Title
Investigating the biosynthesis and function of newly identified intronic small interference RNA in Caenorhabditis elegans

Author
Duong, Trang Huyen

Publication Date
2022

Peer reviewed|Thesis/dissertation

Investigating the biosynthesis and function of newly identified intronic small interference RNA in Caenorhabditis elegans

by
Trang Huyen Duong

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Oral and Craniofacial Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Dengke Ma
0A5FF211220849A...

Dengke Ma
Chair

DocuSigned by:
Noelle L'Etoile

Noelle L'Etoile

DocuSigned by:
K. Mark Ansel

K. Mark Ansel

DocuSigned by:
Hani Goodarzi
FDD44359FCC6487...

Hani Goodarzi

Committee Members

ACKNOWLEDGEMENTS

Eight years flies by: I have been remiss to not exploit every opportunity to thank my principal investigator, Dr. Noelle L'Etoile. Your invaluable mentorship, dedication to science, and joie de vivre has given me the freedom to pursue this project wholeheartedly. You have been both the backstop and chief inquisitor to every nuance and assumption in my work: the end-result is better because of your guidance. Thank you.

I'm indebted to my thesis committee, Drs. Sarah Knox, Mark Ansel, Dengke Ma, Noah Zailent and Hani Goodarzi. Imagine the mire I would be in without your guidance sharpening my focus. I'm equally grateful that you deployed your expertise in humility, which kept me from drowning and emboldened me to persevere.

My parents, Hai Thu Duong and Binh Pham, entrusted me with this life of mine, and gave me their permission to chase it and love it. I would not have tasted the joy of this journey without your love: thank you for believing.

Many thanks to my partner, Florian Muller, and my son, Traian Muller, for perfecting the longsuffering role of supporting me through this project. Hours, days, weeks and months away cannot be redeemed, yet you never wavered in your support.

Thanks also are due to my accomplices, my lab-mates, my cohort, and friends. Whether sharing each other's load in the lab, classwork or coming up for air over too many beers, your camaraderie is priceless. For you I am grateful.

In that vein, Dr. Michael Le, Troy Brant, and Nick Yen were indispensable allies in creating and fine-tuning my pipeline. And, when doubt crept in, Jeff Wang and Ziad Ahmed validated the data, instilling high confidence that I was on the right track.

I am grateful for the hospitable shelter of my mentor and friend, Dr. Daniel Ramos. On many morning runs, you absorbed my rants, rejoiced in my victories, and threshed chaff from grain. Your line of sight was invaluable in clarifying the potential of my work, which freed me to think unconventionally about the implications of this project.

Lastly, nothing is achievable without a supporting cast: My program director, Dr. Pamela Den Besten, co-director, Dr. Ralph Marcucio, and research coordinator, Roger Mraz, were always on point, inciteful and supportive. “We can do it” defines their drive. Thank you for the catalyst.

ABSTRACT

Investigating the biosynthesis and function of newly identified

intronic small interference RNA in *Caenorhabditis elegans*

By Trang Huyen Duong

Small interfering RNA (siRNA) are critical regulators of gene expression in many animals. In the model organism *C. elegans*, our lab found that disruption of the siRNA machinery results in an inability to sense and adapt to specific odorants. Specifically, we found that mutants defective for the RNase III nuclease DCR-1, the double-stranded RNA binding protein RDE-4, the RNA-dependent RNA polymerase RRF-3, or the nuclear RNAi Argonaute NRDE-3 are defective in olfactory adaptation to the odorant, butanone. The lab performed extensive small RNA sequencing experiments to better understand the role of siRNA in olfactory adaptation.

By performing an in-depth survey of these siRNA libraries, we identified a class of small RNA (average size, 21 nucleotides) that mapped to intronic regions of more than 30% of the genes in the *C. elegans* genome. These small RNAs, that map to introns, were mentioned in the Craig Mello lab's 2009 publication [1] but were largely ignored by the field. We term these understudied small RNA–that map to intronic regions–"isiRNA", and have performed extensive bioinformatic analysis to understand their origin and function.

The major findings we reach from this work are:

1. Two separate bioinformatic pipelines confirm the presence of isiRNA in multiple independent, public-domain small RNA seq *C. elegans* datasets.
2. isiRNA map non-randomly to longer introns.
3. A diverse set of common siRNA factors are required for accumulation of both intronic siRNA (isiRNA) and exonic siRNA (esiRNA) in the germline.
4. isiRNA are likely produced via the WAGOs pathway.

5. isiRNA may be amplified by the mutator complex which is at the periphery of P granules.
6. isiRNA levels are dependent on many exo-RNAi factors.
7. isiRNA is independent from the Enhanced for Exogenous RNAi (ERI) endogenous RNAi biosynthesis pathway.
8. isiRNA are tertiary RNA.
9. isiRNA map to genes that have significantly more alternative splice variants.
10. isiRNA binds to the germline specific Argonaute, HRDE-1, in order to repress transcription.
11. HRDE-1 and CSR-1 may compete for isiRNA.
12. isiRNA binds to Argonaute CSR-1 to promote the production of siRNA.
13. Cold shock and exogenous RNAi promote the binding of esiRNA and isiRNA to NRDE-3.
14. Lack of ERGO-1 increased the number of genes with higher isiRNA than esiRNA reads.
15. WAGO-1 is required for biosynthesis of germline specific isiRNA.

Together, these data reveal a previously uncharacterized population of small RNAs with potentially critical functions in diverse aspects of isiRNA maturation and regulation of gene expression. Future wet-lab experiments will be required to probe the biology of this intriguing class of small RNAs.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter two:

Chapter three:

Chapter four:

Chapter five:

Chapter six:

LIST OF TABLES

Chapter four

CHAPTER ONE: INTRODUCTION

Gene expression provides a cell with its identity and allows it to respond characteristically to changes in its environment in order to maintain homeostasis [1,2]. In order to maintain cell identity and homeostasis, gene expression needs to be precisely regulated. Disease can result from loss of gene regulation. For example, loss of the gene expression that controls cellular identity is linked to neoplastic transformation [3,4]; while in aging, changes in gene expression render cells less able to maintain homeostasis [5,6]. Therefore, it is important to understand the many processes by which cells regulate their gene expression. One way that a variety of organisms–such as animals, plants, fungi, and some bacteria–control gene expression is by employing small regulatory RNAs.

In the last century, gene expression was found to be regulated by transcription factors [7,8]. More recently, transcription factors were shown to affect chromatin structure and how genes are packaged. Most recently, it was discovered that RNA feedback can also regulate genes. Specifically, the genes that produce RNA to make protein were also shown to produce small regulatory RNAs that feed back to regulate their own gene's expression [9]. It is critical that we understand this RNA-based regulation of gene expression, so we decided to study small RNAs that map to intronic regions within genes, as this is a species of small RNA that has been overlooked. Indeed, this species may play important gene regulatory roles in health and disease states.

Overall, small regulatory RNAs are about 21-23 nucleotides in length that often are antisense to mRNA. Gene silencing by these small regulatory RNAs can be achieved through two main processes: Post-Transcriptional Gene Silencing (PTGS) and Transcriptional Gene Silencing (TGS). While PTGS is the process of small regulatory RNA targeting mature mRNA in the cytoplasm, the TGS is the process of small regulatory RNA targeting nascent strands of mRNA at the transcriptional site in the nucleus.

Regulatory small RNAs include: micro RNA (miRNA), Piwi-interacting RNA (piRNA), and small interference RNAs (siRNA). MicroRNAs are encoded within the genome as Polymerase II

transcripts that fold back into hairpin double stranded RNA which are processed into ~22 nucleotide noncoding RNA species. These function post-transcriptionally to repress gene expression by interacting with regions in the target gene's 3'UTR that have "seed" sequences with some homology to the miRNA [10]. Piwi-associated RNAs (piRNAs) are 21 nucleotide RNA species that are germline-specific and are encoded in the genome, transcribed by Pol II. They are required for silencing transposable elements and maintaining genome integrity [11]. siRNAs are 22-nucleotide long, begin with a G and are perfectly complementary to mRNA. There are two types of siRNA: exogenous RNAi and endogenous RNAi. Exogenous RNAi is exogenously triggered RNAi elicited by dsRNA supplied from outside the *C. elegans* body and taken up by the intestinal cells. The small interfering RNA species are amplified and transported through the animal where they are taken up by the double stranded RNA import channel, SID-1 [12]. Endogenous RNAi (endo-siRNAs) is triggered by transcripts that are expressed from the genome. Approximately two thirds of the genes within the *C. elegans* genome have endo-siRNA species mapping to them. These siRNAs come from transcription of both coding and noncoding genomic sequences [13]. In this project, I am focusing on the biosynthesis and function of a subset of siRNA that map to the intronic regions of coding genes: We termed these isiRNA.

Most exogenous siRNA processing involves *Dicer*, an exonuclease III protein, while the majority of endogenous siRNA can bypass *Dicer*, instead deriving from short RNA-dependent RNA polymerase transcript [14]. The exogenous double-stranded precursors of siRNA are shortened by one of the exonuclease III protein *Dicer* and other associated factors, such as the RDEs for exo-siRNA or the ERIs for endo-siRNA [15]. The majority of endogenous siRNA tend to bypass *Dicer*, instead deriving from short RNA-dependent RNA polymerase transcript [14].

There are three classes of siRNA: primary siRNA, secondary siRNA and tertiary siRNA. Primary siRNA is the product of double-stranded precursors of siRNA processed by one of the *Dicer* complexes into 26 nucleotides. These primary siRNAs then can be cut to 21-22 nucleotides and separated to single-stranded RNA. The antisense strand that can target mRNA

are termed secondary siRNA. Tertiary siRNA are produced upon triggering of secondary siRNA: This process thought to be initiated in the nucleus [16–18].

There are a few pathways to make the variety of secondary siRNA. The distinction between these different pathways is based on the Argonaute it binds to: For example, there are WAGOs, CSR-1, ERGO-1 and ALG-3/4 pathways [14]. In each pathway, there are additional Argonaute cofactors with overlapping roles to generate siRNA. For example, RdRP RRF-1 and EGO-1 are required for generating 22G in WAGOs pathways [13], while only EGO-1 is required for generating 22G in CSR-1 pathways [19], while only RRF-3 is required for 26 G biosynthesis in both ERGO-1 and ALG-3/4 pathways [20,21].

Unlike miRNA, the siRNA can act in either the cytoplasm or nucleus. In the cytoplasm, the combination of the siRNA, argonauts, and other proteins (RNA induced silencing complexes, or RISC for short). siRNA directs RISC to bind a specific mRNA, the targeting is precise because it is determined by a perfect complementary binding between the siRNA and the targeted mRNA. Once RISC is bound to the targeted mRNA it catalyzes the cleavage of mRNA, which is then degraded, and more siRNA is produced. In the nucleus, the siRNA is required for chromatin modification [22]. The secondary siRNAs bind to NRDE-3 (a somatic Argonaute) or HRDE-1 (a germline specific Argonaute) to be shuttled from the cytoplasm to the nucleus [22,23] where they associate with NRDE-1, NRDE-2, NRDE-4 to form NRDEs-siRNA complex. This NRDEs-siRNA complex is located in the mRNA nascent strand and this triggers the repressive histone 3 lysine 9 trimethylation complex (H3K9me3) loaded on targeted locus [23].

While siRNA is referred to as 22G anti-sense small RNA – it regulates gene expression by targeting exon regions of mRNA in the cytoplasm or the H3K9 complex at the DNA level in the nucleus– there is another subset of small RNA that maps anti-sense to the intronic regions of transcribed mRNA: this has not been studied or understood before. Believing that the understanding of these newly identified isiRNA is important in developing another way to

manipulate gene expression, my thesis project is focused on using bioinformatics to analyze a large volume of sequencing data to understand the biosynthesis and function of these small interference RNA mapped to intronic regions of *C. elegans* genome (isiRNA). Through analysis of hundred of small RNA libraries across dozens of laboratories, I discovered that isiRNA similarly mapped in libraries with the same genetic background. Characterizing the introns with isiRNA mapped to them, we saw that these isiRNA introns are longer compared to introns without isiRNA mapped to them. While the long introns in other species, such as pombe, had poorer information content at the splicing sites compared to the shorter introns, this is not the case with isiRNA intron in *C. elegans*. Rather characterizing the isiRNA genes, I discovered that genes with alternative splicing are more likely to produce isiRNA, especially when compared to that of exonic siRNA.

Next-generation sequencing technology has revolutionized genomic research and discovery by improving throughput while simultaneously cutting cost. Since the first whole-genome sequence of *C. elegans* was completed 2 decades ago we have come very far in improving this technology. For example, from Ilumina, MiSeq, NextSeq, and other short sequence instruments that can do short sequencing of 10 up to 600 nucleotides [24], PacBio and Single-Molecule Real-Time (SMRT) sequencing achieve long read sequencing of single RNA or DNA molecules that often reach up to 10,000 nucleotides [25]. Another advancement of genome sequencing is the transition from whole-tissue sequencing, requiring millions of cells, to single cell sequencing [26].

The sequencing libraries included in this study are *C. elegans* whole-body or embryo sequencing. The adult *C. elegans* body has about 3000 cells, 1000 of which are somatic cells, and 1000-2000 are germline cells. Embryonic is the stage where the germline is not yet developed, therefore there are only about 1000 cells in each embryo. The whole-body sequencing requires that all tissue was homogenized into one mixture. As a result, tissue-specific identification of small RNA reads is very limited. To delineate this, tissue-specific isiRNA

were identified through isiRNA biosynthesis factors or associating Argonautes. Generally, miRNA sequencing is very similar to siRNA sequencing, from sample preparation to library extraction to configuration sequencer platform. The significant difference between miRNA and siRNA library preparation is miRNA has monophosphate at 5' end while siRNA has triphosphate at 5' end: The ligase only recognizes monophosphate. Therefore, in order to capture siRNA it is necessary to produce a monophosphate 5' end: Either by dephosphorylating the triphosphate with Calf Intestinal Alkaline Phosphatase (CIAP) and then phosphorylating to a monophosphate of these short fragments with T4 Polynucleotide Kinase (T4 PNK); or dephosphorylating the triphosphate and diphosphate with Tobacco Acid Pyrophosphatase (TAP) to make monophosphate prior to the ligase step. Notably, all siRNA libraries need to go through CIAP and T4 PNK steps in the preparation protocol, while miRNA does not require this processing. Because of this difference, we can capture miRNA in siRNA libraries, but we can't capture siRNA in miRNA libraries without CIAP and T4 PNK treatment. All of the libraries included in this study have CIAP and T4 PNK or TAP treatment in their preparation protocol.

I describe the workflow of two independent pipelines that we created to map small RNA sequencing libraries. I also describe the deviation in our results with a detailed explanation of these differences. In this project, all of the sequencing datasets are small RNA sequencing data, or Co-Immunoprecipitation (CoIP). These are short-read sequences mostly done by ilumina or MiSeq sequencing instruments from *C. elegans*. These datasets were acquired as unmanipulated sequences published on Gene Expression Omnibus (GEO) from various studies. In 2014, we started to put together a workflow that included different published software such as: FastX, Bowtie 2, HTSeq and other short scripts in python, Java, and R in processing these small RNA sequencing datasets that we included in our study. Bioinformatics, however, is not one of the strengths in our laboratory: We needed an independent eye to help us validate our output and analysis. So, in early 2020, we started our collaboration with the lab of Dr. Goodazi, where processing and analyzing sequencing data is one of their expertise. Jeff Wang and Ziad

Amed are two software engineers who collaborated with me under the guidance of Dr. Goodazi and Dr. L'Etoile in processing all the sequencing datasets included in my study: They performed an independent analysis using the same reference genome and software. The output from our two analyses were compared side by side for the number of reads that each library identified as mapping to intronic and exonic regions. We found that >95% of reads that mapped to intronic regions were found in both pipelines. The 5% of reads that differed between the pipelines were a result of differences in filtering out all other known small RNA and characterizing reads that mapped to the intron-exon junction. Output from both pipelines identified that, in hundred of small RNA libraries from dozens of experiments, isiRNA mapped to approximately 1/3 of the genes in the C. elegans genome. Indeed, we consistently found that isiRNAs account for ~7%-10% of the total siRNA reads in each small RNA library. This reproducibility across libraries and analysis pipelines prompted us to dive further into exploring the properties of this class of understudied small RNA and the introns they map to.

We characterized the introns that these small RNAs map to and found that they map predominantly to introns that are longer than the average intron and may arise more frequently from genes that have alternatively spliced introns. The picture that emerges for these observations is that small RNA that maps antisense to introns is a species of small RNA that might arise from difficult-to-splice, longer transcripts. They may have been selected to limit expression of genes that are prone to expressing unspliced, possibly non-sense coding proteotoxic gene products. Equally likely, this small RNA species may have been selected to limit expression of transposon, or selfish-DNA encoded factors such as transposases, that would wreak havoc with the host genome. We propose that in future studies, combining our pipelines to analyze small RNAs and using long read sequencing to analyze mRNA from the same samples, we would find the isiRNAs mapping predominantly to the unspliced, retained introns.

In chapter 4 of this thesis, I focus on describing my findings on the biosynthesis machinery of isiRNA. Since the first description of inhibition of gene expression by antisense RNA in the early 1980s [26] and the discovery of exogenous and endogenous siRNA in the late 1990s, we have a fairly good understanding of the biosynthesis pathways of those classes of siRNA. However, the biosynthesis of siRNA that targets the intronic regions, the isiRNA, remains unknown. In this study, in order to understand the biosynthesis of isiRNA and thus potential triggers for its production, I looked for mutant backgrounds that disrupt or increase isiRNA levels in publicly deposited data sets (table 1). This candidate screen ocused on those genes that had been shown previously to affect production of endogenous siRNA (endo-siRNA) and exogenous siRNA (exo-siRNA). By doing so, I hoped to uncover the pathways that regulate the less studied isiRNA production. Understanding the genetic factors that are required for biosynthesis of isiRNAs may provide insight into both the function of these small RNAs and what triggers their production.

CHAPTER TWO: WORKFLOW AND PIPELINES

INTRODUCTION

Genome sequence screening has become a necessity in elucidating gene function. Although there are many available package pipelines that allow scientists to conveniently process the sequencing data, these are designed to handle smaller datasets on the local machine or web-based processing [27–29]. Nevertheless, this small-scale alignment-based analysis can still be memory and time consuming on a local computer. For example, an average *C. elegans* sequence library is about 5 Gigabyte: To process this dataset from pre-processing to aligning to post-processing could take up to an hour-to-two hours of machine time. Nowadays, with the increasing publicly available sequencing data, the ability to process large-scale sequences has become crucial in order to sufficiently explore the genome functions.

For a project that needs to process less than 20 sequencing libraries, it is reasonable financially to outsource the sequence processing to a core facility. However, in light of hundreds of sequencing libraries recently becoming available, outsourcing is neither feasible financially nor would it provide the processing details that we need. Therefore, I developed a pipeline that was highly customized to answer our very specific biology questions. This pipeline utilized Python as the main programming language, first to incorporate the existing tools (fastX, Bowtie2, HTseq2, bedtools) to process the small RNA sequencing data; and second, to facilitate our specific needs in exploring the data. My pipeline was validated by comparing the output with that of another, independently built pipeline based on the same criteria. This pipeline has been a useful tool that allowed me to explore hundreds of small RNA libraries across different experiments, and through this, I have gained a better understanding of isiRNA biosynthetic machinery and its possible functions within cells.

SUMMARY

In 2013, our laboratory discovered that the endo-siRNA pathway promotes odor adaptation in *C. elegans* AWC olfactory neurons [30]. Specifically, we found that disruption of siRNA biosynthetic machinery caused odor adaptation defects in worms. In order to understand whether odor adaptation is required for specific species of siRNA, we built a pipeline to analyze the small RNA libraries made from siRNA factor mutants that are available on Gene Expression Omnibus (GEO). In analyzing these datasets, to our surprise, we observed that there is a population of siRNA mapped antisense to the intronic region of nearly half of the coding genes. This small RNA had been previously described in *C. elegans* [13], but its biosynthesis and functions have not been studied nor understood. In this study, I am focusing on analyzing a large quantity of small RNA sequencing data in order to find out how this small RNA is made, and what are the functions it might be serving in the cells. To validate the output of my pipeline, we collaborated with Jeff Wang and Ziad Ahmed in Dr. Goodarzi's lab to reprocess all of the libraries I analyzed with an independent pipeline based on the same criteria.

METHODS

Mapping workflow: Trang's and Jeff-Ziad's pipelines

In this chapter, I describe the two pipelines side by side and then compare the results. In figure 2.1, both pipelines were structured using a fastX toolkit developed in Hannon's lab to do preprocessing steps: clip the adapter and trim the barcode, Bowtie2 for alignment, and bedtools for intersect. The same reference genome of choice, WS253, was used for both pipelines (WS253 was the latest published version when this project started in 2015). Then the alignment results were classified and filtered out the unrelated small RNA (Figure 2.1). All scripts and descriptions of the two pipelines are deposited and available on github.

1. Download WS253 genome in ".fa" format , gff2, gff3, gtf files from Wormbase. The ".fa" file is the raw sequences file, the gff2, gff3, and gtf are annotation files generated by Wormbase.
2. All datasets of interest from GEO are downloaded and converted to ".fa" format. For each data set (or library), I identified the adapter sequences and clipped them off using FastXclipper with the following parameters: -f, -Q33, -i, -o. There is one difference between my pipeline and Jeff and Ziad's pipeline: I only kept reads with adapters to align with the reference genome, while Jeff and Ziad kept all the reads with and without an adapter.
3. The alignment step was done using Bowtie2 software. While I aligned both genome and exome to the reference genome for classification and filtering purposes, Jeff and Ziad aligned only the genome and used annotation to classify and filter the reads. All alignments were done with the following parameters: -f, -v 0, --all, --best, --strata (bowtie-2 manual: http://bowtie-bio.sourceforge.net/manual.shtml). In this step, we set the mismatch = 0 (-v 0). Quality control was done during the alignment using --all, --best, --strata parameters.
4. Classification:

Trang: In order to classify all the antisense reads of interest (intron, intron-exon, exon, and exon-exon) I used the genome annotation to identify intron, exon and exon-intron junction, then I used “blanket” approach to identify reads that mapped to exome-exome junction in mature mRNA, which are not annotated in the reference file (“blanket” = reads mapped to genome - reads mapped to exome). Here below are the four categories of antisense siRNA I classified in my pipeline:

- Antisense intron = reads mapped to genome annotated intron.
- Antisense intron-exon = reads mapped to the splicing junction between intron and exon.
- Antisense ixon = reads mapped to genome annotated exon.
- Antisense exon-exon = reads mapped to exon-exon junction identified using exome alignments.

Jeff and Ziad: used bedtools to intersect the alignments with the annotation file to classify the reads to four categories:

- Antisense intron = reads mapped to genome annotated intron + reads mapped to intron-exon junctions.
- Antisense exon = reads mapped to genome annotated exon.
- Sense exon = reads mapped to genome annotated exon.
- Known = nRNA, tRNA, miRNA, piRNA, snoRNA, rRNA, transposable element, pseudogenic tRNA, pseudogenic rRNA, miRNA primary transcript.

1. Filtering: I using a bedtools program with the following parameters: -wo, -S, -a, stdin, -b (https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html) to intersect the alignment with annotation files to identify and eliminate piRNA, miRNA, snRNA, snoRNA, tRNA, and transposon sequences. I also bioinformatically removed all the reads that were mapped to multiple locations (repetitive region), and reads that were

mapped to multiple genes (some genes are overlapping). Jeff-Ziad's used bedtools to intersect, classify and filter together

2. Output analysis: DEseq2, IGV, heatmaps

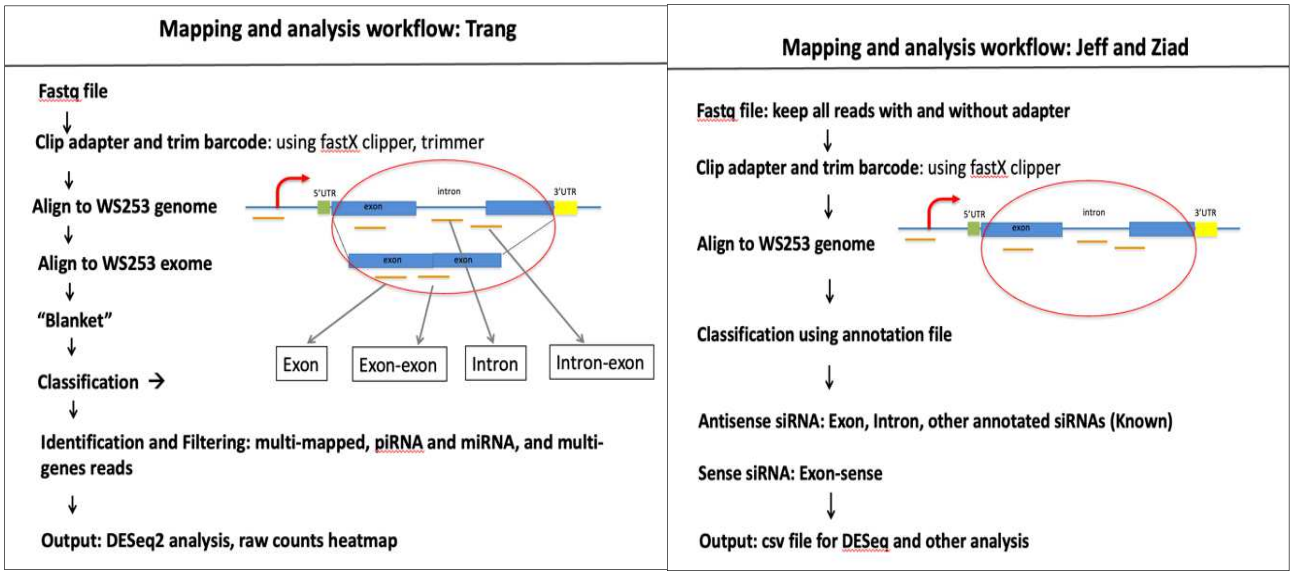


Figure 2.1 Mapping and analysis workflow. Left) Trang's workflow, Right) Jeff and Ziad's workflow. There are two main differences between Trang's workflow from Jeff and Ziad;s workflow are: First, Trang kept only the reads with the adapter, while Jeff kept all the reads with and without the adapter. Second, Trang took all the reads that mapped to the genome minus or "blanket" the reads that mapped to the exome in order to classify the reads to four categories as in boxes, while Jeff and Ziad classified the reads using the WS253 annotation file.

RESULTS

Output different between two pipelines

In this section, I describe the output difference of one library that was processed by both pipelines. This is a wildtype library that was included in Gent et al., 2010 study [31].

In this figure, I overlapped the N2 library output from both pipelines.

The red circle is the “ambiguous” regions, which mean it is annotated as intron of one gene (bottom lines) and as exon in another (red circle). Therefore, the small RNA reads that aligned antisense to this region could be annotated as both intron and exon. While my pipeline preserves those reads, Jeff-Ziad’s discards them.

For this library, above 90% of isiRNA genes are found in the output of both pipelines. There are 679 isiRNA genes in Jeff-Ziad’s output that are not found in my output, while there are 196 isiRNA genes in my output that were not found in Jeff-Ziad’s output (Figure 2.2). There are a few possibilities for the discrepancy between two outputs: First, in the preprocessing step, Jeff and Ziad kept all the reads with and without the adapter, while I only kept the reads with the adapter. Second, I classified the reads that are annotated as both intron and exon to the intron category, while Jeff-Ziad eliminated those reads if they also intersected with the sense exon annotation (Figure 2.2).

Overall, we analyzed 184 libraries in total, and we found that >95% of reads that mapped to intronic regions were found in both pipelines. The 5% of reads that differed between the outputs of the two pipelines were a result of differences in filtering out all other known small RNA, and characterizing reads that mapped to the junction of intron and exon. Output from both pipelines, identified in hundred of small RNA libraries from dozens of experiments, confirmed that isiRNA maps to approximately 1/3 of genes in the *C. elegans* genome. Indeed, we consistently found that isiRNAs account for ~7%-10% of the total siRNA reads in each small RNA library. This reproducibility across libraries gives me the confidence to dive further into exploring the properties of this class of understudied small RNA and the introns they map to.

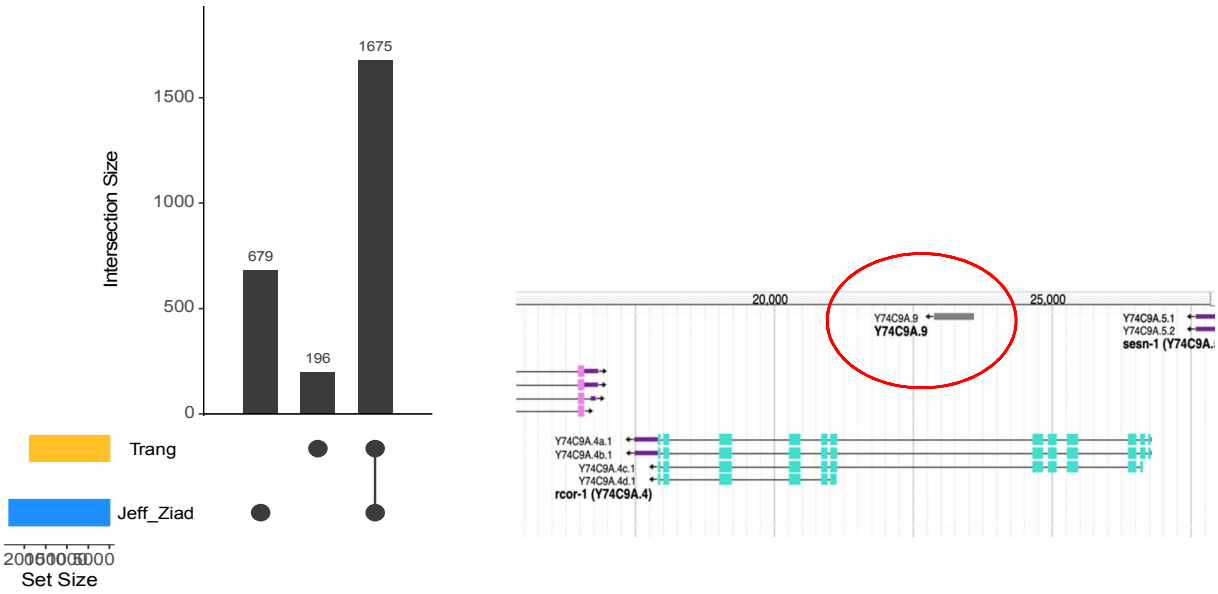


Figure 2.2. Left: upset chart of two outputs. Yellow is Trang's pipeline output and blue is Jeff-Ziad's pipeline output. Y-axis is the number of genes that have isiRNA mapped to them. The first column is 679 isiRNA genes in Jeff and Ziad output but not Trang's output. The second column is 196 isiRNA genes in Trang's output but not in Jeff-Ziad's output. The third column is 1875 isiRNA genes that overlapped between Trang's and Jeff-Ziad's outputs. **Right:** Snapshot of a region of *C. elegans* genome in Wormbase that have two genes overlapped. The red circle indicates a gene's region that is annotated as exon, but this same region is also annotated as the intron of another gene (bottom line).

CHAPTER THREE: PROPERTIES OF isiRNA INTRONS AND STATISTICAL ANALYSIS

INTRODUCTION

In eukaryotic cells, introns from the pre-mRNA are spliced out to make mature mRNA [32]. Splicing is an essential step in gene expression, in which introns are removed and exons are ligated together. In this process, some exons are constitutively present in every mRNA produced by a given pre-mRNA, but some exons are alternatively spliced to generate a variety isoform of mRNA from a single pre-mRNA molecule. Though introns are spliced out in the mRNA maturing process, they are not functionless genetic material in the genome. In fact, introns have been shown to fulfill a broad spectrum of functions in a variety of cellular processes such as: transcription regulation by both modulate splicing and [33], producing the hairpin loop for making miRNA [34,35], and enhancing mRNA nuclear export [36]. In this section, I am focusing on understanding the properties of introns that possibly influence the production of isiRNA.

esiRNA is referred to as 22G anti-sense small RNA that regulates gene expression by targeting the exon regions of mRNA in cytoplasm or the H3K9 complex at the DNA level in the nucleus. In *C. elegans*, more than half of the genes express esiRNA. I found that the other subset of small RNA that maps anti-sense to the intronic regions of transcribed mRNA, the isiRNA, is mapped to almost half of all genes in *C. elegans*. In this study, I found that almost all isiRNA mapped to the same genes that esiRNA maps to.

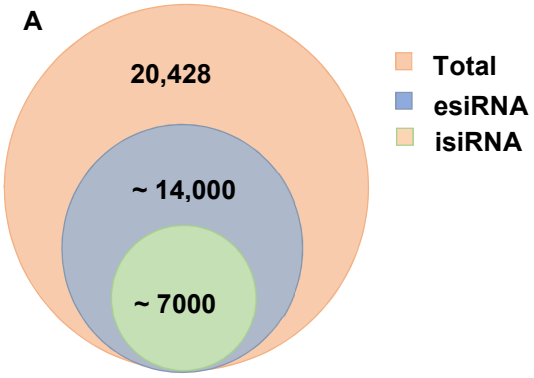
Alternative splicing enhances eukaryote cellular function complexity by increasing the number of unique proteins produced from a single gene [37]. Alternative splicing affects every aspect of cell survival and function. A variety of splicing regulatory elements have been identified to be in intronic regions. For example, in *C. elegans* the conserved sequence at the 5' splice site of the intron guides the binding of U1 and U2 splicing proteins to ensure appropriate splicing. Here, I found that genes with alternative splicing are more likely to have isiRNA mapped to them. This suggested that isiRNA might have a function in regulating alternative splicing.

The size of the intron is optimized to conserve energy [38], as transcription is a slow and energy costly process. It was shown that introns of highly expressing genes are substantially shorter compared to introns of lower expressing genes [39]. Thus, the average size of introns in *C. elegans* is about 50 nucleotides: This is thought to be the result of natural selection in optimizing splicing [39]. Additionally, the longer introns were shown to have a longer splicing time. This prolonged splicing facilitates the formation of siRNA precursors (hairpin loops), thereby promoting the siRNA biosynthetic machinery [40]. Consistent with that, my study showed that isiRNA is produced from a longer intron. This suggested that isiRNA might play a quality-control role by eliminating long, poorly spliced introns in order to ensure the quality of the protein that cell is intended to produce.

RESULTS

A. Genes that have isiRNA mapped to them also have esiRNA mapped to them.

20,428 genes have been identified and annotated in the *C. elegans* genome and more than half of these genes have 22G RNA that map to them. More precisely, about 14,000 genes identified have esiRNA mapped to them [13]. Consistently with our lab result: In the 12 small RNA libraries that our lab made and sequenced, there are on average approximately ~14,000 genes with esiRNA and 7,000 with isiRNA mapped to them. The schematic Figure 3.1A indicate the almost completely overlapped of isiRNA and esiRNA. Then I did the analysis to calculate the number of isiRNA genes overlapped with esiRNA genes for every N2 libraries in my study. Strikingly, I found that in all six N2 libraries, above 90% of genes that produce isiRNA also produce esiRNA (Figure 3.1B). This result suggested that isiRNA might have overlapping functions with esiRNA.


A
20,428
~ 14,000
~ 7000
Total
esiRNA
isiRNA

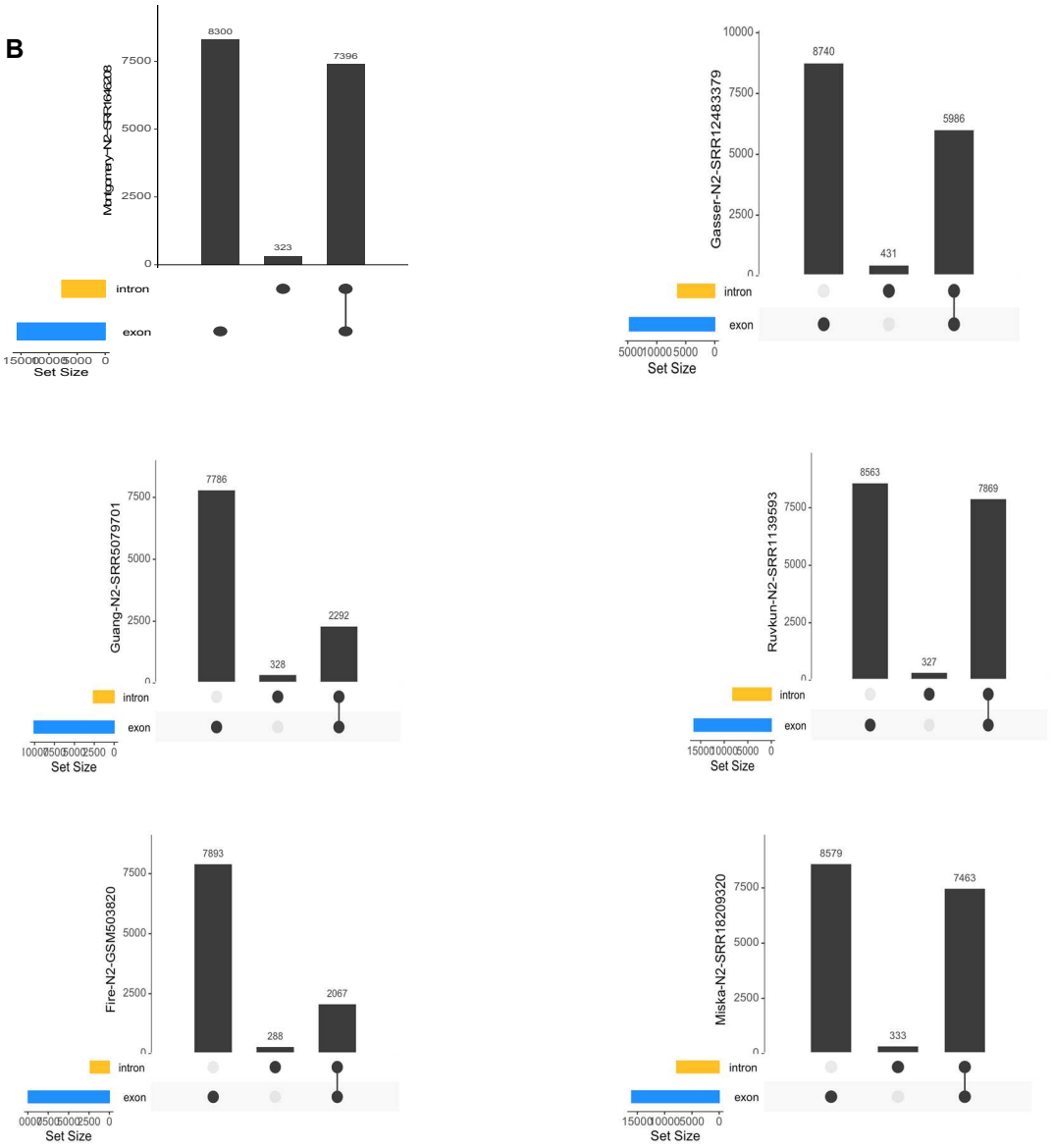

B
8300
7396
323
intron
exon
Set Size
8740
5986
431
Gasser-N2-SRR12483379
7786
2292
328
Guang-N2-SRR5079701
8563
7869
327
Ruvkun-N2-SRR1139593
7893
2067
288
Fire-N2-GSM503820
8579
7463
333
Miska-N2-SRR18209320

Figure 3.1 A. Schematic of genes in *C. elegans*. The outermost circle in pink represents 20,428 genes in *C. elegans*. The blue circle represents ~14000 esiRNA genes. The innermost circle in orange represents ~7000 isiRNA genes. This data is from the 12 libraries that our lab generated in 2013 **B.** Upset charts indicate overlapping of esiRNA and isiRNA genes in 6 N2 libraries from different experiments. X-axis: blue bar represents exon-mapping siRNA, yellow bar represents intron-mapping siRNA. Y-axis is the number of genes. The first column is the number of genes that produce esiRNA, the second column is the number that produce isiRNA, and the third column is the number of genes that produce both esiRNA and isiRNA. Almost all of the isiRNA in this library are also esiRNA.

B. Genes with alternative splicing are more likely to have isiRNA map to them as compared to genes without alternative splicing.

Alternative splicing is a common mechanism to generate multiple isoforms of a protein from the same gene [41]. The splicing reaction entails the removal of introns and the ligating of the exons as mRNA is maturing in the nucleus. In order to remove the intron precisely, the spliceosome needs to recognize the highly conserved sequences at the 5' and the 3' site of the intron. Each gene contains multiple introns and exons. Alternative splicing is the result of ligating different exons together and is an important mechanism to expand a gene's functions [42]. Specifically, alternative splicing allows the same gene to give rise to multiple versions of mRNAs in different tissues, which could give cells different identities [42]. Alternative splicing has been showed to contribute to many processes from sex determination [43] to programing cellular apoptosis [44]. In humans, there are about 25,000 genes in total, and approximately 6,200 of those genes are alternatively spliced [45].

In *C. elegans*, there are 20,428 genes in total, and 4,504 are alternatively spliced. My analysis showed that genes with alternative splicing are more likely to give rise to isiRNA compared to genes without alternative splicing. Specifically, I examined 12 N2 libraries from 9 different labs and found that the percentage of all genes that are alternatively spliced (22.1%) and the percentage of isiRNA genes that are alternatively spliced (38.3%) is significantly different, while there is no difference between the percentage of all genes that are alternatively spliced (22.1%) and the percentage of esiRNA genes that are alternatively spliced (2 tailed Student's t test) (Figure 3.2). I then performed a "one sample proportion test" comparing the probability of any gene giving rise to isiRNA [(7,000/20,000)*100 = 35%] to the probability of an alternatively spliced gene giving rise to isiRNA [(2190/4504)*100 = 49%]: These two percentages are significantly different from each other (P value is 0.0062). I interpret this to mean that genes that are alternatively spliced have a higher probability to give rise to isiRNA compared to genes that are not alternatively spliced.

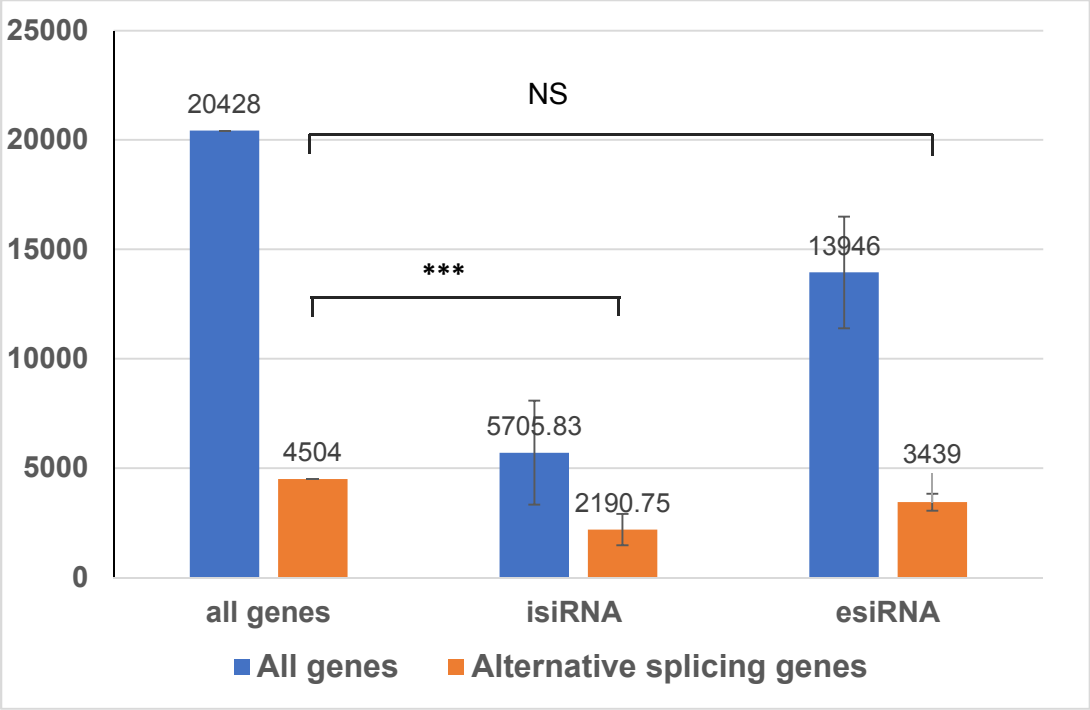
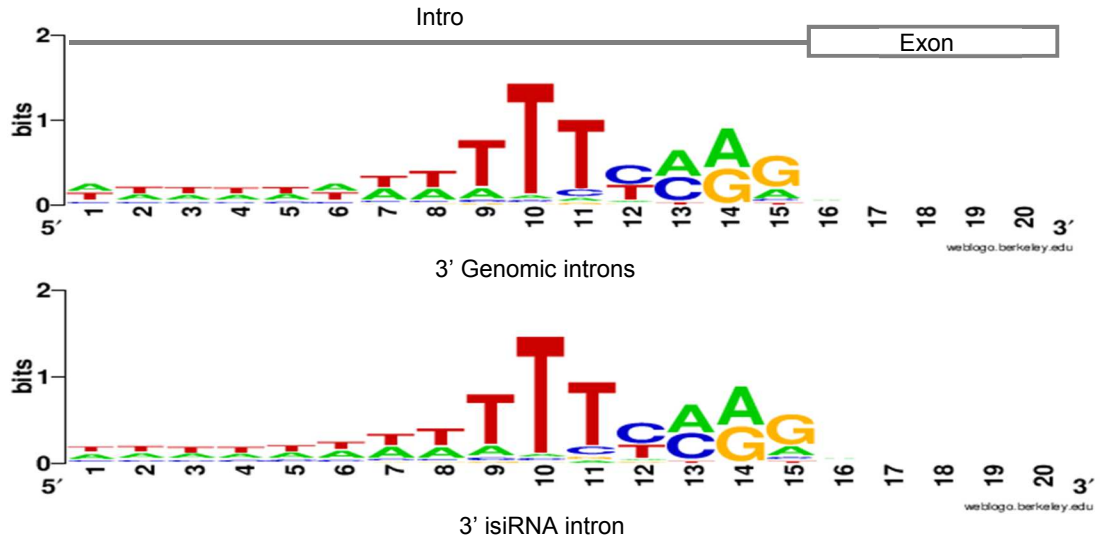


Figure 3.2 Y-axis is the number of genes. X-axis: first blue column is all genes, first red column is all alternative splicing genes. The second blue column is the average number of isiRNA genes. The second red column is the average number of isiRNA-alternatively spliced genes from. The third blue column is the average of esiRNA genes. The third red column is the average of esiRNA-alternatively spliced genes. Average was calculated based on12 N2 libraries across 9 different laboratories. Error bar is standard deviation. NS = not significant

Information content refers to the level of conservation of the splicing site: High information content means the splicing site is well-conserved and low information content means the splicing site is poorly-conserved. Alternatively spliced introns often have less well-conserved 3’ acceptor or 5’ donor sites, lowering their information content. I used a sequence generator software, WebLogo [46], to tested the information content of the 5’ and 3’ splicing sites in those introns giving rise to isiRNA in our 12 libraries, then compared the results to that of the genomic introns. I found that the information content of neither the 5' nor 3’ splicing sites differed between those introns that gave rise to isiRNA and all other genome introns (Figure 3.3A, and 3.3B). In conclusion, introns with and without isiRNA had the same information content at their splice acceptor and donor sites.

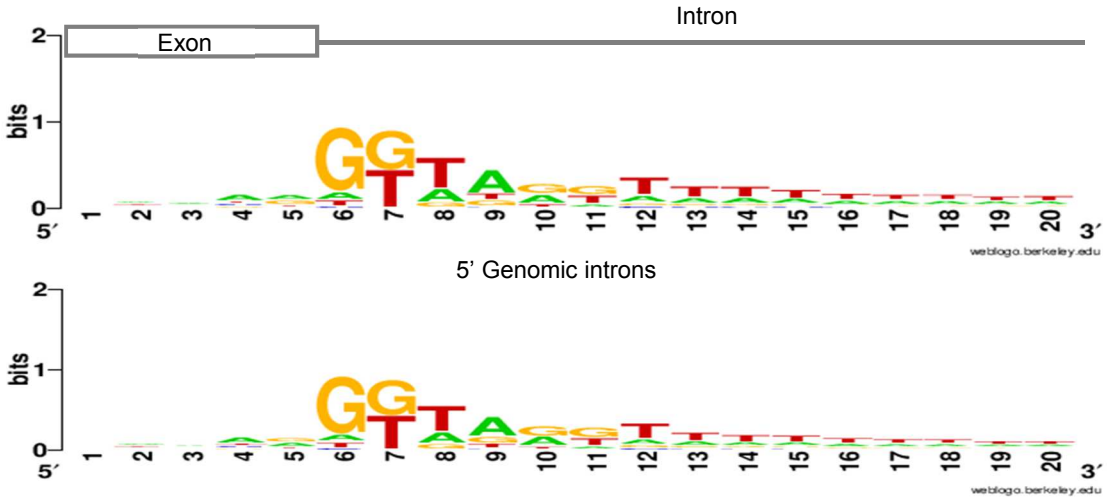
A.



3’ Genomic introns

3’ isiRNA intron

B.



5’ Genomic introns

5’ isiRNA intron

Figure 3.3 Information content calculation using WebLogo software. The larger the letter the more conserve the nucleotide. **A** The information content of 3’ splice site. Top is genomic intron. Bottom is isiRNA intron.
B. Information content calculation of 5’ splicing end. Top is genomic intron. Bottom is isiRNA intron.

C. Median length of isiRNA introns is 10 fold longer than that of genomic introns.

Dumesic et al., 2013 showed that pre-mRNA is the substrate for both spliceosomes and siRNA biosynthetic machinery [40](23415457): Thus, spliceosomes and siRNA biosynthetic machinery are thought to compete for pre-mRNA. In this scenario, the slower the splicing the longer time the mRNA spends as a pre-mRNA; and, this makes the poorly spliced mRNA more likely to be the target of siRNA [40](23415457). The longer the intron the more likely secondary structure hinders spliceosome binding and this can lead to stalling of the splicing machinery, increasing siRNA production [40](23415457).

In *C. elegans*, there are 111,296 introns ranging in length from 15 to 20,249 nucleotides. On average, each gene has about 5 introns. I found that there are approximately ~6500 isiRNA introns from about 6000 genes that produce isiRNA. So not all introns in a gene give rise to isiRNA. I tested to see if the intron's length influences the likelihood that it gives rise to isiRNA.

In order to find the difference in length between introns that give rise to isiRNA and introns that do not, I compared each type of intron's median lengths and their length distributions. This data is based on 12 libraries made from our lab. To handle the large difference in sample sizes between isiRNA introns (6500) and the genomic introns (111,296), I compared the isiRNA intron length to the length of a random set of 6500 genomic introns. I generated this random set of genomic introns by randomly drawing 6500 introns with replacement from 111,296 total genomic introns, repeating this random genomic intron generation 450 times, based on the desired power of 0.9, and CI = 95%. The results showed that the median length of introns that gave rise to isiRNA is 697 nucleotides, while the median length of 6500 random genomic introns is 69 nucleotides (Figure 3.4). In figure 3.4, the Y-axis is the number of genes. The first column is the median of 6500 random genomic introns (69 nt) and the second column is median of 6500 isiRNA introns (693 nt). Student's t test was performed to evaluate the difference between random genomic introns and isiRNA introns. The

result showed that the median length of isiRNA introns is 10 fold longer compared to the median length of genomic introns, p value = 5.29*E-12 (Figure 3.4).

The length of genomic introns and isiRNA introns are not normally distributed. The lengths range from 15 to 20,249 nucleotides in both intron populations. Additionally, the median length of genomic introns is 69 nucleotides while that of isiRNA introns is 693 nucleotides. Therefore, I applied the Kolmogorov-Smirnov test to evaluate the length distribution between these two populations that are not randomly distributed. The result showed that the length distribution of isiRNA introns (blue curve) is separated from the length distribution of genomic introns (red curve). X-axis is intron length. Y-axis is cumulative fraction (Figure 3.5).

I interpret the result above that isiRNA are more likely to target long introns. Perhaps this is one of the functions of isiRNA: Eliminating the long introns that are poorly spliced in order to ensure the quality of the protein that cell is intended to produce.

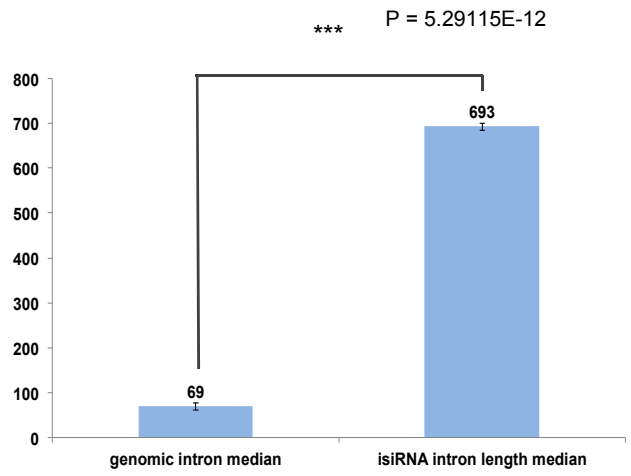


Figure 3.4 Student's t-test showing the different between the median length of genomic intron (69 nucleotides) and the median length of isiRNA intron (693 nucleotides).

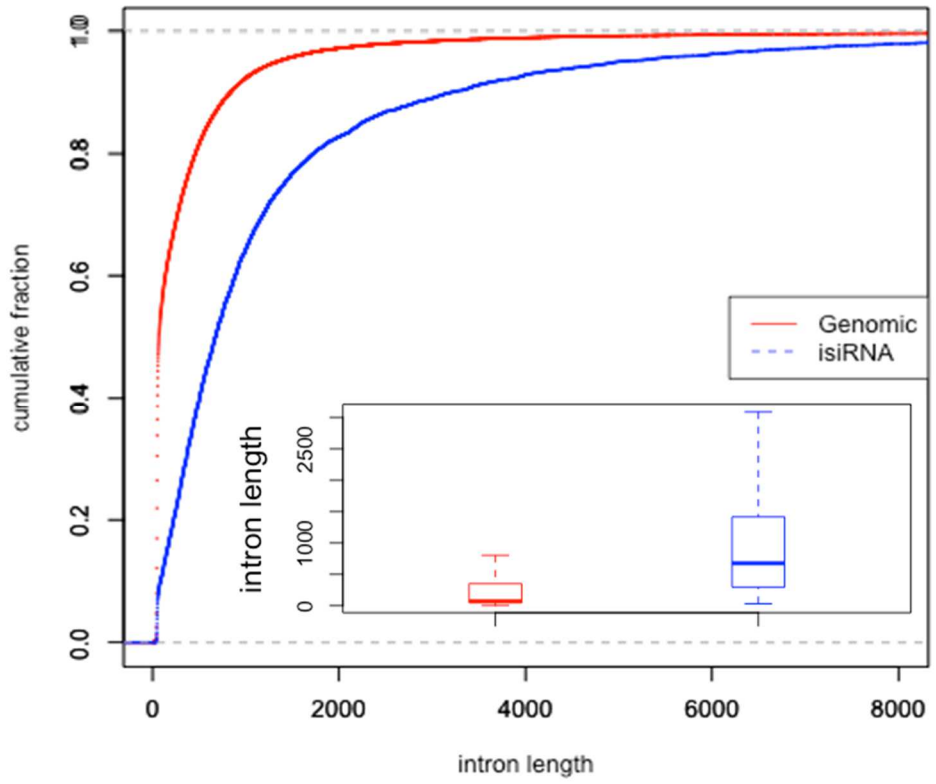


Figure 3.5 Kolmogorov-Smirnov test to show the length distribution difference between isiRNA intron (blue curve) and genomic intron (red curve). X-axis is the intron length. Y-axis is cumulative fraction.

CHAPTER FOUR: INSIGHTS INTO THE BIOSYNTHESIS OF isiRNA

INTRODUCTION

Since the first description of inhibition of gene expression by antisense RNA in the early 1980s [47], and the discovery of exogenous and endogenous siRNA in the late 1990s, we have a fairly good understanding of the biosynthesis pathways of those classes of siRNA. However, the biosynthesis of siRNA that targets the intronic regions, the isiRNA, remains unknown. In this study, in order to understand the biosynthesis of isiRNA and thus potential triggers for its production, I decided to look for mutant backgrounds that disrupt or increase isiRNA levels in publicly deposited data sets (Table 4.1). This candidate screen was focused on those genes that had been shown previously to affect production of endogenous siRNA (endo-siRNA) and exogenous siRNA (exo-siRNA). By doing so, I hoped to uncover the pathways that regulate the less studied isiRNA production. Understanding the genetic factors that are required for biosynthesis of isiRNAs may provide insight into: the function of these small RNAs and possibly what triggers their production.

There are three main small RNA mediated silencing species that are identified in *C. elegans*: microRNAs (miRNA), piwi-associated RNAs (piRNAs), exogenous RNAi, and endogenous RNAi (endo-siRNAs) (Figure 4.1). MicroRNAs are encoded within the genome as transcripts that fold back into hairpin double stranded RNA, which are processed into ~22 nucleotide noncoding RNA species. These function post-transcriptionally to repress gene expression by interacting with regions in the target gene's 3'UTR that have "seed" sequences with some homology to the miRNA [10]. Piwi-associated RNAs (piRNAs) are 21-nucleotide RNA species that are germline-specific and are encoded in the genome. They are required for silencing transposable elements and thus maintain genome integrity [11]. Exogenously triggered RNAi is elicited by dsRNA supplied from outside the worm and taken up by the intestinal cells. The small interfering RNA species are amplified and transported through the animal where they are taken up by the double stranded RNA import channel, SID-1 [12,48]. Endogenous RNAi (endo-

siRNAs) is triggered by transcripts that are expressed from the genome. Approximately two thirds of the genes within the *C. elegans* genome have endo-siRNA species that map to them. These siRNA come from transcription of both coding and noncoding genomic sequences [13,49].

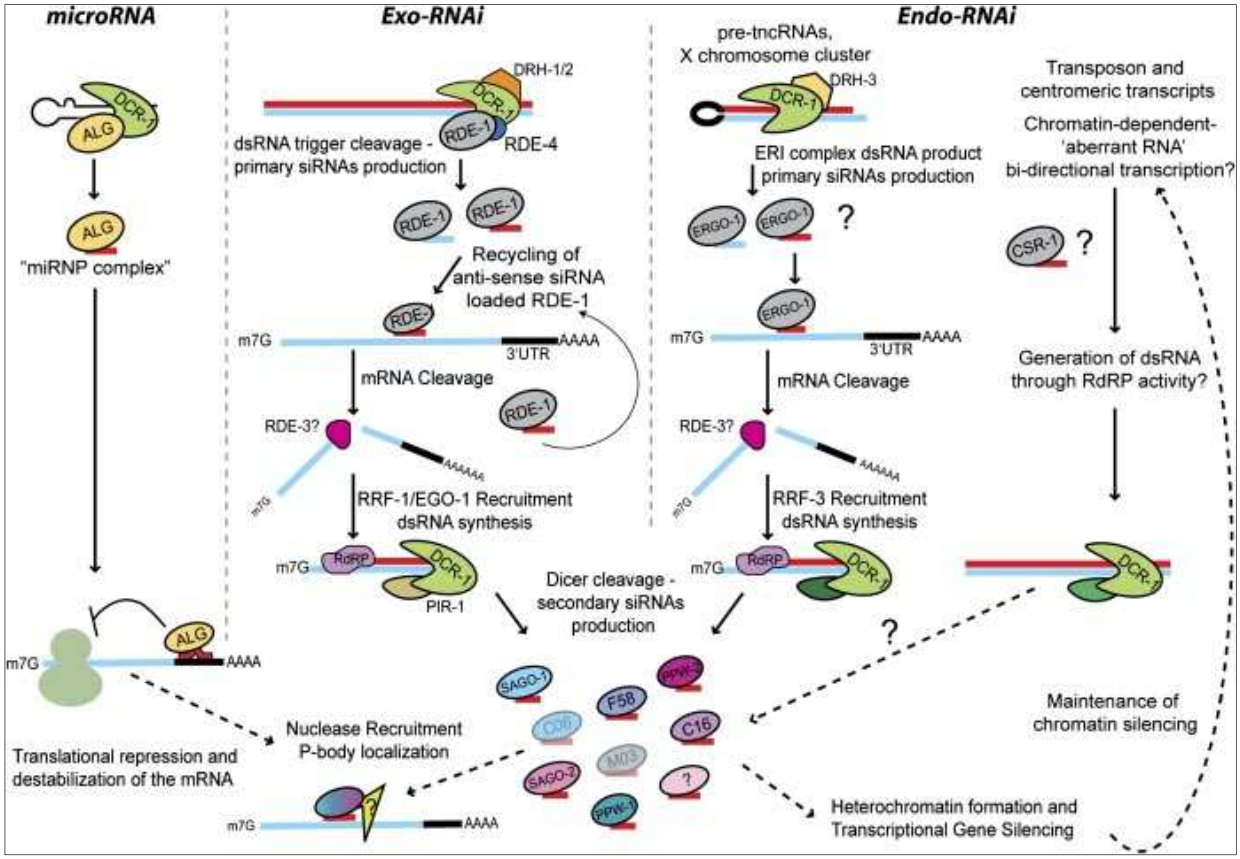


Figure 4.1 Schematic of three main RNAi pathways in *C. elegans*. Yigit, Erbay et al. “Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi.” *Cell* vol. 127,4 (2006): 747-57. doi:10.1016/j.cell.2006.09.033

Table 4.1: Table summary the changes in esiRNA and isiRNA levels in the mutant libraries relatively to the esiRNA and isiRNA levels in the wildtype libraries of the same experiment. The first column is the genotype, second column is esiRNA relative to that of wildtype, the third column is isiRNA relative to that of wildtype, the fourth column is the experiment that the libraries originated from, and the last column is the known phenotype of those mutants.
Since miRNA level are not effected in the small RNA library preparation process, I normalized the esiRNA and isiRNA level in all libraries of this study based on the 11 highest most stable miRNA expression in the small RNA library. The normalization is described below.
Normalization: In each library, read per millions of 11 stably expressed miRNA genes to calculate the normalization fraction. First, take the average expression of those genes then divide them by 10,000. The fraction was then multiplied by each gene. Here below are the 11 highly expressed miRNA genes. WBGene00003315, WBGene00004622, WBGene00003308, WBGene00003299, WBGene00003260, WBGene00003335, WBGene00003279, WBGene00003284, WBGene00003305, WBGene00002993, WBGene00003286.

Factors	esiRNA	isiRNA	PMID	Tissue
glp-1(e2141)	Decrease	Decrease	20116306 Fire	germline
glp-4(bn2)	Decrease	Decrease	20116306 Fire	germline
fem-1(hc17)	Decrease	Decrease	20116306 Fire	germline
ego-1(om84)	Decrease	Decrease	GSE66344 Miska	Germline RdRP
ego-1(om97)	Decrease	Decrease	19804758 Mello	Germline RdRP
ego-1(om97);rrf-1(neC1)	Decrease	Decrease	19800275 Mello	Germline RdRP
ego-1(om97);rrf-1(neC1)	Decrease	Decrease	GSE66344 Miska	Germline RdRP
ekl-1(tm1599)	Decrease	Decrease	19800275 Mello	Germline RdRP complex
csr-1(ADH)	Decrease	Decrease	34108460 Cecere	Germline argonaute
cde-1(tm1021) young & old	Slight Increase	Slight Increase	19804759 Ketting	Germline CSR-1
csr-1 KO (gc017)	Slight increase	Slight increase	34108460 Cecere	Germline Argonaute
prg-1(n4357) x prg-1(n4357)	Decrease	Decrease	24684932 Mont	PIWI argonaute, germline

Factors	esiRNA	isiRNA	PMID	Tissue
prg-1(n4357);mut-14(pk738);smut-1(tm1301) x prg-1(n4357);mut-16(pk710) (Het for mutators) homozygous piArgonaute PRG-1	Decrease	Decrease	24684932 Mont	PIWI argonaute and Mutators, germline
mut-16(pk710) trans-silencing	Decrease	Decrease	GSE66344 Miska	Mutator complex, P Granule Germline
*mut-16(pk710)**	Decrease	Decrease	21245313 Ruvkun	Mutator complex, P Granule Germline
*mut-16(pk710)***	Decrease	Decrease	24684932 Mont	Mutator complex, P Granule Germline
mut-7(pk204)	Decrease	Decrease	19800275 Mello	Mutator complex, P Granule Germline
mut-7(pk204)	Slight decrease	Decrease	24684932 Mont	Mutator complex, P Granule Germline
mut-2(ne298)	Slight decrease	Decrease	24684932 Mont	Mutator complex, P Granule Germline
mut-15(tm1358)	Slight decrease	Decrease	24684932 Mont	Mutator complex, P Granule Germline
mut-16(mg461)	No change	No change	21245313 Ruvkun	Mutator complex P granule Germline
mut-14(pk738) unc119	Decrease	Decrease	24684932 Mont	Mutator complex DEAD box RNA helicase
mut-14(mg464) unc119	No change	No change	24684932 Mont	Mutator complex DEAD box helicase
smut-1(tm1301) unc119	No change	No change	24684932 Mont	Mutator complex DEAD box helicase
mut-14(mg464);smut-1(tm1301) unc119	Decrease	Decrease	24684932 Mont	Mutator complex DEAD box RNA helicases
mut-14(pk738);smut-1(tm1301) unc119	Decrease	Decrease	24684932 Mont	Mutator complex DEAD box RNA helicases
mut-14(pk738);smut-1(tm1301) x mut-16(pk710) (hets)	No change	Decrease	24684932 Mont	Mutator complex DEAD box RNA helicases
ergo-1(gg098), NOT embryo	Decrease	No change	20116306 Fire	26G Argonaute embryo

Factors	esiRNA	isiRNA	PMID	Tissue
ergo-1(tm1860) embryo	No change	Slight increase	20133583 Conte	26G Argonaute embryo
eri-9(gg106)	Decrease	Decrease	20116306 Fire	ERI complex
drh-3(ne4253)GFP sensor	Decrease	Decrease	GSE66344 Miska	ERI complex
drh-3(ne2345)	Decrease	Decrease	24137537 Miska	ERI complex
*dcr-1(*mg375)	Decrease	No change	20116306 Fire	ERI complex
eri-1(mg366)	Decrease	No change	20116306 Fire	ERI complex
rde-1(ne300)	Decrease	Decrease	20116306 Fire	Exo-RNAi Argonaute
rde-4(ne299)	Decrease	Decrease	20116306 Fire	Exo-RNAi
rde-4(ne337)	No change	No change	19800275 Mello	Exo-RNAi
drh-1(ok3495);drh-3(2345)	Slight decrease	Slight decrease	24137537 Miska	Exo- RNAi Endo-RNAi
rrf-1(pk1417) embryo	Increase	Increase	20133583 Conte	RdRP
rrf-1(ok589)	No change	No change	GSE66344 Miska	RdRP
rrf-2(ok210)	No change	No change	GSE66344 Miska	RdRP
*rrf-3(pk1426)**	No change	No change	20116306 Fire	RdRP
*rrf-3(pk1426)***	No change	No change	GSE66344 Miska	RdRP
*rrf-3(pk1426)*** embryo*	No change	No change	20133583 Conte	RdRP
nrde-1(gg088) trans-silencing	Decrease	Decrease	GSE66344 Miska	Nuclear RNAi
nrde-4(gg129) trans-silencing	Decrease	Decrease	GSE66344 Miska	Nuclear RNAi

Factors	esiRNA	isiRNA	PMID	Tissue
nrde-3() trans-silencing	Decrease	Decrease	GSE66344 Miska	Nuclear RNAi Argonaute
nrde-3(tm1116);met-2(n4256) early embryo	Decrease	Decrease	33303642 Gasser	Nuclear RNAi Chromatin RISC
met-2(n4256) early embryo	No change	Decrease	33303642 Gasser	H3K9mono di-me transferase
nrde-3(tm1116) early embryo	No change	No change	33303642 Gasser	Nuclear RNAi Argonaute
set-25(n5021) early embryo	No change	No change	33303642 Gasser	H3K9tri-me transferase
set-25(n5021);nrde-3(tm1116) early embryo	No change	No change	33303642 Gasser	Nuclear RNAi Argonaute H3K9tri-me transferase
rde-2(pk1657)	Decrease	Decrease	24684932 Mont	Mutator complex (MUT-7) nucleotidyltransferase
rde-3(ne3364)	Slight decrease	Slight decrease	19800275 Mello	Mutator complex (MUT-2) nucleotidyltransferase
smg-5(r860)	No change	No change	19800275 Mello	Nonsense mediated decay, cut the poly A tail
mago-12(missing 12 argonautes)	Slight decrease	Decrease	19800275 Mello	Argonaute

RESULTS AND DISCUSSION

A. A diverse set of common siRNA factors disrupt both isiRNA and esiRNA in the germline.

As in other organisms, *C. elegans* germline establishes the growth and maintenance of germ cells throughout different stages of development and in the adult gonad. At the very first cell divisions, the germ cell progenitors are set aside from those that will develop into the soma. Thus, the germline develops and is distinct from somatic cells in early embryogenesis. As *C. elegans* progresses through different stages of development, the mass of the germline becomes a larger proportion of the animal's total mass. The number of germ cells is increased from ten cells at larval 1 (L1 stage), hundred at L2 and L3, and ~1,000 at L4 stage, reaching a maximum number of ~ 3,000 - 4,000 cells in the adult [50]. There is evidence that esiRNA are produced and function in both the soma and germline in *C. elegans* [31,51]. Our data show that normal germline development is important for not only esiRNA but also for isiRNA biosynthesis. Specifically, three of the mutants that ablate the germline, *glp-1(e2141), glp-4(bn2), fem-1(hc17)* [31] all show significantly reduced levels of both esiRNA and isiRNA. These data suggest that isiRNA along with esiRNA are produced in the germline.

RRF-1 is an RNA-dependent RNA Polymerase (RdRP) that is expressed in both the soma and the germline [13]. Our analysis of *rrf-1(pk1417)* embryos that are defective for this RdRP [21] showed an increase in both esiRNA and isiRNA levels, which may indicate that RRF-1 normally opposes production and/or germ cell loading of these small RNAs. Thus, this specific allele *rrf-1 (pk1417)* somehow triggers the esiRNA and isiRNA machinery in the germline to produce more siRNA or to block its handling in germ cells (Figure 4.2).

RdRPs are not the only factors that play an important role in the production of esiRNA and isiRNA in the germline: The mutator factors also contribute to the biosynthesis of esiRNA and isiRNA in these cells. The mutator factors that are normally localized in the germline P-granules are missing in the following mutant strains: *mut-16(pk710), mut-7(pkk204), mut-*

2(ne298), mut-15(tm1358), rde-2(pk1657), mut-14(pk738), mut-14(mg464), smut-1(tm1301) [52] and, when they are missing, the levels of isiRNA and esiRNA are reduced (Figure 4.3). Analysis of small RNA libraries from three independent studies each show that the levels of esiRNA and isiRNA are depleted in the germline and somatic *pk710* allele of *mut-16*: [14,16,52]. By contrast, the *mg461* allele of *mut-16*, which only removes somatic MUT-16 functions [14], did not affect the levels of either esiRNA or isiRNA [14] (Figure 4.4). This strengthens our hypothesis that isiRNA is produced mainly in the germline.

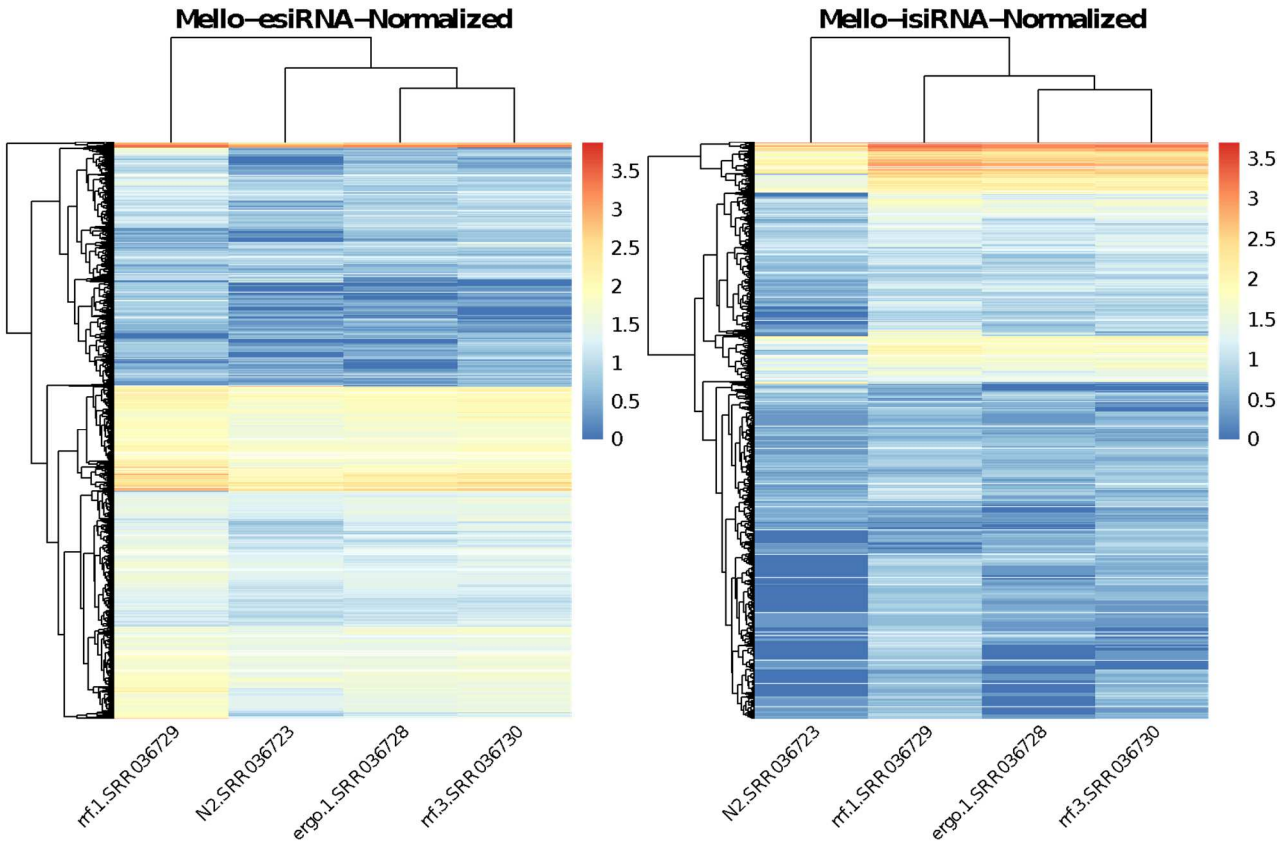


Figure 4.2 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1.

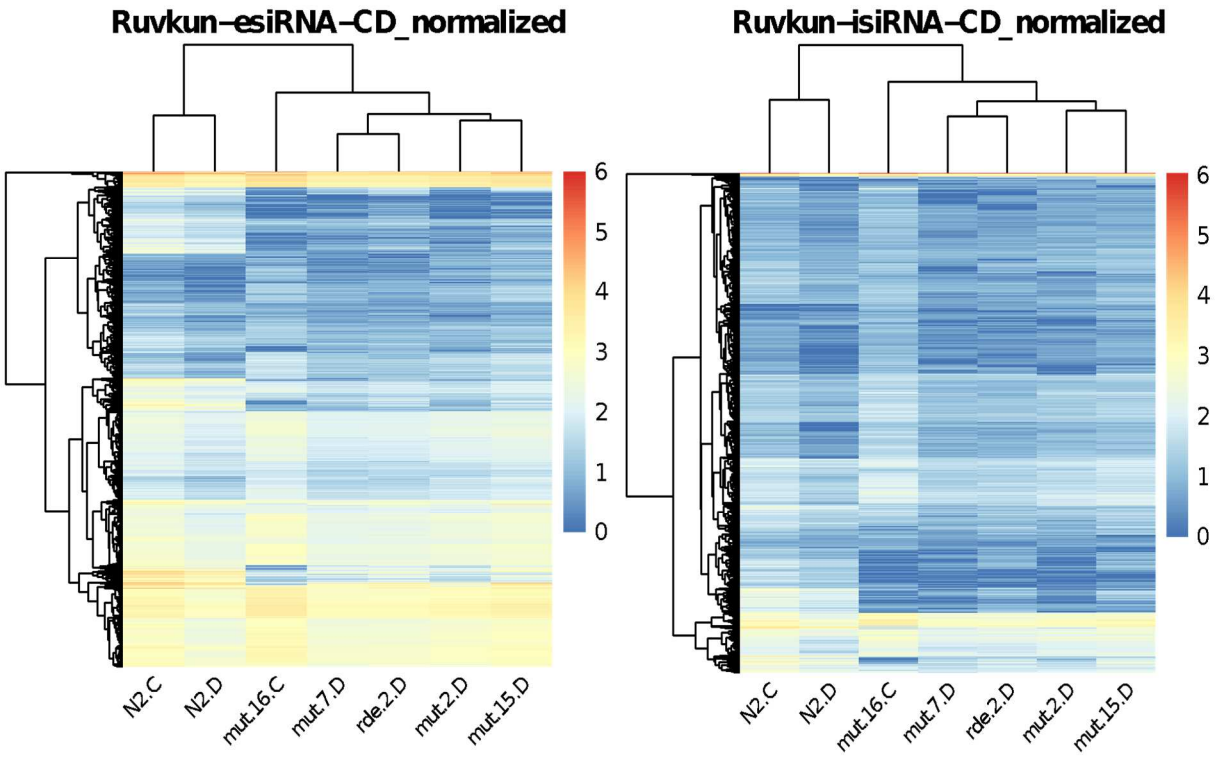


Figure 4.3 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 1.

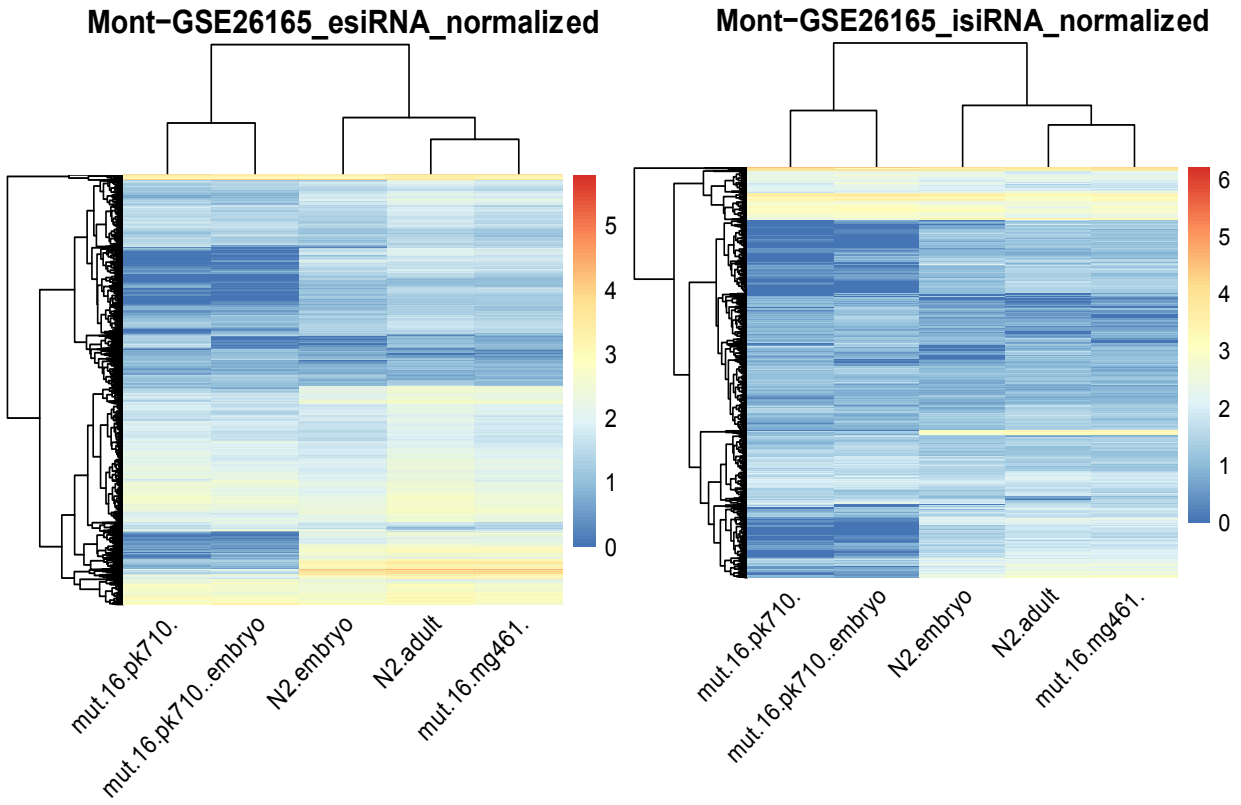


Figure 4.4 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1

B. Evidence that isiRNA are produced via the WAGOs pathway

Small interfering RNA biosynthetic pathways are characterized by the Argonautes that bind the small RNA. The two main pathways in the germline are associated with either the WAGO-1 Argonaute or CSR-1 [13]. The WAGO-1 pathway is dependent on RDE-3 and MUT-7, while the other pathway is dependent on the nucleotidyltransferase, CDE-1 [13]. Both pathways require RdRP RRF-1, the germline specific RdRP EGO-1, and tudor-domain protein EKL-1 22Gs biosynthesis systems [13]. The Mello lab showed that the RdRPs EGO-1 and RRF-1 have a redundant role in 22G biosynthesis [13]. In our analysis of data from multiple, independent studies, we found that *ego-1(om97)* [19], *ego-1(om84),* the double mutant *ego(om97);rrf-1(neC1)* [16], and the germline RdRP complex member *ekl-1(tm1599)* [13] are required for both esi and isiRNA production. The other RdRPs that are not germline-specific, such as: *rrf-1(ok589), rrf-2(ok210), rrf-3(pk1426)* [31], *rrf-3(pk1426)* [16], *rrf-3(pk1426)* [21], do not seem to affect esiRNA or isiRNA production. Thus, the single mutants in the other RdRPs did not alter esiRNA or isiRNA levels. Only loss of *ego-1* reduced the levels of these small RNAs. Therefore, the germline specific RdRP, EGO-1, would seem to be the most important RdRP for production of these small RNA species. As in the WAGO 22G biosynthetic pathway, we also found that RDE-3 and MUT-7 are important for biosynthesis of both esiRNA and isiRNA. By contrast, we saw that esiRNA and isiRNA levels were not altered in the *csr-1(gc017)* knockout strain, which is likely a null mutation in *csr-1* [53] (Figure 4.4). esiRNA and isiRNA levels are in fact elevated. Likewise, in strains that lack the CSR-1-associated co-factor, *cde-1(tm1021)*, the levels of each species is elevated (Figure 4.5). This led us to posit that esiRNA and isiRNA species can be produced by the WAGO pathway when the CSR-1 biosynthetic machinery is missing.

The germline-specific Argonaute CSR-1 is multifaceted. For instance, Claycomb and Mello [19] showed that CSR-1 promotes transcription of germline-specific genes and Singh et al., 2021 reports that CSR-1's slicing activity is required to degrade transcripts that would be poorly translated. We find that complete loss of the germline-specific Argonaute CSR-1 *(csr-1(gc017))*

slightly elevates the level of esiRNA or isiRNA (Figure 4.5). Surprisingly, when the slicing activity of CSR-1 alone is mutated but the protein is expressed, as in the slicing defective *csr-1(ADH1)* allele, we see a decrease in both esiRNA and isiRNA levels (Figure 4.5) [54]. This may mean that the slicing activity is required to create both esiRNA and isiRNA, and when the CSR-1 protein is present but unable to slice RNA these small RNA species are not produced. A parallel pathway for their production may be sensitive to the presence of CSR-1, perhaps because CSR-1(ADH) competes for the esiRNA and isiRNA precursors. Thus, when CSR-1 protein is absent (as in the *csr-1(null)*, the redundant pathway can produce esiRNA and isiRNA. The siRNAs produced by this parallel pathway may build up in CDE-1 mutants that fail to load siRNA onto CSR-1. Indeed, loss of CDE-1 [53] seems to increase the level of esiRNA and, to a lesser extent, isiRNA. This observation is consistent with Ketting's group [53] that found an increase in esiRNA in this mutant background. The redundant pathway might be associated with P granules where WAGO-1 and CSR-1 compete for siRNA loading [55]. PRG-1 may provide the splicing activity needed for isiRNA production when CRS-1 is missing.

Loss of the Argonaute, ERGO-1 that is expressed only in the embryonic stage, and is almost absent in L3, L4 and young adults that do not have embryos [21] does not affect esi or isiRNA levels (Figure 4.7). Through Argonaute CoIP analysis, we found that isiRNA binds to the ERGO-1 Argonaute (Figure 4.8). This association with ERGO-1 suggests the possibility that isiRNAs function in embryos. When we analyzed a small RNA library that was extracted from *ergo-1(gg098)* mutant of all other stages but embryos (Fire, et al), we found that esiRNA levels were decreased relative to wildtype but isiRNA levels were not changed (Figure 4.7). We interpret this to mean that though isiRNAs are bound to ERGO-1, their production, unlike that of eisRNAs does not depend on this Argonaute. A similar trend was seen when we analyzed libraries made from a separate allele of *ergo-1, tm1860*, we found that esiRNA levels were unchanged while the levels of isiRNA were slightly increased [21], This suggests that esiRNA production is dependent on ERGO-1, but isiRNA production is not.

ERGO-1 bound 26 nt small RNAs are primary siRNAs and it would seem that esiRNA production depends more heavily on this trigger species than does isiRNA production. The fact that ERGO-1 is embryonic specific means that the esiRNA load depends more on what is found in the embryo while isiRNA may have an independent trigger that occurs during the later larval stages and continues into the adult. Though isiRNA production does not depend on ERGO-1, ERGO-1 can bind to and potentially use isiRNA as a guide to destroy unspliced mRNA. If one were to sequence the mRNA from an *ergo-1* defective strain, we would expect that the levels of messages with retained introns would increase. This would be another mechanism for quality control to reduce the possibility of translating a message that contains introns.

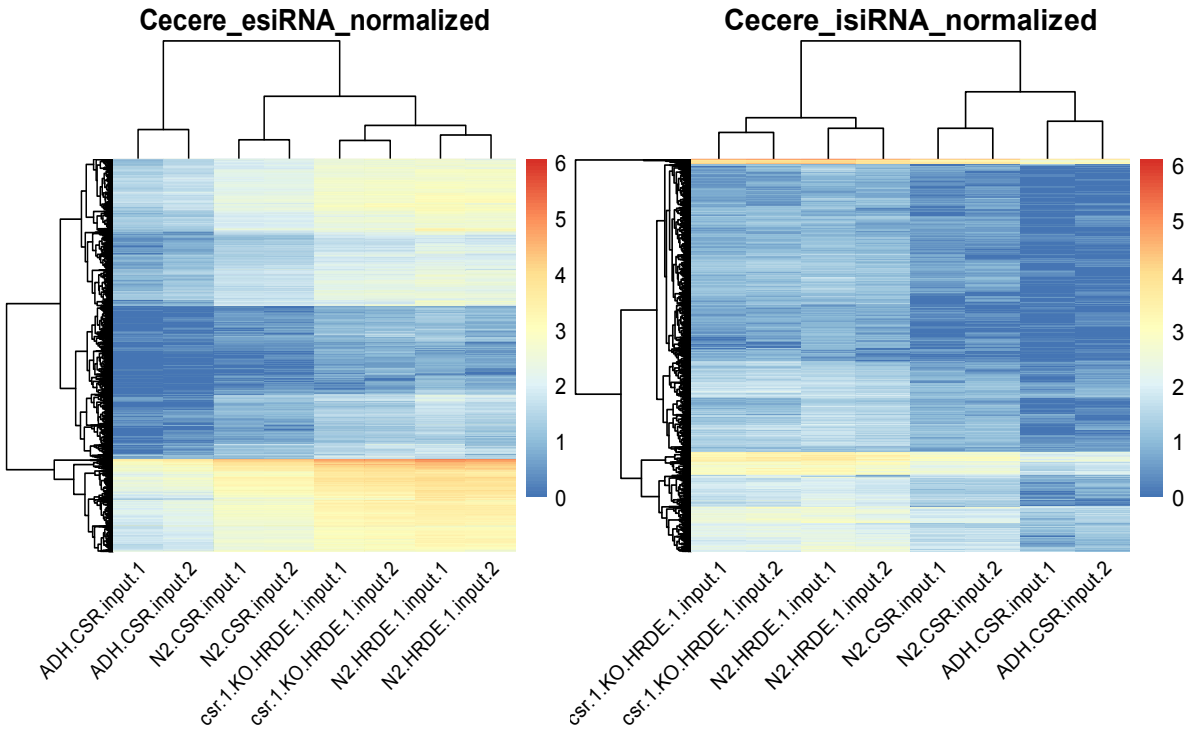


Figure 4.5 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1.

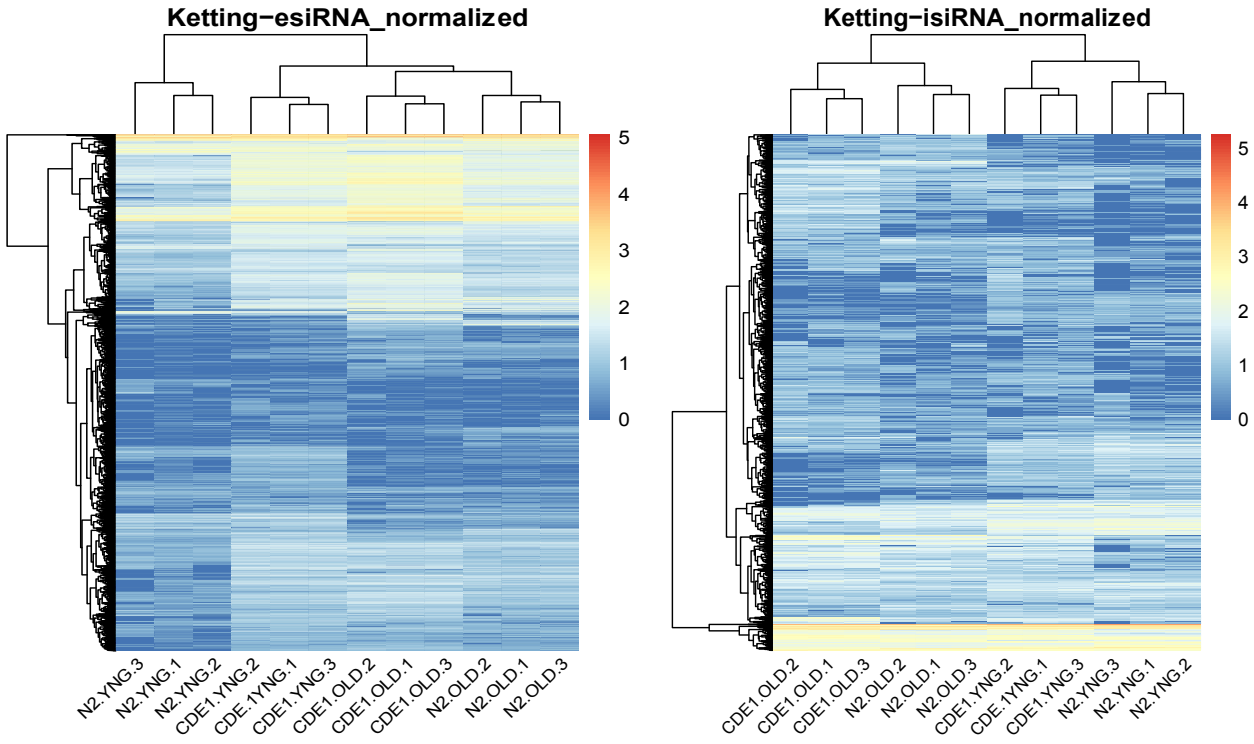


Figure 4.6 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1.

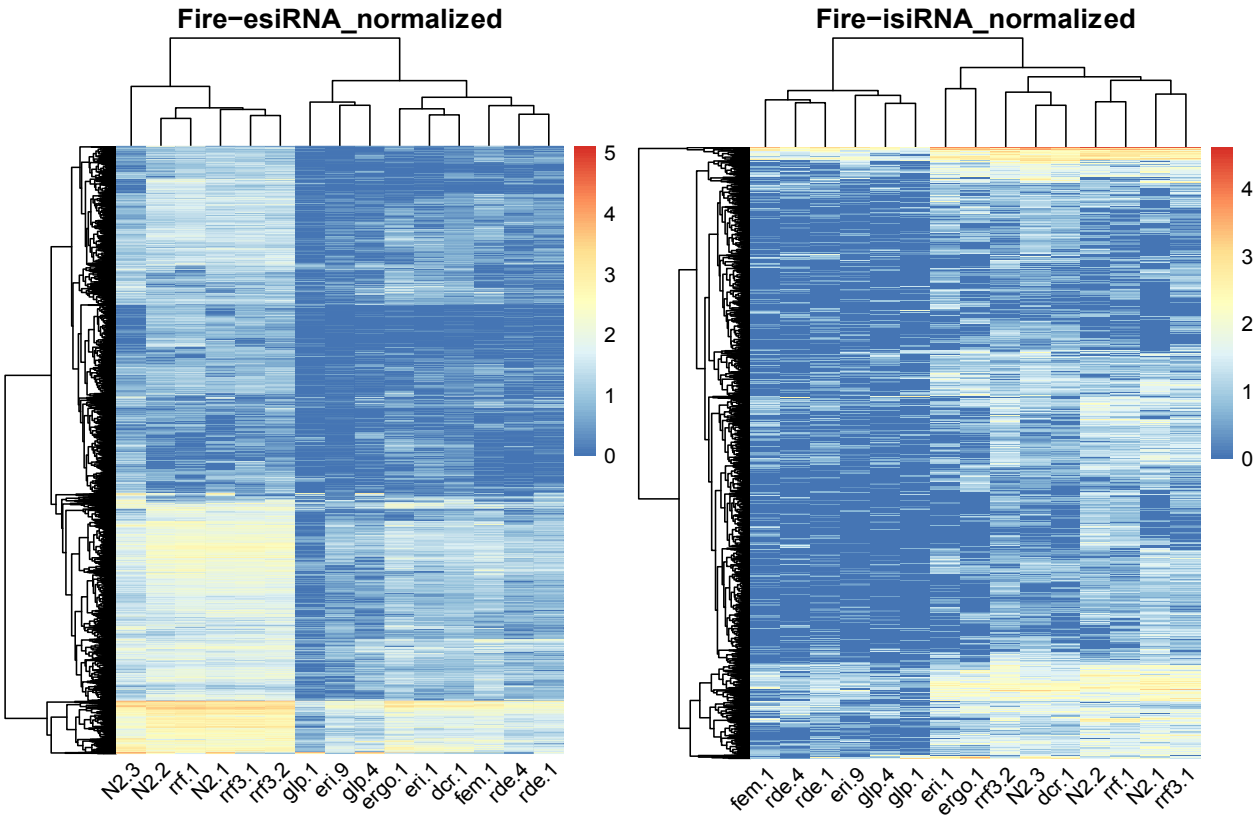


Figure 4.7 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1.

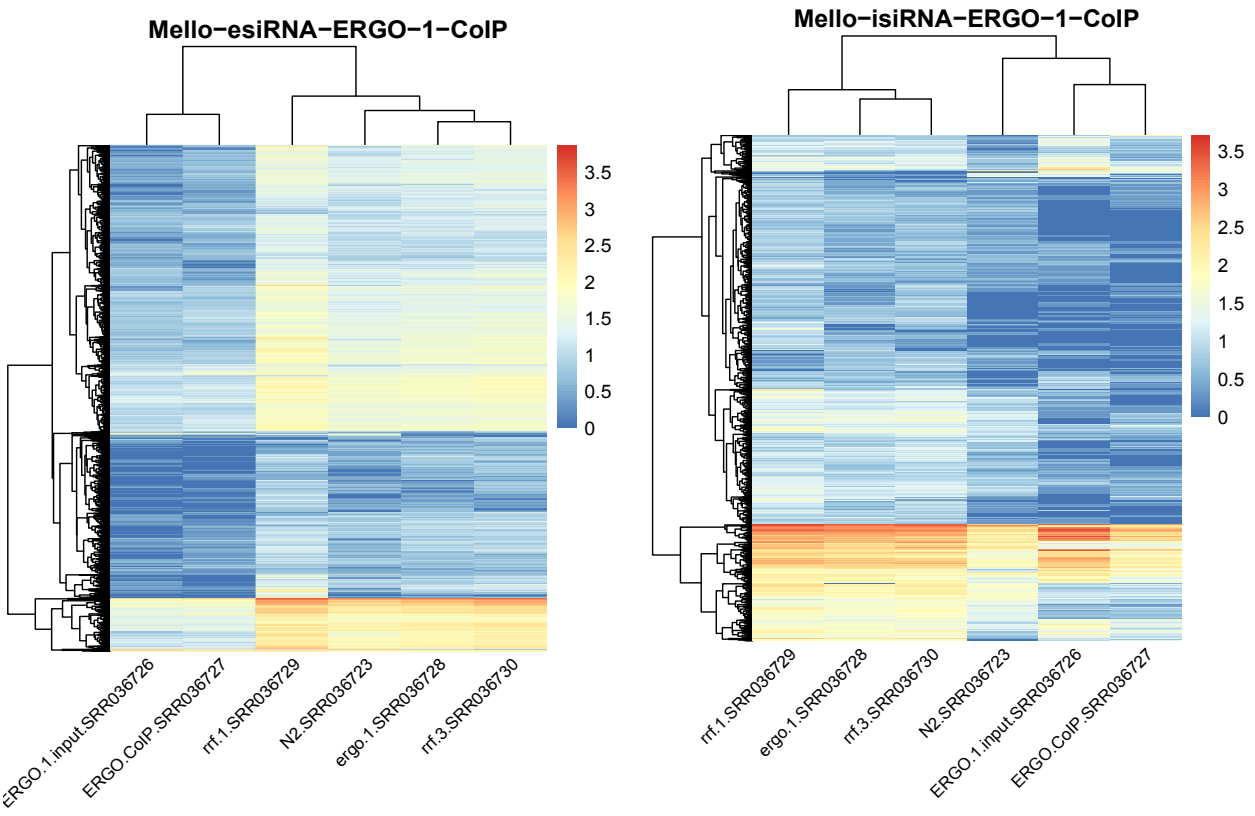


Figure 4.8 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1.

C. isiRNA may be amplified at the mutator complex which is at the periphery of P granules

The six protein factors that make up the mutator complex: MUT-2, MUT-7, MUT-8, MUT-14, MUT-15, MUT-16 co-localize in the punctate mutator foci found at the periphery of P granules. These structures reside right outside of the nuclear pore, through which most mRNA exits the nucleus [51]. This mutator complex is essential for mRNA quality surveillance, siRNA amplification and silencing of transposons [51]. Central to the mutator focus is the Q/N rich, intrinsically disordered protein, MUT-16, which is essential for mutator focus structure and function [56] (Figure 4.9). WAGO - bound 22G esiRNAs have been shown to be depleted in the *mut-2(ne298), mut-7(pk204), or mut-16(pk710)* genetic backgrounds [13]. We found that isiRNA levels are depleted in the mutator-defective *mut-2(ne298), mut-7(pk204), and mut-16(pk710)* strains (Figure 4.3, and figure 4.4). This indicates that these small RNA that map anti-sense to introns may be important for mutator function. Thus, they may be required to silence the transposons that are often harbored within long introns. This may also indicate that siRNA-mediated silencing could not only limit proteotoxic stress that accrues when introns are translated but could also maintain genomic integrity.

In keeping with these findings, when we looked at isiRNA in libraries made from *mut-15(tm1358)* worms that are defective for this relatively uncharacterized mutator focus component, we found that esiRNA was depleted slightly and isiRNA was depleted more. This indicates that *mut-15(tm1358)* is more important for isiRNA (Figure 4.3)

The DEAD box RNA helicase MUT-14 is also found in mutator foci where it may function with SMUT-1, another possibly redundant DEAD box RNA helicase. *mut-14(pk738)* is a dominant negative allele of MUT-14 that fails to carry out RNAi in the germline but does not affect somatic RNAi at all. *mut-14(mg464),* which is likely a null allele, does not affect either germline or somatic RNAi [52]. We found that isiRNA level is depleted in the *mut-14(pk738)* strain which is defective for RNAi in the germline. However, isiRNA levels were not affected in the

mut-14(mg464) strain which cannot carry out somatic RNAi (Figure 4.10). This is also what was seen when esiRNA levels were examined. The production of isiRNA is independent of somatic mutator function but is dependent on the germline mutator function. This is consistent with our hypothesis that isiRNA are produced mostly in the germline.

The *smut-1(tm1301)* mutant showed no change in either esiRNA or isiRNA. The double mutant strains *mut-14(pk738);smut-1(tm1301)* and *mut-14(mg464);smut-1(tm1301)* however, show decreased levels of both esiRNA and isiRNA (Table 4.1). This indicates that these DEAD box helicases act redundantly to produce isiRNA.

The mutator focus also contains the RdRP, RRF-1, leading Phillips and Ruvkun (2014) to postulate that mutator foci are sites of siRNA amplification [52]. We find that this RdRP is not required for esiRNA or isiRNA levels in the adult but that both esiRNA and isiRNA levels are increased in *rrf-1(pk1417)* mutant embryos [21]. When we examined libraries from animals that were heterozygous: *mut-14(pk738);smut-1(tm1301) x mut-16(pk710) (hets)* we found that the levels of isiRNA were more reduced than esiRNA levels (Figure 3.11). This may indicate that in the germline, isiRNA production is more dependent on the full functioning of these mutator genes. We also found that loss of the P-granule associated Argonaute, WAGO-1 in the context of the MAGO12 (missing 12 Argonautes mutant, which included WAGO-1) showed a large decrease in isiRNA but no change in esiRNA levels. Thus, the two species of small RNA may have differentiated and dedicated handling biosynthesis machinery at the mutator complex in the germline though they are otherwise quite similar.

The function of mutator foci is to silence transposons within the germline, to promote RNAi and to provide quality control for mRNA leaving the germline nucleus. When the factors within the mutator foci are lost, animals accumulate high rates of mutations due to unchecked transposon insertion into genes. We predict that in the mutator strains, mRNA with retained introns would be found at higher levels since isiRNA would not be able to provide surveillance and removal of these transcripts or transcriptional silencing of poorly spliced genes. We further

propose that in the MAGO12 mutant strain, silencing of transposons within large introns would also be decreased.

P granules are found adjacent to the mutator foci in the germline. The piRNA biosynthetic pathway resides within these P-granules [55,57]. In species ranging from *M. musculus* to *C. elegans*, piRNAs are required for germline integrity by silencing transposons in worms and maintaining male fertility in mice, zebrafish and worms [58–60]. In both *M. musculus* and *C. elegans*, piRNA is the most abundant small RNA species in the germline. In *C. elegans*, PRG-1 and PRG-2 are germline specific piRNA (piwi) Argonautes. Loss of PRG-1 resulted in development defects that included: abnormal germline, defective spermatogenesis, and increased sensitivity to temperature [57,61,62]. More specifically, mutations in PRG-1 cause defects in meiosis and mitosis in the germline [57]. Additionally, PRG-1 acts upstream of MUT-7 in endogenous siRNA pathway, and it is posited that the endogenous RNAi pathway collaborates with the PRG-1-dependent piRNA pathway to silence piRNA targets such as the transposon, Tc3 [57]. Indeed, the endosiRNA pathway including MUT-7 is required for accumulation of siRNA downstream of piRNA-silenced loci [57].

Because PRG-1 and PRG-2 are the two factors whose functions are specific to the piRNA pathway, we decided to examine the esiRNA and isiRNA in PRG-1 mutant strains. We found that there are reduction in levels of both esiRNA and isiRNA in *prg-1(n4357) x prg-1(n4357)*, and the heterozygous mutant *prg-1(n4357);mut-14(pk738);smut-1(tm1301) x prg-1(n4357);mut-16(pk710)* (Figure 4.11). These siRNA reductions may indicate that intronic RNA could be a template for the piwi pathway, it could also result, non-specifically, from perturbations to germline integrity.

isiRNA production is strongly dependent on both the mutator foci and the P-granules in the germline. Each of these structures sit at the nuclear pore and are important sites of surveillance that could allow detection and possibly destruction of poorly spliced mRNA. I

suggest that isiRNA is amplified within these structures thereby minimizing the number of transcripts with retained introns that exit the nucleus.

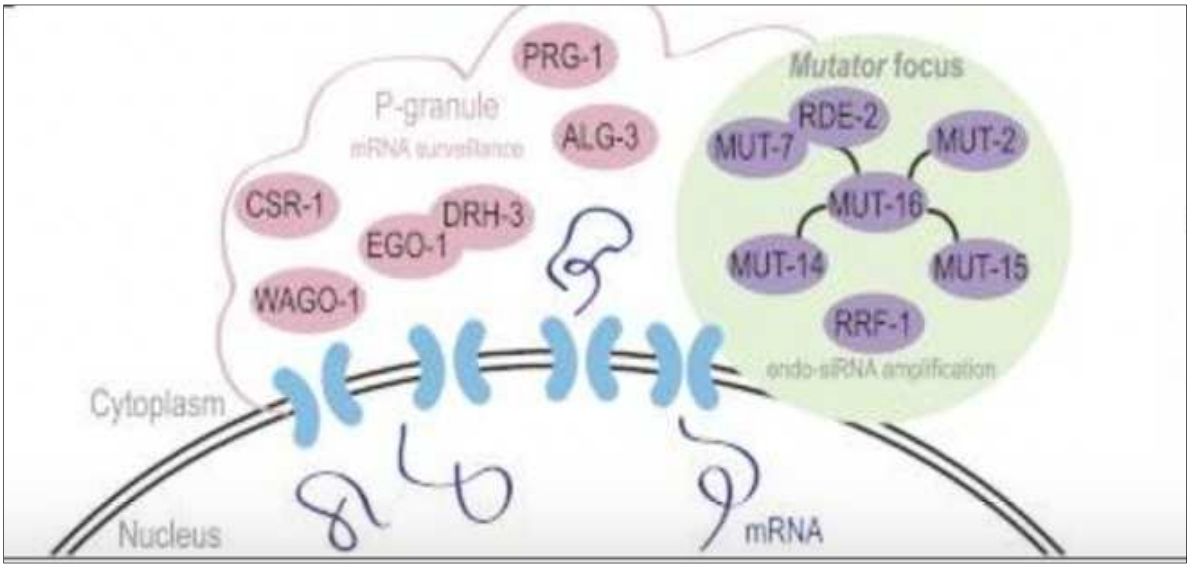


Figure 4.9 Model showing the composition and localization of mutator foci and P granules adjacent to nuclear pore. Philips et al., 2012

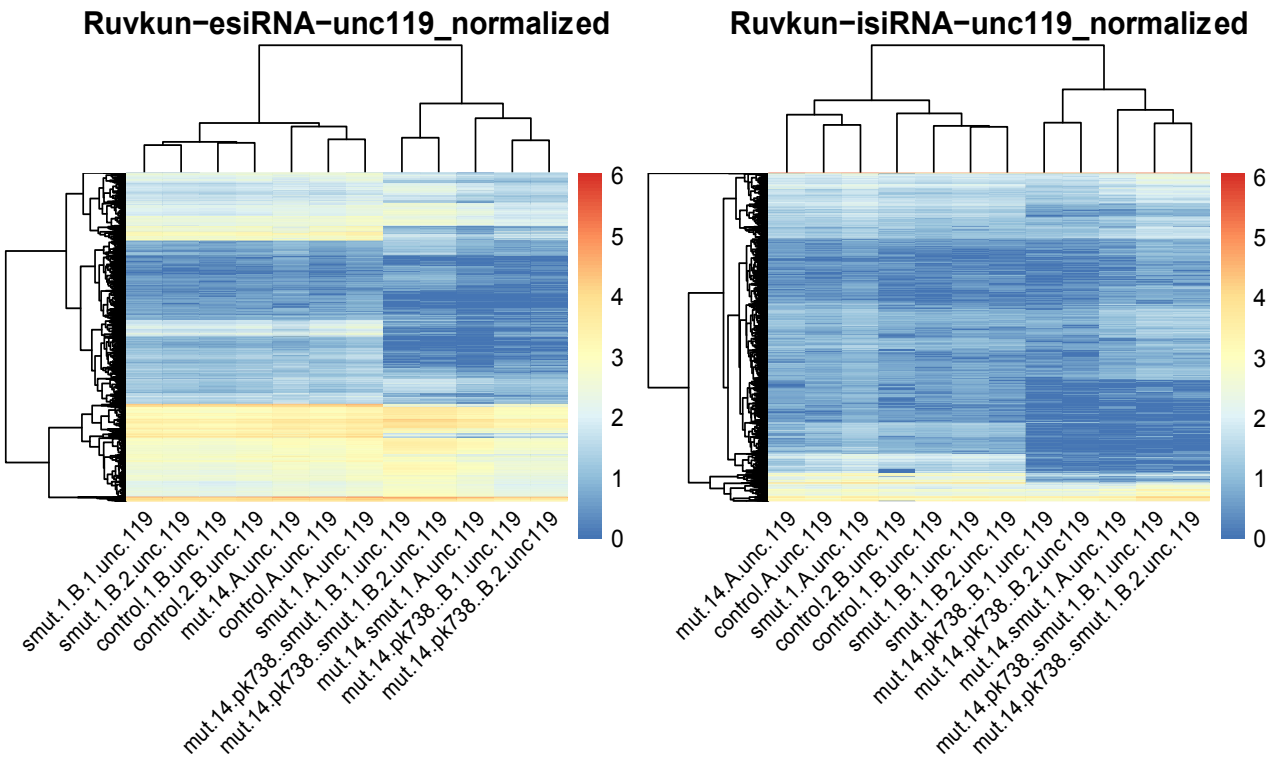


Figure 4.10 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1.

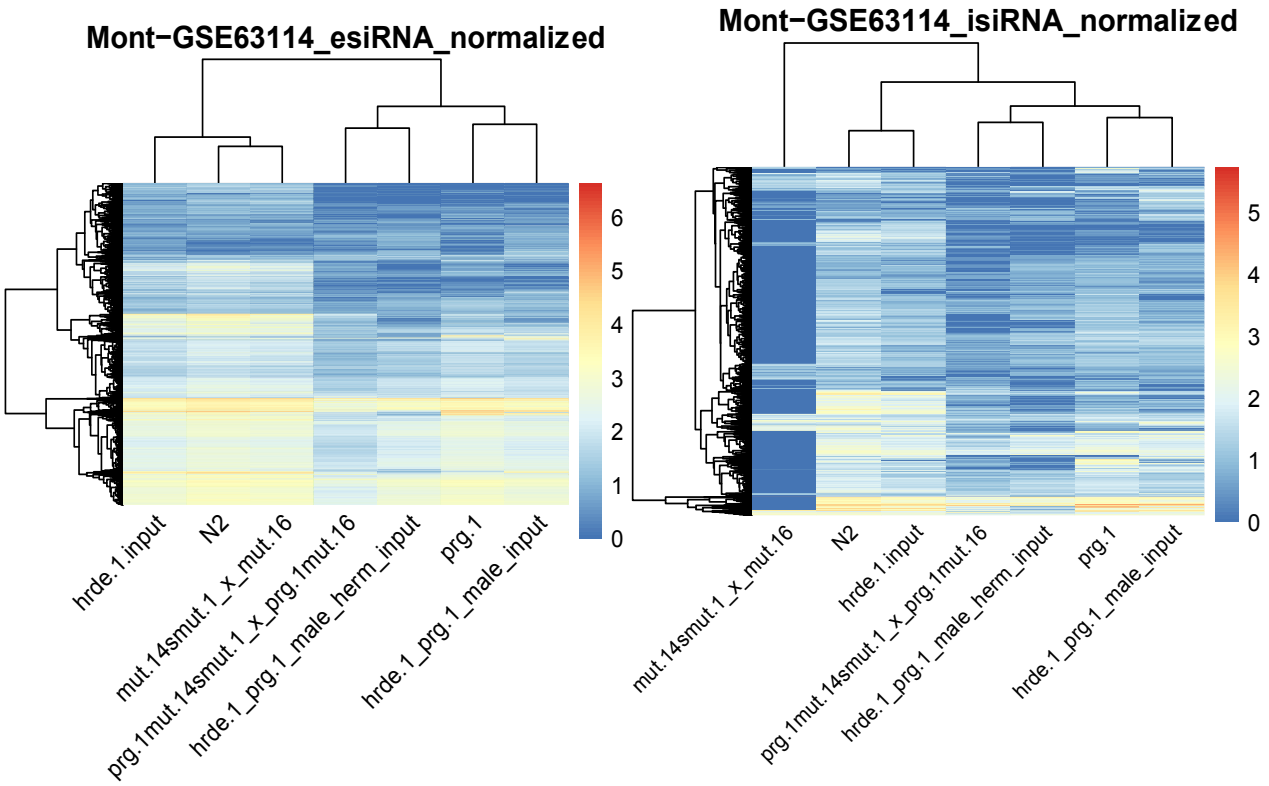


Figure 4.11 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1.

D. isiRNA levels are dependent on many exo-RNAi factors

Mello and Fire found that exogenously provided double stranded RNA was able to trigger silencing of genes that had homology to the dsRNA that were expressed in *C. elegans* [12]. This was called exogenous RNA silencing (exo-RNAi), and screens for genes that blocked this silencing identified: *rde-1, rde-2, rde-3, rde-4, mut-7, mut-2* [63].

As mentioned earlier, four main small RNA species mediate silencing in *C. elegans*: microRNAs (miRNA), piRNA, exogenous RNAi (exo-RNAi), and endogenous RNAi (endo-siRNAs). Subsequent work showed that production of these small interfering RNA species all depend on the exonuclease III, Dicer (DCR-1) for production except for piRNA [57]. Biosynthesis of each species is differentiated from one another by its trigger, and DCR-1 is complexed with a different set of sometimes overlapping factors in each biosynthetic pathway [64]. More specifically, in the miRNA biosynthesis pathway, DCR-1 interacts with ALG-1/2 to produce miRNA and silence mRNA that have miRNA binding sites and these are usually in the 3'UTR [10,65,66]. In exo-RNAi biosynthesis machinery, DCR-1 interacts with RDE-4, RDE-1, and DRH-1/2 to process exogenously provided dsRNA to produce primary siRNA [67]. These primary siRNAs are bound by Argonautes and they template production of secondary siRNA. In the secondary siRNA production process, DCR-1 associates with PIR-1 in complex with an RNA-dependent RNA polymerase (RdRP) complex to amplify secondary siRNA with mRNA serving as a template [64]. In the endo-RNAi biosynthesis pathway, DCR-1 forms a complex with Enhanced for exogenous RNAi (ERI) proteins and a DICER related helicase (DRH-3) to produce primary siRNA. Downstream of these primary siRNAs, DRC-1 associated with the RNA-dependent polymerase, RRF-3 produces secondary siRNA [64,67].

DCR-1, associated with the Argonaute RDE-1 and the double stranded RNA binding protein RDE-4, produces exo-RNAi, which triggers RNA silencing in response to exogenously provided dsRNA [68]. RDE-4 is important in recognizing and invoking the cleavage of dsRNA to produce primary siRNA [68], which is the starting material for both WAGO and CSR-1 mediated

pathways. Consistent with this study, we also observed a decrease in level of esiRNA and isiRNA in *rde-4(ne299)* mutants, which support our hypothesis that similar to esiRNA, isiRNA machinery is also dependent on the exo-RNAi biosynthesis machinery.

In the study by Parrish and Fire (2001), loss of RDE-4 significantly reduced the levels of both ~26 and 22 nt siRNA species that were triggered by exogenously provided dsRNA. However, loss of RDE-1 only affected the 22nt fraction [68]. This means that RDE-1 acts downstream of the 26nt primary siRNA that was formed in the exo-RNAi pathways [68]. Our analysis shows that both esiRNA and isiRNA levels decrease in both *rde-1(ne300)* and *rde-4(ne299)* mutants (Figure 4.7). This could mean that both Argonautes are required for production of isiRNA in the same way that they are required for esiRNA production.

E. isiRNA is independent from endogenous RNAi biosynthesis pathway

Loss of genes that enhanced exogenously triggered RNAi were called Enhanced for RNAi or ERI. DCR-1 protein is limited but required for multiple RNAi pathways, therefore mutations of ERI factors are thought to free DCR-1 to enter other RNAi silencing pathways [64]. DCR-1 associates with ERIs to form ERI/DICER complex to process double stranded RNA and produce the initiating ingredient for both 22G and 26G classes of endogenous RNAi pathways [31,69]. ERI/DICER-complex is a complex of proteins comprises of exonuclease ERI-1, ERI-3, tudor-domain ERI-5, RdRP RRF-3, and endoribonuclease DCR-1 to mediate the RNAi process to negatively regulate mRNA expression [31,64].

dcr-1(mg375) mutant has alteration of the conserved residue within the DCR-1 helicase domain [70]. *dcr-1(mg375)* strain has recently shown to carry a mutation upstream of *mut-16* gene in its genetic background [31]. DCR-1 in *dcr-1(mg375)* loses its contribution to ERI but retains its function in miRNA, piRNA and exoRNA pathways and allows viability [70].
Our analysis showed a very interesting finding that in *dcr-1(mg375)* and *eri-1(mg366)* mutants only esiRNA levels were reduced while isiRNA levels were unchanged. Perhaps, like the germline siRNA, which are shown to have no visible change in its levels in *dcr-1(mg375)*, *eri-1(mg366)* or *ergo-1(gg098)* mutants [13], isiRNA biosynthesis is independent from the ERI pathway.

In the Figure 4.12, isiRNA production does depend on the ERI complex member ERI-9. This putative RNA transferase, ERI-9 is a novel DICER associated protein [70] which is required for endogenous RNAi: ERI-9 mutant strain *eri-9(gg106)* was shown to fail to complement RNAi enhance phenotype, and ERI-9 is required for endo RNAi process in sperm [70]. This could indicate that isiRNA are produced from a complex that is a blend of exo-RNAi associated DICER proteins and some endo-siRNA factors such as ERI-9.

DRH-3 is a DICER related helicase that was shown to be required for RNAi [64]. DRH-3 was shown to be essential for biosynthesis of 22Gs RNA in two major pathways: WAGO and

CSR-1. In the WAGO pathway, DRH-3 requires RDE-3 and MUT-7 as associating factors, while in CSR-1 pathway, DRH-3 requires CDE-1 as a cofactor [13]. DRH-3 is shown to interact with RRF-1 to promote propagation of siRNA biosynthesis by RdRP along the template mRNA, which initiate at 3' terminus [13,71].

Our analysis showed that the level of esiRNA and isiRNA are both decreased in mutant strain *drh-3(ne4253)*. This depletion might be due to the role disruption of DRH-3 in WAGO biosynthetic pathway.

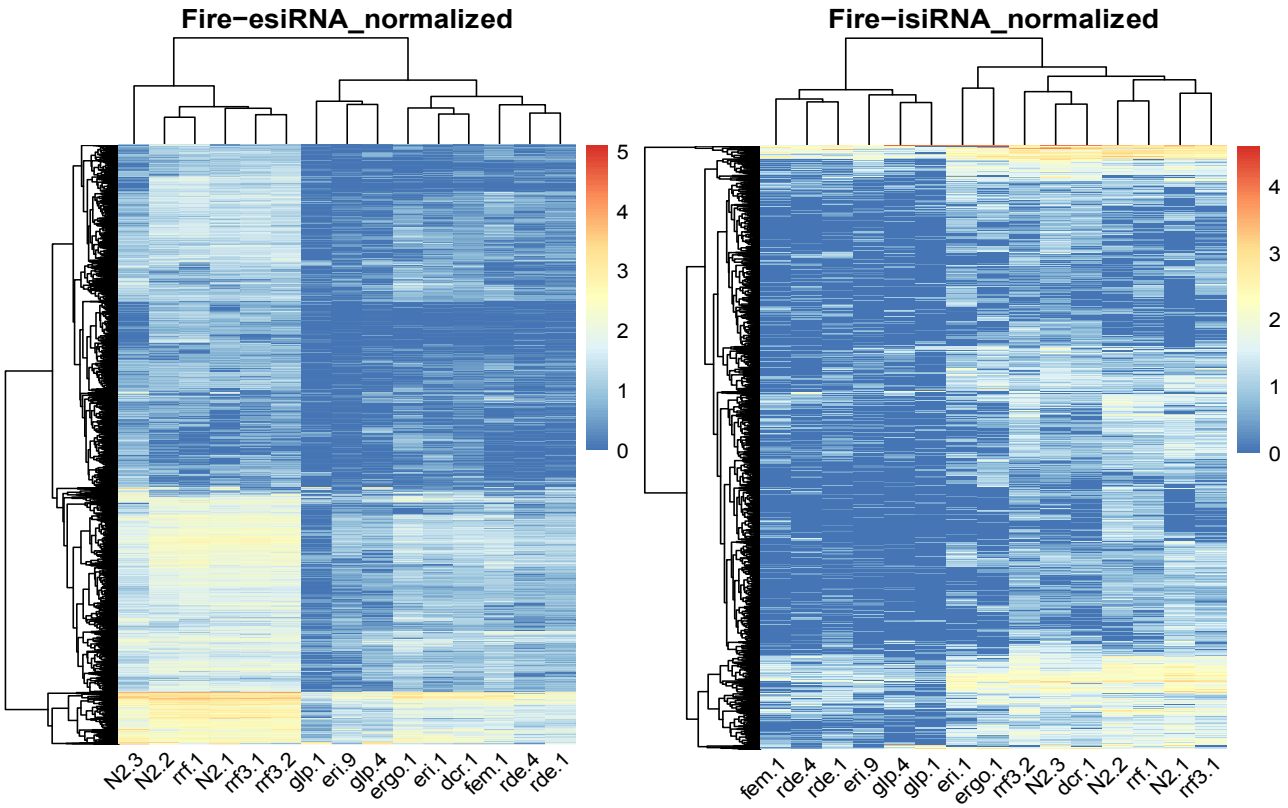


Figure 4.12 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1.

F. isiRNA are tertiary RNA

In *C. elegans*, siRNA is divided into 3 distinct classes: primary siRNA, secondary and tertiary siRNA. The primary siRNA are 26 nucleotides and originate from exogenous dsRNA [15,64,72]. These primary siRNA could then trigger the recruitment of RNA dependent RNA polymerase (RdRP) such as RRF-1, RRF-2, RRF-3, EGO-1 to produce secondary 22G siRNA [71,73]. The secondary siRNA could be imported to the nucleus by the soma specific Argonaute, NRDE-3, or the germline specific Argonaute, HRDE-1 to trigger production of tertiary siRNA in repression of gene transcription [16].

Previously we have discussed many factors that are important for post transcriptional gene silencing in response to exogenous dsRNA. Since these siRNAs target mature mRNAs, they were presumed to occur in cytoplasm [12]. However a genetic screen for factors that defective in RNAi silencing identified some genes that important for transcriptional silencing and that siRNA can target pre mRNA [74], process that carried out in the nucleus [23,75]. These factors are then called Nucleus RNAi Defective, or the NRDE for short.

The nuclear specific RNAi factors NRDE-1/2/4 all work with nuclear Argonautes such as NRDE-3 for somatic tissue or HRDE-1 for germline tissue to set up repressive chromatin structures that silence genes [22,23,76]. SET-25 is a histone methyltransferase, which catalyzes H3K9me1, me2, me3 to establish chromatin repression [77]. SET-25 is required for novel RNA insertion or DNA transposon in two redundant pathways: one required MET-2, (SET-25–MET-2–LIN-61), and the other required soma specific Argonaute NRDE-3 and siRNA, (SET-25–NRDE-3) [77]. The double mutant *nrde-3(tm1116);met-2(n4256)* indicated strong embryonic lethality phenotype) [77].

Our analysis shows that the production of esiRNA and isiRNA are reduced in mutant strains of this repressive complex member such as: *nrde-1(gg088), nrde-4(gg129), nrde-3*, and the double mutant *nrde-3(tm1116);met-2(n4256)* (Figure 4.13, and Figure 4.14). However, there is no change in esiRNA and isiRNA level in single mutant *nrde-3(tm1116), set-25(n5021) or*

double mutant at early embryo stage *set-25(n5021);nrde-3(tm1116)* (Figure 4.13 and Figure 4.14). The fact that isiRNA depending on NRDE-1/4, additionally isiRNA and tertiaries siRNA are both independent of the ERI-1, and *dcr-1(mg375)* means that some if not most of isiRNA are tertiary siRNA.

In summary, there is an accumulation of evidence showing that germline is important for isiRNA production. However, we do not know if isiRNA are germline specific. We also observed in one library that the level of isiRNA was disturbed in the mutation strain of a somatic specific Argonaute, NRDE-3 making us wonder if there are low levels of somatic isiRNA or whether NRDE-3 is required for the intact germline. In order to probe isiRNA levels in the germline in the future, we suggest that techniques such as RNA Fluorescence In Situ Hybridization visualization with isiRNA specific probes in theory could validate the presence of isiRNA in the germline. It could also reveal if these RNA species are found in the soma. Since the germline makes up one half of the cellular mass of the adult worm, one could also make small RNA libraries specifically from the germline. This may have been done and these libraries would be interesting to put through our analysis pipeline.

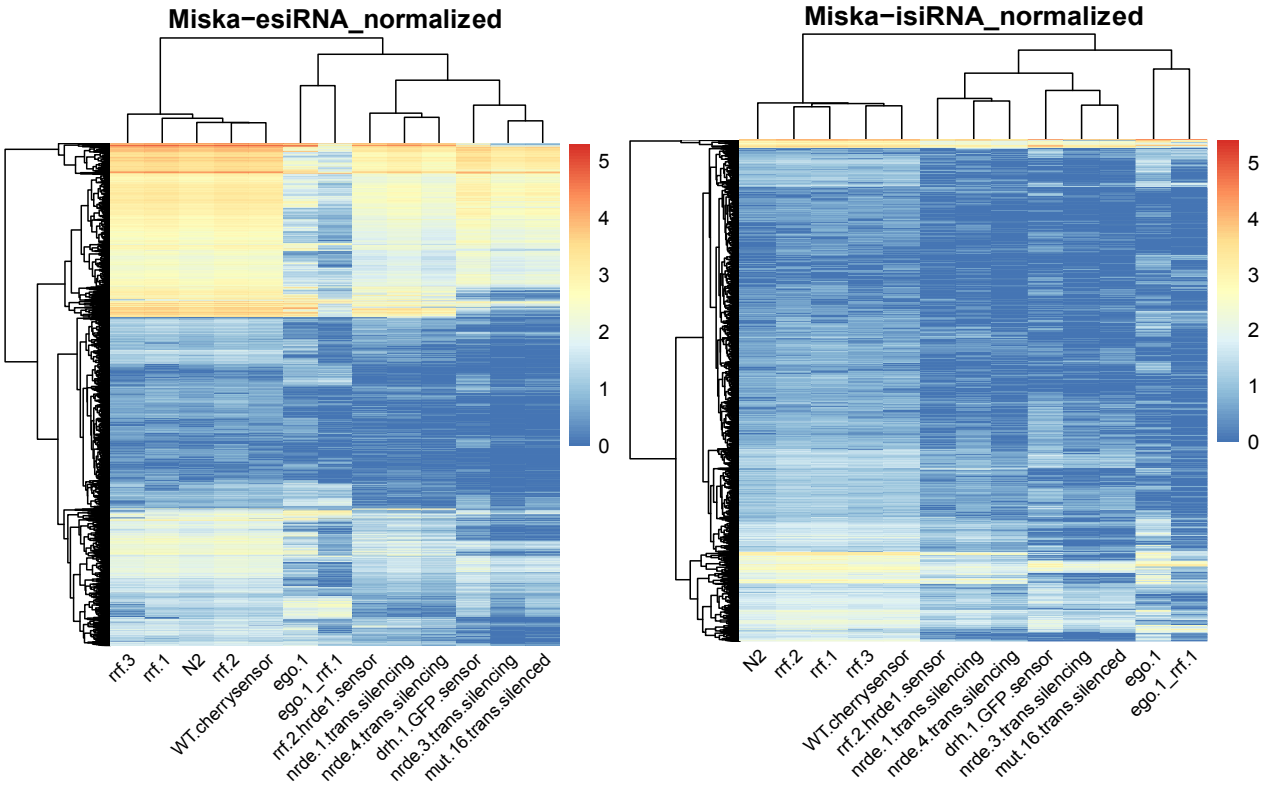


Figure 4.13 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1.

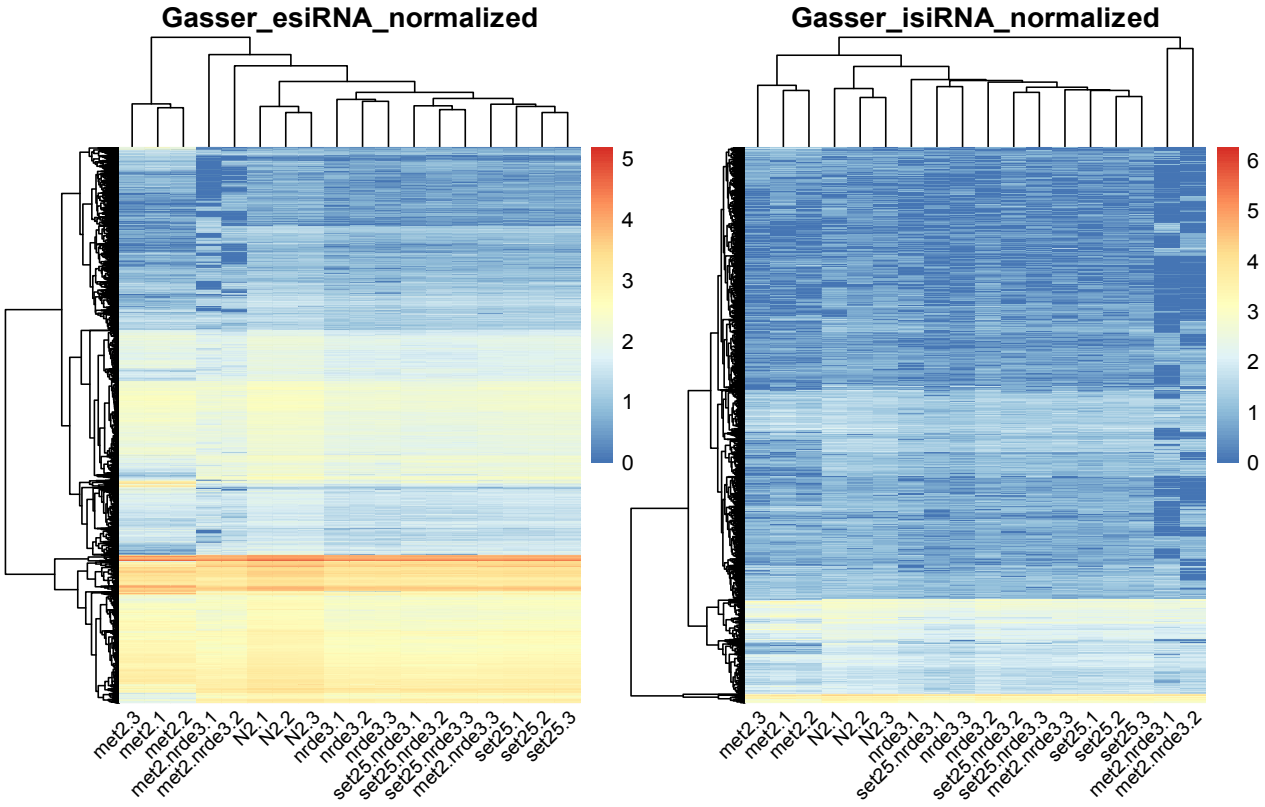


Figure 4.14 Heatmap of log10 normalized esiRNA (left) and isiRNA reads (right). X-axis libraries. Y-axis genes cluster. In each graph, only genes with at least two reads per million are indicated. The color scale: blue is low number of reads, red is high number of reads. Normalization formular please see Table 4.1.

G. There are some genes that have more isiRNA mapped to them than esiRNA and mutation of three specific factors that are required ERI (enhanced exogenous RNAi) increases the numbers of these genes.

Most of the siRNA factors affect the level of both esiRNA and isiRNA, however some of the mutants seem to reduce esiRNA level significantly but impose less reduction to no change on isiRNA level such as: *dcr-1(mg375), ergo-1(gg098), eri-1(mg366)* [31](Figure 4.7). This led me to the question that these mutations cause reduction in a subset of esiRNA or to the overall esiRNA level. To answer this question, I calculate the number of genes that have more isiRNA than esiRNA mapped to them (isiRNA>esiRNA genes) in wildtype animals and compare to that in those mutants. I found that in the wild type libraries, there are 436 isiRNA>esiRNA genes. Interestingly, in those mutants that only have reduction in esiRNA level and unchanged in isiRNA levels, I found that there is an increasing number isiRNA>esiRNA genes. More specifically, in *dcr-1(mg375), ergo-1(gg098), eri-1(mg366)* there are 701, 593, and 694 isiRNA>esiRNA genes respectively (Figure 4.15). This indicates that missing those proteins affect only a subset of esiRNA, perhaps those are specific populations of esiRNA that come from ERGO biosynthetic pathways. This result also suggests that either the level of esiRNA is decreased or there is an increase in the level of isiRNA in those mutants.

Here in this section, I describe some phenotypes of those mutants that have increased number of isiRNA>esiRNA genes. The *mg375* allele of DCR-1 lost its contribution to ERI function but retains miRNA, piRNA, exogenous RNAi function, and allows viability *)* [31]. ERGO-1 is an argonaut that binds and stabilizes 26Gs, the primary endogenous siRNA *)* [31]. ERGO-1 expressed primarily in the germline and embryo [20]. ERI-1 is DEDDh-like 3' to 5' exonuclease. ERI-1 protein has two isoforms, ERI-1a, and ERI-1b. While both isoforms ERI-1a and ERI-1b are involved in 3' processing of 5.8S of ribosome RNA [78], only ERI-1b is required for 26Gs RNA biosynthesis via interaction with DCR-1 [78]. Mutant *eri-1(mg366)* has insertion of 23 base pairs

on exon 6, and has shown a complete lack of 26nt siRNAs, but does not affect the global level of 22G siRNA [20].

Perhaps, mutating *dcr-1(mg375), ergo-1(gg098), eri-1(mg366)* leads to enhanced exogenous RNAi this is postulated to be because the endogenous RNAi machinery is not functional and thus double stranded RNA from outside the worm can trigger a larger RNAi response. We interpret the increasing in number of isiRNA>esiRNA genes in those mutants that like exogenously supplied dsRNA, the isiRNA trigger engages a different complex of the siRNA machinery, and this resulting in increasing the production of isiRNA, while the production of the esiRNA in those mutants are reduced.

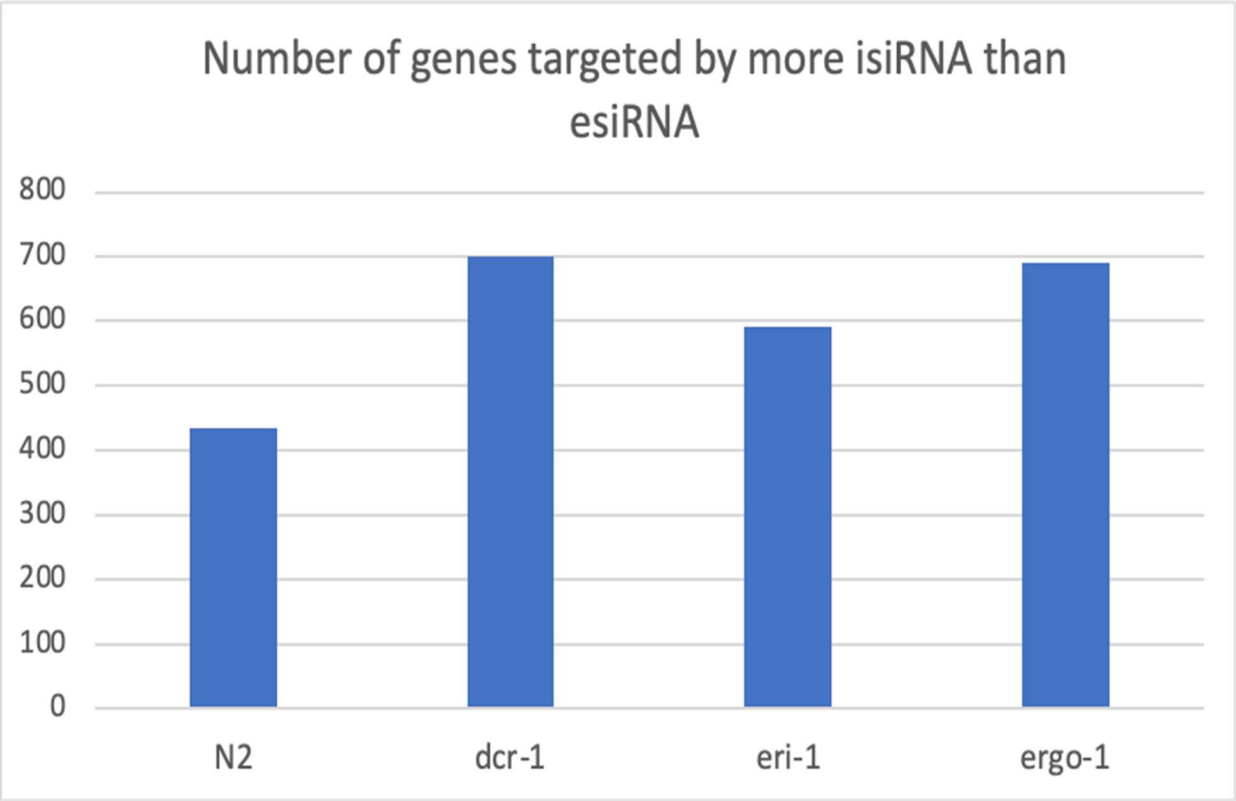


Figure 4.15 Number of gene that produced more isiRNA than esiRNA. Y-axis is gene number. X-axis is mutant strains.

CHAPTER FIVE: INSIGHTS INTO POSSIBLE FUNCTIONS OF isiRNA BY EXAMINING ARGONAUTE ASSOCIATED SMALL RNA

INTRODUCTION

As mentioned in the previous chapter, siRNA can regulate gene expression post-transcriptionally in the cytoplasm or transcriptionally in the nucleus and they do so by bringing protein cofactors into proximity of target cognate mRNA via their ability to bind to such sequences. Although siRNA are believed to be involved in diverse biological processes such as: defense against viral infection [79–81], epigenetic silencing [82,83], influencing germline development [84], the exact target genes and precise biological functions of siRNA are not fully understood. In this study I exploit the biochemical approaches used by other labs to ask how isiRNA might function in regulating gene expression. More specifically, by asking whether isiRNA species are present in libraries made from RNA that are co-Immunoprecipitated with known protein cofactors such as Argonautes, I hoped to infer the function of the bound isiRNA from that of the Argonautes.

Argonautes interact with small RNA to mediate gene silencing [64]. There are 27 different Argonautes in *C. elegans* and each is thought to have distinct functions in different RNAi pathways [64]. More specifically, ALG-1/2 are miRNA specific Argonautes, while RDE-1, RDE-4 are exclusive to exo-siRNA pathways, and ERGO-1 and CSR-1 are endo-siRNA specific Argonautes [64]. Argonaute protein families are keys components of RNA-Induced Silencing Complex (RISC) across different species [85–87]. Yigit et al., 2006 found that Argonautes not only participate in multiple different pathways to produce unique siRNA species to each pathway (Figure 5.1), but also that distinct Argonautes function sequentially during RNAi. This means, if there are multiple Argonautes participating in one RNAi pathway such as RDE-1 and RDE-4, the functions of downstream Argonautes (in this case is RDE-4) depends on the activity of the upstream Argonaute (RDE-1) [64]. In worms some Argonautes are germline and others somatic. More specifically, NRDE-3, SAGO-1 and SAGO-2 are somatic, while HRDE-1, CSR-1 are germline Argonaute.

Co-immunoprecipitation (CoIP) is a technique similar to Immunoprecipitation (IP), which can be used to identify physical interaction between protein-protein or between protein and other molecules such as DNA (ChIP) or RNA (RIP). In this process, a target protein specific antibody can be used to indirectly capture the target protein with other components in its complex. While IP is focused on the primary target, which is the protein that antibodies bind to, the Co-IP is the technique that focuses on identifying the secondary target, which are the molecules that bind to the primary target; it could be protein, DNA, or RNA. In this study, all of CoIP libraries are actually co-RNA Immuno Precipitated (RIP) libraries. However, to be consistent with the terminology that most of the libraries were referred to in their original study, I will also refer to them as Co-IP libraries in my study.

Here I briefly describe the main steps from preparing the sample to sequencing the Co-IP library. First, the animals are synchronized to the same stage of development before lysis for protein extraction. In this method, the protein-RNA complexes are immunoprecipitated with antibodies targeted to the primary protein. RNA was then extracted in both input and the Co-IP. Total RNA was treated with tobacco acid phosphatase to digest triphosphates and diphosphates to monophosphates. RNA was size-selected between 15 - 35 nt on TBE gel. Then the standard small RNA library preparation protocol was applied. In brief, the purified RNA from TBE gel was ligated to the 3' adapter then to 5' adapter. Adapter-ligated RNA was reverse transcribed and Polymerase Chain Reaction (PCR) amplified. The PCR product was then purified. Finally the amplified product will be run through gel for size selection between 140 - 165 base pairs in length before going into the sequencer [86] (Figure 5.1).

Though reports indicate that Argonaute-siRNA interactions are relatively stable (S. Kennedy personal comm), the nature of the Co-IP process requires sonicating and washing steps therefore the interaction between protein and RNA could potentially be disrupted and those siRNAs would be missing from the libraries. Therefore, the amount of siRNA extracted from the Co-IP experiment would be lower than the amount of siRNA associated with the protein

of target in the cell. Because of the small amounts of RNA in the Co-IP, the number of amplification steps in the Co-IP process is always more than in the input to the IP or in the other libraries mentioned in previous chapters.

NRDE-3 and HRDE-1 are both Nuclear RNAi Argonautes. While NRDE-3 is soma specific Argonaute, HRDE-1 is specific to the germline. Both NRDE-3 and HRDE-1 contain two bipartite nuclear localization signals (NLS) and once they are triggered by 22G, the Argonautes will shuttle the 22G to the nucleus, where they can recruit other Nuclear RNAi factors such as NRDE-1/2/4 to load on nascent strand of mRNA to stall its transcription, direct heterochromatin deposition and produce more siRNA [23,75,88]. NRDE-3 and HRDE-1 accumulate in the nucleus in the presence of siRNA, and in the absence of siRNA they are found predominantly in the cytoplasm [23,89].

CSR-1 is a germline specific Argonaute, and it has been shown to have multiple functions. For instance, Claycomb and Mello et al., 2009 showed that CSR-1 promotes transcription of germline specific genes that are required for chromosome segregation [19], and Singh et al., 2021 reports that CSR-1's slicing activity is required to degrade transcripts that would be poorly translated [54]. CSR-1 also can reactivate genes that have been silenced through PGR-1 and mutator activity [90]. CSR-1 also shown to localize together with other components of endo-siRNA and pi-RNA pathways such as DRH-3, EGO-1, ALG-3, and WAGO-1 and PRG-1 near nuclear pore in P granule to surveil mRNA species as they exit the nucleus [51]. Whatever its mechanism of action, CSR-1 is extremely important and loss of CSR-1 activity causes embryonic lethality and other severe developmental defects in other stages [19,64].

In *C. elegans*, PRG-1 and PRG-2 are germline specific piRNA (piwi) Argonautes. Loss of PRG-1 results in development defects that include: abnormal germline, defective spermatogenesis, and increased sensitivity to temperature [57,61,62]. More specifically, mutations in PRG-1 cause defects in meiosis and mitosis in the germline [57]. Additionally, PRG-1 acts upstream of MUT-7 in the endogenous siRNA pathway, and it is posited that the endogenous

RNAi pathway collaborates with the PRG-1-dependent piRNA pathway to silence piRNA targets such as the transposon, Tc3 [57].

ERGO-1 is the Argonaute that binds and stabilizes the 26Gs generated in the oogenic germline [20,21]. ERGO-1 is highly expressed during the embryonic stage, but in L3, L4, and young adult stages its expression is greatly reduced [20,21]. Loss of ERGO-1 results in the enhanced for exogenous RNAi (Eri) phenotype possibly because ERGO-1 competes for the same 26G RNA species or other shared factors such as dicer that are required for both endogenous (ERGO-1 dependent) and exogenous RNA triggered silencing [64,70].

WAGO-1 is a cytoplasmic Argonaute. WAGO-1 is highly expressed in mature spermatids and throughout the germline [91]. WAGO-1 is localized in the P granule, and silences transposon, pseudogenes, and cryptic loci [13]. Loss of WAGO-1 (Quintuple AGO mutant) showed a significant reduction in germline 22G RNA, which suggested that WAGO plays an important role in the germline 22G-RNA function [13]. Indeed, WAGO-1 along with other factors such as RDE-3 and MUT-7 are required for germline esiRNA silencing pathway [13].

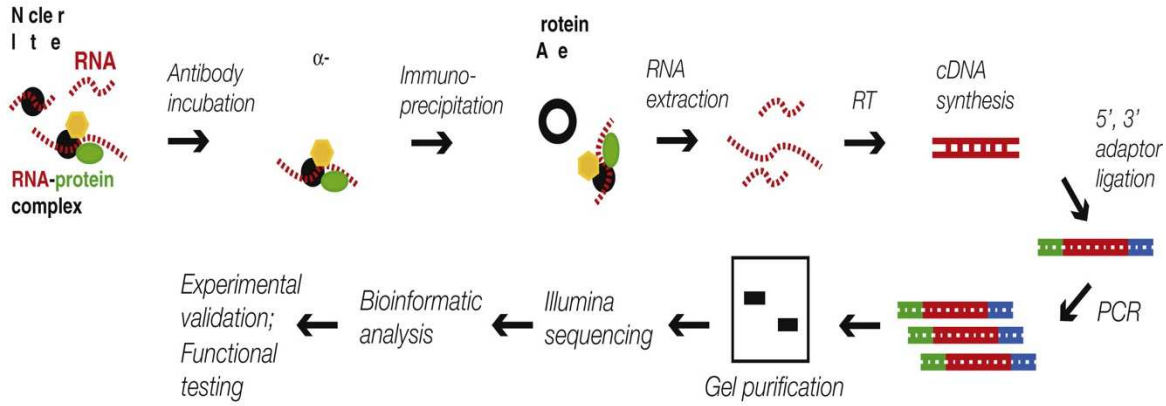


Figure 5.1 RIP-sequence schematic. Image from: 10.1016/j.molcel.2010.12.011

RESULTS AND DISCUSSION

A. HRDE-1 and CSR-1 may compete for isiRNA. The siRNA species bound by HRDE-1 are likely used to repress gene expression and possibly prompt their own amplification via mutator-WAGOs-HRDE-1 feedback loop.

In the germline, 22G RNAs that are produced in the mutator complex bind to WAGO-1 (cytoplasmic Argonaute) and PPW-1 (WAGO-7, a cytoplasmic Arogunaute) and HRDE-1 (a nuclear Argonaute) [84]. In a wild-type genetic background, 22G RNAs can engage with two pathways: CSR-1 to stimulate or HRDE-1 to repress gene expression (Figure 5.2). The HRDE-1 bound 22G RNA species repress gene expression by setting up repressive chromatin. The 22G RNA guides seem to be processed and possibly amplified in the mutator complex-WAGOs-HRDE-1 feedback loop. Albuquerque et al., 2015 showed that loss of PRG-1 shuts down this mutator-WAGOs-HRDE-1 feedback loop. In addition, they also proposed a model in which CSR-1 prevents the 22Gs produced by this pathway from loading onto HRDE-1 [84].

Our analysis of a *prg-1* library from an independent study by Philips et al., 2014 supports the model proposed in Albuquerque et al., 2015 above. More specifically, we found that libraries made from *prg-1* mutants have lower levels of esiRNA and isiRNA compared to libraries from wildtype animals (table 1). The decreased levels of esiRNA and isiRNA in *prg-1* mutants shows their biosynthesis is dependent on a piRNA Argonaute, and might be via the feedback loop that is responsible for amplifying isiRNA similar to the one proposed in Albuquerque et al., 2015's study [52].

Our analysis of libraries made from *csr-1 KO* mutant strains and CSR-1 Co-IP support the model that CSR-1 is competing for 22Gs with HRDE-1. More specifically, our Co-IP analysis shows a very interesting result: in *csr-1 KO* mutant there are more esiRNA and isiRNA bound to HRDE-1 than in libraries from the wildtype. This is especially evident in sample "csr.1.KO.HRDE-1.CoIP.2" (Figure 5.3). Plus, we also found that *csr-1* knockout mutants have slightly elevated levels of both esiRNA and isiRNA. This is perhaps because CSR-1 and HRDE-

1 compete for the same 22Gs, and loss of CSR-1 shuttles these 22Gs to the HRDE-1 WAGO pathway. These shuttled small RNAs are then amplified to result in more esiRNA and isiRNAs.

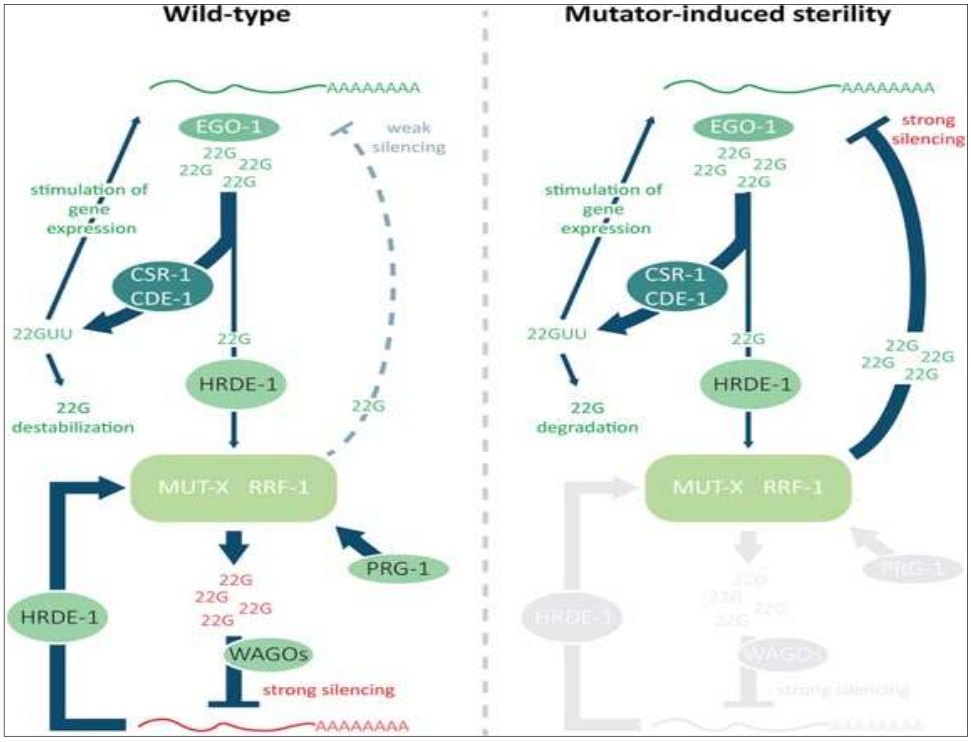


Figure 5.2 model in which CSR-1 prevents the loading of HRDE-1 and in which both PRG-1 and HRDE-1 help to keep mutator activity focused on the proper targets. Image is taken from: *Developmental Cell* 2015 34448-456DOI: (10.1016/j.devcel.2015.07.010).

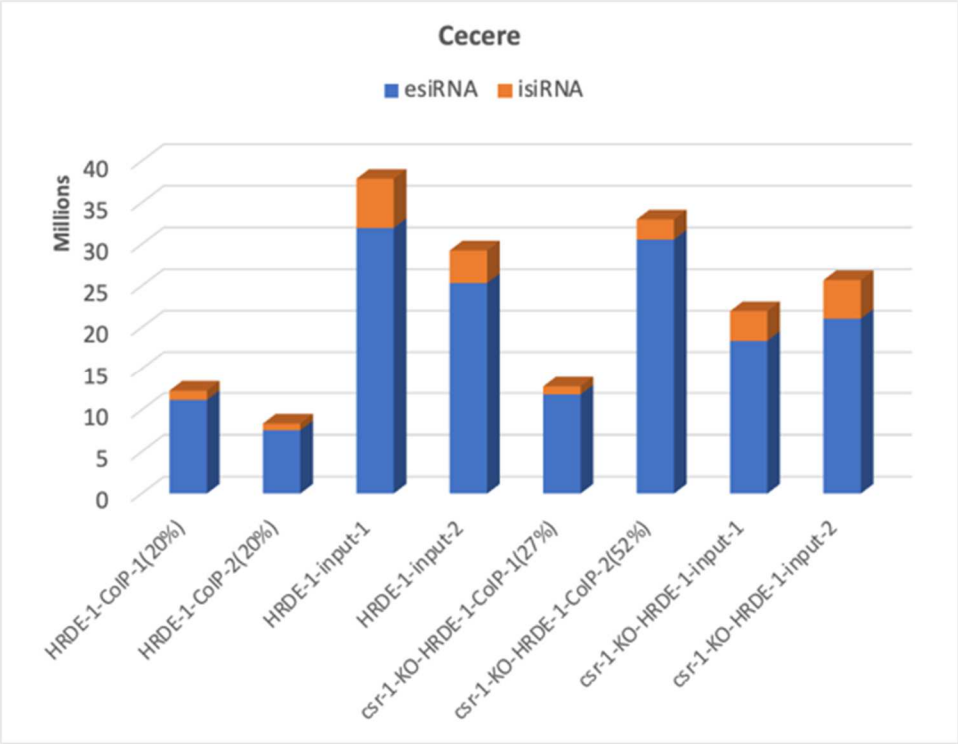


Figure 5.3 Co-IP from Cecere 2021 experiment. X-axis sequence libraries. Y-Axis small RNA read in millions. Percentage is the percentage of isiRNA input that is Co-IPed with HRDE-1

B. isiRNA binds to the germline specific Argonaute, HRDE-1 in order to repress transcription.

HRDE-1 contains two bipartite nuclear localization signals (NLS), once they are triggered by 22G binding, HRDE-1 is shuttled along with the 22G in to the nucleus, where the siRNA serves as a guide to direct the HRDE-1:siRNA complex to the nascent strand of mRNA. Once bound, they can recruit other Nuclear RNAi factors such as NRDE-1/2/4 which load on to the nascent strand of mRNA to stall its transcription, direct heterochromatin deposition and produce more siRNA [23,75,88].

In this section, we show that across four independent libraries from four different labs, isiRNAs are consistently Co-IPed with HRDE-1. I evaluated the binding of esiRNA to HRDE-1. As seen in Singh et al., 2021 [52], the percentage of input that is Co-IPed with HRDE-1 = numbers of reads per million Co-IPed with HRDE-1 divided by numbers of reads per million in input. They found that this percentage ranges between ~30% to ~145% [54]. When I performed the same analysis on the isiRNA reads, using the same formula (isiRNA CoIP with HRDE-1 = CoIP/input), I found that the percentage isiRNA Co-IPed with HRDE-1 ranged from ~20% to ~52% (Figure 5.3). These are similar results indicating that isiRNA is enriched to a similar extent as esiRNA by HRDE-1 Co-IP. Thus, HRDE-1 is likely to bind isiRNA with similar avidity as it binds the esiRNA counterpart

The details from each of the four experiments. First, in Cecere's 2021 datasets, I found that 20% of the isiRNA in the input was pulled down with HRDE-1 in both replicates (1.1, and 0.83 million isiRNA reads in these CoIP libraries) (Figure 5.3). Second, in Montgomery's 2015 datasets, 84%, and 78% of the input isiRNA was pulled down with HRDE-1 from N2 and *prg-1* worm lysates respectively (0.6 and 0.5 million isiRNA reads in N2 and *prg-1* HRDE-1 CoIP respectively) (Figure 5.4). Third, in Miska's datasets, 135% of input isiRNA was pulled down with HRDE-1 (0.2 million isiRNA reads in this CoIP library) (Figure 5.5). And fourth, in Ketting's 2015 datasets, 160% and 109% of input isiRNA was pulled down with HRDE-1 from lysates

made from N2 and *cde-1* mutants respectively (0.16 and 0.15 million isiRNA reads in N2 and *cde-1* HRDE-1 CoIP respectively) (Figure 5.6). Overall, these data provided consistent and strong evidence that HRDE-1 binds isiRNA.

Since isiRNA targets the intronic region specifically and is bound to the nuclear Argonaute HRDE-1, I propose that isiRNA function is primarily but not exclusively carried out in the nucleus, where the introns are being spliced out as mRNA is maturing. Here, I show not only that isiRNA is consistently CoIPed with HRDE-1 across different experiments carried out in different labs and in distinct genetic backgrounds, but also that isiRNA production is dependent on the nuclear RNAi factors such as NRDE-1/2/4 (Chapter Four). I hypothesize that isiRNA loading on HRDE-1 plays a role in regulating genes at the transcriptional level. I propose a model in which cellular stress disrupts splicing, indeed it is known that in cells that express the oncogene MYC, splicing is inhibited [92,93]. These un-spliced introns could potentially trigger proteotoxic stress and under these situations, isiRNA may be produced by the presence of un-spliced messages in the cytoplasm. The cytoplasmic isiRNA would then be loaded onto HRDE-1 thereby triggering nuclear transport of this complex. Once in the nucleus, it would recruit other nuclear RNAi factors such as NRDE-1/2/4 to load on to the nascent strand of mRNA where it stimulates heterochromatin deposition that disrupts transcription. Besides inducing proteotoxic stress if they are translated by virtue of nonsense codons and misfolded proteins, retained introns may also harbor a depress selfish DNA elements such as transposons. In the nucleus, isiRNA produced from these introns would then allow the cell to turn off transcription of these introns and thus limit expression of transposons. In these ways, isiRNA in the cytoplasm and nucleus may protect the cell from the adverse effects of un-spliced messages.

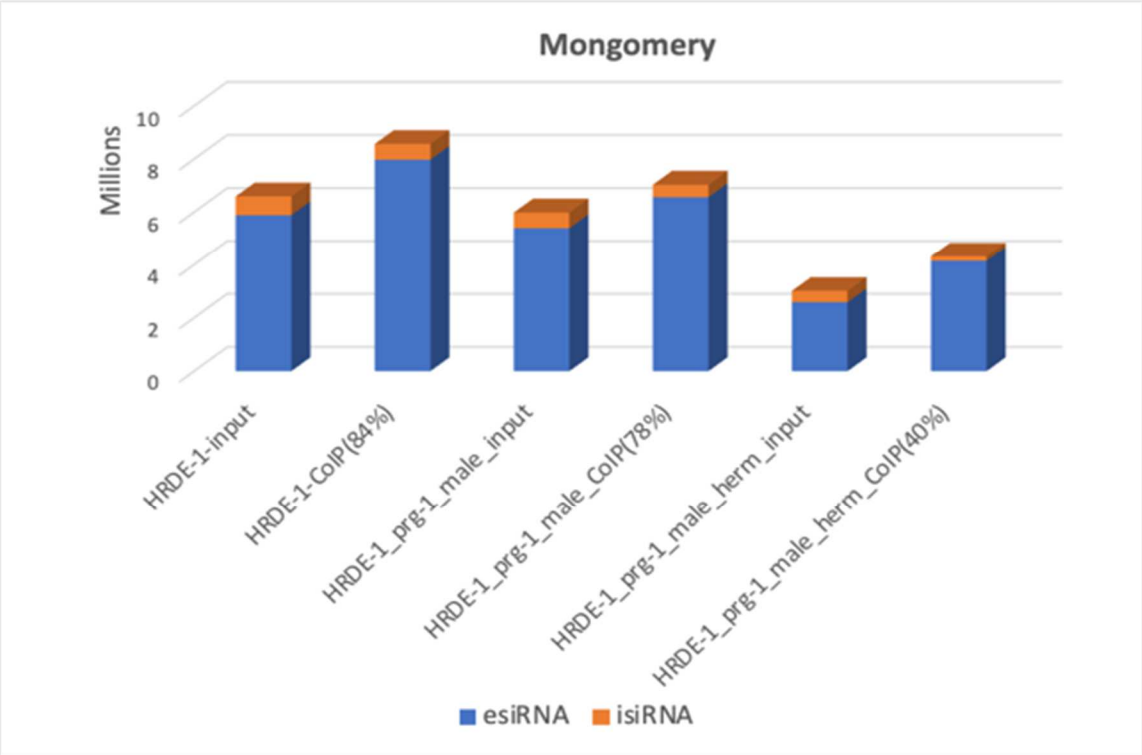


Figure 5.4 Co-IP from Montgomery experiment. X-axis sequence libraries. Y-Axis small RNA read in millions. Percentage is the percentage of isiRNA input that is Co-IPed with HRDE-1

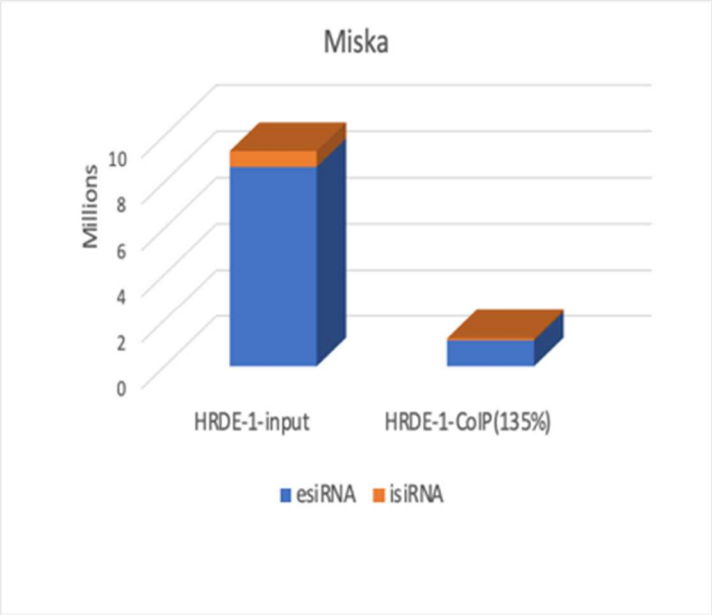


Figure 5.5 Co-IP from Miska experiment. X-axis sequence libraries. Y-Axis small RNA read in millions. Percentage is the percentage of isiRNA input that is Co-IPed with HRDE-1

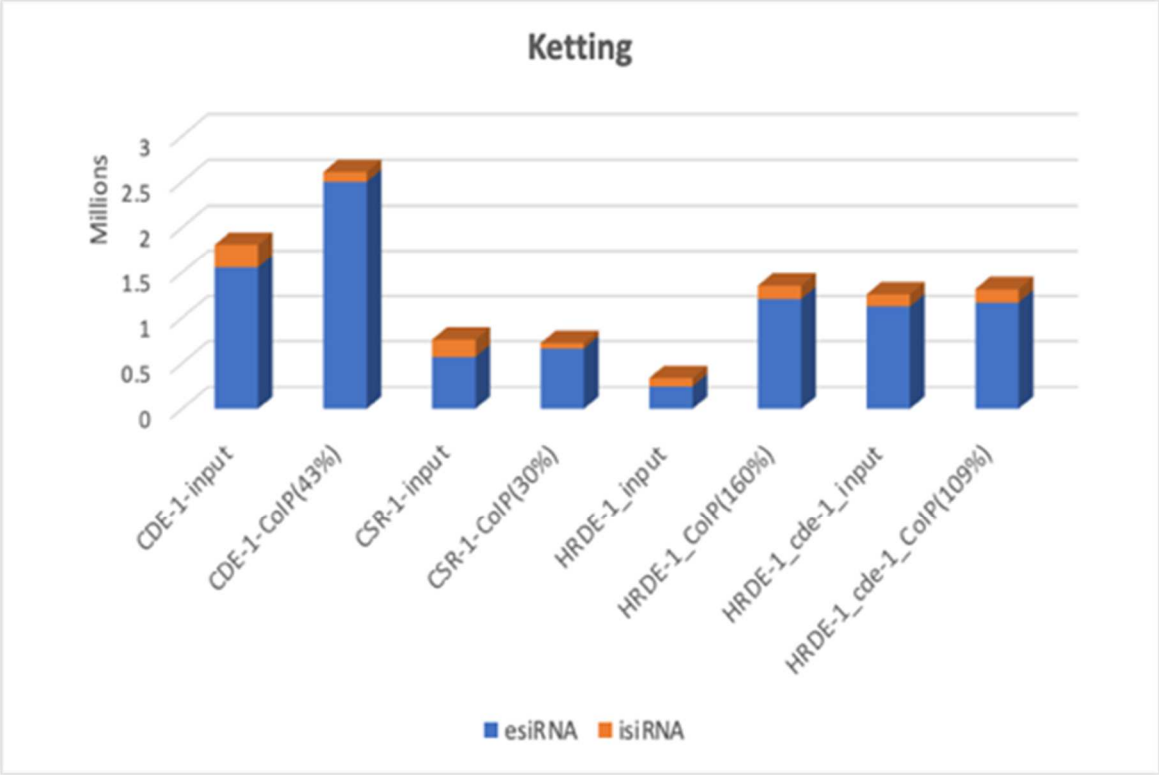


Figure 5.6 Co-IP from Ketting experiment. X-axis sequence libraries. Y-Axis small RNA read in millions. Percentage is the percentage of isiRNA input that is Co-IPed with HRDE-1

C. isiRNA bind to Argonaute CSR-1 to promote the production of siRNA

CSR-1 is a germline specific Argonaute that carries out many functions. CSR-1, like a few other Argonautes, contains a nuclease "slicer" domain and functions in many different cellular processes, ranging from promoting chromosome segregation [19] to surveilling mRNA species as they exit the nucleus [51]. In *csr-1(ADH1)* mutant, the slicing activity of CSR-1 is mutated but the protein is expressed [54]. Co-IP of wild type CSR-1 and mutated CSR-1(ADH1) indicated that isiRNA binds to both wildtype and slicer defective versions of CSR-1 (Figure 5.7).

Unlike HRDE-1 and other Argonautes that associate with 22G to repress gene expression post-transcriptionally and transcriptionally, CSR-1's association with 22G RNA and other factors such as DRH-3, EGO-1, EKL-1 has shown promote proper holocentric chromosome segregation [19]. Though CSR-1 associates with small RNA that targets a subset of germline expressing genes, loss of CSR-1 does not up regulate these mRNA targets [19]. In fact, CSR-1 is shown to promote expression of these genes [84]. CSR-1 slicer activity was shown to participate specifically in triggering the synthesis of siRNAs on the coding sequences of germline mRNA [54]. This is consistent with our analysis of small RNA libraries from *csr-1(ADH1)*, in which we found a decrease in both esiRNA and isiRNA levels (Figure 5.5). This suggests that isiRNA binding to CSR-1 is an important part of siRNA biosynthesis machinery, and that the slicing activity of CSR-1 is required to synthesize both esiRNA and isiRNA.

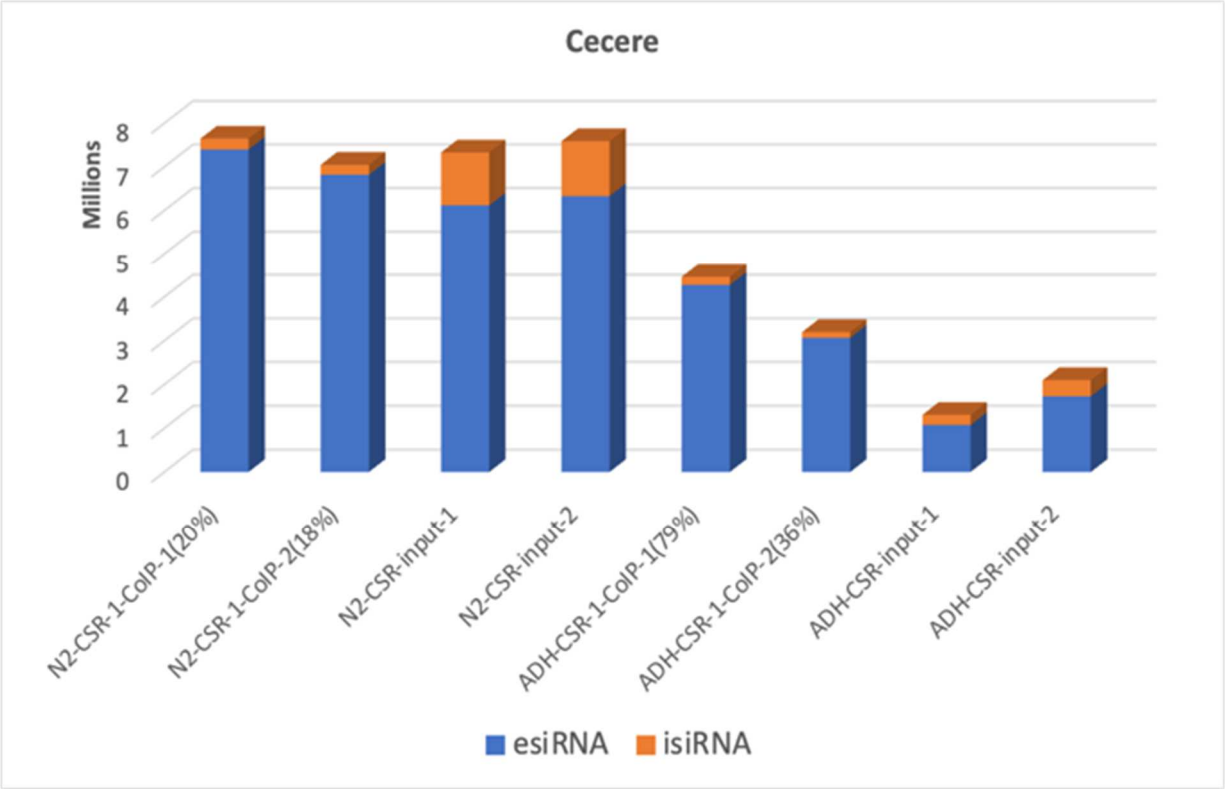


Figure 5.7 Co-IP from Cecere 2021 experiment. X-axis sequence libraries. Y-Axis small RNA read in millions. Percentage is the percentage of isiRNA input that is Co-IPed with HRDE-1

D. Cold shock and exogenous RNAi promotes the binding of esiRNA and isiRNA to NRDE-3

NRDE-3 is a mirror image of HRDE-1, both Argonautes have the same roles in the RNAi processes, in which they shuttle the siRNA from cytoplasm to the nucleus and recruit NRDE-1/2/4 to repress gene expression. The only difference is HRDE-1 is germline specific while NRDE-3 is soma specific. Previously, we showed that isiRNA bind to HRDE-1 in multiple independent datasets from different experiments. In this study, we have only one NRDE-3 CoIP and one *nrde-3* mutant siRNA library to analyze. The Co-IP dataset came from Guang et al., 2017, in which they investigated the effect of UV and cold shock on the association between NRDE-3 and the ribosomal siRNA (risiRNA) in *eri-1 (mg366), and susi-1* mutant backgrounds [89]. Their goal was to investigate the previously understudied small RNA species that are anti-sense to ribosomal RNA (risiRNA). Their work established that these RNA species are bound to the Argonaute NRDE-3, have a Guanosine - methylated 5' end and negatively regulate (along with the nuclear RNAi pathway) ribosomal RNA expression in response to cold shock and UV light exposure [89]. Cold and UV light each disrupt cellular processes and these small RNAs may allow cells to arrest cellular protein synthesis in the face of extreme stress. They focused on NRDE-3 bound small RNAs as these are most likely to be biologically functional rather than remnants of poorly processed "junk" small RNA as they had been supposed to be by others. We used the datasets of NRDE-3 co-IPed small RNAs that they generated to ask similar questions about esiRNA and isiRNA. Although we do not include analysis of risiRNA in our study, we utilize these datasets to investigate the binding of esiRNA and isiRNA to NRDE-3 under cold shock of N2 and UV treatment in *eri-1(mg366)* mutant backgrounds. This is a fair analysis because the libraries are made in the same way and should include all RNA species between 15 and 30 nt regardless of the 5' end modifications.

They used the *eri-1* mutant background because it is enhanced for RNAi in this background. In the *eri-1(mg-366);gfp::nrde-3* genetic background, Zhou et al., 2017 showed that

NRDE-3::GFP is primarily accumulated in the cytoplasm. They found that a *susi-1* mutant background, UV exposure, cold shock treatment and feeding exogenous *dpy-13(RNAi)* promotes the translocation of NRDE-3 to the nucleus [89]. I used my pipeline to quantify both esiRNA and isiRNA from their input and NRDE-3 co-IP libraries to ask whether the binding of esiRNA and isiRNA to NRDE-3 is dependent on the subcellular localization of NRDE-3 which was triggered by these various conditions.

UV treatment of *eri-1(mg-366);gfp::nrde-3* promotes translocation of NRDE-3 to the nucleus [89]. Zhou et al.'s NRDE-3 Co-IP analysis showed that 16 folds more risiRNA is associated with NRDE-3 in *eri-1(mg366)* mutants upon UV treatment. Our analysis of the same datasets showed that esiRNA and isiRNA do not bind to NRDE-3 in either *eri-1(mg366)-no-UV* (0.07 and 0.05 million reads for esiRNA and isiRNA respectively) or *eri-1(mg366)-UV* (0.1 and 0.04 million reads for esiRNA and isiRNA respectively). This suggested that missing ERI-1 prevents esiRNA and isiRNA from loading on to NRDE-3 regardless of whether NRDE-3 accumulated in the nucleus or in the cytoplasm or whether the animals are treated with UV light or not. Importantly, from our analysis of the input, the levels of isiRNA are not changed in the *eri-1(mg366)* mutant as compared to wild type though esiRNA levels are decreased (see Table 4.1 in Chapter Four).

However, feeding exogenous dsRNA targeting gene *dpy-13* to or cold shocking (the two conditions that promote the accumulation of NRDE-3 in the nucleus) *eri-1(mg366)* mutant animals restored the binding of esiRNA and isiRNA to NRDE-3 (Figure 5.8). This suggests that like risiRNA, binding of esiRNA and isiRNA to NRDE-3 is promoted when NRDE-3 is driven into the nucleus by exogenous RNAi or cold treatment. Why isiRNA is not associated with NRDE-3 in UV stressed animals may indicate that it is different from the stress of exogenous RNAi or cold shock.

risiRNA belongs to the 22G class of siRNA, and many components in its biosynthetic machinery are shared with endo-RNA's [89]. There is evidence showing that disrupting

biosynthesis pathways of other endo-siRNAs could shift the shared components to risiRNA biosynthetic machinery to promote risiRNA production [89]. Here, we see that this logic might be applied in the case of *susi-1(R457H)* mutant and to N2 cold shock animals. More specifically, in *susi-1(R457H)* they showed that risiRNA level is increased 10 fold compared to that of wild type [89], while in our analysis, we saw both esiRNA and isiRNA decrease 5 folds. Similarly, in the N2 cold shock input library, they found an increase in risiRNA level [89], while we observed a significant reduction of both esiRNA and isiRNA (Figure 5.8). These results could be that *susi-1(R457H)* and cold shock suppress esiRNA and isiRNA biosynthesis, and the components of esiRNA and isiRNA biosynthetic machinery are became available to be shifted to promote risiRNA production.

Our analysis of NRDE-3 Co-IP showed that cold shock promotes the binding of esiRNA and isiRNA to NRDE-3 (Figure 5.8). However, the library made from input from cold shock treated N2s showed a very low level of esiRNA and isiRNA. This result suggests that cold shock treatment either disrupts the biosynthesis machinery of esiRNA and isiRNA, or promotes the binding of esiRNA and isiRNA to NRDE-3, and that binding of esiRNA and isiRNA to NRDE-3 might protect the level of siRNA species from cold shock treatment.

In conclusion, binding of NRDE-3 to esiRNA and isiRNA are disrupted in *eri-1(mg366)* mutants. UV treatment, cold shock treatment and *dpy-13 RNAi* feeding promoted the nuclear translocation of NRDE-3, which resulted in increasing the production of risiRNA and promoting the association of risiRNA to NRDE-3. However, not all of the treatments that promote translocation of NRDE-3 to the nucleus also promote binding of esiRNA and isiRNA to the NRDE-3. Thus, the nuclear translocation of NRDE-3, in the case of UV treatment, is driven by risiRNA rather than the endogenous species composed of esi and isiRNA. More specifically, *eri-1(mg366)* with UV treatment does not rescue the binding of esiRNA and isiRNA to NRDE-3, but the cold shock, and *dpy-13 RNAi* feeding seem to promote the binding of esiRNA and isiRNA to NRDE-3.

Though our previous examination of isiRNA in Chapter Four revealed that levels of isiRNA are probably highest in the germline, we see from the Zhou et al., 2017 study that isiRNA levels may increase in the soma as reflected by the increased isiRNA binding to the somatic Argonaute, NRDE-3. Cold shock and exogenous RNAi may indeed enhance the levels of isiRNA in the soma and or promote isiRNA association with NRDE-3. This could reflect the fact that only under stressful conditions are levels of isiRNA RNA increased in the soma.

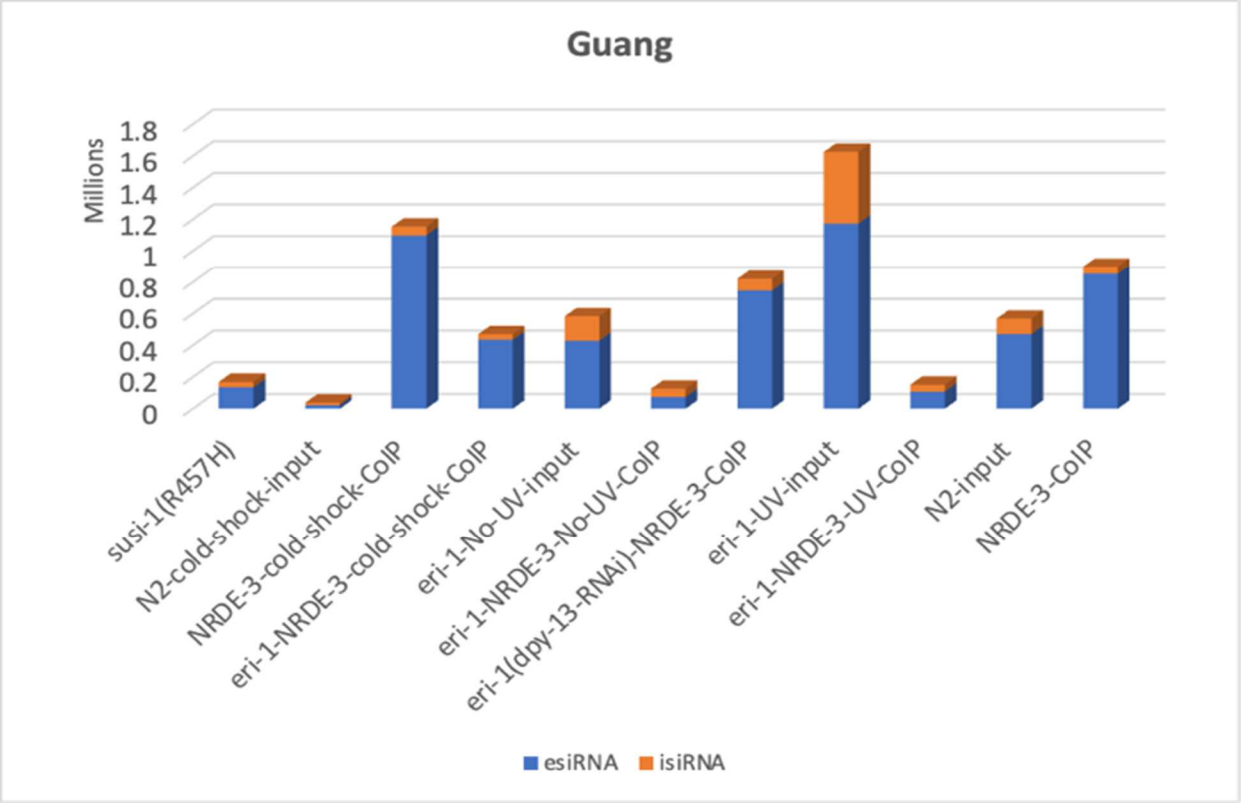


Figure 5.8 Co-IP from Guang 2016 experiment. X-axis sequence libraries. Y-Axis small RNA read in millions. Percentage is the percentage of isiRNA input that is Co-IPed with HRDE-1

E. Lack of ERGO-1 increased the number of genes with higher isiRNA than esiRNA reads

ERGO-1 is the Argonaute that binds and stabilizes the 26G primary siRNAs generated in the oogenic germline [20,21]. ERGO-1 is highly expressed during the embryonic stage, but its expression is reduced in L3, L4, and young adults [20,21]. Loss of ERGO-1 promotes the enhanced exogenous RNAi (Eri) phenotype as the result of losing the 26G class of siRNA and thus the subsequent amplification of secondary endogenous siRNA that carry out gene silencing [64,70].

Analyzing the ERGO-1 Co-IP datasets, I found that isiRNA binds to the ERGO-1 Argonaute (Figure 5.9). This association with ERGO-1 suggests the possibility that isiRNAs function in embryos and may be primary siRNAs. As I described in chapter 3, in the absence of ERGO-1, the level of esiRNA is reduced but the level of isiRNA is unchanged. We interpret this to mean that though isiRNAs are bound to ERGO-1, their production does not depend on this Argonaute. A similar trend was seen when we analyzed libraries made from a separate allele of *ergo-1(tm1860)*. In these libraries, we found that esiRNA levels were unchanged while the levels of isiRNA were slightly increased [21]. This suggests that esiRNA production is dependent on ERGO-1, but isiRNA production is not.

In order to look more closely at the pathways that regulate isiRNA production, I looked at the genes that produced more isiRNA than esiRNA (isiRNA>esiRNA). I found that in the wild type animal, there are about 400 genes that produce isiRNA>esiRNA (Figure 4.15). In *ergo-1(gg098)* mutants this number is increased to 695 (Figure 4.15). This indicates that esiRNA biosynthesis depends more on ERGO-1 than does isiRNA biosynthesis. This increase in number of isiRNA>esiRNA genes could reflect either that the number of esiRNA reads in *ergo-1(gg098)* decrease compared to N2, or the number of isiRNA reads in *ergo-1(gg098)* increase compared to N2. This result suggested that isiRNA biosynthesis is independent of the ERGO-1 Argonaute.

ERGO-1 bound 26 nt small RNAs are primary siRNAs and it would seem that esiRNA production depends more heavily on this trigger species than does isiRNA production. The fact that ERGO-1 is embryonic specific means that the esiRNA load depends more on what is found in the embryo while isiRNA may have an independent trigger that occurs during the later larval stages and continues into the adult. Though isiRNA production does not depend on ERGO-1, ERGO-1 which has slicing activity, may potentially use isiRNA as a guide to destroy unspliced mRNA [94].

In order to investigate this possibility, one could examine long read sequences of mRNAs from an *ergo-1* defective strain and if isiRNA directs ERGO-1 to destroy unspliced mRNA, I would expect that the levels of messages with retained introns would increase. This would be another mechanism for quality control to reduce the possibility of translating a message that contains unspliced introns.

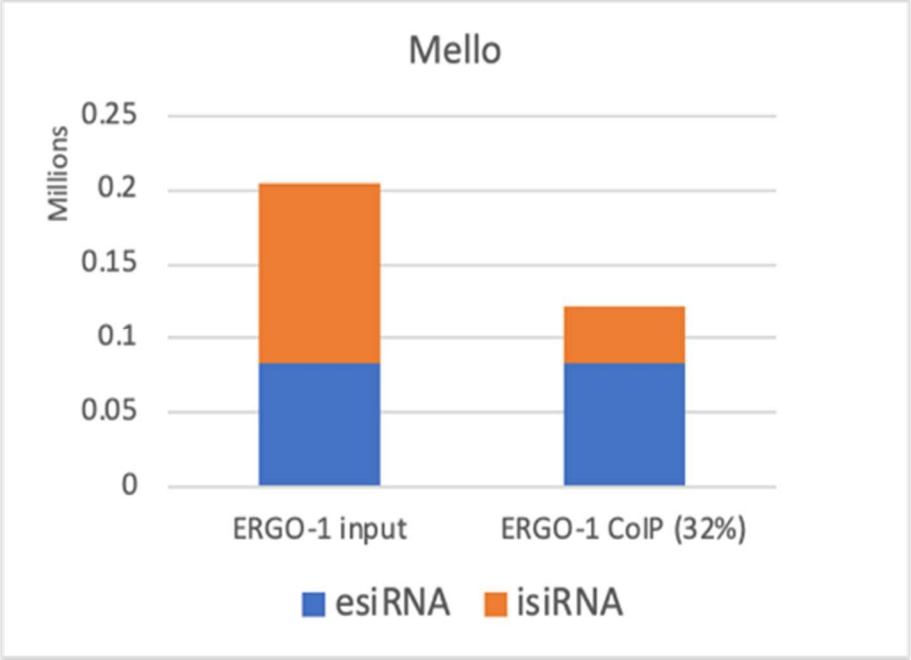


Figure 5.9 Co-IP from Mello experiment. X-axis sequence libraries. Y-Axis small RNA read in millions. Percentage is the percentage of isiRNA input that is Co-IPed with ERGO-1

F. WAGO-1 is required for biosynthesis of germline specific isiRNA

WAGO-1 is a cytoplasmic Argonaute. WAGO-1 is highly expressed in mature spermatids and throughout the germline [91]. WAGO-1 is found to be localized in the P granule, and silences transposon, pseudogenes, and cryptic loci [13]. Missing WAGO-1 resulted in Quintuple AGO mutant with a significant reduction in germline 22G, which suggested that WAGO play an important role in the germline 22G-RNA function [13]. Indeed, WAGO-1 with other factors such as RDE-3 and MUT-7 are required for germline esiRNA silencing pathway [13].

Small RNAs derived from RdRP are 22 nucleotides long and have the propensity for a 5'G residue and thus are called 22Gs [13]. 22Gs are highly expressed in the germline, and ~50% of their targets are coding genes [13]. Biosynthesis of these 22Gs siRNA species is dependent on the Dicer-Related Helicase DRH-3; the RdRPs RRF-1, EGO-1, and the Tudor-domain protein, EKL-1. Based on the associating Argonautes, there are two major systems that can produce 22Gs. One system requires the Argonaute WAGO-1, RDE-3 and MUT-7, and the other system requires the Argonaute CSR-1, and CDE-1. Both systems require the RdRps EGO-1 and RRF-1 along with the helicase DRH-3 [13].

As described in chapter 3, the biosynthesis of isiRNAs depends on DRH-3, RRF-1, EGO-1, EKL-1, RDE-3, MUT-7, and WAGO-1 (in the context of *mago12*), but does not depend on CSR-1 or its cofactor CDE-1. These data suggested that isiRNA biosynthesis is likely to be part of the WAGO-1 system. Surprisingly, there is an abundance of esiRNA Co-IPed with WAGO-1, but I did not detect isiRNA in the WAGO-1 Co-IP library (Figure 5.10). This suggested that the specific isiRNA population that is made from the WAGO-1 system does not bind WAGO-1 and thus is not engaged in the HRDE-1-Mutators-WAGOs-HRDE-1 positive feedback loop proposed in Albuquerque and Ketting's study [84]. This is interesting because isiRNA production or levels depend on the Argonaute WAGO-1 but it is not bound by this Argonaute. This may mean that the WAGO-1 system might be responsible for a specific germline exclusive isiRNA.

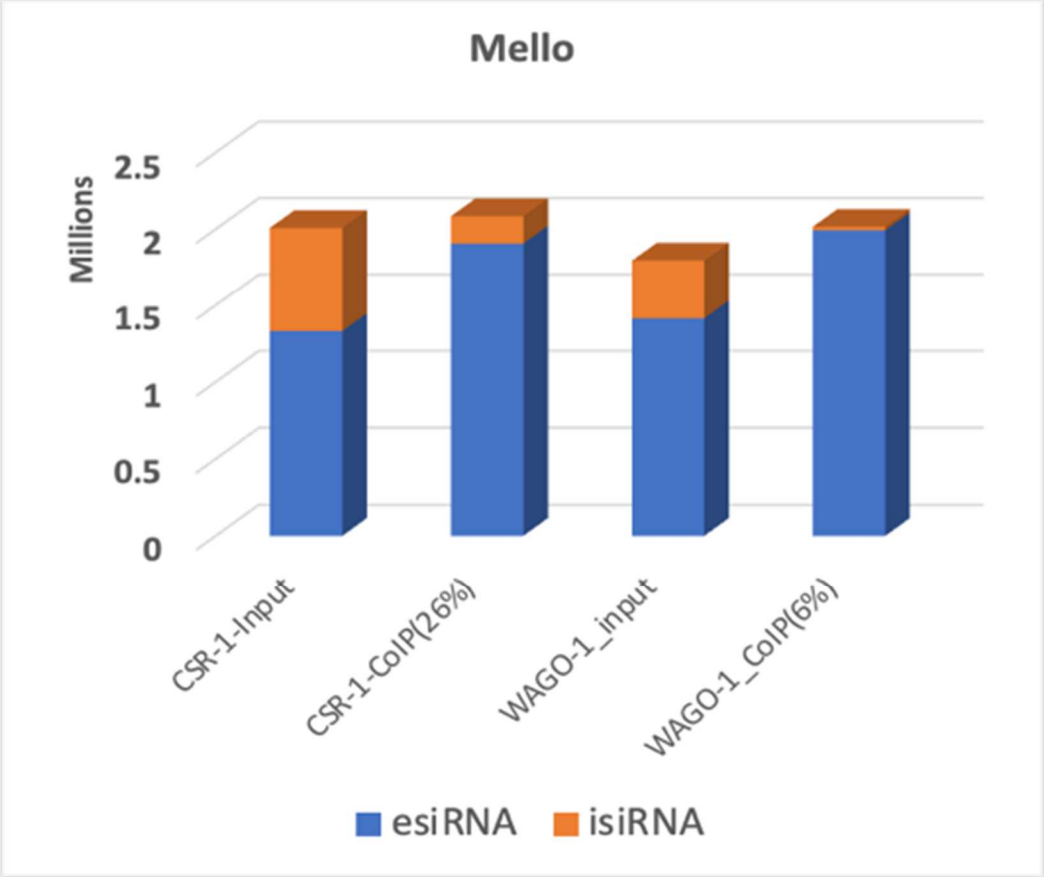


Figure 5.10 Co-IP from Mello experiment. X-axis sequence libraries. Y-Axis small RNA read in millions. Percentage is the percentage of isiRNA input that is Co-IPed with WAGO-1

CHAPTER SIX: CONCLUSION AND FUTURE DIRECTIONS

CONCLUSION AND PROPOSED MODEL

By creating a pipeline to explore public domain RNA sequencing datasets, I was able to characterize a population of small, potentially expression-interfering RNA species that had remained largely understudied (Chapter One). This population of siRNA targets intronic regions, which I termed isiRNAs. I found—using both my pipeline and, subsequently, another one developed in the Goodarzi laboratory—that in two hundred small RNA libraries from dozens of experiments, isiRNA map to approximately 1/3 of all genes in the *C. elegans* genome. Consistently, isiRNAs account for ~7%-10% of total siRNA reads in each small RNA library. That this species is consistently found at similar levels across many different libraries and labs indicates that it is unlikely to be a mere artifact of small RNA library production.

To further investigate the possible significance of this species of small RNA, I asked if it arose from introns with specific sequence properties (Chapter Two). I found that the introns that are most likely to produce isiRNA have a median nucleotide length that is 10-fold longer than that of the average genomic intron. The genes with alternatively spliced introns are also twice as likely to have isiRNAs mapped to them. Longer introns and those that are alternatively spliced are also more likely to have a secondary structure to engage the splicing machinery less effectively. I hypothesize that isiRNA production may favor these poorly spliced introns (retained intron). My hypothesis could be tested by sequencing long mRNA reads which may reveal the identity of retained introns. I expect that those introns would be more likely to have isiRNA mapped to them than introns that are not retained. My hypothesis could also be tested by using CRISPR technology to engineer a poorly spliced intron in a gene that has no isiRNA mapped to it and observing whether that poorly spliced intron subsequently has isiRNA map to it.

Through analysis of libraries made from mutants that are defective in siRNA production, I uncovered some insights into the biosynthesis of the isiRNA. First, a loss of a diverse set of germline specific siRNA factors diminishes isiRNA levels. Second, isiRNA biosynthesis machinery occurs via the Worm Argonautes (WAGOs) pathway. Third, isiRNA may be amplified

at the mutator complex, which is at the periphery of P granules. Fourth, isiRNA levels are dependent on many exo-RNAi factors. Fifth, isiRNA synthesis is independent from the endogenous RNAi biosynthesis pathway. Sixth, isiRNA may be primary (ERGO-1 bound), secondary (HRDE-1 bound) as well as tertiary RNA (independent of *eri-1* and *dcr-1(mg375)*). Finally, the bulk of our evidence points to isiRNA species being found in the germline. This may reflect the need to carefully regulate intron expression in order to limit transposons that are harbored within introns.

I found that isiRNA is bound to specific Argonautes which suggests that they are likely to have specific biological functions. By evaluating the binding of isiRNA to those Argonautes using Co-IP libraries, I discovered that isiRNA may play a role in many cellular processes. First, HRDE-1 and CSR-1 bind these small RNAs and may compete for their binding. The siRNA species bound by HRDE-1 are likely used to repress gene expression and possibly prompt their own amplification via mutator-WAGOs-HRDE-1 feedback loop. Second, isiRNA binds to the germline specific Argonaute, HRDE-1, in order to repress transcription. Third, isiRNA binds to Argonaute CSR-1 to promote the production of siRNA. Fourth, cold shock and exo-RNAi promote the binding of esiRNA and isiRNA to NRDE-3. Fifth, missing ERGO-1 increased the number of genes with higher isiRNA than esiRNA reads. Sixth, WAGO-1 is required for biosynthesis of germline specific isiRNA. The sum of evidence suggests that isiRNA would function in the germline to limit expression of genes that transcribe retained introns.

From my finding, I propose a model that isiRNA biosynthetic machinery can be triggered by either exogenous or endogenous RNAi pathways. The exogenous RNAi pathway includes factors RDE-1, RDE-4, DRH-3 and DCR-1, while the endogenous RNAi pathway includes factors: ERI-1, ERI-9, and DCR-1. Both triggers produce primary isiRNA that is processed further by common factors, MUT-7 and RDE-3, to produce secondary isiRNA. This secondary isiRNA engages in two pathways. First, they are loaded on the germline Argonaute HRDE-1, then this complex is shuttled to the nucleus and recruits NRDE-1, NRDE-2, NRDE-4 to form

NRDEs-siRNA complex to repress gene transcription. Second, the secondary isiRNA engage in a mutator pathway. In this pathway, isiRNA together with mutator foci and other factors (WAGO-1, DRH-3, PRG-1, CSR-1, and EGO-1) at the P granule, where nuclear pores are concentrated, to act as the gatekeeper surveillance for the integrity of mRNA exiting the nucleus (Figure 6.1).

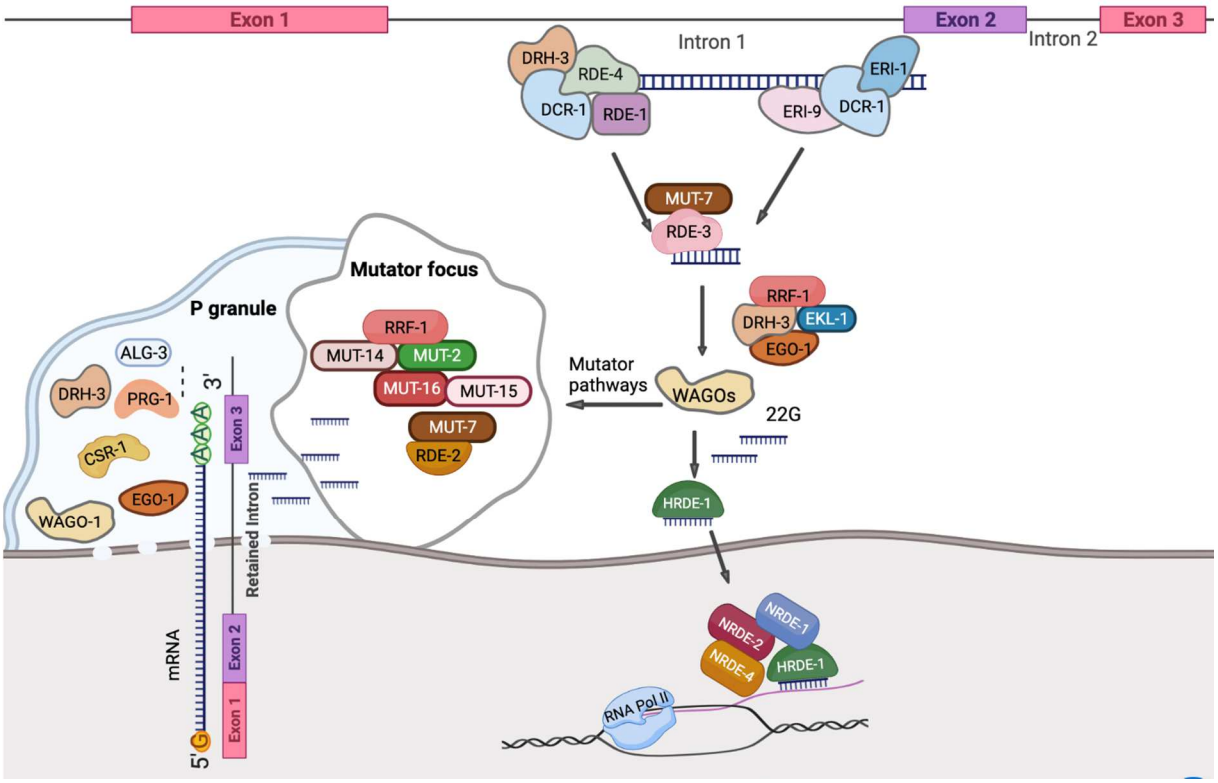


Figure 6.1 Model for isiRNA biosynthetic machinery and function

FUTURE DIRECTION

In conclusion, the consistency of isiRNA identification across hundred of sequencing libraries and their binding to Argonautes of multiple pathways indicates that isiRNA plays important roles in the c ell. There are multiple lines of evidence showing that esiRNA are involved in silencing viral infection in *C. elegans* by targeting the horizontal gene transfer caused by transposon factors such as viral genetic material [95,96]. To test if isiRNA is playing a similar role in antiviral defense, one could explore the small RNA libraries that are made from virus-infected animals to see if isiRNA expression is changed compared to untreated animals. Additionally, post-transcriptional gene expression regulation by esiRNA was shown to be important in maintaining cellular integrity and stability. Disruption of esiRNA regulatory processes could lead to cellular abnormality and disease such as cancer [97]: Antisense oligos therapeutics (an esiRNA mirror image) were used as a means to reverse this gene regulation disruption [98]. However, this approach produced limited efficacy. Perhaps the reason for this lower than desired efficacy is that antisense oligos only targeted mature mRNA post-transcriptionally. I propose that if we expand the targets to include intronic regions using isiRNA we may improve the efficacy of antisense oligos therapeutics.

REFERENCES

1. Yan, J. *et al.* The miR165/166 Mediated Regulatory Module Plays Critical Roles in ABA Homeostasis and Response in Arabidopsis thaliana. *PLoS Genet* **12**, e1006416 (2016).
2. Rué, P. & Martinez Arias, A. Cell dynamics and gene expression control in tissue homeostasis and development. *Mol Syst Biol* **11**, 792 (2015).
3. Rau, A., Flister, M., Rui, H. & Auer, P. L. Exploring drivers of gene expression in the Cancer Genome Atlas. *Bioinformatics* **35**, 62–68 (2019).
4. Delgado, M. D. & León, J. Gene expression regulation and cancer. *Clin Transl Oncol* **8**, 780–787 (2006).
5. Frenk, S. & Houseley, J. Gene expression hallmarks of cellular ageing. *Biogerontology* **19**, 547–566 (2018).
6. Anisimova, A. S. *et al.* Multifaceted deregulation of gene expression and protein synthesis with age. *Proc Natl Acad Sci U S A* **117**, 15581–15590 (2020).
7. Groves, A. K. *et al.* Differential regulation of transcription factor gene expression and phenotypic markers in developing sympathetic neurons. *Development* **121**, 887–901 (1995).
8. Karin, M. & Hunter, T. Transcriptional control by protein phosphorylation: signal transmission from the cell surface to the nucleus. *Curr Biol* **5**, 747–757 (1995).
9. Hutvagner, G. Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. *FEBS Lett* **579**, 5850–5857 (2005).
10. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–854 (1993).

11. Czech, B. *et al.* piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet* **52**, 131–157 (2018).
12. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–811 (1998).
13. Gu, W. *et al.* Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. *Mol Cell* **36**, 231–244 (2009).
14. Zhang, C. *et al.* mut-16 and other mutator class genes modulate 22G and 26G siRNA pathways in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **108**, 1201–1208 (2011).
15. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev* **15**, 2654–2659 (2001).
16. Sapetschnig, A., Sarkies, P., Lehrbach, N. J. & Miska, E. A. Tertiary siRNAs mediate paramutation in C. elegans. *PLoS Genet* **11**, e1005078 (2015).
17. Pak, J., Maniar, J. M., Mello, C. C. & Fire, A. Protection from feed-forward amplification in an amplified RNAi mechanism. *Cell* **151**, 885–899 (2012).
18. Rechavi, O. & Lev, I. Principles of Transgenerational Small RNA Inheritance in Caenorhabditis elegans. *Current Biology* **27**, R720–R730 (2017).
19. Claycomb, J. M. *et al.* The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* **139**, 123–134 (2009).
20. Han, T. *et al.* 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **106**, 18674–18679 (2009).

21. Vasale, J. J. *et al.* Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proc Natl Acad Sci U S A* **107**, 3582–3587 (2010).
22. Burton, N. O., Burkhart, K. B. & Kennedy, S. Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **108**, 19683–19688 (2011).
23. Guang, S. *et al.* An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* **321**, 537–541 (2008).
24. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
25. Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T. & Sandhu, M. S. Long reads: their purpose and place. *Hum Mol Genet* **27**, R234–R241 (2018).
26. Tang, X., Huang, Y., Lei, J., Luo, H. & Zhu, X. The single-cell sequencing: new developments and medical applications. *Cell Biosci* **9**, 53 (2019).
27. Schwartz, S. *et al.* PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res* **10**, 577–586 (2000).
28. Angiuoli, S. V. & Salzberg, S. L. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* **27**, 334–342 (2011).
29. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**, 1394–1403 (2004).
30. Juang, B.-T. *et al.* Endogenous nuclear RNAi mediates behavioral adaptation to odor. *Cell* **154**, 1010–1022 (2013).

31. Gent, J. I. *et al.* Distinct phases of siRNA synthesis in an endogenous RNAi pathway in C. elegans soma. *Mol Cell* **37**, 679–689 (2010).
32. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**, a003707 (2011).
33. Juneau, K., Miranda, M., Hillenmeyer, M. E., Nislow, C. & Davis, R. W. Introns regulate RNA and protein abundance in yeast. *Genetics* **174**, 511–518 (2006).
34. Kim, Y.-K. & Kim, V. N. Processing of intronic microRNAs. *EMBO J* **26**, 775–783 (2007).
35. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
36. Chorev, M. & Carmel, L. The function of introns. *Front Genet* **3**, 55 (2012).
37. Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457–463 (2010).
38. Izban, M. G. & Luse, D. S. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J Biol Chem* **267**, 13647–13655 (1992).
39. Castillo-Davis, C. I., Mekhedov, S. L., Hartl, D. L., Koonin, E. V. & Kondrashov, F. A. Selection for short introns in highly expressed genes. *Nat Genet* **31**, 415–418 (2002).
40. Dumesic, P. A. *et al.* Stalled spliceosomes are a signal for RNAi-mediated genome defense. *Cell* **152**, 957–968 (2013).
41. Wachtel, C. & Manley, J. L. Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. *Mol Biosyst* **5**, 311–316 (2009).

42. Modrek, B., Resch, A., Grasso, C. & Lee, C. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res* **29**, 2850–2859 (2001).
43. Lopez, A. J. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu Rev Genet* **32**, 279–305 (1998).
44. Tudor, G., Aguilera, A., Halverson, D. O., Laing, N. D. & Sausville, E. A. Susceptibility to drug-induced apoptosis correlates with differential modulation of Bad, Bcl-2 and Bcl-xL protein levels. *Cell Death Differ* **7**, 574–586 (2000).
45. Kwan, T. *et al.* Heritability of alternative splicing in the human genome. *Genome Res* **17**, 1210–1218 (2007).
46. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188–1190 (2004).
47. Ecker, J. R. & Davis, R. W. Inhibition of gene expression in plant cells by expression of antisense RNA. *Proc Natl Acad Sci U S A* **83**, 5372–5376 (1986).
48. Shih, J. D. & Hunter, C. P. SID-1 is a dsRNA-selective dsRNA-gated channel. *RNA* **17**, 1057–1065 (2011).
49. Almeida, M. V., Andrade-Navarro, M. A. & Ketting, R. F. Function and Evolution of Nematode RNAi Pathways. *Noncoding RNA* **5**, E8 (2019).
50. Berry, L. W., Westlund, B. & Schedl, T. Germ-line tumor formation caused by activation of glp-1, a Caenorhabditis elegans member of the Notch family of receptors. *Development* **124**, 925–936 (1997).

51. Phillips, C. M., Montgomery, T. A., Breen, P. C. & Ruvkun, G. MUT-16 promotes formation of perinuclear mutator foci required for RNA silencing in the C. elegans germline. *Genes Dev* **26**, 1433–1444 (2012).
52. Phillips, C. M. *et al.* MUT-14 and SMUT-1 DEAD box RNA helicases have overlapping roles in germline RNAi and endogenous siRNA formation. *Curr Biol* **24**, 839–844 (2014).
53. van Wolfswinkel, J. C. *et al.* CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. *Cell* **139**, 135–148 (2009).
54. Singh, M. *et al.* Translation and codon usage regulate Argonaute slicer activity to trigger small RNA biogenesis. *Nat Commun* **12**, 3492 (2021).
55. Youngman, E. M. & Claycomb, J. M. From early lessons to new frontiers: the worm as a treasure trove of small RNA biology. *Front Genet* **5**, 416 (2014).
56. Uebel, C. J., Agbede, D., Wallis, D. C. & Phillips, C. M. Mutator Foci Are Regulated by Developmental Stage, RNA, and the Germline Cell Cycle in Caenorhabditis elegans. *G3 (Bethesda)* **10**, 3719–3728 (2020).
57. Das, P. P. *et al.* Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. *Mol Cell* **31**, 79–90 (2008).
58. Wu, P.-H. *et al.* The evolutionarily conserved piRNA-producing locus pi6 is required for male mouse fertility. *Nat Genet* **52**, 728–739 (2020).
59. Klattenhoff, C. & Theurkauf, W. Biogenesis and germline functions of piRNAs. *Development* **135**, 3–9 (2008).
60. Carmell, M. A. *et al.* MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. *Developmental Cell* **12**, 503–514 (2007).

61. Batista, P. J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol Cell* **31**, 67–78 (2008).
62. Wang, G. & Reinke, V. A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Curr Biol* **18**, 861–867 (2008).
63. Tabara, H. *et al.* The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**, 123–132 (1999).
64. Yigit, E. *et al.* Analysis of the C. elegans Argonaute Family Reveals that Distinct Argonautes Act Sequentially during RNAi. *Cell* **127**, 747–757 (2006).
65. Lai, E. C. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**, 363–364 (2002).
66. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**, 23–34 (2001).
67. Tabara, H., Yigit, E., Siomi, H. & Mello, C. C. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans. *Cell* **109**, 861–871 (2002).
68. Parrish, S. & Fire, A. Distinct roles for RDE-1 and RDE-4 during RNA interference in Caenorhabditis elegans. *RNA* **7**, 1397–1402 (2001).
69. Lee, R. C., Hammell, C. M. & Ambros, V. Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans. *RNA* **12**, 589–597 (2006).

70. Pavelec, D. M., Lachowiec, J., Duchaine, T. F., Smith, H. E. & Kennedy, S. Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in Caenorhabditis elegans. *Genetics* **183**, 1283–1295 (2009).
71. Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K. & Tabara, H. In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. *EMBO J* **26**, 5007–5019 (2007).
72. Zhang, C. *et al.* The Caenorhabditis elegans RDE-10/RDE-11 complex regulates RNAi by promoting secondary siRNA amplification. *Curr Biol* **22**, 881–890 (2012).
73. Smardon, A. *et al.* EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. *Current Biology* **10**, 169–178 (2000).
74. Bosher, J. M., Dufourcq, P., Sookhareea, S. & Labouesse, M. RNA interference can target pre-mRNA: consequences for gene expression in a Caenorhabditis elegans operon. *Genetics* **153**, 1245–1256 (1999).
75. Burkhart, K. B. *et al.* A pre-mRNA-associating factor links endogenous siRNAs to chromatin regulation. *PLoS Genet* **7**, e1002249 (2011).
76. Buckley, B. A. *et al.* A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature* **489**, 447–451 (2012).
77. Padeken, J. *et al.* Argonaute NRDE-3 and MBT domain protein LIN-61 redundantly recruit an H3K9me3 HMT to prevent embryonic lethality and transposon expression. *Genes Dev* **35**, 82–101 (2021).
78. Gabel, H. W. & Ruvkun, G. The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi. *Nat Struct Mol Biol* **15**, 531–533 (2008).

79. Gammon, D. B. *et al.* The Antiviral RNA Interference Response Provides Resistance to Lethal Arbovirus Infection and Vertical Transmission in Caenorhabditis elegans. *Curr Biol* **27**, 795–806 (2017).

80. Moazed, D. Small RNAs in transcriptional gene silencing and genome defence. *Nature* **457**, 413–420 (2009).

81. Schott, D. H., Cureton, D. K., Whelan, S. P. & Hunter, C. P. An antiviral role for the RNA interference machinery in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **102**, 18420–18424 (2005).

82. Castel, S. E. & Martienssen, R. A. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* **14**, 100–112 (2013).

83. Ashe, A. *et al.* piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. *Cell* **150**, 88–99 (2012).

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

85. Filipowicz, W. RNAi: the nuts and bolts of the RISC machine. *Cell* **122**, 17–20 (2005).

86. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).

87. Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437 (2004).

88. Gu, S. G. *et al.* Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat Genet* **44**, 157–164 (2012).

89. Zhou, X. *et al.* RdRP-synthesized antisense ribosomal siRNAs silence pre-rRNA via the nuclear RNAi pathway. *Nat Struct Mol Biol* **24**, 258–269 (2017).

90. Seth, M. *et al.* The C. elegans CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. *Dev Cell* **27**, 656–663 (2013).

91. Conine, C. C. *et al.* Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **107**, 3588–3593 (2010).

92. Kim, T.-H., Leslie, P. & Zhang, Y. Ribosomal proteins as unrevealed caretakers for cellular stress and genomic instability. *Oncotarget* **5**, 860–871 (2014).

93. Lykke-Andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol* **16**, 665–677 (2015).

94. Newman, M. A. *et al.* The surveillance of pre-mRNA splicing is an early step in C. elegans RNAi of endogenous genes. *Genes Dev* **32**, 670–681 (2018).

95. Malone, C. D. & Hannon, G. J. Small RNAs as guardians of the genome. *Cell* **136**, 656–668 (2009).

96. Fischer, S. E. J. *et al.* The ERI-6/7 helicase acts at the first stage of an siRNA amplification pathway that targets recent gene duplications. *PLoS Genet* **7**, e1002369 (2011).

97. Calin, G. A. *et al.* MiR-15a and miR-16-1 cluster functions in human leukemia. *Proc Natl Acad Sci U S A* **105**, 5166–5171 (2008).

98. Mahmoodi Chalbatani, G. *et al.* Small interfering RNAs (siRNAs) in cancer therapy: a nano-based approach. *Int J Nanomedicine* **14**, 3111–3128 (2019).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Trang Duong
96C1472E07CF4A8...
Author Signature

3/14/2022
Date