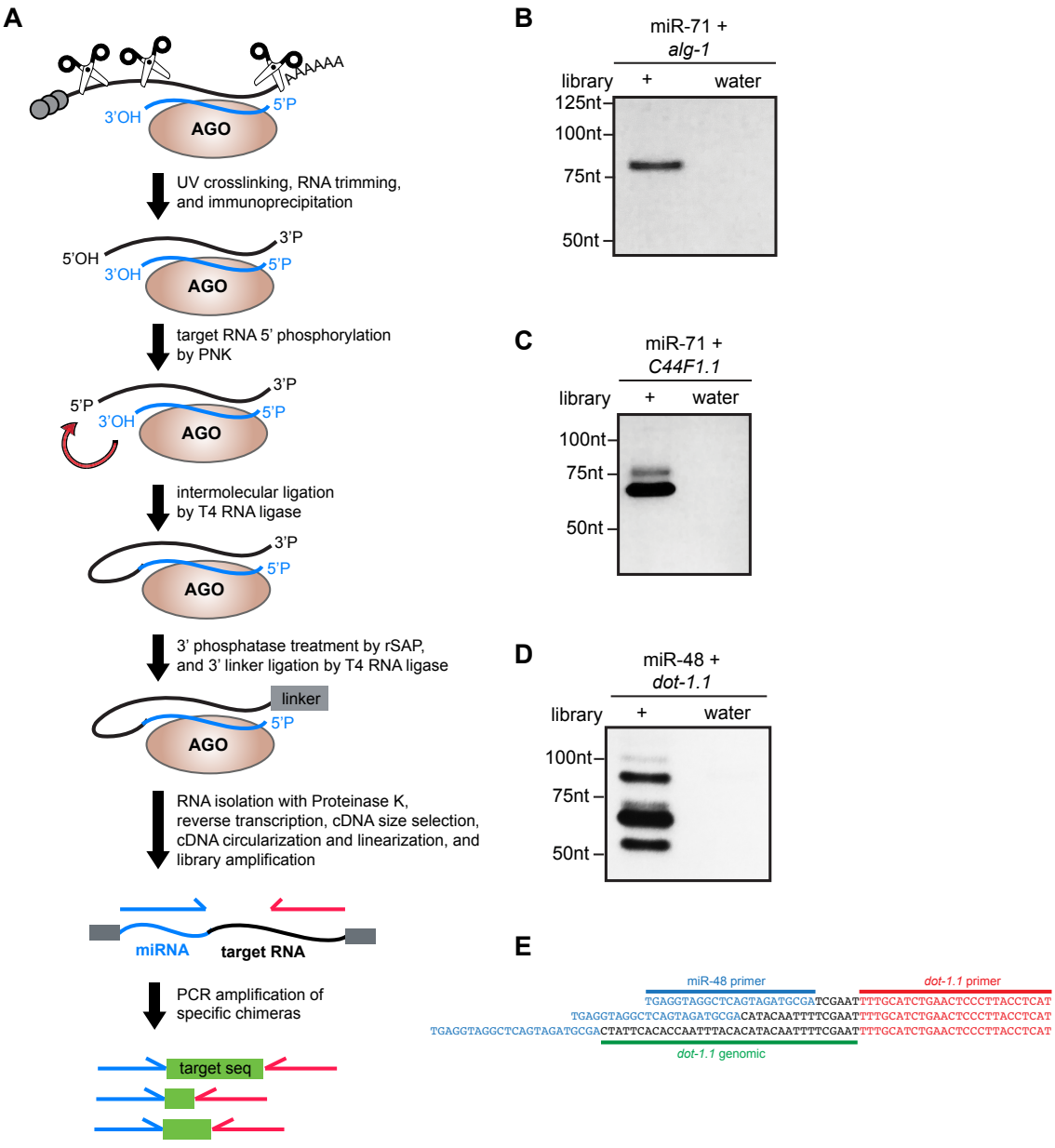


Figure 4.10: Comparison of seed complementarity in shared and specific sites. (A) Comparison of percent of sites with 6mer seed match in shared sites and specific sites for let-7 family and miR-58 family miRNAs. Percent of sites with 6mer seed match in shuffled dinucleotides shown. Error bars represent standard deviation. (B) Percent of 6mer seed matches (nts 2-7) with T1A in shared and specific sites for the let-7 family and miR-58 family of miRNAs.

Figure 4.11: Chimera PCR (ChimP) protocol and additional examples of ChimP. (A) Schematic showing the major biochemical steps in ChimP. Argonaute-miRNA-target RNA complexes are crosslinked using UV-C light in living animals. Lysates are prepared and treated with RNase to digest RNA not protected by Argonaute. Argonaute is then immunoprecipitated using a specific antibody. Target RNAs in purified complexes are then phosphorylated using PNK. Intermolecular ligation is carried out with T4 RNA ligase in the presence of ATP. The 3' phosphate left over after RNase digestion is removed by alkaline phosphatase and a DNA linker is added to the 3' end of the target RNA to allow reverse transcription. RNA is subsequently isolated using Proteinase K and reverse transcribed using a phosphorylated primer. cDNA is then size selected to remove extra reverse transcription primer. The purified cDNA is circularized and re-linearized to allow for amplification of all cDNA products, including those that were prematurely terminated at the crosslinking site. The cDNA is then amplified using Illumina sequencing primers. Chimeric reads are then detected using a forward primer complementary to the entire miRNA and a reverse primer complementary to the predicted target site. (B) Detection of miR-71 binding site in the 3'UTR of *alg-1*. (C) Detection of miR-71 binding site in the 3'UTR of *C44F1.1*. (D) Detection of miR-48 binding to the 3'UTR of *dot-1.1* using a library from a different biological replicate than the examples shown in Figure 5A and E. (E) Sequenced chimeras for miR-48 and *dot-1.1* from the reaction shown in Figure S5D.


A
AAAAAA
3'OH
5'P
AGO
UV crosslinking, RNA trimming,
and immunoprecipitation
5'OH
3'P
target RNA 5' phosphorylation
by PNK
intermolecular ligation
by T4 RNA ligase
3' phosphatase treatment by rSAP,
and 3' linker ligation by T4 RNA ligase
linker
RNA isolation with Proteinase K,
reverse transcription, cDNA size selection,
cDNA circularization and linearization, and
library amplification
miRNA
target RNA
PCR amplification of
specific chimeras
target seq
B
miR-71 +
alg-1
library
+
water
125nt
100nt
75nt
50nt
C
miR-71 +
C44F1.1
D
miR-48 +
dot-1.1
E
miR-48 primer
dot-1.1 primer
TGAGGTAGGCTCAGTAGATGCGATCGAATTTGCATCTGAACTCCCTTACCTCAT
TGAGGTAGGCTCAGTAGATGCGACATACAATTTTCGAATTTGCATCTGAACTCCCTTACCTCAT
TGAGGTAGGCTCAGTAGATGCGACTATTCACACCAATTTACACATACAATTTTCGAATTTGCATCTGAACTCCCTTACCTCAT
dot-1.1 genomic

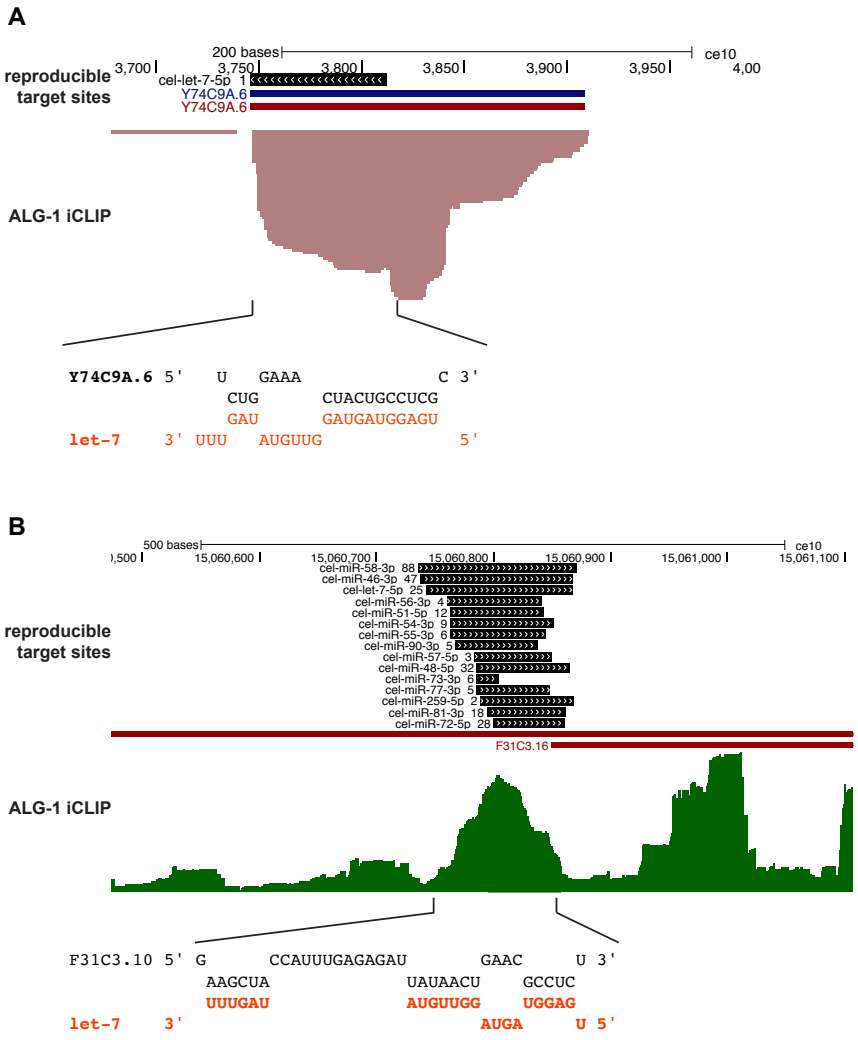


Figure 4.12: Target sites in non-coding RNAs. (A) Reproducible target sites and ALG-1 iCLIP reads that map to *Y74C9A.6*, a snoRNA. The single, reproducible target site for this snoRNA suggests that this interaction may be specific. RNAhybrid duplex prediction of let-7 to the target site shows that let-7 would be able to bind the snoRNA with strong seed pairing. (B) Reproducible target sites and ALG-1 iCLIP reads that map to *F31C3.10*, a ribosomal RNA. The many, unrelated miRNAs bound to this transcript and the weak predicted hybridization of let-7 by RNAhybrid to this site suggests that the interaction is nonspecific.

4.8.2 Supplemental Experimental Procedures

Nematode methods

Caenorhabditis elegans animals were cultured on OP50 bacteria at 25°C, unless otherwise noted, and standard synchronization methods were used (Brenner, 1974).

Generation of *lin-41(ap427)* by CRISPR/Cas9 directed homologous recombination

The *lin-41(ap427)* allele was generated by injecting young-adult wild-type (N2) animals with 25 ng/µL pJB38 (PU6::*lin-41*(LCS1) sgRNA), 25 ng/µL pJB39 (PU6::*lin-41*(LCS2) sgRNA), 2.5 ng/µL Pmyo2::tdTomoato, 10 ng/µL IR101 (HygR), 50 ng/µL Peft-3::cas-9::SV-40::unc-54, and 25 ng/µL pJB40 (homologous recombination template with miR-48 complement sites from *dot-1.1*). pJB40 was constructed using a GBlock from IDT, and 1.5kB homology arms to *lin-41*, and assembled using USER fusion (Nour-Eldin et al., 2010). F1's carrying plasmid arrays were selected with 300 µg/mL HygromycinB. Recombinants were identified by PCR screening. Genomic DNA was sequenced to confirm recombination event. The strain carrying the *lin-41(ap427)* allele was backcrossed to wild-type (N2) animals four times to produce PQ570. Primers used for generating plasmids for CRISPR/Cas9 genome editing are listed below.

Generation of *lin-41(ap427)* double mutants

PQ570 was then crossed to let-7 family mutant animals, *miR-48(n4036)*, *miR-241(n4316)*, *let-7(n2853)*. The double mutants *lin-41(ap427);miR-241(n4316)* (PQ571) and *lin-41(ap427);let-7(n2853)* (PQ577) were used in this study. To confirm

lethality of *lin-41(ap427);miR-48(n4036)* double mutants we carried out 3 independent crosses and genotyped F1 progeny (n=10). Primers used for genotyping of *lin-41(ap427)* and miRNA mutant strains are listed below.

miR-241(n4316) and *miR-48(n4036)* were obtained from the CGC and backcrossed to wild-type (N2) two times to generate PQ553 and PQ554, respectively.

Table 4.1: Oligonucleotides used for generating *lin-41(ap427)*

Identifier	Sequence	Description
A3070	AGCTAGAAAUAGCAAGTTAAAATAAGGCTAG	sgRNA backbone assembly (Addgene #46168)
A3071	ATTTAGAUTTGCAATTCAATTATATAGGGAC	sgRNA backbone assembly (Addgene #46168)
A3427	ATCTAAAUGTTTGacatcgcgttgagtgtagaaGTTTTAGAGCTAGAAAT	lin-41 sgRNA 1 assembly (F)
A3428	ATTTCTAGCUCTAAAACttctacactcaacgcgatgtCAAACATTTAGAT	lin-41 sgRNA 1 assembly (R)
A3429	GCAAATCTAAATGTTTGacatcg	genotype lin-41 sgRNA 1
A3430	ATCTAAAUGTTTGtcaatggttcagaggcagaaGTTTTAGAGCTAGAAAT	lin-41 sgRNA 2 assembly (F)
A3431	ATTTCTAGCUCTAAAACttctgcctctgaaccattgaCAAACATTTAGAT	lin-41 sgRNA 2 assembly (R)
A3432	GCAAATCTAAATGTTTGtcaatg	genotype lin-41 sgRNA 2
A3421	GGCTTAAUGGTGGCTGATAAAGATAATCATCGTGTC	lin-41 homology arm (F)
A3422	GGTTTAAUcagtaaattcctaaactgactgatagtgagtc	lin-41 homology arm (R)
A3423	accaacUcaagtatacct	lin-41 miR-48 complement site assembly (F)

Table 4.1: Oligonucleotides used for generating *lin-41(ap427)* (continued)

Identifier	Sequence	Description
A3424	agaaggUttcaatggttc	lin-41 miR-48 complement site assembly (R)
A3425	agttggUgcaatttgaggaaaagagg	lin-41 upstream homology arm (R)
A3426	accttcUcccgtactcccaccaatag	lin-41 downstream homology arm (F)
48SCS gBLOCK	caaattgcaccaactcaagtatacctTTTGCATCTGAACTCCCTTACCTCATcgcgatgtaaatatcgcaatccttTTTGCATCTGAACTCCCTTACCTCATgaaccattgaaaccttctcccgtac	miR-48 sites from dot-1.1 in lin-41 3'UTR

Table 4.2: Oligonucleotides used for genotyping strains

Identifier	Sequence	Description
A3442	ctactggtatctgtagaacaaagttcag	miR-48(n4097) (FWD)
A3443	cgcgtttgcctcgaaaac	miR-48(n4097) (REV)
A3446	caacaataaacgcaaactcatcagaatg	miR-241(n4316) (FWD)
A3447	ccttattgtctgggggctctatg	miR-241(n4316) (REV)
A75	cccgcggttcgcaacaatggagc	let-7(n2853) (FWD)
A76	tgagagcaagacgacgcagcttcg	let-7(n2853) (REV)
A3488	GAGGACGTGTTATTGTCGCTGATC	lin-41 CDS (FWD)
A3489	cgatatttacatcgcgATGAGGTAAGG	lin-41(ap427) (REV)
A25	GAGGCAGAATGGTTGTATAA	lin-41 3'UTR LCS2 (REV)

PCR-detection of miRNA-target chimeras

miRNA-target chimeras were detected in iCLIP libraries using PCR with a forward primer matching the mature miRNA sequence and a reverse primer complementary to the miRNA target site. Primers used for the detection of chimeras are listed below.

Computational identification of ALG-1 binding sites

ALG-1 binding sites were identified using the CLIPper peak finding algorithm after removing low-quality reads, trimming adapter sequences, collapsing duplicate reads, and removing reads that map to repetitive elements. A scala script to perform this task is available here:
https://github.com/YeoLab/gscripts/blob/master/qscripts/analyze_clip_seq.scala

Computational identification of miRNA-target chimeras

Read sequences were groomed with a two-stage filter through cutadapt (Martin, 2011). In the first filter, reads' 3' ends were truncated where the partial sum of sequencing quality scores fell below 36. The second filter was used to identify miRNAs within sequenced reads. Matches to *C. elegans* miRNAs (miRBase release 20 (Kozomara and Griffiths-Jones, 2011)) were required to be at least 18nt long and miRNA sequences were allowed to begin with an added base of any kind. A Python script to perform this task is available here:
https://github.com/YeoLab/gscripts/blob/master/gscripts/mirna/miR_splitter.py

A scala script to automate the trimming and mapping of chimeric reads is available here:
https://github.com/YeoLab/gscripts/blob/master/qscripts/analyze_miRli.scala

Mapping of miRNA-target chimeric reads

RNA-STAR was used to map non-miRNA portions of reads independently from their ligated miRNA counterparts. Non-miRNA portions of reads were mapped to WBcel220 (ce10). Non-miRNA portions of reads were also mapped to *E. coli* REL606 to determine background ligation frequency.

PCR-duplicate reduction

Duplicate reads / PCR duplication was identified by sequencing adapters with 5 random nucleotides (Konig et al., 2011). Reads that mapped to the same position with identical random-mers were collapsed. Duplicate removal was performed as previously described (Hung et al., 2015) with the Python script available here: https://github.com/YeoLab/gscripts/blob/master/gscripts/clipseq/barcode_collapse.py

Reproducible miRNA-target RNA chimeric sites

Reproducible sites were determined for each miRNA by intersecting target sites between replicates using Bedtools (Quinlan and Hall, 2010) and pybedtools (Dale et al., 2011). Target sites within 5 nucleotides, and on the same strand were considered to be the same site. Reproducible target sites from all combinations of replicates were combined into a single file. All reproducible target sites for a given miRNA were then merged using pybedtools into a single file. Sites were adjusted to be the same size by expanding 25nts upstream and downstream from the center of the site for analysis of base pairing interactions or genic location.

RNA-seq

RNA was collected from mid-L4 wild-type (N2) and *alg-1(gk214)* animals grown at 25°C for 29 hours after L1 synchronization. An Illumina TruSeq mRNA Library Prep Kit was used to generate poly(A) selected RNA-seq libraries. Illumina Solexa Sequencing data was analyzed by mapping reads to WBcel235 assembly of the C. elegans genome using RNA-STAR. Reads were then quantified using htseq-count using Ensembl 81 gene annotations. Differential expression was calculated using DESeq2. Sequencing reads are available through SRA: PRJNA328819.

Classification of genic types

Reproducible and non-reproducible target sites were lifted over to WBcel235 using liftover-utils (https://github.com/AndersenLab/liftover-utils). Sites were then classified by genic type by intersecting the target sites using pybedtools with WormBase.org WS247 gene annotations, published circRNAs (Ivanov et al., 2015; Memczak et al., 2013), and introns and rRNA sequences retrieved from the UCSC genome browser and lifted over to WBcel235.

k-means Clustering

k-means clustering was performed on predicted RNAhybrid duplexes between target RNAs and cognate miRNAs using Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm). For clustering analysis we considered only target sites in mRNAs. The optimal *k* was determined by maximizing BIC, k=5-10 were tested and k=7 produced the most meaningful classes. The clustered RNAhybrid duplexes were visualized using Java TreeView (Saldanha, 2004) and arranged manually into the presented order.

qRT-PCR

RNA was extracted from animals grown at 25°C for 29 hours after L1 synchronization and reverse transcription was performed as previously described (Van Wynsberghe et al., 2011). Fold changes were normalized to *Y45F10D.4* (Zhang et al., 2012) and calculated by $\Delta\Delta$ Ct.

Single worm qRT-PCR

Single worm lysates were generated by first picking worms into ddH2O with a hair pick and then transferring the washed animals into 10µL of single worm lysis buffer (5 mM Tris, pH 8.0; 0.5% Triton X-100; 0.5% Tween 20; 0.25 mM EDTA; and 1 mg/mL proteinase K (NEB)). Lysates were frozen in a dry-ice-ethanol bath for 5 minutes, then incubated at 65°C for 30 minutes and then 85°C for 5 minutes. 2µL of lysate were used to genotype. To the remaining 8µL of lysate, 8µL of 2X RQ-DNase mix was added. Lysates were then incubated at 37°C for 1 hour. RNA was then extracted using standard phenol:chloroform purification methods. The total amount of purified RNA was then used to generate cDNA.

Table 4.3: Oligonucleotides used for qRT-PCR, all oligonucleotides were obtained from IDT

Identifier	Sequence	Description
A2906	CGAGAACCCGCGAAATGTCGGA	Y45F10D.4 (F)
A2907	CGGTTGCCAGGGAAGATGAGGC	Y45F10D.4 (R)
A2203	ACATGTTTCTGGGCGATAGG	lin-41 (F)
A2204	CGTGCTGTTGGCTACTTCAA	lin-41 (R)
A3846	GCACAGTTGAGACGGAGACA	tag-260 (F)
A3847	GCGCCGTTACATGAGGTAGA	tag-260 (R)
A3840	TCCTTTGTACTCGGGTCGTTG	tcl-2 (F)
A3841	CGGTGGAGGATTCGCTTTGA	tcl-2 (R)
A3834	CGGCACCGACGGTACTGACGAAATC	sea-2 (F)
A3835	CTGGCGCATTTAAGTGACTGCTCGTC	sea-2 (R)
A3848	ATCCCGGAATCTGGCATCTC	nhr-7 (F)
A3849	CGACAGTTAGCAGTTGTCAGC	nhr-7 (R)
A3852	TGTGCTTCCCCAATGTTCGT	fozi-1 (F)
A3853	GCTTGACTACCACGTCTCCC	fozi-1 (R)
A3854	GGGATACCGTCTCCGTGTTC	C18B12.4 (F)
A3855	CCACGTACGCACAGTGTAATG	C18B12.4 (R)

Northern blotting

Small and large RNA species were detected using PAGE and agarose northern blotting methods, respectively (Van Wynsberghe et al., 2011). The lin-41 probe was generated using primers A3607 and A3608, miRNAs were detected using Starfire probes complementary to each miRNA of interest (sequences in listed below). Mature let-7 was detected using equal amounts of let-7 and let-7(n2853) complementary probes. All strains were grown for 29 hours at 25°C. Bands were quantified using the IMAGEJ software package (Schneider et al., 2012).

Table 4.4: Oligonucleotides used for northern blots, all oligonucleotides were obtained from IDT

Identifier	Sequence	Description
A3607	ccctcgtaattatccttgtttttcatac	lin-41 3'UTR (FWD)
A3608	cattaggcaattgggacaattaacacc	lin-41 3'UTR (REV)
A1114	AACTATACAACCTACTACCTCA/3StarFire/	let-7 starfire
A1132	TCGCATCTACTGAGCCTACCTCA/3StarFire/	miR-48 starfire
A1134	TCATTTCTCGCACCTACCTCA/3StarFire/	miR-241 starfire
A1216	ACTATACAACCTACTATCTCA/3StarFire/	let-7(n2853) starfire

Bursting Assays

Bursting assays were performed by growing synchronized L1 animals at 25°C for 54 hours and quantifying the number of animals that had burst through the vulva versus the number of surviving animals.

Chimera PCR (ChimP)

Cross-linked lysates were generated as previously described (Broughton and Pasquinelli, 2013) using mid-L4 wild-type (N2) C. elegans animals grown at 25°C for 29 hours after L1 synchronization. Lysates (1mg/mL) were digested with RNaseI

(Ambion) and Turbo DNase (Ambion) for 3 minutes at 37°C.

ALG-1 was immunoprecipitated from lysates using 7.5µg α-ALG-1 (ThermoFisher Scientific) using Protein G dynabeads. Beads were washed with high-salt wash buffer (50mM Tris-HCl, pH 7.4; 1M NaCl; 1mM EDTA; 1% NP-40; 0.1% SDS; 0.5% sodium deoxycholate) and PNK wash buffer (20mM Tris-HCl, pH 7.4; 10mM MgCl2; 0.2% Tween-20).

The 5' end of the target RNA was phosphorylated using T4 PNK (NEB) and 1mM ATP for 20 minutes at 37°C. The beads were then washed with high-salt wash buffer and PNK wash buffer.

Intermolecular ligation of the miRNA to the target RNA was performed using T4 RNA Ligase I (NEB) and 1mM ATP overnight at 16°C. The following morning, additional T4 RNA Ligase I (NEB) and 1mM ATP was added and the reaction allowed to proceed for an additional 5 hours. The beads were then washed with high-salt wash buffer and PNK wash buffer.

The removal of the 3' phosphate and the ligation of the 3' linker were performed as described previously (Broughton and Pasquinelli, 2013).

After linker ligation, the RNA was isolated using a Proteinase K digestion (Van Wynsberghe et al., 2011). The purified RNA was then reverse transcribed. RT primers and small products were removed by size selecting using AMPure XP beads (Beckman Coulter).

The cDNA was then circularized, linearized, and libraries were then amplified using Illumina sequencing primers (Broughton and Pasquinelli, 2013). Libraries containing residual primers were purified using AMPure XP beans (Beckman Coulter) before testing for chimeras. miRNA-target chimeras were detected using PCR by using a forward primer matching the mature miRNA sequence and a reverse primer complementary to the miRNA target site of interest.

Table 4.5: Oligonucleotides used for Chimera PCR and detection of chimeras in iCLIP libraries

Identifier	Sequence	Description
A706	tgaggtagtaggttgtatagtt	cel-let-7-5p
A3693	ATGAGGTAAGGGAGTTCAGATGCAAA	dot-1.1 3'UTR miR-48 site
A3694	ATGAGGAAGGTGATTTGCGAGAG	lgg-2 3'UTR miR-241 site
A3695	cagaggcagaatggttgtataaaaag	lin-41 3'UTR LCS2
A2795	TGAGGTAGGCTCAGTAGATGCGA	cel-miR-48-5p
A2796	TGAGGTAGGTGCGAGAAATGA	cel-miR-241-5p
A3488	GAGGACGTGTTATTGTCGCTGATC	lin-41 CDS (FWD)
A2879	/5phos/NNAGGTNNNAGATCGGAAGAGCGT CGTGgatcCTGAACCGC	iCLIP & ChimP RT
A2895	/5rApp/AGATCGGAAGAGCGGTTCAG/3ddc/	iCLIP & ChimP 3' adapter
A2893	AATGATACGGCGACCACCGAGATCTAC ACTCTTTCCCTACACGACGCTCTTCC GATCT	P5 solexa
A2894	CAAGCAGAAGACGGCATACGAGATCGG TCTCGGCATTCCTGCTGAACCGCTCT TCCGATCT	P3 solexa
A2922	GTTCAGGATCCACGACGCTCTTC/3ddc/	oligo for linearization
A3489	cgatatttacatcgcgATGAGGTAAGG	lin-41(ap427) 3'UTR
A3737	TGAAAGACATGGGTAGTGAGACG	cel-miR-71-5p
A3821	GATGCACTTGACAAGATATATTGGTG	alg-1 3'UTR miR-71 site
A3822	GAAAGAGCTATTGGTAGTTTTTGAGAAG	C44F1.1 3'UTR miR-71 site

4.9 Acknowledgments

We thank G. Pratt for assistance with software development and members of the Pasquinelli Lab for critical reading of the manuscript. Some strains were provided by the *C. elegans* Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). Support for this study was provided by a NSF Graduate Research Fellowship (DGE-1144086) (J.P.B.) and a UCSD Eureka Scholarship (J.L.H.). This work was supported by grants from NIH to G.W.Y. (HG004659, NS075449) and A.E.P. (GM071654).

Chapter 4, in full, is a reprint of the material as it occurs in *Molecular Cell*, Broughton, J.P., Lovci, M.T., Huang, J.L., Yeo, G.W., and Pasquinelli A.E., Cell Press, 2016. Broughton, J.P. was the primary author.

Chapter 5

Aging Argonautes: Building tools to study the microRNA Argonautes in *Caenorhabditis elegans*

5.1 Introduction

Argonautes (AGOs) are a family of proteins that bind various classes of small RNAs and direct small RNA-mediated gene regulatory activities. AGO proteins were initially discovered from a mutation in *Arabidopsis thaliana* that resulted in unusual leaf development that was reminiscent of a squid (Bohmert et al., 1998).

The presence of four domains classify AGO proteins: N-terminal, PAZ, MID, and PIWI. Proteins containing these four domains can be further divided into two subfamilies: AGO and PIWI. In humans, there are four members of the AGO subfamily and four members of the PIWI subfamily. In contrast in the nematode worm, *Caenorhabditis elegans*, the small RNA pathways have undergone extensive expansion and specialization, leading to an increase in the diversity of AGO proteins. In

total, there are 27 AGO family members in *C. elegans* with a variety of small RNA specializations (Youngman and Claycomb, 2014).

MicroRNAs (miRNAs) are a class of small, non-coding RNAs that are bound by AGO proteins. Once bound to AGO, miRNAs direct AGO and its associated factors to target messenger RNAs (mRNAs) through imperfect base-pairing interactions (Pasquinelli, 2012). Generally, the targets of miRNAs are located in the 3' untranslated region (3'UTR) of mRNAs. miRNA targets are post-transcriptionally regulated by the inhibition of translation or the destabilization of the mRNA (Jonas and Izaurralde, 2015). These regulatory processes are mediated by the co-factors that associate with the miRNA-bound AGO, which forms the core of the miRNA induced silencing complex (miRISC).

Two members of the *C. elegans* AGO family, Argonaute-like-gene 1 (ALG-1) and Argonaute-like-gene 2 (ALG-2), specifically function in the miRNA pathway (Grishok et al., 2001). Loss of *alg-1* leads to severe developmental abnormalities, whereas *alg-2* mutants do not exhibit strong phenotypes (Grishok et al., 2001; Tops et al., 2006; Vasquez-Rifo et al., 2012; Zinovyeva et al., 2015). This observation suggests that ALG-1 is the primary effector in the *C. elegans* miRNA pathway during larval development. The expression of fluorescently tagged ALG-1 and ALG-2 from transgenic arrays using their native promoters revealed that both of these proteins are expressed throughout development, with ALG-2 turning on earlier in embryogenesis (Vasquez-Rifo et al., 2012). Knockdown or deletion of both *alg-1* and *alg-2* results in embryonic lethality (Grishok et al., 2001), suggesting that ALG-2 can fill many of the functions of ALG-1 in the miRNA pathway.

In accordance with the important role of the miRNA pathway in development and the ability of miRNAs to regulate genes in many different pathways, it is unsurprising that RNAi knockdown of *alg-1* and *alg-2* result in decreased lifespans

(Kato et al., 2011). Similarly, mutations in other miRNA pathway genes such as *pash-1*, a component of the Microprocessor complex important for the biogenesis of miRNAs, also result in a shortened lifespan (Lehrbach et al., 2012). The deletion of some miRNAs (e.g. miR-71 and miR-238) lead to a decrease in lifespan, whereas the deletion of other miRNAs (e.g. miR-80 and miR-239a/b) lead to lifespan extensions (De Lencastre et al., 2010). The regulation of specific target genes in the individual miRNA deletion strains likely contributes to the different observed lifespan phenotypes.

An outstanding challenge in *C. elegans* research is the limited availability of antibodies that recognize *C. elegans* proteins (Duerr, 2006). Although some antibodies to highly conserved proteins made for vertebrates or invertebrates can cross-react with *C. elegans* proteins, many do not. However, recently the development of specific genome editing tools such as TALENs and CRISPR/Cas9 to generate double-strand breaks in DNA at specific loci in *C. elegans* has allowed for the rapid generation of strains with endogenously tagged proteins (Farboud and Meyer, 2015; Friedland et al., 2013; Kim et al., 2014; Lo et al., 2013; Tzur et al., 2013; Waaijers and Boxem, 2014). These methods leverage the endogenous double-strand break DNA repair pathways, non-homologous end joining or homologous recombination, to generate novel deletion alleles or introduce custom mutations in the DNA. In addition, since these methods target genes at their endogenous loci, the tagged proteins and mutations generated using TALENs or CRISPR/Cas9 provide advantages over transposon-based methods such as MosSCI, which uses the Mos1 transposon to insert DNA at specific pre-selected locations in the genome (Frøkjaer-Jensen et al., 2010). The transgenes generated by MosSCI and similar transposon based methods are limited by the presence of Mos1 sites and may be silenced in the germline.

Here, I describe a set of tools I generated for the exploration of the role ALG-1 and ALG-2 in development and aging. These tools complement a surprising result in our lab, which found that *alg-2* null mutants are long lived, even though previous *alg-2* RNAi experiments had suggested that *alg-2* knockdown results in a shortened lifespan. Using the various strains and tools I developed, our lab was able to demonstrate that ALG-1 and ALG-2 have opposing roles in regulating the lifespan of *C. elegans*.

5.2 Experimental Procedures

To generate the *alg-1* and *alg-2* 3'UTR specific RNAi, a USER cloning (Nour-Eldin et al., 2010) compatible RNAi vector (L4440) was generated by PCR (primers A3370 and A3371). USER cloning was then used to insert *alg-1* and *alg-2* 3'UTR PCR products (*alg-1* primers A3372 and A3373, and *alg-2* primers A3374 and A3375) that contained complementary overhangs to the USER-ready-L4440 backbone. Constructs were transformed into competent cells (DH5α) and confirmed by Sanger sequencing. Verified constructs were then transformed into HT115 bacteria.

To make the PQ530 (3xFLAG::GFP::ALG-1) strain, young adult N2 animals were injected with the following plasmids: 33ng/µL pJB14 (PU6:sgRNA targeting start codon of ALG-1 isoform B), 18ng/µL pCFJ104 (P*myo-3*::*mCherry*::*unc-54*; a gift from Erik Jorgensen; Addgene plasmid # 19328), 2ng/µL pCFJ90 (P*myo-2*::*mCherry*::*unc-54*utr; a gift from Erik Jorgensen; Addgene plasmid # 19327), 37ng/µL *Peft-3*::*cas-9*::*tbb-2*_3'UTR, 10ng/µL P*alg- 1*::3xFLAG::GFP::*alg-1* PCR product. Injected animals were allowed to reproduce at 20°C for 3-4 days and F1's expressing co-injection markers were singled to new plates and allowed to grow at 15°C for one week. Populations of F2 and F3 animals were genotyped for

homologous recombination of 3xFLAG::GFP at the N-terminal of ALG-1 by PCR (primers A2855 and A1001). A successful integrant was backcrossed 4x to N2 to generate PQ530.

To make the PQ549 (ALG-2::mCherry) strain, young adult N2 animals were injected with the following plasmids: 25ng/µL pJB34 (PU6:sgRNA targeting ALG-2), 25ng/µL pJB35 (PU6:sgRNA targeting ALG-2), 50ng/µL *Peft-3*::*cas-9*::*tbb-2*_3'UTR, pJB33 (ALG-2::mCherry homologous recombination template), 25ng/µL pIR101 (HygR plasmid co-splicing GFP), 2.5ng/µL pMyo2::tdTomato, 10ng/µL pGH8 (pRAB-3::mCherry::unc-54utr). Injected animals were allowed to reproduce at 20°C for 2 days and then the plates were then flooded with 1mL of 3mg/mL HygromycinB. Plates were returned to 20°C for 3-4 days. Hyg resistant worms were singled to new plates and allowed to reproduce until F3 animals were on the plate. 50µL worm population lysates were then made and screened for integration of mCherry construct by PCR (primers A3294 and A3295). A successful integrant was then backcrossed 3x to N2 to generate PQ549.

To make the PQ582 strain, young adult N2 animals were injected with the following plasmids: 10ng/µL pJB54 (3xFLAG::mKate2::ALG-2 homologous recombination template based on pDD285 (a gift from Bob Goldstein, Addgene plasmid # 66826)), 50ng/µL pJB53 (Cas9 plasmid with ALG-2 specific sgRNA, a modification of pJW1219 (a gift from Jordan Ward, Addgene # 61250)), 10ng/µL pGH8 (pRAB- 3::mCherry::unc-54utr, a gift from Erik Jorgensen, Addgene plasmid # 19359, pCFJ104 (Pmyo- 3::mCherry::unc-54utr; a gift from Erik Jorgensen; Addgene plasmid # 19328), 2.5ng/µL pMyo2::tdTomato. Isolation of recombinants was performed as described in Dickinson et al. (2015). Briefly, injected animals were grown at 25°C for 3 days. Plates were then flooded 1mL of 3mg/mL HygromycinB. Plates were returned to 25°C for 3 days. Non-glowing roller, Hyg resistant worms were

singled to new plates. L1's from plates with 100% non-glowing, roller worms were moved to new plates and heat shocked at 34°C for 4hrs to remove the cassette. Wild-type worms post heat shock were selected and screened by PCR for integration of 3xFLAG::mKate2 at the ALG-2 locus. Strain was then backcrossed 2x to N2 to generate PQ582.

To make the PQ567, *alg-2(ap426),* strain, young adult N2 animals were injected with the following plasmids: 25ng/µL each of 2 sgRNAs targeting the second exon of *alg-2*, 50ng/µL *Peft-3*::*cas-9*::*tbb-2*_3'UTR, 2.5ng/µL *Pmyo-2*::*tdTomato*, and 25ng/µL IR101. The injected animals were grown at +15°C for 2-3 days, after which the plates were flooded with 3mg/mL HygromycinB to achieve a final concentration of 0.3mg/mL and returned to 15°C for 2 days. HygromycinB resistant animals were moved to single plates, allowed to lay eggs, and genotyped to test for a change in the 3'UTR using primers A3203, A3184, and A3185 (PQ535). The potential PQ567 animals were genotyped for a change in the size of the targeted exon of *alg-2* using primers flanking the region targeted by the two sgRNAs (A3534 and A3535). A successful integrant (PQ535), and a line with a frameshift mutation resulting from an 8 nt deletion in the second exon (PQ567) were each backcrossed 4x to N2.

The functionality of *alg-1* and *alg-2* 3'UTR RNAi was determined by testing for embryonic lethality, as knockdown of one *alg-1* or *alg-2* in the opposite mutant results in eggs that fail to hatch (Grishok et al., 2001). Wild-type (N2), *alg-1(gk214)*, *alg-1(tm492),* and *alg-2(ok304)* animals were plated on 10X bacterial lawns of EV (L4440), *alg-1* library RNAi, *alg-1* 3'UTR RNAi, *alg-2* library RNAi, *alg-2* 3'UTR RNAi at L4. Animals were then grown on RNAi at 20°C for 2-3 days. The ratio of hatched to unhatched larvae was then scored.

The functionality of the fluorescently-tagged AGOs and the *alg-2(ap426)* allele was also tested by embryonic lethality. Wild-type, *alg-1(gk214), alg-2(ok304)*,

PQ562, PQ530, PQ582, and PQ549 were plated on 10X bacterial lawns of EV (L4440), *alg-1* library RNAi, and *alg-2* library RNAi at L4. Animals were then grown on RNAi at 20°C for 2-3 days. The ratio of hatched to unhatched larvae was then scored. AGOs were determined to be functional if embryos hatched at normal ratios, whereas AGOs were determined to be non-functional if the majority embryos failed to hatch.

A double mutant (PQ583) containing both the 3xFLAG::GFP::ALG-1 and the mKate2::3xFLAG::ALG-2 alleles was generated by crossing PQ582 to PQ530. PCR was used to confirm the presence of homozygosity for both of these alleles.

5.3 Results

Work by other labs had previously demonstrated that *alg-1* and *alg-2* knock-down by RNAi results in a shortened lifespan. The RNAi libraries used to target *C. elegans* genes target regions within the coding regions of genes usually around 800-1000 nucleotides in length (Kamath et al., 2003). Considering that *alg-1* and *alg-2* are the result of a recent gene duplication event (Grishok et al., 2001), it may be possible that RNAi to one gene is also targeting the other. To address this possibility, I first analyzed the sequence similarity of the regions targeted by the library *alg-1* and *alg-2* RNAi. In both cases, the library RNAi targets the nucleotides that encode the highly conserved PAZ domains, with 78% nucleotide identity (Figure 5.1A). However, the 3'UTRs of *alg-1* and *alg-2* are not as highly conserved. To generate RNAi constructs that would be specific to *alg-1* and *alg-2*, I cloned sequences from the *alg-1* and *alg-2* 3'UTRs. These two regions have 46% nucleotide identity (Figure 5.1A), and should therefore not cross-target the other AGO.

To resolve whether the shortened lifespan of the *alg-2* library RNAi or the

extended lifespan of the *alg-2* null allele, *alg-2(ok304),* was accurate, we compared the lifespans of animals grown on the library and 3'UTR specific RNAi. As previously described, the *alg-2* library RNAi resulted in a decreased lifespan. Although the lifespan of animals on *alg-2* library RNAi was significantly shortened compared to wild-type, animals on *alg-2* 3'UTR had a significantly longer lifespan compared to wild-type (Figure 5.1B). This finding was in agreement with the lifespan of the *alg-2* null strain. To provide additional evidence in support of the extended *alg-2* lifespan, I used CRISPR/Cas9 genome editing to generate a new *alg-2* null allele, *alg-2(ap426).* This allele contains an eight-nucleotide deletion towards the N-terminus that results in a frame shift and premature stop codon (Figure 5.1C). The lifespan of *alg-2(ap426)* animals was extended compared to wild-type, but interestingly, not as extended as seen in *alg-2(ok304)* animals (Figure 5.1D).

To explore the role of ALG-1 and ALG-2 in lifespan, we looked at the expression of these proteins during larval development and aging. Using an antibody that our lab had previously generated against ALG-1, we found that ALG-1 is constitutively expressed throughout larval development and into adulthood (Figure 5.2A). However, during aging ALG-1 expression is halted and remains off (Figure 5.2B and C). To address whether ALG-1 remained on in specific tissues, I generated a 3xFLAG::GFP::ALG-1 strain using CRISPR/Cas9 genome editing and homologous recombination. Fluorescent microscopy of this strain revealed that ALG-1 expressed decreased in all tissues (Figure 5.2D).

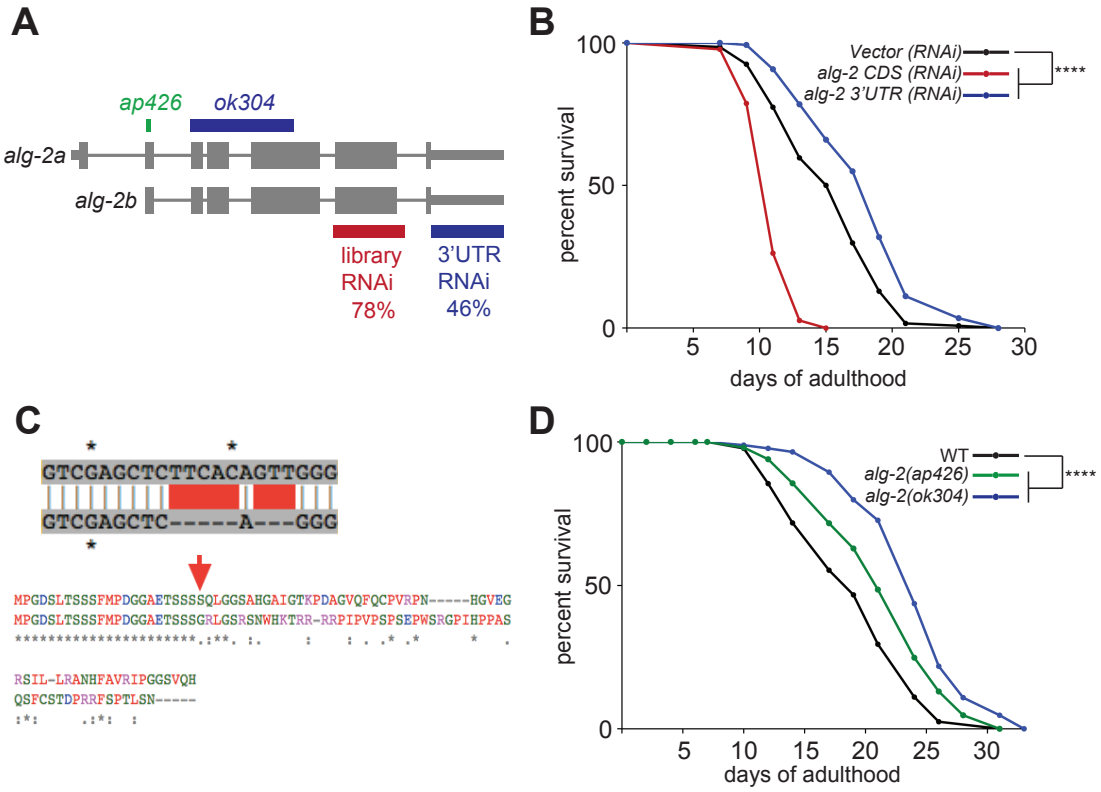
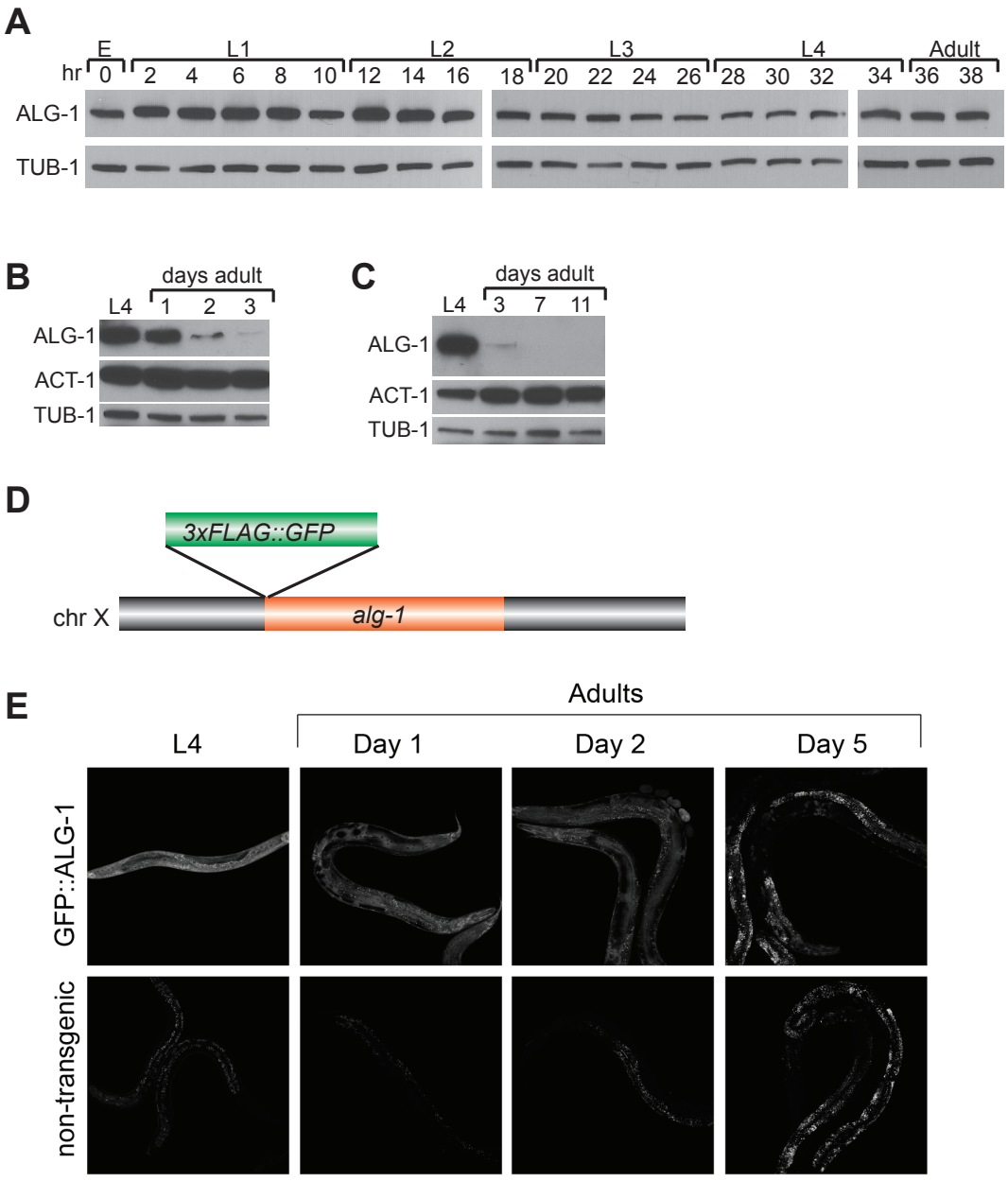


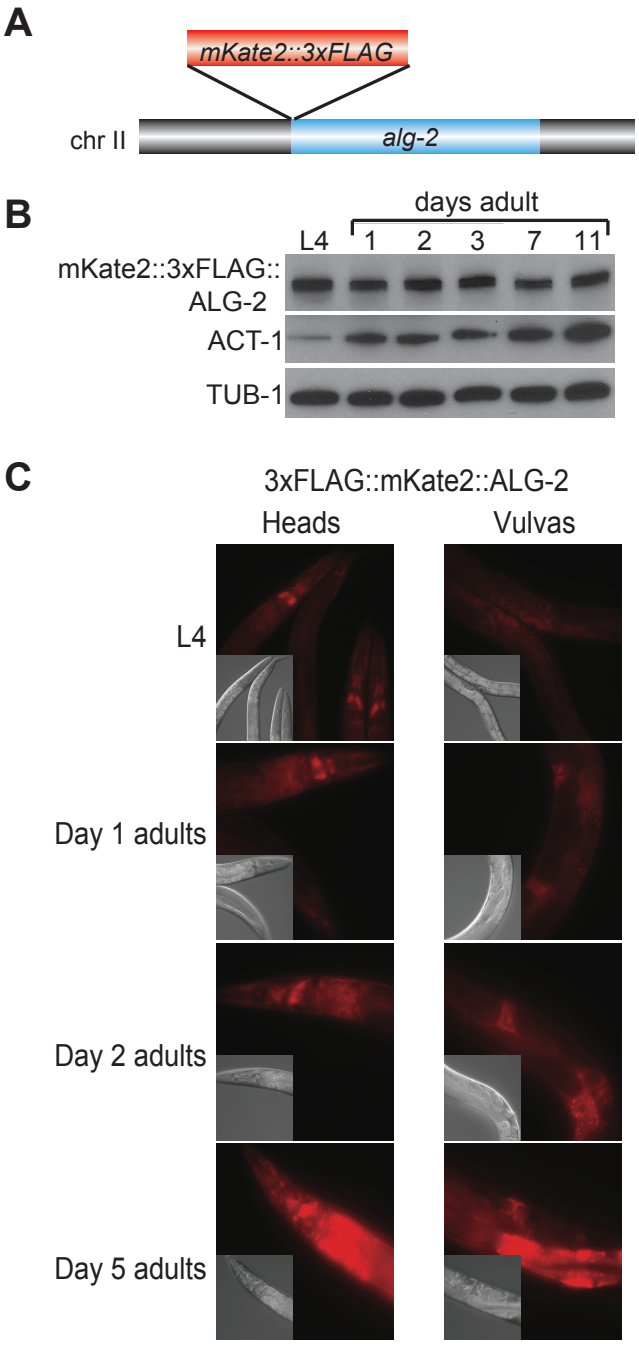
Figure 5.1: Extended lifespan of ALG-2. (A) Diagram showing isoforms of *alg-2* and the region deleted in the *alg-2(ok304)* allele. Locations targeted by library RNAi and 3'UTR specific RNAi are indicated and the percent nucleotide identity with *alg-1* is shown. The location of the new *alg-2* null allele *alg-2(ap426)* is also shown. (B) Representative lifespan of wild-type animals on EV, *alg-2* library RNAi, and *alg-2* 3'UTR RNAi. Lifespan assay performed by A. Alto and I. Nicastro. (C) Schematic showing the eight nucleotide deletion in *alg-2(ap426)*, and comparison of predicted protein product of wild-type *alg-2* and *alg-2(ap426)*, location of the frame shift in *alg-2(ap426)* is indicated by a red arrow. (D) Representative lifespan of wild-type, *alg-2(ok304)*, and *alg-2(ap426)* animals. Lifespan assay performed by A. Alto and I. Nicastro.

Figure 5.2: Loss of ALG-1 expression during aging. (A) Immunoblot of ALG-1 expression throughout larval development and early adulthood. Immunoblot performed by H. Jenq. (B) Immunoblot of ALG-1 in days 1, 2, and 3 of adulthood. Immunoblot performed by A. Pasquinelli. (C) Immunoblot of ALG-1 in days 3, 7, and 11 of adulthood. Immunoblot performed by A. Pasquinelli. (D) Schematic of how 3xFLAG::GFP::ALG-2 strain was generated by CRISPR/Cas9 genome editing and homologous recombination. (E) Confocal microscopy of 3xFLAG::GFP::ALG-1 strain and non-transgenic animals showing the decrease in ALG-1 expression in all tissues during aging. Non-specific autofluorescence common to aging animals is observed at similar levels in non-transgeneic and 3xFLAG::GFP::ALG-1 animals. Microscopy performed by A. Aalto.


A
E
L1
L2
L3
L4
Adult
hr
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38
ALG-1
TUB-1
B
days adult
L4 1 2 3
ALG-1
ACT-1
TUB-1
C
days adult
L4 3 7 11
ALG-1
ACT-1
TUB-1
D
3xFLAG::GFP
chr X
alg-1
E
Adults
L4
Day 1
Day 2
Day 5
GFP::ALG-1
non-transgenic

The loss of ALG-1 suggested a collapse of the miRNA pathway during aging, but the expression of ALG-2 at this time remained unclear. Studies of ALG-2 protein expression have previously been difficult due to the lack of a good antibody against ALG-2. To address this issue, I developed an mKate2::3xFLAG::ALG-2 strain through CRISPR/Ca9 genome editing and homologous recombination (Figure 5.3A). Immunoblots of mKate2::3xFLAG::ALG-2 through adulthood showed that ALG-2 remained expressed in aging animals (Figure 5.3B). Fluorescent microscopy confirmed the results of the immunoblot and also suggested that ALG-2 expression may increase during aging (Figure 5.3C).

Figure 5.3: Expression of ALG-2 during adulthood. (A) Schematic of how mKate2::3xFLAG::ALG-2 strain was generated by CRISPR/Cas9 genome editing and homologous recombination. (B) Immunoblot showing ALG-2 expression throughout adulthood. Immunoblot performed by A. Aalto. (C) Fluorescence microscopy showing the continued expression of ALG-2 in mKate2::3xFLAG::ALG-2 animals from L4 to adult day 5. Microscopy performed by I. Nicastro.


A
mKate2::3xFLAG
chr II
alg-2
B
days adult
L4
1
2
3
7
11
mKate2::3xFLAG::
ALG-2
ACT-1
TUB-1
C
3xFLAG::mKate2::ALG-2
Heads
Vulvas
L4
Day 1 adults
Day 2 adults
Day 5 adults

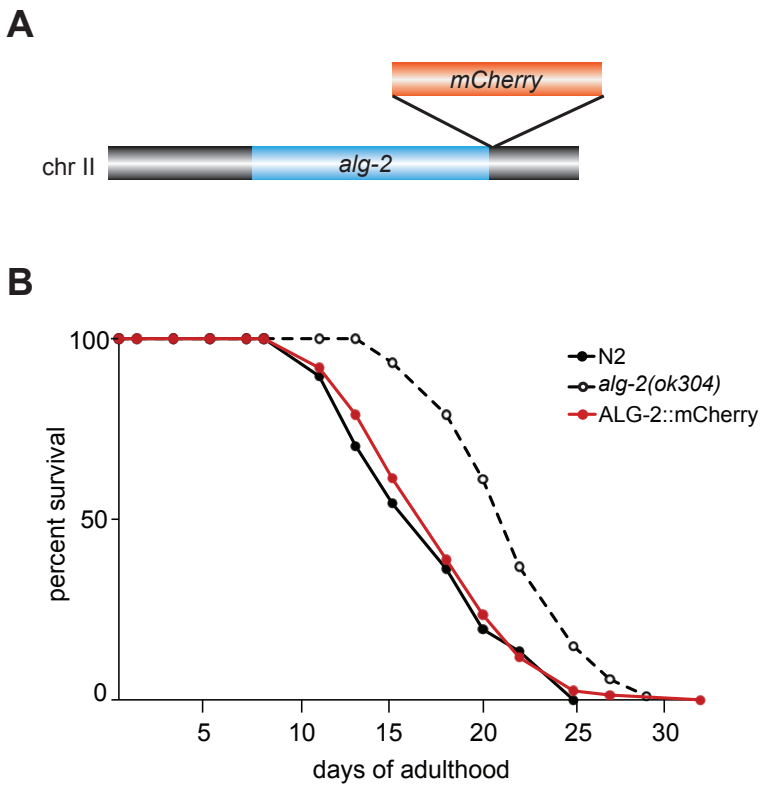


Figure 5.4: No-lifespan extension in ALG-2::mCherry animals. (A) Schematic of how the ALG-2::mCherry strain was generated by CRISPR/Cas9 genome editing and homologous recombination. (B) Lifespan of wild-type, *alg-2(ok304)*, and ALG-2::mCherry animals.

The C-terminal end of AGO proteins is buried with the protein (Elkayam et al., 2012; Nakanishi et al., 2012; Schirle and MacRae, 2012). As a consequence, tagging the C-terminal end of AGO proteins may result in a nonfunctional protein. In the process of generating the N-terminal tagged mKate2::3xFLAG::ALG-2, I also generated a C-terminal tagged ALG-2::mCherry (Figure 5.4A). Treating this strain with either *alg-1* library or *alg-1* 3'UTR RNAi resulted in embryonic lethality, which demonstrated that the C-terminal tagged ALG-2::mCherry was not functional in this assay. However, it remains a possibility that ALG-2::mCherry is retains enough function to not result in the extended lifespan seen in *alg-2(-)* animals. Surprisingly, a lifespan conducted using this C-terminal tagged version of ALG-2 showed no

increase in lifespan compared to wild-type, even though ALG-2::mCherry was no longer able to compensate for the loss of ALG-1 during embryogenesis (Figure 5.4B).

5.4 Discussion

Using the various strains and RNAi constructs that I generated, our lab was able to explore differences in the expression of ALG-1 and ALG-2 during adulthood and observe the opposing lifespan phenotypes caused by deletions of these AGO proteins. The construction of the 3xFLAG::GFP::ALG-1 and mKate2::3xFLAG::ALG-2 strains allowed for one of the first examinations of the expression patterns for both of these proteins from their endogenous loci. Interestingly, we found that ALG-1 expression rapidly decreases in the first few days of adulthood, whereas ALG-2 expression remains on.

In addition to confirming that *alg-2(ok304)* animals are long lived, the 3'UTR specific RNAi constructs that I built reveal an important caveat when studying proteins using RNAi that have highly related homologs. In this case, *alg-2* RNAi appeared to also target *alg-1*, which led to the decreased lifespan phenotype for *alg-2* observed in the initial studies of the miRNA pathway and aging (Samuelson et al., 2007). Furthermore, when comparing animals grown on *alg-1* library RNAi and *alg-1* 3'UTR RNAi, I noticed that the animals on *alg-1* 3'UTR RNAi appeared healthier than those grown on *alg-1* library RNAi. This may result from decreased potency of the 3'UTR RNAi, or may indicate that some of the severe phenotypes observed in with the library RNAi, which are often more severe than seen in *alg-1* null strains, result from additional cross-targeting issues. These observations underscore the importance of generating and confirming phenotypes with genetic mutants, as I did by generating

the novel *alg-2(ap426)* null allele.

By crossing the two fluorescent-AGO strains, I was also able to generate a tool that could be potentially used to identify differential expression of ALG-1 and ALG-2 in a variety of tissues. Preliminary confocal microscopy of this dual tagged strain has shown that while many neurons express both ALG-1 and ALG-2, there exists a subset of neurons that are specific for either ALG-1 or ALG-2. Future analysis of the specificity of ALG-1 and ALG-2 in neurons may contribute to our understanding of how the miRNA pathway can control a variety of pathways in *C. elegans*, including behavior or stress response.

Lifespan assays and other experiments conducted by other members of our laboratory as part of this work identified ALG-1 and ALG-2 as having antagonistic roles in the insulin signaling pathway. Although these analyses reveal ALG-1 and ALG-2 dependent changes during aging, the mechanism by which ALG-1 promotes increased lifespan and ALG-2 decreases lifespan through the insulin signaling pathway remains unclear. An attractive hypothesis is that the direct targets of ALG-1 and ALG-2 are responsible for these different lifespan phenotypes. However, the normal lifespan of animals with a non-functional ALG-2 with a C-terminal tag suggest the possibility that the increased lifespan seen in *alg-2(-)* animals is not the result of miRNA-mediated interactions. However, to further investigate this possibility additional experiments are required to determine whether ALG-2::mCherry is functional during adulthood.

Going forward, the application of iCLIP or other similar methods that can produce miRNA-target chimeric reads can help to identify the specific targets of ALG-1 and ALG-2 during aging. Additionally, CRISPR/Cas9 genome editing can be used to make custom mutations in ALG-1 or ALG-2. One potential mutation to investigate is mutating the PAZ domain of these AGOs to prevent them from

binding miRNAs. This could more specifically reveal whether the miRNA binding and targeting of these proteins is responsible for the opposing lifespan phenotypes. Other mutations to consider include switching the N-terminal domains, which is the least conserved domain between ALG-1 and ALG-2, and inhibiting the slicer activity. These experiments will help to elucidate the mechanisms by which ALG-1 and ALG-2 regulate lifespan in *C. elegans*.

5.5 Acknowledgements

The lifespan assays, microscopy, and immunoblots shown in this chapter were performed by A. Aalto, I. Nicastro, H. Jenq, and A. Pasquinelli. Some of these experiments were performed using reagents and strains that I developed for this project. Parts of the figures shown in this chapter are reproduced from a submitted manuscript with permission of the authors.

Chapter 5 contains material from the submitted manuscript "Opposing Roles of the MicroRNA Argonautes during Aging in *Caenorhabditis elegans*" by Aalto, A.P., Nicastro, I.A., Broughton, J.P., Chen, J.S., and Pasquinelli, A.E. Broughton, J.P. was a co-author on this manuscript.

Chapter 6

Conclusions

The identification of microRNA (miRNA) target sites is essential to understanding the function and regulatory potential of miRNAs. The first clues concerning how miRNAs find and recognize their targets came from the observation that there were complementary regions of the *lin-14* 3' untranslated region (3'UTR) to the lin-4 miRNA (Lee et al., 1993; Wightman et al., 1993). Similarly, studies of the let-7 miRNA also found that there was complementarity in the *lin-41* 3'UTR to let-7 (Pasquinelli et al., 2000; Reinhart et al., 2000; Slack et al., 2000). From these observations, other groups established the basics of miRNA targeting by identifying additional miRNA target sites in 3'UTRs and revealing the importance of target RNA complementarity to miRNA nucleotides 2-8, known as the 'seed' sequence (Bartel, 2009). These first target sites established that individual miRNAs have the potential to regulate hundreds of transcripts in multiple pathways.

However, as the number of confirmed miRNA targets increases, it has become apparent that miRNAs are capable of functionally regulating target RNAs through various base-pairing interactions that include imperfect seed matches (Chi et al., 2012). Furthermore, functional target sites have been noted outside of the 3'UTR. Interestingly, some groups have demonstrated functional noncoding RNA (ncRNA)

target sites, such as in primary-miRNA transcripts (Zisoulis et al., 2012) and in circular RNAs (circRNAs) (Hansen et al., 2013). Additional miRNA-ncRNA target interactions have been observed in chimeric data to small nucleolar RNAs (snoRNAs) and long intergenic noncoding RNAs (lincRNAs) (Grosswendt et al., 2014; Helwak et al., 2013; Moore et al., 2015). The increasing diversity of miRNA targets emphasizes the importance of precisely identifying target sites.

The research described here looks at miRNA function from two levels: the molecular and the phenotypic. At the molecular level, this work provides a detailed map of the miRNA targeting landscaping at the last larval stage of development in *C. elegans*. Using this evidence of miRNA-target interactions, I demonstrate the importance of the miRNA 3' end in providing specificity to miRNA targeting, and I confirm the relevancy of these interactions *in vivo*. At the phenotypic level, I generate a variety of novel tools to help validate the opposing roles of the *C. elegans* miRNA Argonautes (AGOs) in regulating lifespan, and determine the expression of these proteins during larval development and adulthood.

6.1 The importance of the miRNA 3' end in supporting targeting specificity

To identify miRNA target sites, I isolated AGO binding sites using individual-nucleotide resolution crosslinking immunoprecipitation and sequencing (iCLIP) (Broughton and Pasquinelli, 2013; König et al., 2010). This data provided a high-resolution map of sites that are bound by the primary AGO in the *C. elegans* miRNA pathway, ALG-1. Serendipitously, I discovered that miRNA-target chimeric reads are produced during the generation of ALG-1 iCLIP libraries. These chimeric, or hybrid, reads are composed of the mature miRNA sequence ligated to the target RNA.

Several groups have previously developed methods (known as CLASH, modified iPAR-CLIP, and CLEAR-CLIP) to generate miRNA-target chimeric reads and used these approaches to reveal insights into miRNA targeting (Grosswendt et al., 2014; Helwak et al., 2013; Moore et al., 2015). Although ALG-1 iCLIP is not specifically designed to generate chimeric reads, ALG-1 iCLIP libraries contain similar percentages (1.3-5.1%) of chimeric reads to other chimera-generating methods. By analyzing only high-confidence chimeric target sites that were reproducible in at least two biological replicates, I generated a detailed map of the miRNA-targeting landscape in *C. elegans*.

Using this dataset, I first confirmed that these target sites were associated with features of functional miRNA targeting. Notably, the ALG-1 iCLIP chimeric target sites included previously established miRNA targets, such as *lin-41*. Other features of canonical miRNA targeting that were present in the ALG-1 iCLIP chimeric sites included enrichment for seed complementarity to the cognate miRNAs at target sites, the majority of target sites occurring in messenger RNAs (mRNAs), and the identification of target sites that are regulated at the mRNA level in *alg-1(gk214)* mutant animals. Having established that ALG-1 iCLIP chimeric reads revealed features of known miRNA targets, I was then able to use this data to address outstanding questions in the field regarding how miRNAs interact with their target sites.

Structural analyses of AGO proteins bound to miRNAs have revealed that the seed sequence is positioned to facilitate binding to the target RNA, whereas the 3' end of the miRNA is sequestered within the protein and prevented from base pairing with the target RNA (Elkayam et al., 2012; Nakanishi et al., 2012; Schirle and MacRae, 2012). However, the structure of AGO bound to a miRNA and a partial target sequence has suggested that the N-terminal domain of human AGO may

undergo a conformational change after seed pairing. This conformational change could allow the nucleotides along the 3' end of miRNA to base pair with a target RNA (Schirle et al., 2014).

However, the extent of miRNA 3' end complementarity at target sites has remained unclear. To address this outstanding question, I first examined target sites for enrichment of specific nucleotide sequences. According to this analysis, in addition to the presence of seed complementarity, there is enrichment for complementarity to nucleotides 13-16 of the miRNA at target sites. This region of the miRNA is known as the 3' supplementary region. The 3' supplementary region has been suggested to compensate for weak or imperfect seed pairing (Grimson et al., 2007), and is the same region that was suggested to become exposed after a conformational change in AGO (Schirle et al., 2014).

To examine the extent of miRNA 3' end interactions, I paired each mRNA target site to its cognate miRNA. This analysis revealed seven distinct classes of interaction between miRNAs and their mRNA target sites. The majority of these classes featured seed pairing and various degrees of 3' end interactions, with one class featuring seed only pairing and one class featuring no-seed interactions. Interestingly, the majority of target sites had the potential to support base pairing to the miRNA 3' end, which suggests that these 3' end interactions may be an important feature of miRNA targeting. This conclusion is supported by work from two other groups that also showed that many miRNA target sites featured more extensive miRNA 3' end interactions than previously thought (Helwak et al., 2013; Moore et al., 2015). However, it remains unclear how the selection of an RNA hybridization algorithm affects the results from these types of analyses, or whether chimera-generating methods are biased for the capture of target sites with strong 3' end pairing.

From my analysis of miRNA-target RNA interactions, I also found that individual miRNAs have specific patterns of target interaction. For example, the majority of miR-60 sites were seed only, whereas miR-71 target sites were enriched for interactions that featured extensive pairing along the miRNA in addition to seed complementarity. I also found that different classes of target interaction were enriched between family members, such as let-7 and miR-48. Interestingly, one family of miRNA, the miR-51 family, was enriched in non-seed interactions between all family members.

The evidence that the majority of miRNA target sites could support 3' end interactions, suggested that these interactions might play a role in miRNA targeting. To address whether the seed sequence is sufficient for miRNA targeting, I examined target sites of miRNA families. A miRNA family is a group of highly related miRNAs that share the same seed sequence, nucleotides 2-8 from the miRNA 5' end, but often have divergent 3' end sequences. If the seed is the only important feature in miRNA targeting, then miRNA families should generally target the same target sites. The let-7 miRNA family in *C. elegans* offers a convenient genetic model for testing this hypothesis, since three members of the let-7 family of miRNAs are expressed at the same time, in the same tissues, and at relatively high levels (Kato et al., 2009; Martinez et al., 2008).

Surprisingly, I found that the let-7 family members, let-7, miR-48, and miR-241, largely target non-overlapping sets of target sites. These specific target sites were functional in regulating mRNA levels, and in some cases were specifically misregulated in let-7 family member mutants. I did not identify targets for miR-48 or miR-241 that were specifically misregulated by analyzing mRNA levels. The failure to detect specific misregulation of miR-48 or miR-241 specific targets may result from their altered expression in let-7 family member mutants. For example,

miR-48 is significantly up-regulated in *miR-241(n4316)* animals, and miR-241 is also significantly up-regulated in *miR-48(n4097)* animals. Furthermore, analysis of specific miRNA target sites revealed that only 12 let-7 specific targets, 10 miR-48 specific targets, and 0 miR-241 targets did not have binding by other miRNAs or ALG-1 iCLIP peaks. Additionally, miR-48 and miR-241 specific targets may be regulated on the level of translation and not mRNA destabilization. These confounding factors may mask the potential for specific misregulation of many let-7 family specific target sites.

To determine if base pairing could contribute to specific targeting by miRNA family members, I used RNAhybrid (Rehmsmeier et al., 2004) to calculate the minimum free energy (MFE) of hybridization for each let-7 family member to each set of specific target sites. For let-7 specific sites, the let-7 miRNA is more favorably paired than other members of the same family. This pattern continued for miR-48 and miR-241 specific target sites, where miR-48 or miR-241 was more favorably paired, respectively. Similarly, the miR-58 and miR-238 families followed similar trends. These analyses suggested that differences in miRNA targeting could be attributed to differences in the 3' end of the miRNA. However, it is also possible that these effects are due to more favorable chimera formation for miRNA target sites that exhibit 3' end complementarity, leading to the observed trend in pairing favorability for specific sites.

In many studies, functional miRNA targeting has been examined using over-expression of miRNAs or reporter constructs (Thomas et al., 2010). However, it is possible that overexpression of either the miRNA or the reporter may force the system to becoming functional or obscure effects that would otherwise be relevant at endogenous levels of either the miRNA or the target. To test the importance of the miRNA 3' end in miRNA targeting, I employed the specificity of the *lin-41* 3'UTR

for let-7. The *lin-41* 3'UTR contains two sites that are complementary to the let-7 miRNA (Reinhart et al., 2000; Vella et al., 2004), and is misregulated only in let-7 mutants and not in other let-7 family mutants. Furthermore, I only detect chimeras for let-7 at the two let-7 complement sites (LCS) 1 and 2 in *lin-41* 3'UTR and not for other let-7 family members. The specific misregulation of *lin-41* in let-7 mutants has been shown to result in the lethal vulval bursting phenotype associated with let-7 mutants (Ecsedi et al., 2015).

Recently, the CRISPR/Cas9 genome editing technique has allowed for the generation of specific mutations through homologous recombination in *C. elegans* (Dickinson et al., 2013, 2015; Friedland et al., 2013; Lo et al., 2013; Tzur et al., 2013; Waaijers et al., 2013). The programmable specificity of CRISPR/Cas9 targeting allows for custom mutations to be easily generated within the endogenous context. Using this tool, I replaced LCS1 and LCS2 in the *lin-41* 3'UTR with a miR-48 specific site from the *dot-1.1* 3'UTR. These miR-48 specific sites featured perfect seed complementarity to all members of the let-7 family. However, the miR-48 specific sites had strong 3' end complementarity to miR-48, but not to let-7. I used vulval bursting as a readout of *lin-41* misregulation. By crossing my mutant with miR-48 specific target sites, known as *lin-41(ap427),* to let-7 and miR-48 mutants, I demonstrated that the specificity of lin-41 regulation had been switched from let-7 to miR-48. This experiment revealed the functional consequences of 3' end complementarity at the endogenous levels of miRNAs and target RNAs.

Although the phenotypic analysis revealed the functional importance of the 3' end interactions in miRNA targeting specificity, the experiment did not allow me to directly address whether let-7 or miR-48 was binding to the mutant target sites in *lin-41(ap427).* Answering this question with current tools would have required the regeneration and analysis of additional ALG-1 iCLIP libraries in mutants carrying

the *lin-41(ap427)* allele. As an alternative, I developed a novel method, known as Chimera PCR (ChimP), to generate and detect miRNA-target chimeras. This method allows a targeted approach for showing that a miRNA interacts with a specific target site of interest without having to use radioactivity or have the bioinformatics expertise to analyze sequencing data for chimeric events. Using ChimP, I showed that in wild-type animals *lin-41* is specific for let-7, whereas in animals with the *lin-41(ap427)* allele, miR-48, and not let-7, binds to the miR-48 specific target sites in *lin-41*. This result complements the phenotypic data with evidence of direct binding between the miRNA and the target site of interest.

Overall, my work indicates the miRNA 3' end as supporting targeting specificity in addition to the seed sequence. This finding is contrary to early models of miRNA targeting, which suggested that base pairing beyond the seed does not enhance targeting specificity (Bartel, 2009). Interestingly, some recent reports using FRET experiments have suggested that complementarity beyond the seed does not increase the residency time of AGO on target sites (Chandradoss et al., 2015). However, these experiments largely relied on artificial target sites with perfect complementarity. Additionally, the interactions with the 3' end may not be important for the stability of the association and could have functional roles at the target site.

Current models of the mechanism of miRNA targeting suggest a series of steps that interrogate the target sites for complementarity. Initially only nucleotides 2-6 of the seed sequence are exposed by the AGO for binding (Elkayam et al., 2012; Nakanishi et al., 2012; Schirle and MacRae, 2012). Perfect pairing to this partial seed sequence results in a conformational changes in AGO that allows for full seed pairing (nucleotides 2-8) by releasing the steric inhibition of full seed pairing provided by helix-7 (Schirle et al., 2014). After the movement of helix-7, AGO may then be able to undergo additional conformational changes in the N-terminal

domain, which normally prevents sequences on the 3' end of the miRNA from forming base pair interactions. This hypothetical conformational change would then expose nucleotides in the miRNA 3' supplementary region (nucleotides 13-16) for further base pairing interactions (Schirle et al., 2014). My work supports the existence of these additional interactions by demonstrating that miRNA 3' end interactions can serve to support miRNA targeting specificity *in vivo*.

An outstanding challenge in the field of miRNA-RNA targeting is finding methods to increase the number of chimeric reads. Currently, the best methods, including ALG-1 iCLIP, produce at most 5% chimeric reads (Moore et al., 2015). In the future, it will be important to continue to refine methods for the production and isolation of chimeric reads. Several new CLIP-like methods have been developed that significantly increase the yield from CLIP experiments, and by adding intermolecular ligation steps to these methods additional chimeric reads may be recovered (Van Nostrand et al., 2016; Zarnegar et al., 2016). Additionally, some methods for the generation of chimeric reads for other types of RNA-RNA interactions have been developed using psoralen crosslinking (Lu et al., 2016; Sugimoto et al., 2015). Application of the chimera isolation strategies to identifying miRNA-target chimeras may be successful in purifying chimeric reads from libraries. Another approach may be to use poly(U) polymerase to add 4-thiouridine triphosphate (4sUTP) to the miRNA 3' end. 4sU can be reversibly tagged with biotin (Rädle et al., 2013), which would allow for chimeric read purification and prevent inhibition of reverse transcription by biotin.

Other outstanding questions from my work include: *how prevalent is miRNA-targeting of circRNAs?, why do some miRNAs have preferences for coding exon target sites?,* and *what is the minimum number of interactions with the miRNA 3' end to result in specificity in vivo*? In addition, the Chimera PCR (ChimP) method

can be further improved to increase the speed of the protocol and increase the specificity. By including the same improvements that increase the efficiency of enhanced CLIP (eCLIP) and infrared CLIP (irCLIP) (Van Nostrand et al., 2016; Zarnegar et al., 2016), the recovery of miRNA-target chimeras in ChimP can also be enhanced. Furthermore, by using the non-radioactive approaches to isolate protein-RNA complexes of the correct molecular weight that are used in eCLIP and irCLIP, the specificity of ChimP may be further improved while continuing to eliminate the steps that involve radioactivity.

My work on the contribution of the miRNA 3' end in targeting specificity has a variety of implications for the field. First, the computational identification of miRNA target sites often only considers seed complementarity, or assumes that all miRNA family members target the same sets of transcripts. By accounting for cases where extensive 3' end complementarity is present to a specific family member, these target prediction programs may be able to more accurately predict miRNA targets. Additionally, miRNA families are involved in various pathologies and may have differential expression between family members during disease states (Boyerinas et al., 2010), which may have implications on the design and effectiveness of miRNA-based therapeutics. Overall, my work reveals that, although the seed sequence is critical for miRNA targeting, the 3' end of the miRNA supports targeting specificity in the endogenous context.

6.2 The opposing roles of the miRNA Argonautes in regulating lifespan in *C. elegans*

My work has also included building genetic tools to further our understanding of the role of the miRNA pathway in aging. Work by other members of our lab has

shown that the two AGOs involved in the miRNA pathway in *C. elegans*, ALG-1 and ALG-2, have opposite lifespan phenotypes using *alg-1(-)* and *alg-2(-)* mutants. This finding was in contrast to previously published lifespans for the knockdown of these two proteins by RNAi (Kato et al., 2011; Samuelson et al., 2007). To understand the discrepancy between the genetic and the knockdown experiments, I built *alg-1* and *alg-2* specific RNAi constructs, generated several fluorescently-tagged ALG-1 and ALG-2 strains, and made a new *alg-2* null allele.

Considering the high-degree of conservation between ALG-1 and ALG-2 (Tops et al., 2006), one explanation for the different results between the knockout and the knockdown experiments is that the *alg-2* RNAi used in the knockdown experiments also targets *alg-1*. To test this hypothesis, I developed RNAi constructs for *alg-1* and *alg-2* that would target the unconserved 3'UTRs of these mRNAs, rather than the conserved coding exons that are targeted by the library RNAi constructs. Lifespans using *alg-1* and *alg-2* library RNAi replicated the decreased lifespan for both AGOs that had previously been published. However, the *alg-2* 3'UTR RNAi resulted in a lifespan extension, similar to that seen in the *alg-2(ok304*) mutant. This result suggests that ALG-2 has a role in negatively regulating lifespan in *C. elegans*. To further confirm that loss of *alg-2* increases lifespan, I used the CRISPR/Cas9 genome editing method to generate a new *alg-2* null allele. The new alg-2 allele I generated, *alg-2(ap426),* is also long lived compared to wild-type. Curiously, however, it is not as long lived as the *alg-2(ok304)* allele.

Although ALG-1 and ALG-2 have previously been shown to be expressed throughout larval development (Vasquez-Rifo et al., 2012), the expression of these proteins during aging had not been fully investigated. Surprising, we found that ALG-1 expression rapidly decreased after the first day of adulthood and remains off through day 11. To examine the protein levels of ALG-2, I N-terminally tagged

ALG-2 with mKate2::3xFLAG endogenously using CRISPR/Cas9 genome editing and homologous recombination (Dickinson et al., 2015). Using this strain, we were able to demonstrate that ALG-2 remains expressed throughout adulthood. Interestingly, confocal microscopy of a strain containing both 3xFLAG::GFP::ALG-1 and mKate2::3xFLAG::ALG-2 at adult day 1 has shown that while many neurons co-express ALG-1 and ALG-2, some are specific to either AGO.

Additional work revealed that the differing lifespan phenotypes of ALG-1 and ALG-2 both involve the insulin-signaling pathway. However, several important questions remain unaddressed. Primarily, it is unclear which direct targets of ALG-1 and ALG-2 are responsible for the opposing lifespan phenotypes. Despite their high-degree of homology, ALG-1 and ALG-2 may bind different sets of miRNAs, which may contribute to different patterns of targeting. Future work could apply ALG-1 iCLIP or ChimP to confirm the specific targets of these AGOs during aging. Considering that some neurons specifically express ALG-1 or ALG-2, it is also possible that the patterns of expression for these proteins contribute to the lifespan phenotypes. Another outstanding question is: *what mechanism is responsible for the down-regulation of ALG-1 during aging*? An experiment to replace the 3'UTR revealed that although ALG-1 with a different 3'UTR is expressed longer than ALG-1 with its native 3'UTR, ALG-1 expression still declines during the first few days of adulthood. This suggests significant transcriptional regulation of *alg-1* during adulthood. Follow up experiments may focus on switching the *alg-1* and *alg-2* promoters, or overexpressing each AGO. Overall, this work reveals how *alg-1* and *alg-2* act as antagonistic paralogs during aging, and provides a variety of tools that can be applied to a variety of future studies looking at the differences in ALG-1 and ALG-2.

References

Abbott, A. L., Alvarez-Saavedra, E., Miska, E. a., Lau, N. C., Bartel, D. P., Horvitz, H. R., and Ambros, V. (2005). The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans. *Dev. Cell*, 9(3):403–14.

Agarwal, V., Bell, G. W., Nam, J.-W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, 4:1–38.

Alvarez-Saavedra, E. and Horvitz, H. R. (2010). Many Families of C. elegans MicroRNAs Are Not Essential for Development or Viability. *Curr. Biol.*, 20(4):367–373.

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, 455(7209):64–71.

Bailey, T. L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 2:28–36.

Bandyopadhyay, S., Ghosh, D., Mitra, R., and Zhao, Z. (2015). MBSTAR: multiple instance learning for predicting specific functional binding sites in microRNA targets. *Sci. Rep.*, 5:8004.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2):215–33.

Bazzini, A. A., Lee, M. T., and Giraldez, A. J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science*, 336(6078):233–7.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of Arabidopis controlling leaf development. *Curr. Opin. Plant Biol.*, 1(3):188.

Bohnsack, M. T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, 10(2):185–91.

Bosson, A. D., Zamudio, J. R., and Sharp, P. A. (2014). Endogenous miRNA and Target Concentrations Determine Susceptibility to Potential ceRNA Competition. *Mol. Cell*, 56(3):347–359.

Boyerinas, B., Park, S.-M., Hau, A., Murmann, A. E., and Peter, M. E. (2010). The role of let-7 in cell differentiation and cancer. *Endocr. Relat. Cancer*, 17(1):F19–36.

Brennecke, J., Stark, A., Russell, R. B., and Cohen, S. M. (2005). Principles of microRNA-target recognition. *PLoS Biol.*, 3(3):0404–0418.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics*, 77(1):71–94.

Broughton, J. P. and Pasquinelli, A. E. (2013). Identifying Argonaute binding sites in Caenorhabditis elegans using iCLIP. *Methods*, 63(2):119–25.

Cai, X., Hagedorn, C. H., and Cullen, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, 10(12):1957–1966.

Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J., and Joo, C. (2015). A Dynamic Search Process Underlies MicroRNA Targeting. *Cell*, 162(1):96–107.

Chen, Y.-W., Song, S., Weng, R., Verma, P., Kugler, J.-M., Buescher, M., Rouam, S., and Cohen, S. M. (2014). Systematic Study of Drosophila MicroRNA Functions Using a Collection of Targeted Knockout Mutations. *Dev. Cell*, 31(6):784–800.

Chi, S. W., Hannon, G. J., and Darnell, R. B. (2012). An alternative mode of microRNA target recognition. *Nat. Struct. Mol. Biol.*, 19(3):321–7.

Chi, S. W., Zang, J. B., Mele, A., and Darnell, R. B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, 460(7254):479–86.

Dale, R. K., Pedersen, B. S., and Quinlan, A. R. (2011). Pybedtools: A flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics*, 27(24):3423–3424.

De Lencastre, A., Pincus, Z., Zhou, K., Kato, M., Lee, S. S., and Slack, F. J. (2010). MicroRNAs both promote and antagonize longevity in C. elegans. *Curr. Biol.*, 20(24):2159–2168.

Denzler, R., Agarwal, V., Stefano, J., Bartel, D. P., and Stoffel, M. (2014). Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol. Cell*, 54(5):766–76.

Dickinson, D. J., Pani, A. M., Heppert, J. K., Higgins, C. D., and Goldstein, B. (2015). Streamlined Genome Engineering with a Self-Excising Drug Selection Cassette. *Genetics*, 200(August):1–33.

Dickinson, D. J., Ward, J. D., Reiner, D. J., and Goldstein, B. (2013). Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. *Nat. Methods*, 10(10):1028–34.

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

Duerr, J. (2006). Immunohistochemistry.

Ecsedi, M., Rausch, M., and Großhans, H. (2015). The let-7 microRNA directs vulval development through a single target. *Dev. Cell*, 32(3):335–44.

Eichhorn, S. W., Guo, H., McGeary, S. E., Rodriguez-Mias, R. A., Shin, C., Baek, D., Hsu, S.-H., Ghoshal, K., Villén, J., and Bartel, D. P. (2014). mRNA Destabilization Is the Dominant Effect of Mammalian MicroRNAs by the Time Substantial Repression Ensues. *Mol. Cell*, 56(1):104–115.

Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. *Cell*, 150(1):100–10.

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA targets in Drosophila. *Genome Biol.*, 5(1):R1.

Fang, Z. and Rajewsky, N. (2011). The Impact of miRNA Target Sites in Coding Sequences and in 3âĂšUTRs. *PLoS One*, 6(3):e18067.

Farboud, B. and Meyer, B. J. (2015). Dramatic Enhancement of Genome Editing by CRISPR/Cas9 Through Improved Guide RNA Design. *Genetics*, 199(April):959–971.

Finnegan, E. F. and Pasquinelli, A. E. (2012). MicroRNA biogenesis: regulating the regulators. *Crit. Rev. Biochem. Mol. Biol.*, 48(1):51–68.

Friedland, A. E., Tzur, Y. B., Esvelt, K. M., Colaiácovo, M. P., Church, G. M., and Calarco, J. a. (2013). Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nat. Methods*, 10(8):741–3.

Friedman, R. C., Farh, K. K.-H., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, 19(1):92–105.

Frøkjaer-Jensen, C., Davis, M. W., Hollopeter, G., Taylor, J., Harris, T. W., Nix, P., Lofgren, R., Prestgard-Duke, M., Bastiani, M., Moerman, D. G., and Jorgensen, E. M. (2010). Targeted gene deletions in C. elegans using transposon excision. *Nat. Methods*, 7(6):451–3.

Grad, Y., Aach, J., Hayes, G. D., Reinhart, B. J., Church, G. M., Ruvkun, G., and Kim, J. (2003). Computational and experimental identification of C. elegans microRNAs. *Mol. Cell*, 11(5):1253–63.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U. S. A.*, 106(24):9613–8.

Grimson, A., Farh, K. K.-H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, 27(1):91–105.

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., and Mello, C. C. (2001). Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control C. elegans Developmental Timing. *Cell*, 106(1):23–34.

Grosshans, H., Johnson, T., Reinert, K. L., Gerstein, M., and Slack, F. J. (2005). The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in C. elegans. *Dev. Cell*, 8(3):321–30.

Grosswendt, S., Filipchyk, A., Manzano, M., Klironomos, F., Schilling, M., Herzog, M., Gottwein, E., and Rajewsky, N. (2014). Unambiguous Identification of miRNA:Target Site Interactions by Different Types of Ligation Reactions. *Mol. Cell*, 54(6):1042–1054.

Gu, S., Jin, L., Zhang, F., Sarnow, P., and Kay, M. a. (2009). Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat. Struct. Mol. Biol.*, 16(2):144–50.

Guo, H., Ingolia, N. T., Weissman, J. S., and Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308):835–840.

Ha, M. and Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.*, 15(8):509–524.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., and Tuschl, T. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 141(1):129–41.

Hafner, M., Lianoglou, S., Tuschl, T., and Betel, D. (2012). Genome-wide identification of miRNA targets by PAR-CLIP. *Methods*.

Hammell, M., Long, D., Zhang, L., Lee, A., Carmack, C. S., Han, M., Ding, Y., and Ambros, V. (2008). mirWIP: microRNA target prediction based on microRNA-containing ribonucleoproteinâĂŞenriched transcripts. *Nat. . . .*, 5(9).

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature*, pages 1–7.

Hausser, J., Syed, A. P., Bilen, B., and Zavolan, M. (2013). Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res.*, 23(4):604–15.

Hausser, J. and Zavolan, M. (2014). Identification and consequences of miRNA-target interactions - beyond repression of gene expression. *Nat. Rev. Genet.*, 15(September).

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 153(3):654–65.

Hipfner, D. R., Weigmann, K., and Cohen, S. M. (2002). The bantam gene regulates Drosophila growth. *Genetics*, 161(4):1527–1537.

Hristova, M., Birse, D., Hong, Y., and Ambros, V. (2005). The Caenorhabditis elegans heterochronic regulator LIN-14 is a novel transcription factor that controls the developmental timing of transcription from the insulin/insulin-like growth factor gene ins-33 by direct DNA binding. *Mol. Cell. Biol.*, 25(24):11059–72.

Hung, T., Pratt, G., Sundararaman, B., Townsend, M. J., Chaivorapol, C., Bhangale, T., Graham, R. R., Ortmann, W., Criswell, L. A., Yeo, G., and Behrens, T. W. (2015). The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science*, 350(6259):455–459.

Hunter, S. E., Finnegan, E. F., Zisoulis, D. G., Lovci, M. T., Melnik-Martinez, K. V., Yeo, G. W., and Pasquinelli, A. E. (2013). Functional Genomic Analysis of the let-7 Regulatory Network in Caenorhabditis elegans. *PLoS Genet.*, 9(3):e1003353.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H. T., Orejuela, M. R., Piechotta, M., Levanon, E. Y., Landthaler, M., Dieterich, C., and Rajewsky, N. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.*, 10(2):170–7.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816–21.

Jinek, M. and Doudna, J. a. (2009). A three-dimensional view of the molecular machinery of RNA interference. *Nature*, 457(January):405–412.

Jo, M. H., Shin, S., Jung, S.-R., Kim, E., Song, J.-J., and Hohng, S. (2015). Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs. *Mol. Cell*, 59(1):117–24.

Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K. L., Brown, D., and Slack, F. J. (2005). RAS is regulated by the let-7 microRNA family. *Cell*, 120(5):635–47.

Jonas, S. and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.*, 16(7):421–433.

Joung, J. K. and Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*, 14(1):49–55.

Jovanovic, M., Reiter, L., Picotti, P., Lange, V., Bogan, E., Hurschler, B. a., Blenkiron, C., Lehrbach, N. J., Ding, X. C., Weiss, M., Schrimpf, S. P., Miska, E. a., Grosshans, H., Aebersold, R., and Hengartner, M. O. (2010). A quantitative targeted proteomics approach to validate predicted microRNA targets in C. elegans. *Nat. Methods*, 7(10):837–42.

Jungkamp, A.-C., Stoeckius, M., Mecenas, D., Grün, D., Mastrobuoni, G., Kempa, S., and Rajewsky, N. (2011). In vivo and transcriptome-wide identification of RNA binding protein target sites. *Mol. Cell*, 44(5):828–40.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D. P., Zipperlen, P., and Ahringer, J. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, 421(6920):231–7.

Kanamoto, T., Terada, K., Yoshikawa, H., and Furukawa, T. (2006). Cloning and regulation of the vertebrate homologue of lin-41 that functions as a heterochronic gene in Caenorhabditis elegans. *Dev. Dyn.*, 235(4):1142–1149.

Kato, M., Chen, X., Inukai, S., Zhao, H., and Slack, F. J. (2011). Age-associated changes in expression of small , noncoding RNAs , including microRNAs , in C . elegans. *Rna*, 17:1804–1820.

Kato, M., De Lencastre, A., Pincus, Z., and Slack, F. J. (2009). Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during Caenorhabditis elegans development. *Genome Biol.*, 10(5):R54.

Khorshid, M., Hausser, J., Zavolan, M., and van Nimwegen, E. (2013). A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets. *Nat. Methods*, 10(3):253–5.

Kim, H., Ishidate, T., Ghanta, K. S., Seth, M., Conte, D., Shirayama, M., and Mello, C. C. (2014). A Co-CRISPR Strategy for Efficient Genome Editing in Caenorhabditis elegans. *Genetics*.

Kim, Y.-K., Kim, B., and Kim, V. N. (2016). Re-evaluation of the roles of DROSHA, Export in 5, and DICER in microRNA biogenesis. *Proc. Natl. Acad. Sci. U. S. A.*, 113(13):E1881–9.

Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat. Methods*, 8(7):559–64.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D. J., Luscombe, N. M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.*, 17(7):909–15.

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D. J., Luscombe, N. M., and Ule, J. (2011). iCLIP - Transcriptome-wide Mapping of Protein-RNA Interactions with Individual Nucleotide Resolution. *J. Vis. Exp.*, (50):7–13.

Kozomara, A. and Griffiths-Jones, S. (2011). MiRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, 39(SUPPL. 1):152–157.

Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., MacMenamin, P., da Piedade, I., Gunsalus, K. C., Stoffel, M., and Rajewsky, N. (2005). Combinatorial microRNA target predictions. *Nat. Genet.*, 37(5):495–500.

Kudla, G., Granneman, S., Hahn, D., Beggs, J. D., and Tollervey, D. (2011). Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc. Natl. Acad. Sci. U. S. A.*, 108(24):10010–5.

Kwon, S. C., Nguyen, T. A., Choi, Y. G., Jo, M. H., Hohng, S., Kim, V. N., and Woo, J. S. (2016). Structure of Human DROSHA. *Cell*, 164(1-2):81–90.

Lee, R. C. and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. *Science*, 294(5543):862–4.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–54.

Lehrbach, N. J., Castro, C., Murfitt, K. J., Abreu-Goodger, C., Griffin, J. L., and Miska, E. a. (2012). Post-developmental microRNA expression is required for normal physiology, and regulates aging in parallel to insulin/IGF-1 signaling in C. elegans. *Rna*, pages 2220–2235.

Leung, A. K. L. and Sharp, P. a. (2010). MicroRNA Functions in Stress Responses. *Mol. Cell*, 40(2):205–215.

Leung, A. K. L., Young, A. G., Bhutkar, A., Zheng, G. X., Bosson, A. D., Nielsen, C. B., and Sharp, P. a. (2011). Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat. Struct. Mol. Biol.*, 18(2):237–44.

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1):15–20.

Lewis, B. P., Shih, I.-h., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, 115(7):787–98.

Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B., and Bartel, D. P. (2003). The microRNAs of Caenorhabditis elegans. *Genes Dev.*, 17(8):991–1008.

Lin, S. Y., Johnson, S. M., Abraham, M., Vella, M. C., Pasquinelli, A., Gamberi, C., Gottlieb, E., and Slack, F. J. (2003). The C. elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable MicroRNA target. *Dev. Cell*, 4:639–650.

Lo, T.-W., Pickle, C. S., Lin, S., Ralston, E. J., Gurling, M., Schartner, C. M., Bian, Q., Doudna, J. a., and Meyer, B. J. (2013). Precise and Heritable Genome Editing in Evolutionarily Diverse Nematodes Using TALENs and CRISPR/Cas9 to Engineer Insertions and Deletions. *Genetics*, 195(October):331–348.

Lovci, M. T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T. Y., Stark, T. J., Gehman, L. T., Hoon, S., Massirer, K. B., Pratt, G. a., Black, D. L., Gray, J. W., Conboy, J. G., and Yeo, G. W. (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.*, 20(12):1434–42.

Lozzio, C. B. and Wigler, P. W. (1971). Cytotoxic effects of thiopyrimidines. *J. Cell. Physiol.*, 78(1):25–32.

Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., Robinson, J. T., Davidovich, C., Gooding, A. R., Goodrich, K. J., Mattick, J. S., Mesirov, J. P., Cech, T. R., and Chang, H. Y. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell*, 165(5):1267–79.

Marraffini, L. A. and Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.*, 11(3):181–90.

Martinez, N. J., Ow, M. C., Reece-Hoyes, J. S., Barrasa, M. I., Ambros, V. R., and Walhout, A. J. M. (2008). Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity. *Genome Res.*, 18(12):2005–15.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.*, 14(7):447–59.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., le Noble, F., and Rajewsky, N. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 495(7441):333–338.

Min, H. and Yoon, S. (2010). Got target? Computational methods for microRNA target prediction and their extension. *Exp. Mol. Med.*, 42(4):233–244.

Miska, E. a., Alvarez-Saavedra, E., Abbott, A. L., Lau, N. C., Hellman, A. B., McGonagle, S. M., Bartel, D. P., Ambros, V. R., and Horvitz, H. R. (2007). Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. *PLoS Genet.*, 3(12):e215.

Mondol, V., Ahn, B. C., and Pasquinelli, A. E. (2015). Splicing remodels the let-7 primary microRNA to facilitate Drosha processing in Caenorhabditis elegans. *RNA*, 21(8):1396–403.

Moore, M. J., Scheel, T. K. H., Luna, J. M., Park, C. Y., Fak, J. J., Nishiuchi, E., Rice, C. M., and Darnell, R. B. (2015). miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nat. Commun.*, 6(May):8864.

Moss, E. G., Lee, R. C., and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell*, 88(5):637–46.

Mukherji, S., Ebert, M. S., Zheng, G. X. Y., Tsang, J. S., Sharp, P. A., and van Oudenaarden, A. (2011). MicroRNAs can generate thresholds in target gene expression. *Nat. Genet.*, 43(9):854–9.

Nakanishi, K., Weinberg, D. E., Bartel, D. P., and Patel, D. J. (2012). Structure of yeast Argonaute with guide RNA. *Nature*, 486(7403):368–374.

Nehammer, C., Podolska, A., Mackowiak, S. D., Kagias, K., and Pocock, R. (2015). Specific microRNAs regulate heat stress responses in Caenorhabditis elegans. *Sci. Rep.*, 5:8866.

Newman, M. a., Thomson, J. M., and Hammond, S. M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA*, 14(8):1539–1549.

Nguyen, T. A., Jo, M. H., Choi, Y.-G., Park, J., Kwon, S. C., Hohng, S., Kim, V. N., and Woo, J.-S. (2015). Functional Anatomy of the Human Microprocessor. *Cell*, 161(6):1374–87.

Nielsen, C. B., Shomron, N., Sandberg, R., Hornstein, E., Kitzman, J., and Burge, C. B. (2007). Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, 13(11):1894–910.

Nour-Eldin, H. H., Geu-Flores, F., and Halkier, B. A. (2010). USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories. *Methods Mol. Biol.*, 643:185–200.

Pasquinelli, A. E. (2012). MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.*, 13(4):271–82.

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Müller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808):86–9.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–2.

Rädle, B., Rutkowski, A. J., Ruzsics, Z., Friedel, C. C., Koszinowski, U. H., and Dölken, L. (2013). Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. *J. Vis. Exp.*, (78):1–11.

Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA*, 10(2003):1507–1517.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403(6772):901–906.

Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E. E., Nitsch, R., and Wulczyn, F. G. (2008). A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat. Cell Biol.*, 10(8):987–93.

Saldanha, A. J. (2004). Java Treeview - Extensible visualization of microarray data. *Bioinformatics*, 20(17):3246–3248.

Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D., and Serebrov, V. (2015). Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell*, 162(1):84–95.

Samuelson, A. V., Klimczak, R. R., Thompson, D. B., Carr, C. E., and Ruvkun, G. (2007). Identification of Caenorhabditis elegans genes regulating longevity using enhanced RNAi-sensitive strains. *Cold Spring Harb. Symp. Quant. Biol.*, 72:489–97.

Sayed, D. and Abdellatif, M. (2011). MicroRNAs in development and disease. *Physiol. Rev.*, 91(3):827–87.

Schirle, N. T. and MacRae, I. J. (2012). The crystal structure of human Argonaute2. *Science*, 336(6084):1037–40.

Schirle, N. T., Sheu-Gruttadauria, J., Chandradoss, S. D., Joo, C., and MacRae, I. J. (2015). Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. *Elife*, 4(September):1–16.

Schirle, N. T., Sheu-Gruttadauria, J., and MacRae, I. J. (2014). Structural basis for microRNA targeting. *Science (80-.).*, 346:608–613.

Schneider, C. a., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, 9(7):671–675.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209):58–63.

Slack, F. J., Basson, M., Liu, Z., Ambros, V., Horvitz, H. R., and Ruvkun, G. (2000). The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol. Cell*, 5(4):659–69.

Song, J.-J., Smith, S. K., Hannon, G. J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, 305(5689):1434–7.

Stark, A., Brennecke, J., Russell, R. B., and Cohen, S. M. (2003). Identification of Drosophila microRNA targets. *PLoS Biol.*, 1(3).

Stark, T. J., Arnold, J. D., Spector, D. H., and Yeo, G. W. (2012). High-resolution profiling and analysis of viral and host small RNAs during human cytomegalovirus infection. *J. Virol.*, 86(1):226–35.

Subasic, D., Brümmer, A., Wu, Y., Pinto, S. M., Imig, J., Keller, M., Jovanovic, M., Lightfoot, H. L., Nasso, S., Goetze, S., Brunner, E., Hall, J., Aebersold, R., Zavolan, M., and Hengartner, M. O. (2015). Cooperative target mRNA destabilization and translation inhibition by miR-58 microRNA family in C. elegans. *Genome Res.*, pages 1–12.

Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H., and Bartel, D. P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature*, 508(7494):66–71.

Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.*, 13(8):R67.

Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., Luscombe, N. M., and Ule, J. (2015). hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature*, 519(7544):491–494.

Tay, Y., Rinn, J., and Pandolfi, P. P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature*, 505(7483):344–52.

Thomas, M., Lieberman, J., and Lal, A. (2010). Desperately seeking microRNA targets. *Nat. Struct. Mol. Biol.*, 17(10):1169–74.

Tops, B. B. J., Plasterk, R. H. A., and Ketting, R. F. (2006). The Caenorhabditis elegans Argonautes ALG-1 and ALG-2: almost identical yet different. *Cold Spring Harb. Symp. Quant. Biol.*, 71(0):189–94.

Travis, A. J., Moody, J., Helwak, A., Tollervey, D., and Kudla, G. (2014). Hyb: a bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. *Methods*, 65(3):263–73.

Tzur, Y., Friedland, A., Nadarajan, S., Church, G. M., Calarco, J. a., and Colaiácovo, M. P. (2013). Heritable custom genomic modifications in C. elegans via a CRISPR-Cas9 system. *. . .*, pages 2–9.

Urlaub, H., Hartmuth, K., and Lührmann, R. (2002). A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods*, 26(2):170–81.

Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., Blue, S. M., Nguyen, T. B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., and Yeo, G. W. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods*, (November 2015):1–9.

Van Wynsberghe, P. M., Chan, S.-P., Slack, F. J., and Pasquinelli, A. E. (2011). *Analysis of microRNA expression and function.*, volume 106. Elsevier Inc., second edi edition.

Vasquez-Rifo, A., Jannot, G., Armisen, J., Labouesse, M., Bukhari, S. I. A., Rondeau, E. L., Miska, E. a., and Simard, M. J. (2012). Developmental Characterization of the MicroRNA-Specific C. elegans Argonautes alg-1 and alg-2. *PLoS One*, 7(3):e33750.

Vella, M. C., Choi, E.-y., Lin, S.-y., Reinert, K., and Slack, F. J. (2004). The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. *Genes Dev.*, 18(2):132–7.

Vinther, J., Hedegaard, M. M., Gardner, P. P., Andersen, J. S., and Arctander, P. (2006). Identification of miRNA targets with stable isotope labeling by amino acids in cell culture. *Nucleic Acids Res.*, 34(16):2–7.

Viswanathan, S. R., Daley, G. Q., and Gregory, R. I. (2008). Selective blockade of microRNA processing by Lin28. *Science*, 320(5872):97–100.

Waaijers, S. and Boxem, M. (2014). Engineering the Caenorhabditis elegans genome with CRISPR/Cas9. *Methods*.

Waaijers, S., Portegijs, V., Kerver, J., Lemmens, B. B. L. G., Tijsterman, M., van den Heuvel, S., and Boxem, M. (2013). CRISPR/Cas9-targeted mutagenesis in Caenorhabditis elegans. *Genetics*, 195(3):1187–91.

Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, N. M., Rot, G., Zupan, B., Curk, T., and Ule, J. (2010). iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol.*, 8(10):e1000530.

Wee, L. M., Flores-Jasso, C. F., Salomon, W. E., and Zamore, P. D. (2012). Argonaute divides Its RNA guide into domains with distinct functions and RNA-binding properties. *Cell*, 151(5):1055–1067.

Westholm, J. O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., Celniker, S. E., Graveley, B. R., and Lai, E. C. (2014). Genome-wide Analysis of Drosophila Circular RNAs Reveals Their Structural and Sequence Properties and Age-Dependent Neural Accumulation. *Cell Rep.*, 9(5):1966–1981.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, 75(5):855–62.

Worringer, K. A., Rand, T. A., Hayashi, Y., Sami, S., Takahashi, K., Tanabe, K., Narita, M., Srivastava, D., and Yamanaka, S. (2014). The let-7/LIN-41 pathway regulates reprogramming to human induced pluripotent stem cells by controlling expression of prodifferentiation genes. *Cell Stem Cell*, 14(1):40–52.

Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.*, 17(24):3011–6.

Youngman, E. M. and Claycomb, J. M. (2014). From early lessons to new frontiers: The worm as a treasure trove of small RNA biology. *Front. Genet.*, 5(NOV):1–13.

Zarnegar, B. J., Flynn, R. A., Shen, Y., Do, B. T., Chang, H. Y., and Khavari, P. A. (2016). irCLIP platform for efficient characterization of proteinâĂŞRNA interactions. *Nat. Methods*, 13(6):489–492.

Zhang, C. and Darnell, R. B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.*, 29(7):607–14.

Zhang, H., Artiles, K. L., and Fire, A. Z. (2015). Functional relevance of "seed" and "non-seed" sequences in microRNA-mediated promotion of C. elegans developmental progression. *RNA*, pages 1–13.

Zhang, L., Ding, L., Cheung, T. H., Dong, M. Q., Chen, J., Sewell, A. K., Liu, X., Yates, J. R., and Han, M. (2007). Systematic Identification of C.??elegans miRISC Proteins, miRNAs, and mRNA Targets by Their Interactions with GW182 Proteins AIN-1 and AIN-2. *Mol. Cell*, 28(4):598–613.

Zhang, Y., Chen, D., Smith, M. A., Zhang, B., and Pan, X. (2012). Selection of reliable reference genes in caenorhabditis elegans for analysis of nanotoxicity. *PLoS One*, 7(3).

Zinovyeva, A. Y., Veksler-Lublinsky, I., Vashisht, A. a., Wohlschlegel, J. a., and Ambros, V. R. (2015). Caenorhabditis elegans ALG-1 antimorphic mutations uncover functions for Argonaute in microRNA guide strand selection and passenger strand disposal. *Proc. Natl. Acad. Sci.*, 112(38):E5271–E5280.

Zisoulis, D. G., Kai, Z. S., Chang, R. K., and Pasquinelli, A. E. (2012). Autoregulation of microRNA biogenesis by let-7 and Argonaute. *Nature*, 1.

Zisoulis, D. G., Lovci, M. T., Wilbert, M. L., Hutt, K. R., Liang, T. Y., Pasquinelli, A. E., and Yeo, G. W. (2010). Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. *Nat. Struct. Mol. Biol.*, 17(2):173–9.