THE UNIVERSITY OF CHICAGO

MECHANISMS OF SMALL RNA-MEDIATED GENOME SURVEILLANCE THAT DISTINGUISH SELF FROM NON-SELF NUCLEIC ACIDS IN *CAENORHABDITIS ELEGANS*

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON GENETICS, GENOMICS, AND SYSTEMS BIOLOGY

BY

JORDAN SCOT BROWN

CHICAGO, ILLINOIS

MARCH 2023

Don't Panic.

— Douglas Adams, *The Hitchhiker's Guide to the Galaxy*

TABLE OF CONTENTS

LIST OF FIGURES ..... **ix**

LIST OF TABLES ..... **xiii**

LIST OF SUPPLEMENTARY DATA FILES ..... **xiv**

ACKNOWLEDGEMENTS ..... **xv**

ABSTRACT ..... **xvi**

CHAPTER 1: INTRODUCTION **1**

1.1 GENE REGULATION BY SMALL RNA PATHWAYS 1

1.2 SMALL RNA PATHWAYS AND GERM GRANULES 5

1.3 THE DETERMINANTS OF GENE SILENCING 7

1.4 OPEN QUESTIONS AND THIS DISSERTATION 9

CHAPTER 2: IDENTIFICATION AND CHARACTERIZATION OF NOVEL PIRNA PATHWAY MEMBERS USING A REPORTER ASSAY **11**

2.1 ATTRIBUTIONS 11

2.2 ABSTRACT 11

2.3 INTRODUCTION 12

2.4 RESULTS 13

2.3.1 piRNA reporter detects loss of piRNA-dependent silencing factors *13*

2.3.2 RNAi screen identifies P granule factors, mRNA processing factors, and protein transport factors as piRNA silencing components *16*

2.3.3 Integrator Complex resolves piRNA precursor 3' ends *20*

2.3.4 Nuclear pore components maintain P granule factor accumulation and promote 22G RNA silencing *23*

2.3.5 mRNA splicing factor PRP-17 promotes proper mRNA surveillance 28

2.3.6 Importin components promote HRDE-1 silencing and localization 29

2.5 DISCUSSION 31

2.6 METHODS 36

2.6.1 C. elegans strains 36

2.6.2 RNA interference 36

2.6.3 CRISPR 37

2.6.4 Fluorescence Imaging 37

2.6.5 Quantitative real-time PCR 38

2.6.6 Small RNA sequencing 38

2.6.7 Cap-dependent RNA sequencing 40

2.7 SUPPLEMENTARY INFORMATION 41

CHAPTER 3: THE ROLE OF PERINUCLEAR GERM GRANULES IN PIRNA-MEDIATED TRANSCRIPTOME SURVEILLANCE 55

3.1 ATTRIBUTIONS 55

3.2 ABSTRACT 55

3.3 INTRODUCTION 56

3.4 RESULTS 58

3.4.1 GLH/VASA promotes the condensation of piRNA pathway factors 58

3.4.2 GLH/VASA helicase mutants exhibit defects in piRNA silencing 61

3.4.3 P granules promote piRNA silencing 65

3.4.4 Perinuclear P granules promote initiation of piRNA silencing 67

3.4.5 Perinuclear P granules promote silencing of piRNA targets 68

3.4.6 Perinuclear P granules prevent silencing of self genes *73*

3.4.7 PRG-1 is required for the aberrant self silencing *76*

3.5 DISCUSSION 79

3.6 METHODS 82

3.6.1 Caenorhabditis elegans strains *82*

3.6.2 Fluorescence microscopy and image processing *82*

3.6.3 piRNA initiation assay *83*

3.6.4 RNA isolation and quantitative real-time PCR *83*

3.6.5 Small RNA sequencing *84*

3.6.6 RNA immunoprecipitation sequencing (RIP-seq) *85*

3.6.7 Chemical cross-linking and co-immunoprecipitation of GLH-1 ..... *86*

3.6.8 Mass-spectrometry analysis *86*

3.6.9 RNA smFISH and PRG-1 immunohistochemistry *88*

3.6.10 Sequencing data analysis *91*

3.6.11 Data availability *93*

3.7 ACKNOWLEDGEMENTS 94

3.8 SUPPLEMENTARY INFORMATION 95

CHAPTER 4: THE DETERMINANTS OF GENE SILENCING AND LICENSING IN THE *C. ELEGANS* GERMLINE 106

4.1 ATTRIBUTIONS 106

4.2 ABSTRACT 106

4.3 INTRODUCTION 107

4.4 RESULTS 111

4.4.1 piRNAs exhibit a binding preference to the coding regions of target mRNAs in C. elegans *111*

4.4.2 piRNAs preferentially trigger the production of secondary small RNAs at mRNA coding regions *117*

4.4.3 CSR-1 inhibits piRNA binding and WAGO 22G-RNA amplification with distinct modes of action *120*

4.4.4 CSR-1 is not responsible for piRNAs' CDS-binding preference *125*

4.5 DISCUSSION 126

4.6 METHODS 129

4.6.1 Analyses of in vivo miRNA and piRNA binding sites *129*

4.6.2 Prediction of piRNA targeting sites *130*

4.6.3 Measurements of miRNA, piRNA binding sites and 22G-RNA levels at different regions of mRNAs *131*

4.6.4 Metagene analyses *132*

4.6.5 mRNA-seq analyses *132*

4.6.6 Ribo-seq analyses *132*

4.6.7 Data availability *132*

4.7 ACKNOWLEDGEMENTS 133

4.8 SUPPLEMENTARY INFORMATION 134

CHAPTER 5: TOOL DEVELOPMENT FOR *C. ELEGANS* SMALL RNA RESEARCHERS AND THE BROADER RNA COMMUNITY 140

5.1 ATTRIBUTIONS 140

5.2 INTRODUCTION 140

5.3 RESULTS 141

5.3.1 pirScan: a webserver to predict piRNA targeting sites and to avoid transgene silencing in C. elegans 141

5.3.2 piRTarBase: a database of piRNA targeting sites and their roles in gene regulation 151

5.3.3 CLASH Analyst: a webserver to identify in vivo RNA-RNA interactions from CLASH data 165

5.4 DISCUSSION 178

5.5 METHODS 181

5.5.1 pirScan webserver construction.....182

5.5.2 piRTarBase database construction.....182

5.5.3 CLASH Analyst webserver construction.....183

5.6 SUPPLEMENTARY INFORMATION.....184

CHAPTER 6: CONCLUSIONS 188

6.1 THE SIGNIFICANCE OF GERM GRANULES TO SMALL RNA PATHWAY FUNCTION 188

6.2 THE FEASIBILITY OF PREDICTING THE REGULATORY FATE OF GERMLINE TRANSCRIPTS.....191

6.3 TOOL AVAILABILITY AS A MEANS FOR INCLUSIVITY IN SCIENCE.....192

6.4 FUTURE DIRECTIONS 193

6.5 CONCLUDING REMARKS.....196

REFERENCES 197

LIST OF FIGURES

Figure 2.1: piRNA reporter detects loss of biogenesis and effector factors.....15

Figure 2.2: Loss of mRNA processing factors, snRNA processing factors, protein import factors, and P granules components trigger piRNA reporter activation 19

Figure 2.3: The Integrator Complex promotes piRNA biogenesis.....22

Figure 2.4: Nuclear pore component NPP-7 and the nucleolus promote piRNA dependent silencing 27

Figure 2.5: PRP-17 promotes piRNA dependent silencing and CSR-1 target pre-mRNA splicing.....29

Figure 2.6: IMA-3 promotes HRDE-1 nuclear import.....31

Figure 2.7: Multifaceted involvement of conserved and essential machinery work in concert with small RNA-mediated gene silencing in germ cells.....35

Figure 2.8: piRNA reporter screening strategy 41

Figure 2.9: Loss of piRNAs following *snpc-4*, *ints-1*, and *dic-1* knockdown.....42

Figure 2.10: NPP-1/6/7/9 and CSR-1 pathway disruption activates the sensitized piRNA reporter.....43

Figure 2.11: HRDE-1 does not share NRDE-3's nuclear localization signal 44

Figure 3.1: The GLH/VASA helicases GLH-1 and GLH-4 promote the localization of germ granule factors.....60

Figure 3.2: GLH/VASA helicases are required for piRNA-dependent silencing of a piRNA reporter.....64

Figure 3.3: Mutants with defective in either perinuclear or cytoplasmic P granules exhibit defects in gene silencing 67

Figure 3.4: GLH/VASA mutants exhibit defects in producing secondary WAGO-22G-RNAs.. 71
Figure 3.5: Many functional germline genes are silenced in mutants with defects in forming perinuclear PRG-1 granules 74
Figure 3.6: P granules promote piRNA targeting fidelity 82
Figure 3.7: Image analyses corresponding to Figure 3.1A-C 95
Figure 3.8: Image analyses corresponding to Figure 3.2B 97
Figure 3.9: Mutants with defects in PRG-1 perinuclear condensates exhibit defects in triggering piRNA silencing 99
Figure 3.10: GLH/VASA mutants exhibit defects in producing secondary WAGO-22G-RNAs (Corresponding to Figure 3.4) 100
Figure 3.11: Many functional germline genes are silenced in mutants with defects in forming perinuclear PRG-1 granules (Corresponding to Figure 3.5). 102
Figure 3.12: HRDE-1 / CSR-1 IP ratio and CSR-1 IP 3' end enrichment are both associated with enhanced HRDE-1 targeting of CSR-1 genes in VASA mutants 104
Figure 4.1: Transcriptome-wide analyses of piRNA binding sites 110
Figure 4.2: *In vivo* miRNA binding sites are enriched at the 3' UTR of mRNAs in *C. elegans* 113
Figure 4.3: *In vivo* piRNA binding sites are enriched in the coding region (CDS) of germline mRNAs in *C. elegans* 116
Figure 4.4: Secondary WAGO 22G-RNAs are preferentially produced in coding regions ...119
Figure 4.5: CSR-1 inhibits piRNA binding and WAGO 22G-RNA synthesis with distinct modes of action 124

Figure 4.6: A model depicting three distinct mechanisms that control piRNA binding and gene silencing on WAGO and CSR-1 targets 127
Figure 4.7: *In vivo* miRNA binding sites are enriched at the 3' UTR of mRNAs in *C. elegans,* related to Figure 4.2 134
Figure 4.8: piRNAs' CDS binding preference cannot be explained by the predicted piRNA targeting sites, related to Figure 4.3 135
Figure 4.9: The distribution of CSR-1 22G-RNAs, WAGO 22G-RNAs, and piRNA binding sites, related to Figure 4.5 137
Figure 4.10: The distribution of piRNA binding sites, related to Figure 4.5 139
Figure 5.1: An overview of pirScan's workflow 144
Figure 5.2: Avoiding only five piRNAs predicted to target GFP results in germline expression of a *gfp::cdk-1* transgene 147
Figure 5.3: Workflow of piRTarBase 154
Figure 5.4: 22G RNA and mRNA fold expression changes in germline expressed genes with piRNA targeting sites predicted using different methods 159
Figure 5.5: An overview of piRTarBase's user interface 163
Figure 5.6: A model showing the experimental framework underlying CLASH 168
Figure 5.7: A graphical depiction of CLASH Analyst's workflow 170
Figure 5.8: Comparison of RNA–RNA interactions identified from three available searching algorithms in CLASH Analyst 176
Figure 5.9: A step-by-step depiction of the CLASH Analyst workflow 184
Figure 5.10: The distribution of miRNA and piRNA binding sites in C. elegans 185
Figure 5.11: Examples of RNA-RNA interactions identified by CLASH Analyst 186

Figure 5.12: A comparison of RNA-RNA interactions identified by CLAN or Hyb 187

LIST OF TABLES

Table 2.1: Full list of candidates screened using piRNA reporter 45

Table 5.1: List of mutations in *gfp::cdk-1* to avoid targeting by 5 piRNAs in *gfp*$^{\Delta piRNA}$*::cdk-1*. 147

Table 5.2: Database statistics for piRTarBase 183

LIST OF SUPPLEMENTARY DATA FILES[1]

Data file 3.1: List of interacting proteins identified by MS with DTME-crosslinking using GLH-1.....105

Data file 3.2: List of smFISH probes.....105

Data file 4.1: piRNA binding site information for Figures 4.1B and 4.2A.....139

Data file 4.2: Coordinates used for CSR-1 occupancy windows.....139

[1] Note: Due to the large size of some tables, the tables have been provided in a supplementary file accompanying the dissertation. In these cases, the page number provided directs the reader to the table's caption.

ACKNOWLEDGEMENTS

There are many people who have helped me immeasurably over the course of my graduate studies at the University of Chicago. I am very grateful and indebted to my advisor, Heng-Chi Lee, Ph.D. for the incalculable hours he has dedicated to helping me develop as a scientist and the energy he has invested in the pursuit of my goals. Also, other members of the Lee Lab, particularly Wenjun Chen, Ph.D., Donglei Zhang, Ph.D., and Olivia Gaylord, have helped me scientifically and personally over the years. Without any one of them, this work would not have been possible. I would also like to thank the Cummings 8th floor community for making our workplace richer and more vibrant. I have benefitted tremendously from collaboration during my graduate studies. Our frequent collaborator, Wei-Sheng Wu, Ph.D., provided interesting scientific perspectives that were unique from my own and always valuable. The feedback and insight I have gained from the members of my thesis committee, Jon Staley, Ph.D., Alex Ruthenburg, Ph.D., and Jingyi Fei, Ph.D., has also been invaluable. I was lucky to have access to these three scientists whom I greatly admire. I'd also like to thank Allison Abbott, Ph.D., who allowed me to work in her lab as an undergraduate at Marquette and inspired me to pursue scientific research thereafter. I must thank my friends and family who have helped critically during my time as a graduate student. A highlight of my time at University of Chicago has been watching and discussing movies with Jimmy Elias and Meike Lobb-Rabe. My mom – Rhonda, my dad – Dale, and my brother – Jacob, have encouraged and believed in me unendingly. My grandma Joyce fostered an early sense of curiosity and excitement about science by taking my brother and I to the Discover Center Museum in Rockford throughout my childhood. Finally, I thank my partner Emily Hegel, whose constant love and support allowed me to continue past my failures and to celebrate my successes. I cannot possibly express my gratitude for her patience and grace.

ABSTRACT

Organisms rely on a stable genome in their germ cells to successfully establish cellularly competent offspring. However, the mobilization of many transposons and retroviruses called retrotransposons can interrupt important or essential genes in the host's genome, rendering the effected germ cells mutagenized. Such attacks are so common that retrotransposon-derived sequences make up nearly 40% of most mammalian genomes. To combat the possibly deleterious effects of these mobile elements, metazoans have evolved a robust genome defense system housed in their germ cells. This defense system is composed of thousands of endogenously encoded small non-coding RNAs called piRNAs. A PIWI Argonaute protein traffics piRNAs to mRNA targets that share sequence complementarity and helps regulate their expression.

In the first chapter of this dissertation, I review the current literature on small non-coding RNA pathways in general, and how those pathways function in the *C. elegans* germline in particular. I then discuss the connection between germline small RNA pathways and phase separated condensates enriched in animal germ cells called germ granules. Finally, I outline what is currently known about how particular mRNAs become targeted and silenced by small RNAs.

In the second chapter, I demonstrate that a reporter assay can successfully uncover novel components involved in piRNA mediated gene silencing. I discuss where in the pathway each factor likely functions, and further characterize how each factor likely plays a role in gene silencing. This chapter demonstrates that highly conserved and ancient cellular machinery that is essential for snRNA biogenesis, pre-mRNA splicing, nuclear protein import, and nuclear pore formation has evolved to also contribute to small RNA mediated germline defense.

In the third chapter, I discuss the role that the hallmark of animal germ cells, the germ granule, plays in the piRNA pathway. I show that, surprisingly, these highly conserved, phase

separated structures are largely dispensable for small RNA-mediated gene silencing. However, without germ granules, competing small RNA pathways lose robustness, allowing a subset of typically silenced genes to become expressed and a subset of typically expressed genes to become silenced. This mis-regulation likely contributes to the deleterious fertility phenotypes associated with germ granule loss described here and in previous work.

In the fourth chapter, I characterize signals that determine whether particular germline transcripts will become targeted and silenced by the piRNA pathway. The coding region of germline mRNAs rather than the untranslated regions are uniquely vulnerable to piRNA targeting. This finding suggests that the piRNA pathway has evolved to target regions of mRNAs which are less likely to escape detection due to genetic drift.

In the fifth chapter, I present a collaborative effort to make the suite of tools used to perform the complex analyses discussed in this dissertation available to the scientific community. Many analyses involved in parsing next generation sequencing data require advanced domain knowledge to replicate. We have made these tools available in a convenient graphical format for the community to use without specific knowledge of the underlying code required to use them. I hope that the availability of such tools will become a standard in academia to allow for greater transparency and to open barriers closed to many biologists who do not have the bioinformatic expertise necessary to analyze the data already available to the community.

In the final chapter, I discuss several important implications of this work for the field of small RNA biology. I address how the findings presented here influence our understanding of piRNA-mediated transcriptome regulation, and I suggest future work that could further clarify our understanding of how self and non-self is determined for germline mRNAs.

CHAPTER 1

INTRODUCTION

1.1 Gene regulation by small RNA pathways

To preserve the fidelity of development and reproduction, organisms must regulate and fine tune the various gene expression programs which orchestrate essential cellular events. RNA molecules perform many of these functions by transmitting messages from a cell's DNA code to become proteins as mRNA. This class of RNA molecule is referred to as coding RNA, as these molecules will eventually be translated into protein. Remarkably, less than 2% of our genomes are transcribed into coding RNA, while much of the remainder is transcribed into non-coding RNA [18,45]. Non-coding RNA, perform a myriad of other functions by for example forming the structural basis of the ribosome machine that manufactures proteins (tRNA), or by controlling the stability of other RNA molecules by guiding effector proteins to sites of regulation. This last category of gene expression control is performed by small non-coding RNAs including microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs), and small interfering RNAs (siRNAs). Small non-coding RNAs associate with a family of proteins called Argonautes, originally named for the mutant phenotype caused by loss of the first discovered Argonaute protein in *Arabidopsis thaliana* which results in leaves taking on the appearance of a small squid called *Argonauta argo* [19]. This RNA-protein complex, referred to as the RNA induced silencing complex (RISC), regulates gene expression by finding complementary or partially complementary sequences, typically mRNAs, using the bound single stranded small RNA molecule and either cleaving the targeted RNA directly or recruiting effector proteins to degrade the target [28]. The first discovered class of small non-coding RNA that regulates gene expression using this method, miRNAs, regulate the expression of many if not most endogenously expressed mRNAs [13,88]. miRNA driven regulation of gene

expression falls into the broad and highly diverse category of a mode of regulation called RNA interference (RNAi).

In addition to endogenously expressed mRNAs that perform functions for cells at specific times in specific places, there are also many deleterious nucleic acids that can become expressed throughout development, including repetitive elements and transposons. Organisms must defend against the invasion of viral and transposon derived sequences which can disrupt and destabilize the genome by integrating into DNA [114]. Indeed, throughout evolutionary history, repetitive elements have frequently moved about genomes, representing about 15% of the *Drosophila melanogaster* genome, 6% of the *Caenorhabditis elegans* genome, and up to half of human and mouse genomes [74,120].

Organisms have devised methods to defend against the expression of non-self genes that are evolutionarily ancient. In fact, the emergence of siRNA-based RNAi pathways likely predates the divergence of eukaryotes from prokaryotes [136]. siRNA-based RNAi differs from the miRNA pathway in several important ways. As alluded to already, miRNAs principally regulate the expression of endogenous, self genes. There are examples of siRNA-based RNAi pathways that also regulate endogenous gene expression, but siRNA-based mechanisms likely evolved to defend against exogenous, non-self nucleic acids. Additionally, siRNA-based mechanisms typically rely on an array of sequence-distinct small RNAs targeting particular RNA sequences to quickly and effectively shut off gene expression. This potent effect relies on signal amplification to become manifest. In many organisms, signal amplification comes from the use of targeted molecules as potentiators of the silencing signal. Signal amplification results in heritability of the RNAi effect, resulting from amplified siRNA molecules that can both destroy target RNAs post-

transcriptionally, as well as from the recruitment of nuclear factors to modify chromatin and silence targets transcriptionally [67].

Germ cells are particularly vulnerable to non-self elements and important to preserve, as they house the reproductive genome that will be passed on across generations. Genome stability of germ cell DNA is paramount for the survival of a species. As a result, organisms have evolved genome defense pathways specific to the germline to preserve genomic stability. One such small RNA-based germline pathway evolved by metazoans relies on endogenously encoded piRNAs which interact with the highly evolutionarily conserved Argonaute protein PIWI [21,40,53]. Highly diverse organisms that have lost PIWI due to mutation uniformly show fertility defects [38].

Caenorhabditis elegans is an ideal model system to study siRNA-based RNAi pathways. *C. elegans* are highly genetically tractable, allowing for simple and quick development of targeted mutations and generation of reporter strains. Additionally, *C. elegans* have a short generation time, which is a necessity in studying the highly transgenerational processes initiated by small RNA pathways. *C. elegans* are also hermaphroditic, which is an ideal feature for transgenerational studies that can often be hampered by introduced genetic or environmental variation from differences in parentage, confounding the interpretation of epigenetic phenomena. *C. elegans* also share the core conserved RNAi machinery mentioned above. Intriguingly, the piRNA system in *C. elegans*, which relies on the sole PIWI Argonaute in the organism PRG-1, can initiate a highly robust and heritable silencing signal against the expression non-self genes. Understanding components essential for the function of that system and the experimental complications that come from studying such a highly heritable phenomenon are the focus of the second chapter of this dissertation.

Not only does *C. elegans* possess the conserved RNAi machinery referenced so far, but it has also evolved even more diverse Argonaute proteins and effector pathways, making it a highly effective combatant against non-self sequences. Downstream of the exogenous RNAi and piRNA pathways, silencing signals are amplified though the production of secondary siRNAs. Two RNA-dependent RNA polymerases, EGO-1 and RRF-1, produce secondary siRNAs termed 22G-RNAs for their 22 nucleotide length and 5' guanosine bias [95,143]. EGO-1 and RRF-1 act partially redundantly to produce the 22G-RNA pools loaded onto myriad Argonaute proteins, particularly CSR-1 and WAGO Argonautes.

Production of 22G-RNAs triggered by a silencing signal, such as a PRG-1 targeting event, is dependent on the Mutator class of proteins, so called due to their necessity in preventing transposon activation and therefore mutagenesis [118]. One such Mutator protein, MUT-2 (also called RDE-3), has recently been shown to polymerize repetitive UG tails on the 3' ends of mRNA fragments cleaved by RNAi pathway [142]. These polyUG tailed fragments are sufficient to trigger 22G-RNA production through recruitment of RRF-1. Mutator-dependent 22G-RNAs are bound by a wide variety of silencing secondary Argonaute proteins which are evolutionarily expanded in *C. elegans* and thus termed worm-specific AGOs (WAGOs) [183]. These silencing 22G-RNA WAGO complexes can trigger gene silencing by further targeting mRNAs post-transcriptionally in the cytoplasm and recruiting degradation factors, while other WAGO complexes, namely those containing the WAGO Argonautes HRDE-1 or NRDE-3, can translocate into nuclei and recruit chromatin modifiers to transcriptionally silence targeted loci [8,23,141]. The production of 22G-RNAs upon silencing is so robust that when silencing is established by exogenous RNAi or by PRG-1, that silencing signal can be inherited for many generations even in the absence of the initial silencing trigger. One reason for this robust transgenerational silencing effect is due to the

production of tertiary 22G-RNAs, produced in response to WAGO targeting and further loaded onto WAGO Argonautes [8,169,181].

While EGO-1 and RRF-1 can help trigger and potentiate gene silencing by synthesizing Mutator-dependent 22G-RNAs, the majority of 22G-RNAs are actually produced from germline expressed genes specifically by EGO-1 and loaded onto the Argonaute CSR-1 [33]. Because CSR-1 bound 22G-RNAs mostly target mRNAs that are abundantly expressed in the germline, it was initially hypothesized to represent a licensing signal that preserves the expression of endogenous genes in the face of converging silencing small RNA pathways. Indeed, in the absence of CSR-1, many germline expressed mRNAs do become targeted by PRG-1 [137]. However, CSR-1 certainly has other roles in the germline that are at odds with its characterization solely as a licensing factor. For example, like many Argonaute proteins (WAGO family Argonautes excepted) CSR-1 contains an active slicer domain, and CSR-1 slicer activity acts to clear maternal mRNAs following embryogenesis [122]. Also, *csr-1* depletion has been shown to lead to an upregulation of sperm gene expression in adult hermaphrodites, suggesting that CSR-1 helps to repress the expression of these genes in the adult oogenic germline [25].

1.2 Small RNA pathways and germ granules

Remarkably, the robust small RNA pathway activity outlined above that takes place in the adult *C. elegans* germline largely occurs within perinuclear, phase-separated condensates called germ granules. Germ granules are found in all animal germ cells and are critical for fertility [20,46]. Germ granules have been proposed to function as specialized transcriptome surveillance structures, owing to several key features. Germ granule formation and localization closely follow transcriptional activity in *C. elegans* germ cells, where quiescent early embryo germ granules first form in the cytoplasm then as transcription resumes, germ granules become perinuclear [135,162].

Additionally, most mRNA passes through germ granules upon nuclear export in the adult germline [138]. Recently, it has been shown that recruitment of particular mRNAs to germ granules is sufficient to induce silencing [5]. And finally, as alluded to at the beginning of this section, important mRNA surveillance factors discussed so far such as PRG-1, WAGO-1, and CSR-1 Argonautes all co-localize in germ granules [15,33,61].

Adult germ granules are made up of at least four distinct sub-compartments. The earliest studied germ granule substructure, the P granule, sits closest to the nuclear periphery, interfacing directly with nuclear pores [119]. Intriguingly, both PRG-1 and CSR-1 are found in the P granule, suggesting that self and non-self signals compete in this structure. The Z granule and Mutator are two more recently described compartments that are both important for silencing [118,169]. The Mutator factors essential for silencing small RNA production discussed above are all housed in Mutator foci, suggesting that this structure represents a specialized amplification compartment to accumulate silencing signals. The Z granule is more associated with specifically transgenerational silencing. The most recently described sub-compartment is characterized by the presence of a Tudor domain protein called SIMR-1 and is therefore referred to as the SIMR focus [94]. SIMR foci play a role in small RNA silencing by bridging piRNA silencing with Mutator-dependent siRNA accumulation.

Interaction between factors that localize to these granules seems to be key for small RNA pathway function. For example, individual mutation of most Mutator proteins results in breakdown of Mutator foci and major defects in gene silencing [118]. Additionally, as discussed above, RRF-1 and EGO-1 both contribute to 22G-RNA populations in the germline. While the strictly silencing 22G-RNA manufacturer RRF-1 localizes to Mutator foci, EGO-1 seems to associate more with the P granule [160]. Further, EGO-1 has been shown to function even in the absence of a germ

granule environment to produce 22G-RNAs loaded onto CSR-1 in the cytoplasm [144]. These observations suggest that localization to germ granules and association with particular compartments of germ granules can be predictive of the function of particular small RNAs. The degree to which the germ granule itself helps to promote small RNA pathway function is the major focus of the third chapter of this dissertation.

1.3 The determinants of gene silencing

Small RNA pathways uniformly rely on base pairing between the small RNA and its target RNA to regulate gene expression. However, predicting the regulatory targets of small RNAs is not as simple as finding complementarity within possible target sequences. While some small RNA pathways rely on perfect base pairing between small RNA and target (i.e. exogenous RNAi and downstream 22G-RNAs), other small RNAs require comparatively little base pairing to successfully regulate a target. For example, miRNAs require several nucleotides of perfect matching in the so-called seed region, while complementarity within the majority of the small RNA sequence is largely supplementary [13]. piRNAs as well require only selective perfect base pairing, a characteristic that made piRNA targets notoriously difficult to predict as they perfectly complement exceedingly few transposon sequences [15,185].

In addition to partial or complete complementarity, there are other features of either small RNAs or their targets that determine successful regulatory events. The vast majority of miRNA targeting events occur in 3'UTRs of target mRNAs, demonstrating that the gene structure of a potential target plays a role in regulatory potential [13]. Additionally, it has recently been shown that the CSR-1 bound 22G-RNAs produced by EGO-1 in the absence of germ granules tend to act in the cytoplasm to cleave inefficiently translated mRNAs [144]. An extrapolation from this finding is that CSR-1 bound 22G-RNAs localized to germ granules specifically protect against

mRNA silencing, suggesting a functional divergence between CSR-1 complexes found in germ granules compared to those found in the cytoplasm. Also, changing the localization of an mRNA by tethering to P granule enriched PGL-1 proteins leads to small RNA-dependent post-transcriptionally gene silencing of the tethered mRNA, even when its sequence was not changed [5]. Both of these examples illustrate the importance of subcellular localization in determining the regulatory outcome of small RNA pathways on mRNAs. Another example of this same principle has already been referenced in the different localizations of the two 22G-RNA producing RdRPs EGO-1 and RRF-1. Although these two proteins are partially redundant in producing 22G-RNAs, RRF-1 which uniquely localizes to Mutator foci only produces silencing 22G-RNAs while EGO-1, more enriched in the P granule, is essential only for the production of CSR-1 bound 22G-RNAs. Therefore, not only subcellular localization but even germ granule sub-compartment localization can help to predict regulatory outcome.

Although the trends outlined above can help us predict the role of particular small RNAs in the germline, whether those trends are causal or not is more difficult to determine. For example, 22G-RNAs perfectly pair with their target mRNAs as a consequence of their being RdRP products. Mutator localization of RRF-1 correlates strongly with 22G-RNA production, but RRF-1 Mutator localization has only been disrupted by mutating other Mutator components that exist in the same pathway [118,160]. Similarly, the role of the germ granule itself in RNA homeostasis has proven difficult to determine as disruption of the germ granule has not been achieved without mutating necessary components, which conflates the functional impact of the granule itself with the function of its individual regulators. Therefore, the effort to understand what underlies small RNA mediated gene regulation is two-fold. The first step is to characterize trends that can help predict whether particular targeting events occur *in vivo* and lead to a regulatory consequence. The second step is

to determine whether the characteristics of targeting are causal in determining whether the targeting event will occur, or whether they are a consequence of some other feature of the pathway.

1.4 Open questions and this dissertation

Although much is known about small RNA-mediated transcriptome regulation in *C. elegans*, there are still key gaps in our knowledge that necessitate further understanding. In this work, I address four areas in small RNA biology.

First, I identify and characterize the novel role of several components in the piRNA pathway. Finding novel piRNA pathway components has proven difficult due to the robust and multifaceted silencing initiated by piRNAs in the germline. As discussed at length above, piRNA silencing can be conducted independently by common downstream factors following the initiation of silencing. This remarkable feature of the pathway renders reporter assays insufficient to exhaustively identify novel components, as loss of many pathway members often does not lead to an immediate phenotype. In Chapter 2, I describe a specialized reporter that is immediately sensitive to loss of piRNA pathway function. Using this reporter system, I show that the Integrator complex, nuclear pore and nucleolus components, pre-mRNA splicing factors, and protein import components are essential for piRNA pathway function.

Second, I address the role of germ granules in small RNA pathway function. In Chapter 3, I show that surprisingly germ granules are largely dispensable for small RNA silencing and only a small subset of mRNAs silenced by piRNAs rely on the presence of germ granules to maintain that silencing. However, another subset of typically expressed mRNAs become aberrantly targeted and silenced by piRNAs in the absence of germ granules. I show that rather than functioning as a silencing hub, the germ granule functions as an arbiter of self versus non-self in the adult germline,

helping to determine which mRNAs should be silenced and which should be protected from silencing.

Third, I report a novel feature that helps determine piRNA targeting. Analogous to the 3'UTR preference miRNAs show for target mRNAs, piRNAs prefer to bind to the coding regions (CDSs) of target mRNAs. This trend contradicts the pattern predicted by *in silico* methods that only use experimentally-derive sequence homology rules to determine targeting propensity. I show that this trend does not depend on licensing Argonaute CSR-1. I also show that the anti-silencing activity of CSR-1 works differently when combatting piRNA targeting versus WAGO 22G-RNA targeting: CSR-1 protects full transcripts from piRNA targeting uniformly across CSR-1 targeted transcripts, while CSR-1 protection only works locally within targeted transcripts to protect against WAGO 22G-RNA binding.

Fourth, I describe three tools that can allow researchers without advanced knowledge of bioinformatic techniques to perform many of the analysis tasks described throughout this work. pirScan allows researchers to access and manipulate the *in silico* piRNA target prediction rules to predict potential targets of the piRNA pathway and to reduce the likelihood of that targeting by finding silent mutations that can be introduced to a given sequence to avoid piRNA targeting. piRTarBase allows researchers to exhaustively query known and potential piRNA targets within the *C. elegans* and *C. briggsae* transcriptomes. CLASH Analyst allows researchers to analyze complex chimeric read data to determine likely *in vivo* small RNA regulatory events.

Together, this work attempts to address key missing features in our knowledge of germline small RNA pathways and adds valuable insights that will allow for a more complete understanding of how small RNAs promote RNA homeostasis.

CHAPTER 2

IDENTIFICATION AND CHARACTERIZATION OF NOVEL PIRNA PATHWAY MEMBERS USING A REPORTER ASSAY

2.1 Attributions

This chapter has been adapted from: J.S. Brown, *et al.*, Sensitized piRNA reporter identifies multiple RNA processing factors involved in piRNA-mediated gene silencing, bioRxiv (2023). Wenjun Chen performed piRNA reporter strain crosses to *glh-1 glh-4* and *meg-3 meg-4* mutant strains. Donglei Zhang created the *prg-1* mutant piRNA reporter strain. The author conducted the other experiments and analyses.

2.2 Abstract

Metazoans guard their germlines against transposons and other foreign transcripts with PIWI-interacting RNAs (piRNAs). Due to the robust heritability of the silencing initiated by piRNAs in *C. elegans*, previous screens using *Caenorhabditis elegans* were strongly biased to uncover members of this pathway in the maintenance process but not in the initiation process. To identify novel piRNA pathway members, we have utilized a sensitized reporter strain which detects defects in initiation, amplification, or regulation of piRNA silencing. Using our reporter, we have identified Integrator complex subunits, nuclear pore components, protein import components, and pre-mRNA splicing factors as essential for piRNA-mediated gene silencing. We found the snRNA processing cellular machine termed the Integrator complex is required for both type I and type II piRNA production. Notably, we identified a role for nuclear pore and nucleolar components in promoting the perinuclear localization of anti-silencing CSR-1 Argonaute, as well as a role for Importin factor IMA-3 in nuclear localization of silencing Argonaute HRDE-1. Together, we have shown that piRNA silencing is dependent on evolutionarily ancient RNA

processing machinery that has been co-opted to function in the piRNA mediated genome surveillance pathway.

2.3 Introduction

To defend against invading nucleic acids, organisms must identify and silence foreign RNAs while preserving the expression of endogenous RNAs. When organisms fail to combat virus and transposon derived invasive nucleic acids, these elements can destabilize the genome and lead to widespread mutagenesis and infertility (1). One system animals have evolved to combat foreign nucleic acids is comprised of PIWI Argonaute and its associated small noncoding RNAs (piRNAs) (2–4). In *C. elegans*, piRNAs are able to initiate a robust silencing signal against mRNAs deemed as non-self (GFP, for example) which can be inherited for as many as 20 generations after the loss of the initiating PIWI protein complex (5). Because piRNA-initiated gene silencing is heritable independent of the initiation factors, piRNA reporters that rely on activation of a piRNA-silenced GFP transgene are biased to detect components that are required for maintaining and amplifying the already initiated silencing signal. For this reason, some of the requirements for the piRNA pathway in *C. elegans* likely remain unknown.

Here, we present a candidate RNAi screen in *C. elegans* using a reporter that relies on the previously characterized phenomenon that particular GFP transgenes have distinct propensities to be silenced when challenged with a perfectly complementary piRNA (6). We used this piRNA reporter to test for the involvement of snRNA processing factors, mRNA processing factors, germ granules components, and protein transport factors in piRNA-mediated silencing. We further characterized components from each of these categories to better understand how they mediate silencing in *C. elegans*. Our data suggest that the more recently evolved piRNA pathway utilizes highly evolutionarily conserved snRNA processing, pre-mRNA splicing, and protein

import factors to carry out small RNA silencing, including the biogenesis of piRNAs and the regulation of Argonaute localization. Taken together, our results demonstrate that our piRNA reporter can identify various factors involved in distinct steps of the piRNA pathway. Our results also showed that proper piRNA-mediated transcriptome-wide surveillance relies on various regulatory factors that are not directly required for gene silencing.

2.4 Results

2.3.1 piRNA reporter detects loss of piRNA-dependent silencing factors

As described above, because piRNA targeting initiates a robust and stable silencing signal that persists even in the absence of the PIWI Argonaute protein PRG-1, previous reporter-based strategies have been biased to detect downstream components of the piRNA pathway (5,7). To remove this bias and increase the sensitivity of our screen, we aimed to establish a GFP reporter that is silenced in a piRNA dependent manner. We relied on previous observations that different GFP transgenes show distinct capacities to become silenced by the piRNA pathway (6). Our reporter strain contains two separate GFP transgenes with contrasting silencing propensities. One transgene, *cdk-1::gfp,* becomes stably silenced when challenged by an artificial piRNA (integrated into endogenously encoded piRNA locus *21ur-5499*) that perfectly complements a 20nt site within the *gfp* coding sequence (8). A second transgene, *oma-1::gfp,* cannot be silenced by a perfectly complementary piRNA (6). When we crossed these two lines to one another to obtain a strain that contains both GFP transgenes as well as the perfectly complementary piRNA, both transgenes are silenced. We hypothesized that because this strain contains one transgene that is vulnerable to piRNA silencing and another that resists piRNA silencing, then perhaps the silenced state of GFP is less stable and easily perturbed. To test this hypothesis, we introduced a frameshift mutation to the sole PIWI Argonaute *prg-1* using the CRSIPR/Cas9 system. We found

that both transgenes became expressed at the first generation of *prg-1* homozygosity (F2 generation) after injection (Figure 2.1A). Therefore, this reporter can successfully detect loss of *prg-1*, suggesting that a screen using this sensitized reporter would be sensitive to loss of upstream factors in the piRNA pathway.

Although many factors in the *C. elegans* piRNA pathway have been characterized, there are likely unknown components of this pathway that have not yet been discovered due to the use of less sensitive reporters in genetic screens. We therefore sought to use our sensitized reporter to identify components involved in the pathway both at the piRNA biogenesis level and at the effector level (Supplementary Figure 2.8). We used a candidate approach by selecting specific factors to knockdown using RNAi (Figure 2.1B). To ensure that this approach can successfully identify the loss of factors at both of these levels, we performed RNAi against known piRNA biogenesis and effector components. We saw that depletion of piRNA biogenesis factors *snpc-4*, *tofu-2*, and *tofu-5* resulted in significant GFP expression in the reporter, as did effector factors *nrde-1*, *nrde-2*, *set-25*, and *hpl-2* (Figure 2.1C). As expected, we failed to observe significant GFP expression in the reporter when a somatic RNAi factor, *nrde-3,* was depleted with RNAi. These results suggested that our sensitized reporter can be used to detect loss of piRNA components using an RNAi based approach. The full list of candidates tested using our reporter can be found in Supplementary Table 1.

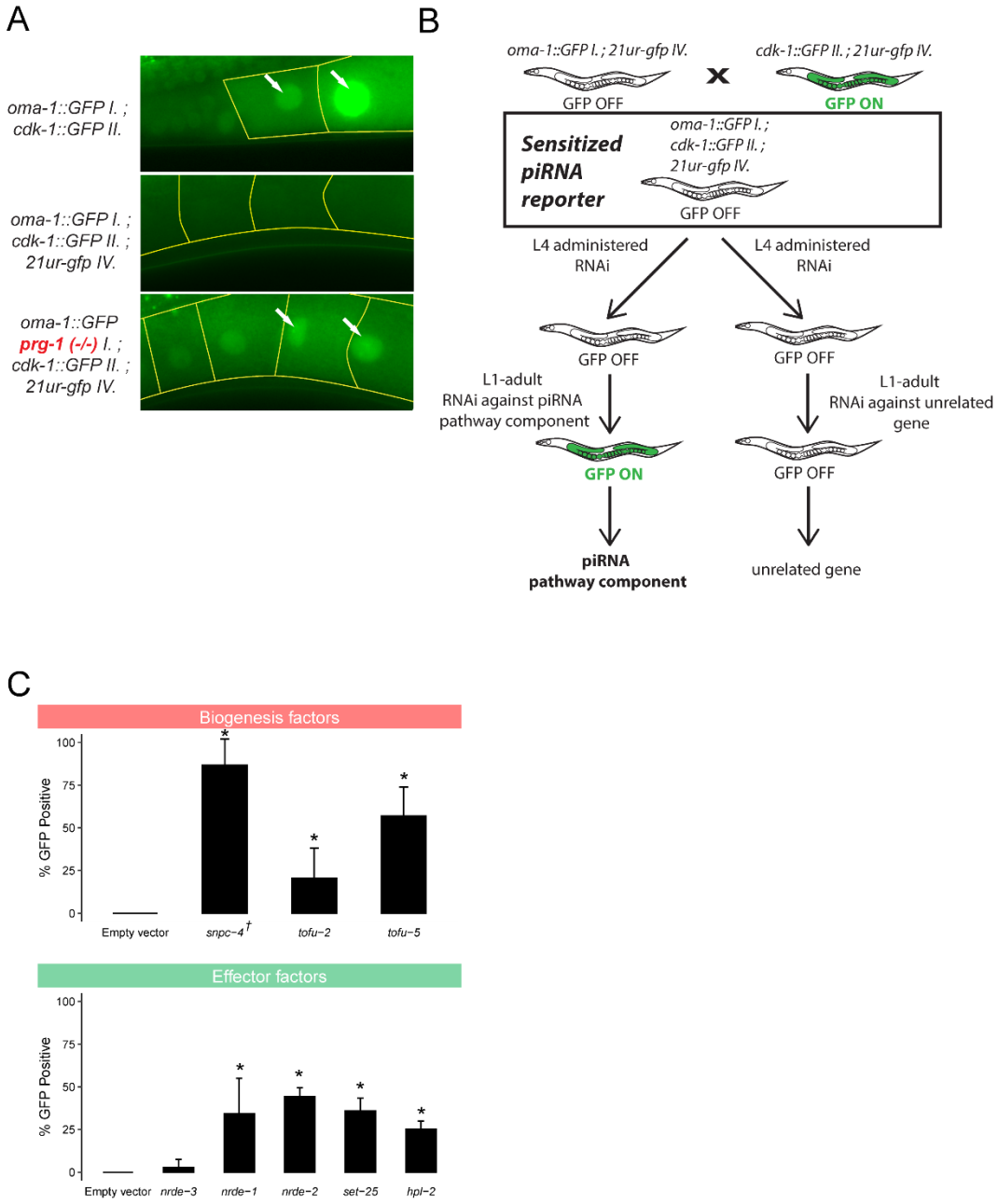


Figure 2.1: piRNA reporter detects loss of biogenesis and effector factors.
(A) Fluorescent micrographs show the expression of GFP transgenes in oocyte cytoplasm and nuclei in the absence of a triggering piRNA or *prg-1*. (B) Schematic shows the screening strategy using the sensitize piRNA reporter. (C) GFP expression in the piRNA reporter in indicated RNAi knockdowns. † indicates L1 to adult same generation RNAi treatment. Asterisks indicate significant reporter activation, see Methods for details.

2.3.2 RNAi screen identifies P granule factors, mRNA processing factors, and protein transport factors as piRNA silencing components

piRNA transcription in *C. elegans* is accomplished by RNA pol II. Since piRNA precursor transcription requires the SNPC-4 complex, which is also involved in initiation of snRNA transcription, we hypothesized that the Integrator complex, a multimeric snRNA processing complex required for snRNA 3' end processing, may be involved in the termination of piRNA transcription. Our model is consistent with the recent report that termination of piRNA precursor transcription involves the Integrator complex (9,10). We performed RNAi against 9 different components of the Integrator complex in the sensitized reporter and found four of the components, *dic-1*, *ints-1*, *ints-9*, and *ints-11*, triggered GFP expression in the reporter following knockdown (Figure 2.2A). We failed to observe GFP expression following knockdown of *ints-4*, *ints-5*, *ints-7*, *ints-12,* or *ints-13* despite the gross morphological phenotypes resulting from most of these treatments. As individual knockdown of *ints-1*, *ints-2*, *ints-4*, *ints-5*, *dic-1*, *ints-7*, *ints-8*, *ints-9*, or *ints-11* has been shown to disrupt snRNA 3' end processing in *C. elegans* (11), yet our results indicate that only individual knockdown of *ints-1*, *dic-1*, *ints-9*, or *ints-11* leads to defects in piRNA silencing, our finding suggests that a subset of the Integrator complex could be functioning at piRNA loci in a fashion distinct from its function at snRNA loci and coding genes. It has been shown recently that the mammalian Integrator complex does indeed form structurally distinct subcomplexes (12). Although a subcomplex composed of *ints-1*, *dic-1*, *ints-9*, and *ints-11* was not reported, our results suggest that these four components promote piRNA biogenesis.

piRNA pathway components are enriched in phase-separated perinuclear organelles called P granules. Previous studies have shown that when P granules are lost, RNAi is compromised (13). It remains unclear, however, whether specific P granule regulators are

important for gene silencing or whether the structures themselves are indispensable. We addressed this question by treated the sensitized reporter with RNAi against genes which have previously been shown to be enriched in or important for the formation of P granules (14). We found that loss of both *glh-1* and *glh-4* or *meg-3* and *meg-4* led to GFP expression in the piRNA reporter (15). In addition, knockdown of nuclear pore components *npp-6. npp-7,* and *npp-9*, phosphatase *cdc-25.1*, and proteasome component *pas-5* all led to GFP expression in the reporter (Figure 2.2A). The specific role of *npp-6/7* in promoting piRNA silencing is further explored below.

Pre-mRNA splicing has recently been shown to act as a signal for endogenous RNAi pathways in *C. elegans* (16–18). In addition, knockdown of some splicing factors leads to defects in P granule formation (14). We therefore wondered whether loss of particular splicing factors would trigger GFP expression in our sensitized reporter. We found that of the 13 splicing factors we tested, only loss of the second step factors *prp-17* and *prp-21* triggered GFP expression (Figure 2.2A), despite the gross morphological phenotypes resulting from treatment with most other RNAi constructs. This finding suggests some specific pre-mRNA splicing factors contribute to piRNA silencing, consistent with reports that splicing factor EMB-4 is essential to piRNA-mediated silencing for reporters containing multiple introns (17,18).

piRNA targeting triggers a silencing signal that relies on siRNAs to function both with secondary Argonaute proteins called WAGOs (Worm-specific ArGOnautes) in the cytoplasm and in the nucleus. The RNA-dependent RNA polymerases that manufacture these siRNAs, termed 22G RNAs due to their possession of a 5’ guanosine and 22nt length, all exist and function in the cytoplasm (19,20). Some of these 22G-RNAs associate with HRDE-1 (WAGO-9) Argonaute which is critical for the inheritance of gene silencing over generations (21). However,

it remains unknown how HRDE-1 and piRNA-dependent 22G RNAs translocate into the nucleus in germ cells. We hypothesized that HRDE-1 binds 22G RNAs in the cytoplasm and subsequently translocates into the nucleus as has been shown with the nuclear PIWI Argonaute in *Drosophila* and the somatic nuclear Argonaute in *C. elegans* NRDE-3 (22,23). To identify the protein import machinery required for nuclear localization of HRDE-1, we depleted several Importin factors in our sensitized reporter and observed GFP expression following *ima-3*, *imb-2*, and *imb-5* knockdown, but not following *ima-1* or *ima-2* knockdown (Figure 2.2A). Since *ima-1* RNAi knockdown led to sterility but not piRNA reporter activation, this suggests that specific protein import machinery is essential for piRNA-mediated gene silencing (see more details on HRDE-1 nuclear localization below).

Because our sensitized reporter can detect piRNA pathway defects at any level of the pathway, such as piRNA biogenesis or production of 22G-RNAs (Figure 2.1C), we next sought to characterize the defects contributed by loss of the factors reported above. To accomplish this goal, we performed RNAi against a subset of these factors and sequenced the small RNAs from a population of depleted animals. While loss of *npp-7*, *prp-17*, *ints-1*, or *dic-1* all led to a reduction in 22G accumulation (Figure 2.2B), only RNAi of Integrator complex components *ints-1* and *dic-1* caused a reduction in piRNAs as well. Reduction of both piRNA and 22G-RNA levels is consistent with the Integrator complex's involvement in piRNA biogenesis, as compromised piRNA production can also lead to reduced 22G-RNA accumulation as is the case in *prg-1* mutants (24). In contrast, the reduction in 22G-RNA but not piRNA levels suggests that *npp-7* and *prp-17* likely function as effectors in the piRNA pathway and do not function in piRNA biogenesis. We then characterized the defects in each of these cases further.

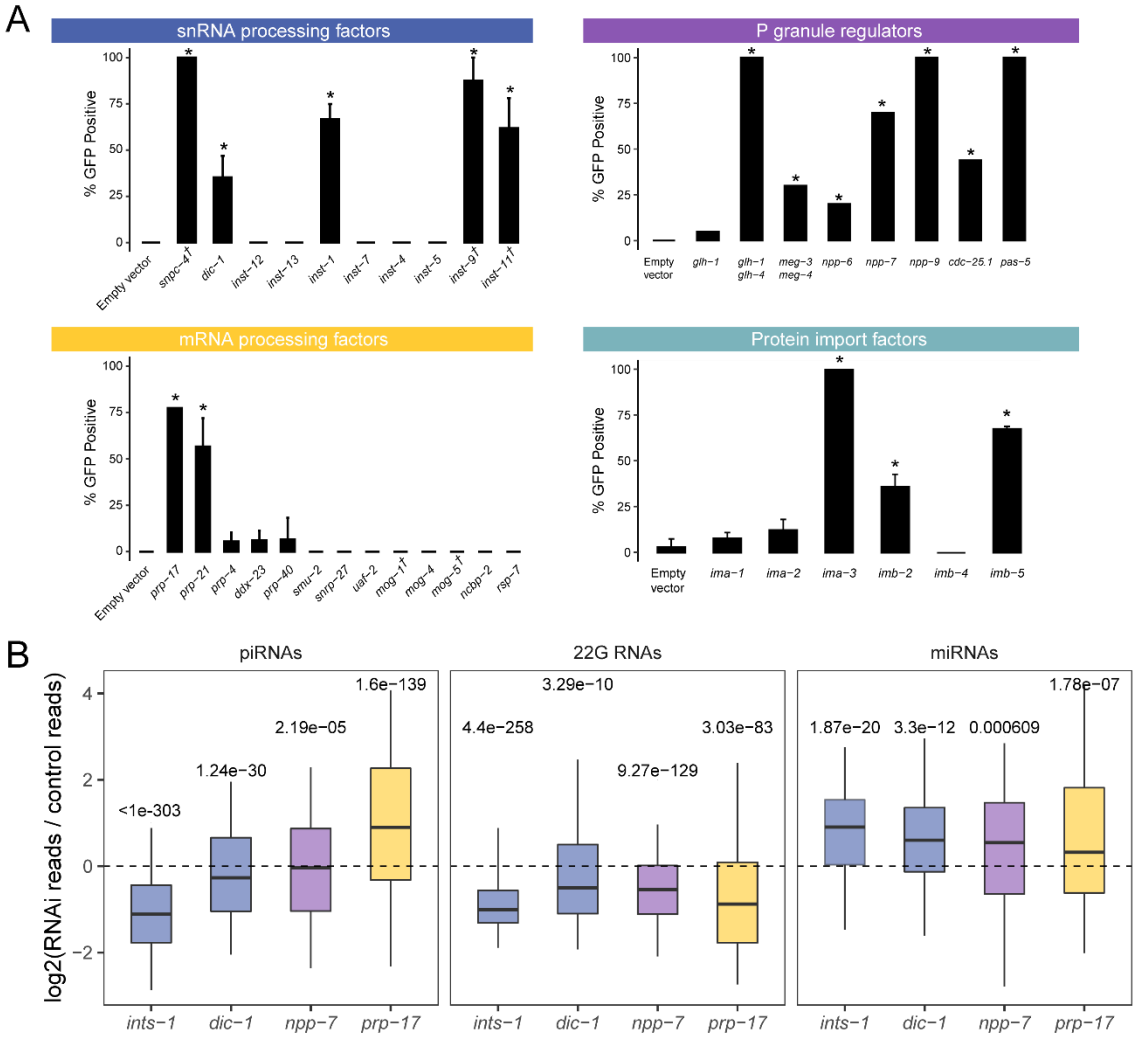


Figure 2.2: Loss of mRNA processing factors, snRNA processing factors, protein import factors, and P granules components trigger piRNA reporter activation.
(A) GFP expression in the piRNA reporter in indicated RNAi knockdowns or mutants (for *glh-1 glh-4* and *meg-3 meg-4*). † indicates L1 to adult same generation RNAi treatment. Asterisks indicate significant reporter activation, see Methods for details. (B) Small RNA fold changes in the indicated RNAi knockdowns relative to empty vector control knockdown. Statistical analysis was performed using a two-tailed Mann–Whitney Wilcoxon test.

2.3.3 Integrator complex resolves piRNA precursor 3' ends and promotes the production of both type I and type II piRNAs

We found that loss of a subset of Integrator complex components caused GFP expression in our sensitized piRNA reporter, consistent with a recent report (9). We sequenced small RNAs from animals depleted for Integrator components *ints-1* or *dic-1* and observed a global reduction in mature piRNA production (Figure 2.2B and Supplementary Figure 2.9A). In addition, piRNA precursor levels are strongly decreased following *ints-1* and *dic-1* knockdown (Supplementary Figure 2.9B). In mammals and *Drosophila*, Integrator complex action at snRNA loci requires the SNAPc complex of proteins to bind the promoter of snRNA genes (25). The SNAPc homolog SNPC-4 in *C. elegans* is also required for piRNA biogenesis and interacts with piRNA promoters (26). We reasoned that if SNPC-4 recruits the Integrator complex to piRNA loci, then piRNAs that rely most on SNPC-4 for their accumulation should also rely on the Integrator. We found that this was the case for both INTS-1 and DIC-1 (Figure 2.3A). Together, these observations suggest that the Integrator plays a role in regulating the production or processing of piRNA precursors.

At snRNA loci in *C. elegans* and *Drosophila*, loss of the Integrator complex results in transcriptional readthrough and consequently elongated snRNA transcripts (11,27). Therefore, we predicted that Integrator depletion would result in readthrough of piRNA precursors at piRNA loci. Since piRNA precursor molecules are rare compared to mature piRNAs and other abundant small RNA species, we performed CapSeq to enrich for capped piRNA precursor molecules (28). In empty vector treated control animals, we found the size of piRNA precursor lengths peak at 25, 41, and 61 nucleotides. Following *ints-1* depletion, animals showed less defined peaks and elongated piRNA precursors (Supplementary Figure 2.9C), consistent with a

role for the Integrator complex in transcriptional termination of piRNA precursors. These findings support a model where at type I piRNA loci, SNPC-4 recruits a subset of the Integrator complex to elongating RNA pol II, leading to the endonucleolytic cleavage and subsequent termination of nascent piRNA precursors (Supplementary Figure 2.9D).

SNPC-4 interacts with an upstream sequence called the Ruby motif, which is only found at type I piRNA loci (2). Surprisingly, we found that both type I and type II piRNA accumulation were globally reduced following *ints-1* and *dic-1* knockdown, although type II accumulation was less affected than type I following depletion of either component (Figure 2.3B). Our finding contradicts the recent report that INTS-11 does not function at type II piRNA loci (9), so we sought to clarify this point. To investigate whether the Integrator is required for gene silencing by type II piRNAs, we built a sensitized piRNA reporter where the GFP-targeting piRNA is made from a type II piRNA locus (28). We found that *snpc-4* depletion leads to only type I reporter activation but not type II reporter activation. However, RNAi depletion of *ints-1*, *dic-1, ints-9*, or *ints-11* resulted in GFP expression in both type I and type II reporters (Figure 2.3C). This suggests that at type I piRNA loci, the Integrator functions with SNPC-4 to promote piRNA accumulation while at type II piRNA loci, the Integrator can function in a SNPC-4 independent manner to promote piRNA biogenesis.

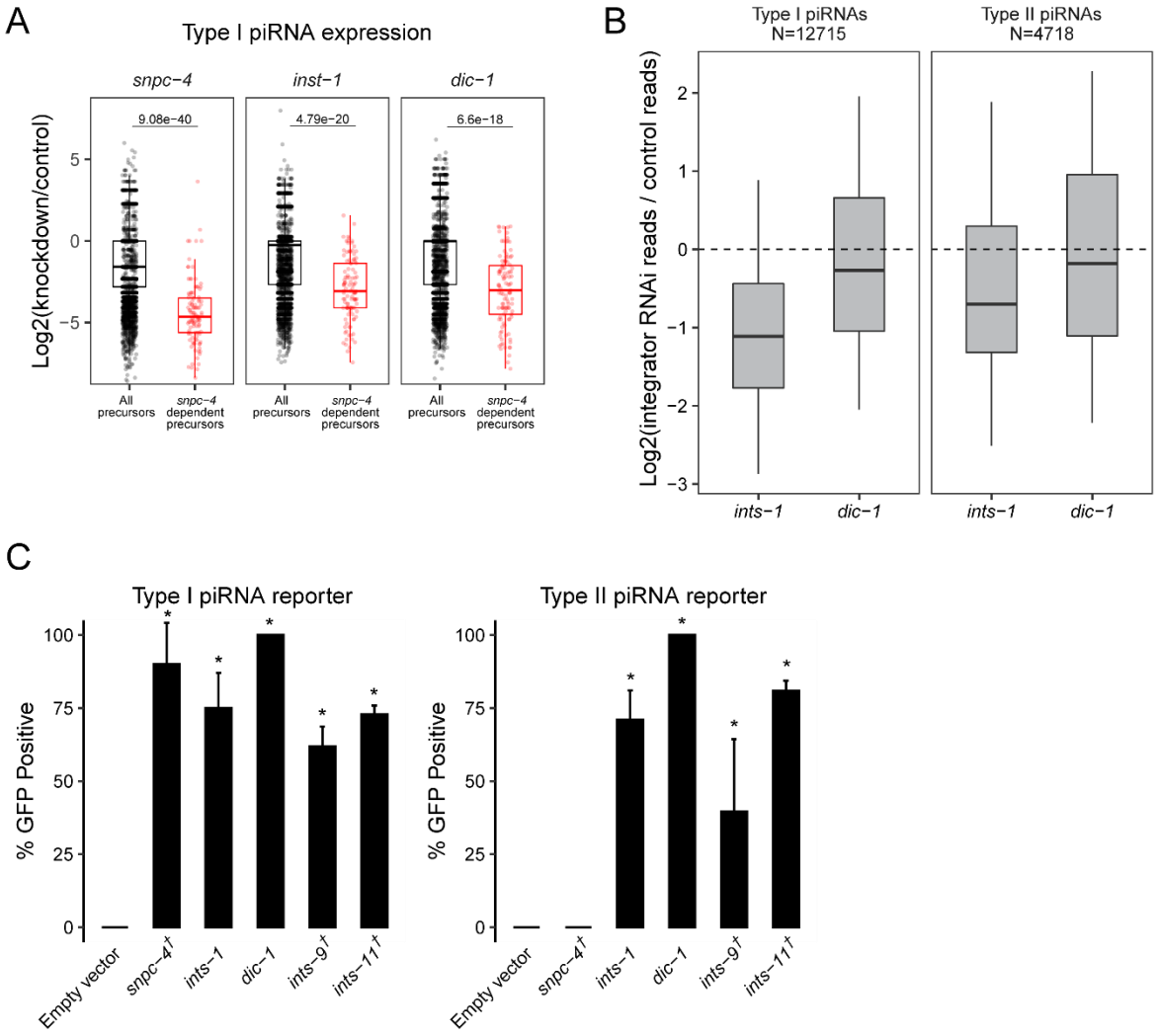


Figure 2.3: The Integrator Complex promotes piRNA biogenesis.
(A) piRNA precursor fold changes in indicated knockdowns relative to empty vector control knockdown. All piRNA precursors are compared to precursors originating from loci which most depend on *snpc-4* for the accumulation of their mature sequences. Statistical analysis was performed using a two-tailed Mann–Whitney Wilcoxon test. (B) piRNA fold changes in the indicated knockdowns relative to empty vector control knockdown. Type I and type II piRNAs are compared separately. (C) GFP expression in the type I and type II piRNA reporters in indicated RNAi knockdowns. † indicates L1 to adult same generation RNAi treatment. Asterisks indicate significant reporter activation, see Methods for details.

2.3.4 Select nuclear pore and nucleolus components maintain CSR-1 perinuclear accumulation and promote piRNA silencing

Out of over 20 nuclear pore components tested, only 4 nuclear pore components (*npp-1*, *npp-6*, *npp-7*, and *npp-9*) caused over 40% GFP expression in our reporter when depleted (Figure 2.4A), despite most treatments resulting in morphologically effected worms (Supplementary Table 1). This suggests that only a subset of the factors that make up the nuclear pore are essential for piRNA silencing. As describe above, we found that loss of *npp-7* resulted in a reduction in 22G-RNAs but not in piRNAs, (Figure 2.2B), suggesting that *npp-7* is not critical for piRNA biogenesis but rather functions in the production of secondary WAGO 22G-RNAs. Furthermore, when we performed RNAi against *npp-1*, *npp-6*, or *npp-7* using the type II piRNA reporter established above, each treatment led to type II reporter activation (Supplementary Figure 2.10A), consistent with nuclear pore components functioning downstream of piRNA biogenesis. Surprisingly, 22G-RNAs that map to CSR-1 targeted genes were significantly more reduced than 22G-RNAs that map to WAGO targeted genes following *npp-7* knockdown (Figure 2.4B), indicating that NPP-7 may play a critical role in regulating CSR-1 function.

CSR-1 is an Argonaute protein that prevents PRG-1 from targeting many endogenous germline genes (29). Consistent with a more pronounced loss of CSR-1 bound 22G RNAs following *npp-7* knockdown, we also observed mis-localization of CSR-1 protein in *npp-7* and *npp-6* RNAi treated animals (Figure 2.4C). These results suggest that the production and/or function of CSR-1 small RNAs are likely compromised in *npp-7* depleted animals. If the GFP expression that we observed in our piRNA reporter following *npp-7* knockdown was caused by a CSR-1 defect, then we reasoned that knockdown of *csr-1* itself by RNAi should also result in GFP expression. Indeed, we found that knockdown of *csr-1* and of the H3K36 methyl transferase

mes-4 both caused GFP expression in our reporter (Supplementary Figure 2.10B). This suggests that disruption of CSR-1 or of chromatin modifiers associated with promoting gene expression can also impact piRNA mediated gene silencing. Therefore, our observations suggest that *npp-6* and *npp-7* may contribute to piRNA silencing by regulating the CSR-1 pathway.

We wondered how predictive dispersal of CSR-1 was on piRNA reporter activation for the remaining nuclear pore components. Therefore, we monitored the expression of CSR-1 and PRG-1 in live adult animals following knockdown of all previously screened nuclear pore components (Supplementary Figure 2.10C). We found that knockdown of *npp-1* caused CSR-1 but not PRG-1 to become mis-localized as with *npp-7* and *npp-6*. These observations suggest that NPP-1, NPP-6, and NPP-7 specifically impact the localization of CSR-1 but not that of other P granule factors at the nuclear periphery, and these three factors also promote piRNA-dependent gene silencing.

On the other hand, *npp-9* knockdown led to piRNA reporter activation but did not cause dispersal of CSR-1 or of PRG-1. This result was surprising, as it has been previously shown that RNAi knockdown of *npp-9* does cause CSR-1 dispersal in *C. elegans* embryos (14). Conversely, knockdown of *npp-24* and *npp-2* cause CSR-1 dispersal but do not lead to piRNA reporter activation. Therefore, there is an imperfect correlation between nuclear pore component loss leading to piRNA reporter activation and CSR-1 dispersal. Nonetheless, our study reveals that select nuclear pore components, including NPP-1/6/7, promote CSR-1 perinuclear localization and are required for proper piRNA silencing.

The *Drosophila* homolog of NPP-7, Nup153, has been shown to function both at the nuclear membrane as a component of the nuclear pore and as a soluble component within the nucleoplasm to promote both gene expression and silencing (30,31). We therefore wondered

whether NPP-7 in *C. elegans* may also be expressed intranuclearly to help CSR-1 define transcriptionally active regions of the genome. We used the CRISPR/Cas9 system to tag endogenous NPP-7 with mCherry and observed its expression in adult animals, and we found that in addition to its strong expression at the nuclear periphery, NPP-7 is also expressed as a single distinct focus in the nucleoplasm (Figure 2.4D). Surprisingly, we found that the single focus of intranuclear NPP-7 in germ cells resides in the nucleolus, and the constitutive nucleolar component FIB-1/fibrillarin surrounds the intranuclear NPP-7 focus. We found that nucleolar NPP-7 accumulates in nearly every germline nucleus but only rarely in somatic nuclei. Additionally, nucleolar but not membrane associated accumulation of NPP-7 relies on FIB-1 (Supplementary Figure 2.10D). We wondered whether reporter GFP expression following *npp-7* RNAi occurred due to loss of membrane-associated or nucleolar NPP-7. If nucleolar NPP-7 is required for piRNA silencing, then we expected that loss of FIB-1, which disrupts nucleolar NPP-7 accumulation, would cause piRNA reporter activation. Indeed, *fib-1* RNAi led to GFP expression in the piRNA reporter, consistent with involvement of the nucleolus in the piRNA pathway (Figure 2.4E). This result does not exclude the possibility that both nucleolar as well as membrane bound NPP-7 are mutually required for piRNA silencing. In vertebrae, the NPP-2 homologue Nup85 is required for nuclear pore assembly following cell division (32). We reasoned that depletion of NPP-2 might disrupt the membrane bound fraction of NPP-7 while leaving the nucleolar fraction unaffected, giving us an opportunity to observe the consequences of membrane bound but not nucleolar loss of NPP-7. Indeed, we found that *npp-2* knockdown affected the enrichment of NPP-7 in the nuclear membrane, resulting in a greater proportion of NPP-7 signal in the nucleolus (Supplementary Figure 2.10D). As shown above, *npp-2* knockdown also leads to CSR-1 dispersal from perinuclear granules. The presence of nucleolar

NPP-7 may explain why *npp-2* knockdown does not compromise piRNA reporter silencing. Furthermore, *npp-1* and *npp-6* knockdown selectively disrupted NPP-7 nucleolar accumulation and also led to sensitized reporter GFP expression and CSR-1 dispersal. Nonetheless, the correlation of nucleolar NPP-7 localization and piRNA silencing was not perfect as *npp-5* knockdown disrupted nucleolar NPP-7 accumulation but did not lead to reporter GFP expression or CSR-1 dispersal. Nevertheless, this correlation suggests that nucleolar and not membrane bound NPP-7 is involved in the piRNA pathway.

We next sought to reconcile the two observations concerning NPP-7 presented here: (1) the involvement of NPP-7 in CSR-1 localization and function and (2) the role of nucleolar NPP-7 in piRNA pathway function. We hypothesized that these two observations may be directly connected if nucleolar NPP-7 is required for CSR-1 localization. To test this, we performed *fib-1* RNAi in a GFP::CSR-1 tagged strain. Remarkably, *fib-1* knockdown disrupts perinuclear CSR-1 accumulation (Figure 2.4F), but not PRG-1 perinuclear accumulation similarly to *npp-7* knockdown (Supplementary Figure 2.10D). Together, our study uncovers roles for select nuclear pore and nucleolus components in promoting CSR-1 localization and piRNA silencing.

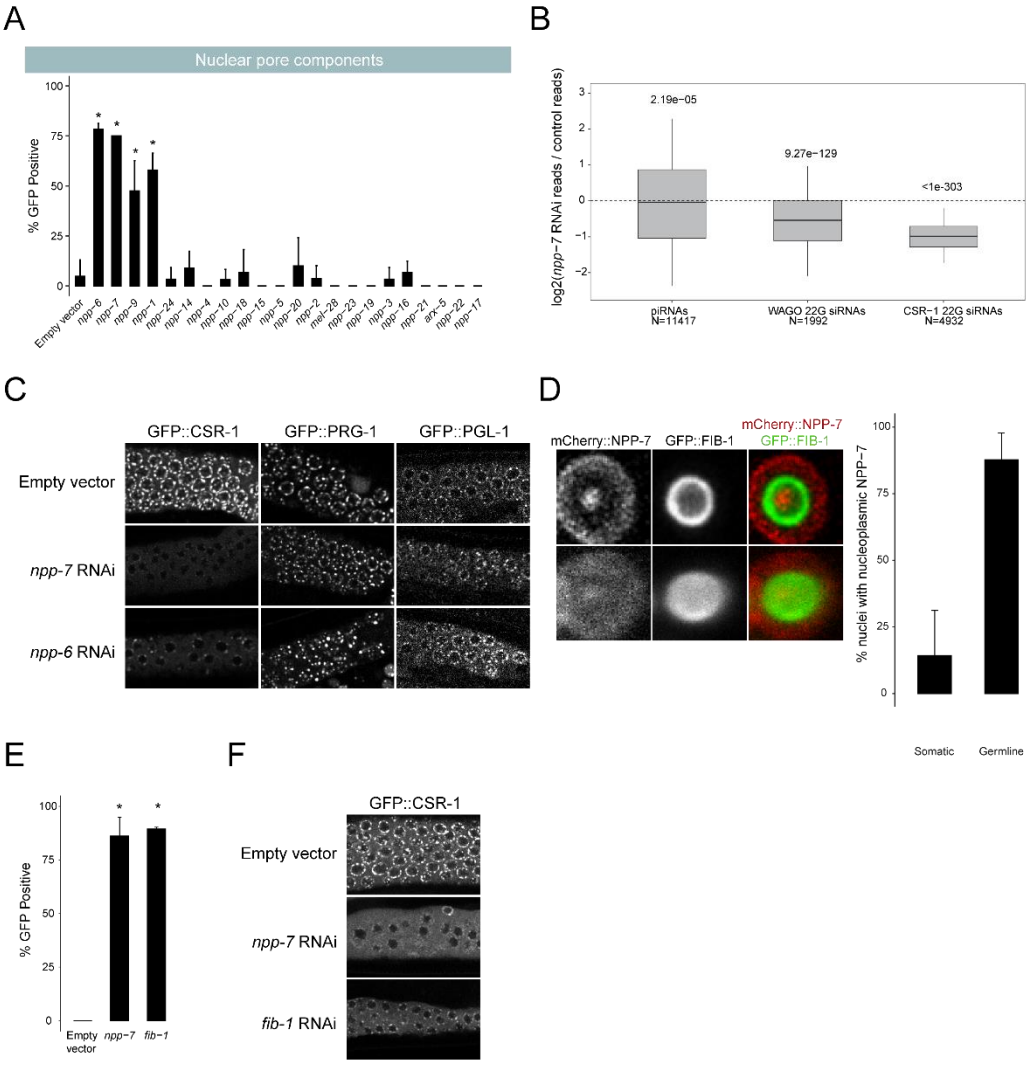


Figure 2.4: Nuclear pore component NPP-7 and the nucleolus promote piRNA dependent silencing.

(A) GFP expression in piRNA reporter in indicated RNAi knockdowns. (B) Small RNA fold changes in *npp-7* RNAi knockdowns relative to empty vector control knockdown. Statistical analysis was performed using a two-tailed Mann–Whitney Wilcoxon test. (C) Fluorescent micrographs of live adult worm gonads show the localization of GFP::CSR-1, GFP::PRG-1, and GFP::PGL-1 in indicated RNAi knockdowns. (D) Fluorescent micrograph of individual live adult worm germline nucleus shows the localization of mCherry::NPP-7 and GFP::FIB-1. Top panels display a single slice through the nucleus center, bottom panels display average projection of all slices though the full nucleus volume. Percentage of soma and germline nuclei with nucleoli that contain NPP-7 are quantified to the right. (E) GFP expression in piRNA reporter in indicated RNAi knockdowns. Asterisks indicate significant reporter activation, see Methods for details. (F) Fluorescent micrographs of live adult worm gonads show the localization of GFP::CSR-1 in indicated RNAi knockdowns.

2.3.5 mRNA splicing factor PRP-17 promotes proper mRNA surveillance

We found that PRP-17 and PRP-21, two splicing factors that function at the second catalytic step of pre-mRNA splicing [73,166], activate GFP expression in the sensitized piRNA reporter when depleted with RNAi. To further characterize the role of pre-mRNA splicing in the piRNA pathway, we sequenced small RNAs from a population of animals treated with *prp-17* RNAi. We found that 22G RNAs were globally reduced following *prp-17* depletion (Figure 2.5A). Further, we saw that CSR-1 associated 22G RNAs were significantly more affected than WAGO associated 22G RNAs. Because PRP-17 is known to function in pre-mRNA splicing, we wondered if there were any disparities in 22G RNA accumulation in introns vs exons. We compared reads that map to intronic regions to those that map to exonic regions on a gene by gene basis and found that for CSR-1 targeted genes but not for WAGO targeted genes, more intronic reads accumulate compared to exonic reads when *prp-17* is depleted compared to a control library (Figure 2.5B). This suggests that PRP-17 both promotes the accumulation of CSR-1 22G RNAs globally, and also promotes CSR-1 target pre-mRNA splicing. Because CSR-1 22G RNAs are produced in the cytoplasm by RNA-dependent RNA polymerase EGO-1, and because improperly splice pre-mRNAs are commonly targeted for degradation by nonsense mediated decay (NMD) machinery, it is likely that a failure for CSR-1 mRNAs to splice can also explain the CSR-1 22G RNA accumulation defect; if CSR-1 mRNA targets are improperly spliced and quickly degraded, then fewer CSR-1 target mRNAs can be templated for 22G RNA production. However, this does not explain why PRP-17 promotes the splicing of CSR-1 target pre-mRNAs but not WAGO target pre-mRNAs. Further work will need to be done to fully characterize this phenomenon.

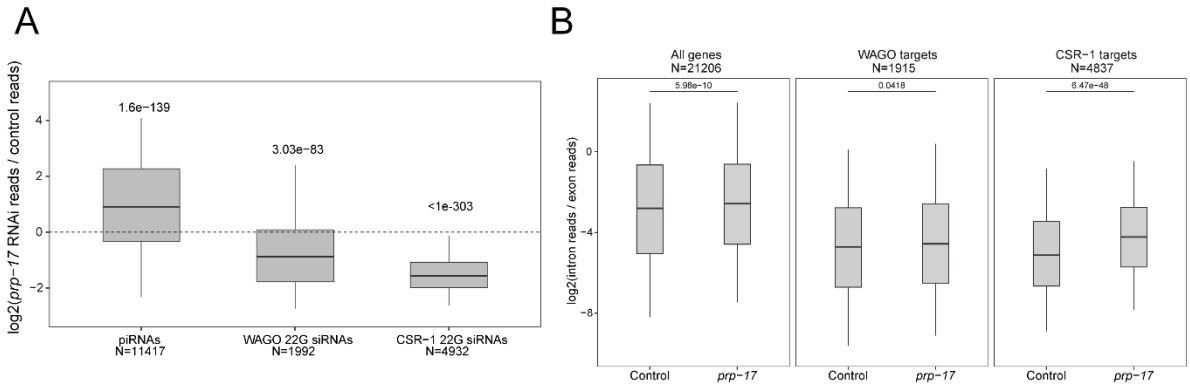


Figure 2.5: PRP-17 promotes piRNA dependent silencing and CSR-1 target pre-mRNA splicing. (A) Small RNA fold changes in *prp-17* RNAi knockdowns relative to empty vector control knockdown. Statistical analysis was performed using a two-tailed Mann–Whitney Wilcoxon test. (B) Ratio of small RNA reads mapping to introns versus exons in indicated RNAi knockdowns. WAGO and CSR-1 targets are separated from all genes. Statistical analysis was performed using a two-tailed Mann–Whitney Wilcoxon test.

2.3.6 Importin components promote HRDE-1 silencing and localization

We found that depletion of 3 Importin family components, *ima-3*, *imb-2*, and *imb-5*, caused GFP expression in the sensitized reporter (Figure 2.2A). Activation during *ima*-3 and *imb*-5 knockdown also occurred using the type II piRNA reporter, supporting a role for the Importin family downstream of piRNA biogenesis (Supplementary Figure 2.11A). Since knockdown of HRDE-1 cofactors NRDE-1 and NRDE-2 led to activation of our piRNA reporter, we hypothesized that HRDE-1 nuclear import might be compromised following depletion of these factors, which would in turn impact piRNA mediated silencing. To test this, we treated a GFP::HRDE-1 tagged animal with *ima-3* and observed HRDE-1 localization in dissected adult gonads using confocal microscopy. Indeed, in the *ima-3* depleted germline, HRDE-1 failed to localize to nuclei. On the contrary, in the *ima-1* depleted germline, which continues to silence the piRNA reporter, HRDE-1 localization was unaffected (Figure 2.6A). These results are consistent with a model in which IMA-3 promotes piRNA-mediated gene silencing by translocating 22G-RNA-bound HRDE-1 from the cytoplasm to the nucleus where it can interact with nascent RNAs

to transcriptionally silence piRNA targets. Interestingly, the nuclear Argonaute protein downstream of the piRNA pathway in *Drosophila*, Piwi, also relies on IMA-3 for translocation into the nucleus following piRNA binding (22).

To further support this model, we asked whether endogenous genes that are silenced by HRDE-1 also rely on IMA-3 to maintain that silenced state. Using RT-qPCR, we found that *B0250.8* and *F15D4.5* are significantly upregulated both in *hrde-1* mutants and following *ima-3* RNAi treatment (Figure 2.6B). Therefore, we conclude that HRDE-1 translocates from the cytoplasm to the nucleus using the Importin α member IMA-3. Importin α members are known to function as adaptor molecules between protein cargo that contains a nuclear localization sequence (NLS) and Importin β members (33). In *C. elegans,* the somatic nuclear Argonaute protein NRDE-3 contains a NLS that, when mutated, results in failure of NRDE-3 to translocate into nuclei (23). Surprisingly, despite sharing significant sequence identity with HRDE-1, the NLS within NRDE-3 poorly aligns with HRDE-1 (Supplementary Figure 2.11B). Therefore, how IMA-3 recognizes HRDE-1 for nuclear translocation requires further study.

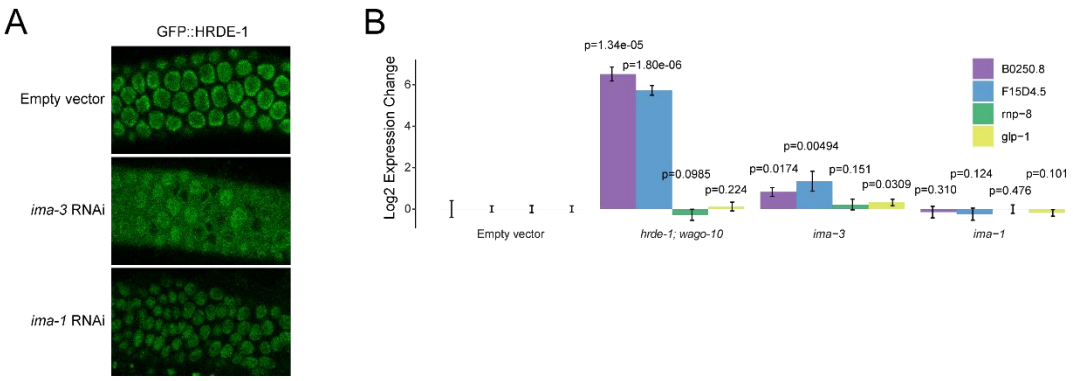


Figure 2.6: IMA-3 promotes HRDE-1 nuclear import.
(A) Fluorescent micrographs of live adult worms gonads show the localization of GFP::HRDE-1 in indicated RNAi knockdowns. (B) RT-qPCR results show expression change in indicated genotypes (*wago-9; wago-10*) or RNAi knockdowns (Empty vector, *ima-3,* and *ima-1*). Statistical significance from two-tailed student's t-test compared to control is displayed above treatments.

2.5 Discussion

We have utilized a sensitized reporter strain which detects defects in both piRNA biogenesis and effector factors to identify and characterize factors involved in piRNA silencing. Using our reporter, we have identified pre-mRNA splicing factors, Integrator complex subunits, protein import components, and nuclear pore components as essential for piRNA-mediated gene silencing. While we have currently used our reporter in a candidate screen, it is likely possible to use this strain in a forward screen, which would allow for even less biased detection of piRNA pathway components. Because the reporter robustly expresses germline GFP upon piRNA pathway disruption, activated worms containing a mutation in essential pathway components could possibly be sorted from silenced worms, for example using a COPAS BIOSORT system (34). Whether the expression of GFP can be distinguished from the significant auto-fluorescence

contributed by intestinal granules adjacent to and overlapping with the adult gonad by a worm sorter remains to be seen and could represent a significant barrier to a high throughput approach.

We found that a subset of the Integrator complex contributes to both type I and type II piRNA biogenesis. Because both snRNA loci and type I piRNA loci require the SNAPc complex for biogenesis (25,26), we expected that the Integrator complex would be involved in piRNA biogenesis as described by recent reports (9,10). Our reporter assay and sequencing data support this model, but unexpectedly and inconsistent with the previously published findings, we also noticed that type II piRNA biogenesis and silencing were compromised following Integrator complex knockdown. Because SNPC-4 is recruited to Ruby motif-containing type I piRNA loci and not type II piRNA loci (2,35), our data suggest that the Integrator also functions at type II piRNA loci in a SNPC-4 independent manner. It has been shown that the Integrator can also function at protein coding genes to terminate transcription, promote RNAP II transcription proximal pausing, and promote enhancer RNA accumulation (36–40). While the canonical function of the Integrator at snRNA loci requires recognition of a conserved 3' box sequence (41), protein-coding genes which rely on the Integrator complex to cleave nascent RNAs do not contain a recognizable 3' box sequence (36). Intriguingly, *C. elegans* snRNA loci are devoid of a 3' box sequence as well (42). Our data suggest that type II piRNA biogenesis represents yet another non-canonical role for the Integrator complex. Further investigation into the commonalities between snRNA loci, type I piRNA loci, type II piRNA loci, and protein coding genes may help to explain how the Integrator complex recognizes and cleaves nascent RNA molecules in *C. elegans*, and also may shed light on the Integrator's non-canonical roles in other organisms.

We found that a nuclear pore component, NPP-7, localizes to both the nuclear envelope and the nucleolus, and that depletion of both NPP-7 and of nucleolar component FIB-1 both activate our piRNA reporter. This suggests a potential role for the nucleolus in the piRNA pathway. In *Drosophila* ovarian somatic cells, nuclear Argonaute Piwi has been shown to localize to nucleoli under heat shock, correlating with the expression of retrotransposons which have integrated into rDNA copies (43). Additionally, it has been shown in *C. elegans* that nuclear RNAi machinery targets rRNAs which are degraded by exosomes in the nucleolus (44). Both of these observations together with our own data suggest that the nucleolus may be involved in specialized small RNA-mediated silencing events. Additionally and possibly connected to nucleolar NPP-7, we found that NPP-7 promotes the localization of germ granule component CSR-1, but not germ granule components PRG-1 or PGL-1. We recently showed that CSR-1 uniquely retains its perinuclear localization when VASA helicases are mutated while PRG-1 and PGL-1 each show significant dispersal, suggesting that CSR-1 is recruited to the perinucleus in a VASA-independent manner (15). Our results suggest that the nuclear pore itself may represent a part of this VASA-independent pathway, and further that CSR-1 recruitment may require involvement of the nucleolus as *fib-1* knockdown similarly led to specifically CSR-1 dispersal.

We observed that knockdown of licensing Argonaute *csr-1* and of the H3K36 methyl transferase *mes-4* both caused GFP expression in our piRNA reporter (Supplementary Figure 2.10B). It has been shown that H3K36me3 occupancy and CSR-1 22G RNA targeting are highly correlated (45). However, it has also been shown that CSR-1 does not directly induce H3K36 methylation (46). Therefore, we hypothesize that our piRNA reporter is activated when pathways that promote gene expression as well as those that control gene repression are disrupted. If

correct, then this model would suggest that a balance between gene silencing and expression exists within the germline and disruption of pathways that control either gene silencing or gene expression can impinge on the alternative. Consistent with this notion, it has been shown that "resetting" germline RNAi pathways by re-introducing RNAi machinery to RNAi deficient *C. elegans* in the absence of piRNAs led to potent sterility due to the silencing of essential genes (47,48). This finding suggests that silencing machinery is essential for keeping pathways that control silencing and expression distinct from one another and balanced in the germline. Our finding, conversely, suggests that factors which promote gene expression are also essential for maintaining this balance.

Altogether, we used a piRNA reporter assay to identify and characterize factors that act at multiple levels of gene silencing, revealing a role for snRNA processing machinery, pre-mRNA splicing factors, the nuclear pore, the nucleolus, and protein import machinery in initiating and maintaining small RNA-mediated gene regulation (Figure 2.7).

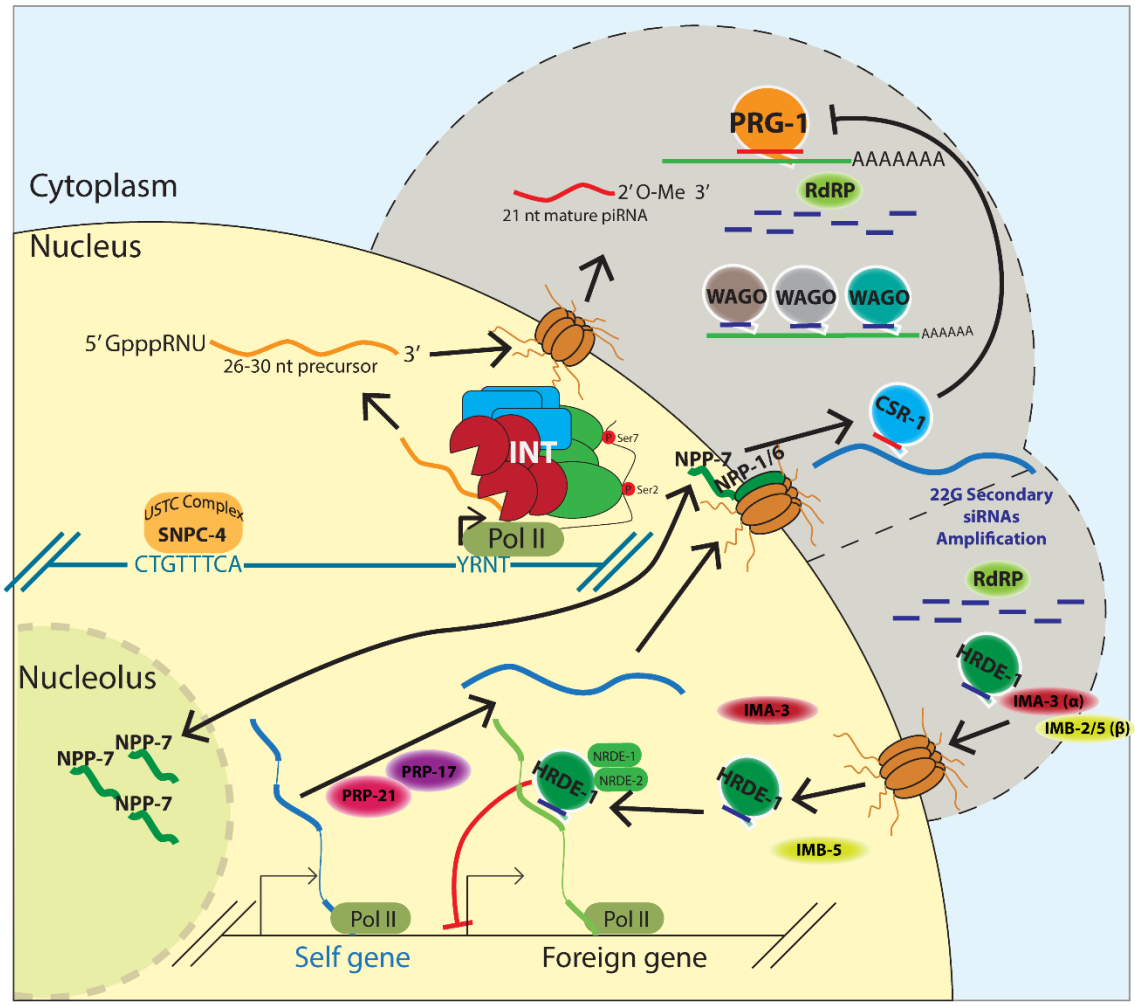


Figure 2.7: Multifaceted involvement of conserved and essential machinery work in concert with small RNA-mediated gene silencing in germ cells.
A model depicting a role for factors identified in our screen. The Integrator complex promotes piRNA biogenesis, a subset of nuclear pore factors promotes CSR-1 localization and interface with the nucleolus, second step splicing factors ensure proper endogenous splicing occurs preferentially, and protein import machinery allow for the nuclear Argonaute HRDE-1 to translocate into the nucleus to transcriptionally silence target genes.

2.6 Methods

2.5.1 C. elegans strains

Animals were grown on standard nematode growth media (NGM) plates seeded with the *Escherichia coli* OP50 strain at 20°C. COP262 (*knuSi221 [fib-1p::fib-1(genomic)::eGFP::fib-1 3′ UTR + unc-119(+)] II.*) and SS747 (*bnIs1[pie-1::GFP::pgl-1 + unc-119(+)]*) and YY538 (*hrde-1 (tm1200) III.*) were obtained from the Caenorhabditis Genetics Center (CGC). YY584 (*ggSi1[hrde-1p::3xflag::gfp::hrde-1] II.*) was a gift from Scott Kennedy. HCL199 (*hrde-1 (tm1200) III; wago-10 (tm1186) V.*) was outcrossed from WM191. HCL105 (*gfp::csr-1 IV.*) and HCL125 (*gfp::prg-1 flag::mCherry::glh-1 I.*) are described in (49). Creation of HCL202 (*mCherry::npp-7(uoc21) I.*) and HCL135 (*prg-1(uoc8) oma-1::gfp I; cdk-1::gfp II; 21ur-anti-gfp(type I) IV.*) are described below.

2.5.2 RNA interference

RNA interference (RNAi) was performed by feeding animals with *E. coli* HT115 (DE3) strains expressing the appropriate double-stranded RNA (dsRNA). RNAi bacterial strains were obtained from the Ahringer *C. elegans* RNAi Collection (Source BioScience). Bacterial cultures were grown in Luria broth supplemented with 100 μg/ml ampicillin and 50 μg/ml tetracycline overnight at 37°C. Cultures were seeded on NGM plates containing 100 μg/ml ampicillin, 50 μg/ml tetracycline, and 1 mM IPTG and incubated at room temperature until dry. L4 hermaphrodites were picked onto the plates for feeding at 20°C and removed after 24 hours. Adult progeny were screened. In instances where L4 treatment resulted in early next generation arrest, L1s were plated and screened as adults. Experiments where same generation screening was necessary are indicated in the relevant figure legend or data table. HT115 (DE3) expressing empty RNAi vector L4440 was used as the control. To assign significance to treatments in the

piRNA reporter activation screen, a threshold of 20% activation was used as control treatment sometime resulted in spurious activation, but never resulted in over 20% naïve activation at 20°C for type I or type II piRNA reporters.

2.5.3 CRISPR

Cas9/sgRNA constructs. We used the online tool sgRNA Scorer 2.0 (https://crispr.med.harvard.edu/) to design sgRNAs. The sgRNAs were cloned into pDD162 [41] by overlapping PCR using pDD162 as the PCR template and the appropriate primers. Overlapping PCR products were inserted into pDD162 linearized with SpeI/BsrBI digestion by the seamless ligation cloning extract (SLiCE) method [184].

Donor constructs. To generate the flag::mCherry::NPP-7 donor construct, 500 bp upstream and 500 bp downstream of the *npp-7* TSS, and the flag::mCherry coding sequences were amplified by PCR using N2 genomic DNA or plasmids containing flag::mCherry as templates. PCR fragments were inserted into pUC19 linearized with HindIII/KpnI digestion by SLiCE. Silent mutations were introduced in guide RNA targeting sites by site-directed mutagenesis in donor constructs using Phusion High-Fidelity DNA Polymerases (Thermo Fisher Scientific).

2.5.4 Fluorescence Imaging

GFP- and RFP/mCherry-tagged fluorescent proteins were visualized in living nematodes by mounting young adult animals on 2% agarose pads with M9 buffer (22 mM KH2PO4, 42 mM Na2HPO4, and 86 mM NaCl) with 10-50 mM levamisole. Fluorescent images used for localization studies were captured using a Zeiss LSM800 confocal microscope with a Plan-Apochromat 40X/1.4 Oil objective. Fluorescent images used for screening for GFP expression were captured using a Zeiss Axio Imager M2 compound microscope with a Plan-Apochromat 40X/1.4 Oil objective.

2.5.5 Quantitative real-time PCR

1 µg of total RNA was reverse transcribed with SuperScript IV Reverse Transcriptase (Invitrogen) in 1x reaction buffer, 2U SUPERase-In RNase Inhibitor (Invitrogen), 0.5 mM dNTPs, and 2.5 µM random hexamers. Each real-time PCR reaction consisted of 3 µL of cDNA, 1 µM forward gene-specific primer and 1 µM reverse gene-specific. The amplification was performed using iTaq Universal SYBR Green Supermix (Bio-Rad) on the Bio-Rad CFX96 Touch Real-Time PCR Detection System. The experiments were repeated for a total of three technical replicates.

2.5.6 Small RNA sequencing

Library preparation. Total RNA was extracted from whole animals of ~100,000 synchronized young adults. Small (<200nt) RNAs were enriched with mirVana miRNA Isolation Kit (Ambion). In brief, 80 µL (200-300 µg) of total RNA, 400 µl of mirVana lysis/binding buffer and 48 µL of mirVana homogenate buffer were mixed well and incubated at room temperature for 5 minutes. Then 176 µL of 100% ethanol was added and samples were spun at 2500 x g for 4 minutes at room temperature to pellet large (>200nt) RNAs. The supernatant was transferred to a new tube and small (<200nt) RNAs were precipitated with pre-cooled isopropanol at -70°C. Small RNAs were pelleted at 20,000 x g at 4°C for 30 minutes, washed once with 70% pre-cooled ethanol, and dissolved with nuclease-free water. 10 µg of small RNAs were fractionated on a 15% PAGE/7M urea gel, and RNA from 17 nt to 40 nt was excised from the gel. RNA was extracted by soaking the gel in 2 gel volumes of NaCl TE buffer (0.3 M NaCl, 10 mM Tris-HCl, 1 mM EDTA pH 7.5) overnight. The supernatant was collected through a gel filtration column. RNA was precipitated with isopropanol, washed once with 70% ethanol, and resuspended with 15 µL nuclease-free water. RNA samples were treated with RppH to convert 22G-RNA 5' triphosphates to monophosphates in 1x reaction buffer, 10U RppH (New England Biolabs), and

20U SUPERase-In RNase Inhibitor (Invitrogen) for 3 hours at 37°C, followed by 5 minutes at 65°C to inactivate RppH. RNA was then concentrated with the RNA Clean and Concentrator-5 Kit (Zymo Research). Small RNA libraries were prepared according to the manufacturer's protocol of the NEBNext Multiplex Small RNA Sample Prep Set for Illumina-Library Preparation (New England Biolabs). NEBNext Multiplex Oligos for Illumina Index Primers were used for library preparation (New England Biolabs). Libraries were sequenced using an Illumina HiSeq4000 to obtain single-end 50 nt sequences at the University of Chicago Genomic Facility.

Analysis. Fastq reads were trimmed using custom perl scripts. Trimmed reads were aligned to the *C. elegans* genome build WS230 using bowtie ver 1.2.1.1 [84] with options -v 0 --best --strata. After alignment, reads that were between 17-40 nucleotides in length were overlapped with genomic features (rRNAs, tRNAs, snoRNAs, miRNAs, piRNAs, protein-coding genes, pseudogenes, transposons) using bedtools intersect [123]. Sense and antisense reads mapping to individual miRNAs, piRNAs, protein-coding genes, pseudogenes, RNA/DNA transposons, simple repeats, and satellites were totaled and normalized to reads per million (RPM) by multiplying be 1e6 and dividing read counts by total mapped reads, minus reads mapping to structural RNAs (rRNAs, tRNAs, snoRNAs) because these sense reads likely represent degraded products. Reads mapping to multiple loci were penalized by dividing the read count by the number of loci they perfectly aligned to. Reads mapping to miRNAs and piRNAs were only considered if they matched to the sense annotation without any overlap. In other words, piRNA and miRNA reads that contained overhangs were not considered as mature piRNAs or miRNAs respectively. piRNA precursors were defined as sequences containing a full mature piRNA sequence plus a 2 nucleotide 5' overhang corresponding to the genomic sequence 2 nucleotides upstream of that piRNA's mature 5' end. 22G-RNAs were defined as 21 to 23 nucleotide long reads with a 5'G that aligned

to protein-coding genes, pseudogenes, or transposons. RPM values were then used in all downstream analyses using custom R scripts using R version 4.0.0 [124], which rely on packages ggplot2 [174], reshape2 [173], ggpubr [71], dplyr [175].

2.5.7 Cap-dependent RNA sequencing

Library preparation. Total RNA isolation was performed as described above for small RNA sequencing. Capped RNA molecules were enriched as described previously [60]. Terminator exonuclease (Epicentre) was used to selectively degrade monophosphorylated RNAs. Quick CIP (New England Biolabs) was used to dephosphorylate non-capped triphosphorylated RNAs. RppH was used with Thermopol Buffer (New England Biolabs) to decap RNAs for ligation. Libraries were prepared according to the manufacturer's protocol of the NEBNext Multiplex Small RNA Sample Prep Set for Illumina-Library Preparation (New England Biolabs). NEBNext Multiplex Oligos for Illumina Index Primers were used for library preparation (New England Biolabs). Libraries were sequenced using an Illumina NovaSeq6000 to obtain single-end 100 nt sequences at the University of Chicago Genomic Facility.

Analysis. Adaptor trimming and alignment were performed as described above for small RNA sequencing. Reads that were between 17-100 nucleotides were retained. Precursor length histograms were constructed based on lengths of unique piRNA precursor sequences (molecules containing a full mature piRNA sequence plus the 2 nucleotides upstream of the mature piRNA 5' end, as described above) from each library. Success of cap-dependent isolation was determined by comparing the enrichment for piRNA precursors versus mature piRNAs in the cap-dependent libraries compared to the previously constructed traditional small RNA libraries.

2.7 Supplementary Information

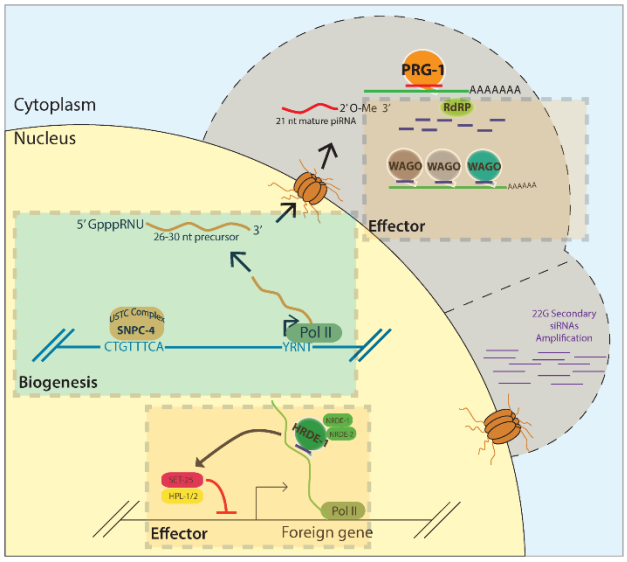


Figure 2.8: piRNA reporter screening strategy.
A model depicts the cellular context of biogenesis and effector steps in piRNA silencing.

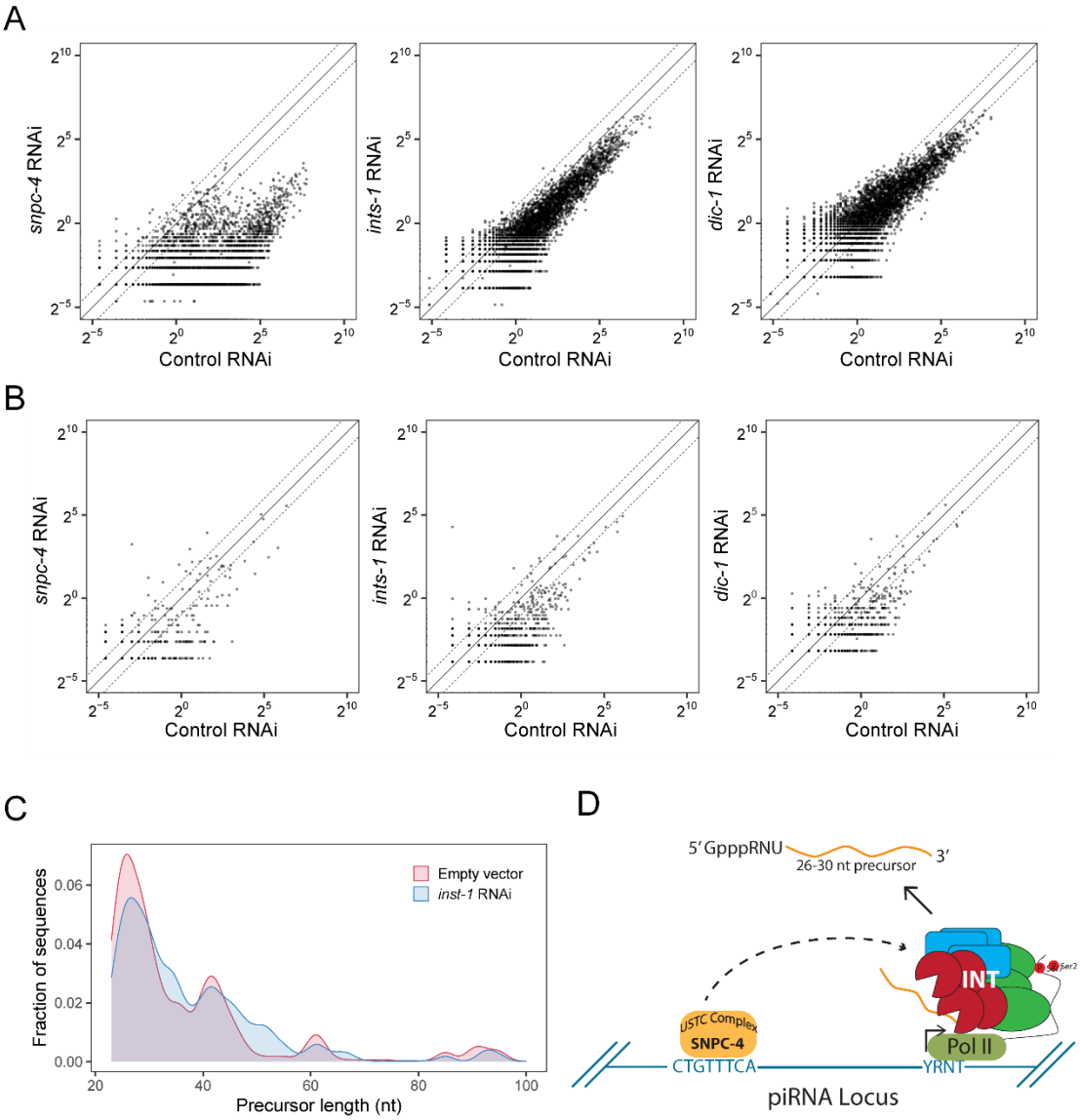


Figure 2.9: Loss of piRNAs following *snpc-4*, *ints-1*, and *dic-1* knockdown.
(A) Scatterplots show the expression of mature type I piRNAs in the indicated RNAi knockdowns. (B) Scatterplots show the expression of type I piRNA precursors in the indicated RNAi knockdowns. (C) A histogram shows the distribution of type I and type II piRNA precursor sequence lengths in cap-enriched sequencing libraries from *ints-1* and empty vector RNAi treated worms. (D) A model shows the proposed action of the Integrator complex at piRNA loci.

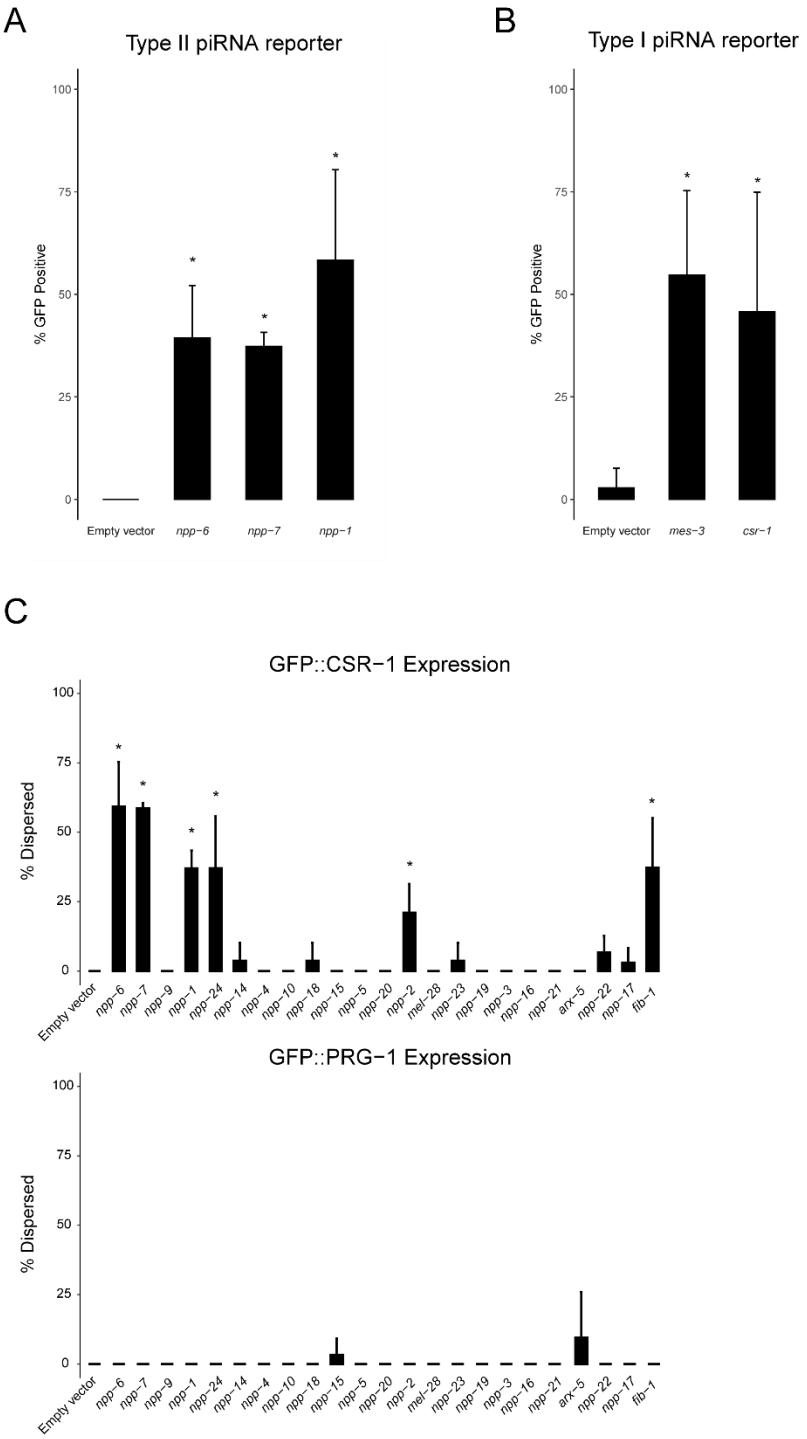


Figure 2.10: NPP-1/6/7/9 and CSR-1 pathway disruption activates the sensitized piRNA reporter. (A) GFP expression in type II piRNA reporter in indicated RNAi knockdowns. Asterisks indicate significant reporter activation, see Methods for details. (B) GFP expression in type I piRNA reporter in indicated RNAi knockdowns. Asterisks indicate significant reporter activation, see Methods for details. (C) Percentage of worms with dispersed GFP::CSR-1 (top) or GFP::PRG-1 (bottom) granules in the indicated RNAi knockdowns. Asterisks indicate $p < 0.05$ in one-tailed Student's t-test.

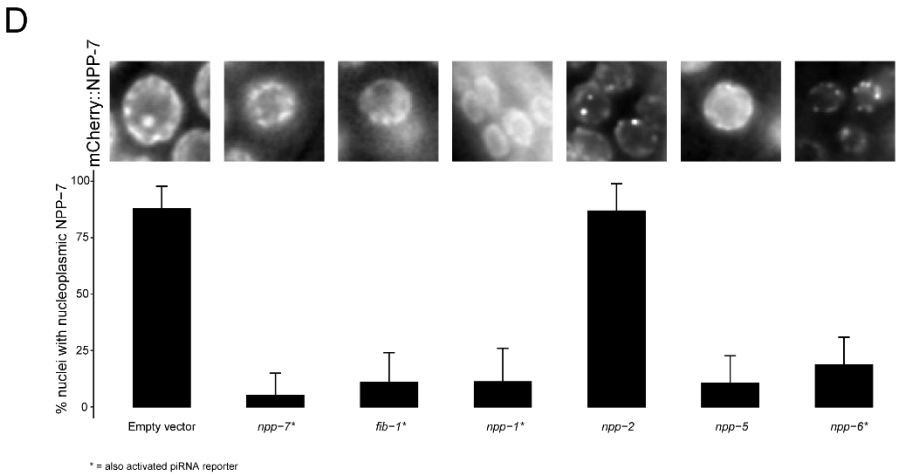


Figure 2.10 (continued): (D) Percentage of germline nuclei with nucleoli that contain NPP-7 following the indicated RNAi knockdowns. Representative fluorescent micrographs show mCherry::NPP-7 localization from indicated treatments. Treatments that also activate the piRNA reported are marked with an asterisk.

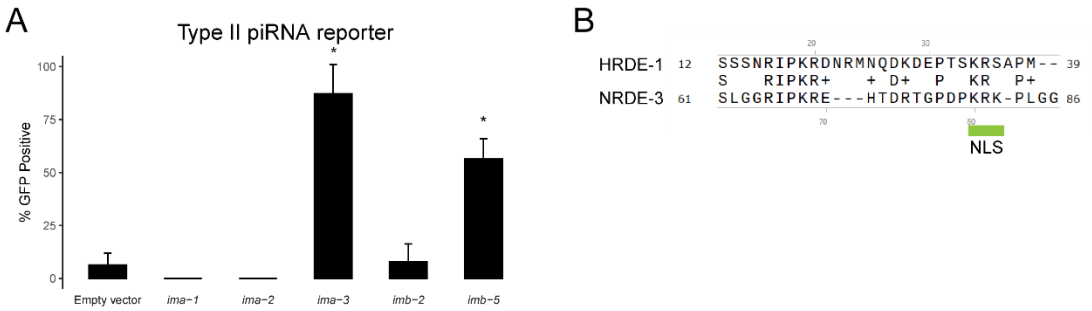


Figure 2.11: HRDE-1 does not share NRDE-3's nuclear localization signal.
(A) GFP expression in type II piRNA reporter in indicated RNAi knockdowns. Asterisks indicate significant reporter activation, see Methods for details. (B) A selection from the Needleman-Wunsch global alignment of HRDE-1 and NRDE-3 amino acid sequences that includes the NRDE-3 nuclear localization signal.

Gene name	Locus name	Wormbase ID	Description	% GFP expression	total worms screened	technical replicates	RNAi strategy	Phenotype
air-1	K07C11.2	WBGene00000098	Cell cycle, P granule regulator	0	5	1	L4	Low broods
arx-5	Y37D8A.1	WBGene00000203	Nuclear pore component	0	33	1	L4	Some dead embryos
bir-1	T27F2.3	WBGene00000249	Cell cycle, P granule regulator	0	12	1	L4	Possibly dumpy
cdc-25.1	K06A5.7	WBGene00000386	Cell cycle, P granule regulator	43.9	41	2	L1	Odd germ cell appearance, cell fate transformation
cdl-1	R06F6.1	WBGene00000411	RNAi resistant	89.3	28	1	L4	Poorly developed germline
cpf-2	F56A8.6	WBGene00000774	RNAi resistant	0	23	1	L4	Burst vulvas
crn-3	C14A4.4	WBGene00000796	Exosome component	0	7	1	L1	Slow development
dcs-1	Y113G7A.9	WBGene00000940	Decapping enzyme homologue	0	32	1	L4	No obvious phenotype
dic-1/ints-6	F08B4.1	WBGene00000994	Integrator complex	67.7	48	4	L4	Poorly developed germline

Table 2.1: Full list of candidates screened using piRNA reporter.
GFP expression and phenotypic notes following knockdown of genes tested with piRNA reporter.

Gene name	Locus name	Wormbase ID	Description	% GFP expression	total worms screened	technical replicates	RNAi strategy	Phenotype
egl-45	C27D11.1	WBGene00001209	Translation initiation, P granule	0	11	1	L2	Very sick, poorly developled germline
eif-3.F	D2013.7	WBGene00001229	Translation initiation, P granule	0	24	1	L2	Poorly developed germline
fib-1	T01C3.7	WBGene00001423	nucleolar protein	89.4	41	4	L4	Poorly developed germline
snpc-4	F32H2.1	WBGene00001568	piRNA biogenesis factor	92	>150	>10	L1	Poorly developed germline
hpl-2	K01G5.2	WBGene00001996	HP1-like protein	25.1	31	3	L4	Slightly delayed development
ima-1	T19B10.7	WBGene00002072	Nuclear transport	7.9	28	2	L4	Some dead embryos
ima-2	F26B1.3	WBGene00002073	Nuclear transport	12.5	28	2	L4	Poorly developed germline, dead embryos
ima-3	F32E10.4	WBGene00002074	Nuclear transport	100	30	2	L4	Poorly developed germline, dead embryos
imb-2	R06A4.4	WBGene00002076	Nuclear transport	36.1	29	2	L4	Poorly developed germline, dead embryos

Table 2.1 (continued): All worms were screened as adults or 3 days post-L1 if adulthood was difficult to determine.

Gene name	Locus name	Wormbase ID	Description	% GFP expression	total worms screened	technical replicates	RNAi strategy	Phenotype
imb-4/xpo-1	ZK742.1	WBGene00002078	Nuclear transport	0	19	1	L4	Poorly developed germline
imb-5/xpo-2	Y48G1A.5	WBGene00002079	Nuclear transport	67.6	36	2	L4	Poorly developed germline, dead embryos
mcm-2	Y17G7B.5	WBGene00003154	P granule regulator	0	8	1	L4	Sick, low broods
unc-32	ZK637.8	WBGene00003154	P granule regulator	0	9	1	L4	Motility defects
mel-28	C38D4.3	WBGene00003210	Nuclear pore component	0	9	1	L4	Some dead embryos
mes-3	F54C1.3	WBGene00003221	H3K27 histone methyl	72.7	22	1	L4	Poorly developed germline, dead embryos
mes-4	Y2H9A.1	WBGene00003222	H3K36 histone methyl	54.6	17	1	L4	Slightly reduced brood
mog-1	K03H1.2	WBGene00003389	pre-mRNA splicing	0	28	3	L1	Very pooly developed germline
mog-4	C04H5.6	WBGene00003392	pre-mRNA splicing	0	18	1	L4	Dead embryos

Table 2.1 (continued): RNAi strategy for each gene is noted. The default RNAi strategy of L4 treatment and screening in subsequent generation is indicated with "L4". If this treatment procedure resulted in early developmentally arrested

Gene name	Locus name	Wormbase ID	Description	% GFP expression	total worms screened	technical replicates	RNAi strategy	Phenotype
mog-5	EEED8.5	WBGene00003393	pre-mRNA splicing	0	21	3	L1	Very pooly developed germline
dcap-2	F52G2.1	WBGene00003582	Decapping enzyme, present in P	0	37	1	L4	Poorly developed germline
npp-1	K07F5.13	WBGene00003787	Nuclear pore component	57.9	23	3	L4	Dead embryos, poorly developed germline
npp-2	T01G9.4	WBGene00003788	Nuclear pore component	3.7	27	3	L4	No embryos, poorly developed germline
npp-3	K12D12.2	WBGene00003789	Nuclear pore component	3.3	25	3	L4	Dead embryos, poorly developed germline
npp-4	Y54E5A.4	WBGene00003790	Nuclear pore component	0	26	3	L4	Dead embryos, poorly developed germline
npp-5	F07A11.3	WBGene00003791	Nuclear pore component	0	33	3	L4	Some dead embryos
npp-6	F56A3.3	WBGene00003792	Nuclear pore component, P granule	67.5	40	4	L4	Poorly developed germline
npp-7	T19B4.2	WBGene00003793	Nuclear pore component, P granule	77.4	85	6	L4	Poorly developed germline

Table 2.1 (continued): progeny, then an alternative strategy was used. Treatment from L1 to adulthood is indicated with "L1". Treatment from L2 to adult is indicated with "L2".

Gene name	Locus name	Wormbase ID	Description	% GFP expression	total worms screened	technical replicates	RNAi strategy	Phenotype
npp-9	F59A2.1	WBGene00003795	Nuclear pore component, P granule	58.1	43	4	L4	Some dead embryos
npp-10	ZK328.5	WBGene00003796	Nuclear pore component, P granule	9.1	28	3	L4	Some dead embryos
npp-14	C03D6.4	WBGene00003800	Nuclear pore component	8.9	29	3	L4	Some dead embryos
npp-15	C29E4.4	WBGene00003801	Nuclear pore component	0	30	3	L4	Some dead embryos
npp-16	Y56A3A.17	WBGene00003802	Nuclear pore component	6.7	28	3	L4	Poorly developed germline
npp-17/rae-1	F10G8.3	WBGene00003803	Nuclear pore component	0	31	1	L4	Some dead embryos
npp-18	Y43F4B.4	WBGene00003804	Nuclear pore component	6.7	26	3	L4	Some dead embryos
npp-19	R06F6.5	WBGene00003805	Nuclear pore component	0	28	3	L4	Some dead embryos
npp-20	Y77E11A.13	WBGene00003806	Nuclear pore component, P granule	10	21	3	L4	Some dead embryos

Table 2.1 (continued)

Gene name	Locus name	Wormbase ID	Description	% GFP expression	total worms screened	technical replicates	RNAi strategy	Phenotype
pas-4	C36B1.4	WBGene00003925	Proteasome, P granule regulator	0	10	1	L2	Very poor development, no obvious germ cells
pas-5	F25H2.9	WBGene00003926	Proteasome, P granule regulator	100	8	1	L2	Sterile, possible cell fate transformation
pbs-3	Y38A8.2	WBGene00003949	Proteasome, P granule regulator	0	16	1	L2	Poorly developed germline
prp-21	W07E6.4	WBGene00004188	pre-mRNA splicing, RNAi	56.7	30	3	L4	Dead embryos
rsp-7	D2089.1	WBGene00004704	pre-mRNA splicing, RNAi	0	27	1	L4	Poorly formed embryos
smu-2	Y49F6B.4	WBGene00004896	pre-mRNA splicing	0	12	1	L4	No obvious phenotype
uaf-2	Y116A8C.35	WBGene00006698	pre-mRNA splicing	0	14	1	L4	Poorly developed germline
npp-22	B0240.4	WBGene00007118	Nuclear pore component	0	29	1	L4	Poorly developed germline
npp-23	C09G9.2	WBGene00007493	Nuclear pore component	0	30	3	L4	~25% worms with poorly developed germline

Table 2.1 (continued)

Gene name	Locus name	Wormbase ID	Description	% GFP expression	total worms screened	technical replicates	RNAi strategy	Phenotype
nrde-1	C14B1.6	WBGene00007577	hrde-1 cofactor	34.2	32	3	L4	No obvious phenotype
prp-4	C36B1.5	WBGene00007972	pre-mRNA splicing	5.6	33	3	L4	Dead embryos
ints-7	D1043.1	WBGene00008361	Integrator complex	0	47	3	L4	Poorly developed germline
E02H1.1	E02H1.1	WBGene00008455	RNAi resistant	0	26	1	L4	Mostly no embryos
ints-11	F10B5.8	WBGene00008642	Integrator complex	71.4	38	4	L1	Poorly developed germline, dead embryos
ncbp-2	F26A3.2	WBGene00009141	pre-mRNA splicing, RNAi	0	27	1	L4	No embryos
snrp-27	R05D11.7	WBGene00011035	pre-mRNA splicing	0	15	1	L4	Slightly delayed development
paa-1	R186.7	WBGene00011308	PP2A subunit A, RNAi	0	30	1	L2	Very poor development, no obvious germ cells
nrde-2	T01E8.5	WBGene00011333	hrde-1 cofactor	44.3	22	3	L4	No obvious phenotype

Table 2.1 (continued)

Gene name	Locus name	Wormbase ID	Description	% GFP expression	total worms screened	technical replicates	RNAi strategy	Phenotype
T25G3.3	T25G3.3	WBGene00012030	rRNA processing, RNAi	0	26	1	L1	No embryos, some burst vulvas
tofu-5	W01A8.5	WBGene00012167	piRNA biogenesis factor	56.9	29	3	L4	Some dead embryos
ints-4	W04A4.5	WBGene00012234	Integrator complex	0	45	3	L4	Very poor development, poorly developed
set-25	Y43F4B.3	WBGene00012802	H3K9 histone methyl	35.9	28	3	L4	No obvious phenotype
ints-5	Y51A2D.7	WBGene00013075	Integrator complex	0	30	3	L4	Very poor development, poorly developed
prp-40	ZK1098.1	WBGene00014218	pre-mRNA splicing	6.7	37	3	L4	Dead embryos
ints-1	C06A5.1	WBGene00015499	Integrator complex	68.8	50	4	L4	Poorly developed germline
ddx-23	F01F1.7	WBGene00017162	pre-mRNA splicing	6.1	32	3	L4	Some dead embryos
cpsf-2	F09G2.4	WBGene00017313	mRNA processing, RNAi	0	23	1	L4	Burst vulvas

Table 2.1 (continued)

Gene name	Locus name	Wormbase ID	Description	% GFP expression	total worms screened	technical replicates	RNAi strategy	Phenotype
ints-9	F19F10.12	WBGene00017608	Integrator complex	63.6	32	4	L1	Poorly developed germline, dead embryos
tofu-2	F20A1.9	WBGene00017620	piRNA biogenesis factor	20.4	31	3	L4	Some dead embryos
csr-1	F20D12.1	WBGene00017641	P granule enriched Argonaute	45.7	36	3	L4	Very poor development, poorly developed
prp-17	F49D11.1	WBGene00018625	pre-mRNA splicing, RNAi	77.4	27	3	L4	Poorly developed germline
pup-2	K10D2.2	WBGene00019628	polyU polymerase	0	7	1	L1	Poorly developed germline
sago-1	K12B6.1	WBGene00019666	RNAi resistant	0	26	1	L4	Some dead embryos
K12H4.5	K12H4.5	WBGene00019680	RNAi resistant	0	31	1	L4	Dead embryos
ints-13	R02D3.4	WBGene00019822	Integrator complex	0	40	3	L4	No obvious phenotype

Table 2.1 (continued)

Gene name	Locus name	Wormbase ID	Description	% GFP expression	total worms screened	technical replicates	RNAi strategy	Phenotype
nrde-3	R04A9.2	WBGene00019862	somatic nuclear Argonaute	2.8	34	3	L4	No obvious phenotype
npp-21	R07G3.3	WBGene00019940	Nuclear pore component	0	31	1	L4	Dead embryos, poorly developed germline
T19B4.5	T19B4.5	WBGene00020558	RNAi resistant	0	28	1	L4	Dead embryos
ints-12/phf-30	T23B12.1	WBGene00020716	Integrator complex	0	45	3	L4	No obvious phenotype, slightly lower broods
npp-24	ZK177.4	WBGene00022672	Nuclear pore component	3.3	32	3	L4	No obvious phenotype
ZK1127.5	ZK1127.5	WBGene00022852	rRNA processing, RNAi	0	28	1	L2	No obvious germ cells

Table 2.1 (continued)

CHAPTER 3

THE ROLE OF PERINUCLEAR GERM GRANULES IN PIRNA-MEDIATED TRANSCRIPTOME SURVEILLANCE

3.1 Attributions

This chapter has been adapted from: W. Chen & J.S. Brown, *et al.*, GLH/VASA helicases promote germ granule formation to ensure the fidelity of piRNA-mediated transcriptome surveillance, Nat. Commun. 13 (2022). Wenjun Chen and the author jointly performed live worm confocal microscopy experiments. Wenjun Chen prepared most small RNA and mRNA libraries for next generation sequencing. Donglei Zhang performed image quantification in live worm confocal experiments. Tao He performed the IP-MS analyses of GLH-1 complex. The author conducted the other experiments and analyses.

3.2 Abstract

piRNAs function as guardians of the genome by silencing non-self nucleic acids and transposable elements in animals. Many piRNA factors are enriched in perinuclear germ granules, but whether their localization is required for piRNA biogenesis or function is not known. Here we show that GLH/VASA helicase mutants exhibit defects in forming perinuclear condensates containing PIWI and other small RNA cofactors. These mutant animals produce largely normal levels of piRNA but are defective in triggering piRNA silencing. Strikingly, while many piRNA targets are activated in GLH mutants, we observed that hundreds of endogenous genes are aberrantly silenced by piRNAs. This defect in self versus non-self recognition was also observed in other mutants where perinuclear P granules are disrupted. Together, our results argue that perinuclear germ granules function critically to promote the fidelity of piRNA-based transcriptome surveillance in *C. elegans* and preserve self versus non-self distinction.

3.3 Introduction

Argonaute proteins use their associated small RNAs as guides to regulate targets with complementary sequences [67]. The PIWI Argonaute and its associated piRNAs are conserved guardians of the animal genome that repress transposons in the germline [15,21,27,40,131]. A prerequisite for any defense system is the ability to distinguish non-self from self. In *C. elegans,* piRNAs trigger gene silencing of non-self RNAs through the recruitment of RNA-dependent RNA Polymerases (RdRPs) to produce WAGO Argonaute-associated 22G-RNAs that mediate transcriptional and posttranscriptional gene silencing [61,141]. While diverse PIWI/piRNAs can recognize both foreign nucleic acids and germline-expressed "self" mRNAs [137,185], self RNAs are protected from piRNA silencing by Argonaute CSR-1 and its associated 22G-RNAs [133,134,170].

Intriguingly, the PIWI-related PRG-1, WAGO-1 and CSR-1 Argonautes are all enriched in germ granules, also known as P granules in *C. elegans* [15,33,61,149]. Germ (P) granules are phase-separated liquid droplets that are found in the germ cells of all animals [20,46]. The localization and formation of P granules are tightly controlled during *C. elegans* development [135,162]. In early embryos, P granules are cytoplasmic and sort to daughter cells of the germ cell lineage. As zygotic transcription begins later in embryogenesis, P granules re-localize to the nuclear periphery and remain perinuclear throughout most of germline development**.** Mutations that affect the formation of cytoplasmic P granules in embryos, such as mutants for germ plasm factors *meg-3 meg-4*, impact the potency and inheritance of RNAi [43,89,112]. However, *meg-3 meg-4* mutant adults have normal perinuclear P granules. The role perinuclear P granules play in small RNA-mediated gene regulation is less clear. Several lines of evidence suggest that perinuclear P granules may be the site of mRNA surveillance by small RNAs. First, perinuclear P

granules have been shown to be the major sites of mRNA transport in germline nuclei [138]. Second, the size of perinuclear P granules shrink soon after inhibition of mRNA transcription or mRNA export [31,138], consistent with the model newly exported mRNAs gather in perinuclear P granules. Third, a recent report showed that tethering an mRNA to P granule component PGL-1 leads to its silencing [5], suggesting the accumulation of mRNA in P granules can trigger silencing by small RNA pathways. In addition, the enzymes required for piRNA 3' end trimming (PARN-1) and the biogenesis of CSR-1 and WAGO associated 22G-RNAs (EGO-1) are both enriched in perinuclear P granules [118,152]. Together, these observations raise the possibility that the production and/or function of small RNAs may require the enrichment of these factors in perinuclear P granules.

The VASA-homologue RNA helicases GLH-1 and GLH-4 play a critical role in the formation of both cytoplasmic and perinuclear P granules [31,59,96,129,147]. In addition, several other P granule factors, including DEPS-1 and PGL-1, have been reported to promote both cytoplasmic and perinuclear P granule assembly [63,147]. These observations make them possible candidates as arbiters for examining P granule function in small RNA mediated gene silencing. Here we show that GLH-1 and GLH-4 play a global role in promoting the liquid condensation of Argonautes and other small RNA factors at perinuclear foci. In addition, we found that the biogenesis of neither piRNAs nor secondary small RNAs, including WAGO or CSR-1 associated 22G-RNAs, broadly require GLH/VASA. In GLH and in other mutants with defects in forming perinuclear P granules, many piRNA targets were activated, with fewer secondary WAGO 22G-RNAs produced at piRNA targeting sites. Additionally, many functional endogenous mRNAs are aberrantly silenced by piRNAs. Together, our results suggest that GLH/VASA helicases and perinuclear P granules are critical for ensuring the fidelity of mRNA surveillance by piRNAs and

that without P granules, small RNA pathways can no longer robustly identify mRNAs as self or non-self.

3.4 Results

3.3.1 GLH/VASA promotes the condensation of piRNA pathway factors

While GLH-1 plays a critical role in controlling the perinuclear localization of PIWI PRG-1 [31,96], we failed to detect PRG-1 in GLH-1 complexes by mass spectrometry under native conditions [31]. We hypothesized that their interaction could be transient and therefore applied the chemical crosslinking reagent dithio-bis-maleimidoethane (DTME) to capture potentially transient interactions [52]. Indeed, using DTME-crosslinked worms, we are able to detect PRG-1 in the GLH-1 complex (Figure 3.1A). In addition, we observed many other small RNA components in the GLH-1 complex, including P granule factors DEPS-1, WAGO-1, CSR-1, and Z granule factor WAGO-4 (Figure 3.1A and Supplementary Data file 3.1). These findings are consistent with previously reported GLH-1 mass spectrometry results [96]. Z granules are derived from P granules during embryogenesis and remain adjacent to P granules in the adult germline [68,169]. Similarly, an interaction between P granule and Z granule factors has previously been shown by PRG-1 mass spectrometry; PRG-1 itself interacts with factors found in both P granules and Z granules [14]. Together with the previous reported results [96], these observations raise the possibility that GLH helicases may play a global role in regulating the localization of small RNA pathway components. As VASA-like helicases GLH-1 and GLH-4 function redundantly to promote the localization of PIWI PRG-1 [31], we examined the localization of various small RNA machinery in the *glh-1 glh-4* mutant. Consistent with previous findings, we found that for PIWI PRG-1, both perinuclear and cytoplasmic foci are greatly reduced in the *glh-1 glh-4* double null mutant (Figure 3.1B and Supplementary Figure 3.7A). For several other small RNA factors, including P granule factors -

DEPS-1, CSR-1, and EGO-1, and Z granule factors - WAGO-4 and ZFNX-1, their perinuclear localization is also significantly reduced in the *glh-1 glh-4* double mutant, although residual foci of some of these factors can still be observed, specifically for CSR-1 where the overall level of perinuclear CSR-1 is reduced while the number of CSR-1 foci is similar to that in wild type animals (Figure 3.1A, C and Supplementary Figure 3.7A, B). We then examined the localization of MUT-16, a key factor in the assembly of Mutator granules [118]. Mutator granules house the small RNA components involved in producing WAGO associated 22G-RNAs. Mutator granules are frequently found in close contact with perinuclear P granules [159]. Previous experiments using RNAi knockdown of *glh-1* and *glh-4* did not lead to disruption of MUT-16 localization [118]. The localization of MUT-16 was significantly disrupted in *glh-1 glh-4* double mutants (Figure 3.1D and Supplementary Figure 3.7C). This is consistent with a recent report demonstrating MUT-16 disruption upon simultaneous RNAi treatment against *glh-1*, *glh-4*, *pgl-1*, and *pgl-3* [144]. Taken together, these results show that GLH/VASA helicases play a global role in enriching small RNA machinery into the distinct liquid condensates observed throughout *C. elegans* germline development, including cytoplasmic and perinuclear P granules, Z granules and Mutator granules.

A

Proteomic Analysis of GLH-1 Complex					
Proteins	Peptide Counts in Indicated Strains				
		mCherry::glh-1		Control	
	Replicate	1	2	1	2
GLH-1		40	27	0	0
PRG-1		11	2	0	0
DEPS-1		6	4	0	0
WAGO-1		18	6	0	0
CSR-1		5	1	0	0
WAGO-4		2	1	0	0

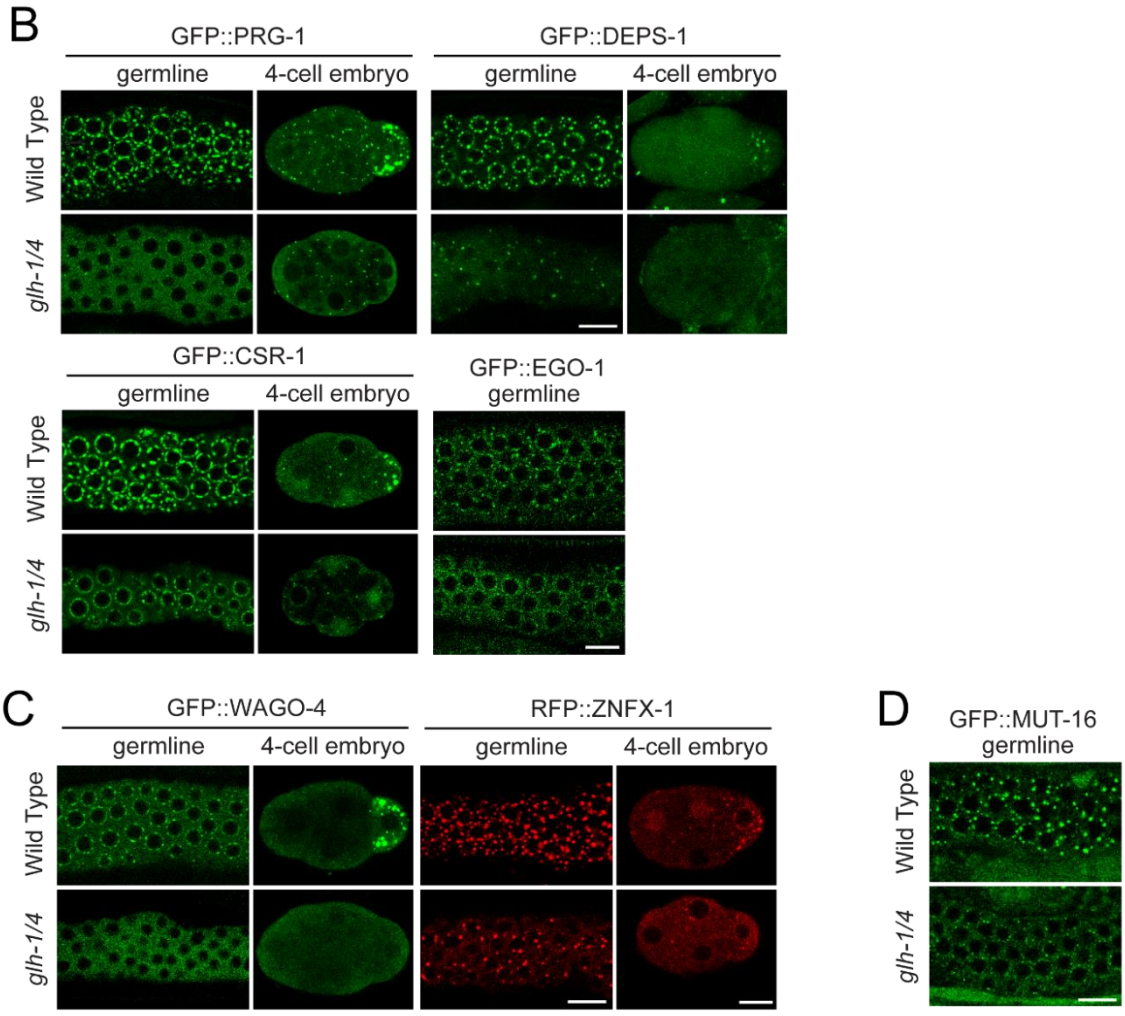


Figure 3.1: The GLH/VASA helicases GLH-1 and GLH-4 promote the localization of germ granule factors.
(A) Proteomic analyses of GLH-1 complex. The numbers of peptides identified in two independent pull-down experiments are shown. (B) Fluorescent micrographs show the localization of the specific proteins known to be enriched in P granules in the wild type or the *glh-1 glh-4* mutant in the adult germline (left) and in the four-cell embryo (right). Bar, 10 micrometers.

Figure 3.1 (continued): (C) Fluorescent micrographs show the localization of the specific proteins known to be enriched in Z granules in the wild type or the *glh-1 glh-4* mutant in the adult germline (left) and in the four-cell embryo (right). Bar, 10 micrometers. (D) Fluorescent micrographs show the localization of the specific proteins known to be enriched in mutator granules in the wild type or the *glh-1 glh-4* mutant in the adult germline. Bar, 10 micrometers.

3.3.2 GLH/VASA helicase mutants exhibit defects in piRNA silencing

To investigate whether GLH-1 and GLH-4 helicases are required for piRNA-mediated gene silencing, we examined whether the silencing of a piRNA reporter [134] requires these GLH/VASA helicases. This piRNA reporter is silenced in wild type animals but is activated in the *prg-1* mutant background (Figure 3.2A). Similarly, we found that the piRNA reporter is activated in the *glh-1 glh-4* double mutant background (Figure 3.2A)**,** suggesting GLH-1 and GLH-4 play a role in piRNA silencing. In addition, we found that the piRNA reporter is also activated in the *glh-1 DQAD* mutant background*.* Previous studies have shown that in *glh-1 DQAD* mutants*,* P granule factors form large aggregates and exhibit defects in the distribution of these granules in the early embryo [31,96]. Indeed, in the *glh-1 DQAD* mutant, large PRG-1 and WAGO-4 aggregates are found in the cytoplasm with a significant reduction in perinuclear PRG-1 and WAGO-4 foci (Figure 3.2B). In addition, these abnormal, cytoplasmic aggregates are not properly sorted to the germ cell lineage and/or not properly degraded in somatic lineages, leading to the presence of these foci in somatic lineages (Figure 3.2B). Together, these data show that GLH mutants are defective in piRNA silencing.

We then wanted to understand why piRNA silencing is defective in GLH/VASA mutants. We found no significant change in expression of the GFP-targeting piRNA in these GLH/VASA mutants compared to wild type animals (Figure 3.2C). In contrast, we observed a pronounced reduction in the 22G-RNAs produced around the GFP-targeting piRNA binding site in both the *glh-1 glh-4* double and *glh-1 DQAD* mutants (Figure 3.2D). These analyses suggest that GLH-1

and GLH-4 are not required for the biogenesis of the GFP-targeting piRNA, but rather promote the production of 22G-RNAs at the piRNA targeting site.

Since perinuclear P granules have been shown to be the major sites of mRNA export [138], we wondered whether GLH/VASA may also contribute to the localization of target mRNAs. We therefore examined the localization of GFP mRNAs, the piRNA target of our piRNA reporter, using single molecule fluorescent *in situ* hybridization (smFISH). In our piRNA reporter strain where GFP is silenced by a GFP-targeting piRNA, we observed large GFP perinuclear foci and little cytoplasmic GFP signal (Figure 3.2E). In the *glh-1 glh-4* double mutant piRNA reporter strain where GFP is activated, the perinuclear GFP foci were greatly reduced, while more cytoplasmic signal was observed (Figure 3.2E). To quantify the extent of silenced *gfp* mRNA colocalization with P granules, we measured the amount of colocalization of *gfp* mRNA signal with PRG-1 foci. Because PRG-1 perinuclear accumulation is reduced in *glh-1 glh-4* mutants, we also analyzed a piRNA reporter strain without the piRNA that triggers *gfp* silencing. We saw that compared to the wild type, non-silenced reporter and the *glh-1 glh-4* mutant reporter, the wild type silenced reporter showed significantly more *gfp* – PRG-1 colocalization (Supplementary Figure 3.8C). We did not observe this increase in colocalization when we monitored an mRNA that does not become silenced in the reporter, *nos-3.* Although this effect was significant, the extent of the colocalization was modest. This could be due to silenced *gfp* mRNA also accumulating in Mutator foci or Z granules. Consistent with this hypothesis, we noticed that many instances of *gfp* – PRG-1 colocalization were only partial, which could be indicative of overlap with granules adjacent to P granules. Therefore, GLH-1 and GLH-4 promotes the location of GFP mRNA at perinuclear foci in the piRNA reporter. These observations are consistent with the model that GLH/VASA

promotes the accumulation of PRG-1, piRNA cofactors, and target RNAs at perinuclear foci to trigger piRNA silencing.

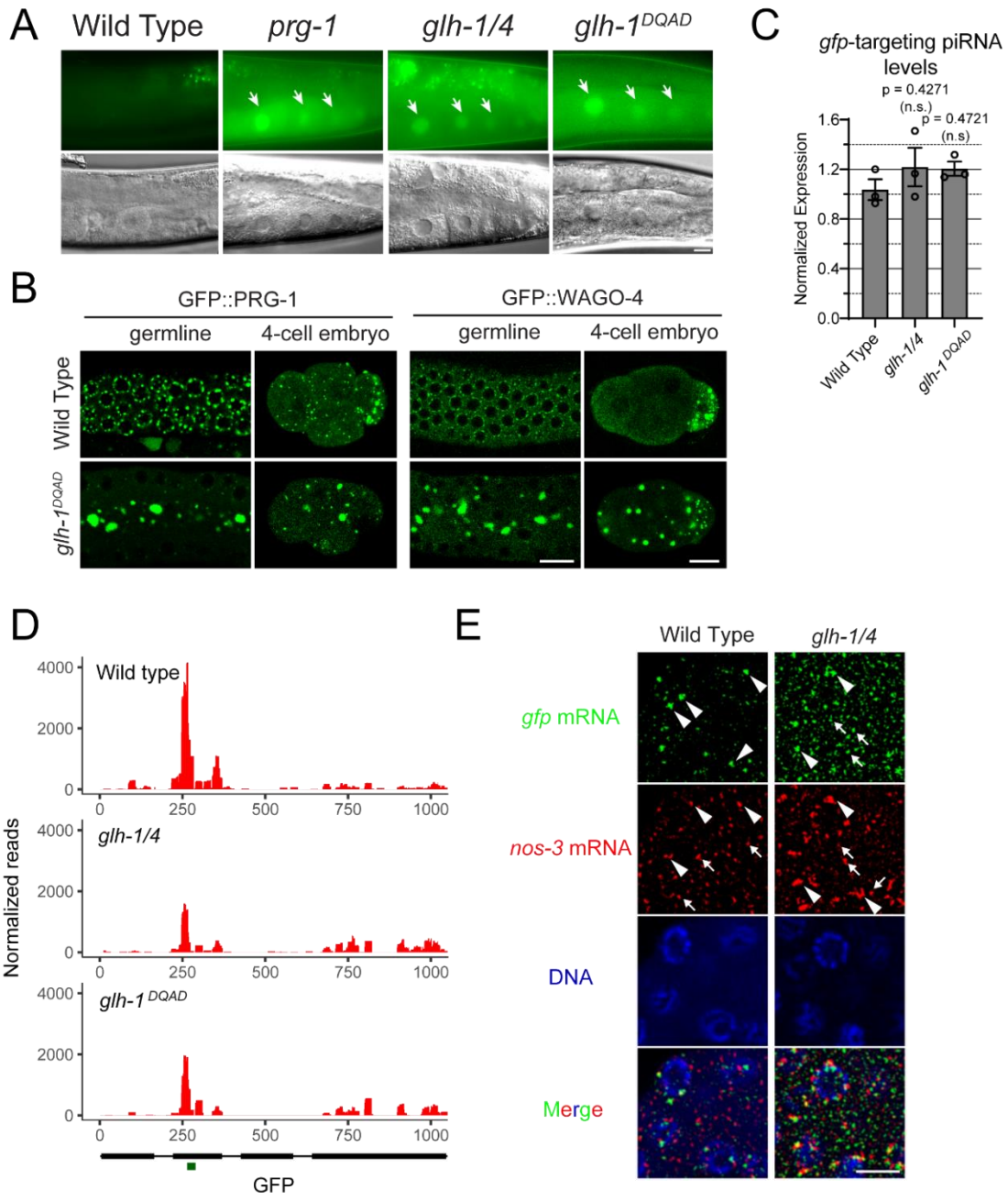


Figure 3.2: GLH/VASA helicases are required for piRNA-dependent silencing of a piRNA reporter.

(A) GFP expression of the piRNA reporter in the indicated strains. Arrows indicate the adult germline nuclei in maturing oocytes with expression of GFP transgenes (top). DIC images of the corresponding reporter images (bottom). Bar, 10 micrometers. (B) Fluorescent micrographs show the localization of PRG-1 and WAGO-4 in the indicated strains in the adult germline (left) and in the four-cell embryo (right). Bar, 10 micrometers. (C) q-RT PCR measurements of the abundance of the GFP-targeting piRNA in the indicated strains. Statistical analysis was performed using a two-tailed Student's t-test. Bars indicate the mean, errors bars indicate the standard deviation, and data points indicate values for the 3 technical replicates from each genotype.

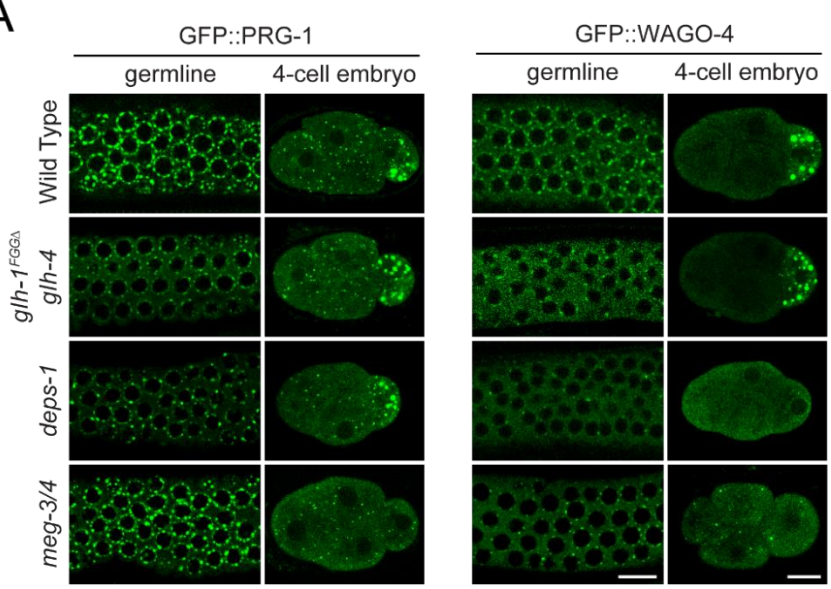
Figure 3.2 (continued): (D) 22G-RNAs distribution at GFP coding sequences in the indicated strains. The green bar (below) indicates the location of the GFP sequence complementary to the GFP-targeting piRNA. (E) Photomicrographs of pachytene nuclei in fixed adult gonads hybridized with single molecule fluorescent (smFISH) probes complementary to *gfp* mRNA (Green) and *nos-3* mRNA (Red) in the indicated strains. Nuclear DNA was stained with DAPI (Blue). The arrowheads indicate perinuclear mRNA foci while arrows indicate cytoplasmic mRNA foci. Bar, 5 micrometers. Photomicrographs are representative of data collected across 3 independent experiments. At least 10 gonads from each experiment were observed at random.

3.3.3 P granules promote piRNA silencing

The GLH mutants examined above are defective in the localization of small RNA machinery in both perinuclear and cytoplasmic P granules. To examine whether piRNA silencing capability correlates with the ability to form either type of liquid condensate, we characterized additional genetic mutants that have been reported to show defects in forming either cytoplasmic and/or perinuclear P granules. The P granule component DEPS-1 has been shown to promote the assembly of perinuclear and cytoplasmic P granules [147]. Indeed, the formation of both perinuclear and cytoplasmic WAGO-4 condensates are compromised in *deps-1* mutants (Figure 3.3A). However, we found *deps-1* mutants exhibit a reduced number of perinuclear PRG-1 condensates but retained normal cytoplasmic PRG-1 condensates. While we have not yet confirmed this finding using our piRNA reporter, a recent study reports that DEPS-1 is required for silencing of a piRNA reporter [150]. In addition, we have recently shown that the N terminal phenylalanine-glycine-glycine (FGG) repeats of GLH-1 promote its perinuclear localization, leading to the recruitment of PIWI PRG-1 at perinuclear P granules [31]. Indeed, we confirmed that PRG-1 and WAGO-4 perinuclear condensates, but not their cytoplasmic condensates, are partially disrupted in the *glh-1* FGGΔ *glh-4* mutant (Figure 3.3A). We also found that the piRNA reporter is activated in this *glh-1 FGGΔ glh-4* strain (Figure 3.3B). These results indicate that mutants defective in forming perinuclear P granules exhibit defects in piRNA silencing. We then examined the *meg-3 meg-4* mutant, in which the localization of the small RNA machinery is

disrupted in cytoplasmic P granules found in the early embryos but still exhibit wild-type like adult perinuclear P granules [112,146] (Figure 3.3A). We found that the piRNA reporter is also activated in ~30 percent of *meg-3 meg-4* mutant animals (Figure 3.3B). Together, these results indicate that the localization of piRNA factors at adult perinuclear P granules and embryonic cytoplasmic P granules both contribute to their function in piRNA silencing.

A



B

Strain Genotype	Perinuclear PRG-1 granules (adult)	Cytoplasmic PRG-1 granules (embryo)	piRNA sensor expression (%)
Wild Type	+++	+++	2
prg-1	-	-	100
glh-1/4	-	-	100
glh-1^{DQAD}	+*	+*	95
glh-1$^{FGG\Delta}$ *glh-4*	++	+++	95
deps-1	++	+++	N/A
meg-3/4	+++	-	30

n>50

Figure 3.3: Mutants with defective in either perinuclear or cytoplasmic P granules exhibit defects in gene silencing.
(A) Fluorescent micrographs show the localization of PRG-1 and WAGO-4 in the indicated strains in the adult germline (left) and in the four-cell embryo (right). Bar, 10 micrometers. Quantification of each genotype's effect on granule properties is provided in Supplementary Figure 3.2A. (B) GFP expression in the piRNA reporter and the ability to form perinuclear or cytoplasmic PRG-1 granules in the indicated strains.

3.3.4 Perinuclear P granules promote initiation of piRNA silencing

Our observations that GLH/VASA promotes the perinuclear localization of piRNA factors and their target mRNAs raises the possibility that their enrichment in P granules allows PIWI PRG-1 and its cofactors to efficiently identify their targets and to trigger gene silencing. To test whether

GLH/VASA mutants exhibit defects in *de novo* piRNA-mediated gene silencing, we microinjected a synthetic piRNA-expressing plasmid into transgenic worms expressing a silencing-prone GFP::CDK-1 transgene [185] and monitored the silencing of the GFP transgene by the GFP-targeting piRNA. While 97% of the injected wild type animals successfully triggered silencing of the GFP transgene by the GFP-targeting piRNA by the F2 generation, only 5% of the *glh-1 glh-4* double mutant animals and 36% of the *glh-1 DQAD* mutant animals triggered silencing of the GFP transgene (Supplementary Figure 3.9A). In addition, we observed that only 40% of the *deps-1* mutant animals and 31% of the *glh-1 FGGΔ glh-4* mutant animals triggered silencing of the GFP transgene. In contrast, *meg-3 meg-4* mutants exhibit a normal, wild type ability to trigger silencing of the GFP transgene by the GFP-targeting piRNA. Taken together, these assays show that mutants defective in forming perinuclear P granules are compromised in their ability to initiate piRNA silencing.

3.3.5 Perinuclear P granules promote silencing of piRNA targets

Our analyses of the piRNA reporter suggest that GLH/VASA is not required for the biogenesis of a GFP-targeting piRNA (Figure 3.2C). Consistent with the reporter analyses, we observed slightly reduced or normal piRNA levels in the *glh-1 glh-4* double mutant or in *glh-1 DQAD* mutant, respectively (Supplementary Figure 3.10A). In addition, we found that the length of piRNAs was not changed in the *glh-1 glh-4* double mutant or in the *glh-1 DQAD* mutant (Supplementary Figure 3.10B), suggesting that the 3' processing of piRNAs by PARN-1 is not affected. On the other hand, over 41% of WAGO targets (genes known to be silenced by WAGO 22G-RNAs) exhibit a two-fold decrease in 22G-RNAs in the *glh-1 glh-4* double mutant and over 51% in the *glh-1 DQAD* mutant (Figure 3.4A, left and Supplementary Figure 3.10C) [61]. To confirm that the reduction in 22G-RNAs is due to decreased production of WAGO-associated 22G-RNAs, we performed an

immunoprecipitation with HRDE-1 (WAGO-9) and compared wild type HRDE-1 22G-RNAs to those in *glh-1* mutants, which exhibit a partial defect in forming perinuclear PRG-1 foci [31]. (Figure 3.4A, right). We used *glh-1* single mutants to perform the HRDE-1 IP experiment due to technical limitations, as it is difficult to accumulate enough material for immunoprecipitation using the nearly sterile *glh-1 glh-4* double mutants. The *glh-1* single mutant has a milder defect in WAGO 22G-RNA accumulation than the *glh-1 glh-4* mutant: 33% of WAGO targets exhibit at least a 2-fold decrease in total 22G-RNAs in the *glh-1* mutant compared to 41% in the *glh-1 glh-4* (Supplementary Figure 3.10D and Figure 3.4A). Our HRDE-1 IP experiment revealed that 48% of WAGO targets exhibit a 2-fold decrease in HRDE-1 22G-RNAs in the *glh-1* mutant. To obtain a more conservative list of affected WAGO genes, we also applied a statistical threshold and observed that 16% of WAGO genes were significantly reduced in HRDE-1 22G-RNA accumulation in *glh-1* mutants, compared to only 5% of WAGO genes significantly increased. Using mRNA sequencing, we found the majority of genes exhibiting significantly reduced HRDE-1 22G-RNA levels in *glh-1* mutants had increased mRNA levels in *glh-1 glh-4* double mutants and in the *glh-1 DQAD* mutants (Figure 3.4B and Supplementary Figure 3.10E). Several transposons were among those shared activated genes, including TC2 and PAL8C_1 (Supplementary Figure 3.10F). These observations indicate that GLH/VASA mutants exhibit defects in silencing WAGO targets. As the levels of 22G RNAs from WAGO targets were reported to be reduced in *meg-3 meg-4* double mutants [112] which exhibit defects only in cytoplasmic but not in perinuclear PRG-1 granules, we wondered whether distinct or common WAGO targets exhibit 22G-RNA defects in mutants defective in cytoplasmic versus perinuclear P granules. We noticed that those WAGO genes with significantly reduced HRDE-1 22G-RNAs in *glh-1* mutants exhibit a greater reduction in 22G-RNAs compared to other WAGO target genes in the *glh-1 glh-4* mutants and in *glh-1*

DQAD mutants. A similar trend of 22G-RNA changes is also found in the *deps-1* mutant and in the *mip-1; mip-2* double mutant, which has also recently been shown to play an important role in P granule assembly[32]. In *meg-3 meg-4* mutants, which exhibit defects in the formation of embryonic cytoplasmic P granules, we see much less difference in 22G-RNA between these WAGO genes, suggesting WAGO 22G-RNA biogenesis is globally compromised as previously reported (Figure 3.4C) [112]. These results suggest that hundreds WAGO target genes are preferentially activated in GLH/VASA mutants and other mutants exhibiting defects in perinuclear P granules.

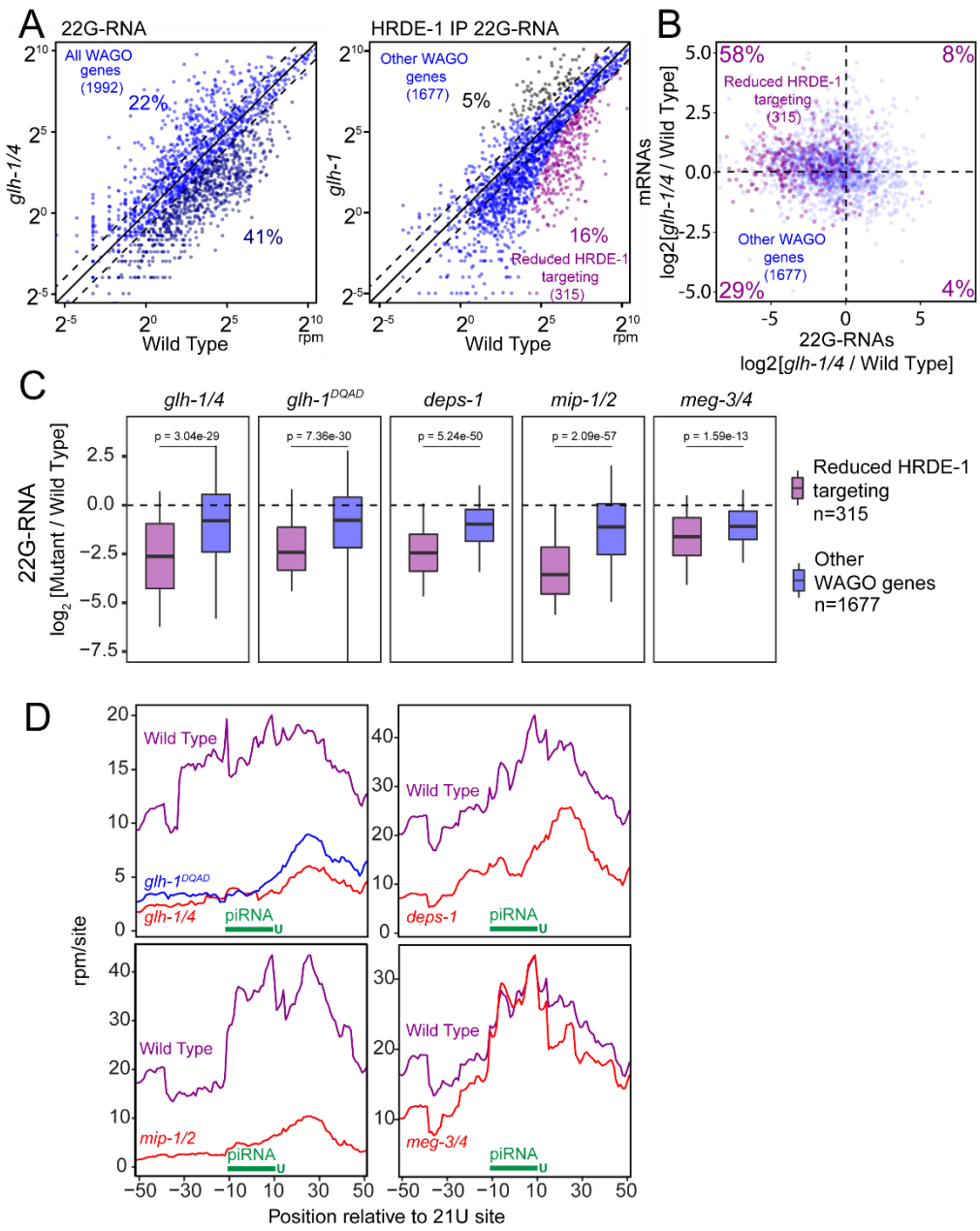


Figure 3.4: GLH/VASA mutants exhibit defects in producing secondary WAGO-22G-RNAs. (A) A scatter plot showing the abundance of all 22G-RNAs (left) or HRDE-1 bound 22G-RNAs (right) mapped to each WAGO target in wild type worms compared to indicated strains. (Left) The percentage of WAGO targets with 2-fold increased or decreased 22G-RNAs in mutants are shown. (Right) The percentage of significantly changed (adjusted $P < 0.05$ [see Methods for details] and 2-fold) WAGO targets are shown. The three diagonal lines indicate a two-fold increase (top), no change (middle), or a two-fold depletion (bottom) in the indicated mutant strains. (B) A scatter plot showing the mRNA vs 22G-RNA $\log_2$ expression changes for WAGO targets in *glh-1 glh-4* mutants vs wild type worms. The upper left quadrant corresponds to WAGO targets that have become activated in the mutant (increased mRNAs and decreased 22G-RNAs). Percentages are shown in each quadrant to indicate the proportion of WAGO targets with reduced HRDE-1-associated 22G-RNAs in the *glh-1* mutant that fall in that quadrant.

Figure 3.4 (continued): (C) 22G-RNA changes for WAGO targets with reduced HRDE-1-associated 22G-RNAs in the *glh-1* mutant versus all other WAGO targets in the indicated strains versus wild type worms. Statistical analysis was performed using a two-tailed Mann-Whitney Wilcoxon test. For all boxplots, lines display median values, boxes display first and third quartiles, and whiskers display 5th and 95th percentiles. (D) Density of 22G-RNAs within a 100 nt window around predicted piRNA target sites in the indicated strains. Computed by summing 22G-RNA density per piRNA targeting site in WAGO targets with reduced HRDE-1 associated 22G-RNAs in the *glh-1* mutant. The plots are centered on the 10th nucleotide of piRNAs.

Since our piRNA reporter analyses suggest that GLH/VASA is critical for the production of downstream 22G-RNAs around piRNA targeting sites, we wondered whether the production of piRNA-dependent 22G-RNAs are preferentially affected in GLH/VASA mutants and other mutants exhibiting defects in forming perinuclear P granules [87]. Indeed, in *glh-1 glh-4, glh-1* DQAD and *deps-1* mutants, those WAGO target genes in which the production of 22G-RNAs most depends on PRG-1 exhibit greater reductions in 22G-RNAs than other WAGO targets (Supplementary Figure 3.10G). A still significant but lesser difference in 22G-RNA production was found for *meg-3 meg-4* mutants (Supplementary Figure 3.10G). To specifically examine the production of 22G-RNA production at piRNA targeting sites, we compared the local production of 22G-RNAs at predicted piRNA target sites [177,185]. For those WAGO targets with significantly reduced HRDE-1 22G-RNAs in *glh-1* mutants, 22G-RNAs are enriched around predicted piRNA target sites in wild type but much less so in *glh-1 glh-4*, *glh-1 DQAD*, or *deps-1* mutants (Figure 3.4D). In contrast, while 22G-RNAs are overall slightly reduced in *meg-3 meg-4* mutants, 22G-RNAs remain enriched around these predicted piRNA target sites (Figure 3.4E). Similar results were observed when we looked at predicted piRNA target sites in all WAGO genes (Supplementary Figure 3.10H). These observations suggest that the production of piRNA-dependent 22G-RNAs is preferentially compromised in the *glh* and *deps-1* mutants. Together, these observations suggest that piRNA-dependent WAGO target genes are preferentially activated in VASA mutants and other mutants exhibiting defects in forming perinuclear P granules.

3.3.6 Perinuclear P granules prevent silencing of self genes

Essentially all germline transcripts are targeted either by 22G-RNAs associated with WAGO or CSR-1 Argonautes, which can silence or license the expression of their targets, respectively [61,133,170]. Since CSR-1 is present in GLH-1 complexes and the formation of CSR-1 perinuclear condensates is promoted by GLHs (Figure 3.1A & B), we wondered whether GLHs may regulate the biogenesis and/or function of CSR-1-associated 22G-RNAs. It has recently been shown that simultaneous RNAi against four P granule factors (*glh-1*, *glh-4*, *pgl-1*, and *pgl-3*) does not lead to global changes to 22G-RNAs antisense to CSR-1 targeted genes [144]. Consistent with this finding, we found that 22G-RNAs antisense to CSR-1 target genes remain mostly unchanged in *glh* mutants compared to wild type (Figure 3.5A, left and Supplementary Figure 3.11A). Intriguingly, in *glh-1 glh-4* double mutants, we noticed that 717 CSR-1 genes (15%) exhibit at least a two-fold increase in 22G-RNAs that mapped to them. Similar to the analyses of WAGO targets, we performed HRDE-1 IP to ask whether these changes were due to increased production of WAGO-associated 22G-RNAs. Again, we used the more fertile but less severely affected *glh-1* single mutants to perform the IP, which from total 22G-RNA sequencing show 330 CSR-1 targets (7%) with two-fold increased 22G-RNA accumulation compared to the 717 CSR-1 targets (15%) upregulated in double mutants (Supplementary Figure 3.11B and Figure 3.5A).

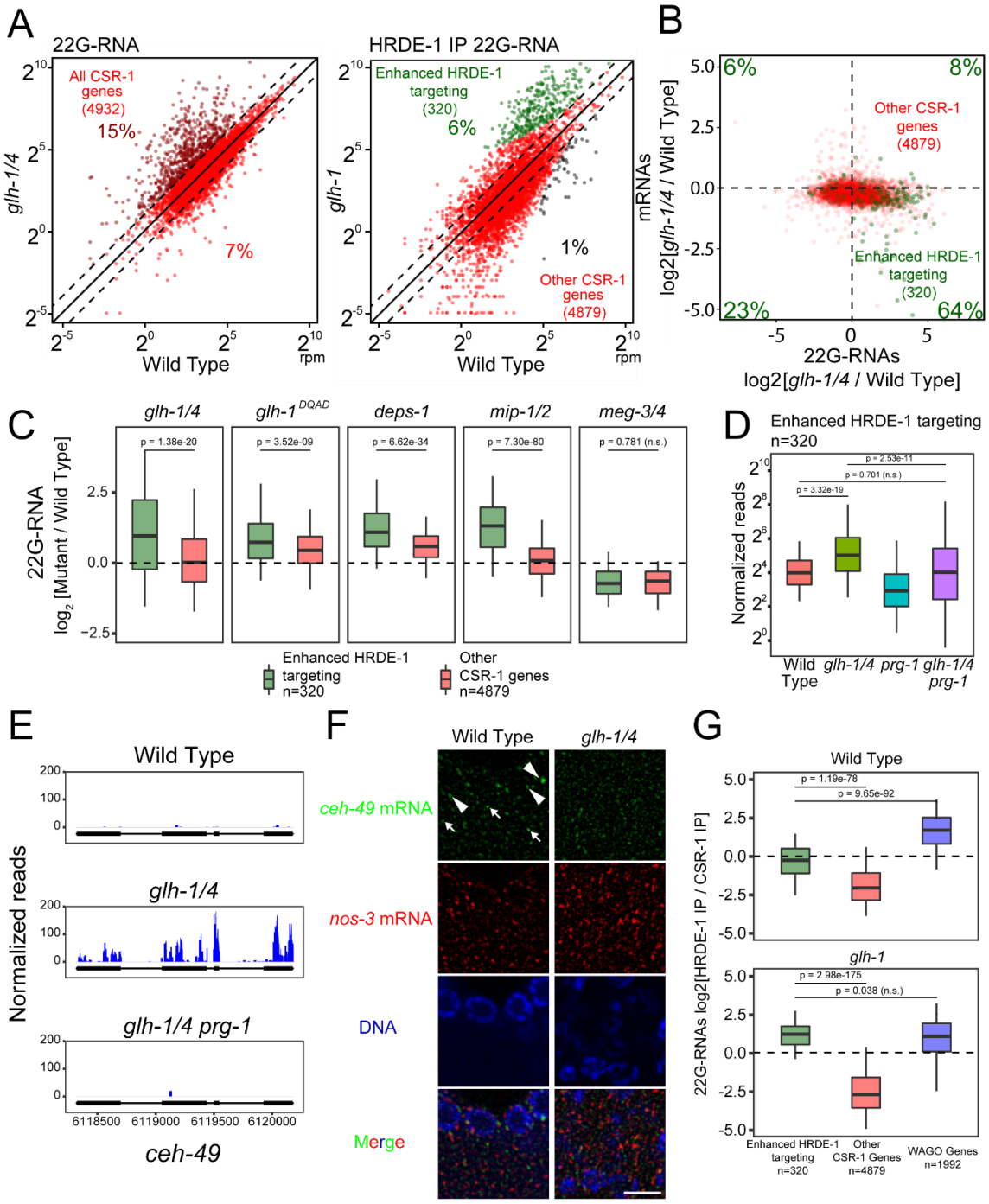


Figure 3.5: Many functional germline genes are silenced in mutants with defects in forming perinuclear PRG-1 granules.

(A) A scatter plot showing the abundance of all 22G-RNAs (left) or HRDE-1 bound 22G-RNAs (right) mapped to each CSR-1 target in wild type worms compared to indicated strains. (Left) Percentages of CSR-1 targets with 2-fold increased or decreased 22G-RNAs in mutants are shown. (Right) The percentage of significantly changed (adjusted $P < 0.05$ [see Methods for details] and 2-fold) CSR-1 targets are shown. Diagonal lines indicate a two-fold increase (top), no change (middle), or a two-fold depletion (bottom) in the indicated mutant strains.

Figure 3.5 (continued): (B) A scatter plot showing the mRNA vs 22G-RNA expression changes for CSR-1 targets in the indicated strain. Percentages in each quadrant indicate the proportion of CSR-1 targets with enhanced HRDE-1 targeting that fall in that quadrant. (C) 22G-RNA fold changes for CSR-1 targets with enhanced HRDE-1 targeting versus all other CSR-1 targets in the indicated strains versus wild type worms. Statistical analysis was performed using a two-tailed Mann-Whitney Wilcoxon test. For all boxplots, lines display median values, boxes display first and third quartiles, and whiskers display 5^{th} and 95^{th} percentiles. (D) 22G-RNA accumulation for CSR-1 targets with enhanced HRDE-1 targeting compared between the indicated strains. Statistical analysis was performed using a two-tailed Mann-Whitney Wilcoxon test. For all boxplots, lines display median values, boxes display first and third quartiles, and whiskers display 5^{th} and 95^{th} percentiles. (E) 22G-RNAs distribution in the enhanced HRDE-1 associated 22G-RNA target ceh-49 in the indicated strains. (F) Photomicrographs of pachytene nuclei in adult gonads hybridized with specified single molecule fluorescent (smFISH) probes in the indicated strains. Nuclear DNA was stained with DAPI (Blue). Arrowheads indicate perinuclear mRNA foci. Arrows indicate cytoplasmic mRNA foci. Bar, 5 micrometers. (G) 22G-RNA enrichment in HRDE-1 versus CSR-1 IP experiments in the indicated strains for CSR-1 targets and WAGO targets. Dotted line indicates no enrichment for either Argonaute. Statistical analysis was performed using a two-tailed Mann-Whitney Wilcoxon test. For all boxplots, lines display median values, boxes display first and third quartiles, and whiskers display 5th and 95th percentiles. Distributions represent data collected across 2 biological replicates for CSR-1 IP and a single experiment for HRDE-1 IP.

We again applied a statistical threshold to our IP data to obtain a more conservative list of affected CSR-1 genes. Strikingly, HRDE-1 (WAGO-9) immunoprecipitation analyses of the glh-1 mutant showed that 320 genes were significantly enhanced in HRDE-1 22G-RNA accumulation in glh-1 mutants (Figure 3.5A). mRNA sequencing analyses confirmed an overall reduction in mRNA levels for these enhanced HRDE-1 targeted genes in glh-1 glh-4 double and in glh-1 DQAD mutants (Figure 3.5B and Supplementary Figure 3.11A). The aberrantly silenced CSR-1 targets include several functional genes involved in germline development, including nos-2, puf-6 and puf-7.

We wondered whether the aberrant silencing of CSR-1 genes occurs in other mutants defective in cytoplasmic and/or perinuclear P granule formation. In *glh-1 DQAD* mutants, a significant increase in 22G-RNAs and decrease in mRNAs are found for those CSR-1 genes with enhanced HRDE-1 22G-RNA levels in *glh-1* mutants. Similar changes of 22G-RNA or mRNA levels are found in other mutants exhibiting defects in the formation of perinuclear P granules,

including *deps-1* and *mip-1/2* mutants. In contrast, in the *meg-3 meg-4* mutant that only disrupts the formation of embryonic cytoplasmic P granules in embryos, we did not observe an increase in 22G-RNAs nor a decrease in mRNAs for these CSR-1 genes with enhanced HRDE-1 22G-RNA levels in *glh-1* mutants (Figure 3.5C). These observations suggest that in mutants with abnormal perinuclear P granules, a group of CSR-1 target genes are aberrantly silenced by WAGO 22G-RNAs.

3.3.7 PRG-1 is required for the aberrant self silencing

Because perinuclear P granule loss affects the localization but not abundance of PIWI PRG-1 [31], we wondered whether PRG-1 may be aberrantly recognizing these CSR-1 genes when perinuclear P granule formation is compromised. Indeed, we found that the aberrant production of 22G-RNAs against CSR-1 genes with enhanced HRDE-1 22G-RNA levels in *glh-1* mutants are significantly reduced in *glh-1 glh-4 prg-1* triple mutants compared to *glh-1 glh-4* double mutants (Figure 3.5D). The mutation of *prg-1* suppresses the aberrant accumulation of CSR-1 22G-RNAs in *glh-1 glh-4* double mutants by 2-fold or more for nearly two-thirds of affected genes (Supplementary Figure 3.11C). These data suggest that most of the aberrant mRNA silencing of CSR-1 genes found in GLH/VASA mutants is triggered by piRNAs. Together, these data suggest VASA not only promotes silencing of non-self (WAGO targets), but also promotes licensing of self (CSR-1 targets) to avoid silencing by piRNAs. Previous studies have shown that improper silencing of CSR-1 targets can occur in *prg-1* mutants[4,106]. In contrast, our observations demonstrate that aberrant silencing of CSR-1 targets can also occur in the presence of PRG-1. In fact, here we showed that in VASA mutants, PRG-1 is responsible for triggering the aberrant silencing of many CSR-1 targets.

If self transcript licensing depends on P granules, then these transcripts should be present in perinuclear condensates. Using RNA smFISH analyses, we examined the localization of *ceh-49*, an aberrantly silenced mRNA normally protected from silencing by CSR-1 (Figure 3.5E), and noticed that *ceh-49* mRNAs were expressed in the pachytene region of the germline in wild type animals and both perinuclear and cytoplasmic signals are detected (Figure 3.5F). In *glh-1 glh-4* mutants, both perinuclear and cytoplasmic *ceh-49* signals were decreased. We performed colocalization analysis with PRG-1 to measure the extent to which *ceh-49* accumulates in P granules compared to an mRNA that does not become aberrantly silenced by piRNAs in VASA mutants. We observed greater colocalization of *ceh-49* mRNA with PRG-1 compared to the colocalization of *nos-3* mRNA with PRG-1 (Supplementary Figure 3.11D). This effect was dependent on *glh-1 glh-4*. These results are consistent with the model that the P granule localization of some mRNA transcripts is critical for their protection from piRNA silencing [112].

We wondered why some CSR-1 targeted genes gained aberrant silencing in GLH/VASA mutants. We hypothesized that these genes might already be prone to silencing in wild type animals. To test this hypothesis, we compared the amounts of HRDE-1 associated 22G-RNAs to CSR-1 associated 22G-RNAs on these genes. Interestingly, in wild type animals, CSR-1 targets with enhanced HRDE-1 22G-RNA levels in *glh-1* mutants already exhibited a significantly higher ratio of HRDE-1 associated 22G-RNAs to CSR-1 associated 22G-RNAs compared to other CSR-1 targets (Figure 3.5G). This ratio for CSR-1 genes with enhanced HRDE-1 22G-RNA levels in *glh-1* mutants favors HRDE-1 even more in *glh-1* mutants (Figure 3.5G). The HRDE-1 to CSR-1 associated small RNA ratio for this group of genes was so far shifted in the *glh-1* mutant that they now more closely resemble WAGO genes (Figure 3.5G). We wondered whether CSR-1 targets which did not meet the stringent cutoff to be called enhanced HRDE-1 targeted genes but still

showed two-fold increased 22G-RNA targeting in *glh-1 glh-4* mutants also showed this trend. We found that the HRDE-1/CSR-1 IP ratio for these targets showed a similar trend of significant elevation compared to other CSR-1 genes, but their ratio was not on average as elevated as the ratio for the targets with elevated HRDE-1 targeting (Supplementary Figure 3.12A), suggesting that while the group of genes which show significant HRDE-1 enhanced targeting in the mutant are distinct, there are likely other genes which become mis-regulated but just to a lesser extent. These data indicate that these aberrantly silenced CSR-1 genes are already more targeted by silencing machinery than other CSR-1 genes in wild type animals. In addition, previous studies have shown CSR-1 22G-RNAs are enriched at their 3' end of transcripts [68,144]. However, CSR-1 22G-RNAs are more distributed across the whole gene body for these aberrantly-silenced CSR-1 targets, similar to the distribution of CSR-1 22G-RNAs observed for WAGO targets (Supplementary Figure 3.12B). To characterize the extent to which these features correlate with enhanced HRDE-1 targeting, we defined the overlap between CSR-1 targets which do not show 3' end targeting enrichment, CSR-1 targets which have a high HRDE-1/CSR-1 IP ratio, and CSR-1 targets which show enhanced HRDE-1 targeting in *glh-1* mutants. We saw that over two-thirds of CSR-1 genes that gain aberrant HRDE-1 22G-RNAs in *glh-1* mutants also share either of these other two features (Supplementary Figure 3.12C). We further characterized this effect by comparing the increase in HRDE-1 targeting following *glh-1* loss for groups of CSR-1 targets that have varying levels of 3' end targeting enrichment and HRDE-1/CSR-1 IP ratios. We saw that while groups of genes with no 3' end enrichment or with a high HRDE-1/CSR-1 IP ratio certainly tend to gain more HRDE-1 22G-RNAs in *glh-1* mutants, the magnitude of these effects are not distinct enough to fully predict which CSR-1 targets would gain HRDE-1 22G-RNAs in P granule mutants (Supplementary Figure 3.12D). Together, these observations suggest that perinuclear P

granules are critical for protecting hundreds of CSR-1 transcripts from aberrant piRNA silencing while at the same time promoting piRNA silencing on WAGO targets.

3.5 Discussion

Mutations in VASA helicase genes have been reported to lead to defects in the localization of PIWI in various animals [31,80,93,96,180]. Whether the localization of other small RNA factors in P granules and other perinuclear granules also relies on GLH/VASA helicases had only been explored sporadically. In *C. elegans*, piRNAs and other small RNA pathway factors are concentrated in distinct but partially overlapping granules, including P granules, Z granules and Mutator granules [118,169], suggesting these granules interact to facilitate gene regulation by small RNAs. Here we expand on these analyses to examine the roles GLH-1 and GLH-4 play in the localization of various small RNA factors, include P granule components - DEPS-1, EGO-1, CSR-1, Z granule components - WAGO-4 and ZNFX-1, and Mutator component MUT-16. We found that the formation of perinuclear condensation of these components are each compromised in *glh-1 glh-4* double mutants. There are some important caveats to our findings. First, not all the components examined are compromised to the same extend in *glh-1 glh-*4 mutants. Second, it is possible that while condensates were observably disrupted upon our genetic manipulation, there could be small RNA factors that remain at the nuclear periphery that, although undetectable by microscopy, still function sufficiently to contribute self versus non-self distinction. Nevertheless, our observations suggest that GLH/VASA plays a global role in promoting the formation of these various liquid condensates.

Phase-separated condensates are capable of concentrating various proteins and RNAs, but whether these condensates indeed play a biological role remains controversial. For example, a previous study has reported that P body formation is a consequence but not a cause of miRNA-

mediated gene silencing [47]. In this study, we found that the production of piRNAs and other 22G-RNAs are not grossly affected in GLH/VASA mutants and other mutants defective in forming perinuclear condensates. Instead, the small RNA-based distinction of self and non-self RNAs are compromised in these mutants. Specifically, piRNA-dependent silencing of non-self is reduced, leading to increased mRNA expression. Simultaneously, hundreds of self RNAs (CSR-1 targets) were aberrantly silenced by piRNAs, leading to reduced mRNA expression. Our results demonstrate that a significant portion of germline transcripts are mis-regulated by HRDE-1 22G RNAs in VASA mutants. As both WAGO and CSR-1 22G-RNAs can establish epigenetic memories[35,141], the silenced or expressed state of many germline transcripts may be preserved in the absence of P granules. In this model, those mRNAs which did not establish robust epigenetic memories would be those that exhibit more mis-regulation in P granule mutants. Indeed, we observed that PRG-1 dependent 22G-RNA targets, which depend on PRG-1/piRNAs at each generation to trigger gene silencing, are those which exhibit a greater reduction of 22G-RNAs on WAGO targets in *glh-1/4* mutants (Supplementary Figure 3.10G). In addition, a previous study from the Ketting lab has demonstrated that re-establishment of the 22G-RNA system in *prg-1* mutants leads to gene-mis-regulation in a stochastic manner that varies between worms[4]. As our measurements of 22G-RNAs or mRNAs are from hundreds of thousands of worms, stochastic activation or silencing that may exist in individual worms may not be detected. Examining whether P granule mutant worms exhibit aberrant activation or silencing of germline transcripts stochastically will be interesting in future studies. Notably, while a similar mis-regulation of CSR-1 and WAGO targets can be found in several mutants defective in P granule assembly, such as VASA, *deps-1*, *pgl-1* and *mip-1/2* mutants, such defects are distinct from those defects observed in mutants defective for Mutator assembly[118] or Z granule function[68,169,181]. Together,

these data argue that perinuclear P granules are critical for the fidelity of piRNA-based surveillance in *C. elegans* and provide an environment that allows distinct Argonaute proteins, such as PRG-1 and CSR-1, to distinguish self from non-self. Our observations support a model where P granules act as the checkpoint for piRNA-mediated gene silencing, where the P granule promotes piRNA target recognition for non-self genes while allowing CSR-1 to guard self genes (Figure 3.6). In GLH/VASA mutants, the P granule fails to form, leading to the dispersal of small RNA factors including PRG-1, WAGOs and CSR-1 into the cytoplasm. When denied the environment of the P granule, hundreds of typically silenced non-self transcripts fail to contact silencing machinery leading to their expression while hundreds of typically expressed self transcripts fail to contact CSR-1 leading to their repression by PRG-1. This model suggests P granules act as a specialized environment where transcripts are allowed residence time to properly contact either silencing or licensing machinery, as evidenced by our smFISH data that demonstrates the GLH/VASA-dependent perinuclear localization of both silenced and expressed mRNAs (Figure 3.2E and Figure 3.6F). Taken together, our study reveals the critical role of perinuclear P granules in promoting the fidelity of self and non-self nucleic acids distinction and has broad implications for the functions of other RNA-enriched liquid condensates.

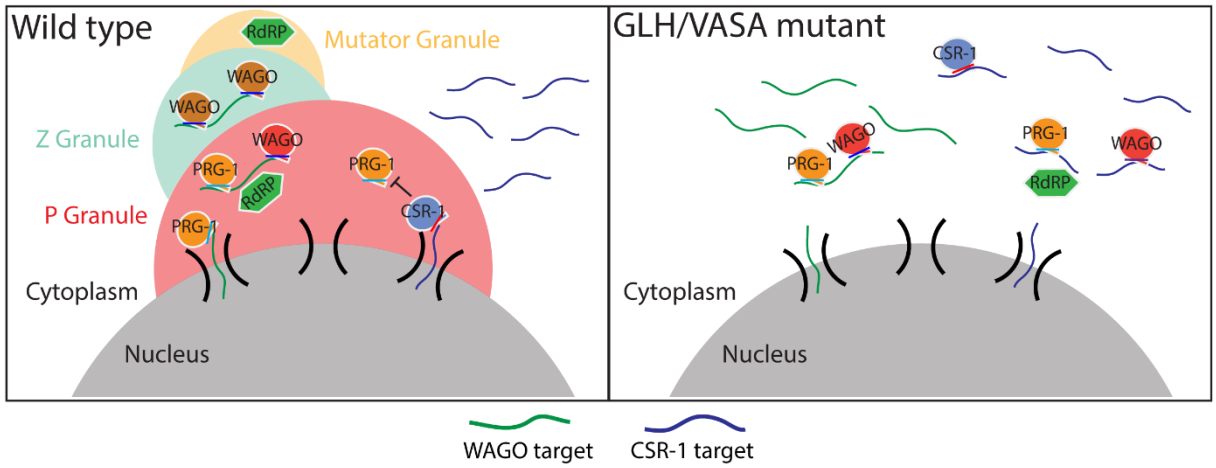


Figure 3.6: P granules promote piRNA targeting fidelity.
Model showing the P granule working in concert with Z and Mutator granules to maintain an environment that protects specific self genes from silencing and promotes the silencing of specific non-self genes (Left). In GLH/VASA mutants where granule integrity is lost, hundreds of typically expressed transcripts become targeted by PRG-1 and silencing machinery while hundreds of typically silenced transcripts become dispersed into the cytoplasm and expressed (Right).

3.6 Methods

3.6.1 Caenorhabditis elegans strains

Animals were grown on standard nematode growth media (NGM) plates seeded with the Escherichia coli OP50 strain at 20°C or temperatures where indicated. Some strains were obtained from the Caenorhabditis Genetics Center (CGC). The strains used in this study are *glh-1(uoc1) I*, *glh-1*^{DQAD}*(uoc3) I*, *glh-1*$^{FGG\Delta}$*(uoc4) I*, *glh-1(uoc5) glh-4(gk225) I/hT2 [bli-4(e937) let-?(q782) qIs48] (I;III)*, *deps-1(bn124) I*, *meg-3(ax3055) meg-4(ax3052) X.*, and *mip-1(uae1) III; mip-2(uae2) V.*

3.6.2 Fluorescence microscopy and image processing

GFP- and RFP-tagged fluorescent proteins were visualized in living nematodes or dissected embryos by mounting young adult animals on 2% agarose pads with M9 buffer (22 mM KH2PO4,

42 mM Na2HPO4, and 86 mM NaCl) with 10-50 mM levamisole, or mounting one-cell embryos on 2% agarose pads by dissecting gravid hermaphrodites into egg salt buffer (5 mM HEPES pH 7.4, 118 mM NaCl, 40 mM KCl, 3.4 mM MgCl2, and 3.4 mM CaCl2). Fluorescent images were captured using a Zeiss LSM800 confocal microscope with a Plan-Apochromat 63X/1.4 Oil DIC M27 objective.

Image processing and quantification of fluorescent puncta were performed using ImageJ. Single slice images of gonads and images from maximum intensity projections of z-series of one-cell embryos were used for quantification. Regions of interest (ROIs) were selected, and areas of ROIs were measured. Image thresholds were set manually, and fluorescent puncta were selected. Puncta density was calculated as the number of fluorescent puncta in the ROI divided by the area of ROI. Integrated intensities and sizes of fluorescent puncta in all ROIs were measured. Data were analyzed by student's t-test or one-way ANOVA followed by Tukey's correction for multiple comparisons.

3.6.3 piRNA initiation assay

Initiation of piRNA silencing by introduction of extrachromosomal arrays was performed as previously described [185]. Briefly, a plasmid carrying a synthetic piRNA was co-injected with pRF4 plasmid carrying the dominant injection marker *rol-6dm* into young adult gonads of indicated genotypes. Progeny were screened for the presence of the injection marker, and progeny of these F1s (F2 progeny of injected worms) were screened for their ability to silence the *gfp::cdk-1* transgene if they were able to inherit the injection marker.

3.6.4 RNA isolation and quantitative real-time PCR

Total RNA was extracted using the standard method with TRIzol reagent (Invitrogen) from whole animals of ~100,000 synchronized young adults. Stem-loop real-time PCR was performed to

measure piRNA levels. 1 μg of total RNA was reverse transcribed with SuperScript IV Reverse Transcriptase (Invitrogen) in 1x reaction buffer, 5U SUPERase-In RNase Inhibitor (Invitrogen), 1 mM dNTPs, and 50 pM stem-loop reverse primer 5'-CTCAACTGGTGTCGTGGAGTCGGCAATTCAGTTGAG-n8-3' (n8=reverse complement sequences of last 8 nucleotide acids in piRNAs). Each real-time PCR reaction consisted 4 μL of cDNA, 1 μM forward piRNA primer 5'-ACACTCCAGCTGGG-n16-3' (n16= first 16 nucleotide acids in piRNAs), and 1 μM universal reverse primer 5'-CTCAAGTGTCGTGGAGTCGGCAA-3'. The amplification was performed using Power SYBR Green (Applied Biosystems) on the Bio-Rad CFX96 Touch Real-Time PCR Detection System. The experiments were repeated for a total of three biological replicates.

3.6.5 Small RNA sequencing

Total RNA was extracted from whole animals of ~100,000 synchronized young adults as described above. Small (<200nt) RNAs were enriched with mirVana miRNA Isolation Kit (Ambion). In brief, 80 μL (200-300 μg) of total RNA, 400 μl of mirVana lysis/binding buffer and 48 μL of mirVana homogenate buffer were mixed well and incubated at room temperature for 5 minutes. Then 176 μL of 100% ethanol was added and samples spun at 2500 x g for 4 minutes at room temperature to pellet large (>200nt) RNAs. The supernatant was transferred to a new tube and small (<200nt) RNAs were precipitated with pre-cooled isopropanol at -70°C. Small RNAs were pelleted at 20,000 x g at 4°C for 30 minutes, washed once with 70% pre-cooled ethanol, and dissolved with nuclease-free water. 10 μg of small RNA was fractionated on a 15% PAGE/7M urea gel, and RNA from 17 nt to 26 nt was excised from the gel. RNA was extracted by soaking the gel in 2 gel volumes of NaCl TE buffer (0.3 M NaCl, 10 mM Tris-HCl, 1 mM EDTA pH 7.5) overnight. The supernatant was collected through a gel filtration column. RNA was precipitated

with isopropanol, washed once with 70% ethanol, and resuspended with 15 µL nuclease-free water. RNA samples were treated with RppH to convert 22G-RNA 5' triphosphate to monophosphate in 1x reaction buffer, 10U RppH (New England Biolabs), and 20U SUPERase-In RNase Inhibitor (Invitrogen) for 3 hours at 37°C, followed by 5 minutes at 65°C to inactivate RppH. RNA was then concentrated with the RNA Clean and Concentrator-5 Kit (Zymo Research). Small RNA libraries were prepared according to the manufacturer's protocol of the NEBNext Multiplex Small RNA Sample Prep Set for Illumina-Library Preparation (New England Biolabs). NEBNext Multiplex Oligos for Illumina Index Primers were used for library preparation (New England Biolabs). Libraries were sequenced using an Illumina HiSeq4000 to obtain single-end 36 nt sequences at the University of Chicago Genomic Facility.

3.6.6 RNA immunoprecipitation sequencing (RIP-seq)

A total of ~100,000 synchronized young adult animals were used for RIP-seq. Worm pellets were resuspended in equal volumes of immunoprecipitation buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40, 80 U/mL RNase Inhibitor (Thermo Fisher Scientific), 1 mM dithiothreitol, and protease inhibitor cocktail without EDTA (Promega)), and grinded in a glass grinder for 8-10 times. Lysates were clarified by spinning down at 15,000 rpm, 4°C, for 15 minutes. Supernatants were incubated with the GFP-Trap magnetic agarose beads (ChromoTek) at 4°C for 1 hour. Beads were washed with wash buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40, and 1 mM dithiothreitol) six times, and then resuspended in TBS buffer for RNA extraction. Total RNA was extracted using the standard method with TRIzol reagent (Invitrogen). Small RNA libraries for RNA-seq were prepared as described above. Libraries were sequenced using an Illumina HiSeq4000 to obtain single-end 36 nt sequences at the University of Chicago Genomic Facility.

3.6.7 Chemical cross-linking and co-immunoprecipitation of GLH-1

Chemical cross-linking of proteins was performed with Dithio-bismaleimidoethane (DTME). ~100,000 synchronized flag::mCherry::GLH-1 and wild type (N2) young adults were collected and washed three times with M9. Two biological replicates for each genotype were used. M9 was discarded to the same amount of the worm volume, then DTME dissolved in Dimethyl Sulfoxide (DMSO) was added to a final concentration of 2 mM. Samples were incubated for 30 minutes at room temperature with occasional shaking before washed three times with M9 to remove the cross-linker. Worm pellets were resuspended in equal volume of immunoprecipitation buffer (20 mM Tri-HCl pH 7.5, 150 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40, 80 U/mL RNase Inhibitor (Thermo Fisher Scientific), 1 mM dithiothreitol, and protease inhibitor cocktail without EDTA (Promega)). Worm pellets were homogenized using glass homogenizer for 15-20 strokes on ice. Lysates were centrifuged at 14000 x g (Eppendorf Centrifuge 5424 R) for 10 minutes to remove insoluble material. Supernatants were incubated with 25 µL of Anti-Flag Magnetic Beads (bioLinkedin) for 2 hours at 4°C on an end-to-end rotator. Supernatant was removed and beads were washed with 1 mL of wash buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40) six times for 10 minutes each time, and with the final wash of 0.05% NP-40. Beads were incubated at 37°C for 30 minutes in 50 µL de-crosslinking buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM MgCl2, 0.2% Tween 20, 10 mM dithiothreitol). The final samples were boiled in 2 x SDS loading buffer at 100°C for 5 minutes before mass-spectrometry analysis.

3.6.8 Mass-spectrometry analysis

Mass spectrometry. The samples were subjected to SDS PAGE gel electrophoresis experiments (120V, 1hr) and visualized by Coomassie staining to obtain gel strip samples. Whole lanes were collected as single samples. 800 µL of 0.1M NH4HCO3/30% acetonitrile (ACN) was added to

samples, decolorized and washed until the protein blue disappeared, and supernatant was removed. 800 μL of H2O was added to samples immediately to terminate reactions for 10 minutes, and the supernatant was removed. 800μL 100mM NH4HCO3 was added and samples were allowed to stand for 20 minutes, and the supernatant was removed. 40μL 100mM DTT and 360μL 100mM NH4HCO3 were added to samples and samples were incubated at 56°C for 30 minutes to reduce the protein; the supernatant was removed, 100μL 100% ACN was added and incubated for 5 min, then aspirated. 280μL 100mM NH4HCO3 and 120μL 200mM 3-indole acetic acid (IAA, freshly prepared, stored in the dark) was added at room temperature and incubated for 20 minutes in the dark; the supernatant was removed,100 μL 100 mM NH4HCO3 was added, and left to stand at room temperature for 15 minutes to remove the supernatant; 100 μL 100% CAN was added and incubated for 5 minutes, aspirated and freeze-dried. 600μL of 20ng/μL Trypsin (Promega, V5113) was added to samples and samples were placed in a refrigerator at 4°C for about 30 minutes to inflate the gel block; then 100μL of 50mM NH4HCO3 buffer was added and digested overnight at 37°C. The enzymatic hydrolysate was aspirated and transferred to a new centrifuge tube, 100μL of 60% ACN/0.1%TFA (trifluoroacetic acid) was added to the gel block samples and sonicated for 15 min, solution was aspirated and added to the previous solution. This extraction was repeated 3 times, combined and lyophilized. After enzymolysis, the peptides were desalted using a C18 StageTip column, concentrated and dried. Then the peptides were reconstituted with 0.1% formic acid aqueous solution for subsequent LC MS/MS analysis

LC-MS/MS analysis. An appropriate amount of peptides from the sample was taken and used on the nanoliter flow rate Easy-nLC 1200 chromatography system (Thermo Scientific) for chromatographic separation. Buffer A was a 0.1% formic acid aqueous solution, and buffer B was a mixed solution of 0.1% formic acid, acetonitrile and water (where acetonitrile was 80%). The

column was equilibrated with 100% buffer A. Samples were injected into the Trap Column (100μm x 20mm, 5μm, C18, Dr. Maisch GmbH) and then passed through the chromatographic analysis column (75 μm × 150 mm, 3 μm, C18, Dr. Maisch GmbH) for gradient separation at a flow rate of 300nL/min. The related liquid phase gradient was as follows: 0-3min, linear gradient of liquid B from 2% to 8%; 3-43 min, linear gradient of liquid B from 8% to 28%; 43-51 min, linear gradient of liquid B from 28% to 40%; 51-52 min, linear gradient of liquid B from 40% to 100%; 52-60 min, linear gradient of liquid B maintained at 100%. After separation of the peptides, a Q Exactive HF-X mass spectrometer (Thermo Scientific) was used for DDA (Data Dependent Acquisition) mass spectrometry analysis. The analysis time is 60 min, the detection mode: positive ion, the precursor ion scan range: 350-1800 m/z, the resolution of the primary mass spectrometer is 60000 @m/z 200, AGC target 3e6, and the primary maximum IT: 50ms. Peptide secondary mass spectrometry analysis is collected according to the following method: after each full scan (full scan), it is triggered to collect the secondary mass spectrum (MS2 scan) of the 20 highest intensity precursor ions, and the resolution of the secondary mass spectrum is 15000@m/z 200, AGC target: 1e5, secondary Maximum IT 25ms MS2 Activation Type: HCD, Isolation window: 1.6m/z, Normalized collision energy: 28. The mass spectrometry database search software MaxQuant 1.6.1.0 was used with the following protein database from Uniprot Protein Database: species *Caenorhabditis elegans* uniprot-*Caenorhabditis elegans* [6239]-27419-20210222.fasta. A PSM FDR < 0.01 and a protein FDR < 0.01 were used to assign positive protein identifications.

3.6.9 RNA smFISH and PRG-1 immunohistochemistry

smFISH of nos-3, gfp, *and* ceh-49 *mRNA.* RNA smFISH was performed on dissected adult *C. elegans* gonads. For particular genotypes and conditions, experiments were performed with at least two technical replicates. Staged young adult worms were washed with M9 and resuspended in

Dissection Buffer (PBS, 1mM EDTA) then deposited onto 18mm circular coverslips that were coated with 0.1% Poly-L-lysine solution. Gonads were dissected from whole worms using 25G x 5/8 hypodermic needles directly onto coverslips in a 12 well tissue culture plate. All subsequent steps were performed directly in the 12 well plate. Gonads were fixed in 4% formaldehyde for 30 minutes at room temperature. Gonads were then washed with PBS and dehydrated with 70% ethanol at 4C for one or two overnights. Gonads were re-hydrated with PBS and washed once with FISH wash buffer (2xSSC, 10% formamide) at 37C. FISH probes were suspended in Hybridization Buffer (10% formamide, 2mM vanadyl ribonucleoside complex, 20mg/mL BSA, 10mg/mL dextran sulfate, 2mg/mL *E.coli* tRNA) 1:50. smFISH probes were created by Biosearch Technologies to target *gfp* mRNA, *nos-3* mRNA, or *ceh-49* mRNA. Oligo sequences can be found in Supplementary Data 3.2. *gfp* mRNA probes were conjugated to Quasar670 (Custom probe, Biosearch Technologies Cat. No. SMF-1065-5), *nos-3* to CalFluor Red 610 (Custom probe, Biosearch Technologies Cat. No. SMF-1082-5), and *ceh-49* to Quasar670 (Custom probe, Biosearch Technologies Cat. No. SMF-1065-5). Gonads were hybridized overnight at 37C in 12 well tissue plate wrapped tightly in parafilm. Gonads were washed with FISH wash buffer, stained with DAPI, and mounted onto 25mm x 75mm x 1mm plain glass slides with ProLong Diamond Antifade Mountant. Samples were imaged on a Zeiss LSM800 Confocal Microscope at 40X magnification using a Plan-Apochromat 40X/1.4 Oil Objective.

smFISH of nos-3, gfp, *and* ceh-49 *mRNA and antibody staining of PRG-1.* RNA smFISH and antibody staining was performed on dissected adult *C. elegans* gonads. For particular genotypes and conditions, experiments were performed with at least two technical replicates. Staged young adult worms were washed with M9 and resuspended in Dissection Buffer (PBS, 1mM EDTA) then deposited into a watch glass. Adults were dissected using 25G x 5/8 hypodermic

needles and then transferred to 1.5mL microcentrifuge tubes. All subsequent steps were performed in individual microcentrifuge tubes. Gonads were fixed in ice cold methanol for five minutes, freeze cracked in liquid nitrogen for one minute, and submerged in ice cold acetone for five minutes. Gonads were washed once with Antibody Wash Buffer (PBS, 1mM EDTA, 0,1% Tween-20) and 50uL of anti-PRG-1[15] suspended 1:200 in Antibody Suspension Buffer (PBS, 1mM EDTA, 0.1% Tween-20, 20mg/mL, 2mM vanadyl ribonucleoside complex) was added. Tubes were shaken at 850rpm overnight at 4C. Gonads were washed twice with Antibody Wash Buffer and 50uL of anti-Rabbit Alexa488 (Jackson Labs Cat. No. 711-547-003) suspended 1:400 in Antibody Suspension Buffer was added. Tubes were shaken at 850rpm for two hours at room temperature. Gonads were wash once with Antibody Wash Buffer and once with 2xSSC. Samples were suspended in FISH Wash Buffer for 10 minutes at 37C. smFISH protocol, mounting, and imaging was then performed as above.

Image processing and analysis. Image processing and quantification of fluorescent puncta were performed using ImageJ. Single slice images of gonads were used for quantification. Regions of interest (ROIs) were selected, and areas of ROIs were measured. For smFISH and PRG-1 channels (Alexa488, CalFluor Red 610, and Quasar670), a Difference of Gaussian filter was applied to full image stacks with $\sigma_1 = 4$ and $\sigma_2 = 1$. Image thresholds were set manually and uniformly across all samples. Colocalization analysis was performed by measuring the total number of pixels in a stack which showed colocalization as a fraction of the total number of pixels in a stack pertaining to the molecule being measured. For example, colocalization of *gfp* mRNA with PRG-1 protein was measured by dividing the number of thresholded pixels in the *gfp* channel which overlap with thresholded pixels in the PRG-1 channel by the total number of thresholded

pixels in the *gfp* channel. Images from 7-30 gonads were collected and quantified. Data were compared by student's t-test.

3.6.10 Sequencing data analysis

RNA-seq. Fastq reads were trimmed of adaptors using cutadapt [97]. Trimmed reads were aligned to the *C. elegans* genome build WS230 using bowtie2 ver 2.3.0 [83]. After alignment, reads were overlapped with genomic features (protein-coding genes, pseudogenes, transposons) using bedtools intersect [123]. Reads per kilobase million (RPKM) values were then calculated for each individual feature by summing the total reads mapping to that feature, multiplied by 1e6 and divided by the product of the kilobase length of the feature and the total number of reads mapping to protein-coding genes. Protein-coding genes were used to normalize by sequencing depth because mRNA libraries were prepared by polyA tail selection, so reads mapping to features devoid of polyA tails are likely contaminants. RPKM values were then used in all downstream analyses using custom R scripts, which rely on packages ggplot2[174], reshape2 [173], ggpubr [71], dplyr [175].

sRNA-seq. Fastq reads were trimmed using custom perl scripts. Trimmed reads were aligned to the *C. elegans* genome build WS230 using bowtie ver 1.2.1.1 [84] with options -v 0 --best --strata. After alignment, reads that were between 17-40 nucleotides in length were overlapped with genomic features (rRNAs, tRNAs, snoRNAs, miRNAs, piRNAs, protein-coding genes, pseudogenes, transposons) using bedtools intersect [123]. Sense and antisense reads mapping to individual miRNAs, piRNAs, protein-coding genes, pseudogenes, RNA/DNA transposons, simple repeats, and satellites were totaled and normalized to reads per million (RPM) by multiplying be 1e6 and dividing read counts by total mapped reads, minus reads mapping to structural RNAs (rRNAs, tRNAs, snoRNAs) because these sense reads likely represent degraded products. Reads

mapping to multiple loci were penalized by dividing the read count by the number of loci they perfectly aligned to. Reads mapping to miRNAs and piRNAs were only considered if they matched to the sense annotation without any overlap. In other words, piRNA and miRNA reads that contained overhangs were not considered as mature piRNAs or miRNAs respectively. 22G-RNAs were defined as 21 to 23 nucleotide long reads with a 5'G that aligned to protein-coding genes, pseudogenes, or transposons. RPM values were then used in all downstream analyses using custom R scripts using R version 4.0.0 [124], which rely on packages ggplot2 [174], reshape2 [173], ggpubr [71], dplyr [175]. To determine GLH-dependent 22G-RNA lists and PRG-1 dependent 22G-RNA lists, a Bayesian approach as described previously [95] was used: two models were compared to determine p-values. The first model states that a given gene has the same probability of accumulating positive reads between two samples. The second model states that a given gene's probability of accumulating positive reads is unique between two samples. The probability of each model for each gene was calculated and compared to yield the final probability (p-value) for each gene's read accumulation being the same between two samples. Significance was determined with Bonferroni correction $P < 0.05$.

Metagene analysis. Metagene profiles across gene lengths were calculated by computing the depth at each genomic position using 21 to 23 nucleotide long small RNA reads with a 5'G using bedtools genomecov [123]. A custom R script was then used to divide genes into 100 bins and sum the normalized depth within each bin. Groups of genes were then plotted using the sum of the normalize depth at each bin. To determine transcripts with particular CSR-1 IP enrichment at their 3' ends, the proportion of 21 to 23 nucleotide long small RNA reads with a 5'G in averaged wild type CSR-1 IP libraries mapping to the last 15% of each transcript's length was compared to the proportion of reads mapping to the remaining 85% of each transcript's length. Non-enriched

transcripts were accordingly defined as those with less than 15% of the reads mapping to the full-length transcript falling into that last 15% bin. Medium enrichment were those transcripts with 16% to 30% of their reads mapping to the last 15% of their mRNA length, and high enrichment were those transcripts with greater than 30% of their reads mapping to the last 15% of their mRNA length. Metagene profiles relative to piRNA targeting sites were calculated as the mean normalized reads per million at each nucleotide position using the indicated 22G-RNA reads. piRNA targeting positions were determined using the stringent piRNA targeting rules for the indicated group of transcripts according to the previously published piRNA targeting rules[185].

HRDE-1 / CSR-1 ratio. Normalized reads (RPM) were calculated as described above. RPM values were compared between HRDE-1 IP and CSR-1 IP libraries and ratio levels were assigned as follows: highly CSR-1 favored transcripts were those CSR-1 targeted transcripts with a $\log_2$ (HRDE-1 IP / CSR-1 IP) of less than -2.8, CSR-1 favored were those with a $\log_2$ of between -2.8 and -1.82, slightly CSR-1 favored were those with a $\log_2$ of between -1.82 and 0, and HRDE-1 favored were those CSR-1 targeted transcripts with a $\log_2$ greater than 0.

3.6.11 Data availability

All sequencing data generated in this study (mRNA-seq, small RNA-seq, and RIP-seq) are available at the NCBI SRA database with accession number PRJNA802581 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA802581). Small RNA-seq from the *parn-1* mutant [152] can be found with accession number SRS1021265 (https://www.ncbi.nlm.nih.gov/sra/?term=SRS1021265). The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the accession number PXD033506 (https://www.ebi.ac.uk/pride/archive/projects/PXD033506). Source data are provided with this paper.

3.7 Acknowledgements

We thank Dr. Edwin Ferguson, Dr. Karen Bennett, and members of the Lee lab for critical comments on the manuscript. Some strains used in this study were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). This work is supported in part by NIH predoctoral training grant T32 GM07197 to J.B.; the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH) grant 1UL1TR002389-01 to the Institute for Translational Medicine (ITM); the National Natural Science Foundation of China (grants 31771500 and 31922019) and the program for HUST Academic Frontier Youth Team (grant 2018QYTD11) to DZ; the Ministry of Science of Technology of Taiwan (MOST 108-2628-E-006-004-MY3 and MOST 110-2221-E-006-198-MY3 grants to W.-S.W. the NIH P01 grant (HD078253) to Z.W, the NIH grant R01-GM132457 to H.-C.L.

3.8 Supplementary Information

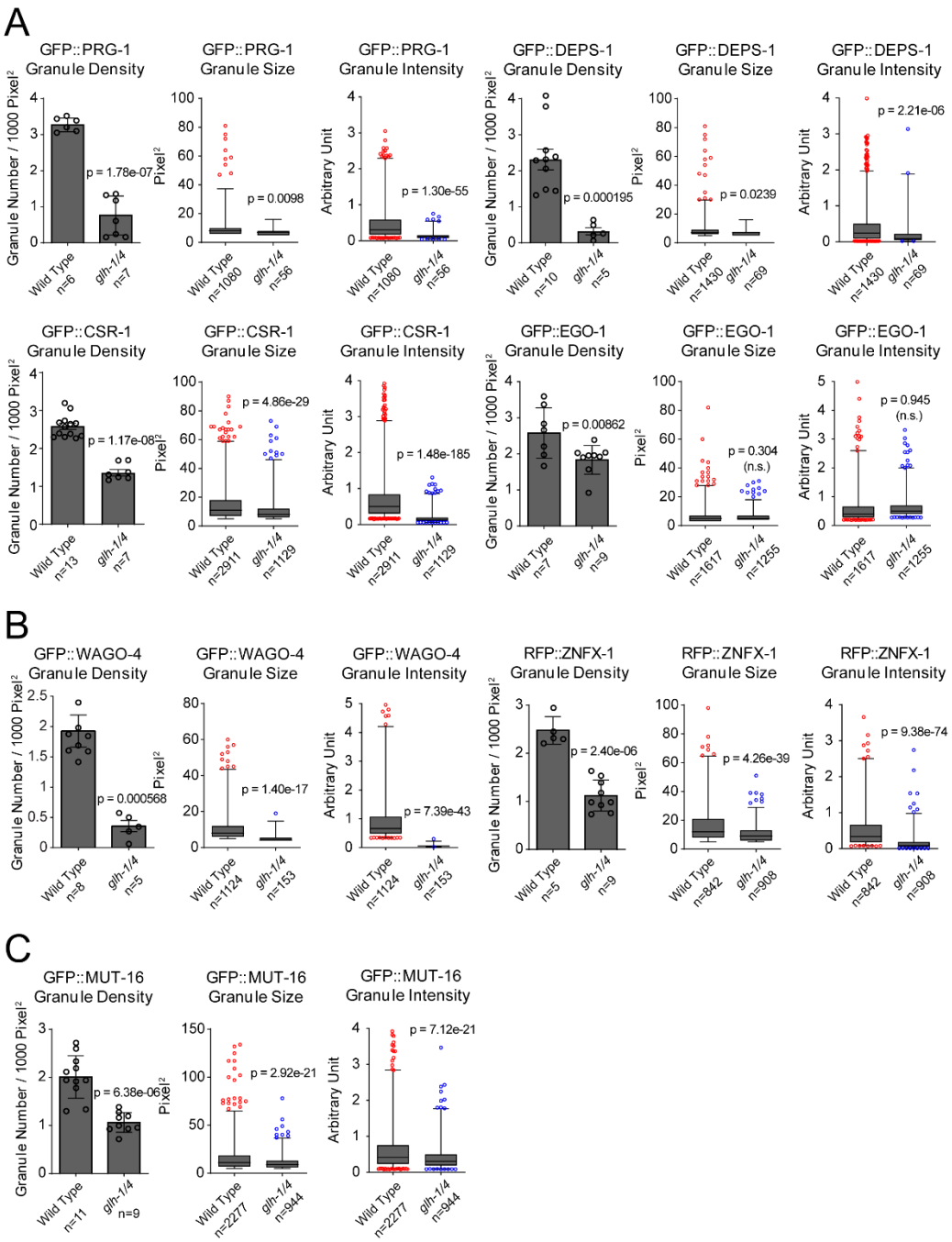


Figure 3.7: Image analyses corresponding to Figure 3.1A-C.
(A) Granule density, size, and intensity quantification of proteins known to be enriched in P granules in the wild type or the *glh-1 glh-4* mutant, corresponding to micrographs in Figure 1b. Statistical analysis was performed using a one-tailed Student's t-test. Bars indicate the mean, errors bars indicate the standard deviation, and data points indicate values for individual worms (density) or granules (size and intensity) from each genotype. N values indicate number of worms (density) or granules (size and intensity) examined.

Figure 3.7 (continued): (B) Granule density, size, and intensity quantification of proteins known to be enriched in Z granules in the wild type or the *glh-1 glh-4* mutant, corresponding to micrographs in Figure 1c. Statistical analysis was performed using a one-tailed Student's t-test. Bars indicate the mean, errors bars indicate the standard deviation, and data points indicate values for individual worms (density) or granules (size and intensity) from each genotype. N values indicate number of worms (density) or granules (size and intensity) examined. (C) Granule density, size, and intensity quantification of proteins known to be enriched in mutator granules in the wild type or the *glh-1 glh-4* mutant, corresponding to micrographs in Figure 1d. Each bar represents n = 8~10 worms. Statistical analysis was performed using a one-tailed Student's t-test. Bars indicate the mean, errors bars indicate the standard deviation, and data points indicate values for individual worms (density) or granules (size and intensity) from each genotype. N values indicate number of worms (density) or granules (size and intensity) examined.

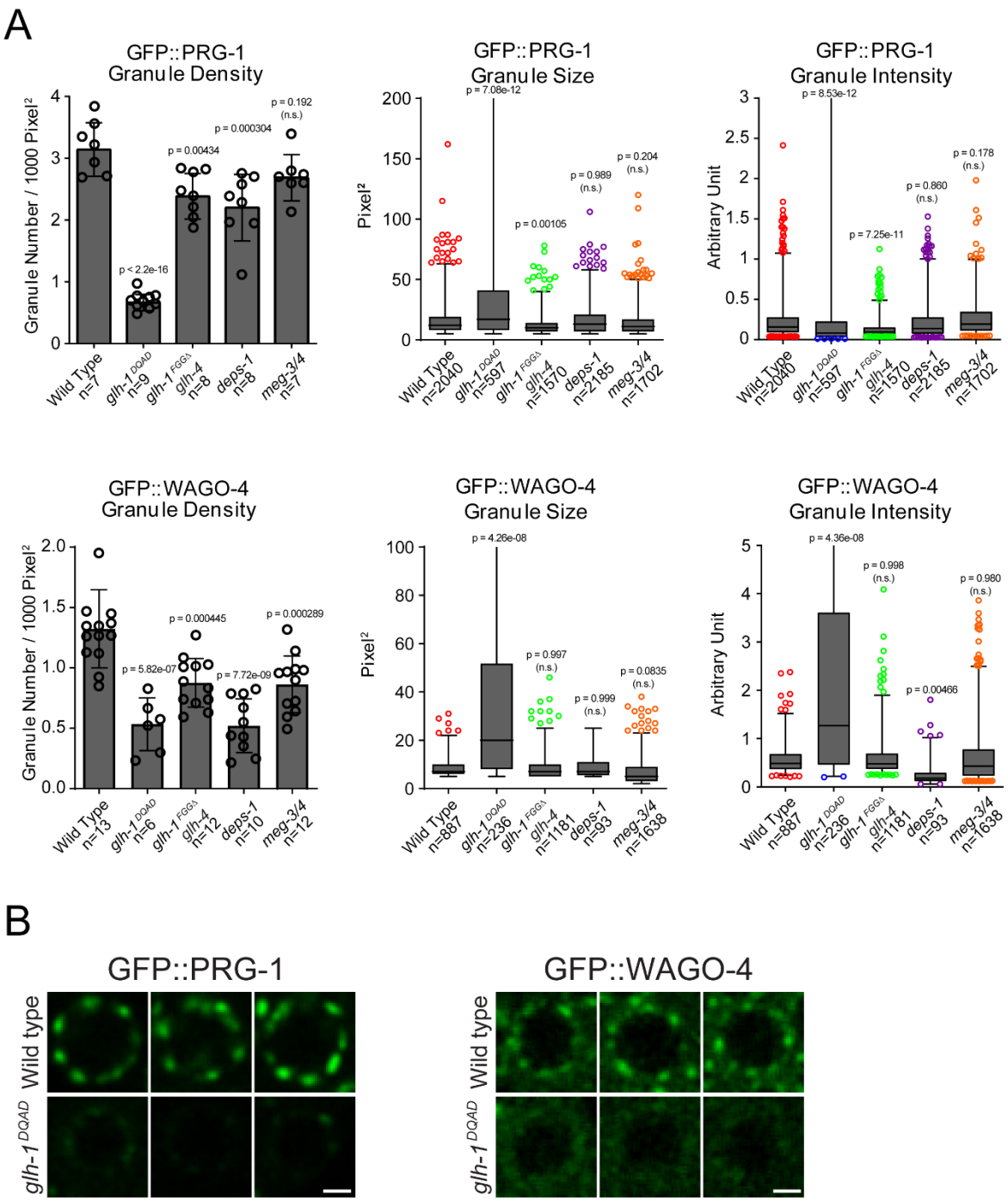


Figure 3.8: Image analyses corresponding to Figure 3.2B.
(A) Granule density, size, and intensity quantification of GFP::PRG-1 and GFP::WAGO-4 in the indicated genotypes, corresponding to micrographs in Figures 3.2B and 3.3A. Statistical analysis was performed using a one-way ANOVA and Tukey's correction for multiple comparisons were used. Bars indicate the mean, errors bars indicate the standard deviation, and data points indicate values for individual worms (density) or granules (size and intensity) from each genotype. N values indicate number of worms (density) or granules (size and intensity) examined. (B) Single nuclei from micrographs presented in Figure 3.2B. Bar, 2 micrometers.

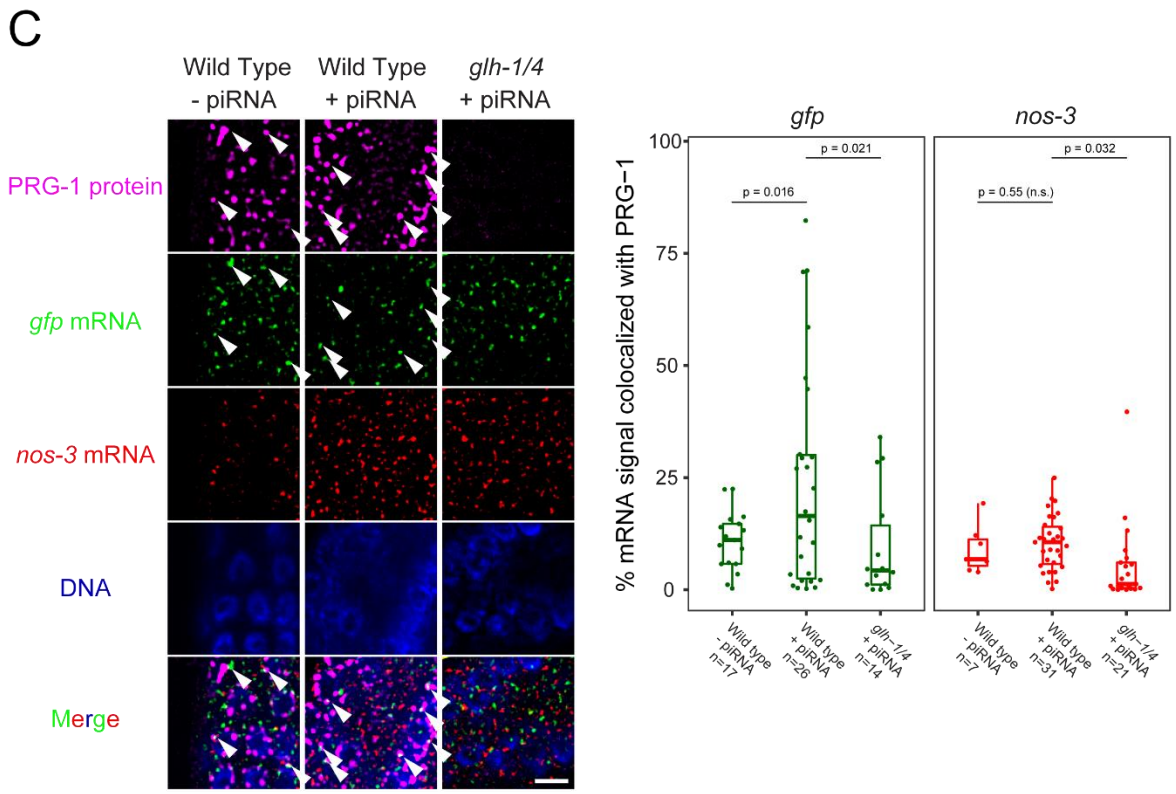


Figure 3.8 (continued): (C) Colocalization analysis of mRNA signal (*gfp* and *nos-3*) with PRG-1 signal. An example single z-slice image from each genotype is presented to the left. Arrowheads indicate pixels showing *gfp* mRNA signal colocalizing with PRG-1 signal. Colocalized pixels are highlighted in white in the merged image. Bar, 5 micrometers. Quantification is shown to the right. Each dot represents a z-series from a single gonad of the indicated genotype, probed under the indicated conditions. Data points represent biological replicates in two independent experiments. Statistical analysis was performed using a two-tailed Student's t-test. For all boxplots, lines display median values, boxes display first and third quartiles, and whiskers display 5th and 95th percentiles.

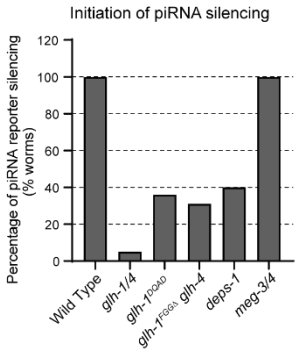


Figure 3.9: Mutants with defects in PRG-1 perinuclear condensates exhibit defects in triggering piRNA silencing.
The percentage of F2 generation worms that exhibit silencing of a GFP transgene in the indicated strains. For *meg-3 meg-4* mutants, n = 14. For all other strains, n > 50.

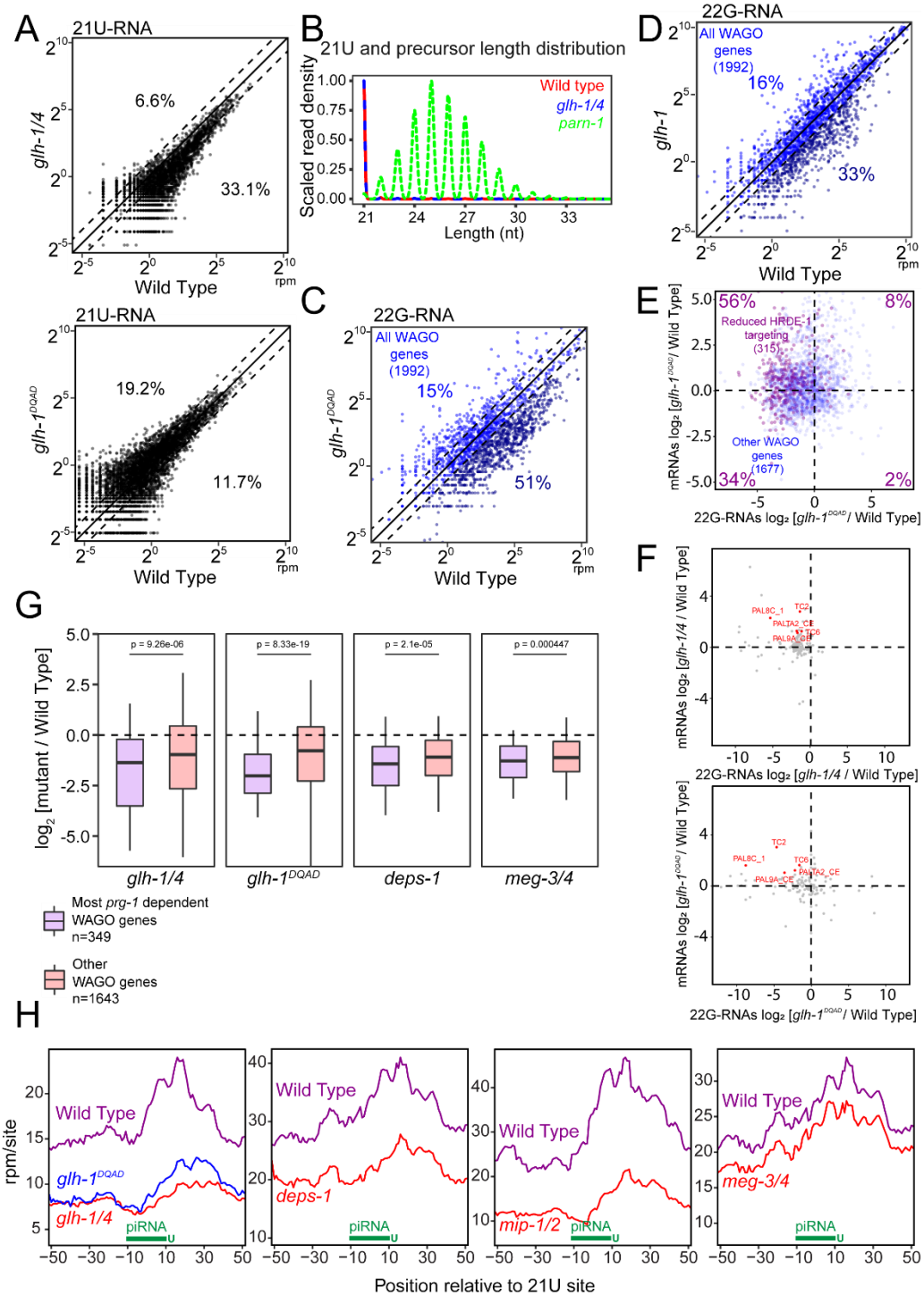


Figure 3.10: GLH/VASA mutants exhibit defects in producing secondary WAGO-22G-RNAs (Corresponding to Figure 3.4).

(A) A scatter plot showing the abundance of each piRNA (21U-RNA) in *glh-1 glh-4* and *glh-1 DQAD* mutant worms compared to wild type. Diagonal lines indicate a two-fold increase (top), no change (middle), or a two-fold depletion (bottom) in the indicated mutant strains. Percentages indicate the proportion of piRNAs increased or decreased 2-fold compared to wild type.

Figure 3.10 (continued): (B) Scaled kernel density estimation of mature piRNAs (21nt) and their precursors (>21nt) in the indicated strains. (C-D) Scatter plots showing the abundance of all 22G-RNAs mapped to each WAGO targets in wild type worms compared to the indicated mutants. The percentage of WAGO targets with 2-fold increased or decreased 22G-RNAs in mutants are shown. Diagonal lines indicate a two-fold increase (top), no change (middle), or a two-fold depletion (bottom) in the mutant. (D) A scatter plot showing the mRNA vs 22G-RNA expression changes for WAGO targets in *glh-1 DQAD* mutants vs wild type worms. Percentages indicate the proportion of WAGO targets with reduced HRDE-1-associated 22G-RNAs that fall in that quadrant. (E) A scatter plot showing the mRNA vs 22G-RNA expression changes for transposon mRNA in the indicated mutants vs wild type worms. Transposons with increased mRNAs and decreased 22G-RNAs in both *glh-1 glh-4* and *glh-1 DQAD* mutants are highlighted in red. 22G-RNA data reflect data collected from a single experiment. mRNA data reflect data collected across 2 biological replicates. (F) 22G-RNA fold changes for WAGO targets most dependent on *prg-1* for 22G-RNA accumulation in the indicated mutants compared to wild type. Most *prg-1* dependent WAGO targets defined as those WAGO targets with a significant (adjusted $P < 0.05$ [see Methods for details] and 2-fold) decrease in 22G-RNA accumulation in *prg-1* mutants. Statistical analysis was performed using a two-tailed Mann-Whitney Wilcoxon test. For all boxplots, lines display median values, boxes display first and third quartiles, and whiskers display 5th and 95th percentiles. (G) Density of 22G-RNAs within a 100 nt window around predicted piRNA target sites in the indicated strains. Computed by summing 22G-RNA density per piRNA targeting site in all WAGO targets. Plots are centered on the 10th nucleotide of piRNAs.

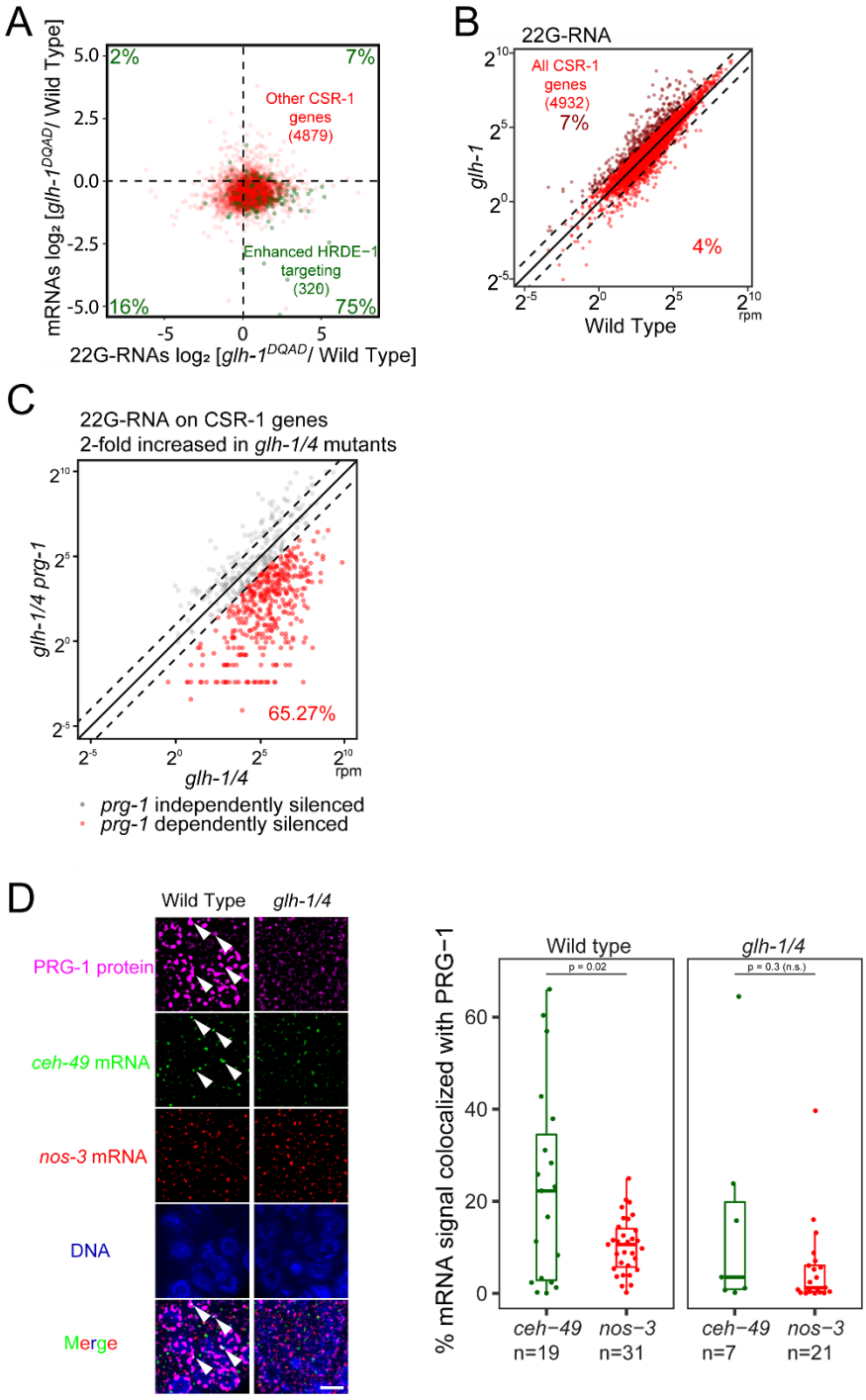


Figure 3.11: Many functional germline genes are silenced in mutants with defects in forming perinuclear PRG-1 granules (Corresponding to Figure 3.5).
(A) A scatter plot showing the mRNA vs 22G-RNA $\log_2$ expression changes for CSR-1 targets in *glh-1 DQAD* mutants vs wild type worms. The bottom right quadrant corresponds to CSR-1 targets that have become silenced in the mutant (decreased mRNAs and increased 22G-RNAs). Percentages are shown in each quadrant to indicate the proportion of CSR-1 targets with enhanced HRDE-1-associated 22G-RNAs in the *glh-1* mutant that fall in that quadrant.

Figure 3.11 (continued): (B) A scatter plot showing the abundance of all 22G-RNAs mapped to each CSR-1 targets in wild type worms compared to *glh-1* mutants. The percentage of CSR-1 targets with 2-fold increased or decreased 22G-RNAs in mutants are shown. The three diagonal lines indicate a two-fold increase (top), no change (middle), or a two-fold depletion (bottom) in *glh-1* mutants. (C) A scatter plot showing 22G-RNA accumulation on CSR-1 genes with a 2-fold increase in 22G-RNA accumulation in *glh-1 glh-4* mutants. The *glh-1 glh-4 prg-1* triple mutant is compared to the *glh-1 glh-4* double mutant. In red, enhanced CSR-1 targets that are suppressed by *prg-1* loss are shown. The three diagonal lines indicate a two-fold increase (top), no change (middle), or a two-fold depletion (bottom) in the indicated mutant strains. Distributions represent data collected across 2 biological replicates. (D) Colocalization analysis of mRNA signal (*ceh-49* and *nos-3*) with PRG-1 signal. An example single z-slice image from each genotype is presented to the left. Arrowheads indicate pixels showing *ceh-49* mRNA signal colocalizing with PRG-1 signal. Colocalized pixels are highlighted in white in the merged image. Bar, 5 micrometers. Quantification is shown to the right. Each dot represents a z-series from a single gonad of the indicated genotype, probed under the indicated conditions. Data points represent biological replicates in two independent experiments. Statistical analysis was performed using a two-tailed Student's t-test. For all boxplots, lines display median values, boxes display first and third quartiles, and whiskers display 5th and 95th percentiles.

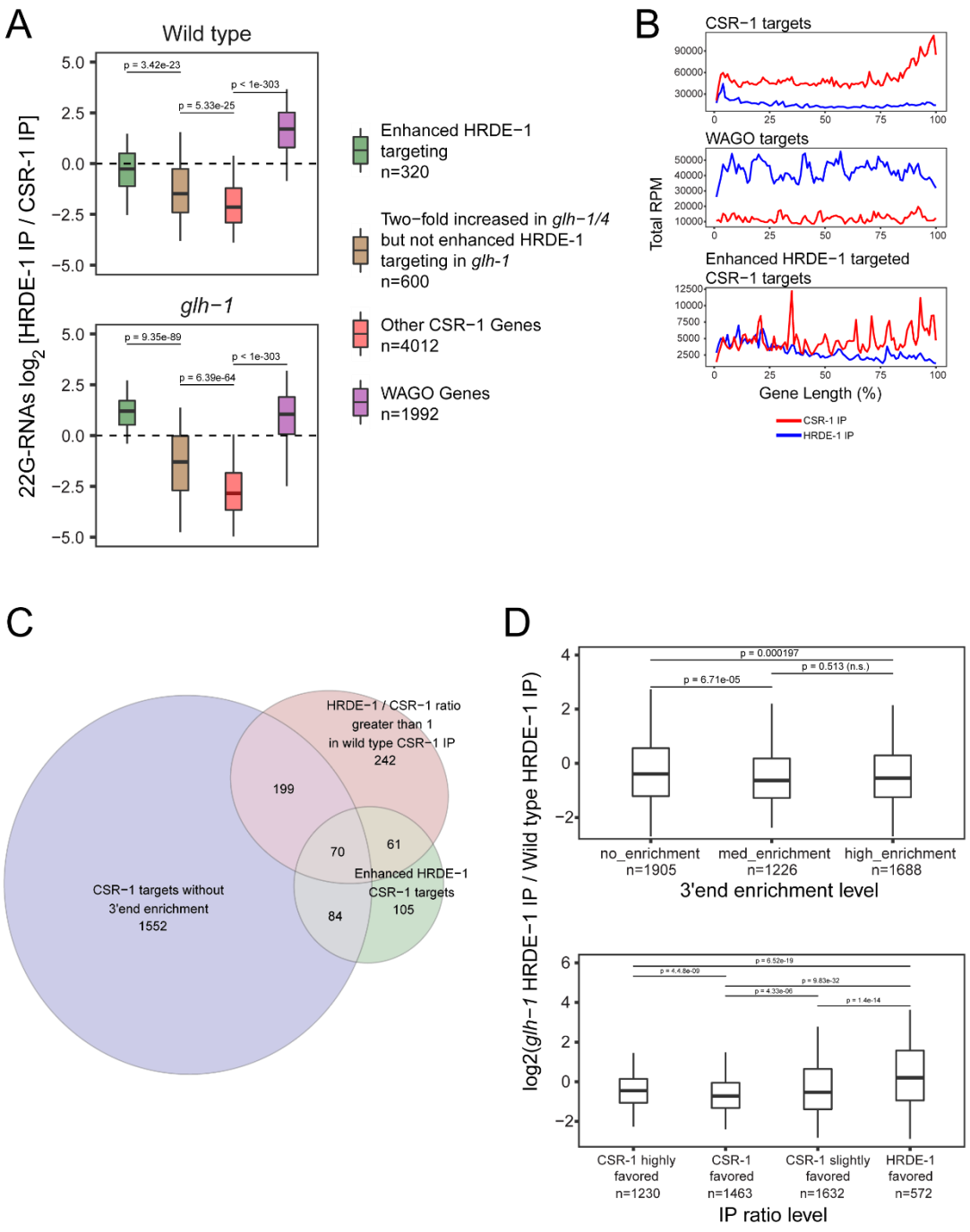


Figure 3.12: HRDE-1 / CSR-1 IP ratio and CSR-1 IP 3' end enrichment are both associated with enhanced HRDE-1 targeting of CSR-1 genes in VASA mutants.
(A) 22G-RNA enrichment in HRDE-1 versus CSR-1 IP experiments in the indicated strains for CSR-1 targets and WAGO targets. Dotted line indicates no enrichment for either Argonaute. Statistical analysis was performed using a two-tailed Mann-Whitney Wilcoxon test. For all boxplots, lines display median values, boxes display first and third quartiles, and whiskers display 5th and 95th percentiles. Distributions represent data collected across 2 biological replicates for CSR-1 IP and a single experiment for HRDE-1 IP.

Figure 3.12 (continued): (B) Metagene traces showing the total accumulation of 22G-RNAs by percentage of gene length in IP experiments. Traces are shown for three classes of genes: all CSR-1 targets (left), all WAGO targets (middle), and CSR-1 targets with enhanced HRDE-1 targeting (right). Traces from CSR-1 IP experiments are shown in red and traces from HRDE-1 IP experiments are shown in blue. (C) Venn diagram showing the overlap between CSR-1 targets that fail to show a 3' end enrichment in wild type CSR-1 IP (see Methods), CSR-1 targets with a HRDE-1 to CSR-1 IP ratio greater than 1, and CSR-1 targets with enhanced HRDE-1 targeting in *glh-1* mutants. (D) Characterization of the effect of CSR-1 22G-RNA 3' end enrichment on the level of HRDE-1 enhancement in *glh-1* mutants (top), and the effect of HRDE-1 to CSR-1 IP ratio on the level of HRDE-1 enhancement in *glh-1* mutants (bottom). For 3' end enrichment and HRDE-1 to CSR-1 IP ratio characterization, see Methods. Statistical analysis was performed using a two-tailed Mann-Whitney Wilcoxon test. For all boxplots, lines display median values, boxes display first and third quartiles, and whiskers display 5th and 95th percentiles. Distributions represent data collected across 2 biological replicates for CSR-1 IP derived values and a single experiment for HRDE-1 IP derived values.

Data file 3.1: List of interacting proteins identified by MS with DTME-crosslinking using GLH-1.

Data file 3.2: List of smFISH probes.

CHAPTER 4

THE DETERMINANTS OF GENE SILENCING AND LICENSING IN THE *C. ELEGANS* GERMLINE

4.1 Attributions

This chapter has been adapted from: W.-S. Wu & J.S. Brown, *et al.*, Transcriptome-wide analyses of piRNA binding sites suggest distinct mechanisms regulate piRNA binding and silencing in C. elegans, RNA, rna-079441 (2023). Wei-Sheng Wu and members of his lab jointly performed bioinformatic analyses with the author.

4.2 Abstract

PIWI-interacting RNAs (piRNAs) protect genome integrity by silencing transposon mRNAs and some endogenous mRNAs in various animals. However, *C. elegans* piRNAs only trigger gene silencing at select predicted targeting sites, suggesting additional cellular mechanisms regulate piRNA silencing. To gain insight into possible mechanisms, we compared the transcriptome-wide predicted piRNA targeting sites to the *in vivo* piRNA binding sites. Surprisingly, while sequence-based predicted piRNA targeting sites are enriched in 3' UTRs, we found that *C. elegans* piRNAs preferentially bind to coding regions (CDS) of target mRNAs, leading to preferential production of secondary silencing small RNAs in the CDS. However, our analyses suggest that this CDS binding preference cannot be explained by the action of anti-silencing Argonaute CSR-1. Instead, our analyses imply that CSR-1 protects mRNAs from piRNA silencing through two distinct mechanisms – by inhibiting piRNA binding across the entire CSR-1 targeted transcript, and by inhibiting secondary silencing small RNA production locally at CSR-1 bound sites. Together, our work identifies the CDS as the critical region that is uniquely competent for piRNA binding in *C. elegans.* We speculate the CDS binding preference may have evolved to allow the piRNA pathway

to maintain robust recognition of RNA targets in spite of genetic drift. Together, our analyses revealed that distinct mechanisms are responsible for restricting piRNA binding and silencing to achieve proper transcriptome surveillance.

4.3 Introduction

Small non-coding RNAs, such as miRNAs and piRNAs, play important roles in gene regulation by guiding Argonaute proteins to target RNAs with sequence complementarity [67]. miRNAs play critical roles in various developmental processes by regulating the expression of endogenous genes [13]. PIWI-interacting RNAs (piRNAs) are germline-enriched small RNAs found in diverse animals that silence non-self nucleic acids, such as transposons and viruses [15,21,27,40,131]. One unique characteristic of piRNAs is their sequence diversity - tens of thousands of sequence-distinct piRNAs are produced. To identify mRNA targets regulated by these diverse piRNAs, both sequence-based approaches as well as experimental approaches have been reported; using bioinformatic analyses and *in vivo* reporter assays, previous studies have identified the piRNA targeting rules in *C. elegans* [137,185]. In *C. elegans*, piRNAs identify their targets through base pairing, requiring nearly perfect complementarity at seed regions and tolerating several mismatches in the remaining sequence. Several predicted piRNA targeting sites have been confirmed to play a regulatory role *in vivo*, and removal of predicted piRNA targeting sites allows for stable expression of several silencing-prone transgenes [137,185]. piRNAs are predicted to bind various germline-silenced mRNAs, and surprisingly, many germline-expressed mRNAs as well [177,185]. CLASH (crosslinking, ligation and sequencing of hybrids) is a powerful biochemical approach to identify *in vivo* small RNA binding sites [65]; protein-bound small RNAs are ligated to their target mRNAs and these hybrid molecules, comprised of small RNAs and mRNA fragments, are sequenced to identify putative direct interactions. A recent study has applied

CLASH analyses of *C. elegans* PIWI (PRG-1) to identify *in vivo* piRNA binding sites [137]. Analysis of this data revealed that piRNAs bind both germline-silenced and germline-expressed mRNAs in *C. elegans*. Therefore, both bioinformatic and experimental approaches suggest that piRNAs alone are not sufficient to distinguish self from non-self nucleic acids. Two types of endogenous small RNAs work with piRNAs to achieve proper transcriptome surveillance. WAGO associated small RNAs are produced at piRNA targeting sites to silence non-self nucleic acids [87,141]. In addition, CSR-1 Argonaute and its associated small RNAs have been shown to target germline-expressed RNAs and protect self nucleic acids from piRNA silencing [133,137,170]. While the mechanism that controls CSR-1 licensing remains largely unknown, these studies suggest an interesting model where CSR-1 and its small RNAs act as a gene-licensing pathway, critical for distinguishing self nucleic acids from non-self in *C. elegans*.

Our genome-wide analyses of piRNA targeting sites have identified surprisingly few overlapping sites between experimentally-identified piRNA binding sites and sequence-based predicted piRNA targeting sites [177]. These observations suggest that piRNA binding to mRNA targets may be regulated, such as by the CSR-1 pathway. However, CSR-1 protection alone cannot fully explain gene licensing, as CSR-1 knockdown does not result in silencing of many CSR-1 targets [25,144]. Currently it is unclear whether additional mechanisms exist to control piRNA recognition.

To gain insights into the mechanisms that regulate piRNA silencing, we compared the *in vivo* binding sites of piRNAs to the *in silico* predicted piRNA targeting sites on target mRNAs (Figure 4.1A). Specifically, we performed metagene analyses to examine the distribution of piRNA binding sites on mRNAs, and we compared binding density across various genomic features. We also examined the distribution of silencing WAGO small RNAs and licensing CSR-

1 small RNAs, and their relationship with piRNA binding sites. Surprisingly, we found that piRNAs preferentially bind to coding regions (CDS) of mRNAs. As a consequence, piRNAs trigger the production of secondary WAGO silencing small RNAs preferentially at the CDS. In contrast, our analyses showed that the sequence-based predicted piRNA targeting sites are slightly enriched in 3' UTRs, thus the preference for CDS binding by piRNAs cannot be explained by the distribution of predicted piRNA targeting sites. In addition, we found that CSR-1 licenses germline expression by inhibiting both piRNA binding and secondary silencing WAGO small RNA production, but does so through distinct modes of action. Nonetheless, our analyses suggest CSR-1 is not responsible for restricting piRNA binding to the CDS. Together, our analyses identify CDSs as the critical regions that are uniquely susceptible to piRNA silencing in *C. elegans* and revealed distinct mechanisms that regulate piRNA silencing *in viv*o.

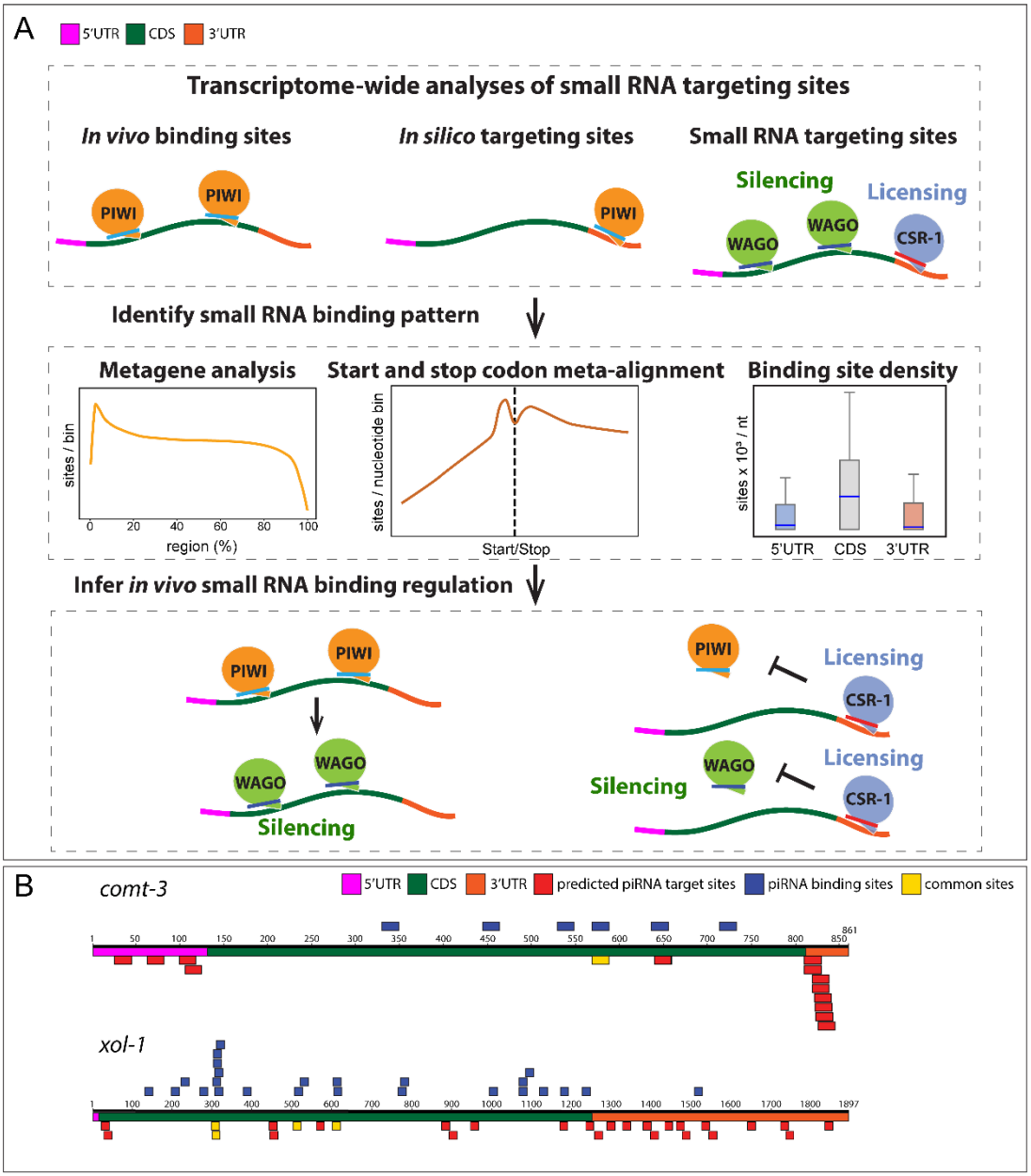


Figure 4.1: Transcriptome-wide analyses of piRNA binding sites.
(A) To investigate *in vivo* small RNA binding regulation, we analyzed *in vivo* binding site data, *in silico* predicted piRNA targeting site data, and small RNA targeting site data. Synthesizing these datasets allowed us to identify small RNA binding patterns across groups of genes and within genomic features. By comparing such pattern in distinct genetic mutants, we infer and propose possible regulatory mechanism of piRNA silencing. (B) Two examples highlight distinct patterns observed between *in vivo* piRNA binding sites and in silico predicted piRNA binding sites. The location of predicted piRNA targeting sites are shown in red, the experimentally identified piRNA binding sites are shown in blue, and the common sites (defined as sites where the same piRNA was determined by prediction and experimentally to target overlapping mRNA regions) are shown in yellow. The specific regions of the indicated mRNAs, including 5' UTR, CDS and 3' UTR are also labeled.

4.4 Results

4.3.1 piRNAs exhibit a binding preference to the coding regions of target mRNAs in C. elegans

As described above, only a small proportion of predicted piRNA targeting sites were found to be bound by piRNAs in vivo, suggesting that piRNA binding to mRNA targets is regulated by cellular mechanisms [177]. Intriguingly, while examining two well-characterized piRNA germline target mRNAs [137,151], we noticed that despite the presence of several predicted piRNA targeting sites in various regions of these mRNAs, including CDS, 5' UTRs and 3' UTRs (Figure 4.1B and for complete binding site information, see Supplementary Data file 4.1), the experimentally identified piRNA binding sites are mostly present in the CDSs, with some of these binding sites also predicted as piRNA targeting sites. We therefore wondered whether the observed piRNA CDS-binding preference could be a global trend, applicable to other germline mRNAs.

To test whether our analyses of Argonaute/small RNA binding sites are consistent with previously published results, we first examined the distribution of experimentally identified miRNA binding sites in *C. elegans*, which are known to be enriched in 3' UTRs of mRNAs [22]. Using previously published miRNA Argonaute ALG-1 iCLIP data, we identified transcriptome-wide *in vivo* miRNA binding sites. As expected, when we analyzed *lin-4* miRNA binding sites on *lin-14* mRNA, all *in vivo lin-4* binding sites are located in the 3' UTR of *lin-14* (Figure 4.2A and Supplementary Data file 4.1). We then calculated the density of all miRNA binding sites in different regions of the mRNAs, including 5' UTR, CDS or 3' UTR [176]. We observed a higher density for both the number of miRNA binding sites and the miRNA binding events at the 3' end of mRNAs, especially at the 3' UTR of mRNAs in *C. elegans* (Figure 4.2B & 4.2C and Supplementary Figure 4.7A & 4.7B). To more closely examine the local change in miRNA binding

preference between different regions, we examined the distribution of miRNA binding sites at the borders of CDS and UTRs. As expected, we observed that the number of miRNA binding sites did not change around the start codon but increased sharply immediately downstream of the stop codon (Figure 4.2D). Together, these analyses confirmed the previous observations that miRNAs predominately bind to their target mRNAs at their 3' UTRs [22].

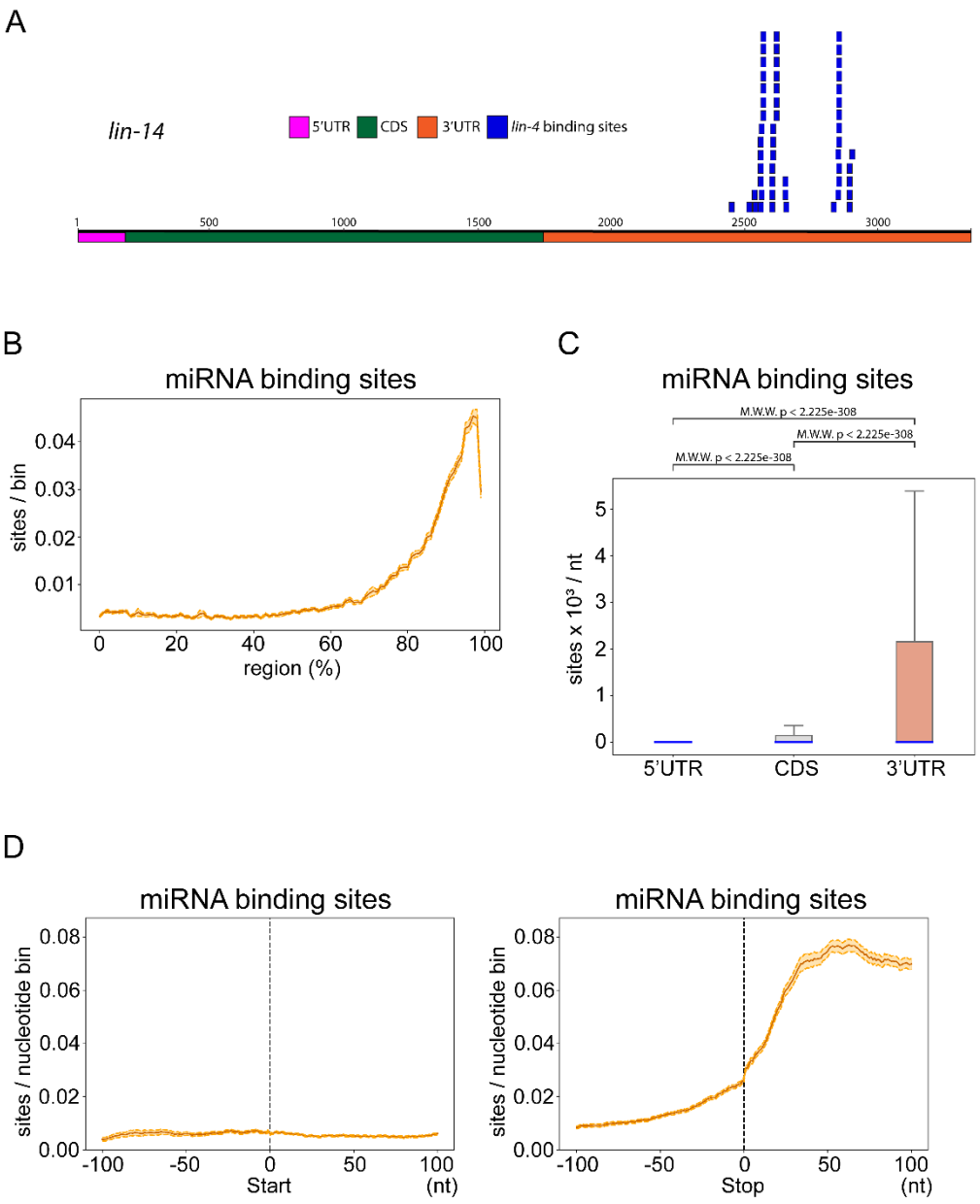


Figure 4.2: *In vivo* miRNA binding sites are enriched at the 3' UTR of mRNAs in *C. elegans*. (A) *lin-4* miRNA *in vivo* binding sites found in *lin-14* mRNA. (B) Metagene trace shows the distribution of miRNA binding sites along an mRNA. The solid line indicates the average number of miRNA binding sites in all mRNAs. The dotted lines indicate the average number plus/minus one standard error. (D) The density of miRNA binding sites in the indicated regions. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (D) The distribution of miRNA binding sites around the start codon (left) or stop codon (right). A 200 nucleotide-window centered at start or stop codon is shown. The solid line indicates the average number of miRNA binding sites in all mRNAs. For each position, the average is calculated from all mRNAs which possess that position. For example, not all mRNAs have 5'UTRs but they all have CDSs. The dotted lines indicate the average number plus/minus one standard error.

We then compared the genome-wide distribution of sequence-based predicted piRNA targeting sites and experimentally identified piRNA binding sites on germline mRNAs, since piRNAs are mostly expressed in the germline [111,178]. Using the stringent piRNA targeting rule identified in previous studies [178,185], we observed that the predicted piRNA targeting sites are enriched at the 3' end of germline mRNAs (Figure 4.3A), and their densities are highest in the 3' UTR of germline mRNAs (Figure 4.3B). Using published PIWI PRG-1 CLASH data [137], we then analyzed the transcriptome-wide *in vivo* piRNA binding sites on germline mRNAs [176]. Interestingly, we observed a very distinct distribution of *in vivo* piRNA binding sites compared to that of the predicted piRNA targeting sites; the overall distribution of piRNA binding sites on germline mRNAs exhibits a peak at the 5' end, followed by a steady level across the gene body before a decrease at the 3' end (Figure 4.3C). Consistent with the CDS binding preference found in the well-characterized piRNA targeted mRNAs, we found that the density of *in vivo* piRNA binding sites is highest in the CDS of germline mRNAs (Figure 4.3D). Specifically, the piRNA binding sites increased most notably around the start codon and decreased most notably around the stop codon (Figure 4.3E) – opposite to the distribution of the predicted piRNA targeting sites (Supplementary Figure 4.8A). These observations suggest that piRNAs preferentially bind to CDSs, and this preference cannot be explained by the distribution of predicted piRNA targeting sites.

As some miRNAs are also expressed in the germline, we wondered whether the enrichment of piRNA binding sites in CDSs is a feature shared by germline-expressed miRNAs [101]. We therefore restricted our miRNA binding site analysis to germline-enriched miRNAs at germline mRNAs. However, we found that miRNA binding sites for these germline-enriched miRNAs remain enriched at the 3' UTR and 3' end of germline mRNAs (Supplementary Figure 4.8B).

Therefore, the piRNA CDS-binding preference is not shared by miRNAs in the germline. We wondered whether binding of miRISC complexes in the 3' UTR could explain the exclusion of piRNA binding to 3' UTRs. If so, those mRNAs without miRNA binding sites at their 3' UTRs should exhibit a reduced CDS-binding preference and enhanced 3' UTR binding. We analyzed piRNA binding sites in germline mRNA transcripts which contain predicted binding sites of germline miRNAs in their 3'UTRs and those that do not. We found that piRNA binding is heavily enriched in the CDS for both of these groups (Supplementary Figure 4.8C). Additionally, we analyzed piRNA binding site density in germline transcripts with experimentally identified germline miRNA binding sites in their 3'UTRs and those without (Supplementary Figure 4.8D). Again, we found that piRNAs still heavily favor CDS binding even without 3'UTR germline miRNA binding. Therefore, despite germline miRNAs' preference for binding 3' UTRs (Supplementary Figure 4.7B), miRNA binding at 3' UTRs is not sufficient to explain the piRNA binding preference for the CDS. Since previous studies have shown that germline-expressed transcripts are protected from piRNA silencing by CSR-1 and its associated small RNAs [133,137,170], we also examined whether there is a difference in piRNA binding distribution on germline-expressed mRNAs (here defined as CSR-1 targets) or germline-silenced mRNAs (here defined as WAGO targets, see more descriptions of WAGO Argonautes below). We found that the CDS enrichment for piRNA binding sites exists for both WAGO targets as well as for CSR-1 targets (Supplementary Figure 4.8E) [33,61]. This observation suggests that the CDS binding preference is likely not caused by CSR-1 (see more about CSR-1 function below). Together, we conclude that the *in vivo* piRNA binding sites are enriched in the CDSs of germline-expressed mRNAs in *C. elegans*.

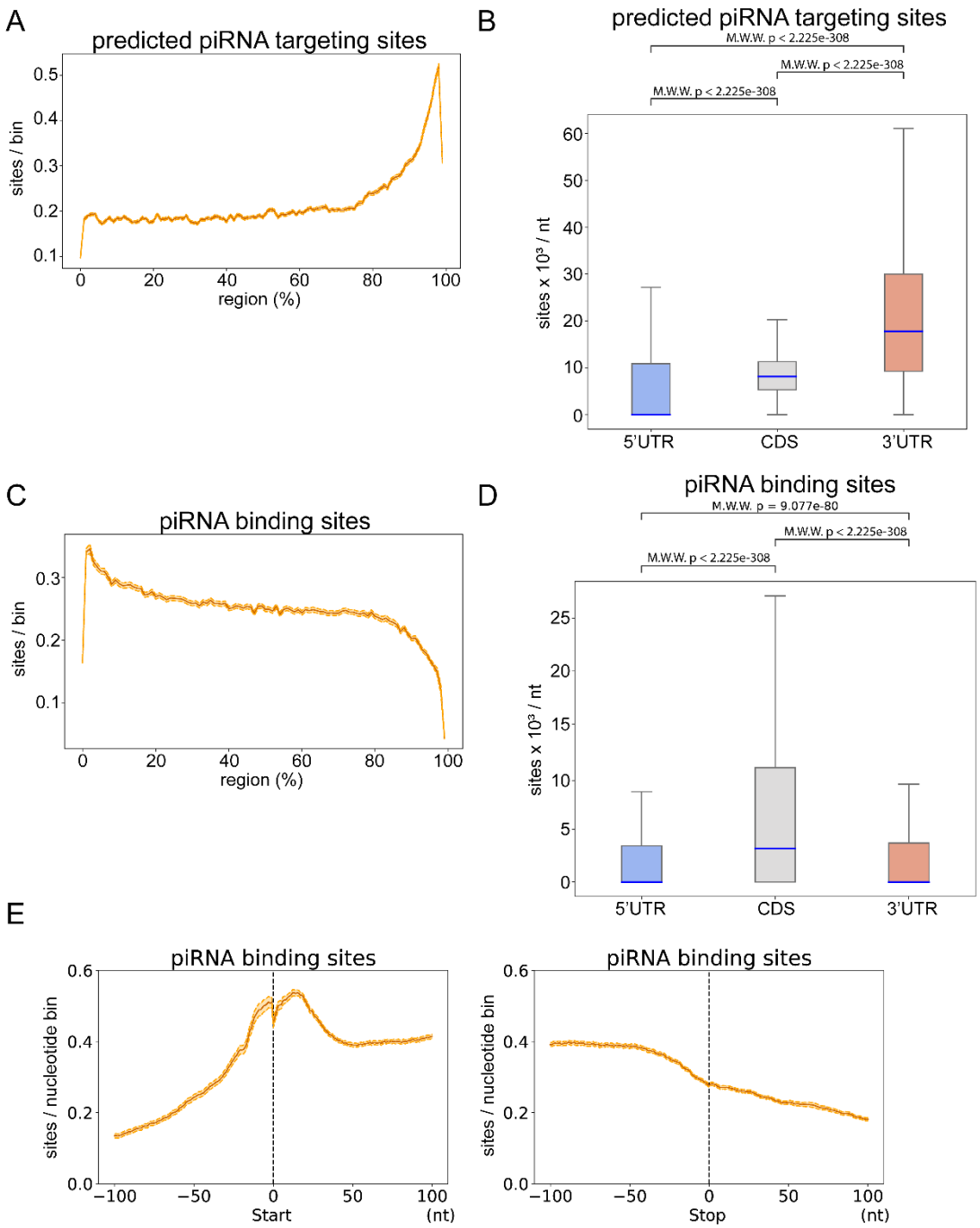


Figure 4.3: *In vivo* piRNA binding sites are enriched in the coding region (CDS) of germline mRNAs in *C. elegans*.

(A) Metagene trace shows the distribution of predicted piRNA targeting sites along a germline mRNA. The solid line indicates the average number of piRNA targeting sites in all germline mRNAs. The dotted lines indicate the average number plus/minus one standard error. (B) The density of predicted piRNA targeting sites in the indicated regions of germline mRNAs. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test.

Figure 4.3 (continued): (C) Metagene trace shows the distribution of experimentally identified piRNA binding sites along a germline mRNA. The solid line indicates the average number of piRNA binding sites in all germline mRNAs. The dotted lines indicate the average number plus/minus one standard error. (D) The density of experimentally identified piRNA binding sites in the indicated regions of germline mRNAs. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (E) The distribution of piRNA binding sites around the start codon (left) or stop codon (right). A 200 nucleotide-window centered at start or stop codon is shown. The solid line indicates the average number of piRNA binding sites in all germline mRNAs. For each position, the average is calculated from all germline mRNAs which possess that position.

4.3.2 piRNAs preferentially trigger the production of secondary small RNAs at mRNA coding regions

piRNAs induce gene silencing by triggering the production of secondary silencing small RNAs that associate with Worm specific ArGOnautes (WAGOs), also known as WAGO 22G-RNAs [8,10,87]. These WAGO 22G-RNAs are produced by RNA-dependent RNA polymerase locally at piRNA binding sites. Since piRNAs preferentially bind to CDSs, we wondered whether the production of WAGO 22G-RNAs is also enriched at CDS regions. To test this hypothesis, we compared the density of WAGO 22G-RNAs in distinct regions of WAGO targeted mRNAs. Indeed, for germline-silenced mRNAs (WAGO targets) [61], we observed that both WAGO-9 (also known as HRDE-1) and WAGO-1, bound 22G-RNAs are significantly enriched in the CDS (Figure 4.4A). In addition, for CSR-1 targeted mRNAs, WAGO 22G-RNAs are also enriched in the CDS (Figure 4.4A), despite their overall WAGO 22G-RNA levels being much lower in CSR-1 targets than in WAGO targets. WAGO 22G-RNA production can be induced by piRNAs and other small RNA pathways [34,49]. Therefore, if piRNA-induced WAGO 22G-RNAs are produced preferentially at the CDS, we expect that in PIWI mutants, which lose all piRNAs, the WAGO 22G-RNAs should be preferentially reduced at coding regions. Indeed, when we compared the levels of WAGO-1 22G-RNAs between wild type and PIWI *prg-1* mutants [14], we observed a greater reduction of WAGO-1 22G-RNAs in the CDS than that in 5' or 3' UTRs, for both

germline-silenced and germline-expressed mRNAs (Figure 4.4B). These observations suggest that the preferential piRNA binding at the CDS of mRNAs is functionally relevant, as piRNA binding preference in the CDS leads to a corresponding enrichment of secondary WAGO 22G-RNAs produced at the CDS. Furthermore, this confirms that the piRNA binding pattern we observed was not simply due to a technical bias of the CLASH method, as this independent signature of piRNA binding corroborates our findings. Interestingly, a recent study compared the ability of synthetic piRNAs to trigger gene silencing and showed that those synthetic piRNAs targeting the CDS can trigger robust gene silencing, but for synthetic piRNA targeting 5' or 3'UTRs, gene silencing is not as successful [121]. These reporter-based experiments support our transcriptome-wide analyses that the CDS is the critical region that is uniquely competent for piRNA silencing in *C. elegans*.

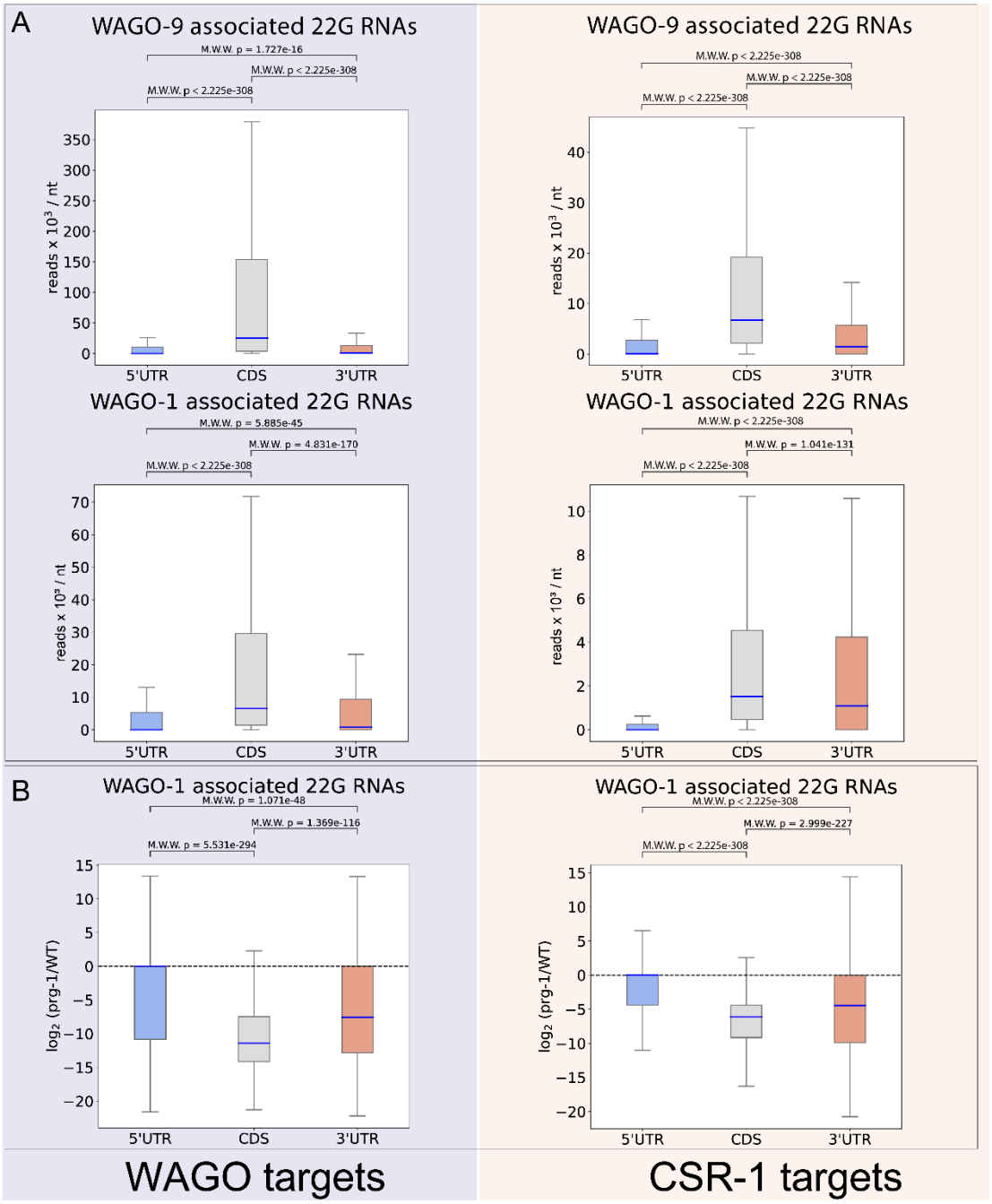


Figure 4.4: Secondary WAGO 22G-RNAs are preferentially produced in coding regions.
(A) The density of WAGO-9 (top) and WAGO-1 (bottom) 22G-RNAs in the indicated regions of WAGO targets (left) or CSR-1 (right) targets. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (B) The ratio of WAGO-1 22G-RNAs in the *prg*-1 mutant over those in wild type in the indicated regions of WAGO (left) or CSR-1 (right) targets. Blue lines represent median values. The statistical significance (P-value) of the difference in fold change (*prg-1* mutant/WT) between two regions was calculated by the Mann-Whitney U test.

4.3.3 CSR-1 inhibits piRNA binding and WAGO 22G-RNA amplification with distinct modes of action

CSR-1 Argonaute and its associated 22G-RNAs have been reported to protect germline transcripts from piRNA silencing [133,170]. However, the mechanism behind CSR-1 Argonaute's anti-piRNA silencing function remains unclear. A previous study suggests that CSR-1 inhibits the piRNA pathway by interfering with piRNA binding [137]. Nonetheless, it is unclear whether CSR-1 also inhibits downstream WAGO 22G-RNA synthesis. In addition, it is unknown whether CSR-1 binding leads to protection of the whole transcript from piRNA silencing, or CSR-1 only interferes with piRNA silencing locally, proximal to CSR-1 targeting sites. To gain insights into CSR-1's anti-piRNA silencing function, we first examined the distribution of CSR-1 22G-RNAs on germline mRNAs. We found that the levels of CSR-1 22G-RNAs are higher at the 3' end of CSR-1 targets (Figure 4.5A, top) and are produced at both CDSs and 3' UTRs (Supplementary Figure 4.9A, top), consistent with a recent report [144]. In addition, we found a low level of CSR-1 22G-RNAs produced from WAGO targets. These CSR-1 22G-RNAs are more equally distributed across the gene body, with their levels reduced at both 5' and 3' ends (Figure 4.5A, bottom), and enriched at the CDS (Supplementary Figure 4.9A, bottom).

To examine the relationship between CSR-1 binding and piRNA binding, we then examined the effect of CSR-1 depletion on piRNA binding [137]. Our analyses showed that CSR-1 depletion leads to a greater increase of piRNA binding in CSR-1 targets than in WAGO-1 targets, indicating that the greater level of CSR-1 22G-RNAs leads to greater levels of protection from piRNA binding (Figure 4.5B and Supplementary Figure 4.9B). If CSR-1 inhibits piRNA binding locally, we expect a greater increase of piRNA binding at regions where CSR-1 levels are higher. However, despite the very different distribution of CSR-1 22G-RNA accumulation on WAGO

targets and CSR-1 targets, we observed a similar increase of piRNA binding in all regions of the transcripts for both WAGO and CSR-1 targets (Figure 4.5B and Supplementary Figure 4.9B) upon CSR-1 depletion. Therefore, while the overall level of CSR-1 targeting a given transcript correlates with its protection level against piRNA binding, the location of CSR-1 targeting within the transcript does not seem to matter. To further examine the relationship between CSR-1 localization and piRNA binding, we defined subgroups of CSR-1 targets for CSR-1 targets where CSR-1 22G-RNAs are either enriched at their 3' UTRs (n=2082) or are not enriched at any specific regions (n=763) (Figure 4.5C and Supplementary Figure 4.9C). Again, upon CSR-1 depletion, we found that piRNA binding sites increased in all regions of both subgroups of CSR-1 targets despite their different CSR-1 distributions (Figure 4.5D and Supplementary Figure 4.9D). These results suggest that CSR-1's inhibition of piRNA binding does not correlate with local CSR-1 22G-RNA levels.

While our analyses show that CSR-1 inhibits piRNA binding, we observed that piRNAs bind germline-expressed CSR-1 targets with higher density than they bind germline-silenced WAGO targets (Supplementary Figure 4.9E). This apparent contradiction could be explained by a significantly higher expression level for CSR-1 targets, providing more opportunities for piRNA binding. To test this hypothesis, we normalized the piRNA binding density to mRNA expression levels. Indeed, we found piRNA density is significantly greater for WAGO targets than for CSR-1 targets on a per transcript basis (Supplementary Figure 4.9F). Nonetheless, significantly more WAGO 22G-RNAs (both WAGO-1 and WAGO-9 22G-RNAs) were produced from all regions of WAGO targets compared to CSR-1 targets (Supplementary Figure 4.9G), indicating that piRNA binding on CSR-1 targets does not trigger WAGO 22G-RNA production as effectively as piRNA binding on WAGO targets. This observation suggests that CSR-1 not only inhibits piRNA binding,

it may also inhibit the downstream production of WAGO 22G-RNAs. To examine the relationship between CSR-1 binding and WAGO 22G-RNA synthesis, we examined the change of WAGO-9 (HRDE-1) 22G-RNAs upon CSR-1 depletion [144]. Interestingly, for both CSR-1 and WAGO targets, there is a greater increase in WAGO-9 22G-RNAs preferentially at regions where CSR-1 22G RNAs are enriched, including the 3' end of CSR-1 targets and at the gene body of WAGO targets, respectively (Figure 4.5E and Supplementary Figure 4.9H). To further examine the relationship between CSR-1 distribution and WAGO 22G-RNA production, we compared those CSR-1 target subgroups mentioned above that have either CSR-1 22G-RNAs enriched at 3' UTRs or not enriched at a specific region. Upon CSR-1 depletion, we found a greater increase in WAGO 22G-RNA production at 3' UTRs for those CSR-1 targets with a 3' CSR-1 targeting enrichment (Figure 4.5F and Supplementary Figure 4.9I). These observations show that the local CSR-1 22G-RNA levels contribute to the inhibitory strength against WAGO 22G-RNA production, indicating that CSR-1 acts locally to inhibit the production WAGO 22G-RNAs.

Our analyses suggest that CSR-1 acts to combat piRNA binding in a transcript-wide manner, evidenced by CSR-1 depletion leading to uniformly increased piRNA binding across transcripts (Figure 4.5D). However, we also showed that CSR-1 acts locally to prevent WAGO 22G-RNA accumulation, where the most CSR-1 dense regions of a transcript are most protected from WAGO 22G-RNA amplification (Figure 4.5F). To directly compare the effects of CSR-1 targeting on piRNA binding and WAGO 22G-RNA synthesis, we identified 100nt windows within each germline mRNA that are most heavily and most poorly targeted by CSR-1. We then compared the impact of CSR-1 depletion on piRNA binding and WAGO 22G-RNA production within each window. If CSR-1 combats piRNA binding in a transcript-wide manner, we would expect there to be little difference in the increase in piRNA binding density upon CSR-1 depletion between the

most heavily and poorly CSR-1 targeted regions. Conversely, if CSR-1 prevents local WAGO 22G-RNA amplification, we expect there would be a significantly greater increase in WAGO 22G-RNA production in heavily CSR-1 targeted regions compared to the production in poorly targeted regions. Indeed, the increase in piRNA binding following CSR-1 depletion is not significantly different between high-density and low-density CSR-1 targeting windows (Figure 4.5G). In addition, we found that for WAGO 22G-RNA production, there is strikingly more 22G-RNA accumulation in high-density CSR-1 windows following CSR-1 depletion compared to low-density CSR-1 windows for germline expressed mRNAs (Figure 4.5G). A similar trend of piRNA binding and WAGO 22G-RNA production was found (Figure 4.5G) for CSR-1 target genes. For WAGO targets, we found that depletion of CSR-1 has little effect on piRNA binding, but leads to a slight insignificant increase in WAGO-22G-RNAs in high density CSR-1 windows. Taken together, our analyses suggest that CSR-1 can counteract the piRNA pathway through the inhibition of both piRNA binding and WAGO 22G-RNA production, but these two functions of CSR-1 exhibit distinct modes of action, where only the inhibition of WAGO 22G-RNA synthesis is sensitive to the levels of local CSR-1 22G-RNAs.

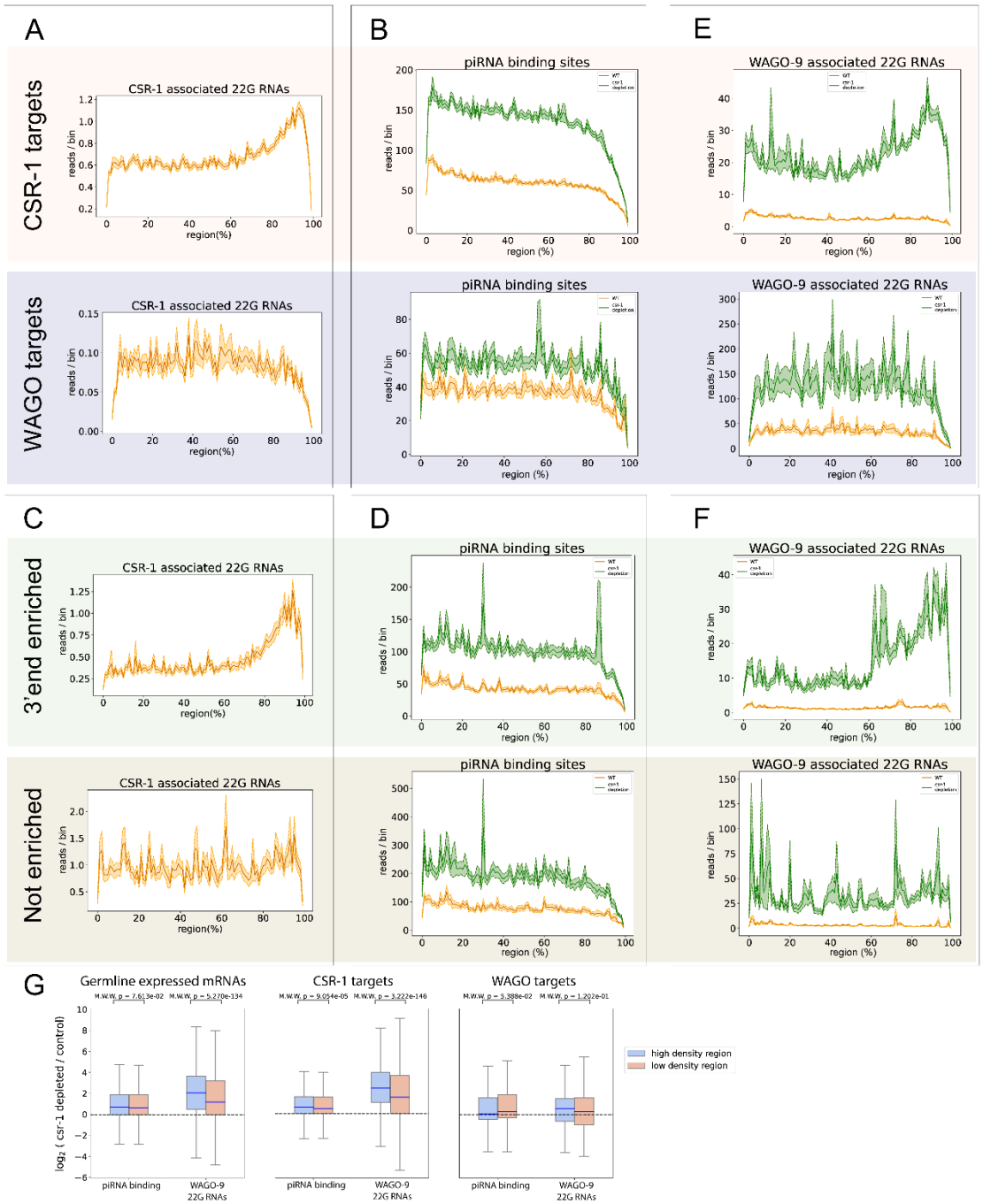


Figure 4.5: CSR-1 inhibits piRNA binding and WAGO 22G-RNA synthesis with distinct modes of action.

(A) Metagene trace shows the distribution of CSR-1 22G-RNAs along CSR-1 (top) or WAGO (bottom) targets. The solid line indicates the average number of CSR-1 22G-RNA read counts in CSR-1 or WAGO targeted mRNAs. The dotted lines indicate the average number plus/minus one standard error.

Figure 4.5 (continued): (B) Metagene trace shows the distribution of piRNA binding events in wild type (orange) or in CSR-1 depleted animals (green) mapped to CSR-1 (top) or WAGO (bottom) targets. The piRNA-tRNA hybrid reads, which are expected to be non-specific ligation products and thus remain unchanged between samples, are used for normalization. The solid line indicates the average number of piRNA binding events (read counts) in CSR-1 or WAGO targets. The dotted lines indicate the average number plus/minus one standard error. (C) Metagene trace shows the distribution of CSR-1 22G-RNAs mapped to the indicated CSR-1 subgroup genes, including 3' end 22G enriched (top) or not enriched (bottom) CSR-1 targets. The solid line indicates the average number of CSR-1 22G-RNAs read counts in the indicated CSR-1 subgroup genes. The dotted lines indicate the average number plus/minus one standard error. (D) Metagene trace shows the distribution of piRNA binding events in wild type (orange) or in CSR-1 depleted animals (green) mapped to the indicated CSR-1 subgroup genes, including 3' end 22G enriched (top) or not enriched (bottom) CSR-1 targets. The solid line indicates the average number of piRNA binding events (read counts) in the indicated CSR-1 subgroup genes. The dotted lines indicate the average number plus/minus one standard error. (E) Metagene traces show the distribution of WAGO-9 (HRDE-1) 22G-RNAs in wild type (orange) or in CSR-1 depleted animals (green) mapped to CSR-1 (top) or WAGO (bottom) targets. The miRNA mapped reads, which are expected to be non-specific RNA cloned from these experiments and thus remain unchanged between samples, are used for normalization. The solid line indicates the average number of WAGO-9 22G-RNA read counts in CSR-1 or WAGO targets. The dotted lines indicate the average number plus/minus one standard error. (F) Metagene traces show the distribution of WAGO-9 (HRDE-1) 22G-RNAs in wild type (orange) or in CSR-1 depleted animals (green) mapped to the indicated CSR-1 subgroup genes, including 3' end 22G enriched (top) or not enriched (bottom) CSR-1 targets. The miRNA mapped reads are used for normalization. The solid line indicates the average number of WAGO-9 22G-RNAs read counts in the indicated CSR-1 subgroup genes, including 3' end 22G enriched (top) or not enriched (bottom) CSR-1 targets. The dotted lines indicate the average number plus/minus one standard error. The miRNA mapped reads are used for normalization between samples. (G) The ratio of piRNA binding events or WAGO-9 22G-RNAs in *csr-1* depletion over control in high-density CSR-1 targeted windows and low-density CSR-1 targeted windows for germline mRNAs, CSR-1 targets and WAGO targets. The piRNA-tRNA hybrids and the miRNA reads are used for normalization of piRNA binding and WAGO-9 22G-RNAs, respectively. Blue lines represent median values. The statistical significance (P-value) of the difference in fold change (*csr-1* depletion over control) between two windows was calculated by the Mann-Whitney U test.

4.3.4 CSR-1 is not responsible for piRNAs' CDS-binding preference

We then investigated whether CSR-1 contributes to the piRNA CDS binding preference. If CSR-1 is responsible for the piRNA CDS-binding preference, we would expect this preference to diminish upon CSR-1 depletion. However, upon CSR-1 knockdown, piRNAs retain their CDS binding preference for both CSR-1 targets and WAGO targets (Supplementary Figure 4.9B). Specifically, the knockdown of CSR-1 did not change the pattern of piRNA binding sites around

start or stop codons for either CSR-1 targets or WAGO-1 targets (Supplementary Figure 4.10A). As translation has been shown to antagonize CSR-1 22G-RNA synthesis in CDS regions [144], we wondered whether translation may also play a role in regulating piRNA binding to the CDS. For both highly-translated and lowly-translated CSR-1 or WAGO-1 targets [1], we found that piRNAs retain their CDS binding preference **(**Supplementary Figure 4.10B**)**. However, since translation of germline mRNAs is a highly regulated process throughout germline development, future studies are needed to further investigate whether translation plays a role in regulating piRNA binding. Together, our analyses suggest that the piRNA CDS binding preference cannot be explained by CSR-1 licensing and indicates that an unknown mechanism promotes or restricts piRNA binding to the CDS.

4.5 Discussion

Small RNAs, such as miRNAs and piRNAs, regulate gene expression through base-pairing with their target mRNAs. It has been shown that miRNA targeting sites are enriched at 3' UTRs and that 3' UTR sites exhibit more gene regulatory effects than those at the CDS [9]. While piRNAs are known for their roles in genome defense, it was previously unknown whether piRNA targeting or regulatory potential operate differently among different regions of their mRNA targets. By analyzing transcriptome-wide *in vivo* piRNA binding sites, we uncovered a new feature regulating the piRNA defense system in *C. elegans*: *C. elegans* piRNAs preferentially bind and trigger gene silencing at the CDS of germline mRNAs. Consequently, secondary WAGO 22G-RNAs are preferentially produced at the CDS as well (Figure 4.6). Our finding has important implications on using synthetic piRNAs to silence target genes of interest, where piRNAs targeting the CDS are likely going to be more effective. Indeed, a recent publication has confirmed this trend, albeit not on a transcriptome-wide scale [121]. Nonetheless, it is worth noting that synthetic piRNAs

targeting the 3' UTRs of transgenes can trigger transgene silencing [8,87]. In addition, the CDS of *oma-1*, but not its UTRs has been reported to carry specific licensing properties that counter piRNA silencing [134]. Future studies will be necessary to further understand the underlying mechanisms that confer such differences in sensitivity to piRNA silencing. Furthermore, our study also provides insights into the applications in the expression of foreign nucleic acids in the germline of *C. elegans*. piRNAs are known to trigger gene silencing of transgenes that carry foreign nucleic acids, and removal of piRNA sites from the coding regions of these foreign nucleic acids are thus likely to be critical for the successful expression of these foreign nucleic acids in *C. elegans*. Indeed, previous studies have optimized the coding regions of foreign nucleic acids, including GFP, mCherry or Cas9, which successfully prevent these silencing-prone transgenes from piRNA-induced gene silencing [185].

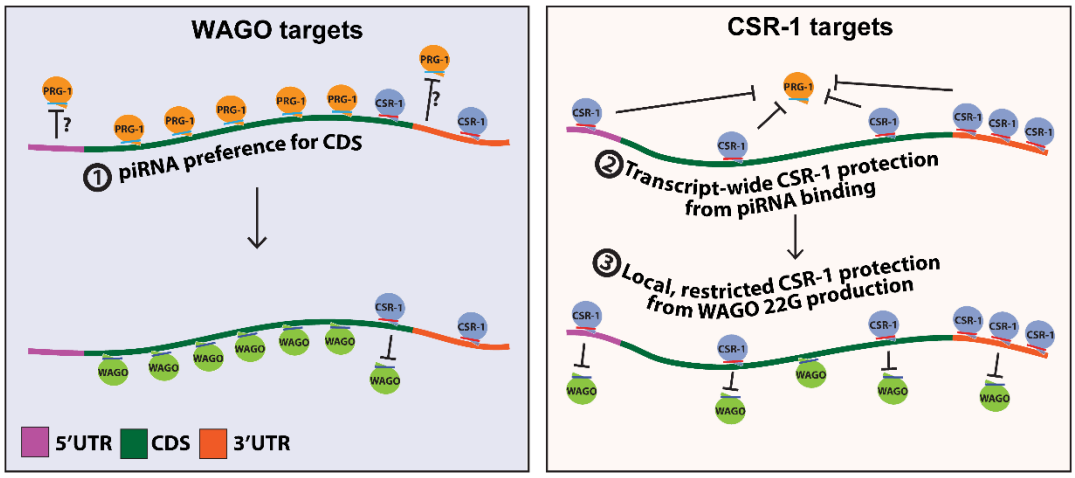


Figure 4.6: A model depicting three distinct mechanisms that control piRNA binding and gene silencing on WAGO and CSR-1 targets.
For WAGO targets (left), a CSR-1-independent mechanism promotes piRNAs binding at CDS (mechanism 1). WAGO 22G-RNA production, which is initiated by piRNA targeting, is therefore following a similar pattern of enrichment. For CSR-1 targets (right), CSR-1 blocks piRNA binding in a transcript-wide mode of action (mechanism 2). For downstream WAGO 22G-RNA production, CSR-1 acts locally to block WAGO 22G-RNA accumulation (mechanism 3).

In addition, our analyses show that while the reported piRNA targeting rules are effective in preventing germline transgene silencing by mutating potential targeting sites, they can be further

improved to identify functional piRNA targeting sites by considering other features of targeting events. For example, currently many CDS targeting events may have poor hybridization and are therefore dropped in the *in silico* prediction models. For these sites, lowering the stringency of hybridization may better recapitulate the *in vivo* target site landscape. In addition, target sites with high local CSR-1 levels may not be effective to trigger WAGO 22G-RNA production. Furthermore, future analyses are needed to examine and identify non-canonical piRNA binding rules.

Our work also provides insights into the mechanisms by which CSR-1 inhibits the piRNA pathway. Specifically, we found that CSR-1 depletion leads to increased piRNA binding across whole CSR-1 targeting transcripts, while increased secondary WAGO 22G-RNA synthesis preferentially occurs locally at CSR-1 targeting sites. In addition, our analyses showed that CSR-1 poorly targeted regions are just as effective in inhibiting piRNA binding compared to CSR-1 highly targeted regions. While we cannot fully rule out the possibility that low levels of CSR-1 also act locally to inhibit piRNA binding, our analyses showed that two anti-silencing functions of CSR-1 exhibit two distinct modes of action, where only the inhibition of WAGO 22G-RNAs but not piRNA binding are correlated with local CSR-1 levels. Taken together, our observations support a model where the mechanisms underlying these two CSR-1 anti-silencing processes are distinct. In this model, CSR-1 binding at various locations in a transcript can result in the protection of the entire transcript from piRNA recognition while CSR-1 targeted regions are locally protected against WAGO 22G-RNA production (Figure 4.6). Several previous experimental reports have provided evidence that supports our model. First, tethering of CSR-1 to a single locus at the 3' UTR of an endogenous mRNA can protect that mRNA from piRNA silencing [36]. Second, in transgenic worms carrying a silenced GFP::CDK-1 transgene, only foreign GFP sequences

produce high levels of WAGO 22G-RNA, while the CDK-1 sequence (which is identical to the endogenous CDK-1 gene and therefore targeted by CSR-1 22G-RNAs) produces low levels of WAGO 22G-RNAs, indicating a local inhibition of WAGO 22G-RNAs at the CSR-1 targeted region [133]. Nonetheless, understanding the specific mechanisms by which CSR-1 counteracts piRNA binding and WAGO 22G-RNA synthesis requires future investigation. We speculate that CSR-1 binding can somehow remove its bound mRNA from piRNA recognition subcellular hotspots, while CSR-1 inhibits WAGO 22G-RNAs synthesis locally, such as by competing for RNA-dependent RNA polymerases – resources known to be shared by these two Argonautes to produce their associated small RNAs [61].

Finally, our analyses also suggest that the CSR-1 pathway is not responsible for piRNAs' CDS binding preference, suggesting that an unknown mechanism promotes CDS surveillance by piRNAs. As the CDSs of foreign RNAs are expected to be under more selective pressure than UTRs, we speculate that the preferential recognition of CDSs by piRNAs likely offers a mechanism for the piRNA surveillance system to reduce the chance of evolution of foreign RNA variants that evade piRNA recognition.

4.6 Methods

4.5.1 Analyses of in vivo miRNA and piRNA binding sites

The iCLIP data of ALG-1 (SRR3882949) and the CLASH data of PRG-1 (SRR6512652/WT and SRR6512654/CSR-1 depletion) were used in these analyses [22,137]. To identify *in vivo* miRNA and piRNA binding sites, hybrids of miRNA/piRNA with their target mRNAs were identified by CLASH analyst [176] with default settings for data pre-processing and searching algorithms. To compare the amount of piRNA binding in wild type or CSR-1 depleted animals, the levels of piRNA/tRNA hybrid reads were used for normalization. The *C. elegans* miRNA/piRNA sequences

and the mRNA sequences (n=43040) of Wormbase version WS275 were used. Once hybrid reads are identified, miRNA and piRNA binding sites are defined by the regions of interacting mRNAs with lowest binding energy or highest binding score to corresponding miRNA using RNAup [92] or to corresponding piRNA using pirScan [178], respectively. When the mRNA interacting sequences (CLASH identified regions) are shorter than miRNA or piRNA, they are first extended to the size of miRNA and piRNAs using both the upstream and downstream sequences before they are examined for sites with best pairing energy/score.

Germline mRNAs (n=27792) are defined as mRNAs that are detected either in the spermatogenic or oogenetic transcriptome [111]. WAGO targets (n=3644) are defined as genes whose mapped 22G-RNAs exhibit over two-fold enrichment from either WAGO-1 IP than that from input 22G-RNAs [61] or WAGO-9 IP than that from input 22G-RNAs [141]. CSR-1 targets (n=15821) are defined as transcripts whose mapped 22G-RNAs exhibit over two-fold enrichment from CSR-1 IP than that from input 22G-RNAs [33]. To compare the WAGO-9 IP levels of wild type or CSR-1 depleted animals, the total miRNA mapped reads were used for normalization. The list of germline-expressed miRNAs are obtained from the previous report [105] for prediction and identification of the mRNA targets of germline miRNAs.

4.5.2 Prediction of piRNA targeting sites

Predicted piRNA targeting sites (with stringent targeting rules) on *C. elegans* germline mRNAs were obtained from piRTarbase [177]. For stringent rules, the following criteria are used; at seed region, no non-GU mismatches and no more than 2 GU mismatches are allowed. At non-seed region, no more than 2 non-GU mismatches and no more than 3 GU mismatches are allowed. Furthermore, no more than 6 total mismatches are allowed from both seed and non-seed regions [178].

4.5.3 Measurements of miRNA, piRNA binding sites and 22G-RNA levels at different regions of mRNAs

The *C. elegans* transcriptome data (WS275) annotation was used to define the location of 5' UTRs, CDSs, and 3' UTRs. The number of sites or read counts from each region were then divided by the nucleotide length of the corresponding regions for each mRNA to obtain the density. Where sites were spanning between two distinct regions, the sites or read counts were split according to the portion that mapped to each region. For measurements of sites/reads around start and stop codons, the 200 nt (+/- 100 nt) window centered at start or stop codon of each transcript was aligned and the number of the sites/reads per transcript mapped at each nucleotide was calculated. As transcripts have different 5' and 3'UTR lengths, the mapped sites/read counts at each position were divided by the number of transcripts that contain such position to obtain the sites/read counts per transcript. The following small RNA data were used to analyze the distribution and levels of CSR-1 and WAGO 22G-RNAs; CSR-1 associated small RNAs data (SRR12318132), WAGO-1 associated small RNAs (SRR8482951/WT, SRR8482949/*prg-1* mutant) [14], and WAGO-9 (HRDE-1) associated small RNAs (SRR12318140/WT, SRR12318144/CSR-1 depletion) [144].

We defined the highly or poorly CSR-1 targeted regions by identifying 100 nucleotide windows in each germline mRNA transcript which contain the most or the least (non-zero) CSR-1 associated small RNAs (SRR12318132). We also required that these highly or poorly CSR-1 targeted regions are not overlapping with each other and exhibit at least a two-fold difference in levels of CSR-1 associated small RNAs. The list of CSR-1 regions used in these analyses are provided in Supplementary Data file 4.2.

4.5.4 Metagene analyses

A custom script was used to divide mRNA transcripts into 100 bins and the normalized reads within each bin was calculated. The average number of sites or read counts at each bin per mRNA transcript was then calculated.

4.5.5 mRNA-seq analyses

The mRNA-seq data from wild type *C. elegans* young adult samples (SRR6512642 and SRR6512643) (Shen et al. 2018) were used to define RPKM-standardized piRNA binding densities. Reads were mapped to the WS275 transcriptome with bowtie2. Normalized reads were calculated for each library by summing sense-mapping reads to each transcript using bedtools, standardizing by mRNA length to calculate reads per kilobase million (RPKM), then averaging the two libraries together.

4.5.6 Ribo-seq analyses

The ribo-seq data from the wild type *C. elegans* late L4/young adult sample (36 hours post hatching, SRR3356499) [1] were used to identify the number of ribosome protected fragments (RPF) mapped to each CSR-1 target or WAGO-1 target. The transcripts with top 10% or bottom 10% amounts of RPF were defined as highly-translated and poorly-translated transcripts, respectively.

4.5.7 Data availability

All sequencing data analyzed in the manuscript are available at NCBI GEO or ENA database. The SRR numbers of sequencing data used in specific analyses are provided in the Materials and Methods section above. Scripts to analyze reads within 5’ UTRs, CDSs, and 3’ UTRs, to analyze read accumulation proximal to start and stop codons, and to generate metagene distributions are available at https://github.com/RyanCCJ/RDT.

4.7 Acknowledgements

This work is supported in part by NIH predoctoral training grant T32 GM07197 to J.B.; the Ministry of Science of Technology of Taiwan (MOST 108-2628-E-006-004-MY3, MOST 110-2221-E-006-198-MY3 and MOST-111-2221-E-006-151-MY3 grants to W.-S.W., the NIH grant R01-GM132457 to H.-C.L.

4.8 Supplementary Information

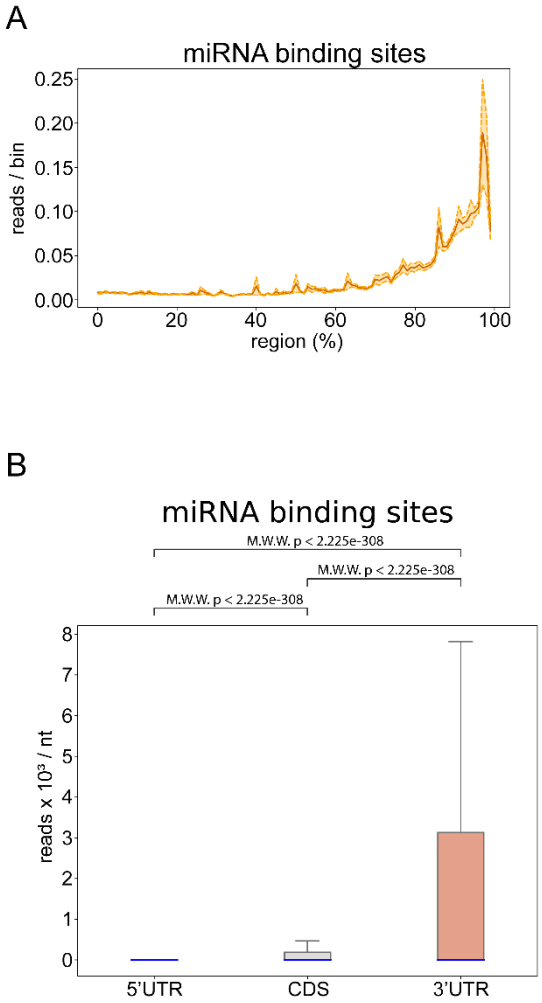


Figure 4.7: *In vivo* miRNA binding sites are enriched at the 3' UTR of mRNAs in *C. elegans,* related to Figure 4.2.
(A) Metagene trace shows the distribution of miRNA binding events along an mRNA. The solid line indicates the average number of miRNA binding events (read counts) in all mRNAs. The dotted lines indicate the average number plus/minus one standard error. (B) The density of miRNA binding events in the indicated regions. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test.

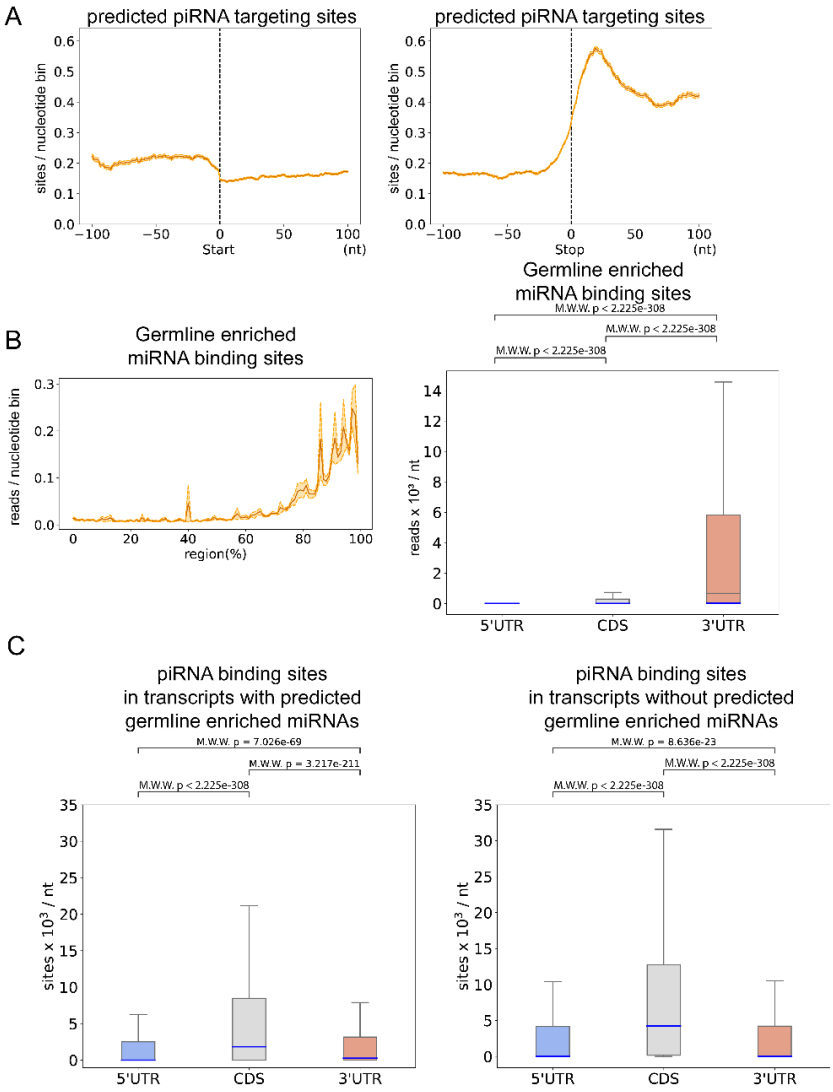


Figure 4.8: piRNAs' CDS binding preference cannot be explained by the predicted piRNA targeting sites, related to Figure 4.3.
(A) The distribution of predicted piRNA targeting sites around the start codon (left) or stop codon (right). A 200 nucleotide-window centered at start or stop codon is shown. The solid line indicates the average number of predicted piRNA targeting sites in all germline mRNAs. For each position, the average is calculated from all germline mRNAs which possess that position. (B) Metagene trace shows the distribution of germline-expressed miRNA binding sites along a germline mRNA (left). The solid line indicates the average number of miRNA binding events (read counts) in germline mRNAs. The dotted lines indicate the average number plus/minus one standard error. The density of experimentally identified germline-expressed miRNA binding sites in the indicated regions of germline mRNAs (right). Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (C) The density of experimentally identified piRNA binding sites in the indicated regions of germline mRNAs that have predicted miRNA binding sites for germline-expressed miRNAs in 3' UTRs (left) or do not (right). Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test.

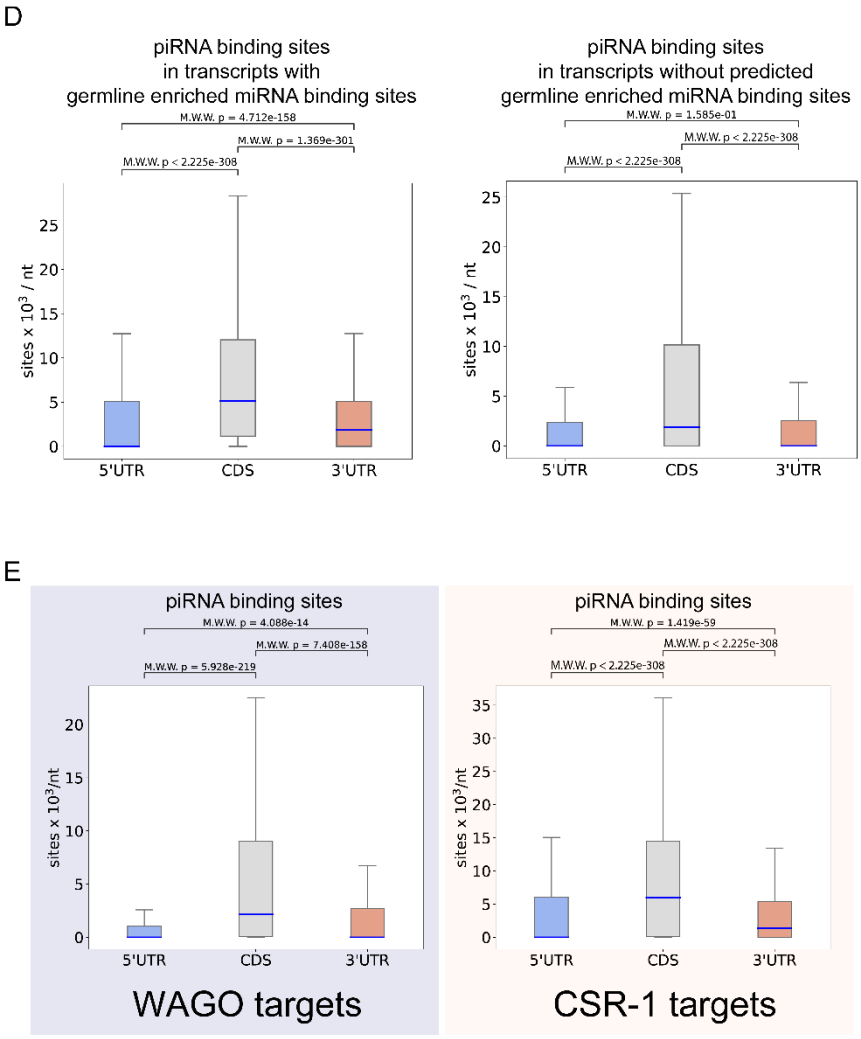


Figure 4.8 (continued): (D) The density of experimentally identified piRNA binding sites in the indicated regions of germline mRNAs that have experimentally identified miRNA binding sites for germline-expressed miRNAs in 3' UTRs (left) or do not (right). Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (E) The density of piRNA binding sites in the indicated regions of germline-silenced mRNAs (WAGO targets, left) and germline-expressed mRNAs (CSR-1 targets, right). Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test.

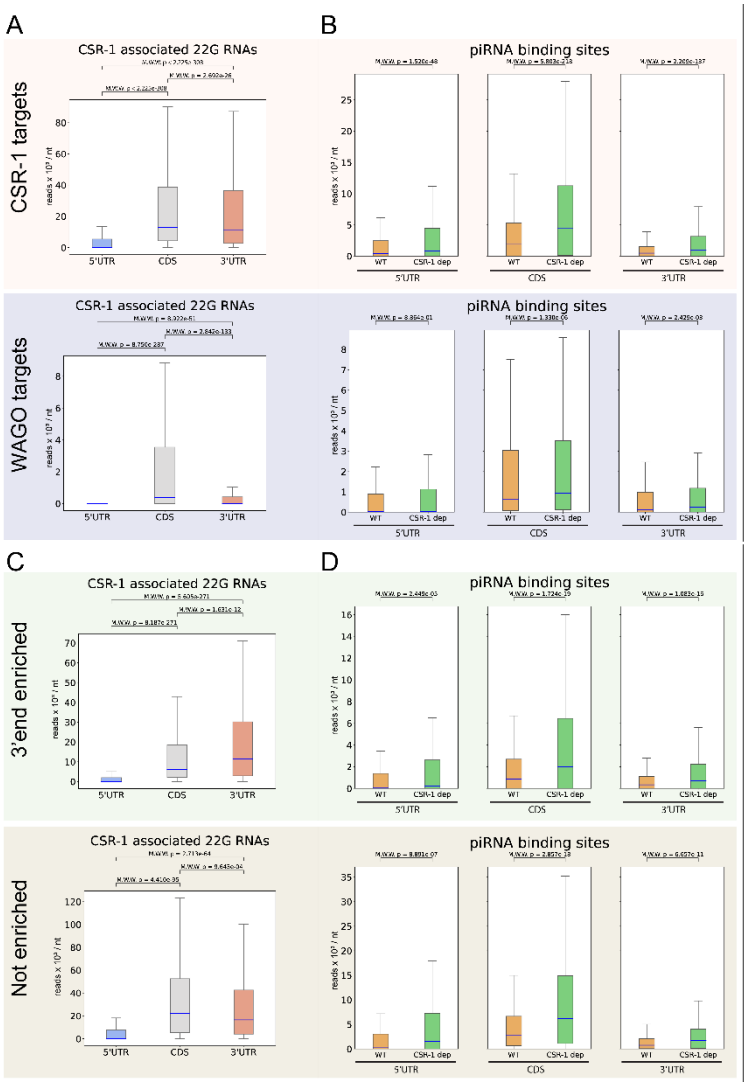


Figure 4.9: The distribution of CSR-1 22G-RNAs, WAGO 22G-RNAs, and piRNA binding sites, related to Figure 4.5.

(A) The density of CSR-1 22G-RNAs in the indicated regions of CSR-1 (top) or WAGO (bottom) targets. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (B) The density of piRNA binding events in the indicated regions in wild type (orange) or in CSR-1 depleted animals (green) mapped to CSR-1 (top) or WAGO (bottom) targets. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (C) The density of CSR-1 22G-RNAs mapped to the indicated regions of CSR-1 subgroup genes, including 3' end 22G-RNA enriched (top) or not enriched (bottom) CSR-1 genes. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (D) The density of piRNA binding events mapped to the indicated regions in wild type (orange) or in CSR-1 depleted animals (green) of CSR-1 subgroup genes, including 3' end 22G-RNA enriched (top) or not enriched (bottom) CSR-1 genes. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test.

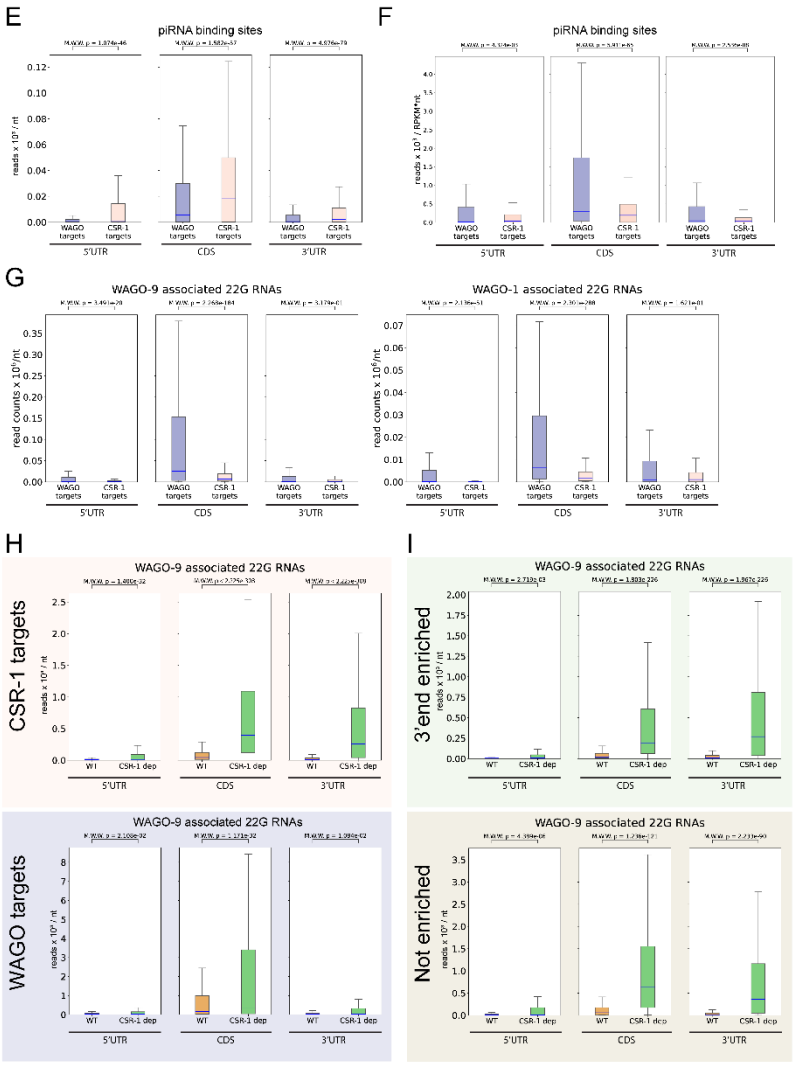


Figure 4.9 (continued): (E) The density of piRNA binding events (read counts) in the indicated regions of WAGO targets or CSR-1 targets. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (F) The density of piRNA binding events (read counts) standardized by normalized mRNA reads (RPKM) in the indicated regions of WAGO targets or CSR-1 targets. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (G) The density of WAGO-9 (left) or WAGO-1 (right) 22G-RNAs mapped to the regions of WAGO targets or CSR-1 targets. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (H) The density of WAGO-9 (HRDE-1) 22G-RNAs mapped to the indicated regions in wild type or in CSR-1 depleted animals of CSR-1 (top) or WAGO (bottom) targets. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test. (I) The density of WAGO-9 (HRDE-1) 22G-RNAs in wild type or in CSR-1 depleted animals mapped to the indicated regions of CSR-1 subgroup genes, including 3' end 22G-RNA enriched (top) or not enriched (bottom) CSR-1 genes. Blue lines represent median values. The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test.

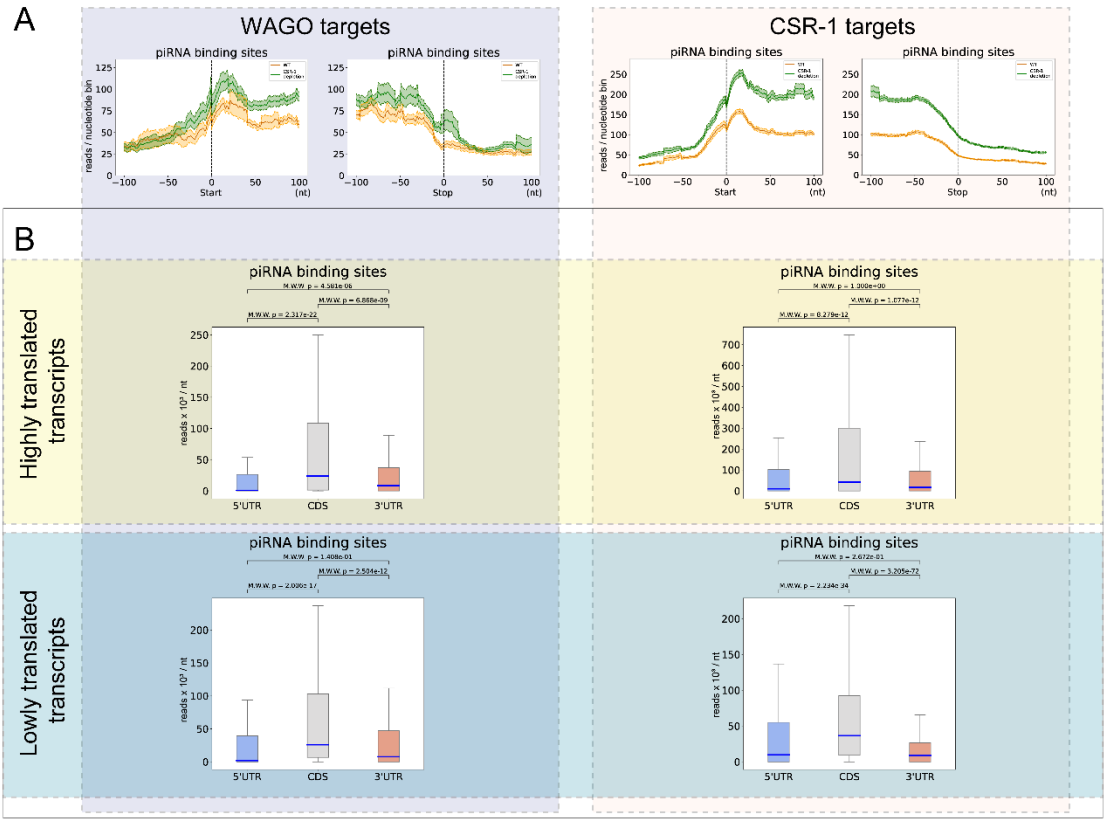


Figure 4.10: The distribution of piRNA binding sites, related to Figure 4.5.
(A) The distribution of piRNA binding events around the start codon (left) or stop codon (right) in wild type or in CSR-1 depleted animals. A 200 nucleotide-window centered at start or stop codon is shown. The solid line indicates the average number of WAGO-9 22G-RNAs read counts in WAGO-1 targets (left) or CSR-1 targets (right). (B) The density of piRNA binding events (read counts) in the indicated regions of highly-translated or poorly-translated WAGO targets (left) and highly-translated or poorly-translated CSR-1 targets (right). The statistical significance (P-value) of the difference in density between two regions was calculated by the Mann-Whitney U test.

Data file 4.1: piRNA binding site information for Figures 4.1B and 4.2A.

Data file 4.2: Coordinates used for CSR-1 occupancy windows.

CHAPTER 5

TOOL DEVELOPMENT FOR *C. ELEGANS* SMALL RNA RESEARCHERS AND THE BROADER RNA COMMUNITY

5.1 Attributions

This chapter has been adapted from three independently published works:

(1) W.-S. Wu, *et al.*, pirScan: a webserver to predict piRNA targeting sites and to avoid transgene silencing in C. elegans, Nucleic Acids Res. 46 (2018) W43–W48.

(2) W.S. Wu & J.S. Brown, *et al.*, PiRTarBase: A database of piRNA targeting sites and their roles in gene regulation, Nucleic Acids Res. 47 (2019) D181–D187.

(3) W.-S. Wu & J.S. Brown, *et al.*, CLASH Analyst: A Web Server to Identify In Vivo RNA–RNA Interactions from CLASH Data, Non-Coding RNA. 8 (2022) 6.

Wei-Sheng Wu and members of his lab (Tsung-Te Chen, Yu-Han Chu, Wei-Che Huang, Hao Chen, Pin-Hao Chen, Sheng-Cian Shiue, and Dong-En Lee) adapted analyses that were performed jointly with the author for the web.

5.2 Introduction

With the advent of next-generation sequencing technology, new datasets with increasingly complex and rigorous tools to analyze them have dominated the scientific landscape. In the *C. elegans* community, researchers that have relied on traditional molecular techniques to ask questions about RNA regulation are increasingly left behind due to the inaccessibility of this new and abundant source of data. We have adapted several applications involving complex analytical tasks to an easy-to-use graphical environment, enabling researchers from diverse backgrounds to access data directly that would otherwise require specialized knowledge of advanced bioinformatics to reproduce. Our goals in developing these tools are (1) to allow a wide array of

small RNA and *C. elegans* researchers access to techniques that are becoming ubiquitous in the field of RNA biology, and (2) to make our own work more accessible and reproducible by presenting the complex techniques used in a more user-friendly and easier to follow context.

5.3 Results

5.2.3 pirScan: a webserver to predict piRNA targeting sites and to avoid transgene silencing in C. elegans

5.2.3.1 Motivation and purpose

PIWI-interacting RNAs (piRNAs) are a class of small noncoding RNAs that guard animal genomes against mutation by silencing transposons [171]. piRNAs guide the Argonaute proteins of the PIWI subclass to target mRNAs for silencing [6,53,58,85]. The majority of the approximately 15,000 piRNAs encoded by the *C. elegans* genome, however, do not match transposon sequences [15,130]. It has been proposed accordingly that piRNAs target additional mRNAs with diverse functions. The rules governing piRNA targeting have proven enigmatic [57,155]. It has been shown, however, that piRNAs appear to tolerate at least three mismatches between their sequences and likely prefer pairing at a seed region [10,55,87,126,186]. Our group has further refined the piRNA targeting rules in *C. elegans* by examining how a single piRNA recognizes its target, and by analyzing piRNA reporter assays [185]. These analyses have confirmed that piRNAs require near perfect matching within the 2nd to 7th nucleotide seed-region, much like microRNAs [13]. Unlike microRNAs, however, piRNAs can only tolerate a few mismatches outside of the seed region. Lastly, the first nucleotide does not contribute to the piRNA targeting. Therefore, although the piRNA targeting rules share similarities with the miRNA targeting rules, the numerous tools available for miRNA target prediction are insufficient to successfully predict piRNA targeting sites [37,66,78].

Significantly, the elucidation of the piRNA targeting rules can help resolve a decades-long impediment to *C. elegans* researchers: transgene silencing in the germline. Transgenes carrying GFP or mCherry tags are frequently silenced in *C. elegans* germlines [72,103]. Additionally, this transgene silencing phenomenon is dependent on the PIWI Argonaute PRG-1 [141]. By introducing silent mutations in silencing-prone transgenes, such as GFP, mCherry and Cas9, at sites predicted to be targeted by endogenous piRNAs, we have successfully observed strong germline expression of these transgenes in wildtype animals [185]. In total, the modified transgenes are expressed in all 14 independently isolated transgenic worm lines, while the unmodified transgenes are silenced in all 15 worm lines. Not only do these results validate that our piRNA targeting rules are correct, but they also provide an avenue to avoid the persistent silencing of transgenes in *C. elegans* germlines [185].

To make our piRNA targeting rules available to the worm community, we have developed pirScan (http://cosbi4.ee.ncku.edu.tw/pirScan/), which predicts the endogenous piRNA targeting sites in an input sequence using our established targeting rules. pirScan intuitively displays where piRNA targeting sites are located within an input sequence, as well as the pairing information at each piRNA targeting site. The results of each piRNA target prediction can subsequently be downloaded. pirScan then provides suggestions of how to break piRNA targeting rules and therefore render a given sequence more likely to be expressed. Once the suggested silent mutations are chosen by the user using the graphical interface, the revised sequence is re-scanned to verify that it is now free of piRNA-targeting sites. The final modified sequences can then be downloaded.

5.2.3.2 General Framework

pirScan is a tool that predicts the presence of piRNA targeting sites within a given sequence in *C. elegans* and suggests silent mutations that can be introduced to avoid silencing of transgenes by

piRNAs (Figure 5.1). The user can input a mature mRNA or spliced DNA sequence and choose to use default or customized piRNA targeting rules to search for targeting sites. The results are provided with a concise graphical and tabular representation displaying all predicted piRNA targeting sites within the input sequence. Each of these views can be downloaded in Portable Network Graphic (.PNG) and comma separated value (.csv) format, respectively. The user can then choose between suggested silent mutations that will allow the input sequence to escape predicted piRNA targeting, but still retain the same amino acid sequence. If more than one piRNA targeting site were present in the input sequence, then each piRNA will have its own separate table complete with the possible silent mutations to introduce. Upon the user's completion of the proposed mutagenesis, the modified sequence are immediately re-submitted for scanning to verify that the sequence will successfully escape piRNA targeting and ensure that new targeting sites were not inadvertently introduced. Once the user is satisfied with the adjusted input sequence, the modified version can be downloaded as a text file (.txt).

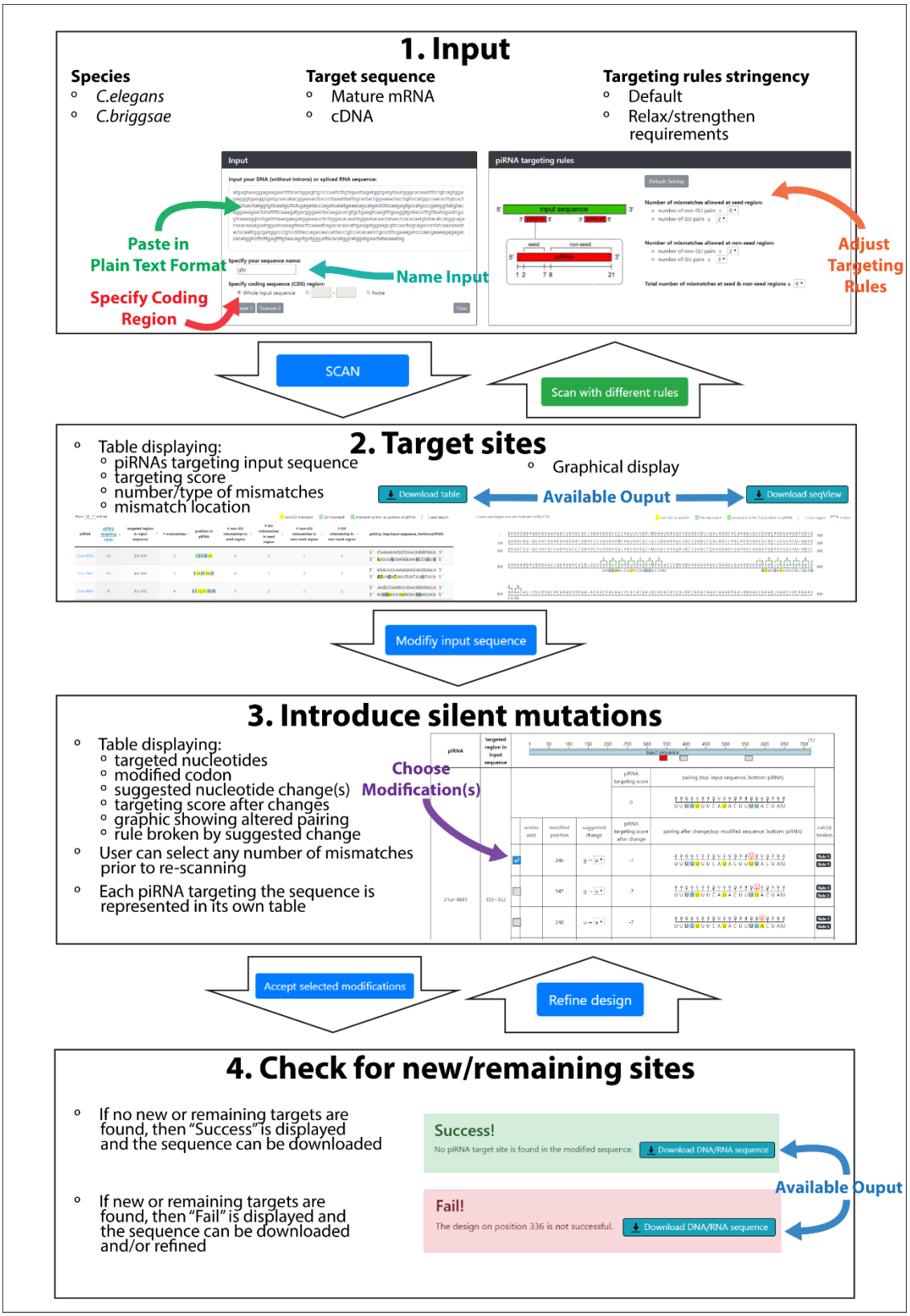


Figure 5.1: An overview of pirScan's workflow.

5.2.3.3 Input

pirScan will accept an input sequence in plain text format (Figure 5.1, 1. Input). The user must also specify an input name, which will be included in the output following modifications. The input can be a mature mRNA sequence or spliced DNA sequence. By default, pirScan will assume that the first nucleotide present is the start of the open reading frame (see Example 1 in the input window on pirScan). If the first nucleotide present is not the start of the open reading frame, users can also specify the numerical positions of the open reading frame to provide coding information of the input sequence by changing the selection in the input window (see Example 2 in the input window on pirScan). As piRNAs can also target 5' and 3' untranslated regions (UTRs), we suggest the user to include the sequence of 5' and/or 3' UTRs and simply specify the coding region. The input sequence should not contain introns. 'None' can be selected as the coding sequence region if the input is a noncoding RNA/DNA.

5.2.3.4 Target Stringency

pirScan will identify piRNA sites in *C. elegans* using the default target stringency setting if not specified (Figure 5.1**,** 1. Input). The default setting will predict the confident piRNA targeting sites according to our reporter assay which allows 0 non-GU mismatches in seed, less than or equal to 2 GU pairs in seed, less than or equal to 2 non-GU mismatches in non-seed, less than or equal to 3 GU pairs in non-seed, and less than or equal to 6 mismatches and GU pairs in seed and non-seed regions combined [185]. To predict more piRNA targeting sites, we have used a slightly relaxed rule (0 non-GU pairs in seed, less than or equal to 2 GU pairs in seed, less than or equal to 3 non-GU pairs in non-seed, less than or equal to 4 GU pairs in non-seed, and less than or equal to 6 mismatches in seed and non-seed regions combined). Both stringencies have enabled our group to obtain transgenic strains that stably express three silencing-prone transgenes in the germline,

including GFP::CDK-1, mCherry::ANI-1, and Cas9 [157,168,185]. The stringency of the targeting rules can be adjusted by the user prior to the first submission. If an adjustment to the target stringency is made, then those revised targeting rules are automatically used for rescanning of a modified sequence. If the chosen stringency results in over 100 predicted targeting sites, then the search will abort and ask the user to either strengthen the stringency or shorten in input sequence.

5.2.3.5 Targeting Site Output

Upon submission of an input sequence, the user can download a table in .csv format that contains the list of predicted piRNA targeting sites in the input sequence, the targeting score, the location of targeting sites, the numbers of mismatches, and some aspects about those mismatches (Figure 5.1, 2. Target sites). The targeting score is based on the results of our reporter assay [185] and is calculated using a penalty system where a perfect targeting site is assigned a score of 10. Each non-GU mismatch in the seed region, GU pair in the seed region, non-GU mismatch in the non-seed region, and GU pair in the non-seed region incurs a -7, -1.5, -2, and -1.5 penalty, respectively. We recommend changing any site predicted by the default targeting rules which results in targeting scores of 0 and above, as these are confident sites that are more likely to be critical for silencing. In fact, we have successfully expressed a silencing-prone *gfp::cdk-1* transgene by modifying all four confident sites predicted with default targeting rules and one additional site found in the junction between the *gfp* and *cdk* sequence (Figure 5.2 and Table 5.1). This suggests that, in contrast to the more relaxed rules used in our previous study [185], it is possible to selectively mutate only a few sites and still observe transgene expression. Additionally, a visual representation of the predicted piRNA targeting to the input sequence is available for downloading in .PNG format. The pairing between the 21 nucleotide long piRNA and the corresponding position in the input sequence is displayed for clarity, with the target sites of the input sequence listed according

to targeting score. Either of these outputs can allow a user to customize his/her input sequence outside of our pirScan framework if he/she chooses.

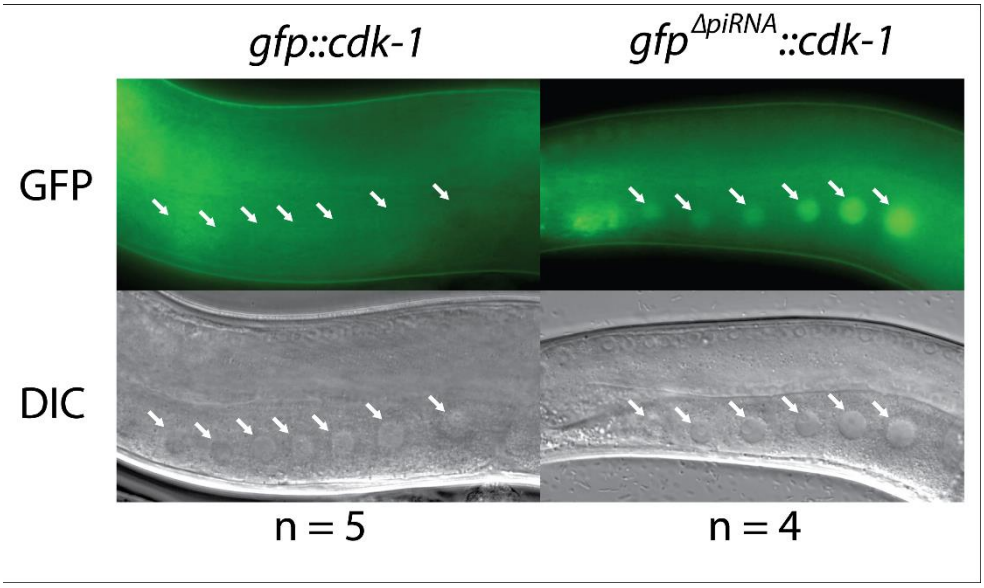


Figure 5.2: Avoiding only five piRNAs predicted to target GFP results in germline expression of a *gfp::cdk-1* transgene.
gfp::cdk-1 is not expressed prior to introduction of silent mutations, but *gfp*$^{\Delta piRNA}$*::cdk-1* is expressed when five predicted piRNA target sites are modified to break the targeting rules. Nuclei are indicated by white arrows.

piRNA	Targeted region in spliced GFP	Mutation(s)	Targeting score before/after
21ur-9758	241-261	G245A, T255C	6 / -3
21ur-9756	556-576	T570A, C573A	3.5 / -9
21ur-7942	390-410	A408C	1.5 / -5.5
21ur-4841	338-358	T354A, T357C	0 / -14
21ur-1953*	703-723	G722C	2.5 / -4.5

Table 5.1: List of mutations in *gfp::cdk-1* to avoid targeting by 5 piRNAs in *gfp*$^{\Delta piRNA}$*::cdk-1*. *21ur-1953 is present in the linker region between *gfp* and *cdk-1*.

5.2.3.6 Modification of Input Sequences

At the end of the output page of piRNA target sites, a link is offered for users to introduce silent mutations in their input sequences to avoid piRNA recognition (Figure 5.1, 3. Introduce silent mutations). The system displays piRNA targeting sites by their targeting score from highest to lowest and lists up to 6 suggested changes at each site. Each piRNA will have its own separate table complete with the possible silent mutations to introduce. The user can select/deselect any modifications or select multiple modifications at each piRNA site. The system also indicates which rule(s) each mutation will break. Note that sometimes multiple mutations are needed for avoiding piRNA recognition. Importantly, pirScan will not suggest mutations that will result in the incorporation of a rarely used codon, defined as a codon present in less than 10% of the occurrences of a given amino acid [113]. Once the user chooses silent mutations to introduce into the input sequence, the program will automatically rescan any sites with modified sequence(s) and display a table summarizing the changes (Figure 5.1, 4. Check for new/remaining sites). pirScan will display "Success" if no piRNA targeting sites are found in the modified sequence, or "Fail" if additional piRNA targeting sites are found. The modified sequences (either success or failure) can be downloaded in .txt format. In the case of failure, the user can choose to refine the design and the program will first display the piRNA sites that remain to allow further modification. We suggest the user to choose additional sites if a single mutation is not sufficient to avoid piRNA recognition, or to choose different sites if a new piRNA targeting site is found. Additionally, a visual representation which highlights the chosen substitutions can be downloaded in .PNG format.

5.2.3.7 Potential Applications

Our goal in developing pirScan is to make our discoveries concerning the rules of piRNA targeting available to the scientific community and to provide researchers with a method of avoiding the germline silencing of transgenes in *C. elegans*. pirScan can be employed for similar purposes, as well as for more disparate analyses. Examples of tangential goals are outlined below.

We have described our success in expressing GFP, mCherry and Cas9 in C. elegans germline [185]. pirScan will be useful to recode other foreign sequences, such as RNA/DNA binding proteins, enzymes or additional fluorescent markers. A recent study has reported an alternative approach by adding periodic A_n/T_n clusters (PATCs) to the promoter and/or introns of the transgenic [50]. Although the function of PATCs in gene expression is not understood, this strategy has also been employed by our group to stimulate gene expression in the germline [50,185]. It remains to be determined whether transgene expression can be further promoted by combining both strategies.

Interpreting the lack of expression from a previously published or previously used transgenic construct can be an important indicator of the merit in a particular hypothesis or research question. For example, when miRNA promoters are used to drive GFP expression, GFP is detectable in most tissue types with the exception of the germline [98]. This result suggests that either miRNAs are not expressed in the C. elegans germline, or that the transgenes used in this study are prone to germline silencing. pirScan can be applied to these constructs to suggest the best interpretation of this or a similar negative result.

pirScan is designed so that the input sequence can ultimately be modified to escape silencing of the indicated endogenous piRNAs. In some cases, the user may be unable to introduce even silent mutations in his/her transgenic sequence. pirScan can be used to identify which piRNAs

are responsible for silencing a particular construct, and the user can apply CRISPR/Cas9 or a similar method to delete or modify the indicated piRNA sequence in the C. elegans genome.

pirScan accepts sequences of any origin as input. Therefore, the piRNA targeting site prediction is not limited to exogenously constructed genes. If a user hypothesizes that the lack of germline expression or germline-specific regulation of an endogenous gene could be a result of piRNA targeting, then he/she can simply submit the endogenous sequence and look for piRNA targeting sites.

5.2.3.8 Conclusions

The discovery of small noncoding RNAs has changed how we understand the factors contributing to gene expression. In a similar vain to the numerous miRNA target prediction tools that are available to predict miRNA binding, we have developed a tool to predict piRNA targeting sites using the piRNA targeting rules. This piRNA target prediction tool is the first of its kind for *C. elegans* and we believe it will be highly valuable in assisting the *C. elegans* community to solve the decades old dilemma of germline-specific transgene silencing. Our piRNA targeting rules have only been experimentally validated in *C. elegans*; however, we will expand pirScan once the piRNA rules are identified in other animal systems. In addition, we can update our tool to incorporate any additional factors that influence targeting. For example, the influence of piRNA abundance on silencing efficiency is not yet understood, but our tool can be easily updated in the future to consider expression level if this is established as a deterministic factor. As the piRNA targeting rules are likely to be operant in *C.briggsae* as well, we have also provided the option of using the piRNA data of *C. briggsae* by simply choosing *C.briggsae* in the input page [139]. As

PIWI proteins are highly conserved among animals, pirScan offers an exciting platform of piRNA target prediction that can be expanded to various organisms in the future.

5.2.3.9 Acknowledgements

This work is supported by Ministry of Science of Technology of Taiwan grants (MOST-105-2221-E-006-203- MY2 and MOST-106-2628-E-006–006-MY2) to W.-S.W., a NIH predoctoral training grant (T32 GM07197) to J.B., a NIH P01 grant (HD078253) to Z.W., and a NIH R00 grant (GM108866) to H.-C.L.

5.3.4 piRTarBase: a database of piRNA targeting sites and their roles in gene regulation

5.3.4.1 Motivation and purpose

Animals have evolved a robust mechanism to silence transposons and thereby guard their genomes against mutation using small noncoding RNAs called PIWI interacting RNAs (piRNAs) [171]. piRNAs guide the PIWI subclass of Argonaute protein to search for mRNA targets and regulate their expression [6,53,58,85]. Surprisingly, tens of thousands of piRNAs produced by animals, from nematodes to mice, do not match transposon mRNA sequences [7,15,130]. Recently, piRNAs have been shown to regulate the expression of endogenous mRNAs. For example, it has been reported in mice that piRNAs produced during the pachytene stage of meiosis initiate the destruction of thousands of mRNAs to prepare spermatocytes for sperm production [57]. In flies, piRNAs regulate the localization and degradation of mRNAs in the early embryo [127,167]. Additionally, recent studies have also reported the function of piRNAs in the regulation of sex determination, axonal regeneration, and nictation in silkworms and nematodes [75,76,86]. However, predicting transcriptome-wide piRNA targeting sites has been difficult due to limited knowledge of piRNA targeting.

Our group has recently determined the targeting rules that piRNAs use to find mRNA targets in *C. elegans* [185]. We found that piRNAs, like microRNAs [13], possess a 2nd to 7th nucleotide seed sequence that requires nearly perfect complementarity between the piRNA and its target. Though unlike most miRNA targeting events, piRNAs can only tolerate a few mismatches outside of the seed sequence. We have used the piRNA targeting rules to successfully express otherwise frequently silenced transgenes in the *C. elegans* germline by avoiding piRNA targeting and constructed a tool, pirScan, that allows other researchers to do the same [178,185]. Additionally, the cross-linking, ligation, and sequencing of hybrids (CLASH) approach has recently been applied to the PIWI Argonaute in *C. elegans*, which has revealed endogenous piRNA-mRNA interactions *in vivo* [137]. Both of these methods predict that many endogenous mRNAs can be targeted by piRNAs. However, both methods have been shown to identify many miRNA target sites that do not repress the expression of their targets, suggesting additional approaches are needed to reveal functional piRNA sites [2].

Our purpose in constructing piRTarBase (http://cosbi6.ee.ncku.edu.tw/piRTarBase/) is to synthesize the recent developments in piRNA targeting site identification from predictions and experiments into a user-friendly interface that will allow researchers to explore the piRNA targeting sites and their regulatory effects on endogenous genes (Figure 5.3). piRTarBase displays various information about these piRNA sites [137], allowing users to search for and browse desired endogenous genes for piRNA targeting sites, as well as to identify the predicted mRNA targets of desired piRNAs. Additionally, piRTarBase contains expression data from wildtype and PIWI mutant animals, allowing users to assess the effects that piRNA targeting has on a given gene. Importantly, our analysis suggests that piRNA targeting sites that are both predicted by our targeting rules and identified by CLASH, referred to as common sites, are significantly better at

predicting mRNAs that are regulated by PIWI *in vivo* than either method alone. While numerous miRNA targeting site databases are available, there is currently only one database for piRNA targeting sites, and it only inventories piRNA sequences and the limited published targets [187]. By contrast, piRTarBase integrates the results derived from the piRNA targeting rules, PIWI CLASH data, and expression data of mRNAs and small RNAs, allowing researchers of different fields to identify candidates of functional piRNA targeting sites. Database entries that pertain to piRNA target sites were collected from transcriptome-wide prediction of piRNA target sites using pirScan [178] and published PIWI PRG-1 CLASH data [137]. Database entries corresponding to mRNA and small RNA expression profiles in WT and PIWI mutant animals were collected from published data [15,54,87,102,137]. We will continue to update piRTarBase with new datasets as they become available.

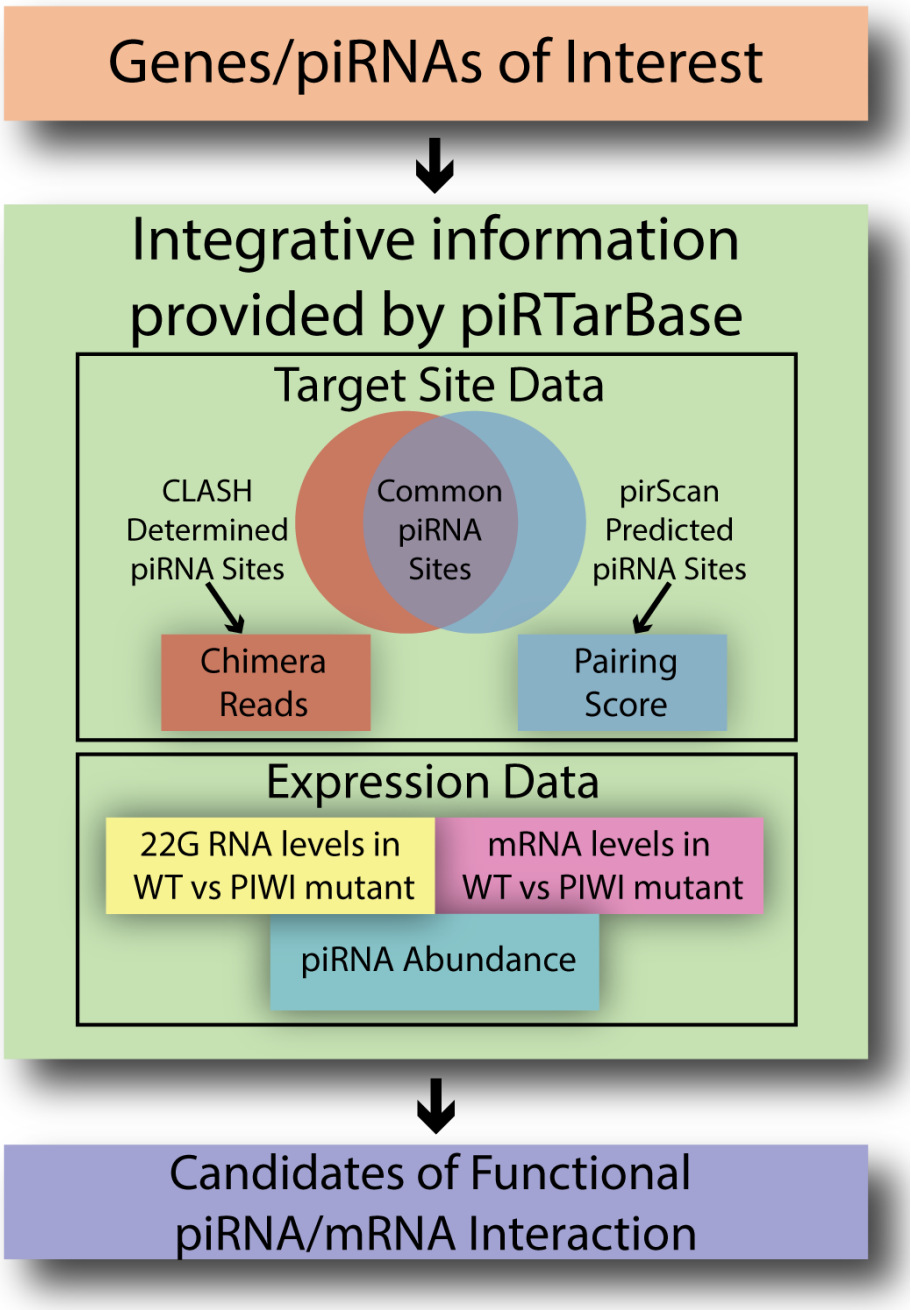


Figure 5.3: Workflow of piRTarBase.

5.3.4.2 Compendium of pirScan predicted piRNA targeting sites

pirScan predicts transcriptome-wide piRNA targeting sites of 17849 *C. elegans* piRNAs (15364 type 1 and 2485 type 2 piRNAs) based on stringent and relaxed rules, resulting in 571204 and 1420256 piRNA targeting sites, respectively [178]. The stringent rules are based on an *in vivo*

reporter assay in *C. elegans* that reveals which sites, when mutated, are sufficient to avoid silencing [185]. Therefore, predicted sites based on stringent rules are more likely to contribute to gene silencing. The piRNA targeting score is based on the data obtained in the same reporter assay [178]. Alternatively, the relaxed rules are based on genome-wide monitoring of secondary siRNA (22G RNA) expression changes surrounding presumptive piRNA targeting sites [185]. In *C. elegans*, when PIWI-piRNA complexes recognize a target gene, they induce the production of 22G RNAs by RNA-dependent RNA polymerases around the targeted site [10,87]. The piRNA rules derived from this approach are slightly less stringent than those derived from the piRNA reporter assay and are therefore defined as the relaxed rules.

5.3.4.3 Compendium of CLASH predicted piRNA targeting sites

The cross-linking, ligation, and sequencing of hybrids (CLASH) approach is a powerful method that can be used to identify RNA-RNA interactions *in vivo* [65]. Using the CLASH method culminates in the sequencing of chimeric molecules representing a ligation event between the small RNA and a fragment of its mRNA target [79]. Recently, CLASH was performed on the *C. elegans* PIWI Argonaute PRG-1, revealing *in vivo* interactions between piRNAs and their mRNA targets [137]. However, the paper does not provide a list of sites identified by CLASH. Here we reanalyzed their published CLASH data to provide transcriptome-wide piRNA targeting sites. As CLASH data are known to be noisy and chimera reads can result from random RNA ligations, we restricted our analysis to chimeric reads that appear at least 5 times in the ligated libraries. We analyzed CLASH data by first finding a piRNA whose full mature sequence perfectly matches a portion of a chimera. To identify the mRNA target of that piRNA, we mapped the remainder of the chimera sequences that were over 15 nucleotides in length to the *C. elegans* transcriptome. As steps of CLASH experiments require RNase treatment, the chimeras may experience RNA

degradation; therefore, to identify the best possible pairing between the piRNA and the target mRNA, we extended the length of the mapped mRNA region by 21 nucleotides from both the 5' and 3' ends of the read and predicted the piRNA targeting site as the interaction with the highest piRNA targeting score, similar to the approach taken in the previous study [65]. piRTarBase displays the targeted mRNA region of each chimera, while the predicted pairing between the mRNA and the mapped piRNA is found in a separate table as described below. In total, piRTarBase inventories 10116 CLASH-identified piRNA target sites.

5.3.4.4 mRNA and small RNA sequencing data

In *C. elegans* piRNAs trigger silencing of their targets using amplified siRNAs known as 22G RNAs. Therefore, in *prg-1* PIWI mutants, which lose all piRNAs, piRNA targets have been shown to have increased levels of mRNAs and reduced levels of 22G RNAs [10,87]. In order to allow users to better evaluate how piRNA targeting affects a particular gene, piRTarBase collects mRNA expression data from wildtype (N2) and PIWI mutant (*prg-1*) animals. Normalized mRNA expression data is displayed for each gene in wildtype and in PIWI mutant animals. piRTarBase also displays the mRNA fold change (averaged between replicates) in PIWI mutant vs. wildtype animals for each gene. Therefore, genes with mRNA fold changes greater than 1 are more likely to be regulated by piRNA targeting events. For example, it has recently been shown that *xol-1*, a master regulatory gene controlling dosage compensation and sex determination in *C. elegans,* is regulated by piRNAs [151]. Looking at the piRTarBase entry for *xol-1*, a user can see that the mRNA fold change in PIWI mutants is over 2, meaning that *xol-1* is more than twice as highly expressed in PIWI mutants. This confirms that *xol-1* is likely to be regulated by piRNAs.

piRTarBase also collects small RNA expression data from wildtype and PIWI mutant animals. "22G RNA seq data" displayed in piRTarBase reflects the accumulation of 22G RNAs

on a given mRNA species. This data is also available in wildtype and PIWI mutant animals. The calculated fold change compares the accumulation in 22G RNAs in PIWI mutants to that in wildtype animals. Because successful piRNA targeting events lead to the production of 22G RNAs [10,87], a reduction in 22G RNAs along a given gene in a PIWI mutant would be expected if that gene was being regulated by piRNAs. Therefore, a fold change less than 1 would suggest possible piRNA regulation of that gene. Taking *xol-1* as an example again, it can be seen that the 22G RNA fold change is 0.0215, which shows that 22G-RNAs are dramatically reduced in PIWI mutants and suggests that *xol-1* is being regulated by piRNAs.

5.3.4.5 Defining common piRNA targeting sites

pirScan's prediction algorithm is based on a robust reporter assay that allowed our group to optimize transgene sequences for germline expression in *C. elegans* [185]. While the sites predicted using pirScan robustly identifies sites that contribute to silencing *in vivo*, it also considers piRNAs that do not actually interact with their predicted mRNA targets. For example, it may be that some piRNAs are expressed at low levels, or a piRNA and its presumptive target are not expressed at the same developmental time, or that mRNAs are protected for piRNA interaction by anti-silencing signals such as periodic An/Tn repeats (PATCs) or CSR-1 Argonaute interaction [50,137]. Conversely, the targets predicted based on CLASH data may identify targets that interact with mRNA targets but do not actually contribute to silencing, as shown in the case of miRNAs [2]. Therefore, we examined whether the piRNA targeting sites identified by both methods (904 sites and 2273 sites using stringent and relaxed rules, respectively), referred to as common piRNA target sites, are more likely to represent targeting events that are sufficient for regulating mRNA targets. Indeed, for reported targets *xol-1* and Y40B10A.2a, removing only a single common piRNA targeting site displayed in piRTarBase (*21ur-4863* and *21ur-8264* targeting sites,

respectively) in each of these genes was sufficient to increase these genes' expression levels [137]. Furthermore, overall mRNA expression is significantly higher in PIWI mutants for genes whose targets were predicted using both pirScan and CLASH, but not for genes identified using each method separately (Figure 5.4A). Additionally, when we compare the 22G RNA fold changes in genes whose targets were predicted by pirScan, CLASH, or by both independently, we see that overall 22G RNAs are reduced significantly in PIWI mutants for target genes identified using both methods, but not for target genes identified by the individual methods (Figure 5.4B). Taken together, these observations suggest that the common piRNA target sites provided by piRTarBase are more likely to have regulatory effect on their targets *in vivo* than sites that are identified only by prediction or experimental approach alone.

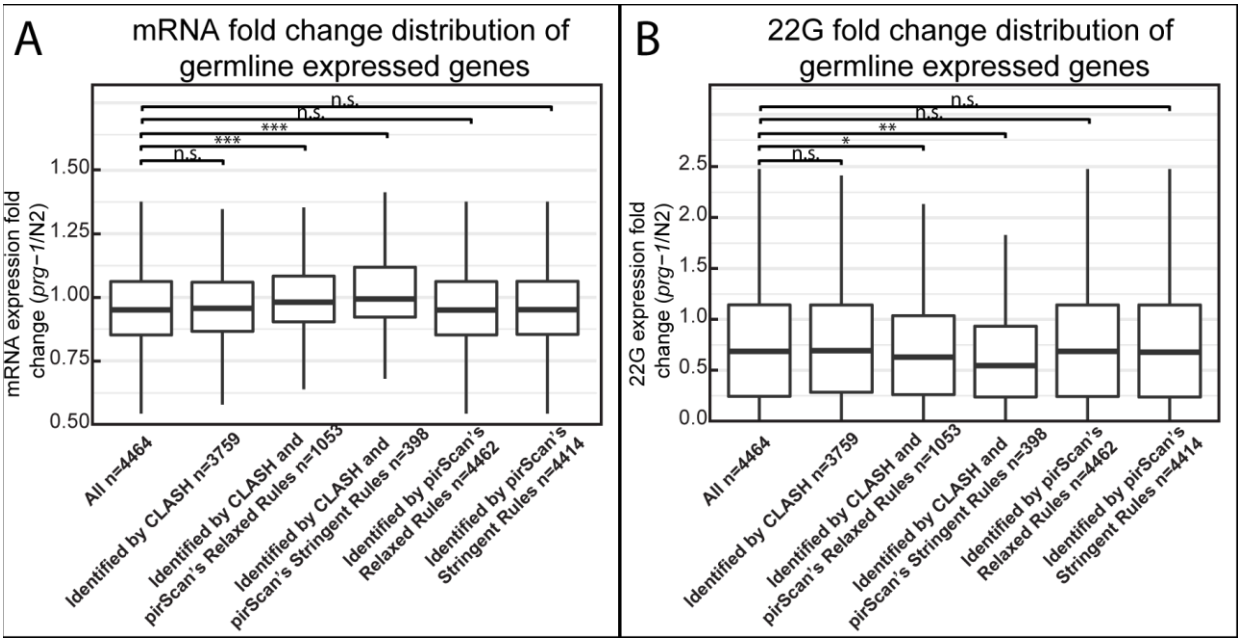


Figure 5.4: 22G RNA and mRNA fold expression changes in germline expressed genes with piRNA targeting sites predicted using different methods.
Only genes found to be expressed in the *C. elegans* germline according to [110] were considered. Distributions are displayed without outliers for clarity. Outliers were defined as points less than Q_1-1.5*IQR or greater than Q_3+1.5*IQR. Statistics were performed using the Mann-Whitney U test. Comparisons were made between each distribution and the full germline expression gene set. *** indicates p < 1e-9, ** indicates p < 1e-3, * indicates p < 0.05, n.s. indicates not significant. (A) mRNA fold expression changes, from [102], in genes with predicted piRNA targeting sites. (B) 22G RNA fold expression changes, from [87], in genes with predicted piRNA targeting sites. Note that *prg-1* mutant animals have a reduced number of germ cells compared to wildtype, which likely contributes to the reduction of overall levels of 22G RNAs and mRNAs among all germline expressed genes.

5.3.4.6 Comparison with other piRNA databases

Currently only one database (piRBase) of piRNA targeting sites is established. This piRNA database focuses on curating a list of piRNAs and their targets [187]. While such databases successfully condense the existing literature into a more easily searchable format, they lack the predictive power of piRTarBase. piRTarBase uniquely provides candidates of piRNA target sites on endogenous transcripts using pirScan's prediction algorithm and PIWI CLASH data. This allows piRTarBase to collect piRNA targeting sites genome-wide at a much higher confidence level than existing databases. In addition, piRTarBase offers a better integrative platform that

allows users to examine various information about piRNA targeting sites, including expression level of mRNA targets, piRNA and 22G-RNA. In addition, users can easily examine the position of mismatches, pairing score, abundance of chimera reads/piRNAs for each piRNA targeting site with sortable tables. While piRTarBase does not inventory the limited piRNA target sites reported in mice or flies, we plan to expand piRTarBase to other organisms once the corresponding piRNA targeting rules and PIWI CLASH data are available. Taken together, we believe that piRTarBase offers a user-friendly interface for researchers of diverse fields to explore the potential regulatory function of piRNAs.

5.3.4.7 Searching by gene in the interface

piRTarBase supports querying either by gene or by piRNA (Figure 5.5, 1. Input). The user can select the desired species (currently *C. elegans* or *C.briggsae*) and the desired stringency of the targeting rules. For gene/transcript querying, clicking search following input of the gene/transcript name results in a table displaying the queried terms, a table displaying mRNA and small RNA expression data, and a table displaying isoform-specific piRNA targeting site information (Figure 5.5, 2a. Targeting piRNAs). In the mRNA expression data, expression levels for the input gene are displayed in reads per kilobase million (RPKM) for wildtype (N2) animals and PIWI (*prg-1*) mutant animals, with two biological replicates for each strain. This information is summarized for the user in a mRNA fold change field, which calculates the average mRNA fold change in PIWI mutants relative to wildtype animals. The user can use this field to determine whether loss of PIWI results in an elevated level of expression (fold change > 1) or a reduced level of expression (fold change < 1). Additionally, 22G RNA sequencing data is available in reads per million (RPM) for wildtype and PIWI mutant animals. The 22G RNA fold change is displayed in a field, which will allow the user to determine whether loss of PIWI results in a reduction of 22G RNAs mapping to

that gene (fold change <1) or an accumulation of 22G RNAs mapping to that gene (fold change >1). As of the publication of this manuscript, piRTarBase curates all four publicly available 22G RNA sequencing datasets in *C. elegans* [15,54,87,137] and the single available 22G RNA sequencing dataset in *C.briggsae* [151]. The last table on the page displays the piRNA targeting sites for each mRNA isoform of the queried gene. Targeting sites are grouped into three categories: (1) Predicted piRNA target sites: targeting sites predicted by pirScan's algorithm (2) CLASH identified piRNA target sites: targeting sites predicted by CLASH data (3) Common piRNA target sites: targeting sites that were predicted by both pirScan and CLASH data. Additional information about each identified targeting site can be found by clicking "Show target sites" (Figure 5.5, 3a. Target Site Data). On this page, piRNA targeting sites are first displayed as a graphic, with pirScan predicted target sites in red, CLASH identified target sites in blue, and common piRNA target sites in yellow. Tables corresponding to each type of identified targeting site follow. Hovering over any displayed targeting site gives the user additional information about that site, including which piRNA is being displayed, the expression level of that piRNA, the number of hybrids/chimeras sequenced (if the piRNA was predicted by CLASH), the region targeted in the mRNA identified in the chimera and the predicted pairing region, the piRNA targeting score, and the mismatches and pairing between the piRNA and the targeted site. This information is additionally displayed in the tables that follow the graphic. By clicking the sorting icons, the order of piRNA sites displayed can be sorted by the values in various columns, such as piRNA abundance or piRNA targeting score.

5.3.4.8 Searching by piRNA in the interface

For querying by piRNA, the user can input a piRNA name and click search. The average piRNA expression data displayed in this case corresponds to the expression of that piRNA as determined

by PIWI immunoprecipitation (PRG-1 IP) small RNA sequencing studies (Figure 5.5, 2b. Targeted mRNAs). All mRNA targets are displayed in a single table, with each row containing all of the information outlined above for querying by gene, including mRNA and small RNA expression fold changes and numbers of target sites predicted for that particular piRNA on a given gene. Users can click the sorting icons to display the mRNA targets in ascending or descending order of the values in various columns, such as mRNA or 22G RNA fold change. To find more information about the piRNA-mRNA information displayed in each row, the user can click 'Show target sites' to open an analogous page as described for querying by gene, except only targeting sites involving the queried piRNA are displayed (Figure 5.5, 3b. Target Site Data).

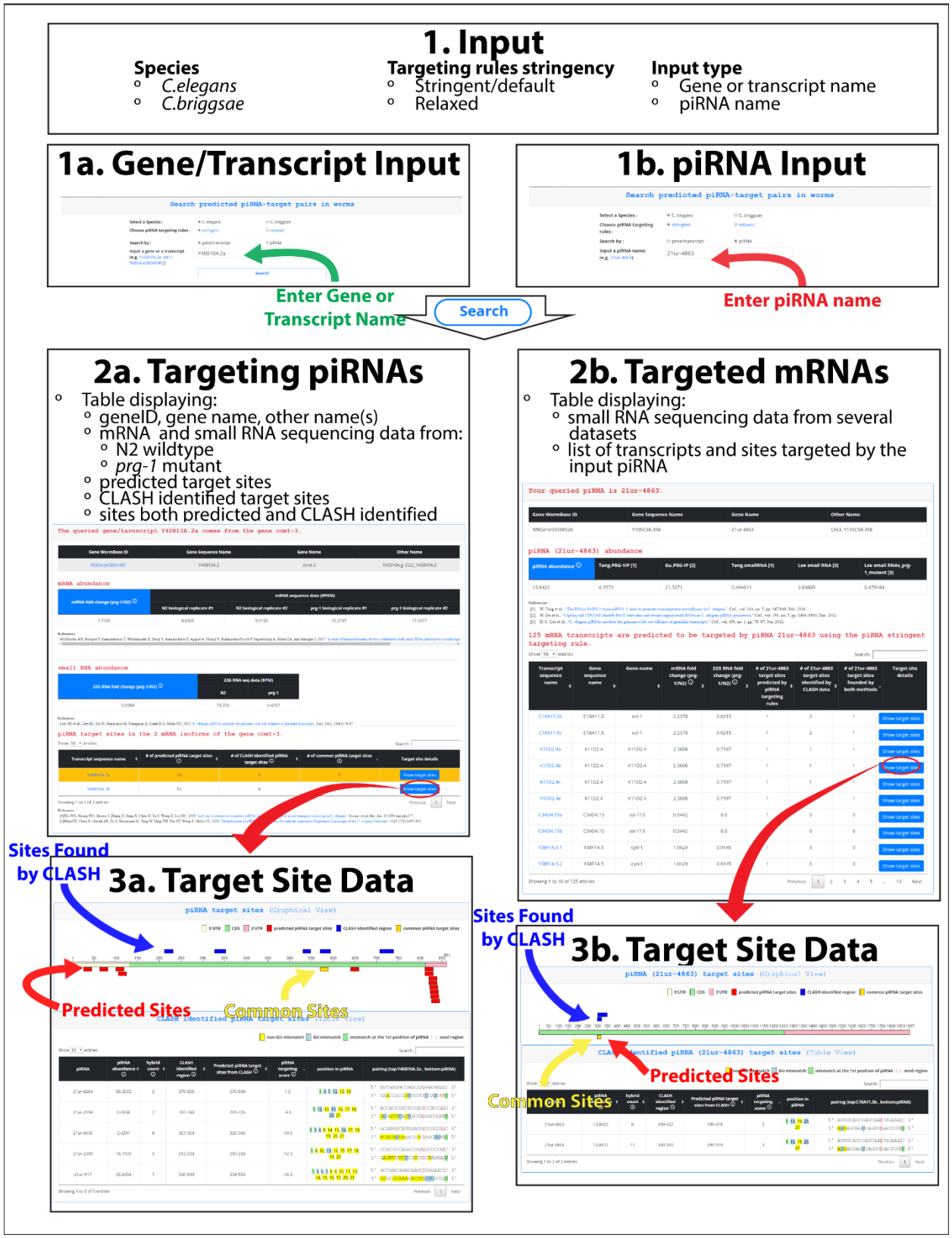


Figure 5.5: An overview of piRTarBase's user interface.

5.3.4.9 Browsing by genes or piRNAs in the interface

All transcripts or piRNAs can be displayed in a single sortable table. Browsing by transcript will allow users to find the mRNA and 22G RNA fold change information and the piRNA targeting site information, and to access the target site specific pages for each transcript. Browsing by piRNA will allow users to find expression information for each piRNA and its corresponding mRNA targets, and to access the target site specific pages for each target of the queried piRNA. Rows can be sorted by any field, allowing users to identify which mRNA transcripts contain the most target sites, or whose expression is most affected by loss of PIWI for example.

5.3.4.10 Summary and Prospective Works

piRTarbase provides the first integrative database allowing users to access both predicted and experimentally identified piRNA targeting sites and to explore their regulatory function. Importantly, our analyses demonstrate that the common piRNA sites identified by both bioinformatic and experimental approaches are more likely to have regulatory effects on their targets than those identified by the individual approaches. The user can further examine the change in mRNA and 22G-RNA expression of each specific target gene to assess the likelihood that piRNAs might regulate that target. Together, such an integrative platform uniquely allows users to sort through the hundreds of thousands of possible piRNA target sites and focus hypotheses on those sites that are more likely to be functional *in vivo*.

Currently, *C. elegans* is the only species whose piRNA targeting rules have been elucidated and the only species where PIWI CLASH has been performed. As prediction algorithms or CLASH data become available in other organisms, we will expand piRTarBase to include piRNA target sites for those organisms as well. For example, since the *C. elegans* piRNA targeting rules

are likely conserved in *C.briggsae* due to the high similarity between their PIWIs, we have expanded the piRTarBase to include the piRNA target sites in *C.briggsae*.

We believe that piRTarBase provides a user friendly and informative interface for researchers of diverse backgrounds to investigate whether genes of interest are regulated by piRNAs or vice versa. For example, recent quantitative trait locus (QTL) analyses in *C. elegans* have mapped the nictation behavior- controlling and other gene regulatory loci to regions inside the piRNA clusters on chromosome IV [86,128]. piRTarBase will allow researchers to identify candidates for the mRNA targets of the piRNAs located within these QTLs that may be contributing to the phenotype. Similarly, the authors of a recent study that linked piRNA regulation to axonal growth in *C. elegans* [75] will benefit from the ability to search through genes critical for neuronal regeneration and ask whether those genes may be targeted by specific piRNAs that underlie the regeneration defects. Such selective analyses were not possible prior to piRTarBase.

5.3.4.11 Acknowledgements

We thank Hao Chen for providing technical help on analyzing CLASH data. This work is supported by Ministry of Science of Technology of Taiwan grants (MOST-105-2221-E-006-203-MY2, MOST-106-2628-E-006–006-MY2, and MOST 107-2221-E-006 -225 -MY3) to Wei-Sheng Wu., a NIH predoctoral training grant (T32 GM07197) to the author., and a NIH R00 grant (GM108866) to Heng-Chi Lee.

5.3.5 CLASH Analyst: a webserver to identify in vivo RNA-RNA interactions from CLASH data

5.3.5.1 Motivation and purpose

RNA molecules perform some of the most fundamental and important functions in organisms, including coding for proteins (mRNAs), providing structural support (tRNAs, lncRNAs), acting

as vital enzymes (snRNAs, ribozymes), and guiding proteins to regulate other RNAs (miRNAs, piRNAs, siRNAs). Although discovered only relatively recently, this last category of RNA function has been shown to play a prominent role in posttranscriptional and transcriptional gene regulation. By guiding Argonaute family proteins to RNA targets using base-pairing interactions, miRNAs, piRNAs, and siRNAs can effectively and specifically target diverse mRNAs for regulation. For example, it has been estimated that ~50% of human genes are regulated by miRNA [9,13,91]. To attempt to identify functionally significant RNA–RNA interactions, researchers have historically relied on bioinformatic prediction to locate possible regulatory events based on sequence complementarity, sequence conservation of target sites, and by experimentally validated results of each regulatory site individually [115].

Recently, a technique has been developed to capture in vivo RNA–RNA interactions, called CLASH (crosslinking, ligation, and sequencing of hybrids) [65,79,137]. UV crosslinking and immunoprecipitation (CLIP) is a powerful biochemical approach to map protein-RNA interactions *in vivo* [161]. CLASH builds off of the idea of capturing RNA–protein complexes in vivo, as developed with CLIP, and adds a ligation step to produce chimeric RNA molecules made up of the regulatory RNA, which was held by the protein and the target RNA in proximity to regulatory RNAs in the cell (Figure 5.6, a model showing how CLASH works). In addition, several other techniques, such as iCLIP and CLEAR-CLIP [22,107], also generate similar chimeric RNA reads that reveal in vivo RNA–RNA interactions. Although CLASH and other CLASH-like data have proven powerful in identifying various types of RNA–RNA interactions, including miRNAs, piRNAs, and other non-coding RNAs with their targets, the data are not trivial to analyze. Sequenced CLASH reads represent fusions of two different RNA molecules that cannot be mapped using traditional approaches as accomplished with RNA-seq. Furthermore, currently developed

chimera-identifying programs such as CLAN and Hyb are limited in their accessibility to researchers with little bioinformatic knowledge [156,188]. CLAN identifies RNA–RNA pairs from chimeric reads but does not integrate other critical components of the analyses, including preprocessing of sequencing reads and postprocessing analyses such as calculation of base-pairing energy, leaving these necessary steps up to the user. Hyb is a command-line program that requires some degree of computational experience to use. Furthermore, neither of these tools include a visual display of these RNA–RNA interactions, nor do they have the capability to search for sites with specific criteria.

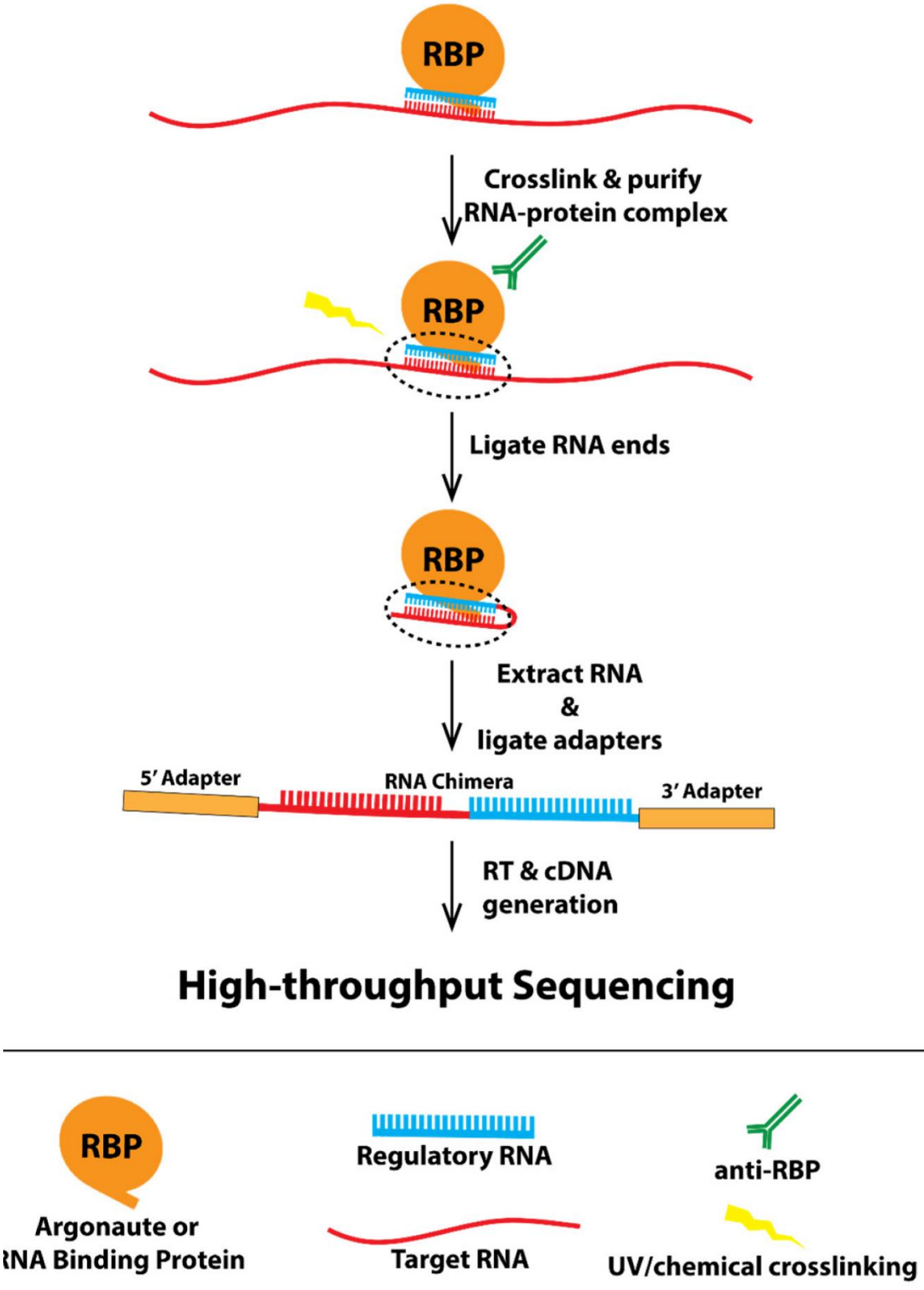


Figure 5.6: A model showing the experimental framework underlying CLASH.

5.3.5.2 General Framework

CLASH Analyst is a tool that identifies RNA–RNA interactions using user-provided raw CLASH data in three simple steps (Figure 5.7). First, the user will upload three files: CLASH raw data, a file containing the presumed regulatory RNA sequences, and a file containing the presumed target RNA sequences (Supplementary Figure 5.9A). Second, the user will choose to apply default or customized parameters to preprocess the raw sequencing reads, which removes RNA linker sequences and filters the sequenced reads based on desired sequencing quality. Third, the user can choose how CLASH Analyst will search for chimeric reads using a specific searching algorithm, such as pirTarBase, CLAN, or Hyb [156,177,188] (Supplementary Figure 5.9B). A Job ID and a link are provided upon job submission to access the results of the analyses. As the process takes some time to run, users can provide an email address so that they can be informed once the analysis is complete. Upon completing the analysis, the CLASH Analyst will provide two summary tables that indicate the parameters chosen for the search and the numbers of chimera and RNA–RNA interactions found (Supplementary Figure 5.9C). CLASH Analyst allows users to browse or search through the identified chimeric reads according to the names of their input regulatory RNAs, the names of their target RNAs, or by RNA–RNA pair according to the nature of the chimeric reads (abundance or pairing energy). For example, if users are interested in identifying the targets of a given regulatory RNA (such as *lin-4* miRNA), they can immediately sort all chimeric reads containing such RNA of interest, based on the abundance of the interaction in chimeric read count (Supplementary Figure 5.9D). For each RNA–RNA interaction, such as *lin-4* miRNA and its target *lin-14* (*T25C12.1a.1*), the user can click to reveal comprehensive information of each chimeric read, including where along each molecule the chimeric reads originated, how many reads were

captured, and where mismatches occurred within the base-pairing interaction (Supplementary Figure 5.9E). The user can also download the complete analysis results.

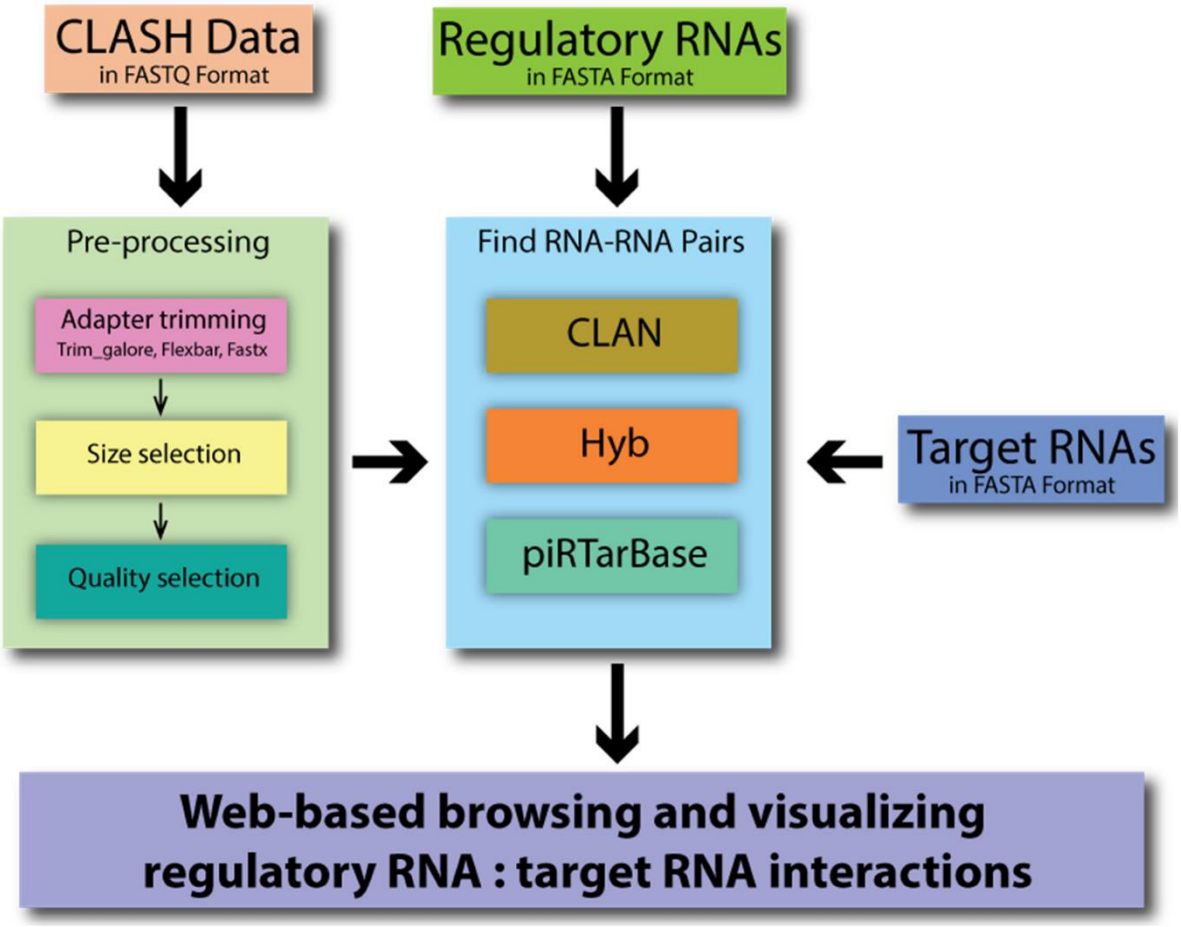


Figure 5.7: A graphical depiction of CLASH Analyst's workflow.

5.3.5.3 Data Input

The user must upload a compressed FASTQ file and two compressed FASTA files (Supplementary Figure 5.9A). The FASTQ file contains the raw data from the CLASH experiment that the user wishes to analyze. The first compressed FASTA file must correspond to the regulatory RNA sequences, and the second FASTA file should correspond to all RNA sequences the user wants CLASH Analyst to query as the target sequences. Example uploaded files are included and can be loaded by clicking "Load example files". In addition, common regulatory RNA sequences (e.g.,

human miRNAs, *C. elegans* miRNAs, *C. elegans* piRNAs) and target RNA sequences (e.g., human mRNAs, *C. elegans* mRNAs) are preloaded and can be selected by the user using a drop-down menu to skip uploading/downloading those data.

5.3.5.4 CLASH Read Pre-processing

CLASH Analyst performs all necessary steps to preprocess the raw reads (including adaptor/barcode sequence removal and reads quality control, etc.) so that chimeras can be identified. The user can choose default settings or select specific settings according to how the CLASH libraries were prepared (Supplementary Figure 5.9B). Users should provide adaptor information to allow precise trimming of adaptor sequences from the sequencing reads. Some library preparations include the ligation of a multiplexing barcode to the 5′ end of each RNA. This sequence should be provided so that CLASH Analyst can trim it from reads before downstream analysis. Some library preparations may also include some random nucleotides in the 5′ adaptor to act as unique molecular identifiers (UMIs). In this case, the user should provide the number of random nucleotides included in the 5′ adaptor to allow CLASH Analyst to trim UMIs. For example, if the 5′ adaptor includes 6 random nucleotides, then the user should input NNNNNN as part of the 5′ barcode input. Similarly, the 3′ adaptor sequence can be provided for precise trimming. Since Trim galore is capable of detecting 3′ adaptor sequences without prompting (although providing the sequence can help), Trim galore is chosen as the default trimming tool [97]. If the user chooses this tool, then 3′ adaptor sequence input is optional. The user can also select from two other tools, flexbar or fastx, to perform adaptor trimming [44,81]. In addition, the user provides the range of RNA sizes that will be analyzed. Based on how molecules were selected during library construction, the user should have an expectation of the length of reads following adaptor trimming and select/filter only reads of a certain length range for analyses. For example,

if RNA molecules between 17 nt and 70 nt long were selected and ligated to adaptors, then the trimmed reads should be 17–70 nt in length. Finally, the user can filter reads that do not meet a quality threshold, which is encoded in the FASTQ file called the Phred score. Phred scores are assigned during sequencing based on the probability of errant base calling. The default threshold for the Phred score is 30. A Phred score of 30, 40, or 50 for a particular base corresponds to a 99.9%, 99.9%, or 99.999% likelihood that the called base is accurate, respectively. CLASH Analyst can also accept data that have been preprocessed prior to being uploaded if the user wishes to perform these steps themselves.

5.3.5.5 Identify Chimeras

Once data are preprocessed, CLASH Analyst identifies chimeric reads using default or user-defined settings. The user can choose among three different algorithms to assign chimeric reads: CLAN [188], Hyb [156], and piRTarBase [177] (Supplementary Figure 5.9C). These algorithms differ on how they search for chimeric reads, resulting in some differences in the set of interactions they each identify (details can be found in "Section 5.3.5.11 Analysis Results"). The default searching tool in CLASH Analyst is CLAN.

If the user chooses to use CLAN, they can use default settings or specify criteria for searching chimeric reads, including (1) the minimum size of each fragment within a chimeric read for analyses; (2) how much overlap between the two fragments is allowed (if the default value of 4 is selected, then 4 nt would be allowed to map to each fragment simultaneously); (3) how many hits will be allowed for each fragment, as each fragment could match multiple user-input FASTA sequences. Similarly, the user can use default or specify criteria for searching chimeric reads using Hyb or piRTarBase. We recommend users first analyze their data with CLAN if they are unsure of which algorithm to choose, as CLAN performs the least stringent analysis and identifies more

RNA–RNA interaction candidates (for details, see Section 5.3.5.11). Then, to further refine the results, users can choose to rerun their data using piRTarBase or Hyb. To reanalyze data using a different algorithm quickly, the user can select the "Reanalyze old data?" option and adjust the algorithm settings accordingly after providing a Job ID (provided upon original submission). Users can access analysis results using the provided Job ID for 30 days. Extensions to this 30-day default can be provided upon reasonable request.

5.3.5.6 User Information

Once the user has uploaded all data and selected all necessary settings, they can submit the analysis job by pressing submit at the bottom of the page. A Job ID and a website link are provided for the users to check the status and examine the results. The analysis can typically take several hours to run (sometimes a day for larger datasets), so we provided an option to input an email address for the user to be notified once the analysis is complete. This is an optional, step and the email address is used only to notify and send the user a link to the results page.

5.3.5.7 Chimeric Reads Output

The user can input the Job ID or use the website link to browse the results within CLASH Analyst. The user can choose one of the three ways to display their results, which include browsing by (1) regulatory RNAs within the chimera, (2) target RNAs within the chimera, or (3) individual RNA–RNA pair ranked by chimera abundance or pairing energy. Regardless of what the user chooses to browse by, a summary of the analysis is first displayed in two tables (Supplementary Figure 5.9D). The first shows all the user-specified settings that were used to perform the analysis, while the second shows the number of various reads identified by CLASH Analyst, including (1) the number of reads provided by the user, (2) the number of reads and unique reads after trimming the adaptors, (3) the number of those trimmed reads that were identified as chimeric, and (4) the number of

RNA–RNA interactions identified. The browsable and searchable table containing the remainder of the output will depend on which browse mode was selected and will be discussed separately here. By selecting any of these three options, users can also download the full output files organized by CLASH read, regulatory RNA name, target RNA name, and RNA–RNA interaction in .csv format.

5.3.5.8 Browse by Regulatory RNA Name

The browsable and searchable output table will contain three columns: A column corresponding to regulatory RNA names, a column displaying the number of identified targets, and an option to show more details about the targets for each regulatory RNA (Supplementary Figure 5.9D). The user can choose to show more details for any regulatory RNA that was included in the input. This reveals all target RNA sequences identified within chimeras containing the selected regulatory RNA in a tabular format. Within the table, users can find the number of chimeric reads representing the number of unique chimeras containing that RNA–RNA interaction, such as *lin-4* and *lin-14 (T25C12.1a.1*). Additionally, any of the regulatory RNA–target RNA pairs can be further examined by clicking "show interaction". A graphic showing the interactions between target RNA and regulatory RNA is available for viewing (Supplementary Figure 5.9E). Additionally, a table that shows each identified interaction by chimeric read can be viewed. This table shows information about each hybrid, including the number of reads (read count), the thermodynamic favorability of the identified regulatory RNA and target RNA hybrid, where the interacting site mapped, and the predicted base pairing between the regulatory RNA and its target RNA.

5.3.5.9 Browse by Target RNA Name

The output table contains four columns: A column corresponding to the target RNA name, a column displaying the number of unique chimeras that contained that target RNA, a column

displaying the number of unique regulatory RNAs found to pair with that target, and an option to show interactions involving each target RNA (Supplementary Figure 5.9D). By clicking "show interaction", the user can view a detailed graphic and table outlining all interactions involving that target RNA. The graphic and table contain the same information outlined at the end of the Browse by Regulatory RNA Name section.

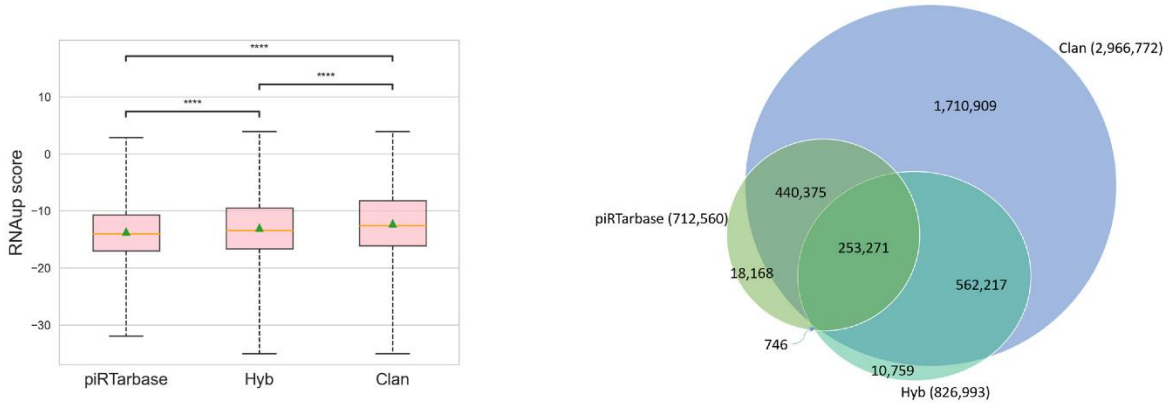
5.3.5.10 Browse by RNA-RNA Pair

Users can also browse by RNA–RNA pair. The results can either be sorted according to read count or according to the most favorable binding energy (Supplementary Figure 5.9D). The output table has rows corresponding to each unique chimera and presents information about each hybrid, including the number of reads (read count), the thermodynamic favorability of the identified regulatory RNA and target RNA hybrid, where the interacting site was mapped, and the predicted base pairing between the regulatory RNA and its target RNA. The webpage only shows the first 10,000 results. To view the rest, the user can download the result table in .csv format, as discussed above.

5.3.5.11 Analysis Results

Among the three searching algorithms available in CLASH Analyst, piRTarBase is the most stringent, and CLAN is the least stringent but most sensitive algorithm. Specifically, piRTarBase first searches for the regulatory RNA sequences within each read, and the default settings do not tolerate mismatches or deletions. It then uses the remaining sequence to identify targets with perfect matches. By using its default setting, piRTarBase requires the presence of intact regulatory RNA and does not allow insertions between regulatory RNA and target RNAs. Hyb tolerates short deletions and insertions in regulatory RNAs and target RNAs, and it also tolerates overlapping fragments. Nonetheless, Hyb will then choose one RNA–RNA interaction with the highest score

among multiple possible matches, which increases its searching stringency. The least stringent algorithm, CLAN, will identify more interactions using more sensitive searching and will report all possible matches, resulting in multiple interpretations of RNA–RNA interactions for some individual chimeric reads. To compare the performance of each of the three searching algorithms (CLAN, Hyb, and piRTarBase) [156,177,188] available in CLASH Analyst, we used a previously published *C. elegans* PRG-1 CLASH dataset (SRR6512652) [137] and compared the outputs of these three algorithms. We identified 712,560 RNA–RNA interactions from this dataset by piRTarBase, compared with 826,993 identified by Hyb and 2,966,772 identified by CLAN. We found that using the more stringent options (piRTarBase and Hyb), the identified chimeras generally have more favorable free energies of hybridization (Figure 5.8A), which suggests that they are more likely to represent true in vivo interactions [92]. In these analyses, CLAN can identify the majority of interactions identified by piRTarBase or Hyb and report many more possible interactions (Figure 5.8B), although some will represent lower confidence hits.



(**A**) (**B**)

Figure 5.8: Comparison of RNA–RNA interactions identified from three available searching algorithms in CLASH Analyst.
(A) the distribution of RNAup scores from RNA–RNA interactions identified in three searching algorithms. The lower the RNAup score, the more thermodynamic favorable interactions are predicted. ***: $p < 0.0001$; (B) a Venn diagram showing the unique and overlapping RNA–RNA interactions identified from three searching algorithms.

To compare CLASH Analyst's performance with previously analyzed results, we used the default parameter setting of CLASH Analyst to analyze three CLASH datasets including human miRNAs, *C. elegans* miRNAs, and *C. elegans* piRNAs [22,65,137]. Since previous CLASH studies of *C. elegans* miRNAs and piRNAs did not provide the full details of their results, we were not able to compare the number of targets found in CLASH Analyst to those from previous studies. Nonetheless, we found that *C. elegans* miRNA-binding sites are enriched at 3′ untranslated regions (UTRs) of their mRNAs, consistent with the previous report (Supplementary Figure 5.10) [22]. In addition, we found that regulatory RNA-binding sites identified by CLASH Analyst include all the key examples shown in the original CLASH studies, such as the interactions between *lin-4* miRNA and *lin-14* mRNA, as well as *21ur-1 (type 2)* piRNA and *xol-1* mRNA (Supplementary Figure 5.11). In addition, using a human miRNA CLASH dataset (SRR959751), we found that the default searching algorithm (CLAN) in CLASH Analyst identified significantly more miRNA–target interactions (13,860 vs. 1436) than using the Hyb searching algorithm, which was used to identify RNA–RNA interactions in the previous report (Supplementary Figure 5.12) [65]. Another recently published tool, ChiRA, was reported to identify approximately fourfold more interactions than the original human miRNA CLASH mRNA–target interactions [165]. Since ChiRA groups RNA–RNA interactions with similar interactions into common reads loci (CRLs), it is difficult to directly compare the results of ChiRA to CLASH Analyst. Taken together, we conclude that CLASH Analyst can identify similar or more miRNA targets than the currently published methods.

5.3.5.12 Acknowledgements

Ministry of Science of Technology of Taiwan (MOST 108-2628-E-006-004-MY3 and MOST 110-2221-E-006-198-MY3 to W.-S.W.); NIH predoctoral training grant (T32 GM07197 to J.B.); NIH R01 grant (GM132457 to H.-C.L.).

5.4 Discussion

As high throughput next-generation sequencing techniques continue to grow in popularity and specificity, the ability to interact with newly generated datasets will become increasingly necessary to researchers from diverse backgrounds. Routinely developing tools to simplify and parameterize complex analysis tasks will allow for the wide swath of small RNA researchers to keep pace and make use of valuable data.

The discovery of small noncoding RNAs has changed how we understand the factors contributing to gene expression. In a similar vain to the numerous miRNA target prediction tools that are available to predict miRNA binding, we have developed a tool to predict piRNA targeting sites using the piRNA targeting rules. This piRNA target prediction tool is the first of its kind for *C. elegans* and we believe it will be highly valuable in assisting the *C. elegans* community to solve the decades old dilemma of germline-specific transgene silencing. Our piRNA targeting rules have only been experimentally validated in *C. elegans*; however, we will expand pirScan once the piRNA rules are identified in other animal systems. In addition, we can update our tool to incorporate any additional factors that influence targeting. For example, the influence of piRNA abundance on silencing efficiency is not yet understood, but our tool can be easily updated in the future to consider expression level if this is established as a deterministic factor. As the piRNA targeting rules are likely to be operant in *C. briggsae* as well, we have also provided the option of using the piRNA data of *C. briggsae* by simply choosing *C. briggsae* in the input page [139]. As PIWI proteins are highly conserved among animals, pirScan offers an exciting platform of piRNA target prediction that can be expanded to various organisms in the future.

piRTarbase provides the first integrative database allowing users to access both predicted and experimentally identified piRNA targeting sites and to explore their regulatory function.

Importantly, our analyses demonstrate that the common piRNA sites identified by both bioinformatic and experimental approaches are more likely to have regulatory effects on their targets than those identified by the individual approaches. The user can further examine the change in mRNA and 22G-RNA expression of each specific target gene to assess the likelihood that piRNAs might regulate that target. Together, such an integrative platform uniquely allows users to sort through the hundreds of thousands of possible piRNA target sites and focus hypotheses on those sites that are more likely to be functional *in vivo*. Currently, *C. elegans* is the only species whose piRNA targeting rules have been elucidated and the only species where PIWI CLASH has been performed. As prediction algorithms or CLASH data become available in other organisms, we will expand piRTarBase to include piRNA target sites for those organisms as well. For example, since the *C. elegans* piRNA targeting rules are likely conserved in *C. briggsae* due to the high similarity between their PIWIs, we have expanded the piRTarBase to include the piRNA target sites in *C. briggsae*. We believe that piRTarBase provides a user friendly and informative interface for researchers of diverse backgrounds to investigate whether genes of interest are regulated by piRNAs or vice versa. For example, recent quantitative trait locus (QTL) analyses in *C. elegans* have mapped the nictation behavior- controlling and other gene regulatory loci to regions inside the piRNA clusters on chromosome IV [86,128]. piRTarBase will allow researchers to identify candidates for the mRNA targets of the piRNAs located within these QTLs that may be contributing to the phenotype. Similarly, the authors of a recent study that linked piRNA regulation to axonal growth in *C. elegans* [75] will benefit from the ability to search through genes critical for neuronal regeneration and ask whether those genes may be targeted by specific piRNAs that underlie the regeneration defects. Such selective analyses were not possible prior to piRTarBase.

CLASH is a critical approach to identify RNA–RNA interactions, including the interaction between small non-coding RNAs and their targets. However, the analyses of CLASH data are not trivial, and current tools are limited to researchers with advanced bioinformatic experience. Our goal in developing the CLASH Analyst was to allow researchers of varying levels of bioinformatics expertise to harness the power of CLASH to identify RNA–RNA interactions easily and quickly. CLASH Analyst can take raw sequencing reads and perform comprehensive analysis on CLASH data. CLASH Analyst further provides browser and virtualization tools that are critical for researchers to identify RNA–RNA interactions of interest. By analyzing published CLASH datasets using three available searching algorithms offered by CLASH Analyst, we showed that CLAN identified more RNA–RNA interactions, while pirTarbase identified high confidence interactions whose pairing are generally thermodynamically more stable. Since CLASH Analyst offers three searching algorithms with different stringencies, it empowers its users to perform customized searching by choosing the searching algorithm and selecting specific parameters based on their preferences and needs. In addition, we found that by repeating analysis of published CLASH datasets using CLASH Analyst, we could identify as many or more RNA–RNA interactions as reported in the respective original studies. We also confirmed that we could readily identify all RNA–RNA interaction signatures relevant to the respective studies, such as 3′ UTR enrichment for miRNA-targeting events and physiologically relevant miRNA and piRNA-targeting events. During the development of CLASH Analyst, our analyses of piRNA CLASH datasets revealed genome-wide, piRNA-binding sites in *C. elegans* [177]. In addition to exploring RNA–RNA interactions from user-generated CLASH datasets, CLASH Analyst will be valuable for researchers of diverse fields to identify RNA–RNA interactions of interest from published CLASH datasets. Many published CLASH studies highlight specific aspects of the RNA–RNA

interactions. The majority of RNA–RNA interactions were not mentioned due to space constraints or lack of interest. Although some of these papers include supplemental material that contains the processed results from the CLASH experiment, the list of RNA–RNA interactions is not exhaustive and may not include specific details of the interactions that are critical for further analyses. We envision CLASH Analyst as the perfect tool for researchers to retread previously published data, to find unappreciated results that can be used to support or formulate new hypotheses. In addition, CLASH Analyst can be used to identify noncanonical RNA–RNA interactions. For example, it has been proposed that Argonaute proteins can bind tRNA fragments (i.e., tRFs) to translationally repress mRNAs [140]. Researchers interested in this hypothesis can search through publicly available Argonaute CLASH datasets from various organisms to find evidence of tRF–mRNA interactions using CLASH Analyst. As RNA–RNA interactions have proven to be increasingly present in diverse biological contexts, novel techniques have been developed to capture these interactions. Although CLASH Analyst was developed to assist researchers in finding interactions from CLASH data, we anticipate our tool to be equally as useful in identifying RNA–RNA interactions in datasets from other methods which generate chimeric reads, such as recently developed techniques MARIO and RIC-seq [21,22]. Therefore, CLASH Analyst will lower the barrier for researchers who are interested in employing these exciting new tools but reticent due to their inexperience with advanced bioinformatics. We are confident that CLASH Analyst will allow users from any background to analyze their own data or previously published data to identify potential in vivo RNA–RNA interactions. We believe that with such advanced bioinformatic techniques available to all researchers in the RNA community, CLASH Analyst will allow for significant and novel advancement in our understanding of RNA regulation.

5.5 Methods

5.5.1 pirScan webserver construction

pirScan identifies piRNA targets by searching for homology between the input sequences and the *C. elegans* piRNA sequence data, which includes 15364 type I piRNAs that can be downloaded from WormMine (http://intermine.wormbase.org/tools/wormmine/) and 2485 type II piRNAs, which was collected from PRG-1 IP data [60] with at least 1 read per million reads in [152]. For *C. briggsae* piRNAs, a 14453 piRNA-containing database is composed of piRNAs that are cloned at least 1 read from one of three small RNA libraries separately on three strains, i.e., adult hermaphrodites, adult males, and embryos, and contain an upstream motif score of at least 5, based on the data in [139]. The search for piRNA target sites is performed using our Python script. The web interface of pirScan was constructed using Django, a Python web framework that encourages rapid web development. All figures in pirScan were generated using D3.js, a JavaScript library which provides powerful visualization components. All tables in pirScan were generated using DataTables, a table enhancing plug-in for the jQuery Javascript library which adds sorting, paging and filtering abilities to plain HTML tables with minimal effort. pirScan is available at http://cosbi4.ee.ncku.edu.tw/pirScan/. (main site), http://cosbi2.ee.ncku.edu.tw/pirScan/ (backup site 1) or http://cosbi5.ee.ncku.edu.tw/pirScan/ (backup site 2).

5.5.2 piRTarBase database construction

piRTarBase curates target sites that correspond to the WS230 *C. elegans* genome assembly and the WS230 *C.briggsae* assembly. The number of target sites compiled in piRTarBase are displayed in Table 5.2. Expression data was taken from [15,54,87,102,137,151].

The web interface of piRTarBase was constructed using the Python language with the Django MTV framework. The figures were generated by D3.js, which is a JavaScript library for manipulating documents based on data.

Method	Number of target sites	Number of genes	Number of transcripts
Stringent rules	571204	19781	32825
Relaxed rules	1420256	20097	33288
CLASH	10116	7880	16076
Stringent Rules & CLASH	904	903	1898
Relaxed Rules & CLASH	2273	2272	4876

Table 5.2: Database statistics for piRTarBase.

5.5.3 CLASH Analyst webserver construction

CLASH Analyst analysis pipeline behind the web interface was implemented using awk, bash, and Python scripts. The web interface of CLASH Analyst was constructed using Django, a Python web framework that encourages rapid web development. All figures in CLASH Analyst were generated using D3.js, a JavaScript library that provides powerful visualization components. All tables in CLASH Analyst were generated using DataTables, a table enhancing plug-in for the jQuery JavaScript library, which adds sorting, paging, and filtering abilities to plain HTML tables with minimal effort. CLASH Analyst is freely available for academic use at https://cosbi7.ee.ncku.edu.tw/CLASHanalyst or https://cosbi.ee.ncku.edu.tw/CLASHanalyst (backup site). The source code of CLASH Analyst can be downloaded from Github at https://github.com/t50504/CLASHanalyst.

5.6 Supplementary Information

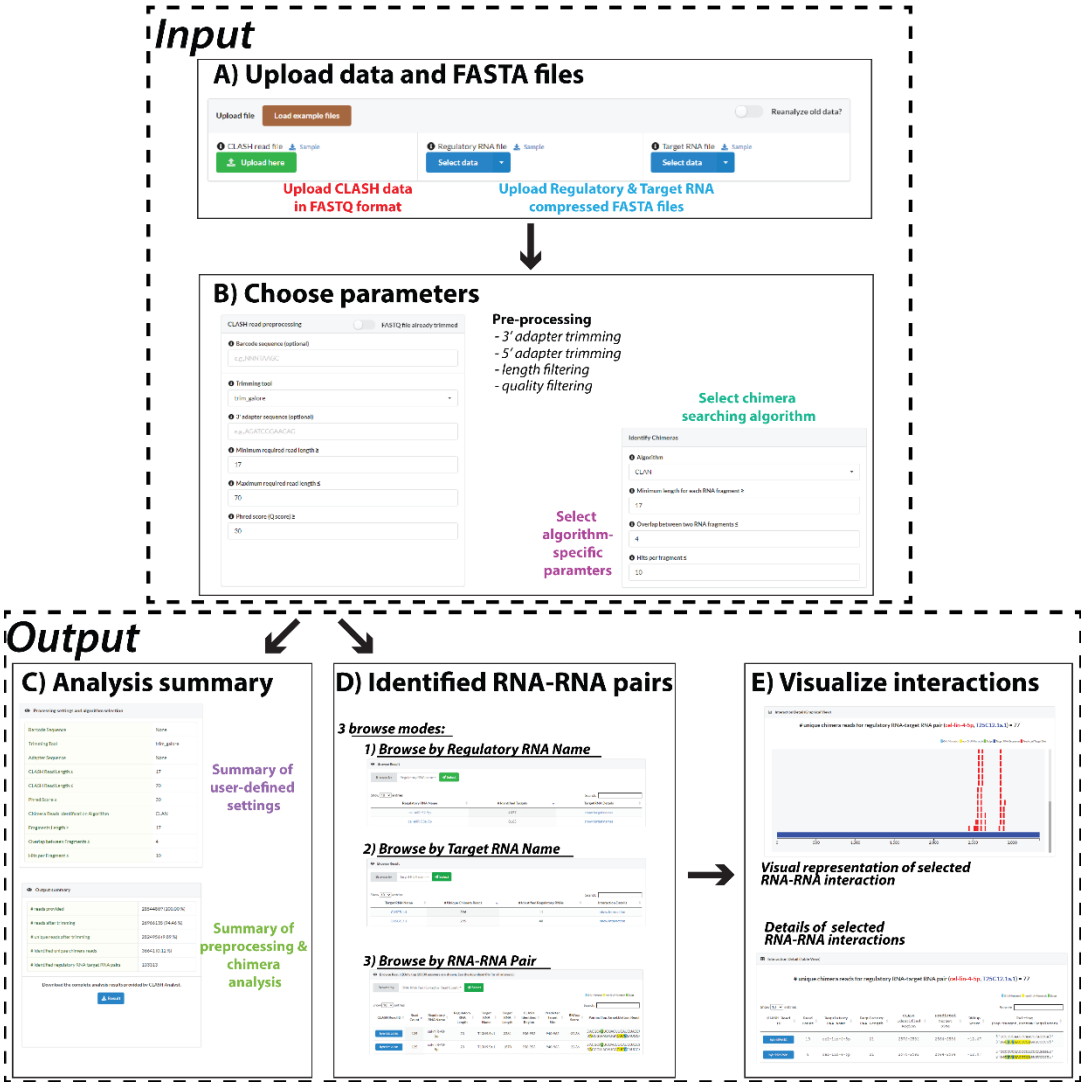


Figure 5.9: A step-by-step depiction of the CLASH Analyst workflow.

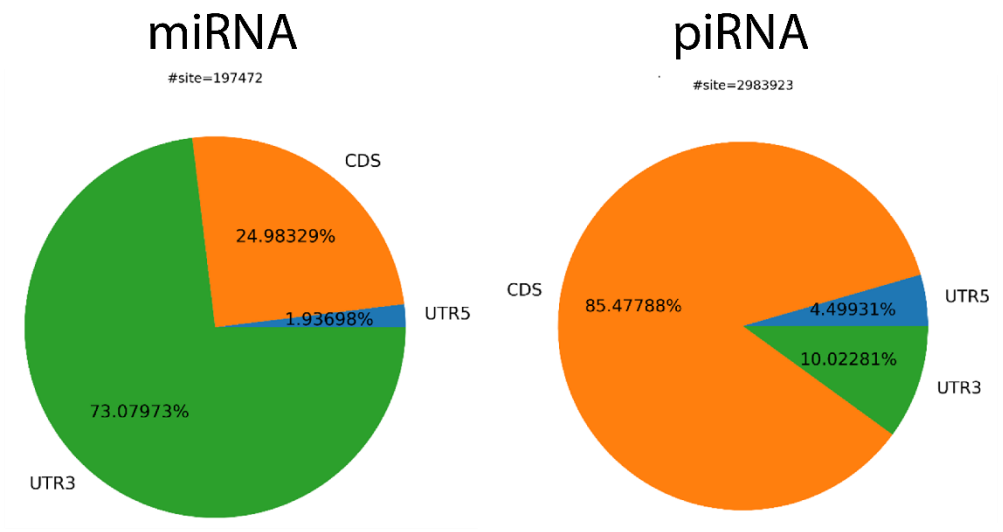


Figure 5.10: The distribution of miRNA and piRNA binding sites in C. elegans.
Note that while more piRNA targeting sites are found in CDSs, more miRNA targeting sites are found in 3' UTRs.

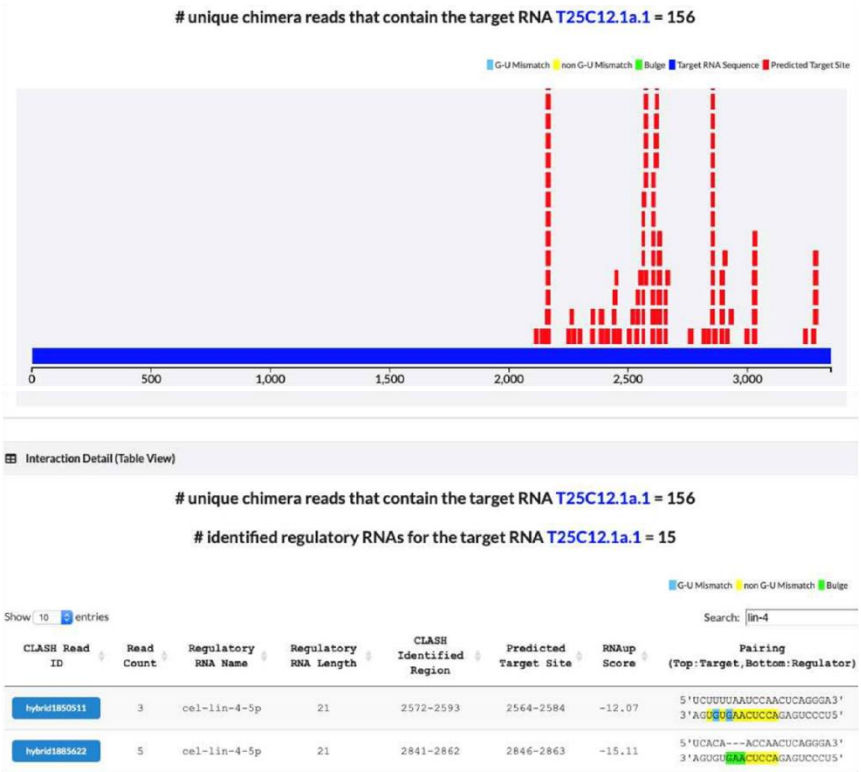


Figure 5.11: Examples of RNA-RNA interactions identified by CLASH Analyst.
The RNA-RNA interactions identified by CLASH Analyst, including *lin-4* miRNA and *lin-14* (*T25C12.1a.1*) mRNA (top), and *21u-x1* piRNA and *xol-1* (*C18A11.5c.1*) mRNA (bottom).

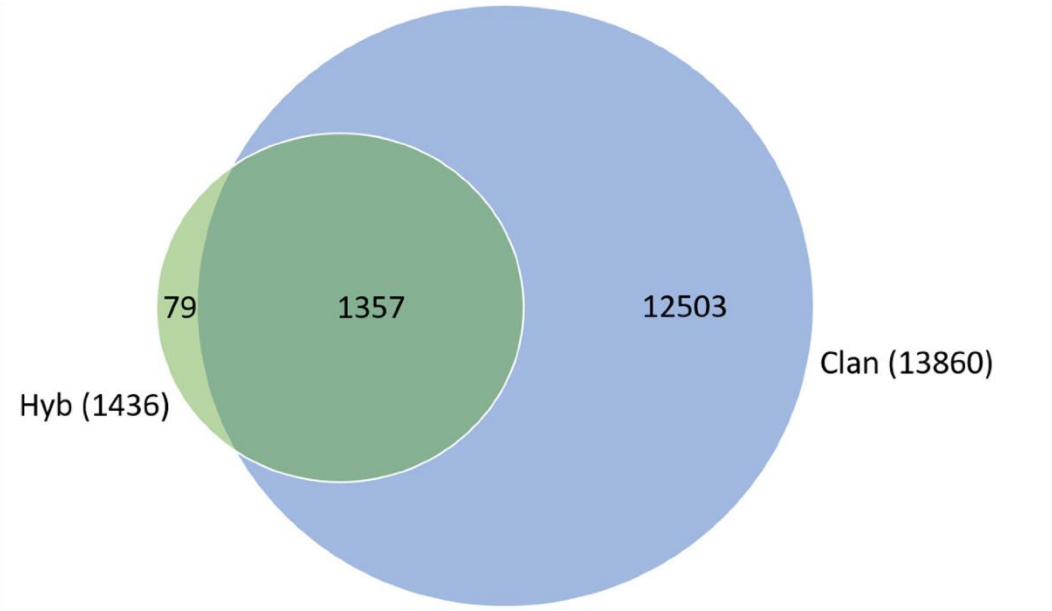


Figure 5.12: A comparison of RNA-RNA interactions identified by CLAN or Hyb.
A Venn diagram showing the unique and overlapping RNA-RNA interactions identified by CLAN or Hyb in analyzing the human miRNA CLASH data set (SRR959751).

CHAPTER 6

CONCLUSIONS

In this dissertation, I have discussed several facets of small RNA-mediated transcriptome surveillance in *C. elegans*. I have described novel factors involved in piRNA-mediated gene silencing, the functional impact of germ granules to small RNA pathway function, features common to endogenous piRNA targeting events, and several tools available to assist researchers in the small RNA community. In this final chapter, I will discuss the implications and outlook of this work. I will close by suggesting several future studies that will be essential to extend this work further.

6.1 The significance of germ granules to small RNA pathway function

A longstanding question in the field of *C. elegans* small RNA biology is what role, if any, do germ granules play in small RNA-mediated gene silencing. In Chapter 3, I reported several lines of evidence that in the absence of germ granules, self versus non-self distinction becomes blurred, so that a subset of typically silenced mRNAs become activated while a subset of typically expressed mRNAs become aberrantly silenced. This suggests that the germ granule does not represent a cellular locale for gene silencing, but rather plays a more complex role as an arbiter between competing signals that are generated and acted upon in the same place, namely at the sites of mRNA export.

There are some surprising implications of this work. One finding that was very surprising to us was that piRNA accumulation is relatively unaffected in germ granule mutants (Figure 3.10). Not only does PRG-1 accumulate in germ granules, but so too do key piRNA processing enzymes such as PARN-1 [152]. One naïve expectation we had was that the germ granule could function as a convenient localization structure to allow processing of piRNA precursors to occur as immature

piRNAs are exported into the cytoplasm and bound by PRG-1. It is easy to imagine that without some strategy of keeping newly exported precursors in a central location, PARN-1 could act too early and fully degrade piRNA precursors. Contrary to this idea, mature piRNA levels were unaffected in germ granule mutants, indicating that PRG-1 concentration at the site of nuclear export is not essential for piRNA maturation. This leads to the prediction that either the concentration of PRG-1 at sites of piRNA precursor nuclear export are sufficient to protect precursors prior to PARN-1 action even when germ granules are dispersed, or there is alternative mode of protection afforded to piRNA precursors between the time of their nuclear export to their association with PRG-1. To follow up on the implications of our finding, it will be important to understand more completely what factors are necessary for piRNA loading onto PRG-1 and whether piRNA precursors are actively protected from PARN-1 or other exonucleases following their transcription and until PRG-1 loading.

Similarly, the longstanding finding that small RNA factors, and piRNA pathway factors in particular, co-localize to germ granules has led to the hypothesis that the germ granule is necessary for efficient gene silencing. Under this model, germ granules concentrate silencing factors such as RdRPs and Argonautes so that siRNAs can efficiently find their targets, and so that degradation factors can rapidly degrade targeted mRNAs prior to their contact with the ribosome. Consistent with this subcellular partitioning model, it has been shown that germ granules do not contain appreciable amounts of rRNA and are therefore not likely sites of translation [132]. Our data contradict this model, as only a subset of WAGO 22G-RNAs fail to accumulate in the absence of germ granules (Figure 3.4). This does not exclude the possibility that the germ granule does indeed enhance the efficiency of silencing, but it suggests that small RNA silencing efficiency does not completely rely on germ granule partitioning.

The role of subcellular localization on regulatory outcome has become a major area of focus for small RNA biologists. It has recently been shown that manipulating the localization of a reporter mRNA by tethering it to a germ granule enriched protein is sufficient to change its expression state [5]. Additionally, it has been shown that cytoplasmic rather than germ granule localized CSR-1 regulates mRNAs post-transcriptionally for poorly translated mRNAs [144]. The data presented in Chapter 3 similarly show that when PRG-1 is not restricted to the germ granule, it begins to silence self mRNAs in the cytoplasm. An open question is whether aberrant PRG-1 silencing in the absence of germ granules occurs because CSR-1 fails to protect these transcripts, or rather is the function of PRG-1 in germ granules fundamentally different than cytoplasmic PRG-1. An interesting experiment to distinguish between these two models would be to restrict PRG-1 from entering the germ granule when the germ granule is still intact. An impediment to this experiment currently is that we do not understand how PRG-1 is recruited to the germ granule.

Another group published data similar to the work presented in Chapter 3 at nearly the same time that our group published that story, but with several key differences that led to very different interpretations [39]. The authors were focused on the same observation foundational to Chapter 3 of this work: the aberrant accumulation of 22G-RNAs targeting CSR-1 transcripts in *glh-1* mutants. They made two observations that contradict our work. First, they saw that mutation of GLH paralogs *glh-2*, *glh-3*, or *glh-4* in the presence of the *glh-1* mutation can suppress aberrant 22G accumulation. We saw that most mRNAs that accumulate aberrant 22G-RNAs in the absence of *glh-1* still do so in the absence of *glh-1* and *glh-4*, and that many of these targets aberrantly accumulate 22G-RNAs to an even higher degree (Figure 3.5). Second, they did not observe an aberrant accumulation of 22G-RNAs in the *glh-1 (DQAD)* mutant which fails to release RNA, directly in contrast to our observations (Figure 3.5). These differences led to an alternative model

to explain the aberrant accumulation of 22G-RNAs in the absence of GLH-1: GLH-1 itself prevents WAGO 22G-RNA accumulation against CSR-1 targets and in the absence of GLH-1, GLH paralogs promote 22G-RNA accumulation against these typically expressed mRNAs. The observation that the *glh-1 (DQAD)* mutant fails to accumulate 22G-RNAs aberrantly against CSR-1 targets is key to this interpretation, as its failure to release targeted mRNAs prevents GLH paralogs from promoting 22G-RNA accumulation in their model. Our interpretation of this similar data is that the germ granule and not GLH-1 itself prevents aberrant 22G-RNA accumulation. Key pieces of evidence in support of this model are the aforementioned continued accumulation of aberrant 22G-RNAs in the *glh-1 (DQAD)* mutant that we saw, and our finding that loss of other germ granule regulators leads to similar aberrant 22G-RNA accumulation (Figure 3.5). This contradicts the model proposed by the other group because in the *deps-1* and *mip-1/2* mutants, GLH-1 is still present and presumably still functioning, yet 22G-RNAs still aberrantly accumulate. Therefore, the key feature to explain aberrant 22G-RNA accumulation is the absence of germ granules. These two models, however, are not mutually exclusive. It could be that GLH paralogs regulate 22G-RNA accumulation against very different sets of targets, and also that germ granule integrity is important to preserve those differences and promote small RNA silencing fidelity.

6.2 The feasibility of predicting the regulatory fate of germline transcripts

The holy grail in demonstrating our understanding of how mRNAs are regulated by small RNAs in the germline would be to predict *de novo* whether a given mRNA will be silenced or expressed. To accomplish this goal, we would need a complete understanding of all intrinsic features of mRNAs sampled by small RNAs, and we would need complete mechanistic insight into how small RNAs interact with mRNAs. In Chapter 4 of this dissertation, I report one previously undescribed feature of piRNA targeting events: that piRNAs preferentially target the coding regions of

germline transcripts. Findings such as these inch us closer to that complete picture of transcript features that small RNAs interrogate. This observation currently does not provide additional mechanistic insight into how piRNAs and PRG-1 can use this intrinsic property of germline transcripts to enhance targeting potential. We are confident, however, that the anti-silencing Argonaute CSR-1 does not play a mechanistic role in this process (Figure 4.9).

The findings reported in Chapter 4 and Chapter 5 illustrate how wide the gap is between our current understanding of small RNA-based mRNA regulation and that holy grail. The elucidation of the piRNA targeting rules in *C. elegans* was a significant step forward in our ability to predict whether given sequences will be targeted by piRNAs [178,185]. However, we also saw that only relatively few endogenous sites predicted to be bound by piRNAs are actually bound *in vivo* [177]. This contradiction illustrates that additional and as yet unknown modes of regulation exist in the germline. One currently known regulator of piRNA targeting, anti-silencing Argonaute CSR-1, can explain some of the inconsistency between *in silico* predicted piRNA targeting sites and *in vivo* bound sites, but cannot explain our recent discovery that coding regions but not untranslated regions are highly enriched for piRNA targeting *in vivo* as described above. As more features of piRNA targets are found, it will be important to re-characterize our ability to predict piRNA targets *de novo* as a means of assessing how exhaustive our understanding of piRNA biology is.

6.3 Tool availability as a means for inclusivity in science

In biology and in science more broadly, experimentation has become higher throughput and datasets have become larger and increasingly more complex. Interaction with next-generation sequencing data has become a necessity for virtually any researcher in life sciences, even those that do not generate sequencing datasets themselves. For scientists that fall into this category, it

may be difficult to justify the hundreds of hours and sustained effort necessary to learn the bioinformatic techniques necessary to analyze and interpret next-generation sequencing, even when the insights gained from that data could contribute to their own research goals. In Chapter 5 of this work, I have described efforts we have made to allow scientists without any bioinformatic expertise to interact with complex sequencing data in an easy-to-use graphical format.

Publication of the servers and databases described in Chapter 5 help the community in two important and distinct ways. First, they help bridge the gap between valuable published datasets and the scientists who could benefit from exploring them but lack the technical knowhow to process the complex publicly available raw data. Second, they help increase transparency in our own work by taking an analytical pipeline that we have developed and wrapping it into a parametrized format that can be used to easily reproduce our results. As NGS applications become more diverse and specific, it will be important to foster this inclusive mindset so that the larger scientific community can gain insights from published data and also verify claims made about that data by the originators of it. It is not difficult to imagine that as more NGS applications are developed, even those of us who are intimately familiar with bioinformatics will not have the expertise necessary to analyze datasets of diverse and increasingly esoteric experimental origin. Novel RNA isolation techniques and library design constantly introduce new difficulties and complications in genomic data analysis. The upkeep of dedicated webservers and databases will allow data to be accessed and explored inclusively by the scientific community at large.

6.4 Future directions

A key question referenced many times throughout this work is how sufficiently does mRNA localization predict regulatory outcome. Much of the data presented and referenced reports correlations, where some mutation disrupts mRNA regulation and that disruption coincides with a

change in mRNA localization. These correlations suggest a relationship between localization and mRNA regulation, but a definitive causal link is missing. One promising experimental technique that could be used to address this gap is mRNA tethering, which has been used successfully in the *C. elegans* germline to bring particular proteins of interest into direct proximity with particular mRNAs [5,36,42,170]. This technique involves appending RNA hairpins derived from the lambda phage *box b* gene to an mRNA of interest that can be tightly bound by lambda phage N anti-termination protein fragment tagged to a protein of interest. Groups that have used this technique in *C. elegans* were mainly interested in the role of their tagged protein in gene regulation. This technique could also be used more generally to shuttle mRNAs of interest to different subcellular locales. For example, if a particular protein always accumulates in the cytoplasm and rarely enters the germ granule, then one could test the effect of excluding an mRNA from the germ granule by tethering that mRNA to the cytoplasmic protein. As the atlas of proteins grows, we know with increasing confidence and precision where particular proteins tend to reside in the cell. Therefore, it will become increasingly easier to enrich particular mRNAs in particular subcellular locations using a tethering approach. With a system like this in place, mRNA, small RNA, and peptide production could all be monitored to gain a full regulatory picture of a particular mRNA based solely on localization differences. In this way, localization as a causal determinant of mRNA regulation could be fairly assessed.

Another open question that I hoped to address more completely in my work is what determines the fate of 22G-RNAs manufactured by RRF-1 and EGO-1. As previously described, the RdRP EGO-1 is solely responsible for CSR-1 22G-RNA production, but it also contributes to WAGO 22G-RNA production as well. If these molecules are being produced by the same polymerase, presumably sharing the same molecular characteristics, then what determines whether

a given 22G-RNA will be bound by a WAGO Argonaute or by CSR-1? This decision is of great consequence, as CSR-1 targeted transcripts tend to be highly expressed while WAGO targeted transcripts tend to be silenced. Additionally, the decision to be bound by WAGO or CSR-1 happens in a relatively reproducible way, as there are few population level differences in the transcripts enriched for CSR-1 versus WAGO targeting when immunoprecipitation experiments are performed independently. Also, it has been shown previously that by resetting germline RNAi pathways by mutating key RNAi pathway factors and then reintroducing them, CSR-1 and WAGO bound 22G-RNAs become mis-sorted and confused, suggesting that the key determinant in this sorting is not likely to be a characteristic of the 22G-RNA molecule itself [4,117]. One simple model is that subcellular localization is the key determinant. CSR-1 bound 22G-RNA molecules do not depend on germ granules for their accumulation [144]. Therefore, EGO-1 activity in the germ granule could feed 22G-RNAs into the WAGO pathway while EGO-1 activity in the cytoplasm could feed 22G-RNAs into the CSR-1 pathway. If this model is correct, then mRNA residence time in the germ granule should be highly predictive of that mRNA's fate as a WAGO or CSR-1 target. We already know with high confidence which mRNAs are targeted by WAGOs versus by CSR-1. We are lacking information about transcriptome-wide mRNA subcellular localization. One very promising technique that could address this is multiplexed error-robust fluorescent *in situ* hybridization (MERFISH) [179]. MERFISH allows for simultaneous spatial quantification of thousands of mRNAs using a barcoding approach. Performing this technique in the *C. elegans* germline would give a snapshot of mRNA localization for nearly the entire germline transcriptome. Taking an array of snapshots would allow for high confidence quantification of where particular mRNAs tend to reside in subcellular space. Comparing the information gained

from a MERFISH experiment to the CSR-1/WAGO occupancy data that we have already could reveal much about the relationship between mRNA localization and templated 22G-RNA fate.

6.5 Concluding remarks

In this dissertation, I have described a reporter assay that revealed evolutionarily ancient factors involved in piRNA-mediated transcriptome surveillance. I have shown that germ granule loss leads to a breakdown of self versus non-self distinction in germ cells. I have reported that the coding regions of germline mRNAs are uniquely vulnerable to piRNA targeting, and I have helped increase the availability of the techniques used to draw these conclusions. This work has helped contribute to our understanding of small RNA pathways in the *C. elegans* germline, and more broadly has increased our understanding of gene regulation at large.

REFERENCES

[1] F. Aeschimann, P. Kumari, H. Bartake, D. Gaidatzis, L. Xu, R. Ciosk, H. Großhans, LIN41 Post-transcriptionally Silences mRNAs by Two Distinct and Position-Dependent Mechanisms, Mol. Cell. 65 (2017) 476-489.e4.

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

[3] A. Akay, T. Di Domenico, K.M. Suen, A. Nabih, G.E. Parada, M. Larance, R. Medhi, A.C. Berkyurek, X. Zhang, C.J. Wedeles, K.L.M. Rudolph, J. Engelhardt, M. Hemberg, P. Ma, A.I. Lamond, J.M. Claycomb, E.A. Miska, The Helicase Aquarius/EMB-4 Is Required to Overcome Intronic Barriers to Allow Nuclear RNAi Pathways to Heritably Silence Transcription, Dev. Cell. 42 (2017) 241-255.e6.

[4] B.F.M. de Albuquerque, M. Placentino, R.F. Ketting, Maternal piRNAs Are Essential for Germline Development following De Novo Establishment of Endo-siRNAs in Caenorhabditis elegans, Dev. Cell. 34 (2015) 448–456.

[5] S.T. Aoki, T.R. Lynch, S.L. Crittenden, C.A. Bingman, M. Wickens, J. Kimble, C. elegans germ granules require both assembly and localized regulators for mRNA repression, Nat. Commun. 12 (2021).

[6] A. Aravin, D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, P. Morris, M.J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J.J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan, T. Tuschl, A novel class of small RNAs bind to MILI protein in mouse testes, Nature. 442 (2006) 203–207.

[7] A.A. Aravin, G.J. Hannon, J. Brennecke, The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race, Science (80-.). (2007) 761–765.

[8] A. Ashe, A. Sapetschnig, E.-M. Weick, J. Mitchell, M.P. Bagijn, A.C. Cording, A.-L. Doebley, L.D. Goldstein, N.J. Lehrbach, J. Le Pen, G. Pintacuda, A. Sakaguchi, P. Sarkies, S. Ahmed, E.A. Miska, piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans., Cell. 150 (2012) 88–99.

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

[10] M.P. Bagijn, L.D. Goldstein, A. Sapetschnig, E.-M. Weick, S. Bouasker, N.J. Lehrbach, M.J. Simard, E.A. Miska, Function, Targets, and Evolution of Caenorhabditis elegans piRNAs, Science (80-.). 337 (2012) 574–578.

[11] D. Baillat, M. Hakimi, A.M. Näär, A. Shilatifard, N. Cooch, R. Shiekhattar, S. Louis, Integrator , a Multiprotein Mediator of Small Nuclear RNA Processing , Associates with the C-Terminal Repeat of RNA Polymerase II, 123 (2005) 265–276.

[12] D. Baillat, E.J. Wagner, Integrator: Surprisingly diverse functions in gene expression,

Trends Biochem. Sci. 40 (2015) 257–264.

[13] D.P. Bartel, MicroRNAs: Target Recognition and Regulatory Functions, Cell. 136 (2009) 215–233.

[14] G. Barucci, E. Cornes, M. Singh, B. Li, M. Ugolini, A. Samolygo, C. Didier, F. Dingli, D. Loew, P. Quarato, G. Cecere, Small-RNA-mediated transgenerational silencing of histone genes impairs fertility in piRNA mutants, Nat. Cell Biol. 22 (2020) 235–245.

[15] P.J. Batista, J.G. Ruby, J.M. Claycomb, R. Chiang, N. Fahlgren, K.D. Kasschau, D.A. Chaves, W. Gu, J.J. Vasale, S. Duan, D. Conte, S. Luo, G.P. Schroth, J.C. Carrington, D.P. Bartel, C.C. Mello, PRG-1 and 21U-RNAs Interact to Form the piRNA Complex Required for Fertility in C. elegans, Mol. Cell. 31 (2008) 67–78.

[16] T. Beltran, E. Pahita, S. Ghosh, B. Lenhard, P. Sarkies, Integrator is recruited to promoter-proximally paused RNA Pol II to generate Caenorhabditis elegans piRNA precursors, EMBO J. 40 (2021) 1–17.

[17] A.C. Berkyurek, G. Furlan, L. Lampersberger, T. Beltran, E. Weick, E. Nischwitz, I.C. Navarro, F. Braukmann, A. Akay, J. Price, F. Butter, P. Sarkies, E.A. Miska, The RNA polymerase II subunit RPB- 9 recruits the integrator complex to terminate Caenorhabditis elegans piRNA transcription, (2021) 1–21.

[18] E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigó, T.R. Gingeras, E.H. Margulies, Z. Weng, M. Snyder, E.T. Dermitzakis, R.E. Thurman, M.S. Kuehn, C.M. Taylor, S. Neph, C.M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J.A. Greenbaum, R.M. Andrews, P. Flicek, P.J. Boyle, H. Cao, N.P. Carter, G.K. Clelland, S. Davis, N. Day, P. Dhami, S.C. Dillon, M.O. Dorschner, H. Fiegler, P.G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K.D. James, B.E. Johnson, E.M. Johnson, T.T. Frum, E.R. Rosenzweig, N. Karnani, K. Lee, G.C. Lefebvre, P.A. Navas, F. Neri, S.C.J. Parker, P.J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F.S. Collins, J. Dekker, J.D. Lieb, T.D. Tullius, G.E. Crawford, S. Sunyaev, W.S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I.L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H.A. Hirsch, E.A. Sekinger, J. Lagarde, J.F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermüller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J.S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M.C. Dickson, D.J. Thomas, M.T. Weirauch, J. Gilbert, J. Drenkow, I. Bell, X. Zhao, K.G. Srinivasan, W.K. Sung, H.S. Ooi, K.P. Chiu, S. Foissac, T. Alioto, M. Brent, L. Pachter, M.L. Tress, A. Valencia, S.W. Choo, C.Y. Choo, C. Ucla, C. Manzano, C. Wyss, E. Cheung, T.G. Clark, J.B. Brown, M. Ganesh, S. Patel, H. Tammana, J. Chrast, C.N. Henrichsen, C. Kai, J. Kawai, U. Nagalakshmi, J. Wu, Z. Lian, J. Lian, P. Newburger, X. Zhang, P. Bickel, J.S. Mattick, P. Carninci, Y. Hayashizaki, S. Weissman, T. Hubbard, R.M. Myers, J. Rogers, P.F. Stadler, T.M. Lowe, C.L. Wei, Y. Ruan, K. Struhl, M. Gerstein, S.E. Antonarakis, Y. Fu, E.D. Green, U. Karaöz, A. Siepel, J. Taylor, L.A. Liefer, K.A. Wetterstrand, P.J. Good, E.A. Feingold, M.S. Guyer, G.M. Cooper, G. Asimenos, C.N. Dewey, M. Hou, S. Nikolaev, J.I. Montoya-Burgos, A. Löytynoja, S.

Whelan, F. Pardi, T. Massingham, H. Huang, N.R. Zhang, I. Holmes, J.C. Mullikin, A. Ureta-Vidal, B. Paten, M. Seringhaus, D. Church, K. Rosenbloom, W.J. Kent, E.A. Stone, S. Batzoglou, N. Goldman, R.C. Hardison, D. Haussler, W. Miller, A. Sidow, N.D. Trinklein, Z.D. Zhang, L. Barrera, R. Stuart, D.C. King, A. Ameur, S. Enroth, M.C. Bieda, J. Kim, A.A. Bhinge, N. Jiang, J. Liu, F. Yao, V.B. Vega, C.W.H. Lee, P. Ng, A. Yang, Z. Moqtaderi, Z. Zhu, X. Xu, S. Squazzo, M.J. Oberley, D. Inman, M.A. Singer, T.A. Richmond, K.J. Munn, A. Rada-Iglesias, O. Wallerman, J. Komorowski, J.C. Fowler, P. Couttet, A.W. Bruce, O.M. Dovey, P.D. Ellis, C.F. Langford, D.A. Nix, G. Euskirchen, S. Hartman, A.E. Urban, P. Kraus, S. Van Calcar, N. Heintzman, T. Hoon Kim, K. Wang, C. Qu, G. Hon, R. Luna, C.K. Glass, M.G. Rosenfeld, S.F. Aldred, S.J. Cooper, A. Halees, J.M. Lin, H.P. Shulha, X. Zhang, M. Xu, J.N.S. Haidar, Y. Yu, V.R. Iyer, R.D. Green, C. Wadelius, P.J. Farnham, B. Ren, R.A. Harte, A.S. Hinrichs, H. Trumbower, H. Clawson, J. Hillman-Jackson, A.S. Zweig, K. Smith, A. Thakkapallayil, G. Barber, R.M. Kuhn, D. Karolchik, L. Armengol, C.P. Bird, P.I.W. De Bakker, A.D. Kern, N. Lopez-Bigas, J.D. Martin, B.E. Stranger, A. Woodroffe, E. Davydov, A. Dimas, E. Eyras, I.B. Hallgrímsdóttir, J. Huppert, M.C. Zody, G.R. Abecasis, X. Estivill, G.G. Bouffard, X. Guan, N.F. Hansen, J.R. Idol, V.V.B. Maduro, B. Maskeri, J.C. McDowell, M. Park, P.J. Thomas, A.C. Young, R.W. Blakesley, D.M. Muzny, E. Sodergren, D.A. Wheeler, K.C. Worley, H. Jiang, G.M. Weinstock, R.A. Gibbs, T. Graves, R. Fulton, E.R. Mardis, R.K. Wilson, M. Clamp, J. Cuff, S. Gnerre, D.B. Jaffe, J.L. Chang, K. Lindblad-Toh, E.S. Lander, M. Koriabine, M. Nefedov, K. Osoegawa, Y. Yoshinaga, B. Zhu, P.J. De Jong, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature. 447 (2007) 799–816.

[19] K. Bohmert, AGO1 defines a novel locus of Arabidopsis controlling leaf development, EMBO J. 17 (1998) 170–180.

[20] C.P. Brangwynne, C.R. Eckmann, D.S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher, A.A. Hyman, Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation, Science (80-.). 324 (2009) 1729–1732.

[21] J. Brennecke, A.A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, G.J. Hannon, Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila, Cell. 128 (2007) 1089–1103.

[22] J.P. Broughton, M.T. Lovci, J.L. Huang, G.W. Yeo, A.E. Pasquinelli, Pairing beyond the Seed Supports MicroRNA Targeting Specificity, Mol. Cell. 64 (2016) 320–333.

[23] B.A. Buckley, K.B. Burkhart, S.G. Gu, G. Spracklin, A. Kershner, H. Fritz, J. Kimble, A. Fire, S. Kennedy, A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality, Nature. 489 (2012) 447–451.

[24] N.O. Burton, K.B. Burkhart, S. Kennedy, Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans, Proc. Natl. Acad. Sci. 108 (2011) 19683–19688.

[25] A.C. Campbell, D.L. Updike, CSR-1 and P granules suppress sperm-specific transcription in the C. elegans germline, Development. 142 (2015) 1745–1755.

[26] C. Cao, Z. Cai, R. Ye, R. Su, N. Hu, H. Zhao, Y. Xue, Global in situ profiling of RNA-RNA spatial interactions with RIC-seq, Nat. Protoc. 16 (2021) 2916–2946.

[27] M.A. Carmell, A. Girard, H.J.G. van de Kant, D. Bourc'his, T.H. Bestor, D.G. de Rooij, G.J. Hannon, MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline, Dev. Cell. 12 (2007) 503–514.

[28] M.A. Carmell, Z. Xuan, M.Q. Zhang, G.J. Hannon, The Argonaute family: Tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis, Genes Dev. 16 (2002) 2733–2742.

[29] G. Cecere, S. Hoersch, S. O'keeffe, R. Sachidanandam, A. Grishok, Global effects of the CSR-1 RNA interference pathway on the transcriptional landscape, Nat. Struct. Mol. Biol. 21 (2014) 358–365.

[30] W. Chen, J.S. Brown, T. He, W.S. Wu, S. Tu, Z. Weng, D. Zhang, H.C. Lee, GLH/VASA helicases promote germ granule formation to ensure the fidelity of piRNA-mediated transcriptome surveillance, Nat. Commun. 13 (2022).

[31] W. Chen, Y. Hu, C.F. Lang, J.S. Brown, S. Schwabach, X. Song, Y. Zhang, E. Munro, K. Bennett, D. Zhang, H.C. Lee, The dynamics of P granule liquid droplets are regulated by the caenorhabditis elegans germline RNA helicase GLH-1 via its ATP hydrolysis cycle, Genetics. 215 (2020) 421–434.

[32] P.G. Cipriani, O. Bay, J. Zinno, M. Gutwein, H.H. Gan, V.K. Mayya, G. Chung, J. Chen, H. Fahs, Y. Guan, T.F. Duchaine, M. Selbach, F. Piano, K.C. Gunsalus, Novel LOTUS-domain proteins are organizational hubs that recruit C. elegans Vasa to germ granules, Elife. 10 (2021) 1–34.

[33] J.M. Claycomb, P.J. Batista, K.M. Pang, W. Gu, J.J. Vasale, J.C. van Wolfswinkel, D.A. Chaves, M. Shirayama, S. Mitani, R.F. Ketting, D. Conte, C.C. Mello, The Argonaute CSR-1 and Its 22G-RNA Cofactors Are Required for Holocentric Chromosome Segregation, Cell. 139 (2009) 123–134.

[34] C.C. Conine, P.J. Batista, W. Gu, J.M. Claycomb, D.A. Chaves, M. Shirayama, C.C. Mello, Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in Caenorhabditis elegans, Proc. Natl. Acad. Sci. 107 (2010) 3588–3593.

[35] C.C. Conine, J.J. Moresco, W. Gu, M. Shirayama, D. Conte, J.R. Yates, C.C. Mello, Argonautes promote male fertility and provide a paternal memory of germline gene expression in C. Elegans, Cell. 155 (2013) 1532–1544.

[36] E. Cornes, L. Bourdon, M. Singh, F. Mueller, P. Quarato, E. Wernersson, M. Bienko, B. Li, G. Cecere, piRNAs initiate transcriptional silencing of spermatogenic genes during C. elegans germline development, Dev. Cell. 57 (2022) 180-196.e7.

[37] C. Coronnello, P. V. Benos, ComiR: combinatorial microRNA target prediction tool,

Nucleic Acids Res. 41 (2013) W159–W164.

[38] D.N. Cox, A. Chao, H. Lin, Piwi Encodes a Nucleoplasmic Factor Whose Activity Modulates the Number and Division Rate of Germline Stem Cells, Development. 127 (2000) 503–514.

[39] S. Dai, X. Tang, L. Li, T. Ishidate, A.R. Ozturk, H. Chen, A.L. Dube, Y.-H. Yan, M.-Q. Dong, E.-Z. Shen, C.C. Mello, A family of C. elegans VASA homologs control Argonaute pathway specificity and promote transgenerational silencing, Cell Rep. 40 (2022) 111265.

[40] P.P. Das, M.P. Bagijn, L.D. Goldstein, J.R. Woolford, N.J. Lehrbach, A. Sapetschnig, H.R. Buhecha, M.J. Gilchrist, K.L. Howe, R. Stark, N. Matthews, E. Berezikov, R.F. Ketting, S. Tavaré, E.A. Miska, Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to Suppress Tc3 Transposon Mobility in the Caenorhabditis elegans Germline, Mol. Cell. 31 (2008) 79–90.

[41] D.J. Dickinson, J.D. Ward, D.J. Reiner, B. Goldstein, Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination, Nat. Methods. 10 (2013) 1028–1034.

[42] Y.-H. Ding, H. Ochoa, T. Ishidate, C. Mello, The nuclear Argonaute HRDE-1 directs target gene re-localization and shuttles to nuage to promote small RNA mediated inherited silencing, BioRxiv. (2022) 2022.10.04.510877.

[43] A.E. Dodson, S. Kennedy, Germ Granules Coordinate RNA-Based Epigenetic Inheritance Pathways, Dev. Cell. 50 (2019) 704-715.e4.

[44] M. Dodt, J.T. Roehr, R. Ahmed, C. Dieterich, FLEXBAR-flexible barcode and adapter processing for next-generation sequencing platforms, Biology (Basel). 1 (2012) 895–905.

[45] M.G. Dozmorov, C.B. Giles, K.A. Koelsch, J.D. Wren, Systematic classification of non-coding RNAs by epigenomic similarity, BMC Bioinformatics. 14 (2013).

[46] E.M. Eddy, Germ plasm and the differentiation of the germ cell line., Int. Rev. Cytol. 43 (1975) 229–80.

[47] A. Eulalio, I. Behm-Ansmant, D. Schweizer, E. Izaurralde, P-body formation is a consequence, not the cause, of RNA-mediated gene silencing., Mol. Cell. Biol. 27 (2007) 3970–81.

[48] N. Ezzeddine, J. Chen, B. Waltenspiel, B. Burch, T. Albrecht, M. Zhuo, W.D. Warren, W.F. Marzluff, E.J. Wagner, A Subset of Drosophila Integrator Proteins Is Essential for Efficient U7 snRNA and Spliceosomal snRNA 3'-End Formation, Mol. Cell. Biol. 31 (2010) 328–341.

[49] S.E.J. Fischer, T.A. Montgomery, C. Zhang, N. Fahlgren, P.C. Breen, A. Hwang, C.M. Sullivan, J.C. Carrington, G. Ruvkun, The ERI-6/7 helicase acts at the first stage of an siRNA amplification pathway that targets recent gene duplications, PLoS Genet. 7 (2011).

[50] C. Frøkjær-Jensen, N. Jain, L. Hansen, M.W. Davis, Y. Li, D. Zhao, K. Rebora, J.R.M. Millet, X. Liu, S.K. Kim, D. Dupuy, E.M. Jorgensen, A.Z. Fire, An Abundant Class of Non-coding DNA Can Prevent Stochastic Gene Silencing in the C. elegans Germline, Cell. 166 (2016) 343–357.

[51] A. Gardini, D. Baillat, M. Cesaroni, D. Hu, J.M. Marinis, E.J. Wagner, M.A. Lazar, A. Shilatifard, R. Shiekhattar, Integrator regulates transcriptional initiation and pause release following activation, Mol. Cell. 56 (2014) 128–139.

[52] D.T. Ge, W. Wang, C. Tipping, I. Gainetdinov, Z. Weng, P.D. Zamore, The RNA-Binding ATPase, Armitage, Couples piRNA Amplification in Nuage to Phased piRNA Production on Mitochondria, Mol. Cell. 74 (2019) 982-995.e6.

[53] A. Girard, R. Sachidanandam, G.J. Hannon, M.A. Carmell, A germline-specific class of small RNAs binds mammalian Piwi proteins, Nature. 442 (2006) 199–202.

[54] W.-S.S. Goh, J.W.E. Seah, E.J. Harrison, C. Chen, C.M. Hammell, G.J. Hannon, A genome-wide RNAi screen identifies factors required for distinct stages of C. elegans piRNA biogenesis, Genes Dev. 28 (2014) 797–807.

[55] W.S.S. Goh, I. Falciatori, O.H. Tam, R. Burgess, O. Meikar, N. Kotaja, M. Hammell, G.J. Hannon, piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis, Genes Dev. 29 (2015) 1032–1044.

[56] E. Gómez-Orte, B. Sáenz-Narciso, A. Zheleva, B. Ezcurra, M. de Toro, R. López, I. Gastaca, H. Nilsen, M.P. Sacristán, R. Schnabel, J. Cabello, Disruption of the Caenorhabditis elegans Integrator complex triggers a non-conventional transcriptional mechanism beyond snRNA genes, 2019.

[57] L.-T. Gou, P. Dai, J.-H. Yang, Y. Xue, Y.-P. Hu, Y. Zhou, J.-Y. Kang, X. Wang, H. Li, M.-M. Hua, S. Zhao, S.-D. Hu, L.-G. Wu, H.-J. Shi, Y. Li, X.-D. Fu, L.-H. Qu, E.-D. Wang, M.-F. Liu, Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis, Cell Res. 24 (2014) 680–700.

[58] S.T. Grivna, A novel class of small RNAs in mouse spermatogenic cells, Genes Dev. 20 (2006) 1709–1714.

[59] M.E. Gruidl, P.A. Smith, K.A. Kuznicki, J.S. McCrone, J. Kirchner, D.L. Roussell, S. Strome, K.L. Bennett, Multiple potential germ-line helicases are components of the germ-line-specific P granules of Caenorhabditis elegans., Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 13837–42.

[60] W. Gu, H.-C. Lee, D. Chaves, E.M. Youngman, G.J. Pazour, D. Conte, C.C. Mello, CapSeq and CIP-TAP Identify Pol II Start Sites and Reveal Capped Small RNAs as C. elegans piRNA Precursors, Cell. 151 (2012) 1488–1500.

[61] W. Gu, M. Shirayama, D. Conte, J. Vasale, P.J. Batista, J.M. Claycomb, J.J. Moresco, E.M. Youngman, J. Keys, M.J. Stoltz, C.-C.G. Chen, D.A. Chaves, S. Duan, K.D.

Kasschau, N. Fahlgren, J.R. Yates, S. Mitani, J.C. Carrington, C.C. Mello, Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the C. elegans Germline, Mol. Cell. 36 (2009) 231–244.

[62] S. Guang, A.F. Bochner, D.M. Pavelec, K.B. Burkhart, S. Harding, J. Lachowiec, S. Kennedy, An Argonaute Transports siRNAs from the Cytoplasm to the Nucleus, Science (80-.). 321 (2008) 537–541.

[63] M. Hanazawa, M. Yonetani, A. Sugimoto, PGL proteins self associate and bind RNPs to mediate germ granule assembly in C. elegans, J. Cell Biol. 192 (2011) 929–937.

[64] A. Harel, A. V. Orjalo, T. Vincent, A. Lachish-Zalait, S. Vasu, S. Shah, E. Zimmerman, M. Elbaum, D.J. Forbes, Removal of a single pore subcomplex results in vertebrate nuclei devoid of nuclear pores, Mol. Cell. 11 (2003) 853–864.

[65] A. Helwak, G. Kudla, T. Dudnakova, D. Tollervey, Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding, Cell. 153 (2013) 654–665.

[66] J.C. Huang, T. Babak, T.W. Corson, G. Chua, S. Khan, B.L. Gallie, T.R. Hughes, B.J. Blencowe, B.J. Frey, Q.D. Morris, Using expression profiling data to identify human microRNA targets, Nat. Methods. 4 (2007) 1045–1049.

[67] G. Hutvagner, M.J. Simard, Argonaute proteins: key players in RNA silencing., Nat. Rev. Mol. Cell Biol. 9 (2008) 22–32.

[68] T. Ishidate, A.R. Ozturk, D.J. Durning, R. Sharma, E. zhi Shen, H. Chen, M. Seth, M. Shirayama, C.C. Mello, ZNFX-1 Functions within Perinuclear Nuage to Balance Epigenetic Signals, Mol. Cell. 70 (2018) 639-649.e6.

[69] F. V Jacinto, C. Benner, M.W. Hetzer, The nucleoporin Nup153 regulates embryonic stem cell pluripotency through gene silencing, Genes Dev. 29 (2015) 1224–1238.

[70] D.M. Kasper, G. Wang, K.E. Gardner, T.G. Johnstone, V. Reinke, The C. elegans SNAPc Component SNPC-4 Coats piRNA Domains and Is Globally Required for piRNA Abundance, Dev. Cell. 31 (2014) 145–158.

[71] A. Kassambara, ggpubr: “ggplot2” Based Publication Ready Plots, (2020).

[72] W.G. Kelly, A. Fire, Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans., Development. 125 (1998) 2451–6.

[73] J.A. Kerins, M. Hanazawa, M. Dorsett, T. Schedl, PRP-17 and the pre-mRNA splicing pathway are preferentially required for the proliferation versus meiotic development decision and germline sex determination in Caenorhabditis elegans, Dev. Dyn. 239 (2010) 1555–1572.

[74] M.G. Kidwell, Transposable elements and the evolution of genome size in eukaryotes,

Genetica. 115 (2002) 49–63.

[75] K.W. Kim, N.H. Tang, M.G. Andrusiak, Z. Wu, A.D. Chisholm, Y. Jin, A Neuronal piRNA Pathway Inhibits Axon Regeneration in C. elegans, Neuron. 97 (2018) 511-519.e6.

[76] T. Kiuchi, H. Koga, M. Kawamoto, K. Shoji, H. Sakai, Y. Arai, G. Ishihara, S. Kawaoka, S. Sugano, T. Shimada, Y. Suzuki, M.G. Suzuki, S. Katsuma, A single female-specific piRNA is the primary determiner of sex in the silkworm, Nature. 509 (2014) 633–636.

[77] J. Kreher, T. Takasaki, C. Cockrum, S. Sidoli, B.A. Garcia, O.N. Jensen, S. Strome, Distinct Roles of Two Histone Methyltransferases in Transmitting H3K36me3-Based Epigenetic Memory Across Generations in Caenorhabditis elegans, Genetics. 210 (2018) 969–982.

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

[79] G. Kudla, S. Granneman, D. Hahn, J.D. Beggs, D. Tollervey, Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast, Proc. Natl. Acad. Sci. 108 (2011) 10010–10015.

[80] S. Kuramochi-Miyagawa, T. Watanabe, K. Gotoh, K. Takamatsu, S. Chuma, K. Kojima-Kita, Y. Shiromoto, N. Asada, A. Toyoda, A. Fujiyama, Y. Totoki, T. Shibata, T. Kimura, N. Nakatsuji, T. Noce, H. Sasaki, T. Nakano, MVH in piRNA processing and gene silencing of retrotransposons, Genes Dev. 24 (2010) 887–892.

[81] H. Lab, FASTX-Toolkit (RRID:SCR_005534), (n.d.).

[82] F. Lai, A. Gardini, A. Zhang, R. Shiekhattar, Integrator mediates the biogenesis of enhancer RNAs, Nature. 525 (2015) 399–403.

[83] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods. 9 (2012) 357–359.

[84] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (2009).

[85] N.C. Lau, Characterization of the piRNA Complex from Rat Testes, Science (80-.). 313 (2006) 363–367.

[86] D. Lee, H. Yang, J. Kim, S. Brady, S. Zdraljevic, M. Zamanian, H. Kim, Y. Paik, L. Kruglyak, E.C. Andersen, J. Lee, The genetic basis of natural variation in a phoretic behavior, Nat. Commun. 8 (2017) 273.

[87] H.-C. Lee, W. Gu, M. Shirayama, E. Youngman, D. Conte, C.C. Mello, C. elegans piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts, Cell. 150 (2012) 78–87.

[88] R.C. Lee, R.L. Feinbaum, V. Ambros, the C. elegans\rheterochronic gene lin-4 encodes small RNAs with antisense\rcomplementarity to lin-14., Cell . 75: 843–85 (1993) 843–854.

[89] I. Lev, I.A. Toker, Y. Mor, A. Nitzan, G. Weintraub, O. Antonova, O. Bhonkar, I. Ben Shushan, U. Seroussi, J.M. Claycomb, S. Anava, H. Gingold, R. Zaidel-Bar, O. Rechavi, Germ Granules Govern Small RNA Inheritance, Curr. Biol. (2019) 1–12.

[90] S. Liao, X. Chen, T. Xu, Q. Jin, Z. Xu, D. Xu, X. Zhou, C. Zhu, S. Guang, X. Feng, Antisense ribosomal siRNAs inhibit RNA polymerase I-directed transcription in C. elegans, Nucleic Acids Res. 49 (2021) 9194–9210.

[91] W. Liu, X. Wang, Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data, Genome Biol. 20 (2019) 1–10.

[92] R. Lorenz, S.H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, I.L. Hofacker, ViennaRNA Package 2.0, Algorithms Mol. Biol. 6 (2011) 26.

[93] C.D. Malone, J. Brennecke, M. Dus, A. Stark, W.R. McCombie, R. Sachidanandam, G.J. Hannon, Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary, Cell. (2009).

[94] K.I. Manage, A.K. Rogers, D.C. Wallis, C.J. Uebel, D.C. Anderson, D.A.H. Nguyen, K. Arca, K.C. Brown, R.J. Cordeiro Rodrigues, B.F. de Albuquerque, R.F. Ketting, T.A. Montgomery, C.M. Phillips, A tudor domain protein, SIMR-1, promotes siRNA production at piRNA-targeted mRNAs in C. elegans, Elife. 9 (2020).

[95] J.M. Maniar, A.Z. Fire, EGO-1, a C. elegans RdRP, modulates gene expression via production of mRNA-templated short antisense RNAs, Curr. Biol. 21 (2011) 449–459.

[96] E.A. Marnik, J.H. Fuqua, C.S. Sharp, J.D. Rochester, E.L. Xu, S.E. Holbrook, D.L. Updike, Germline Maintenance Through the Multifaceted Activities of GLH/Vasa in Caenorhabditis elegans P Granules, Genetics. 213 (2019) 923–939.

[97] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet.Journal. 17 (2011) 10.

[98] N.J. Martinez, M.C. Ow, J.S. Reece-Hoyes, M.I. Barrasa, V.R. Ambros, A.J.M. Walhout, Genome-scale spatiotemporal analysis of Caenorhabditis elegans microRNA promoter activity, Genome Res. 18 (2008) 2005–2015.

[99] I.W. Mattaj, L. Englmeier, Nucleocytoplasmic transport: The soluble phase, Annu. Rev. Biochem. 67 (1998) 265–306.

[100] A. Mattout, D. Gaidatzis, J. Padeken, C.D. Schmid, F. Aeschimann, V. Kalck, S.M. Gasser, LSM2-8 and XRN-2 contribute to the silencing of H3K27me3-marked genes through targeted RNA decay, Nat. Cell Biol. 22 (2020) 579–590.

[101] T.J. McEwen, Q. Yao, S. Yun, C.-Y. Lee, K.L. Bennett, Small RNA in situ hybridization in Caenorhabditis elegans, combined with RNA-seq, identifies germline-enriched microRNAs., Dev. Biol. 418 (2016) 248–257.

[102] A.N. McMurchy, P. Stempor, T. Gaarenstroom, B. Wysolmerski, Y. Dong, D. Aussianikava, A. Appert, N. Huang, P. Kolasinska-Zwierz, A. Sapetschnig, E.A. Miska, J. Ahringer, A team of heterochromatin factors collaborates with small RNA pathways to combat repetitive elements and germline stress, Elife. 6 (2017) 560–564.

[103] C. Merritt, G. Seydoux, Transgenic solutions for the germline., 2010.

[104] E.A. Mikhaleva, T.A. Leinsoo, The nucleolar transcriptome regulates Piwi shuttling between the nucleolus and the nucleoplasm, (2018).

[105] A.L. Minogue, M.R. Tackett, E. Atabakhsh, G. Tejada, S. Arur, Functional genomic analysis identifies miRNA repertoire regulating C. elegans oocyte development, Nat. Commun. 9 (2018) 1–11.

[106] B.E. Montgomery, T. Vijayasarathy, T.N. Marks, C.A. Cialek, K.J. Reed, T.A. Montgomery, Dual roles for piRNAs in promoting and preventing gene silencing in C. elegans, Cell Rep. 37 (2021) 110101.

[107] M.J. Moore, T.K.H. Scheel, J.M. Luna, C.Y. Park, J.J. Fak, E. Nishiuchi, C.M. Rice, R.B. Darnell, miRNA–target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity, Nat. Commun. 6 (2015) 8864.

[108] M.A. Newman, F. Ji, S.E.J. Fischer, A. Anselmo, R.I. Sadreyev, G. Ruvkun, The surveillance of pre-mRNA splicing is an early step in C. elegans RNAi of endogenous genes., Genes Dev. (2018) 1–12.

[109] T.C. Nguyen, X. Cao, P. Yu, S. Xiao, J. Lu, F.H. Biase, B. Sridhar, N. Huang, K. Zhang, S. Zhong, Mapping RNA-RNA interactome and RNA structure in vivo by MARIO, Nat. Commun. 7 (2016) 1–12.

[110] M.A. Ortiz, D. Noble, E.P. Sorokin, J. Kimble, A New Dataset of Spermatogenic vs. Oogenic Transcriptomes in the Nematode Caenorhabditis elegans, G3 GENES, GENOMES, Genet. 4 (2014) 1765–1772.

[111] M.A. Ortiz, D. Noble, E.P. Sorokin, J. Kimble, A new dataset of spermatogenic vs. oogenic transcriptomes in the nematode Caenorhabditis elegans., G3 (Bethesda). 4 (2014) 1765–72.

[112] J.P.T. Ouyang, A. Folkmann, L. Bernard, C.-Y. Lee, U. Seroussi, A.G. Charlesworth, J.M. Claycomb, G. Seydoux, P Granules Protect RNA Interference Genes from Silencing by piRNAs, Dev. Cell. 50 (2019) 716-728.e6.

[113] Paul M Sharp and Keith R Bradnam, No Title, 2nd ed., Cold Spring Harbor Laboratory Press, New York, 1997.

[114] M. Percharde, T. Sultana, M. Ramalho-santos, What Doesn ' t Kill You Makes You Stronger : Transposons as Dual Players in Chromatin Regulation and Genomic Variation, 1900232 (2020) 1–10.

[115] S.M. Peterson, J.A. Thompson, M.L. Ufkin, P. Sathyanarayana, L. Liaw, C.B. Congdon, Common features of microRNA target prediction tools, Front. Genet. 5 (2014) 1–10.

[116] M.M. Pfleiderer, W.P. Galej, Structure of the catalytic core of the Integrator complex, Mol. Cell. 81 (2021) 1246-1259.e8.

[117] C.M. Phillips, K.C. Brown, B.E. Montgomery, G. Ruvkun, T.A. Montgomery, PiRNAs and piRNA-Dependent siRNAs Protect Conserved and Essential C. elegans Genes from Misrouting into the RNAi Pathway, Dev. Cell. 34 (2015) 457–465.

[118] C.M. Phillips, T.A. Montgomery, P.C. Breen, G. Ruvkun, MUT-16 promotes formation of perinuclear Mutator foci required for RNA silencing in the C. elegans germline, Genes Dev. 26 (2012) 1433–1444.

[119] J.N. Pitt, J.A. Schisa, J.R. Priess, P granules in the germ cells of Caenorhabditis elegans adults are associated with clusters of nuclear pores and contain RNA., Dev. Biol. 219 (2000) 315–33.

[120] R.N. Platt, M.W. Vandewege, D.A. Ray, Mammalian transposable elements and their impacts on genome evolution, Chromosom. Res. 26 (2018) 25–43.

[121] M. Priyadarshini, J.Z. Ni, A.M. Vargas-Velazquez, S.G. Gu, C. Frøkjær-Jensen, Reprogramming the piRNA pathway for multiplexed and transgenerational gene silencing in C. elegans, Nat. Methods. 5 (2022).

[122] P. Quarato, M. Singh, E. Cornes, B. Li, L. Bourdon, F. Mueller, C. Didier, G. Cecere, Germline inherited small RNAs facilitate the clearance of untranslated maternal mRNAs in C. elegans embryos, Nat. Commun. 12 (2021) 1441.

[123] A.R. Quinlan, I.M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features, Bioinformatics. 26 (2010) 841–842.

[124] R Core Team, R: A Language and Environment for Statistical Computing, (2020).

[125] K.J. Reed, J.M. Svendsen, K.C. Brown, B.E. Montgomery, T.N. Marks, T. Vijayasarathy, D.M. Parker, E.O. Nishimura, D.L. Updike, T.A. Montgomery, Widespread roles for piRNAs and WAGO-class siRNAs in shaping the germline transcriptome of Caenorhabditis elegans, Nucleic Acids Res. 48 (2020) 1811–1827.

[126] M. Reuter, P. Berninger, S. Chuma, H. Shah, M. Hosokawa, C. Funaya, C. Antony, R. Sachidanandam, R.S. Pillai, Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing, Nature. 480 (2011) 264–267.

[127] N. Robine, E.C. Lai, C. Rouget, C. Papin, A. Boureux, Maternal mRNA deadenylation

and decay by the piRNA pathway in the early Drosophila embryo ′, (2010).

[128] M. V. Rockman, S.S. Skrovanek, L. Kruglyak, Selection at Linked Sites Shapes Heritable Phenotypic Variation in C. elegans, Science (80-.). 330 (2010) 372–376.

[129] D.L. Roussell, K.L. Bennett, glh-1, a germ-line putative RNA helicase from Caenorhabditis, has four zinc fingers., Proc. Natl. Acad. Sci. 90 (1993) 9300–9304.

[130] J.G. Ruby, C. Jan, C. Player, M.J. Axtell, W. Lee, C. Nusbaum, H. Ge, D.P. Bartel, Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in C. elegans, Cell. 127 (2006) 1193–1207.

[131] K. Saito, K.M. Nishida, T. Mori, Y. Kawamura, K. Miyoshi, T. Nagami, H. Siomi, M.C. Siomi, Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome, Genes Dev. 20 (2006) 2214–2222.

[132] J.A. Schisa, J.N. Pitt, J.R. Priess, Analysis of RNA associated with P granules in germ cells of C. elegans adults., Development. 128 (2001) 1287–98.

[133] M. Seth, M. Shirayama, W. Gu, T. Ishidate, D. Conte, C.C. Mello, The C. elegans CSR-1 Argonaute Pathway Counteracts Epigenetic Silencing to Promote Germline Gene Expression, Dev. Cell. 27 (2013) 656–663.

[134] M. Seth, M. Shirayama, W. Tang, E.-Z. Shen, S. Tu, H.-C. Lee, Z. Weng, C.C. Mello, The Coding Regions of Germline mRNAs Confer Sensitivity to Argonaute Regulation in C. elegans, Cell Rep. 22 (2018) 2254–2264.

[135] G. Seydoux, The P Granules of C. elegans: A Genetic Model for the Study of RNA–Protein Condensates, J. Mol. Biol. 430 (2018) 4702–4710.

[136] S.A. Shabalina, E. V. Koonin, Origins and evolution of eukaryotic RNA interference, Trends Ecol. Evol. 23 (2008) 578–587.

[137] E.-Z. Shen, H. Chen, A.R. Ozturk, S. Tu, M. Shirayama, W. Tang, Y.-H. Ding, S.-Y. Dai, Z. Weng, C.C. Mello, Identification of piRNA Binding Sites Reveals the Argonaute Regulatory Landscape of the C. elegans Germline, Cell. 172 (2018) 937-951.e18.

[138] U. Sheth, J. Pitt, S. Dennis, J.R. Priess, Perinuclear P granules are the principal sites of mRNA export in adult C. elegans germ cells, Development. 137 (2010) 1305–1314.

[139] Z. Shi, T.A. Montgomery, Y. Qi, G. Ruvkun, High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes, Genome Res. 23 (2013) 497–508.

[140] M. Shigematsu, Y. Kirino, tRNA-derived short non-coding RNA as interacting partners of argonaute proteins, Gene Regul. Syst. Bio. 9 (2015) 27–33.

[141] M. Shirayama, M. Seth, H.-C. Lee, W. Gu, T. Ishidate, D. Conte, C.C. Mello, piRNAs

Initiate an Epigenetic Memory of Nonself RNA in the C. elegans Germline, Cell. 150 (2012) 65–77.

[142] A. Shukla, J. Yan, D.J. Pagano, A.E. Dodson, Y. Fei, J. Gorham, J.G. Seidman, M. Wickens, S. Kennedy, poly(UG)-tailed RNAs in genome protection and epigenetic inheritance, Nature. 582 (2020) 283–288.

[143] T. Sijen, J. Fleenor, F. Simmer, K.L. Thijssen, S. Parrish, L. Timmons, R.H.A. Plasterk, A. Fire, On the Role of RNA Amplification in dsRNA-Triggered Gene Silencing, Cell. 107 (2001) 465–476.

[144] M. Singh, E. Cornes, B. Li, P. Quarato, L. Bourdon, F. Dingli, D. Loew, S. Proccacia, G. Cecere, Translation and codon usage regulate Argonaute slicer activity to trigger small RNA biogenesis, Nat. Commun. 12 (2021) 3492.

[145] J.R. Skaar, A.L. Ferris, X. Wu, A. Saraf, K.K. Khanna, L. Florens, M.P. Washburn, S.H. Hughes, M. Pagano, The Integrator complex controls the termination of transcription at diverse classes of gene targets, Cell Res. 25 (2015) 288–305.

[146] J. Smith, D. Calidas, H. Schmidt, T. Lu, D. Rasoloson, G. Seydoux, Spatial patterning of P granules by RNA-induced phase separation of the intrinsically-disordered protein MEG-3, Elife. 5 (2016) 1–18.

[147] C.A. Spike, J. Bader, V. Reinke, S. Strome, DEPS-1 promotes P-granule assembly and RNA interference in C. elegans germ cells, Development. 135 (2008) 983–993.

[148] B. Stadelmayer, G. Micas, A. Gamot, P. Martin, N. Malirat, S. Koval, R. Raffel, B. Sobhian, D. Severac, S. Rialle, H. Parrinello, O. Cuvier, M. Benkirane, Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes, Nat. Commun. 5 (2014).

[149] S. Strome, W.B. Wood, Immunofluorescence visualization of germ-line-specific cytoplasmic granules in embryos, larvae, and adults of Caenorhabditis elegans., Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 1558–62.

[150] K.M. Suen, F. Braukmann, R. Butler, D. Bensaddek, A. Akay, C. Lin, N. Doshi, A. Sapetschnig, A. Lamond, J.E. Ladbury, E.A. Miska, DEPS-1 is required for piRNA-dependent silencing and PIWI condensate organisation in Caenorhabditis elegans, (2020).

[151] W. Tang, M. Seth, S. Tu, E.-Z. Shen, Q. Li, M. Shirayama, Z. Weng, C.C. Mello, A Sex Chromosome piRNA Promotes Robust Dosage Compensation and Sex Determination in C. elegans, Dev. Cell. 44 (2018) 762-770.e3.

[152] W. Tang, S. Tu, H.-C. Lee, Z. Weng, C.C. Mello, The RNase PARN-1 Trims piRNA 3′ Ends to Promote Transcriptome Surveillance in C. elegans, Cell. 164 (2016) 974–984.

[153] D.C. Tatomer, N.D. Elrod, D. Liang, M. Xiao, J.Z. Jiang, M. Jonathan, K. Huang, E.J. Wagner, S. Cherry, J.E. Wilusz, The Integrator complex cleaves nascent mRNAs to

attenuate transcription, Genes Dev. 33 (2019) 1525–1538.

[154] J. Thomas, K. Lea, E. Zucker-aprison, T. Blumenthal, The spliceosomal snRNAs of Caenorhabditis elegans, 18 (1990) 2633–2642.

[155] J.A. Toombs, Y.A. Sytnikova, G. Chirn, I. Ang, N.C. Lau, M.D. Blower, Xenopus Piwi proteins interact with a broad proportion of the oocyte transcriptome, RNA. 23 (2017) 504–520.

[156] A.J. Travis, J. Moody, A. Helwak, D. Tollervey, G. Kudla, Hyb: A bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data, Methods. 65 (2014) 263–273.

[157] Y.C. Tse, A. Piekny, M. Glotzer, Anillin promotes astral microtubule-directed cortical myosin polarization, Mol. Biol. Cell. 22 (2011) 3165–3175.

[158] K.M. Tyc, A. Nabih, M.Z. Wu, C.J. Wedeles, J.A. Sobotka, J.M. Claycomb, The Conserved Intron Binding Protein EMB-4 Plays Differential Roles in Germline Small RNA Pathways of C. elegans, Dev. Cell. 42 (2017) 256-270.e6.

[159] C.J. Uebel, D. Agbede, D.C. Wallis, C.M. Phillips, Mutator Foci Are Regulated by Developmental Stage, RNA, and the Germline Cell Cycle in Caenorhabditis elegans, G3 Genes|Genomes|Genetics. 10 (2020) 3719–3728.

[160] C.J. Uebel, D.C. Anderson, L.M. Mandarino, K.I. Manage, S. Aynaszyan, C.M. Phillips, Distinct regions of the intrinsically disordered protein MUT-16 mediate assembly of a small RNA amplification complex and promote phase separation of Mutator foci, (2018) 1–22.

[161] J. Ule, K.B. Jensen, M. Ruggiu, A. Mele, A. Ule, R.B. Darnell, CLIP Identifies Nova-Regulated RNA Networks in the Brain, Science (80-.). 302 (2003) 1212–1215.

[162] D. Updike, S. Strome, P Granule Assembly and Function in Caenorhabditis elegans Germ Cells, J. Androl. 31 (2010) 53–60.

[163] D.L. Updike, S. Strome, A genomewide RNAi screen for genes that affect the stability, distribution and function of P granules in Caenorhabditis elegans, Genetics. 183 (2009) 1397–1419.

[164] J.M. Vaquerizas, R. Suyama, J. Kind, K. Miura, N.M. Luscombe, A. Akhtar, Nuclear pore proteins Nup153 and megator define transcriptionally active regions in the Drosophila genome, PLoS Genet. 6 (2010).

[165] P. Videm, A. Kumar, O. Zharkov, B.A. Grüning, R. Backofen, ChiRA: An integrated framework for chimeric read analysis from RNA-RNA interactome and RNA structurome data, Gigascience. 10 (2021) 1–13.

[166] U. Vijayraghavan, M. Company, J. Abelson, Isolation and characterization of pre-mRNA

splicing mutants of Saccharomyces cerevisiae., Genes Dev. 3 (1989) 1206–1216.

[167] A. Vourekas, P. Alexiou, N. Vrettos, M. Maragkakis, Z. Mourelatos, Sequence-dependent but not sequence-specific piRNA adhesion traps mRNAs to the germ plasm, Nature. 531 (2016) 390–394.

[168] S. Waaijers, V. Portegijs, J. Kerver, B.B.L.G. Lemmens, M. Tijsterman, S. van den Heuvel, M. Boxem, CRISPR/Cas9-Targeted Mutagenesis in Caenorhabditis elegans, Genetics. 195 (2013) 1187–1191.

[169] G. Wan, B.D. Fields, G. Spracklin, A. Shukla, C.M. Phillips, S. Kennedy, Spatiotemporal regulation of liquid-like condensates in epigenetic inheritance, Nature. (2018).

[170] C.J. Wedeles, M.Z. Wu, J.M. Claycomb, Protection of germline gene expression by the C. elegans argonaute CSR-1, Dev. Cell. 27 (2013) 664–671.

[171] E.-M. Weick, E.A. Miska, A.S. Team, piRNAs: from biogenesis to function., Development. 141 (2014) 3458–71.

[172] C. Weng, A. Kosalka, A.C. Berkyurek, P. Stempor, X. Feng, H. Mao, C. Zeng, W.-J. Li, Y.-H. Yan, M.-Q. Dong, C. Zuliani, O. Barabas, J. Ahringer, S. Guang, E. Miska, The USTC complex co-opts an ancient machinery to drive piRNA transcription in C. elegans, BioRxiv. (2018).

[173] H. Wickham, Reshaping Data with the {reshape} Package, J. Stat. Softw. 21 (2007) 1--20.

[174] H. Wickham, ggplot2: Elegant Graphics for Data AnalysisTitle, Springer-Verlag New York, 2016.

[175] H. Wickham, R. François, L. Henry, K. Müller, dplyr: A Grammar of Data Manipulation, (2021).

[176] W.-S. Wu, J.S. Brown, P.-H. Chen, S.-C. Shiue, D.-E. Lee, H.-C. Lee, CLASH Analyst: A Web Server to Identify In Vivo RNA–RNA Interactions from CLASH Data, Non-Coding RNA. 8 (2022) 6.

[177] W.-S. Wu, J.S. Brown, T.-T. Chen, Y.-H. Chu, W.-C. Huang, S. Tu, H.-C. Lee, piRTarBase: a database of piRNA targeting sites and their roles in gene regulation, Nucleic Acids Res. 47 (2019) D181–D187.

[178] W. Wu, W.-C. Huang, J.S. Brown, D. Zhang, X. Song, H. Chen, S. Tu, Z. Weng, H.-C. Lee, pirScan: a webserver to predict piRNA targeting sites and to avoid transgene silencing in C. elegans, Nucleic Acids Res. 46 (2018) W43–W48.

[179] C. Xia, J. Fan, G. Emanuel, J. Hao, X. Zhuang, Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression, Proc. Natl. Acad. Sci. U. S. A. 116 (2019) 19490–19499.

[180] J. Xiol, P. Spinelli, M.A. Laussmann, D. Homolka, Z. Yang, E. Cora, Y. Couté, S. Conn, J. Kadlec, R. Sachidanandam, M. Kaksonen, S. Cusack, A. Ephrussi, R.S. Pillai, RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts, Cell. 157 (2014) 1698–1711.

[181] F. Xu, X. Feng, X. Chen, C. Weng, Q. Yan, T. Xu, M. Hong, S. Guang, A Cytoplasmic Argonaute Protein Promotes the Inheritance of RNAi, Cell Rep. 23 (2018) 2482–2494.

[182] R. Yashiro, Y. Murota, K.M. Nishida, L. Negishi, H. Siomi, M.C. Siomi, R. Yashiro, Y. Murota, K.M. Nishida, H. Yamashiro, K. Fujii, A. Ogai, Piwi Nuclear Localization and Its Regulatory Mechanism in Drosophila Ovarian Somatic Cells, CellReports. 23 (2018) 3647–3657.

[183] E. Yigit, P.J. Batista, Y. Bei, K.M. Pang, C.C.G. Chen, N.H. Tolia, L. Joshua-Tor, S. Mitani, M.J. Simard, C.C. Mello, Analysis of the C. elegans Argonaute Family Reveals that Distinct Argonautes Act Sequentially during RNAi, Cell. 127 (2006) 747–757.

[184] D. Zhang, M. Glotzer, The RhoGAP activity of CYK-4/MgcRacGAP functions non-canonically by promoting RhoA activation during cytokinesis, Elife. 4 (2015) 1–25.

[185] D. Zhang, S. Tu, M. Stubna, W.-S. Wu, W.-C. Huang, Z. Weng, H.-C. Lee, The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes, Science (80-.). 359 (2018) 587–592.

[186] P. Zhang, J.Y. Kang, L.T. Gou, J. Wang, Y. Xue, G. Skogerboe, P. Dai, D.W. Huang, R. Chen, X.D. Fu, M.F. Liu, S. He, MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes, Cell Res. 25 (2015) 193–207.

[187] P. Zhang, X. Si, G. Skogerbø, J. Wang, D. Cui, Y. Li, X. Sun, L. Liu, B. Sun, R. Chen, S. He, D.-W. Huang, piRBase: a web resource assisting piRNA functional study, Database. 2014 (2014).

[188] C. Zhong, S. Zhang, Accurate and Efficient Mapping of the Cross-Linked microRNA-mRNA Duplex Reads, IScience. 18 (2019) 11–19.