University of Massachusetts Medical School
eScholarship@UMMS

GSBS Dissertations and Theses | Graduate School of Biomedical Sciences

2010-06-14

Roles of Cellular RNA-Dependent RNA Polymerases in Endogenous Small RNA Pathways in Caenorhabditis elegans: A Dissertation

Jessica J. Vasale
University of Massachusetts Medical School

Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/gsbs_diss

 Part of the Amino Acids, Peptides, and Proteins Commons, Animal Experimentation and Research Commons, Genetic Phenomena Commons, Molecular Biology Commons, Molecular Genetics Commons, and the Nucleic Acids, Nucleotides, and Nucleosides Commons

Repository Citation

Vasale JJ. (2010). Roles of Cellular RNA-Dependent RNA Polymerases in Endogenous Small RNA Pathways in Caenorhabditis elegans: A Dissertation. GSBS Dissertations and Theses. https://doi.org/10.13028/rwy3-kh64. Retrieved from https://escholarship.umassmed.edu/gsbs_diss/481

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Dissertations and Theses by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Graduate School of Biomedical Sciences
GSBS Dissertations

University of Massachusetts Medical School • Year 2010

ROLES OF CELLULAR RNA-DEPENDENT RNA POLYMERASES IN ENDOGENOUS SMALL RNA PATHWAYS IN *CAENORHABDITIS ELEGANS*: A Dissertation

Jessica Joy Vasale

University of Massachusetts Medical School

ROLES OF CELLULAR RNA-DEPENDENT RNA POLYMERASES IN ENDOGENOUS SMALL RNA PATHWAYS IN *CAENORHABDITIS ELEGANS*:

A Dissertation Presented

By

Jessica Joy Vasale

Submitted to the Faculty of the

University of Massachusetts

Graduate School of Biomedical Sciences, Worcester, MA

In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

June 14th, 2010

ROLES OF CELLULAR RNA-DEPENDENT RNA POLYMERASES IN ENDOGENOUS SMALL RNA PATHWAYS IN *CAENORHABDITIS ELEGANS*

A Dissertation Presented By

Jessica J. Vasale

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

---

Craig Mello, Ph.D., Thesis Advisor

Kirsten Hagstrom, Ph.D., Member of Committee

Danesh Moazed Ph.D., Member of Committee

Craig Peterson, Ph.D., Member of Committee

Sean Ryder, Ph.D., Member of Committee

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

---

Thoru Pederson, Ph.D., Chair of Committee

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

---

Anthony Carruthers, Ph.D.
Dean of the Graduate School of Biomedical Sciences

Interdisciplinary Graduate Program
June 14, 2010

Dedication

To my Family, who has been my source of inspiration.

Mimi, I miss you.

Acknowledgements

As I reflect over the years I have spent at UMass, I have so many people to thank for their support and encouragement. To begin, I would like to thank my Advisor, Craig Mello, for giving me the pleasure of working in his lab and sharing with me his knowledge and passion for science. I also would like to thank Thoru Pederson, Kirsten Hagstrom, Danesh Moazed, Craig Peterson, and Sean Ryder, who have all supported me by participating on my Defense Committee. I am also extremely grateful for my remarkable lab mates, past and present, who have all pushed me, taught me, inspired me, supported me, believed in me, encouraged me, and touched my life in their individual ways, to make me a better scientist. I also need to express my gratitude to my friends, some of who have descended with me to over three atmospheres, and of course, my family. Thank you to my parents, Greg & Suzanne, my sister, Heather, my brother-in-law, Luke, my niece, CaraMia, and my grandmother, Josephine -who I think about so often- for their unconditional love and support. Finally, I would like to thank Mike for his love and support over the last five years of my life. I can't wait to see how you're going to top the limo ride for my next birthday.

Abstract

The RNA interference (RNAi) pathway in *Caenorhabditis elegans* is a two-step, small RNA-mediated silencing pathway. Unlike in other organisms, Dicer processing of double-stranded RNA into small interfering (si) RNAs is not sufficient in worms to induce gene silencing. The activity of cellular RNA-dependent RNA polymerase (RdRP) is necessary to synthesize a secondary pool of siRNAs, which interact with a unique class of Argonaute proteins to form the functional effector complexes that mediate silencing. The aims of this thesis were to: 1) characterize the role of RdRP family members in endogenous small RNA biogenesis; 2) identify the Argonaute proteins that interact with RdRP-dependent small RNAs; and 3) investigate the biological function of RdRP-dependent small RNA pathways in *C. elegans*.

In this thesis, I describe genetic, deep sequencing, and molecular studies, which identify 22G-RNAs as the most abundant class of endogenous small RNA in *C. elegans*. The 22G-RNAs resemble RdRP-dependent secondary siRNAs produced during exogenous RNAi, in that they possess a triphosphorylated 5' guanine residue and exhibit a remarkable strand bias at target loci. Indeed, I show that 22G-RNAs are dependent on the activity of the RdRPs RRF-1 and EGO-1 and function in multiple distinct endogenous small RNA pathways. Interestingly, I have found that RRF-1 and EGO-1 function redundantly in the germline to generate 22G-RNAs that are dependent on and interact with members of an expanded family of worm-specific Argonaute (WAGO) proteins.

The WAGO/22G-RNA pathway appears to be a transcriptome surveillance pathway that silences coding genes, pseudogenes, transposons, and non-annotated, or cryptic, transcripts. In contrast, I have found that EGO-1 alone is required for the biogenesis of a distinct class of 22G-RNAs that interact with the Argonaute CSR-1. Surprisingly, the CSR-1/22G-RNA pathway does not appear to silence its targets transcripts. Instead, the CSR-1/22G-RNA pathway is essential for the proper assembly of holocentric kinetochores and chromosome segregation.

Lastly, I show that a third endogenous small RNA pathway, the ERI pathway, is a two-step silencing pathway that requires the sequential activity of distinct RdRPs and Argonautes. In the first step of this pathway, the RdRP, RRF-3, is required for the biogenesis of 26G-RNAs that associate with the Argonaute, ERGO-1. In the second step, RRF-1 and EGO-1 generate 22G-RNAs that associate with the WAGO Argonautes.

This work demonstrates how several *C. elegans* small RNAs pathways utilize RdRPs to generate abundant populations of small RNAs. These distinct categories of small RNAs function together with specific Argonaute proteins to affect gene expression, to play essential roles in development, and in the maintenance of genome and transcriptome integrity.

Table of Contents

List of Tables

Tables included in Thesis:

The tables listed below have not been included and can be found online in supplementary information in *Cell* Oct 2;139(1):123-34 (II) and in *Molecular Cell* Oct 23;36(2):231-144 (III)

List of Figures

Copyright © Information

The following chapters were published in the following peer-reviewed journals, with permission for use according to copyright licenses below.

Chapter II of this thesis was published as "The Argonaute CSR-1 and its 22G-RNA Cofactors are Required for Holocentric Chromosome Segregation.", *Claycomb JM, *Batista PJ, Pang KM, Gu W, Vasale JJ, van Wolfswinkel JC, Chaves DA, Shirayama M, Mitani S, Ketting RF, Conte D, Jr., Mello CC. (2009); *Cell* Oct 2;139(1):123-34; License number 2374831415649

Chapter III of this thesis was published as "Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the *C. elegans* Germline.", *Molecular Cell* 2009, Oct 23;36(2):231-44; *Gu W, *Shirayama M, *Conte D, Jr., Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, Chen CC, Chaves DA, Duan S, Kasschau KD, Falgren N, Yates JR, 3rd, Mitani S, Carrington JC, Mello CC. License number 2374831415649

Chapter IV of this thesis was published as "Sequential rounds of RNA-dependent RNA Transcription Drive Endogenous Small-RNA Biogenesis in the ERGO-1/Argonaute Pathway", *Vasale JJ, *Gu W, Thivierge C, Batista PJ, Claycomb JM, Youngman EM, Duchaine TF, Mello CC, Conte D, Jr. (2010); Feb 23;107(8):3582-7 *Proceedings of the National Academy of Sciences of the United States of America*. Authors need not obtain permission for use as part of their dissertations

* These authors contributed equally

CHAPTER I

General Introduction

RNA interference

In 1998, Craig Mello and Andrew Fire published their landmark discovery that long double-stranded RNA (dsRNA) triggers potent and specific gene silencing in *C. elegans*, in a process called RNA Interference (RNAi) (Fire et al., 1998; Rocheleau et al., 1997). Since the initial discovery, RNAi has become an invaluable tool to inhibit gene expression and study gene function in numerous biological systems, including those where genetics is intractable (Fjose et al., 2001). Furthermore, advances in RNAi technology have developed beyond basic science research into a strategic tool used in biomedical drug therapeutics (He et al., 2009a; Milhavet et al., 2003; Shrey et al., 2009). The impact of Mello and Fire's discovery was recognized in eight years when they were awarded the Nobel Prize in Physiology or Medicine in 2006 (Zamore, 2006).

The work of Mello and Fire built on several previous studies, which collectively suggested that either sense or antisense RNA could act to silence a target mRNA. For example, Fire showed that transgenic *C. elegans* strains expressing either sense or antisense *unc-22* RNA phenocopied the *unc-22* loss-of-function phenotype (Fire et al., 1991). Homology-dependent gene silencing (HDGS) was also observed by scientists attempted to enhance the pigmentation of petunia plants by over-expressing the enzyme required for pigment biosynthesis (Napoli et al., 1990; van der Krol et al., 1990). Remarkably, these genetically engineered transgenic plants produced flowers that lacked pigment altogether and were completely white or, at best, variegated in color. This HDGS

phenomenon was dubbed “co-suppression”, because it resulted in the silencing of both the transgenic and the endogenous copies of the gene. Molecular analysis revealed that co-suppression was a result of post-transcriptional gene silencing (PTGS), because mRNA levels for the enzyme were markedly reduced while the rate of transcription appeared normal (de Lange et al., 1995). Thus, RNAi is believed to have emerged as a natural defense mechanism to protect the genome from invading nucleic acids, such as viruses or transposable elements (Baulcombe, 1999; Kumar, 2008; Ratcliff et al., 1997; Ratcliff et al., ; Schutz and Sarnow, 2006; Tabara et al., 1999; Voinnet et al., 1999; Wu-Scharf et al., 2000).

Since the initial description of RNAi in *C. elegans*, it has become clear that RNAi and related silencing pathways function in a multitude of eukaryotic organisms including plants, trypanosomes, hydra, fungi, insects, fish and mammals (Caplen et al., 2001; Kennerdell and Carthew, 1998, 2000; Li et al., 2000; Lohmann et al., 1999; Ngo et al., 1998; Romano and Macino, 1992; Sanchez Alvarado and Newmark, 1999; Wianny and Zernicka-Goetz, 2000). The fact that gene expression could be inhibited by dsRNA in diverse organisms implied the existence of a conserved gene silencing mechanism. Organisms use RNAi-related pathways to regulate development, transposon silencing and chromatin structure. While some small RNA pathways mediate PTGS, via translational repression or mRNA degradation, other pathways mediate transcriptional gene silencing (TGS), via chromatin remodeling and formation of heterochromatin (Almeida and Allshire, 2005; Buhler and Moazed, 2007; Meister

and Tuschl, 2004; Mochizuki and Gorovsky, 2004b; Morris, 2008; Pal-Bhadra et al., 2002; Pal-Bhadra et al., 2004). Surprisingly, in ciliated protozoa, small RNA pathways even mediate DNA elimination (Mochizuki and Gorovsky, 2004a, b; Yao et al., 2003).

A number of lines of evidence suggested that RNAi and other gene silencing phenomena were mechanistically related, and were likely to rely on conserved families of proteins. First, the identification of small RNA species that accumulated under RNAi conditions implicated these molecules in gene silencing functions (Hamilton and Baulcombe, 1999; Zamore, 2001). Second, the characterization of these small RNAs and their generation from longer dsRNA implied the existence of an RNAse enzyme to generate small RNAs (Elbashir et al., 2001b; Zamore et al., 2000). Biochemical studies identified such an RNAse, which came to be known as Dicer (Bernstein et al., 2001). Third, genetic screens in plants, *C. elegans*, and *N. crassa* identified Argonaute proteins as key factors required for RNAi and related silencing phenomena (Cogoni, 2001; Tabara et al., 1999). Subsequent biochemical and structural studies established the pivotal role of Argonaute proteins as the binding partners of small RNAs and effectors of gene silencing (Bohmert et al., 1998; Fagard et al., 2000; Tabara et al., 1999). Collectively, numerous elegant genetic and biochemical studies have revealed the molecular details of these silencing pathways.

Small RNAs

Biochemical studies established that dsRNA is cleaved into 18- to 25-nucleotide (nt) small RNAs, which are the sequence-specific modulators of RNAi. Small RNAs of ~25 nt were first observed to accumulate during PTGS in plants (Hamilton and Baulcombe, 1999). The PTGS-associated small RNAs were of both sense and antisense polarities with respect to the mRNA, suggesting that they were not simply degradation products from the targeted mRNA (Hamilton and Baulcombe, 1999). Using *Drosophila melanogaster* syncytial blastoderm extracts, Sharp and colleagues developed the first *in vitro* RNAi system, demonstrating that dsRNA triggers the specific degradation of a homologous mRNA (Tuschl et al., 1999). *In vitro* experiments further showed that dsRNA is a precursor that is processed into 21-23 nt small RNA products and that these small RNAs, referred to as short interfering RNAs (siRNAs), mediate silencing (Elbashir et al., 2001b; Elbashir et al., 2001c; Parrish et al., 2000; Zamore et al., 2000). Indeed, delivery of synthetic 21- or 23 nt RNA duplexes in *D. melanogaster*, plants, and vertebrate cells was capable of effectively triggering RNAi, indicating that siRNAs are directly involved in the RNAi process (Caplen et al., 2001; Elbashir et al., 2001a; Klahre et al., 2002; Lewis et al., 2002).

Chemical alterations in the sense strand were more tolerated than antisense, indicating that the two strands were not functionally equivalent (Khvorova et al., 2003; Parrish et al., 2000). It was ultimately determined that the two strands possess mechanistically distinct roles in triggering silencing. Only

one strand of an siRNA duplex (the “guide strand”) is incorporated into the functional RNA-induced silencing complex, or RISC, and provides the sequence specificity required for silencing (Hammond et al., 2000). Loading of this guide strand into RISC is determined in part by thermodynamic stability of small RNA duplex (Khvorova et al., 2003; Schwarz et al., 2003). The other strand, called the passenger strand, is ultimately degraded (Nykanen et al., 2001; Schwarz et al., 2003).

Dicer

The underlying catalytic mechanism of small RNA biogenesis was quickly revealed by the identification of Dicer, an RNAse III like enzyme that is capable of cleaving long dsRNA into siRNA duplexes and is required for RNAi (Bernstein et al., 2001; Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001) Dicer belongs to a highly conserved family of dsRNA-specific RNAses, which encode an N-terminal DExH/DEAH RNA helicase domain, a PAZ (Piwi/Argonaute/Zwille) domain, two tandem RNase III catalytic domains (RNase IIIa and RNase IIIb) and a dsRNA-binding domain at the C-terminus (Zhang et al., 2004) (Figure I-1). Some organisms, like flies and plants, possess multiple Dicer orthologs, while *C. elegans* only has one Dicer, DCR-1 (Liu et al., 2009). In *C.elegans*, the product of Dicer cleavage is a ~21 nt small RNA duplex with 3′ dinucleotide overhangs (which are characteristic of RNAse III enzymes) and each strand bearing 5′ monophosphates and 3′ hydroxyls (Elbashir et al., 2001b)

Figure I-1

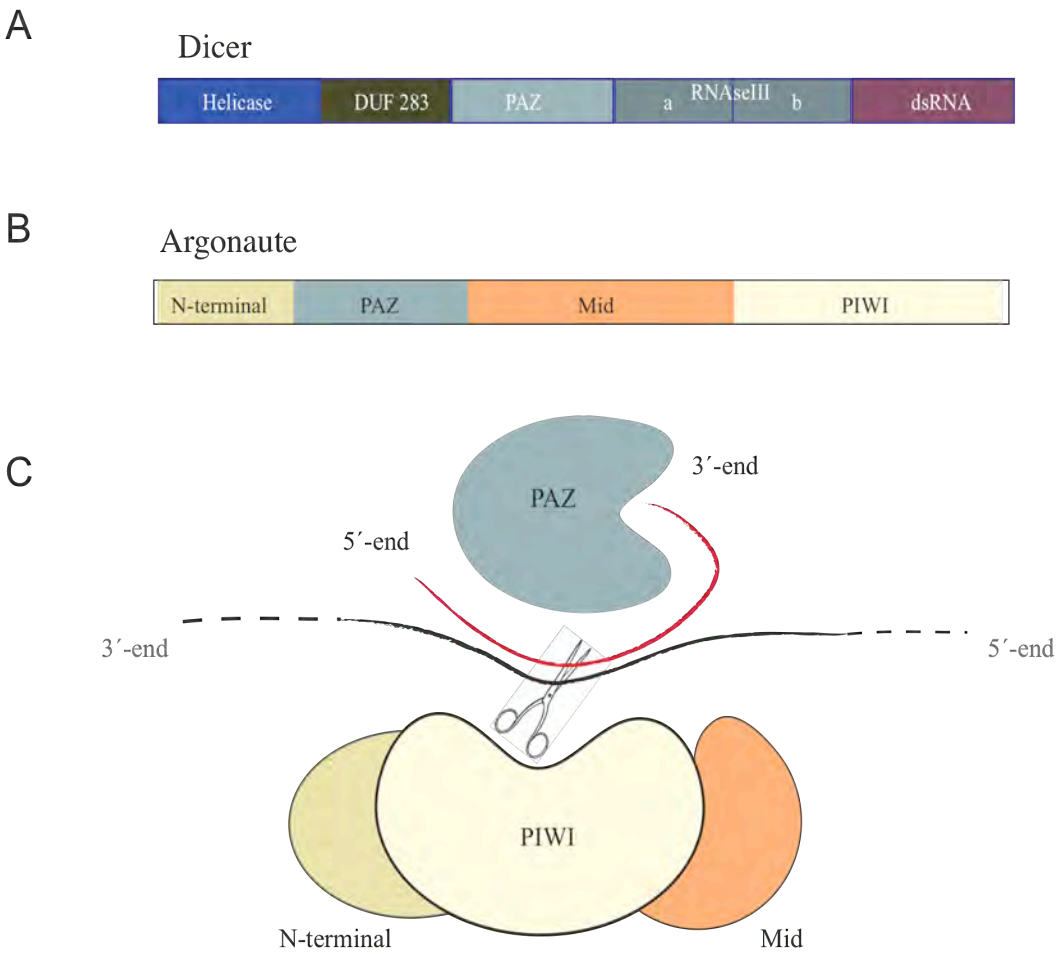

Figure I-1. Domains present in Dicing and Slicing RNAi Proteins

(A) Protein domains of Dicer and (B) Argonautes in *C. elegans*
(C) Schematic representation of of Argonaute association with small RNA during mRNA cleavage (adapted from (Liu and Paroo, 2010))

Dicer products can vary in length. For example, in *A. thaliana* the activity of the Dicer-like proteins DCL2, DCL3 and DCL4 generate 22-, 24-, and 21- nt long siRNA species, respectively (Xie et al., 2004). Structural studies of Dicer suggest that this difference is a result of the distance between the PAZ and the catalytic centers of the RNase III domains, which thus acts as a molecular ruler (Gan et al., 2006; Ji, 2008; MacRae and Doudna, 2007).

Argonautes

The catalytic heart of the RISC effector complex is a protein of the Argonaute family. Argonaute proteins are an evolutionarily conserved family of proteins that are essential for small RNA-mediated silencing pathways. Because of their role as effectors in RNAi, Argonautes have been studied in many different organisms, including worms (*C. elegans*), fruit flies (*D. melanogaster*), yeast (*S. pombe*), plants (*A. thaliana*) and mammals, among others (Boisvert and Simard, 2008; Ghildiyal and Zamore, 2009). Argonaute proteins can be further divided phylogenetically into the Argonaute subfamily, the Piwi subfamily, and the *C. elegans* worm-specific Argonautes (WAGO) subfamily (Boisvert and Simard, 2008; Hock and Meister, 2008; Yigit et al., 2006). The number of Argonaute genes varies widely among species. The *Drosophila* genome encodes at least five distinct proteins, humans possess eight, and *C. elegans* encode the largest number, at 23, including 12 WAGOs (Hutvagner and Simard, 2008). Although multiple Argonaute proteins are often expressed in a single organism, many

proteins are functionally distinct or expressed in specific tissues, allowing them to interact with different classes of small RNAs, and rendering them necessary at different steps of small RNA silencing mechanisms (Okamura et al., 2004; Yigit et al., 2006).

Argonaute proteins consist of four distinct structural domains: the N-terminal domain, a central PAZ domain (which is also found in Dicer proteins), a Mid domain, and C-terminal PIWI domain (Collins and Cheng, 2006) (Figure I-1B). Argonaute proteins interact with small RNAs in a sequence-independent manner via the PAZ domain, Mid domain, and PIWI domains (Collins and Cheng, 2006) (Figure I-1C). The PIWI domain faces the PAZ domain and is structurally similar to RNase H enzymes; two conserved aspartates and a conserved histidine define the DDH motif and are essential for cleavage of mRNAs, referred to as “Slicer” activity. Argonautes that possess these catalytic residues cleave the complementary mRNA between nucleotide positions +10 and +11 of the respective siRNA (with the 5′-end of the guide siRNA labeled the +1 position and acting as the ruler to determine the position of the target mRNA cleavage) (Elbashir et al., 2001b; Elbashir et al., 2001c; Schwarz et al., 2004). The human Argonaute hAGO2, the *D. Melanogaster* AGO2, and the *C. elegans* CSR-1 have all been shown to possess siRNA-directed Slicer activity *in vitro* (Aoki et al., 2007; Rivas et al., 2005).

Features of RNAi

In *C. elegans,* several experimental methods can be used to trigger RNAi. Worms can be soaked in dsRNA solution or fed with bacteria that express dsRNA (Tabara et al., 1998; Timmons et al., 2001; Timmons and Fire, 1998). Additionally, dsRNA can be expressed *in vivo* via hairpin or sense and antisense transgenes (Timmons et al., 2003). Lastly, dsRNA can be administered by microinjection into the body cavity of *C. elegans* (Fire et al., 1998).

The initial studies of RNAi in worms and plants described several noteworthy observations that are important to understanding the molecular mechanism of gene silencing. First, the silencing effect is systemic. In *C elegans*, injection of dsRNA into the intestine of a worm, or feeding on bacteria that express dsRNA, elicits an RNAi response throughout the body of the worm (Fire et al., 1998; Timmons et al., 2001). The ability of dsRNA to trigger a silencing response and spread to other cells is not unique to *C. elegans* (Newmark et al., 2003; Roignant et al., 2003; Tabara et al., 1998; Voinnet, 2005; Voinnet et al., 1998). For example, in tobacco plants, systemic silencing can spread from silenced stocks (root systems) to non-silenced scions (vegetative tissues) (Palauqui et al., 1997). These observations implicate active RNA uptake or transport systems in *C. elegans* and plants. The second key observation is that injection of dsRNA at a concentration of only a few molecules per cell was sufficient to trigger a potent RNAi response in *C. elegans* (Fire et al., 1998), indicating that RNAi is a catalytic process or that silencing signals are amplified.

Third, although RNAi appeared to be relatively transient, it could be transmitted into subsequent generations. That is, the progeny of an animal or plant exposed to dsRNA were also affected (Chuang and Meyerowitz, 2000; Fire et al., 1998).

Studies aimed at investigating the systemic nature of RNAi in *C. elegans* indentified several mutants that were defective in systemic RNAi (Hunter et al., 2006; Tijsterman et al., 2004). Two of these genes, *sid-1* and *sid-2*, encode transmembrane proteins required for uptake of dsRNA (Hunter et al., 2006; Winston et al., 2007). A homolog of SID-1 has been identified in the human genome, raising the possibility of a conserved mechanism of dsRNA transport (Duxbury et al., 2005).

Steps of RNAi in *C. elegans*

In *C. elegans*, exogenous RNAi (exo-RNAi) requires two distinct phases, referred to as the initiation (or primary) phase and the maintenance (or secondary) phase (Grishok et al., 2000) (Figure I-2). In the primary phase, Dicer-dependent siRNAs are loaded onto an Argonaute complex resulting in gene silencing that is similar to the canonical RNAi pathway, described above, which occurs in *D. melanogaster* and mammals (Ghildiyal and Zamore, 2009). However, in *C. elegans* the primary step in the pathway is not sufficient for gene silencing, and amplification of silencing signals is necessary to achieve gene silencing. Although the mechanistic details of the secondary step are

Figure I-2

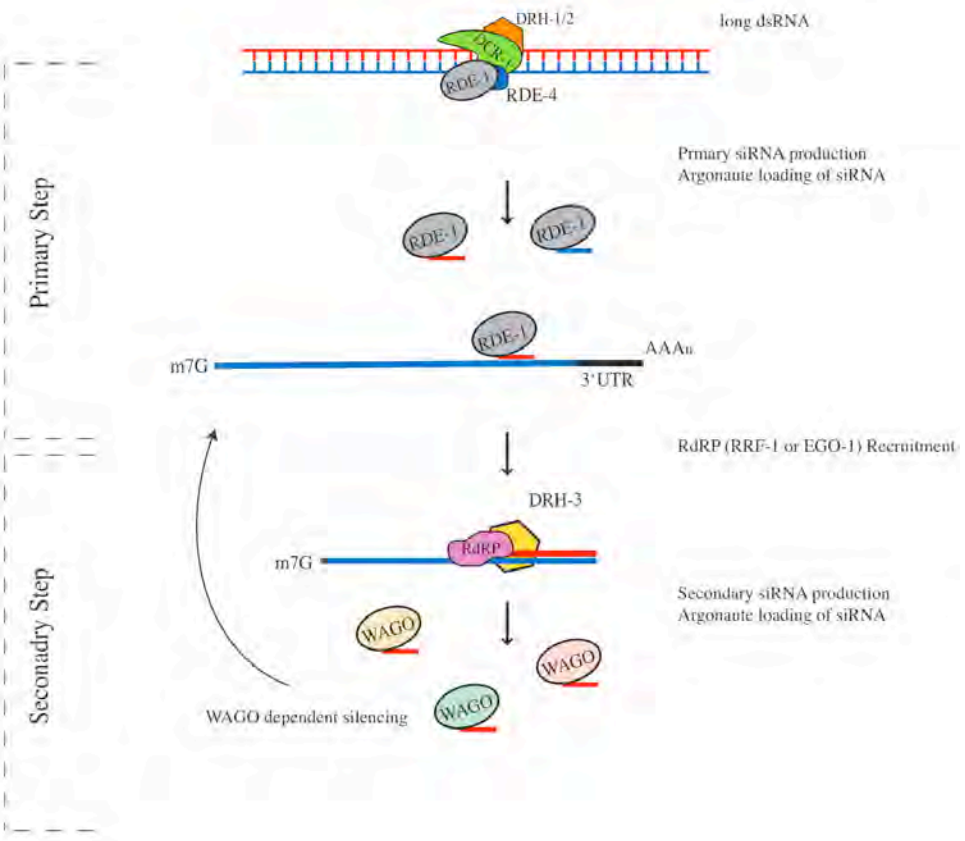

Figure I-2. RNAi Pathway in *C. elegans*

The two steps of the exo-RNAi pathway of *C. elegans*. Primary siRNAs are generated by Dicer and interact with RDE-1. mRNAs identified by the RDE-1/primary siRNA complex are utilized by RdRPs as a template in the generation of WAGO interacting secondary siRNAs (22G-RNAs).

unknown, it is now understood that RNA-dependent RNA Polymerases (RdRP) amplify the silencing process by generating secondary siRNAs. These secondary siRNAs interact with the WAGO subfamily of Argonautes and are thought to promote silencing of the targeted mRNA. In addition, genes of the *mutator* class are also involved in the secondary phase of RNAi. In the sections below, I will discuss the known roles of these proteins in small RNA biogenesis in *C. elegans,* as well as provide links to conserved RNAi-related silencing pathways in other organisms.

Genetic screens to isolate RNAi-deficient (*rde*) mutants were pioneered in *C. elegans* and yielded several complementation groups, including *rde-1*, *rde-2*, *rde-3*, and *rde-4* (Tabara et al., 1999). Some of the mutants that were found to be defective for RNAi were also indentified to be part of the mutator (*mut*) class of genes, as they are defective in silencing transposons in the germline (Ketting et al., 1999; Tabara et al., 1999). Genes in the *mut* class include *mut-2/rde-3*, *mut-7*, *mut-8/rde-2*, *mut-9, mut-14, mut-15/rde-5* and *mut-16/rde-6*. Additional phenotypes associated with *mut*-class Rde mutants include elevated X chromosome non-disjunction (leading to a high incidence of males or Him phenotype), sterility at 25°C, and de-repression of normally silenced transgene arrays in the germline (Ketting et al., 1999; Ketting and Plasterk, 2000; Tabara et al., 1999).

Subsequent analyses, examining the genetic requirements for heritability of RNAi, in addition to biochemical studies, elucidated that the RNAi pathway

acts in sequential molecular steps. RDE-4, a double-stranded RNA-binding protein, is required for binding to the initial trigger dsRNA (Parker et al., 2006) and forms a complex with Dicer, and the Argonaute RDE-1 (Grishok et al., 2000; Tabara et al., 1999; Tabara et al., 2002). This complex, together with the Dicer-related helicase proteins, DRH-1 and DRH-2 (DRH-1/2), make up the recognition and initiation complex that is required to mount a silencing response using exogenous dsRNA in *C. elegans* (Tabara et al., 2002).

Dicer-dependent siRNAs, referred to as primary siRNAs in *C. elegans*, are loaded onto RDE-1 (Sijen et al., 2007; Yigit et al., 2006). RDE-1 interacts with the targeted mRNA through the homology between the mRNA and the guide siRNA. RDE-1 contains the conserved catalytic residues required for RISC slicer activity and is thought to cleave the mRNA, although it has not yet been shown to do so (Aoki et al., 2007; Yigit et al., 2006).

In *C. elegans,* primary siRNAs are not sufficient to elicit a silencing response. RdRPs play an essential role in RNAi (Table I-1). Not only are mRNAs targets of gene silencing, they are also used by RdRP proteins as templates for *de novo* synthesis of a secondary pool of antisense small RNAs, referred to as secondary siRNAs (Aoki et al., 2007; Sijen et al., 2001; Sijen et al., 2007). These secondary siRNAs form a distinct class of siRNAs that have a 5´ triphosphate

Table I-1

RNA-dependent RNA Polymerase	Mutant Phenotype	Primary Tissue of Function	Small RNA Pathway		Required for Small RNA Classes
EGO-1	sterile at all temperatures	germline	exo-RNAi	endo-RNAi, CSR-1 pathway	22G-RNAs
RRF-1	wild-type, fertile	soma, some germline redundancy with EGO-1	exo-RNAi	endo-RNAi, WAGO-1, Eri pathway	22G-RNAs
RRF-2	fertile	N/A	N/A	N/A	N/A
RRF-3	sterile at non-permissive temperatures	germline and soma	N/A	endo-RNAi, Eri pathway	26G-RNAs

Table I-1. Functional Characteristics of RNA-dependent RNA Polymerase Family Members of *C. elegans*

and a bias for guanosine (G) at the 5´-end (Aoki et al., 2007; Pak and Fire, 2007; Sijen et al., 2007). The secondary siRNAs are then loaded onto the functionally redundant WAGO Argonautes (Yigit et al., 2006).

Other proteins required for RNAi, such as RDE-2 (Tops et al., 2005) RDE-3 (a nucleotidyltransferase family member) (Chen et al., 2005), MUT-7 (an RNAse D 3′-5′ exonuclease) (Ketting et al., 1999) and MUT-14 (a DEAD-box RNA helicase) (Tijsterman et al., 2002) are not required for the initial phase of RNAi, but are required for the secondary phase (Grishok et al., 2000; Tops et al., 2005). The roles of these factors have not been defined at a mechanistic level.

Classes of Small RNAs in *C. elegans*

Each class of small RNA can be distinguished by the mechanism of biogenesis, length of small RNA, composition of 5′ nucleotide, modifications on 5′ or 3′ nucleotides and the Argonaute proteins with which they associate (Bartel, 2004; Ghildiyal and Zamore, 2009; Golden et al., 2008; Kim et al., 2009) (Table I-2). Four main classes of small RNAs have been identified in *C. elegans*, including microRNAs (miRNAs), Piwi-interacting small RNAs (21U/piRNAs) and the endogenous 22G- and 26G-RNAs.

miRNAs

The founding family of endogenous small RNAs, miRNAs were identified in *C. elegans* as temporal regulators of development and have since been found

Table I-2

Classes of small RNA	Average Nucleotide Length	RdRP-Dependent	Dicer-Dependent	Associated Argonautes	Biochemical Modification 5'-end	3'-end
miRNAs	21-22nt	No	Yes	ALG-1, ALG-2, RDE-1	monophosphate	hydroxyl
piRNAs	21nt	No	No	PRG-1, PRG-2	monophosphate	2'-O-methyl
22G-RNAs	22nt	Yes	No	CSR-1, WAGO	triphosphate	hydroxyl, uridylation
26G-RNAs	26nt	Yes	Yes	ERGO-1	monophosphate	2'-O-methyl

Table I-2 Classes and Characteristics of Endogenous Small RNAs of *C. elegans*
See review (Ghildiyal and Zamore, 2009)

to be a highly conserved class of small RNAs in many organisms, including humans (Carthew and Sontheimer, 2009; Frasch, 2008; Lee et al., 1993; Lim et al., 2003a; Lim et al., 2003b; Resnick et al., 2010). miRNAs regulate the expression of protein coding genes by binding with imperfect complementarity to the 3′ untranslated regions of target mRNAs to inhibit translation (Lee et al., 1993). In some cases, when there is nearly perfect base-pairing between a miRNA and a target mRNA, silencing results from mRNA degradation (Bartel, 2009; Chekulaeva and Filipowicz, 2009; Meister et al., 2004; Olsen and Ambros, 1999).

In animals, miRNA biogenesis requires sequential processing steps by the RNAse III related proteins Drosha and Dicer. Within the nucleus, primary miRNAs (pri-miRNAs) are processed into premature miRNAs (pre-miRNAs) by a DROSHA (Denli et al., 2004; Han et al., 2004; Yeom et al., 2006). Pre-miRNAs are exported to the cytoplasm (Yi et al., 2003) and processed by Dicer into mature miRNAs, which are subsequently loaded onto the miRNA-specific Argonautes, ALG-1 and ALG-2 in *C. elegans* (Grishok et al., 2001; Hutvagner et al., 2001; Jannot et al., 2008; Tops et al., 2006). In animals, miRNAs that reside within introns, known as mirtrons, are processed in a DROSHA-independent manner (Okamura et al., 2007; Ruby et al., 2007). Interestingly, mirtron pre-miRNAs are generated during intron processing by the splicing machinery.

piRNAs

piRNAs are a unique class of animal-specific small RNAs. In *C. elegans* piRNAs are 21nt long and are ~30nt long in other animals (Thomson and Lin, 2009). They are highly expressed in germline cells where they are required for proper germline development and gametogenesis (Hartig et al., 2007; He et al., 2009b; Klattenhoff and Theurkauf, 2008; Lau et al., 2006; Nishida et al., 2007; Updike and Strome, 2009). Generally originating from genomic repeats or transposons, piRNAs physically interact with Piwi clade of Argonaute proteins (Aravin and Hannon, 2008; Batista et al., 2008; Brennecke et al., 2007; Das et al., 2008; Girard et al., 2006; Lau et al., 2001; Vagin et al., 2006). Biogenesis of piRNAs is distinct from that of miRNAs, in that the accumulation of piRNAs is independent of Dicer activity. In *Drosophila*, retrotransposon-derived piRNA biogenesis occurs via reciprocal Slicer cleavage by two distinct PIWI Argonautes, Argonaute-3 (AGO3) and Aubergine (AUB). The AGO-3 sense-piRNA cleavage event specifies the 5′-end of the AUB associated antisense piRNAs, and the AUB antisense piRNA cleavage event defines the 5′-end of AGO-3 associated piRNAs, resulting in a so-called “ping-pong” feedback loop (Brennecke et al., 2007; Gunawardane et al., 2007). It is unclear how the 3′-ends of these piRNAs are defined, although it is known that 3′ nt is modified, possessing 2′-O-methyl group (Horwich et al., 2007; Saito et al., 2007).

The *C. elegans* piRNAs have a bias for a 5′ monophosphorylated uracil (U) and are also methylated at the 3′ end (Ruby et al., 2006; Saito et al., 2007).

Individual piRNAs share a common upstream motif and primarily map to two large genomic clusters on chromosome IV (Batista et al., 2008; Ruby et al., 2006). In *C. elegans*, the biogenesis of piRNAs is currently unclear, but there is no evidence for a “ping-pong” amplification cycle in 21U-RNA biogenesis (Batista et al., 2008; Das et al., 2008).

Among the 23 Argonaute proteins encoded in *C.elegans*, only PRG-1 is essential for the biogenesis of piRNAs (Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008). PRG-1 is expressed in the germline throughout development, where it localizes to nuage-like structures called P-granules (Batista et al., 2008). P-granules are germ cell-specific cytoplasmic bodies that associated with nuclear pores and are linked to the nuclear export of nascent mRNA (Sheth et al., 2010).

Endo-siRNAs

Over the last several years a large number of endo-siRNAs that are distinct from miRNAs and piRNAs have been identified in *C. elegans* (Ambros et al., 2003; Ruby et al., 2006). These endo-siRNAs fall into two broad classes, based on their average size and the requirements for their biogenesis: the 22nt long 22G-RNAs and the 26nt long 26G-RNAs (Table I-2). Directly produced by the RdRPs EGO-1 and RRF-1, the 22G-RNAs possess a strong bias for a 5’ guanosine (G) that is triphosphorylated, and target protein coding genes, pseudogenes, transposable elements and noncoding regions of the genome

(Ambros et al., 2003; Ruby et al., 2006). The 26G-RNAs target protein coding genes, pseudogenes and noncoding genomic elements. Although 26G-RNAs possess a 5′ G that is a characteristic of RdRP products, they are 5'-monophosphorylated, require Dicer for their production, and rely on the activity of a different RdRP, known as RRF-3. The data presented in this thesis serve to clarify the role of distinct RdRPs in these different small RNA pathways.

The RNA-dependent RNA polymerases of *C. elegans*

There are four RdRPs encoded in the *C. elegans* genome*: ego-1, rrf-1, rrf-2*, and *rrf-3* (Sijen et al., 2001) (Table I-1). Thus far, EGO-1, RRF-1, and RRF-3 have been shown to play roles in both exogenous RNAi and endogenous small RNA pathways in *C. elegans*. A role for RRF-2 in gene silencing, if any, remains a mystery.

EGO-1 and RRF-1 in RNAi and *C. elegans* development

EGO-1 (enhancer of *glp-1*) was the first RdRP shown to be required for RNAi in *C. elegans*. It was identified in a screen designed to isolate genes required for germline proliferation (Qiao et al., 1995). EGO-1 is required for proper development of sperm and oocytes, and *ego-1* mutants display defects in mitotic and meiotic germline chromosome organization and segregation (Qiao et al., 1995; She et al., 2009; Smardon et al., 2000). Subsequent RNAi assays in *ego-1* mutants showed sensitivity to RNAi targeting the somatically-expressed

gene *unc-22*. In contrast, *ego-1* mutants displayed varying degrees of resistance to dsRNA targeting germline-specific genes (Smardon et al., 2000). These observations suggested that EGO-1 functions specifically in germline RNAi and in endogenous silencing pathways in the germline that are essential for viability.

Interestingly, the *rrf-1* and *ego-1* genes are tightly linked, as they are positioned in tandem on chromosome I with less than 1000 nt between them. While EGO-1 is essential for germline RNAi, studies of RRF-1 revealed that it is required for RNAi in the soma. Remarkably, RRF-1 was also required for transitive RNAi, or the spreading of secondary siRNAs upstream of the sequence targeted by the dsRNA trigger (Sijen et al., 2001). RRF-1 RdRP activity is dependent on its interaction with the Dicer-related helicase, DRH-3 (Aoki et al., 2007), a paralog of the DRH-1/-2 proteins that interact with DCR-1 (Tabara et al., 2002). Loss-of-function mutations in *rrf-1,* however, display no defects in fertility or viability (Sijen et al., 2001).

RRF-3 and the ERI Pathway

In contrast to the Rde phenotypes exhibited in the absence of EGO-1 or RRF-1, loss-of-function mutations in the RdRP *rrf-3* display an enhanced RNAi phenotype (Eri) (Simmer et al., 2002). Mutations in *rrf-3* also display enhanced silencing of transgenes in the soma suggesting that RRF-3 functions in a pathway that antagonizes both the exogenous RNAi pathway and the transgene silencing pathway (Simmer et al., 2002). One model for the enhanced RNAi

phenotype is that an endogenous small RNA pathway involving RRF-3 competes for factors that also function in the exogenous RNAi pathway. When RRF-3 activity is lost, the endogenous small RNA pathway is disabled, and the shared factors are liberated leading to increased activity in the exogenous RNAi pathway, i.e. enhanced RNAi. In support of this model, a sextuple *wago* mutant is defective for RNAi and fails to accumulate ERI-dependent small RNAs. In contrast, over-expression of two WAGOs, SAGO-1/WAGO-8 and SAGO-2/WAGO-6, results in an enhanced RNAi phenotype (Yigit et al., 2006). These results suggested that the exogenous RNAi and ERI pathways converge at the point of WAGO activity (the effector step of both pathways) and that at the WAGOs are present at limiting levels.

Proteomic studies have shown that RRF-3 interacts with Dicer in a putative “ERI complex.” This ERI complex possesses additional proteins, some of which are involved in RNAi (DRH-3 and RDE-4), as well as other factors that are involved specifically in the ERI pathway (ERI-1, ERI-3 and ERI-5) (Duchaine et al., 2006) (Figure I-3). Because of its composition, this complex is likely to be the ERI pathway initiation complex, akin to the RDE-1::RDE-4::DRH-1/2::DICER complex required for the initial steps of RNAi. Although the WAGO Argonautes appear to function in the downstream effector steps of both the RNAi and ERI pathways, two questions persisted: Which small RNAs do RRF-3 and the ERI complex generate? And, what is the functional equivalent of the Argonaute RDE-1 in the ERI pathway? Both questions are discussed in detail in Chapter IV.

Figure I-3

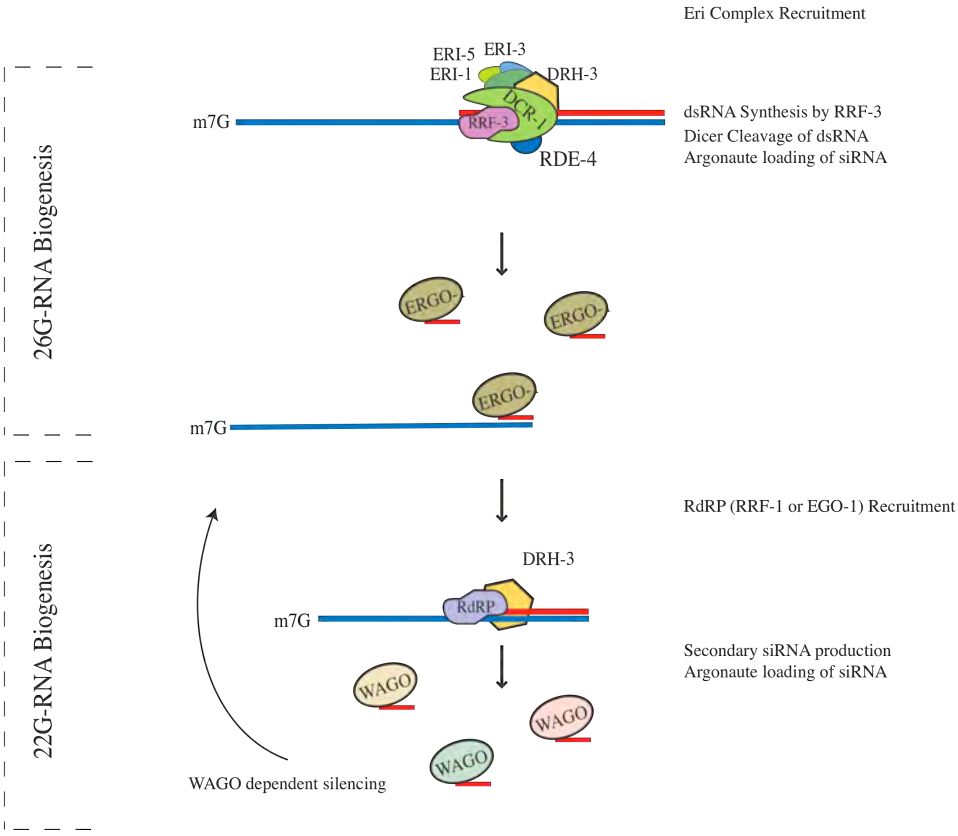

Figure I-3. ERi Pathway in *C. elegans*

Similarly to the exo-RNAi pathway, the endogenous ERI pathway involves two steps of small RNA biogenesis. At the first step, 26G-RNAs are generated through the activity of the ERI complex. At the second step, RNA transcripts targeted by 26G-RNAs are used as templates by RRF-1 and/or EGO-1.

Briefly, clues about the function of RRF-3 and the ERI pathway came from the observation that *rrf-3* mutants are deficient in the accumulation of several species of endogenous siRNAs (Duchaine et al., 2006; Lee et al., 2006). These particular endo-siRNAs were then used to define “ERI” class small RNAs to probe Argonaute mutants, including the ERI Argonaute ERGO-1, for loss of these same small RNAs. Like *rrf-3*, loss of *ergo-1* leads to a depletion of the ERI class of 22G- and 26G-RNAs. In contrast, *rde-1* mutants displayed normal levels of ERI class 22G- and 26G-RNAs, while multiple-fold *wago* mutants showed a loss of the ERI class 22G-RNAs, but not the ERI class 26G-RNAs. Finally, I demonstrated that ERGO-1 directly interacts with a set of ERI class 26G-RNAs, but not 22G-RNAs (see Chapter IV). Collectively, these results indicate that ERGO-1 serves as the primary Argonaute, interacting with 26G-RNAs in the two-step ERI pathway, while WAGOs function with the secondary 22G-RNAs in the downstream steps of this pathway.

RdRPs in other gene silencing pathways

RdRPs were originally discovered as components of RNA viruses where they function to transcribe and replicate the viral genome in host cells. Subsequently, non-viral, or “cellular”, RdRPs have been found to act in gene silencing pathways in a wide range of eukaryotic organisms.

The first cellular RdRP biochemically isolated and subsequently cloned was from tomato plants (T-RdRP) (Schiebel et al., 1993a, b). QDE-1 was the first

RdRP family member to be isolated in a screen for genes required for quelling, a form of co-suppression, in the fungus *N. crassa*. Shortly thereafter, RdRP proteins were isolated in other organisms, including *A. thaliana* (Dalmay et al., 2000), *D. discoideum* (Martens et al., 2002), *S. pombe* (Volpe et al., 2002) and *C. elegans* (Sijen et al., 2001; Simmer et al., 2002; Smardon et al., 2000), and found to be essential in various gene silencing pathways. There are no obvious cellular RdRP homologues in *Drosophila* or mammalian genomes. However, *Drosophila* ELP-1, the largest subunit of the RNA polymerase II elongator complex, was recently shown to possess RdRP activity and to be required for RNAi and transposon silencing (Lipardi and Paterson, 2009). In addition, the human telomerase TERT was recently shown to interact with the mitochondrial RNA processing endoribonuclease RMRP, resulting in a complex with RdRP activity (Maida et al., 2009). The TERT-RMRP RdRP complex appears to negatively regulate RMRP expression by producing RMRP dsRNA that is processed by Dicer into AGO2-asociated siRNAs.

In plants, the RdRP RDR6 (also known as SDE1/SGS2) functions in the trans-acting (TAS) pathway, which is initiated by miRNAs (Vazquez et al., 2004) In this pathway, RDR6 converts single-stranded TAS cleavage products into dsRNA that is processed by Dicer into siRNAs called trans-acting siRNAs (ta-siRNAs). A separate RdRP complex containing RDR2 is involved in RNA-dependent DNA methylation (RdDM) in plants (Herr et al., 2005). In the ciliated protozoan, *T. thermophila*, the RdRP, Rdr1, and Dicer, Dcr2, form a complex that

generates 23-24 nt small RNAs, which function in genome rearrangement and elimination during development (Lee and Collins, 2007),

Extensive studies in *S. pombe* have shown that a sole RdRP, Rdp1, acts with Dicer and an Argonaute in a pathway that is required for centromeric heterochromatin assembly and is essential for centromere function (Buhler and Moazed, 2007; Verdel et al., 2004; Volpe et al., 2003). Transcripts derived from pericentromeric repeats are substrates for Rdp1-dependent synthesis of dsRNA that is processed in a concerted manner by Dcr1 to form siRNAs (Colmenares et al., 2007; Motamedi et al., 2004). The targeting of pericentromeric transcripts by Argonaute-associated siRNAs results in recruitment of heterochromatin modifying and binding proteins and heterochromatin formation. The work described in Chapter II, suggests that the RdRP EGO-1 may play a similar role in regulating *C. elegans* centromeres, albeit through a different mechanism and different type of endo-siRNA.

Scope of Thesis

The overall goal of my thesis work was to understand the distinct and overlapping roles of the different cellular RdRPs (EGO-1, RRF-1 and RRF-3) in RNAi-related silencing pathways in *C. elegans*. One major objective of my project was to uncover endogenous small RNAs that are dependent on the activities of the RdRPs for their biogenesis. Along with others in the lab, I also endeavored to determine which Argonaute effector complexes function together with these distinct classes of small RNAs generated by the different RdRPs. Together,

these studies have led us to understand which RNA transcripts are targeted in different tissues in *C. elegans*, and have provided insights into the functional and biological consequences of silencing particular transcripts throughout development.

In these studies, I used a combination of biochemical, genetic, and cell biological approaches. In addition, the development of cutting edge molecular cloning strategies and techniques aimed at comprehensive analysis of nucleic acids on a genome-wide scale, referred to here as deep sequencing (Fox et al., 2009; Morozova et al., 2009), have been an essential component of the studies detailed in this dissertation. Deep sequencing approaches have allowed us to analyze noncoding small RNAs to an extent that would have otherwise been impossible through traditional genetics and phenotypic analysis alone. By analyzing small RNA populations and mapping the small RNA sequences to specific loci within the genome, we and other researchers have identified novel types of small RNAs and have predicted previously unknown gene targets, thus elucidating the genome landscape of gene regulation by small RNAs. Through comparative small RNA profiling of gene silencing mutants and Argonaute protein-small RNA complexes, deep sequencing has become an invaluable tool in examining the genetic requirements for small RNA biogenesis and elucidating small RNA functions.

By deep sequencing the small RNA complement from wild-type and RdRP mutants, I determined that the RdRPs, EGO-1 and RRF-1 are required for the

biogenesis of an abundant class of antisense endo-siRNAs, called 22G-RNAs. Through detailed molecular, genetic, and cell biology studies, we elucidated that EGO-1 functions in a pathway with the Argonaute CSR-1 to regulate chromosome segregation (Chapter II), while EGO-1 and RRF-1 function redundantly with the WAGOs to modulate gene expression and regulate the transcriptome in the germline (Chapter III).

Through additional deep sequencing and genetic studies, I, and others, have determined that RRF-3 is the sole RdRP required for the production of the 26G-RNAs. We also elucidated that 26G-RNA pathways involve a “two-step” mechanism, where the 26G-RNAs produced by RRF-3 lead to the production of 22G-RNAs by EGO-1 and RRF-1 (Chapter IV). I determined that one of the 26G-RNA pathways regulates the levels of repetitive and cryptic transcripts during embryogenesis. I found that the Argonaute ERGO-1 functions specifically with the 26G-RNAs in this pathway, while other studies have shown that the WAGOs interact with the 22G-RNAs.

To conclude this thesis (Chapter V), I discuss the potential role(s) of the RRF-3-dependent 26G-RNAs that associate with ERGO-1 during embryogenesis. For instance, although I have shown that this embryonic 26G-RNA pathway regulates the levels of repetitive and cryptic transcripts, the functional consequences of regulating such transcripts in the embryo are not understood. My published work, along with recent observations that have arisen

from further analysis of my data, suggest exciting roles for this pathway in chromosome organization, and provide testable models for future studies.

In summary, my work has greatly advanced our understanding of how distinct cellular RdRPs use expressed RNA transcripts as templates to generate a plethora of small RNAs. In turn, these diverse small RNAs play essential roles in the regulation of the transcriptome and chromosome structure throughout development in *C. elegans*.

CHAPTER II

The Argonaute CSR-1 and Its 22G-RNA Cofactors Are Required for Holocentric Chromosome Segregation

Author Contributions

Chapter II: J.Vasale contributed the following to this section: She generated balanced a *ego-1* mutant strain with appropriate genetic markers to allow large-scale selection of *ego-1* homozygous mutants. She prepared large scale populations of *ego-1(om97)* homozygous animals by counter-selection. Also she performed small RNA extraction, purification, and preparation of cDNA libraries in parallel with the congenic wild-type control strain (DA1316). She also contributed to analyses and interpretation of deep-sequencing results. These data are presented in Figure II-5, Figure II-S7, and Table II-S-5. This chapter II has been published as "The Argonaute CSR-1 and its 22G-RNA Cofactors are Required for Holocentric Chromosome Segregation." *****Claycomb JM, *****Batista PJ, Pang KM, Gu W, Vasale JJ, van Wolfswinkel JC, Chaves DA, Shirayama M, Mitani S, Ketting RF, Conte D, Jr., Mello CC. (2009); *Cell* Oct 2;139(1):123-34; License number 2374831415649. ***** These authors contributed equally.

SUMMARY

RNAi-related pathways regulate diverse processes, from developmental timing to transposon silencing. Here we analyze several *C. elegans* factors required for RNAi and chromosome segregation. We show that the Argonaute CSR-1, the RNA-dependent RNA polymerase EGO-1, the Dicer-related helicase DRH-3, and the Tudor-domain protein EKL-1 localize to chromosomes, and are required for the proper alignment of the holocentric kinetochores. In the absence of these factors, the kinetochores appear twisted and fail to orient to opposing spindle poles. Deep-sequence analysis reveals that small RNAs depleted in the mutants and enriched in the CSR-1 immunoprecipitation (IP) complex are antisense to thousands of germline-expressed protein-coding genes that are distributed along the length of each chromosome. Interestingly, in spite of its role in experimentally-induced RNAi, CSR-1 does not appear to down-regulate the mRNA or protein levels of its endogenous targets. Instead, CSR-1 is enriched at target genomic loci in a small RNA-dependent manner by Chromatin IP (ChIP) experiments, supporting a model in which CSR-1 and its interacting small RNAs target protein-coding regions distributed along the length of each chromosome to promote the proper organization of chromatin and alignment of the holocentric kinetochores at metaphase.

INTRODUCTION

In many organisms, centromeric regions are flanked by repetitive sequences that assemble into densely packed heterochromatin (Carroll and Straight, 2006; Vos et al., 2006). These pericentromeric heterochromatin domains are thought to play a role in stabilizing kinetochores, the proteinaceous structures to which spindle attachments are made (Cheeseman and Desai, 2008; Welburn and Cheeseman, 2008). In plants (*A. thaliana*) (Kasschau et al., 2007), fission yeast (*S. pombe*) (Buhler et al., 2008; Reinhart and Bartel, 2002), and fruit flies (*D. melanogaster)* (Brennecke et al., 2007), deep-sequencing studies have identified abundant endogenous small RNAs derived from repetitive regions, including the pericentromeric heterochromatin.

In *S. pombe*, transcripts generated from the repetitive pericentromeric regions become substrates for an RNA-dependent RNA polymerase (RdRP). After processing by the ribonuclease Dicer, small RNAs derived from these transcripts are loaded into an Argonaute complex (the RNA-induced transcriptional silencing complex; RITS). The RITS complex targets pericentromeric heterochromatin and is thought to function in a feedback loop to reinforce chromatin marks that stabilize centromeres during mitosis (Buhler et al., 2007).

Not all organisms exhibit repetitive heterochromatin domains associated with centromeric regions. A striking example of this is the organization of the holocentric chromosomes of nematodes (Dernburg, 2001). Holocentric, or

holokinetic, chromosomes were first described over 100 years ago, in a series of elegant cytological studies by Theodor and Marcella Boveri. In these classic studies, the large presomatic germline chromosomes of the parasitic nematode, *Parascaris*, were shown to make multiple spindle attachments along their length (Pimpinelli and Goday, 1989; Satzinger, 2008). Remarkably, in the somatic cells of the early embryo, the large germline chromosomes were observed to undergo fragmentation resulting in the elimination of heterochromatin and the production of over 40 small, euchromatic chromosomes that comprise the somatic genome. Despite the elimination of heterochromatin, these newly formed chromosomes continued to exhibit holocentric features including continuous kinetochores and multiple spindle attachments along their lengths (Goday et al., 1992).

Although *C. elegans* chromosomes do not exhibit chromosomal fragmentation, they are similar to the somatic chromosomes of *Parascaris* in that they are largely euchromatic and exhibit a well-defined holokinetic structure (Albertson and Thomson, 1982). Despite superficial differences, the kinetochores of holocentric and monocentric chromosomes are assembled from a set of highly conserved proteins (Maddox et al., 2004), including the histone variant HCP-3/ CENP-A. However, unlike monocentric chromosomes, HCP-3/CENP-A is incorporated into nucleosomes along the entire poleward face of condensed holocentric chromosomes (Buchwitz et al., 1999; Nagaki et al., 2005). The underlying sequences required for the assembly of holokinetic centromeres, and the potential involvement of Argonaute/small-RNA pathways in their assembly

and function, have not yet been explored.

In *C. elegans*, previous studies have shown that depletion of *drh-3*, a Dicer-related helicase, or *csr-1*, an Argonaute, result in similar anaphase bridging and chromosome segregation defects (Duchaine et al., 2006; Nakamura et al., 2007; Yigit et al., 2006). Both factors are also required for RNAi (Duchaine et al., 2006; Yigit et al., 2006), and in vitro studies suggest that DRH-3 is required for the synthesis of small RNAs by RdRPs, whereas CSR-1 has been shown to cleave complementary RNA targets when loaded with triphosphorylated small RNAs (Aoki et al., 2007).

Here we have analyzed the role of DRH-3 and CSR-1 in chromosome segregation and have identified endogenous small RNAs that interact with CSR-1. The CSR-1-interacting small RNAs are members of a class of endogenous small RNAs that are neither microRNAs nor piRNAs (Ambros and Lee, 2004; Ambros et al., 2003; Gu et al., 2009; Guang et al., 2008; Pak and Fire, 2007; Ruby et al., 2006). These abundant small RNAs (termed 22G-RNAs) are primarily 22 nucleotides in length, with a 5′ triphosphate and a strong bias for a 5′ guanosine (Ambros et al., 2003; Gu et al., 2009; Ruby et al., 2006). Together with Gu et al. (Gu et al., 2009), we demonstrate that the CSR-1-interacting small RNAs comprise one of two major 22G-RNA pathways. The second 22G-RNA system is dependent on the worm-specific Argonautes (WAGOs) and functions to silence transposons, pseudogenes, and cryptic loci, as well as certain protein-coding genes.

We provide evidence that EGO-1, an RdRP (Smardon et al., 2000), and EKL-1, a tudor-domain protein (Rocheleau et al., 2008), function along with DRH-3 and CSR-1 to promote chromosome segregation. Together, these factors are required for the biogenesis of CSR-1-interacting 22G-RNAs, which, surprisingly, are antisense to thousands of germline-expressed genes. CSR-1 interacts with chromatin at its target loci but does not appear to silence mRNA or protein expression. We hypothesize that the role of CSR-1 in chromosome segregation in *C. elegans* is analogous to that of Ago1 in the *S. pombe* chromosome segregation pathway. However, instead of targeting repetitive pericentromeric heterochromatin, CSR-1 targets protein-coding euchromatic domains to promote their proper organization within the holocentric chromosomes of *C. elegans*.

RESULTS

A Set of RNAi-Related Factors Required for Chromosome Segregation

To identify additional genes that function with *drh-3* and *csr-1* to promote chromosome segregation, we examined the mutant phenotypes of genes previously implicated in RNAi-related pathways for evidence of chromosome segregation defects. We found that one of four RdRP genes, *ego-1* (Smardon et al., 2000), and the tudor-domain-containing gene, *ekl-1* (Rocheleau et al., 2008), exhibited defects in fertility and chromosome segregation, similar to those described previously for *drh-3* and *csr-1* (Duchaine et al., 2006; Nakamura et al., 2007; Yigit et al., 2006). EKL-1 had been implicated in several silencing pathways by RNAi-based screens (Kim et al., 2005; Robert et al., 2005; Rocheleau et al., 2008). We found that a null allele of *ekl-1(tm1599)* was deficient for both germline and somatic RNAi and in addition caused a fully penetrant sterile phenotype (Figure II-S1).

Mutation or RNAi depletion of *drh-3*, *csr-1*, *ego-1*, and *ekl-1* resulted in a similar spectrum of meiotic and mitotic defects. The germlines of each mutant are underproliferated, with nuclei of abnormal shape and size (Duchaine et al., 2006; Maine et al., 2005; She et al., 2009; Vought et al., 2005) (data not shown and see Figure II-3E). Chromosomal abnormalities were evident in DAPI-stained oocytes, which occasionally possessed more than six DAPI-staining bodies (Figure II-1A) (Nakamura et al., 2007; She et al., 2009). One measure of

Figure II-1

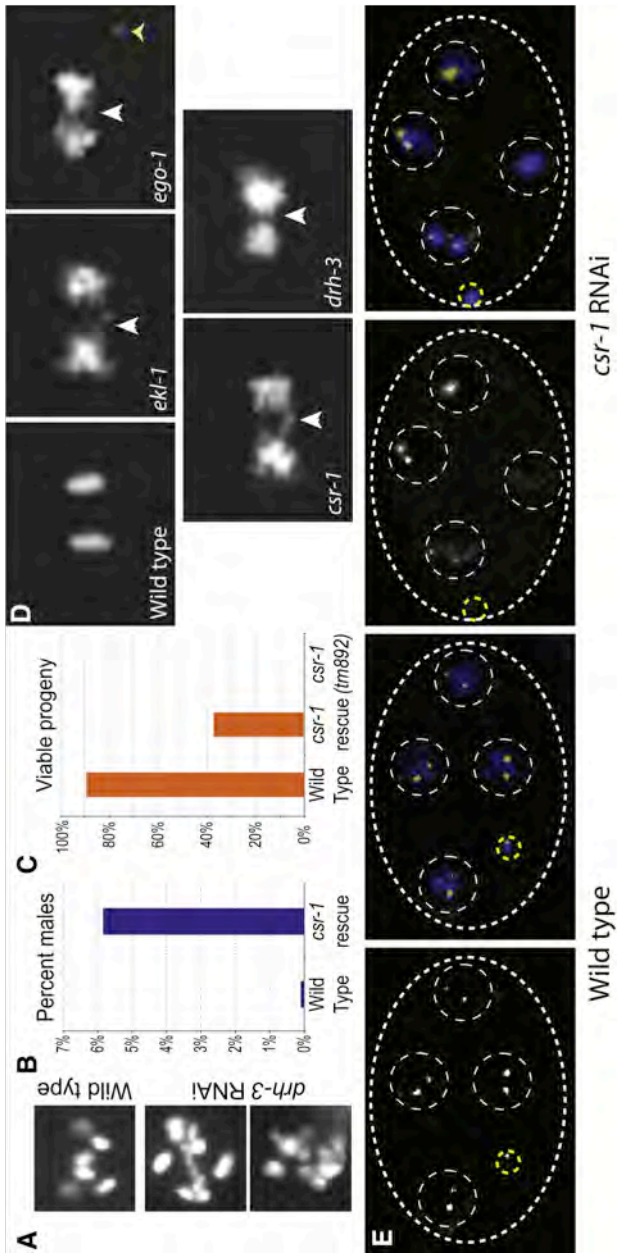
A
Wild type
drh-3 RNAi
B
Percent males
7%
6%
5%
4%
3%
2%
1%
0%
Wild Type
csr-1 rescue
C
Viable progeny
100%
80%
60%
40%
20%
0%
Wild Type
csr-1 rescue
csr-1 (tm892)
D
Wild type
ekl-1
ego-1
csr-1
drh-3
E
Wild type
csr-1 RNAi

Figure II-1
***csr-1*, *ego-1*, *ekl-1*, and *drh-3* Mutants Display Chromosome Segregation Defects in Mitosis and Meiosis**

(A) Diakinetic oocyte chromosomes in wild-type and *drh-3* or *ego-1* RNAi-depleted animals. Six discrete DAPI figures are observed in wild-type, whereas greater than six figures are present in mutant oocytes.(

(B) Incidence of males in wild-type (N2) and 3× Flag *csr-1* rescue.

(C) Viable progeny per brood in wild-type (N2), 3× Flag *csr-1* rescue, and *csr-1(tm892)*.

(D) DAPI-stained wild-type (N2) and RNAi-depleted embryos undergoing the first mitotic division. Anaphase bridging is evident (white arrowhead). An aberrant piece of DNA is visible in *ego-1* (yellow arrowhead).

(E) Fluorescence in situ hybridization with probes for chromosome V 5S rDNA in wild-type and *csr-1* RNAi-depleted embryos (DNA, blue; FISH signal, green). Left panels in each set show FISH signal alone. White dotted lines indicate embryo (large oval) and nuclei (circles). Yellow dotted lines indicate polar bodies. Images are projections of Z stacks through the entire embryo after deconvolution

chromosome segregation defects in the hermaphrodite germline is the proportion of XO male progeny, which arise via spontaneous loss of the X chromosome at a frequency of 0.1%–0.2% in wild-type populations (Meneely et al., 2002). We found that a partially rescued transgenic *csr-1(tm892)* strain (Figure II-S1) generated approximately 6% male progeny (a high incidence of males, or *him*, phenotype) (Figure II-1B). A similar him phenotype was also observed in strains homozygous for hypomorphic alleles of *drh-3* (Gu et al., 2009). These observations suggest that the loss of *csr-1* or *drh-3* can lead to defects in chromosome segregation during either mitotic or meiotic divisions in the germline. Despite the evidence described above for chromosomal abnormalities in the germline, we failed to directly observe mitotic or meiotic chromosome mis-segregation (n = greater than 100 germlines examined, data not shown). In most cases, the dividing nuclei exhibited either wild-type segregation or already contained an abnormal DNA complement. The relative paucity of abnormalities observed in *csr-1(tm892)* germlines could reflect a perdurance of maternally loaded CSR-1.

In addition to the him phenotype, dead embryos were also prevalent in both the *csr-1(tm892)* rescued strain and the hypomorphic *drh-3* strains (Gu et al., 2009). For instance, the *csr-1 (tm892)* rescued strain only generated approximately 38% viable progeny (Figure II-1C). The dead embryos produced by this strain arrested at various points in embryogenesis, up to approximately

the 100-cell stage, and accumulated nuclei with abnormal DNA content. To better examine chromosome morphology and segregation defects in the absence of *csr-1*, *ekl-1*, *ego-1*, and *drh-3*, we used DAPI as well as histone-GFP and tubulin-GFP (Figure II-1D) in RNAi-depleted embryos. Chromosomes appeared to condense during prophase with wild-type timing and morphology. However, as the cell cycle progressed, the following defects were evident during essentially every cell division, beginning with the first cell division of the embryo. At metaphase, chromosomes failed to align into well-organized plates perpendicular to the long axis of the spindle. At anaphase, chromosomal bridging was evident in the spindle midzone (Figure II-1D), and at cytokinesis the lagging chromosomes were bisected by the cleavage furrow. As embryogenesis progressed, abnormally shaped nuclei, with greater or less than wild-type chromosomal complements, accumulated until cell division arrested at about the 50-cell stage (visible in Figures II-S2, II-S3, and II-S5).

To examine the chromosome segregation abnormalities resulting from loss of these RNAi factors at the molecular level, we utilized fluorescence in situ hybridization (FISH) with 5S rDNA probes to chromosome V. Of 32 wild-type embryos, only two showed aberrant FISH signals in one or more nuclei (van Wolfswinkel et al., 2009). In contrast, more than half (10/19) of the *csr-1*-depleted embryos displayed abnormal numbers of FISH-positive chromosomes along with a range of additional abnormalities including aberrantly sized and shaped nuclei (Figures II-1E and II-S2).

DRH-3, EKL-1, EGO-1, and CSR-1 Promote the Proper Organization and Alignment of Metaphase Chromosomes

We next examined three related aspects of chromosome structure that are essential for faithful chromosome segregation: kinetochore formation, condensin loading, and cohesin loading. During mitotic divisions in wild-type C. elegans embryos, HCP-3 localizes to the poleward faces of metaphase chromosomes (Buchwitz et al., 1999; Oegema et al., 2001). In *csr-1*, *drh-3*, *ekl-1*, and *ego-1* RNAi-depleted embryos, HCP-3 was loaded onto chromosomes but was dramatically disorganized. Instead of poleward localization on both sides of the metaphase plate, HCP-3 was distributed over the metaphase chromosomes in an interrupted pattern that extended throughout the spindle midzone (Figures II-2A and II-2B). This pattern could reflect a defect in chromosome alignment and/or compaction or could indicate that, even though HCP-3 is loaded, it is not targeted to the appropriate regions of the chromosome. Finally, to assess whether the kinetochores were fully assembled in *csr-1*, *drh-3*, *ekl-1*, and *ego-1* RNAi-depleted embryos, we examined the outer kinetochore protein KLP-7/MCAK (a kinesin) and the conserved spindle checkpoint protein BUB-1 (Oegema et al., 2001). Both were loaded onto mitotic chromosomes in the RNAi-depleted embryos but were disrupted in a manner similar to HCP-3 (Figures II-2C and II-S3).

Because the observed chromosome segregation defects could result from

Figure II-2

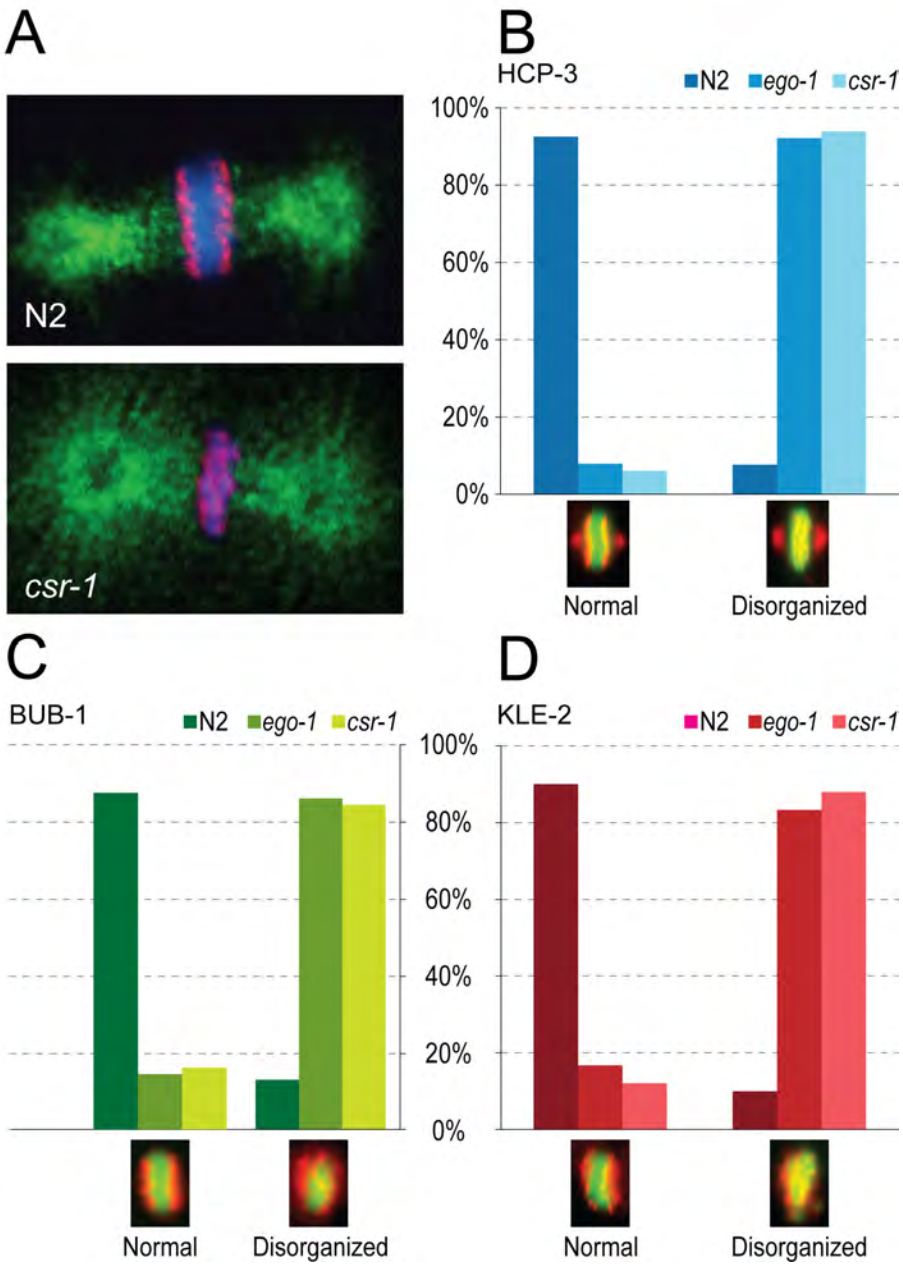
A
N2
csr-1
B
HCP-3
N2
ego-1
csr-1
100%
80%
60%
40%
20%
0%
Normal
Disorganized
C
BUB-1
N2
ego-1
csr-1
100%
80%
60%
40%
20%
0%
Normal
Disorganized
D
KLE-2
N2
ego-1
csr-1
Normal
Disorganized

Figure II-2.
***csr-1*, *ego-1*, *ekl-1*, and *drh-3* RNAi-Depleted Embryos Display Defects in Chromsome Organization**

(A) Single confocal sections showing kinetochore organization in the first cell division of wild-type (N2) and *csr-1* RNAi-depleted embryos (HCP-3, red; tubulin, green; DNA, blue).

(B) HCP-3/inner kinetochore disorganization frequency in wild-type (N2), versus *ego-1* and *csr-1* RNAi-depleted embryos (example metaphase images: HCP-3, red; DNA, green).

(C) BUB-1/outer kinetochore disorganization frequency in wild-type (N2), versus *ego-1* and *csr-1* RNAi-depleted embryos (example metaphase images: BUB-1, red; DNA, green).

(D) KLE-2/condensin disorganization frequency in wild-type (N2), versus *ego-1* and *csr-1* RNAi-depleted embryos (example metaphase images: KLE-2, red; DNA, green).

problems in chromosome condensation or cohesion, we examined the localization of proteins involved in these processes. Both the Condensin I/Condensin IDC protein CAPG-1 and the Condensin II protein KLE-2 (Csankovszki et al., 2009), as well as the cohesins SCC-1 and SCC-3 (Mito et al., 2003), were loaded onto mitotic chromosomes in *csr-1*, *drh-3*, *ekl-1*, and *ego-1* RNAi-depleted embryos but displayed highly disorganized localization patterns, in a manner similar to HCP-3 (Figures II-2D and II-S3).

Expression Studies Reveal Localization to P Granules and Mitotic Chromosomes

To explore the role of these RNAi components in chromosome segregation, we examined the expression and localization patterns of DRH-3, EGO-1, EKL-1, and CSR-1. Western blot analyses revealed that DRH-3, EKL-1, and two isoforms of CSR-1 are present at all developmental stages, and that EGO-1 and CSR-1 are most enriched in young adults, gravid adults, and embryos (Figure II-3A) (Vought et al., 2005). DRH-3 and EKL-1 were detected in *glp-4(bn2)* adults, which fail to develop a germline and are thus greatly enriched in post-mitotic cells (Beanan and Strome, 1992). This finding is consistent with the role of DRH-3 and EKL-1in the biogenesis of a broader set of somatically expressed 22G-RNAs (Gu et al., 2009). The larger CSR-1 isoform was expressed throughout larval development and was also present at low levels in post-mitotic populations lacking a germline. Quantitative real-time RT-PCR analysis of both *csr-1* colocalize in the germline

Figure II-3

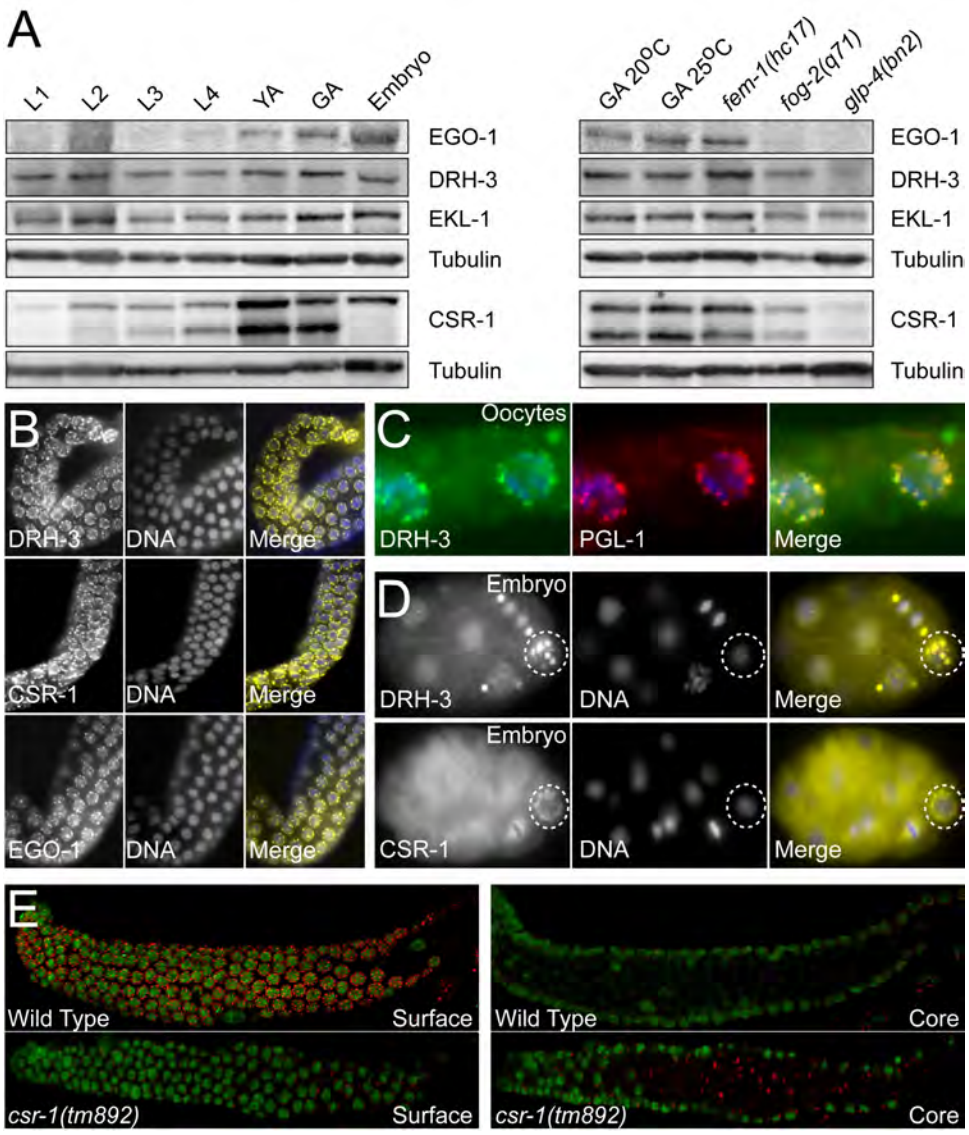
A
L1
L2
L3
L4
YA
GA
Embryo
EGO-1
DRH-3
EKL-1
Tubulin
CSR-1
Tubulin
GA 20°C
GA 25°C
fem-1(hc17)
fog-2(q71)
glp-4(bn2)
EGO-1
DRH-3
EKL-1
Tubulin
CSR-1
Tubulin
B
DRH-3
DNA
Merge
CSR-1
DNA
Merge
EGO-1
DNA
Merge
C
Oocytes
DRH-3
PGL-1
Merge
D
Embryo
DRH-3
DNA
Merge
Embryo
CSR-1
DNA
Merge
E
Wild Type
Surface
csr-1(tm892)
Surface
Wild Type
Core
csr-1(tm892)
Core

Figure II-3. CSR-1, DRH-3, EKL-1, and EGO-1 Are Expressed in the Germline

(A) Western blots of developmentally staged protein lysates (left) or various germline mutant lysates (right) probed for EGO-1, DRH-3, EKL-1, CSR-1 (multiple isoforms), and tubulin (as a loading control). L1, L2, L3, and L4 are larval stages; YA, young adults; GA, gravid adults; Embryos, mixed stage embryos. GA 25°C, gravid adults grown at 25°C; *fem-1(hc17)*, no sperm at 25°C; *fog-2(q71)*, enriched to 95% males by filtration (20°C); and *glp-4(bn2)*, no germline at 25°C.

(B) Wild-type perinuclear germline localization of DRH-3, CSR-1, and EGO-1 (left, yellow) (DNA, center, blue).

(C) DRH-3 (left, green) colocalizes with the P granule component, PGL-1 (center, red; DNA, blue).(D) DRH-3 and CSR-1 (left, yellow) remain localized to P granules in the embryonic P cell lineage (dashed circles; DNA, center, blue).

(E) Single confocal sections of PGL-1 (red) in wild-type and *csr-1(tm892)* mutant germlines through the germline surface and core. P granules become detached from the nuclear periphery in *csr-1(tm892)* (DNA, green; distal is to the left).

Quantitative real-time RT-PCR analysis of both *csr-1* transcripts indicated that their expression recapitulates the protein expression pattern (Figure S4). DRH-3, EGO-1, and CSR-1 colocalize in the germline with PGL-1, a previously characterized component of the germline nuage structures called P granules (Figures II-3B and II-3C) (Kawasaki et al., 1998). EKL-1 was not detected in P granules (data not shown). While many developmentally important factors transiently localize to P granules, DRH-3 and CSR-1 maintained their P granule localization in germ cells throughout the life cycle (Figure II-3D). As was previously shown for *ego-1* mutants (Vought et al., 2005), mutations in *ekl-1*, *csr-1*, and *drh-3* also caused a striking disruption in the perinuclear localization of P granules (Figure II-3E and data not shown), indicating that these factors function more intimately in promoting or maintaining P granule structure and association with the nuclear periphery.

As oocytes matured, EGO-1 was lost from the P granules, while DRH-3 (Figure II-3C) and CSR-1 (Figure II-4A) maintained P granule association. In mature oocytes, CSR-1 (Figure II-4A) and EGO-1 both became enriched in nuclei, where CSR-1 was enriched on the diakinetic chromosomes. In the mitotic cells of embryos, each factor became enriched in prophase nuclei. As chromosomes condensed, DRH-3, EGO-1, and EKL-1 became enriched along the length of each chromosome, while CSR-1 remained nuclear (Figures II-4B–II-4E). All four proteins exhibited robust localization around the metaphase plate (Figures II4F–II-4I). CSR-1 and DRH-3 displayed a pattern similar to cohesins

Figure II-4

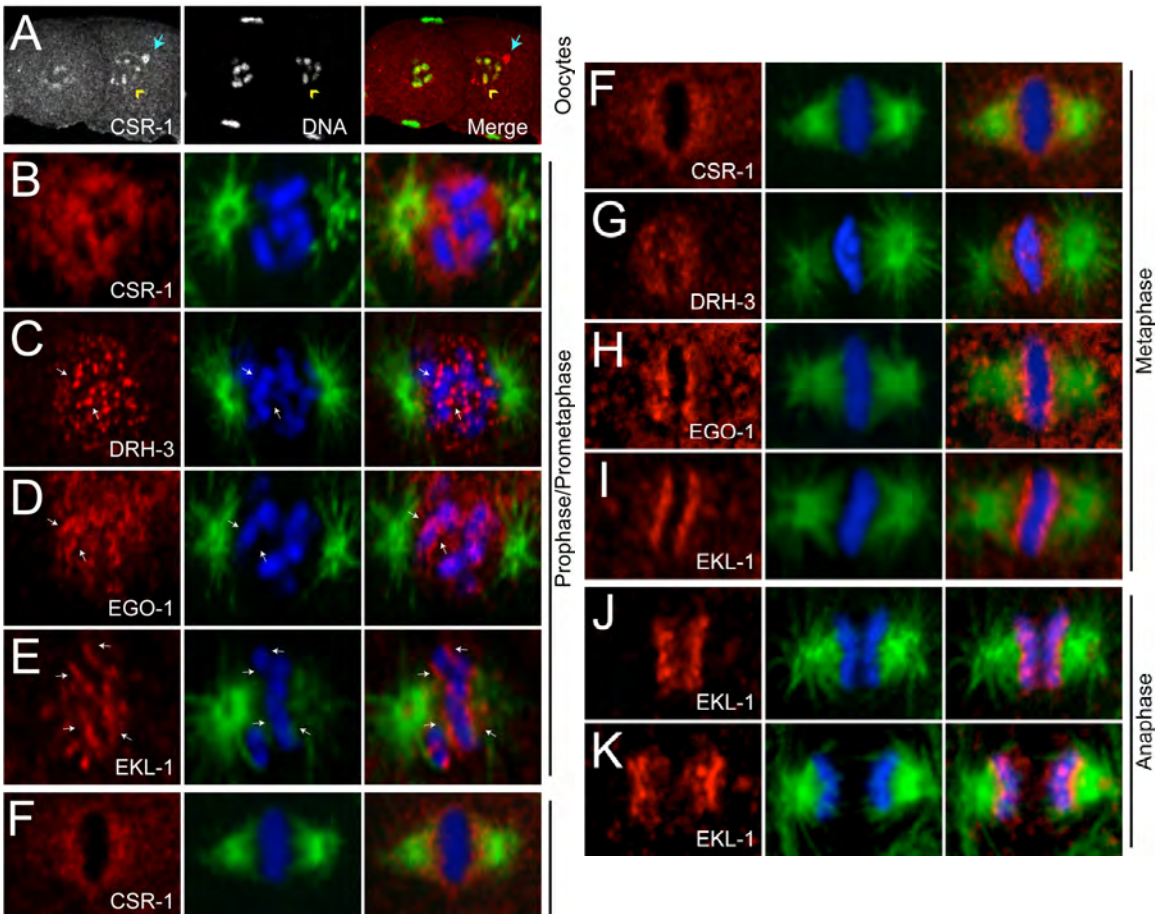
A
CSR-1
DNA
Merge
B
CSR-1
C
DRH-3
D
EGO-1
E
EKL-1
F
CSR-1
Oocytes
Prophase/Prometaphase
F
CSR-1
G
DRH-3
H
EGO-1
I
EKL-1
J
EKL-1
K
EKL-1
Metaphase
Anaphase

Figure II-4. CSR-1, DRH-3, EKL-1, and EGO-1 Localize to Chromosomes

(A) Single confocal sections of CSR-1 (left, red) in wild-type oocytes. CSR-1 is enriched on diakinetic chromosomes as oocytes mature (yellow arrowhead) and remains in some P Granules (blue arrow) (DNA, center, green; distal is to the left).

(B–E) Single confocal sections of CSR-1 (B), DRH-3 (C), EGO-1 (D), and EKL-1 (E) (red) in wild-type embryo prophase/prometaphase (tubulin, green; DNA, blue).

(F–I) Single confocal sections of CSR-1 (F), DRH-3 (G), EGO-1 (H), and EKL-1 (I) (red) in wild-type embryo metaphase (tubulin, green; DNA, blue).

(J and K) Single confocal sections of EKL-1 in wild-type embryo early (J) and late (K) anaphase (tubulin, green; DNA, blue).

(Mito et al., 2003), whereas EKL-1 (and to a lesser degree, EGO-1) appeared to be more closely associated with chromosomes in a pattern similar to kinetochore proteins. In fact, EKL-1 retained a robust association with chromosomes during anaphase, whereas the other RNAi factors became more difficult to detect (Figures II-4J and II-4K). Cytoplasmic localization was also detected for each protein. Finally, all aspects of the localization patterns were absent in each respective mutant background (Figure II-S5).

We then asked whether DRH-3, EGO-1, CSR-1, and EKL-1 depend on each other's wild-type activities for their expression and localization. Consistent with the idea that these factors function together, we found a codependence for proper localization both to metaphase chromosomes and to the P granules. Whereas western blotting demonstrated that the expression of EGO-1, CSR-1, and EKL-1 was undiminished in *drh-3* mutants (Figure II-S5) (Gu et al., 2009), the localization of each protein to chromosomes at metaphase was nearly abolished (Table II-S1), and EGO-1 and CSR-1 lost their association with germline P granules. In *ekl-1* and *ego-1* RNAi-depleted embryos only CSR-1 exhibited greatly reduced association with the metaphase plate and with P granules. Finally, DRH-3, EGO-1, and EKL-1 localized to the disrupted metaphase plates in *csr-1* RNAi-depleted embryos, and DRH-3 and EGO-1 associated with mislocalized P granules in *csr-1* RNAi-depleted germlines. Taken together, these data indicate a hierarchy in the RNAi/chromosome segregation pathway, in which the wildtype activity of DRH-3 was necessary for the proper

targeting of EKL-1, EGO-1, and CSR-1 to chromosomes.

CSR-1 Associates with Small RNAs that Are Antisense to Germline-Expressed Genes

The targets of Argonaute proteins can be deduced by analyzing the sequences of the Argonaute-associated small RNAs. Therefore, we recovered CSR-1 complexes and analyzed the associated small RNAs using a deep-sequencing approach. CSR-1 complexes were enriched 2-fold or greater over the input control for a class of *drh-3*-, *ego-1*-, and *ekl-1*- dependent 22G-RNAs that are antisense to at least 4191 protein-coding genes. These gene-targeted 22G-RNAs collectively represented greater than 99% of all 22G-RNA reads matching loci with a 2-fold or greater increase in read count in the CSR-1 IP complex (Figures II-5A, II-5B, and II-S6 and Table II-S2). MicroRNAs, 21U-RNAs, and nearly all other 22G-RNA species, including those targeting transposons and other repetitive sequences, pseudogenes, and intergenic or nonannotated regions, were depleted in CSR-1 complexes (Figures II-5A and II-5B). The exceptions were 22G-RNAs targeting seven families of repetitive elements and 23 loci annotated as pseudogenes. Altogether, repeat-targeted 22G-RNAs accounted for only 0.25% of the total reads enriched in the CSR-1 IP complex, whereas pseudogene-targeted reads represented less than 0.5% (Table II-S3). 22G-RNAs corresponding to at least 80% of the CSR-1-

Figure II-5

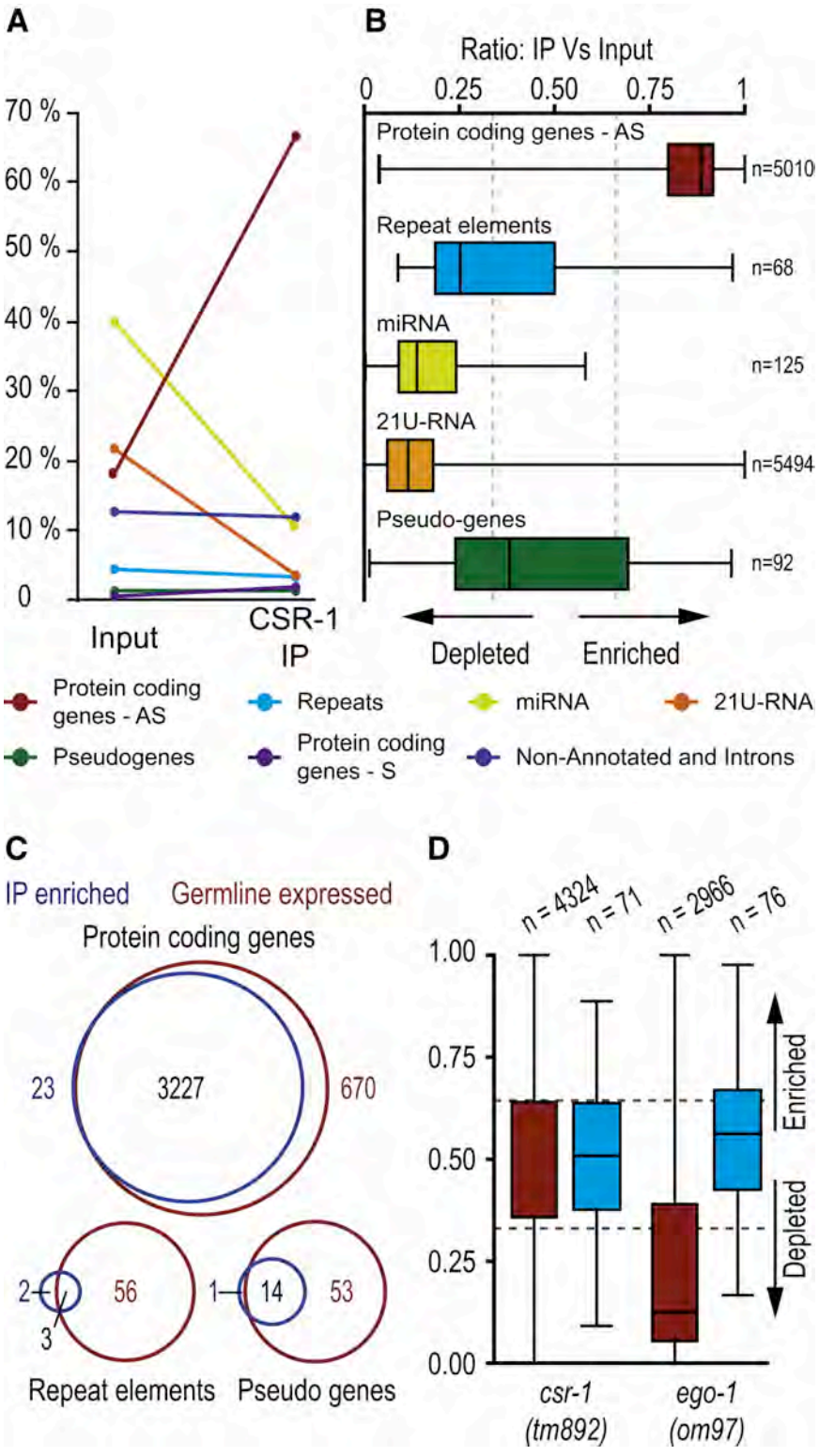
A
70 %
60 %
50 %
40 %
30 %
20 %
10 %
0
Input
CSR-1 IP
B
Ratio: IP Vs Input
0
0.25
0.50
0.75
1
Protein coding genes - AS
n=5010
Repeat elements
n=68
miRNA
n=125
21U-RNA
n=5494
Pseudo-genes
n=92
Depleted
Enriched
Protein coding genes - AS
Repeats
miRNA
21U-RNA
Pseudogenes
Protein coding genes - S
Non-Annotated and Introns
C
IP enriched
Germline expressed
Protein coding genes
23
3227
670
2
56
3
1
14
53
Repeat elements
Pseudo genes
D
n = 4324
n = 71
n = 2966
n = 76
1.00
0.75
0.50
0.25
0.00
Enriched
Depleted
csr-1 (tm892)
ego-1 (om97)

Figure II-5
Analysis of Small RNAs Enriched in CSR-1 IP Complexes

(A) Line plot comparing the relative proportions of small RNA classes between wild-type (N2) input (left) and CSR-1 IP (right) samples (AS = antisense, S = sense).

(B) Box and whisker plot of the relative proportion of small RNA reads for each locus targeted within each small RNA class, in the CSR-1 IP relative to input. Loci with values closer to 1 indicate enrichment of small RNA reads in the IP, a value of 0.5 indicates equal proportions of reads in the IP and input, and values closer to 0 indicate loci depleted of small RNA reads in the IP. Boxes contain 50% of siRNA loci (between the 25th and 75th percentiles), with the line inside each box representing the median value. Lines extending to the right of the box represent the most enriched value, and lines extending to the left of the box represent the most depleted value in the IP. X axis is relative proportion of reads (measured as IP value divided by input plus IP values for any given locus). Dotted lines indicate the values corresponding to 2-fold enrichment (a value of 0.66) or depletion (a value of 0.33). Calculations were made with small RNA cutoffs as described in the Supplemental Experimental Procedures.

(C) Venn diagram depicting the proportion of loci that possess a 2-fold or greater depletion of 22G-RNAs in the *glp-4(bn2)* mutant that are also enriched 2-fold or more in the CSR-1 IP. Only loci present in both datasets with 25 reads per million or more are represented.

(D) Box and whisker plot of the relative proportion of small RNA reads for each locus in the *csr-1(tm892)* and *ego-1(om97)* relative to a congenic wild-type strain (DA1316). Protein-coding genes (red) and repeat elements (blue) are represented. *drh-3* and *ekl-1* small RNA analyses are described in Gu et al. (2009).

targeted mRNAs were strongly depleted in the *glp-4(bn2)* mutant (Figure II-5C) (Gu et al., 2009), which lacks a germline, indicating that the CSR-1 22G-RNAs are expressed in the germline. Finally, consistent with the involvement of the ß-nucleotidyl transferase CDE-1 in the uridylation of CSR-1-associated 22G-RNAs, approximately 40% of the 22G-RNA reads enriched in the CSR-1 IP were extended at the 3´ end with at least one uridine (II-Figure S6) (van Wolfswinkel et al., 2009).

When factors involved in Argonaute-mediated small RNA biogenesis are absent or nonfunctional, the corresponding small RNAs are also depleted (Batista et al., 2008; Grishok et al., 2001; Yigit et al., 2006). Thus, we prepared small RNA libraries from *csr-1(tm892*) and *ego-1(om97)* mutants and compared them to libraries from *drh-3(ne4253)* and *ekl-1(tm1599)* mutant populations (Gu et al., 2009). Consistent with the IP analysis described above, *csr-1* and *ego-1* mutants were depleted for a set of 22G-RNAs that are antisense to protein-coding genes (Figures II-5D and II-S7 and Tables II-S4 and II-S5). To be scored as depleted in the mutants, an arbitrary cut off of 25 reads per million in the wild-type data set was used. As a consequence, many loci for which read counts were significantly increased in the IP studies above were excluded from this analysis. Nevertheless, approximately 900 loci exhibited 22G-RNAs that were dependent on *csr-1*, as well as on *ego-1*, *drh-3*, and *ekl-1*. Consistent with a germline origin for these 22G-RNAs, the majority were depleted in *glp-4(bn2)* animals, which lack a germline. While the proportion of 21U-RNAs was

unaltered in the four mutants, microRNA populations, overall, appeared slightly decreased in *csr-1* and *ego-1*, relative to the total read count, possibly due to a dearth of embryos in these mutant populations (Figure II-S7) (Gu et al., 2009).

As expected, based on their broad role in 22G-RNAbiogenesis, all 22G-RNAs, including those targeting repetitive elements, were depleted in *drh-3* and *ekl-1* samples (Gu et al., 2009). 22G-RNAs targeting repeats (including those targeting the seven repeat families that were enriched in CSR-1 complexes) were unaltered in small RNA populations from the *csr-1* and *ego-1* mutants (Figure II-5D). Furthermore, those 22G-RNAs, which were not associated with, or dependent on CSR-1, were instead dependent on the activity of the *ego-1* paralog, *rrf-1*, or on a combination of *ego-1* and *rrf-1* activities but exhibited no other distinguishing biochemical properties. These remaining CSR-1-independent 22G-RNAs, including those produced by RRF-1, engage a distinct family of Argonautes that mediate transposon silencing and other silencing activities unrelated to chromosome segregation (Gu et al., 2009). These data are consistent with CSR-1 IP data and suggest that *csr-1* and *ego-1* are specifically involved in the expression of a particular subset of gene-targeted 22G-RNAs.

CSR-1 Targets Are Not Misregulated in *csr-1* Mutants

The genes targeted by CSR-1 22G-RNAs include numerous genes whose mRNAs are expressed in the germline, oocytes, and embryos. To determine if CSR-1 regulates its targets at the mRNA level, we performed transcriptional

profiling on *csr-1 (tm892)* mutant versus wild-type (N2) adult worms. Previous work demonstrated that CSR-1 is capable of degrading target mRNAs in vitro (Aoki et al., 2007). However, strikingly, the global profile of gene expression for *csr-1*, including the profile of CSR-1 22G-RNA targets, was very similar to that for wild-type (Figure II-6A and Table II-S6). Thus, CSR-1 does not downregulate its target mRNAs. Similar results were reported for the expression of CSR-1 targets in transcriptional profiling studies performed on *drh-3* (Figure II=S8) (Gu et al., 2009) and *cde-1* mutants (van Wolfswinkel et al., 2009).

With available antibodies for the protein products of several CSR-1 22G-RNA targets, we next used immunofluorescence and western blotting to examine protein expression levels in *csr- 1*, *drh-3*, and *cde-1* mutants. There were no significant changes in the protein levels of the CSR-1 22G-RNA targets we examined, including those of the small RNA pathway components PRG-1 and DCR-1; the P granule factors PGL-1, CAR-1, and CGH-1; the cohesin SCC-3; and the dosage-compensation factors DPY-27 and CAPG-1 (Figures II-6B and II-S8). Together, these data suggest that CSR-1 22G-RNA complexes do not act globally to significantly alter target gene expression.

CSR-1 Is Bound to Chromatin at 22G-RNA Target Loci

In *S. pombe*, the Argonaute Ago1 associates directly with chromatin as a part of the RITS complex (Buhler et al., 2006; Motamedi et al., 2004). A large-scale proteomics study identified CSR-1 associated with fractions of sperm and oocyte

Figure II-6

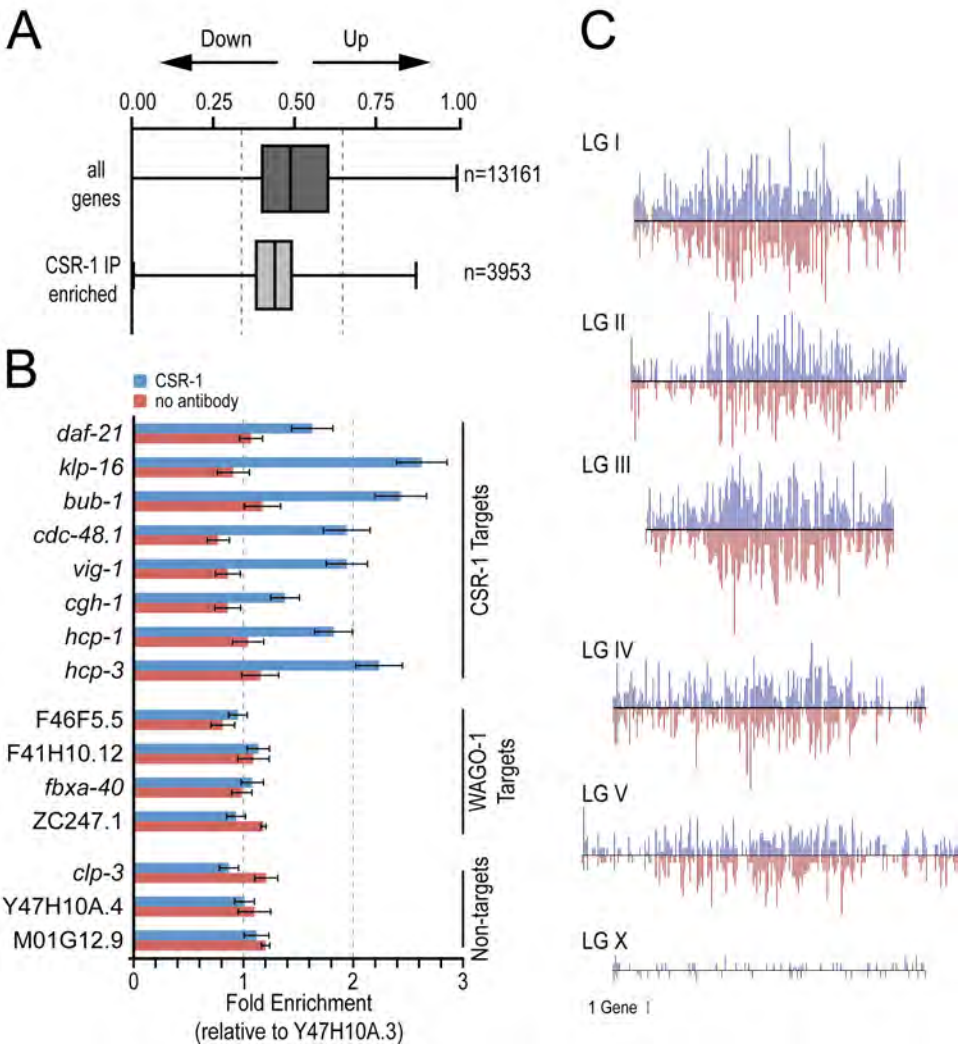
A
Down
Up
0.00
0.25
0.50
0.75
1.00
all genes
n=13161
CSR-1 IP enriched
n=3953
B
CSR-1
no antibody
daf-21
klp-16
bub-1
cdc-48.1
vig-1
cgh-1
hcp-1
hcp-3
F46F5.5
F41H10.12
fbxa-40
ZC247.1
clp-3
Y47H10A.4
M01G12.9
CSR-1 Targets
WAGO-1 Targets
Non-targets
0
1
2
3
Fold Enrichment
(relative to Y47H10A.3)
C
LG I
LG II
LG III
LG IV
LG V
LG X
1 Gene

Figure II-6. CSR-1 22G-RNA Complexes Bind to Target Genomic Loci

(A) Box and whisker plot of mRNA expression from microarray experiments in wild-type versus *csr-1(tm892)* mutants. The analysis was done for all genes measured by the array (left) and the subset of only CSR-1 22G-RNA target genes (right).

(B) Western blot analysis of wild-type and *csr-1(tm892)* protein lysates, probed for CSR-1 22G-RNA target proteins. EKL-1 is a loading control.

(C) ChIP/quantitative real-time PCR analysis of CSR-1 enrichment at CSR-1 22G-RNA or WAGO-1 22G-RNA target loci. Fold enrichment is calculated relative to the Y47H10A.3 locus, which, like *clp-3,* Y47H10A.4, and M01G12.9, is not targeted by small RNAs. Data from a single, representative set of experiments are presented; error bars are the standard deviation from the mean of three replicates of a single ChIP sample. (IP with CSR-1, blue; IP with beads only/no antibody, red.)

(D) Density of CSR-1 22G-RNA target genes on each chromosome. Each bar represents the numbers of genes in a 100 kb bin. (Watson strand, blue; Crick strand, red.) Chromosome number is as indicated. Scale bar represents one gene.

chromatin (Chu et al., 2006). Using a similar method (Chu et al., 2006), we have determined that CSR-1 associates with chromatin in embryos (Figure II-S9). These observations led us to examine whether CSR-1 complexes directly bind to the genomic loci of the CSR-1 22G-RNA targets.

Using chromatin immunoprecipitation (ChIP), we found an enrichment of CSR-1 at target loci when compared to several other genomic loci that are not targeted by small RNAs. RNA polymerase II was used as a positive control and consistently showed enrichment at many CSR-1 target loci (Figure II-S9). In contrast, negative control experiments using agarose beads alone (without CSR-1 antibody) never displayed enrichment (Figures II-6C and II-S9). Of the 12 CSR-1 22G-RNA target loci examined, 10 showed 1.5-fold or greater enrichment of CSR-1 binding in five or more independent experiments (Figure II-6C). Conversely, CSR-1 was never enriched at the targets of another germline-expressed Argonaute, WAGO-1 (Figure II-6C). CSR-1 was not detected in chromatin fractions treated with RNase A (data not shown), nor did we detect CSR-1 enrichment by ChIP at target loci in the *drh-3(ne4253*) mutant, in which 22G-RNAs are depleted (Figure II-S9). These findings indicate that CSR-1 interacts with its target genomic loci in a 22G-RNA-dependent manner. Furthermore, CSR-1 22G-RNA target loci are distributed relatively uniformly along the chromosomes (Figure II-6D), suggesting that the CSR-1 22G-RNA pathway could act in a genome wide manner to influence chromosome segregation

DISCUSSION

Here we have investigated the role of the *C. elegans* Argonaute CSR-1 in promoting proper chromosome segregation. We have shown that CSR-1 interacts with a class of 22-nucleotide RNAs, called 22G-RNAs, which are antisense to at least 4191 protein-coding genes, seven repeat element families, and 23 pseudogenes distributed throughout the genome. A parallel study by Gu et al. (Gu et al., 2009) has shown that a distinct Argonaute, WAGO-1, interacts with a nonoverlapping set of 22G-RNAs that primarily target transposons, cryptic elements, and pseudogenes (see below). The biogenesis of both CSR-1- and WAGO-1- bound 22G-RNAs is dependent on a core set of factors, including DRH-3, EKL-1, an RdRP, and a ß-nucleotidyl transferase. However, WAGO-1-associated 22G-RNAs appear to down regulate their mRNA targets, whereas CSR-1 22G-RNAs do not. Whole-genome microarray studies showed that the mRNA targets of CSR-1 22G-RNAs are not misregulated in the *csr-1*, *drh-3*, and *cde-1* mutant backgrounds (Gu et al., 2009; van Wolfswinkel et al., 2009) (Figures II-6 and II-S8). In addition, immunoflourescence and western blot analysis on the protein products of several CSR-1 targets revealed no change in expression (Figures II-6 and II-S8). Based on these findings, it seems unlikely that perturbed expression of CSR-1 22G-RNA targets results in the observed chromosome segregation defects. Instead, our findings support a model in which the CSR-1 pathway may directly contribute to holocentric chromosome organization by ensuring that the expressed, euchromatic domains within the

genome support the proper juxtaposition and alignment of the kinetochores, which must span these domains (II-Figure 7).

How Does CSR-1 Influence Chromosome Segregation?

Several lines of evidence, including ChIP, chromatin isolation, and immunolocalization studies, indicate that CSR-1 pathway components associate directly with chromatin in an RNA-dependent manner. These data support a direct role for CSR-1 22G-RNA complexes in promoting chromosome segregation, perhaps through a mechanism that is similar to the Ago1 pathway that regulates centromere formation in *S. pombe*. Indeed, both the CSR-1 and the Ago1 pathways utilize similar components for small RNA biogenesis. These include a helicase, an RdRP, and a ß-nucleotidyl transferase. However, these pathways target dramatically different loci: the Ago1 system targets repetitive, pericentromeric heterochromatin, whereas the CSR-1 pathway overwhelmingly targets protein-coding euchromatic domains distributed throughout the genome. Despite this difference, perhaps the small RNAs produced in both systems perform analogous functions. The targeting of CSR-1 22G-RNA complexes to chromosomal loci in the germline could recruit chromatin modifiers that mark CSR-1 22G-RNA-targeted domains and provide boundaries that define the adjacent centromeric domains of HCP-3 incorporation. Consistent with this notion, a preliminary comparison indicates that the domains targeted by CSR-1 22G-RNAs are, in large part, mutually excluded from regions that are enriched

Figure II-7

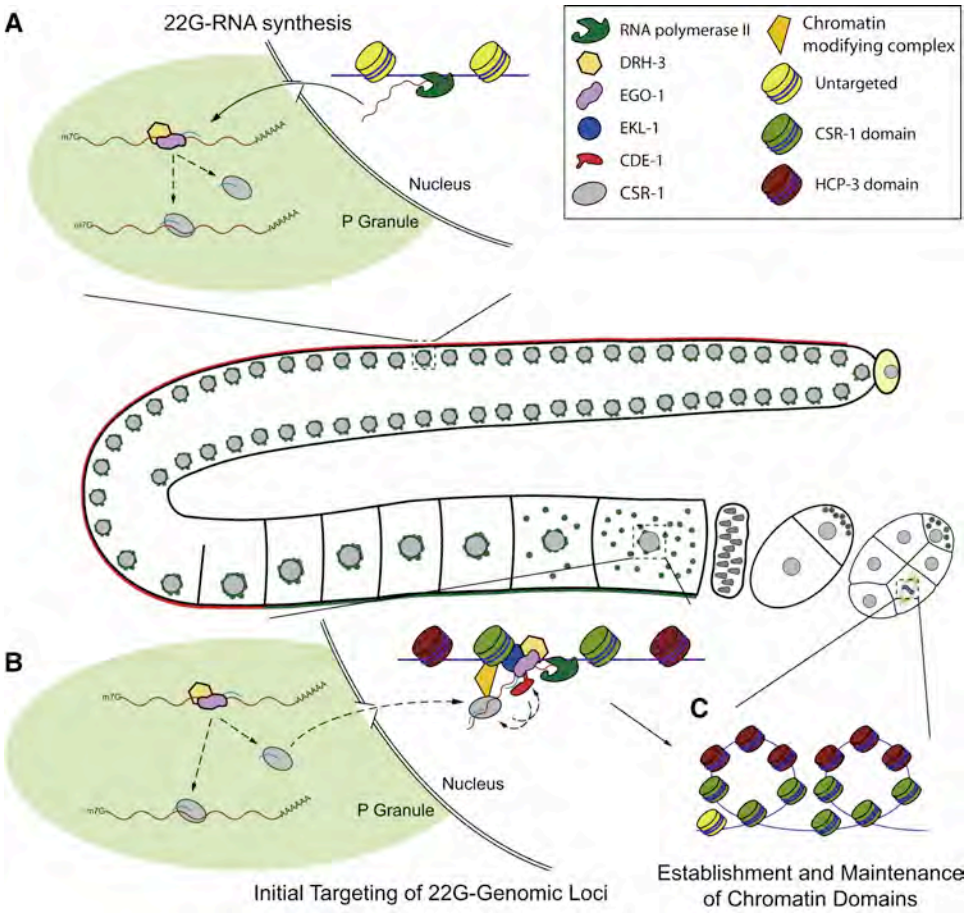
A
22G-RNA synthesis
Nucleus
P Granule
RNA polymerase II
DRH-3
EGO-1
EKL-1
CDE-1
CSR-1
Chromatin modifying complex
Untargeted
CSR-1 domain
HCP-3 domain
B
Nucleus
P Granule
Initial Targeting of 22G-Genomic Loci
C
Establishment and Maintenance of Chromatin Domains

Figure II-7. Model for the Activity of the CSR-1 22G-RNA Pathway in Chromosome Segregation

(A) 22G-RNA synthesis: In the germline, DRH-3, EGO-1, and CSR-1 localize to perinuclear P granules, where DRH-3 and EGO-1 initiate the synthesis of 22G-RNAs from transcripts that are important for germline development and early embryogenesis. These 22G-RNAs are loaded onto CSR-1 and can guide the complex to its targets.

(B) Initial targeting of genomic loci: In oocytes, CSR-1 22G-RNA complexes move into the nucleus where they target nascent transcripts, possibly by cleaving them. Chromatin-modifying factors may associate with CSR-1 complexes to promote local modification of histones at and near CSR-1 target loci, establishing pericentromeric chromatin domains (green nucleosomes). A complex containing EGO-1, DRH-3, and possibly EKL-1 is proposed to amplify the signal in a positive feedback loop, by generating more 22G-RNAs in the nucleus with the CSR-1 22G-RNA-targeted nascent transcripts as the template.

(C) Establishment and maintenance of chromatin domains: The CSR-1 22G-RNA-dependent chromatin domains containing modified histones (green nucleosomes) may promote the proper binding and organization of other components such as condensins and cohesins in embryo mitotic divisions. Furthermore, these chromatin domains could help to both recruit and restrict the incorporation of the centromeric histone H3 variant HPC-3/CENP-A (red nucleosomes) in chromatin domains adjacent to those targeted by CSR-1 22G-RNA complexes. Regions of the chromatin loop out and self-associate, permitting the assembly of a proper planar, rigid kinetochore on the poleward faces of condensed chromosomes. As cell divisions continue, chromatin domains could be maintained epigenetically, possibly even by EKL-1, thus becoming less reliant on CSR-1 22G-RNA activity throughout development.

for the conserved, centromeric histone variant HCP-3/CENP-A (R. Gassmann and A. Desai, personal communication). Thus, like the *S. pombe* Ago1 system, the CSR-1 pathway may help to define adjacent domains of HCP-3/CENP-A incorporation but does so by targeting protein-coding genes rather than repetitive heterochromatin.

CSR-1 22G-RNA targets are distributed relatively uniformly on each chromosome, as would be expected if these targets serve in the positioning or alignment of kinetochores along the length of each chromosome. The one notable exception is the X chromosome, which is depleted of genes expressed in the germline (Reinke et al., 2000), and which possesses fewer CSR-1 targets than the autosomes (~70 versus 500–900 per autosome). It is not clear how this lower number of CSR-1 target sites might impact X chromosome segregation. The X chromosome is the only chromosome whose loss is tolerated by the organism (resulting in spontaneous males within hermaphrodite populations). Indeed, the stability of the X chromosome is more than an order of magnitude lower than that of the autosomes (the loss of which is generally not detected in wild-type populations) (Meneely et al., 2002). Whatever the explanation for the reduced fidelity of X chromosome segregation, clearly the limited number of CSR-1 targets is sufficient, or there are other pathways governing segregation of the X chromosome.

P Granules and 22G-RNA Biogenesis

CSR-1 and the other protein components of the 22G-RNA pathway localize to P granules. P granules are found in close apposition on the cytoplasmic face of nuclear pores (Pitt et al., 2000) and are thought to be sites of accumulation for many mature polyadenylated mRNAs (Schisa et al., 2001). The nuclear association of P granules is lost in *csr-1*, *ego-1, ekl-1*, and *drh-3* mutant backgrounds, suggesting that the association of CSR-1 and its cofactors with mRNA targets emerging from the nuclear pore may help to drive the perinuclear association of P granules (Figure II-7A). Perhaps consistent with this idea, P granules also lose their perinuclear association in transcriptionally quiescent or nearly quiescent germ cells, e.g., oocytes and early embryo germ cells.

If their initial biosynthesis occurs in P granules, 22G-RNAs may subsequently guide CSR-1 back to chromatin or to chromatin associated nascent transcripts (Figure II-7B). Because CSR-1 targets are robustly expressed in the maternal germline, it is possible that CSR-1 complexes initially engage nascent transcripts during gametogenesis. Once established, these hypothetical CSR-1-chromatin domains could be preserved throughout embryogenesis, perhaps even in the absence of additional transcription (Figure II-7C). Consistent with this idea, we found that CSR-1-chromatin localization was most prominent in the two or three most mature oocytes in each gonad arm (Figure II-4A). The retention of CSR-1 complexes at target loci could occur through direct binding to other chromatin components, possibly even through EKL-1, as tudor domains have

been shown to interact with the methyl-arginine and -lysine moieties of histone tails (Taverna et al., 2007).

Distinct Roles for Argonautes in RNAi and 22G-RNA Pathways

Our studies indicate that at least two distinct germline 22G-RNA pathways with several overlapping core components exist in *C. elegans*: the CSR-1 and WAGO-1 pathways. Like CSR-1, WAGO-1 prominently localizes to P granules. However, the perinuclear distribution of P granules and chromosome segregation are not altered by the loss of *wago-1*, even within the context of a 12-fold WAGO mutant (composed of null alleles of *wago-1* and 11 related WAGO Argonautes) (Gu et al., 2009). How are these Argonautes loaded with distinct 22G-RNA species, despite their shared localization and reliance on upstream factors? One attractive scenario is that mRNA targets are sorted into distinct P granule subcompartments, wherein the amplification of 22GRNAs takes place. Additional protein factors, such as CDE-1, and/or structural elements within target transcripts may be involved in the recognition and compartmentalization of target mRNAs (Gu et al., 2009).

Recombinant CSR-1 protein has been shown to exhibit Slicer activity in vitro (Aoki et al., 2007), and CSR-1 has been implicated in down regulating genes in response to foreign dsRNA (Yigit et al., 2006). However, endogenous CSR-1 22G-RNA targets do not appear to be down regulated by CSR-1. CSR-1 22G-RNAs are expressed at low levels relative to WAGO-1 22G-RNAs (Figure II-

S10), perhaps below a threshold to trigger mRNA turnover. Consistent with this idea, not all WAGO-1 22G-RNA targets exhibit mRNA silencing, but those that do typically exhibit the highest levels of corresponding 22G-RNA accumulation (Gu et al., 2009).

It is tempting to speculate that the incompletely penetrant effects of *csr-1* mutants on RNAi are indirect, perhaps arising as a consequence of the dramatic disruption of P granules in *csr-1* mutants. There are already two distinct Argonaute systems implicated in the RNAi pathway, RDE-1 (Tabara et al., 1999) and the WAGO system (Yigit et al., 2006), and at least the WAGO-1 protein is localized to P granules (Gu et al., 2009). In *csr-1* mutants, perhaps the dissociation of P granules from germ nuclei disrupts access to target mRNAs or other activities required for the full activity of the germline RNAi response.

Our findings together with those of Gu et al. (Gu et al., 2009) indicate that the majority of the genome is targeted by Argonaute systems that provide diverse surveillance functions. Expressed genes are targeted by CSR-1, while classically heterochromatic domains including transposons and pseudogenes are targeted by WAGO-1. Both of these systems contribute to the physical maintenance of the genome by promoting, respectively, (1) chrochromosome segregation and (2) the suppression of mobile or otherwise potentially deleterious elements.

Correlates of these pathways are likely to function in other nematodes and, indeed, could help explain the classic observations in *Parascaris* made by Theodor and Marcella Boveri more than 100 years ago (Pimpinelli and Goday,

1989; Satzinger, 2008). By targeting heterochromatic domains, a system analogous to the WAGO-1 pathway could promote chromosome fragmentation and the elimination of the heterochromatin in *Parascaris*. This could occur via an Argonaute pathway similar to that which mediates chromosome fragmentation during macronuclear formation in *Tetrahymena* (Yao and Chao, 2005). By targeting genes, correlates of the CSR-1 22G-RNA system could ensure the proper higher-order assembly of the holocentric kinetochores found in diverse nematode species and could provide this function even after fragmentation and the elimination of heterochromatin as in the tiny somatic chromosomes of *Parascaris*. Additional insights into the underlying molecular mechanisms through which Argonaute systems promote the higher-order structure of chromosomes will require further study. The observation that such pathways, however different, exist in nematodes and fungi suggests that similar activities are likely to be ubiquitous in eukaryotes.

MATERIALS AND METHODS

Worm Strains

Worms were cultured according to (Brenner, 1974). Alleles used in this study are as follows, in order of chromosome: LGI: *ekl-1(tm1599), ego-1(om97), drh-3(tm 1217), glp-4(bn2), avr-14(ad1302), hT2[qIs48](I; III)*; LGIII: *cid-1/cde-1(tm1021); qC1[neIs(myo-2::avr-15, rol-6(su1006), unc-22(RNAi))]*, LGIV: *fem-1(hc17), csr-1(tm892), DnT1[unc(n754dm) let](IV;V)*; LGV: *fog-2(q71), avr-15(ad1051), glc-1(pk54)*. Strains: **DA1316**: *avr-14(ad1302)LGI; avr-15(ad1051)LGV, glc-1(pk54)LGV*; **AZ212*:*** *(unc-119(ed3) ruIs32[unc-119(+) pie-1::GFP::H2B] III)*; **XA3501:** *(unc-119(ed3) ruIs32[unc-119(+) pie-1::GFP::H2B] III*; *ojIs1[unc-119(+) pie-1::GFP::tbb-2])*; **WM193:** *csr-1(tm892)LGIV; neIs19[pie-1::3xflag::csr-1, unc-119(+)]*; **WM194:** *csr-1(tm892)LGIV; neIs19[pie-1::gfp::csr-1, unc-119(+)]*.

Brood Size, Viability, and *him* quantitations were performed as described in (Batista et al., 2008).

Creation of *csr-1* Transgenes

GFP or 3x Flag *csr-1* rescuing transgenes were first constructed in a yeast artificial chromosome (YAC), as described in (Rocheleau et al., 1999). The *csr-1* genomic locus was PCR amplified from these YACs and cloned into pDONR201 (Invitrogen), then transferred to pID2.02 (D'Agostino et al., 2006) using the Gateway cloning system (Invitrogen). The resulting plasmids were introduced into *unc-119(ed3)* strain using biolistic transformation according to (Praitis et al.,

2001). Transgenic strains were identified and integrated lines were crossed into the *csr-1(tm892)* background. PCR was used to identify rescued *csr-1(tm892)* animals. Primer sequences are available upon request. Rescue levels were similar for all GFP and 3x Flag transgenic strains tested.

RNAi

1 mg/ml dsRNA targeting *drh-3*, *csr-1*, or *ekl-1* was injected into young adult Bristol N2 worms. After 36–48 hr at 20°C, worms containing embryos were dissected and fixed for immunostaining.

Antibody Generation

A rabbit antibody, used in immunostaining, was generated against the CSR-1 polypeptide from amino acids E462 to E987 (containing the PAZ and most of the PIWI domain) (Capralogics, Inc.). Additional rabbit antibodies, used in IP experiments, were generated and purified by Anaspec using the following peptides: VDYNAPKDPEFRQKYPNLKFP and QRCKDKGMHIGSYSMDQHN GERGSENFL. A GST-fusion protein containing an EKL-1 N-terminal fragment (L58 to S309) was used to generate rabbit antisera. DRH-3 and EGO-1 antibodies are described in (Gu et al., 2009).

Immunostaining and Microscopy

Gonads and embryos were excised from worms in 1x sperm salts/2mM levamisole on poly-L-lysine coated slides, frozen and cracked on dry ice for greater than 10 minutes, and fixed at –20 ^{o}C for 5 minutes each (15 minutes total) in each of the following, respectively: 100% Methanol, 50% Methanol/50% Acetone, and 100% Acetone. All sample incubations were performed in a humid chamber. Samples were blocked for one hour in 1xPBS/0.1% Tween-20/3%BSA (PBST+BSA) at room temperature, and then incubated with primary antibody overnight at 4^{o}C. Slides were washed 3 times 10 minutes with PBST, and then incubated for 15 minutes in PBST+BSA. Secondary antibodies were from Jackson Immunoresearch and Molecular Probes. Incubation with secondary antibodies was performed for one hour in PBST+BSA at room temperature. Slides were washed 3 times ten minutes in PBS, and then mounted in Vectashield with DAPI (Vector Labs). All images were collected using a Hamamatsu Orca-ER digital camera mounted on a Zeiss Axioplan 2 microscope and with Openlab software, unless noted.

In time-lapse microscopy, embryos from strain AZ212 and/or XA3501 were dissected from gravid adults in M9 and placed on 2% agarose pads for imaging. 10 Z sections of 2µm thick were collected every 5 or 10 seconds using the Perkin Elmer Ultraview RS spinning disc confocal microscope system mounted on a Zeiss Axiovert 200M microscope. Stacks of frames from every time point were overlaid and compiled into time-lapse movies using the Ultraview

software (Perkin Elmer). Images in Figure 2A were acquired using the Leica TCS SP2 confocal microscope system and software. Z sections of 1μm thick were acquired simultaneously from embryos stained with anti-HCP-3 (Oegema et al., 2001), anti-alpha- tubulin (Accurate Chemical and Scientific corp, clone YOL1/34) and DAPI. Images in Figure 3E and 4A were acquired using the Zeiss LSM 700 point scanning confocal microscope attached to a Zeiss Axio Observer Z1 stand, with Zeiss software. Z sections of 0.8μm thick were acquired simultaneously from germlines stained with anti-PGL-1 (Kawasaki et al., 1998) and DAPI. All Images in Figure 4 (except 4A) were acquired using Solamere Technology Group CSU10B Spinning Disk Confocal System scan head mounted on a Nikon TE-2000E2 inverted microscope with a 100x Plan-APOCROMAT NA1.4 Oil lens and a Roper Coolsnap HQ2 camera. Metamorph software was used to analyze the images. Z sections ranging from 0.1 to 0.3μm were collected from embryos. Quantitation of kinetochore disorganization was performed on at least 25-50 metaphase plates per genotype. Metaphases were counted as being either normal or disorganized (twisted) only. Mitotic cells were identified by their tubulin staining.

FISH

Embryos were dissected in egg salts with 0.1% Tween-20, followed by brief 2% formaldehyde fixation, permeabilization by freeze crack, and fixation for 1 min in -20°C methanol. Slides were washed in PBST and gradually transferred to 100%

ethanol. Slides were dried and incubated in 2x SSC/50% formamide at 37°C for 1 hr. The probe was sealed on the slide, DNA was denatured at 95°C for 3 min, and hybridization was performed overnight at 37ºC. Slides were washed in 2x SSC/50% formamide, 2x SSC, 1X SSC, and PBST, then counter stained with DAPI.

Western Blot Analysis- Proteins were resolved by SDS-PAGE on Criterion Precast gradient gels (4%–15%, Biorad) and transferred to Hybond-C membrane (Amersham Biosciences). The membrane was incubated overnight at 4°C with anti-CSR-1 either: (i) affinity purified antibody, 1 µg/ml, (ii) Full-Length A.v. Polyclonal Antibody (BD Bioscience), diluted 1:1000, or (iii) anti alpha-tubulin (Accurate Chemical) diluted 1:2000, in PBST-5%milk solution (137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4, and 5% [w/v] dried milk). The membrane was incubated 1 h at room temperature with HRP-conjugated secondary antibody (Jackson Immunoresearch) diluted 1:5,000 in PBST and then visualized by Western Lightening ECL Kit from Perkin Elmer. Images were collected on a LAS-3000 Intelligent Dark-Box (Fujifilm).

Small RNA Cloning and Data Analysis

Cloning and data analysis are as described in Batista et al. (Batista et al., 2008) for Terminator exonuclease (Epicenter Technologies) treated samples.

Analysis of deep sequencing data is as described in (Gu et al., 2009). In addition, for reads that did not match the *C. elegans* genome, the last 3´ thymine(s) was/were removed until a different nucleotide was the last 3´ end nucleotide (reads ending with a different nucleotide were not considered for analysis). Sequences that were at least 17 nucleotides long were blasted against the *C. elegans* genome, and included for analysis. Reads in the sense orientation for ribosomal and tRNA genes are considered to be degradation products and were not included in the analysis. For each deep-sequencing library described, basic statistics are listed as follows: total numbers of sequencing reads, number of genome-matching reads (perfect match), number of total genome matching reads after removal of U nucleotides at the 3´ end of non genome matching reads and number of reads after removal of reads considered to be degradation products. CSR-1 IP: 3864681; 1347558; 1848228; 1799685. Wild-type Input: 5003742; 3124393; 3237929; 3193213. DA1316 (AVR triple): 5903016; 3295762; 3377578; 2760559. *csr-1(tm892)*: 5139346; 3306716; 3367248; 2182697. *ego-1(om97)*: 5080570; 2773473; 2799340; 1561121.

Small RNA reads matching unique loci (generally, those sequences targeting protein coding genes, pseudogenes, microRNAs, 21U-RNAs, and non-annotated loci or introns) were first normalized to the number of times they matched the genome. To compare unique loci between different libraries, the number of reads for each locus was normalized to the total number of reads in the library (excluding those reads considered to be degradation products). A

cutoff of 25 reads per million (for protein coding genes, pseudogenes, and non-annotated loci or introns) or 5 reads per million (for microRNAs, 21U-RNAs) was used to perform each analysis. In the comparison between CSR-1 IP and input or *glp-4(bn2)* versus wild-type, for each locus examined, at least one of the samples was required to have 25 reads per million for that locus to be included in the analysis.

The analysis of transposable and repetitive elements (including simple repeats) is complicated by the fact that these elements generally map to many loci throughout the genome, with various degrees of sequence divergence. In addition, the number of these loci throughout the genome varies among different genetic backgrounds. Thus, in the analysis of repeat elements (not including simple repeats), we considered the reads that match each reference element sequence in Repbase (Jurka et al., 2005), without normalizing each read to the number of times it matches the genome. Instead a simple cutoff of 25 reads per million for the analysis of this particular class of small RNA was used.

For comparisons between loci targeted by 22G-RNAs enriched in the CSR-1 IP and loci targeted by small RNAs expressed in the germline it was necessary to compare two data sets, the CSR-1 IP data set (CSR-1 IP library compared to wild-type input library), and the *glp-4(bn2)*, germline depleted, data set (*glp-4(bn2)* library compared to wild-type library). Only genes, pseudogenes or repeat elements present in both the IP and *glp-4(bn2)* datasets (loci above the cutoff value) were included. Using these criteria, defines a set of 944 protein

coding genes, 3 repeat elements and 15 pseudo-genes [present at 25 reads per million in the CSR-1 IP and the input but absent in *glp-4(bn2)*]. 22G-RNAs targeting another 615 protein coding genes, 13 repeat elements and 53 pseudo-genes were enriched in the CSR-1 IP datasets but were not present or did not meet the cutoff of 25 reads per million in the input dataset.

Tiling Microarray Procedures

Synchronous populations of wild type and *csr-1(tm892)* animals were grown for 54 hours post-hatching at 20ºC on OP-50 *E. coli* at a density of approximately 50,000 animals per 15cm Petri dish. The worms were harvested as young adults without oocytes. RNA extraction was performed using TRI-Reagent (MRC Laboratories). Instead of pelleting and resuspending the RNA (as described in the TRI Reagent protocol), RNA was recovered, washed and eluted using the RiboPure total RNA isolation kit (Ambion). Reverse transcritption was performed on 7µg of each sample using the GeneChip WT Double-Stranded cDNA Synthesis Kit. The dsDNA was then purified using the GeneChip Sample Cleanup Module (Affymetrix) and quantified. 7.5µg of each dsDNA sample were used for the subsequent fragmentation and labeling reactions, using the GeneChip WT Double Stranded DNA Terminal Labeling Kit (Affymetrix). Hybridization to the arrays was made using standard Affymetrix protocols and reagents, and scanning was done with GeneChip Scanner 3000 7G at the UMass Medical School's Genomics Core Facility. Experiments were all

conducted in triplicate from independent plates and sample preparations. The arrays used in this study are the GeneChip *C. elegans* Tiling 1.0R chips from Affymetrix (comprising over 3.2 million perfect match/mismatch 25 nt probes covering the complete non-repetitive *C. elegans* genome with a 25 base pair resolution). Detailed protocols can be found in the GeneChip Whole Transcript (WT) Double-Stranded Target Assay Manual from Affymetrix. Signal values for each array probe were calculated using Affymetrix Tiling Analysis Software 1.1.2 (bandwidth: 30; intensities: PM/MM) with three *csr-1(tm892)* replicates as the experimental datasets and three wt replicates as the controls. Probe overlap with annotations was assessed using the Affymerix-provided ce4 coordinate, which indicates the genomic position matching the center of the array probe. Only genes with signal for at least 10 different probes in either the wild type or *csr-1(tm892)* samples were included for analysis.

Sub-cellular Fractionation/Chromatin Isolation

Sub-cellular fractionation and chromatin isolation is as described (Chu et al., 2006), with the following modifications: 500ml of early embryos were resuspended in 2 volumes of Buffer A (as described, with the addition of 1% SUPERNaseIN (Ambion), dounced in a Wheaton metal Dounce homogenizer only until nuclei were visible under the dissecting microscope, then the extracts were centrifuged at 1500xg for 1 min. at 4°C. The rest of the protocol was as

described until the isolated chromatin was resuspended in 50ml 2X SDS-PAGE Sample Buffer. Western blot analysis is described above.

For RNAse A treatment of chromatin fractions, RNAse A (50µg per ml lysate) was added prior to dounce homogenization. After douncing, extracts were incubated for 30 minutes at room temperature to allow for RNAse A to act. In untreated control experiments, this room temperature incubation was also added to the procedure described above.

Chromatin Immunoprecipitation

Embryos were prepared by using 20% hypochlorite and extracted from gravid N2 adults grown for 58-60 hours at 20°C. Live embryos were washed five times with M9 buffer, and then (when indicated, Figure 6) treated with 10mM dimethyl 3,3´-dithiobispropionimidate (DTBP, Thermo Fisher Scientific) diluted in M9 buffer (50ml total volume), for 30 minutes at room temperature with rotating. DTBP was quenched by the addition of 2.5ml of 2.5M glycine for 5 minutes at room temperature. Embryos were washed once with M9 before proceeding to formaldehyde cross-linking.

Embryos were cross-linked using 2.6% formaldehyde for 30 minutes at room temperature (50ml total volume) followed by quenching with 2.5ml 2.5M glycine for 5 minutes at room temperature. Embryos were then washed three times with M9 buffer, once with FA buffer (50mM HEPES/KOH pH 7.5, 1mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate; 150mM NaCl), and frozen

in 500ml aliquots at −80°C. Extracts were prepared by resuspending embryo pellets in 1 volume FA Buffer supplemented with protease and phosphatase inhibitors, followed by dounce homogenization and sonication (15x, 15sec., 15% output, with a power setting of 5 on a Fisher Sonic Dismembrator 550) in a volume of 2ml. Protein concentration was determined by Lowry method and 3.3mg extract was used for each ChIP in a total volume of 500 µl. 10% of each IP was removed as input (50µl), and frozen until the next day.

10µg (anti-RNA Pol II, Abcam, #5408) or 20µg (anti-CSR-1) of antibody or buffer alone (no antibody control) was added to each IP sample and incubated overnight at 4°C. Immune complexes were recovered using 50µl of a 50% slurry of protein-A/G agarose beads (Santa Cruz Biotechnology) and washed at room temperature with 1ml of each of the following solutions: FABuffer (2x 5 minutes), FA Buffer with 1M NaCl (1x 5 minutes), FA Buffer with 500mM NaCl (1x 10 minutes), TEL (0.25M LiCl, 1% NP-40, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris-HCl, pH 8.0) (1x 10 minutes), and TE (1mM EDTA, 10mM Tris-HCl, pH 8.0) (2x 5 minutes). Samples were eluted twice with 150µl elution buffer (1% SDS in TE with 250mM NaCl) for 15 minutes at 65°C with shaking. Eluates were combined and treated with 1µl (20mg/ml) Proteinase K for 2 hours at 55°C with shaking.

Input samples were thawed and treated with 10µg of RNAse A (Ambion) for 2 hours at room temperature, before adding 150µl of elution buffer and treating with 1µl Proteinase K for 2 hours at 55°C with shaking. Then, crosslinks

were reversed for all samples by incubation overnight at 65°C with shaking. DNA was recovered by phenol chloroform extraction and ethanol precipitation. All samples were resuspended in 50µl of ultrapure water and stored at -20°C. ChIP samples were analyzed by quantitative real-time PCR.

Quantitative Real-time PCR

Is as described in (Batista et al., 2008). For mRNA analysis: cDNA was generated from 1mg *C. elegans* total RNA using random hexamers with Superscript III Reverse Transcriptase (Invitrogen). qRT-PCR was performed on the ABI Prism 7500 Sequence Detection System using Applied Biosystems SYBR Green PCR Master mix. Thermocycling was done for 40 cycles, reactions were 15ml total volume (7.5ml SYBR master mix, 0.6ml of 10mM primer, 2ml cDNA, 4.3ml dH_20). Primer sequences are available upon request. Expression levels of *csr-1* isoforms are determined relative to *act-3* mRNA levels.

For ChIP analysis: qRT-PCR was performed on the ABI Prism 7500 Sequence Detection System using Applied Biosystems FAST SYBR Green PCR Master mix. Thermocycling was done for 40 cycles, reactions were 15ml total volume (7.5ml SYBR master mix, 0.6ml of 10mM primer, 2ml Input DNA and 4.3ml dH_20 or 4ml IP DNA and 2.3ml dH_20). Primer sequences are available upon request. Fold enrichment was determined relative to the control, Y47H10A.3, levels (This gene not appreciably targeted by small RNAs.). Error is calculated as described in (Claycomb et al., 2002).

ACCESSION NUMBERS

All RNA sequences extracted from Illumina reads as described were deposited in the NCBI's Gene Expression Omnibus (GEO) (Edgar et al., 2002) and are accessible through GEO Series accession number GSE18165. Included under this accession number are the following data: Small RNAs that coimmunoprecipitate with CSR-1 and the corresponding wild-type input control, 50 ligation dependent (TAP); and small RNA populations from *csr-1(tm892*), *ego-1(om97)*, and a congenic wild-type strain (DA1316), 5´ ligation-dependent (CIP/PNK). Microarray data were deposited in the NCBI's GEO and are accessible through GEO Series accession number GSE18141.

Supplemental Information

Figure II-S1

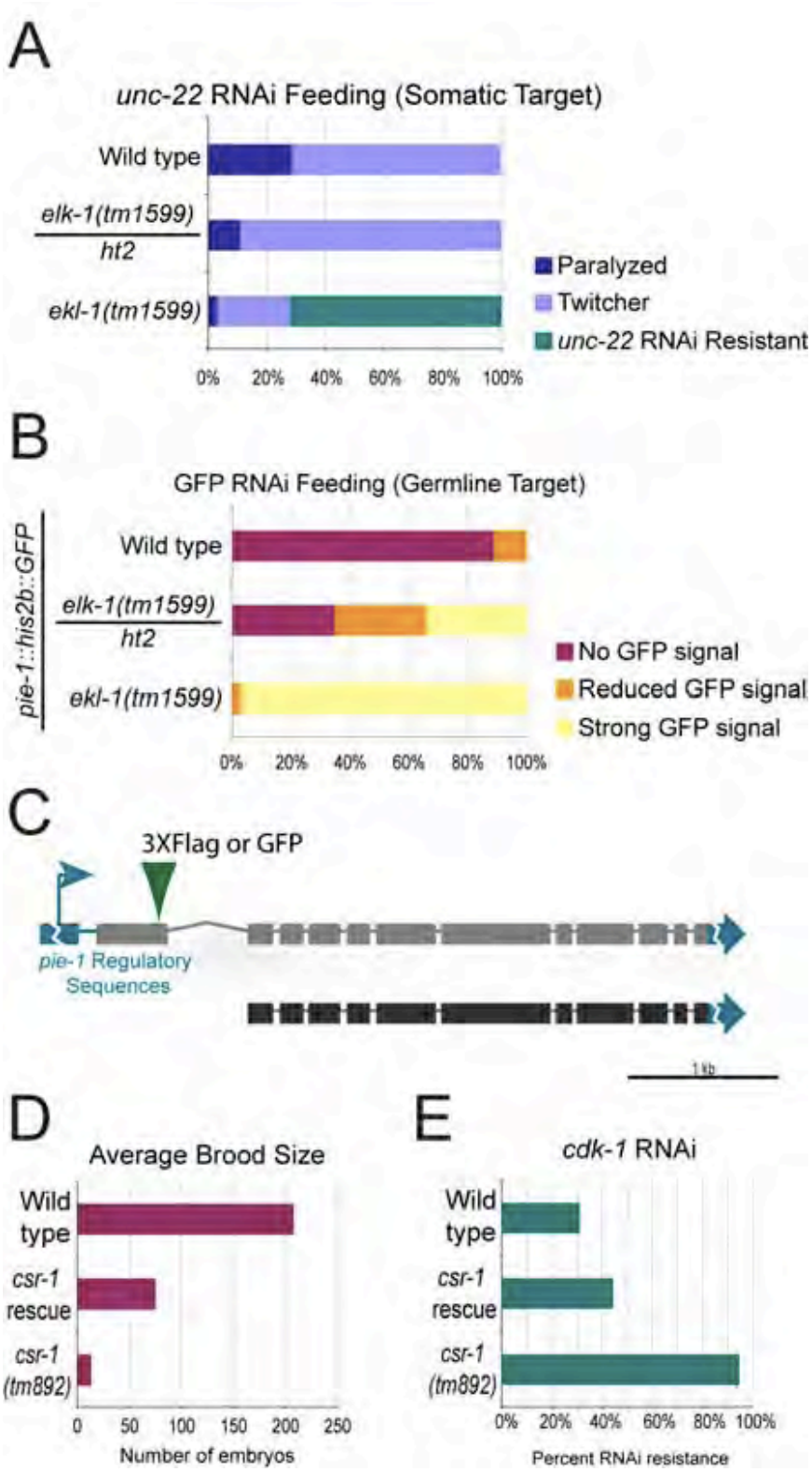
A
unc-22 RNAi Feeding (Somatic Target)
Wild type
elk-1(tm1599)
ht2
ekl-1(tm1599)
Paralyzed
Twitcher
unc-22 RNAi Resistant
0%
20%
40%
60%
80%
100%
B
GFP RNAi Feeding (Germline Target)
pie-1::his2b::GFP
Wild type
elk-1(tm1599)
ht2
ekl-1(tm1599)
No GFP signal
Reduced GFP signal
Strong GFP signal
0%
20%
40%
60%
80%
100%
C
3XFlag or GFP
pie-1 Regulatory Sequences
1 kb
D
Average Brood Size
Wild type
csr-1 rescue
csr-1 (tm892)
0
50
100
150
200
250
Number of embryos
E
cdk-1 RNAi
Wild type
csr-1 rescue
csr-1 (tm892)
0%
20%
40%
60%
80%
100%
Percent RNAi resistance

Figure II-S1. RNAi deficiency of *ekl-1(tm1599)* and rescue of *csr-1(tm892)*

(A) Somatic RNAi Deficiency of *ekl-1(tm1599)* homozygotes, *ekl-1(tm1599)/hT2* heterozygotes, and wild type (N2) animals fed *unc-22* RNAi food.

(B) Germline RNAi Deficiency of *ekl-1(tm1599)* homozygotes, *ekl-1(tm1599)/hT2* heterozygotes, and otherwise wild-type animals carrying a *pie-1::h2b::gfp* transgene, fed *gfp* RNAi food.

(C) Diagram of the *csr-1* transcripts encoded by *csr-1* rescuing transgenes. Only the long isoform possesses an epitope tag and *pie-1* regulatory sequences were used.

(D) Brood size (number of embryos laid) analysis in wild type (N2), *csr-1(tm892)*, and 3x Flag *csr-1* rescue.

(E) RNAi deficiency in wild type (N2), *csr-1(tm892)*, and 3x Flag *csr-1* rescue grown on *cdk-1* RNAi food. *cdk-1* RNAi produces embryos with a one-cell stage arrest that is distinct from the multicellular arrest of *csr-1(tm892)* mutant embryos.

Figure II-S2

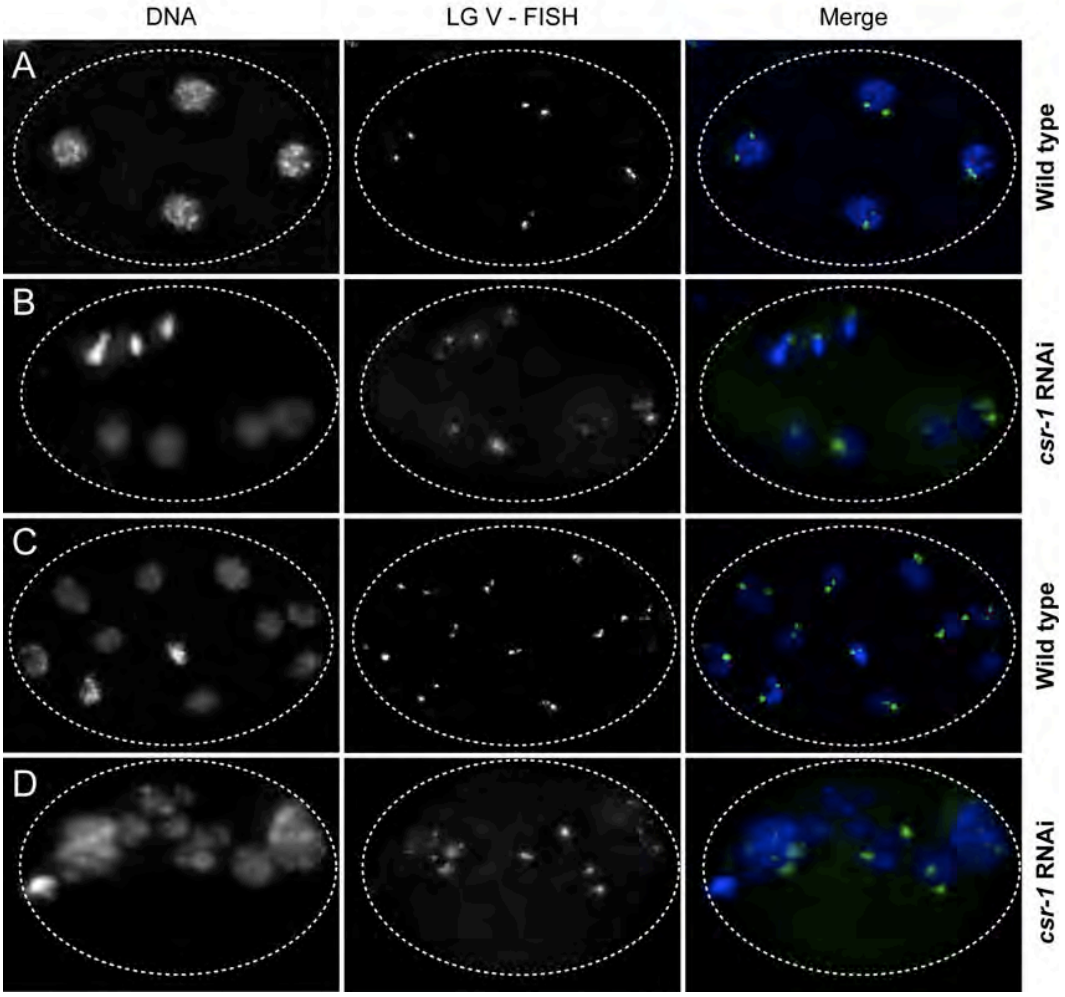
DNA
LG V - FISH
Merge
A
B
C
D
Wild type
csr-1 RNAi
Wild type
csr-1 RNAi

Figure II-S2. Fluorescence *in situ* hybridization with probes against chromosome V

(A) Additional examples of Fluorescence *in situ* hybridization with probes for chromosome V in wild type (A and C) and *csr-1* RNAi (B and D) embryos of approximately the same stage (DNA, blue; LG V FISH signal, green). White dotted lines indicate bondaries of the embryo. More than the appropriate number of FISH signals are evident in morphologically abnormal and aneuploid nuclei of *csr-1* RNAi embryos. Images show projections of Z-stacks through the entire embryo, after deconvolution.

Figure II-S3

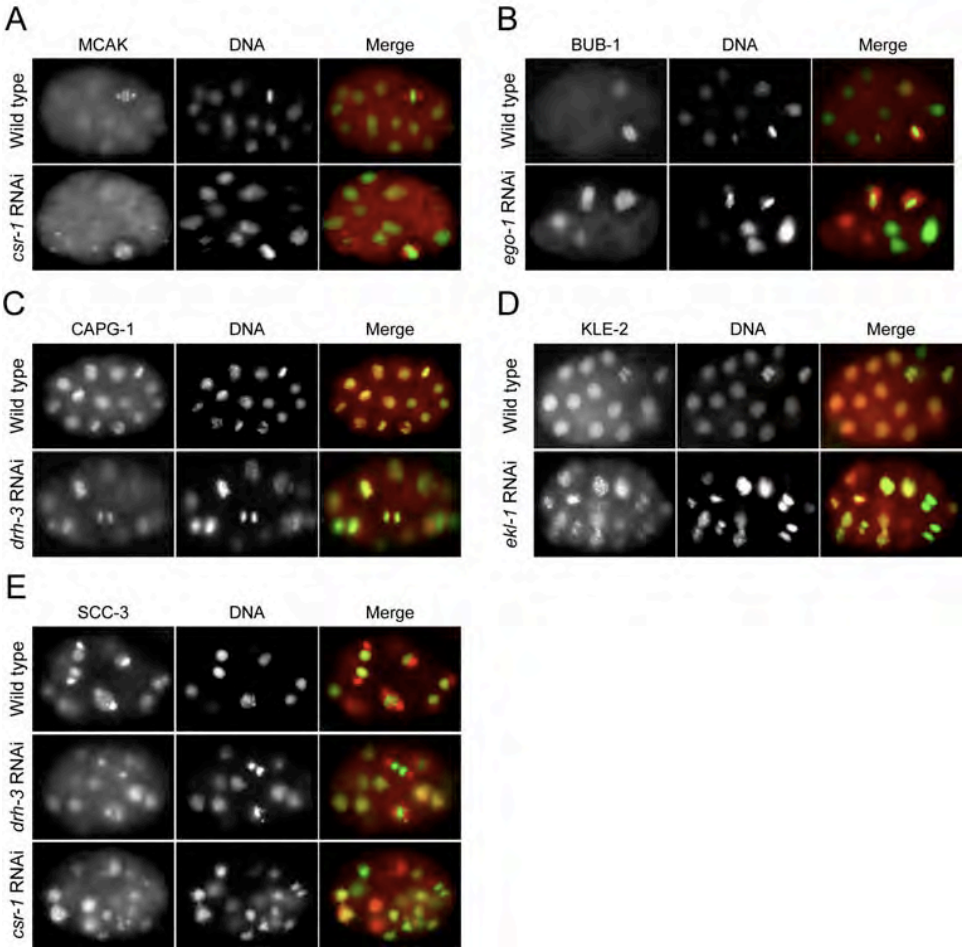
A
MCAK
DNA
Merge
Wild type
csr-1 RNAi
B
BUB-1
DNA
Merge
Wild type
ego-1 RNAi
C
CAPG-1
DNA
Merge
Wild type
drh-3 RNAi
D
KLE-2
DNA
Merge
Wild type
ekl-1 RNAi
E
SCC-3
DNA
Merge
Wild type
drh-3 RNAi
csr-1 RNAi

Figure II-S3. Localization of outer kinetochore, condensin, and cohesin proteins in wild type and RNAi depleted embryos

(A) Localization of the outer kinetochore component MCAK/KLP-7 (left, red in merge) in wild type and *csr-1* RNAi embryos. DNA was stained with DAPI (center, green in merge).

(B) Localization of the outer kinetochore/mitotic checkpoint component BUB-1 (left, red in merge) in wild type and *ego-1* RNAi embryos. DNA was stained with DAPI (center, green in merge).

(C) Localization of the condensin CAPG-1 (left, red in merge) in wild type and *drh- 3* RNAi embryos. DNA was stained with DAPI (center, green in merge).

(D) Localization of the condensin KLE-2 (left, red in merge) in wild type and *ego-1* RNAi embryos. DNA was stained with DAPI (center, green in merge).

(E) Localization of the cohesin SCC-3 (left, red in merge) in wild type and *drh-3* and *csr-1* RNAi embryos. DNA was stained with DAPI (center, green in merge).

Figure II-S4

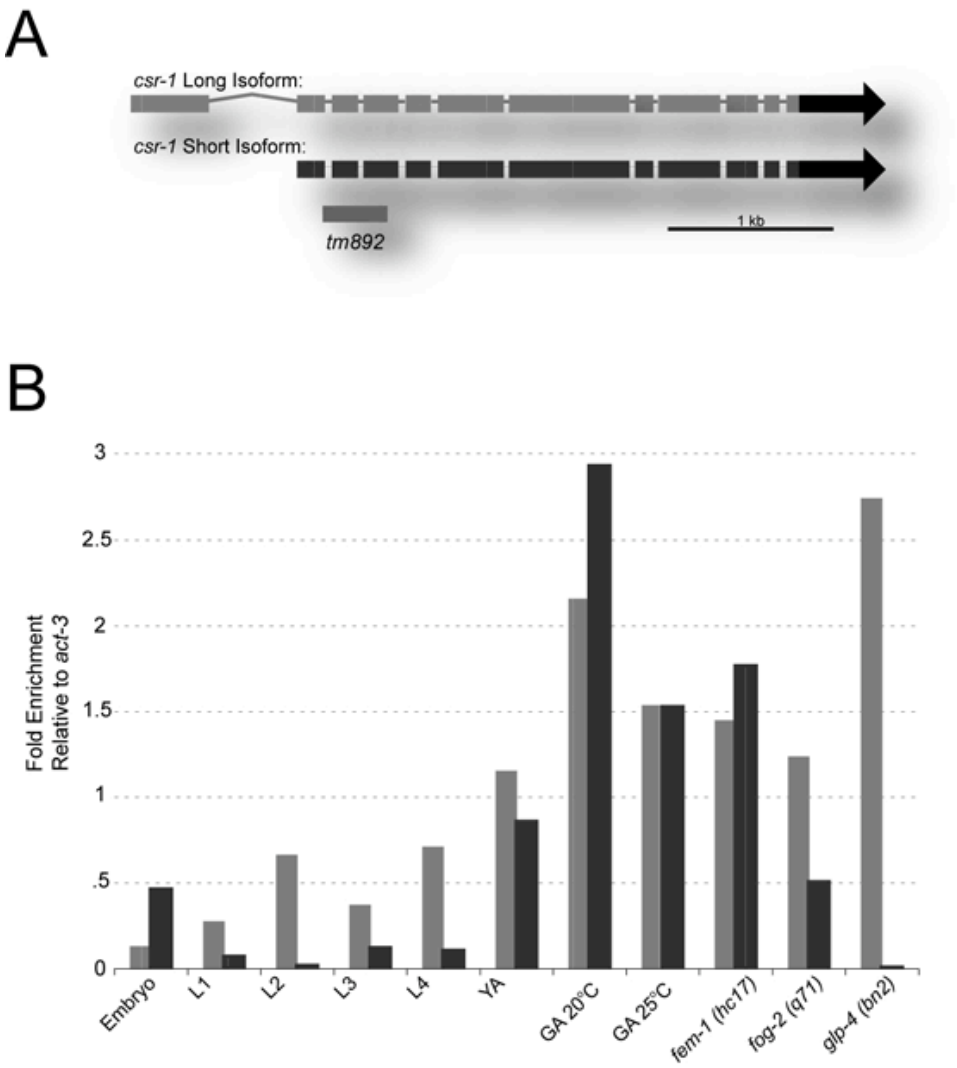
A
csr-1 Long Isoform:
csr-1 Short Isoform:
tm892
1 kb
B
Fold Enrichment Relative to act-3
3
2.5
2
1.5
1
.5
0
Embryo
L1
L2
L3
L4
YA
GA 20°C
GA 25°C
fem-1 (hc17)
fog-2 (q71)
glp-4 (bn2)

Figure II-S4. Quantitative real-time RT-PCR analysis of *csr-1* transcripts

(A) A diagram of the two transcripts generated from the *csr-1* locus, drawn to scale. The position of the deletion allele used in these studies *(tm892)* is marked. Scale bar is 1kb.

(B) *csr-1* mRNA levels of the long isoform (light gray), and the short isoform (dark gray) in various stages of development and germline mutant backgrounds, as described in Figure 3, relative to *act-3* mRNA levels. Forward real-time primers were specific for each isoform, by the use of the SL1 splice leader sequence, which is added to the 5´ end of each *csr-1* transcript, along with 6-10 nucleotides of *csr-1* sequence at the 5´ end of either isoform.

Figure II-S5

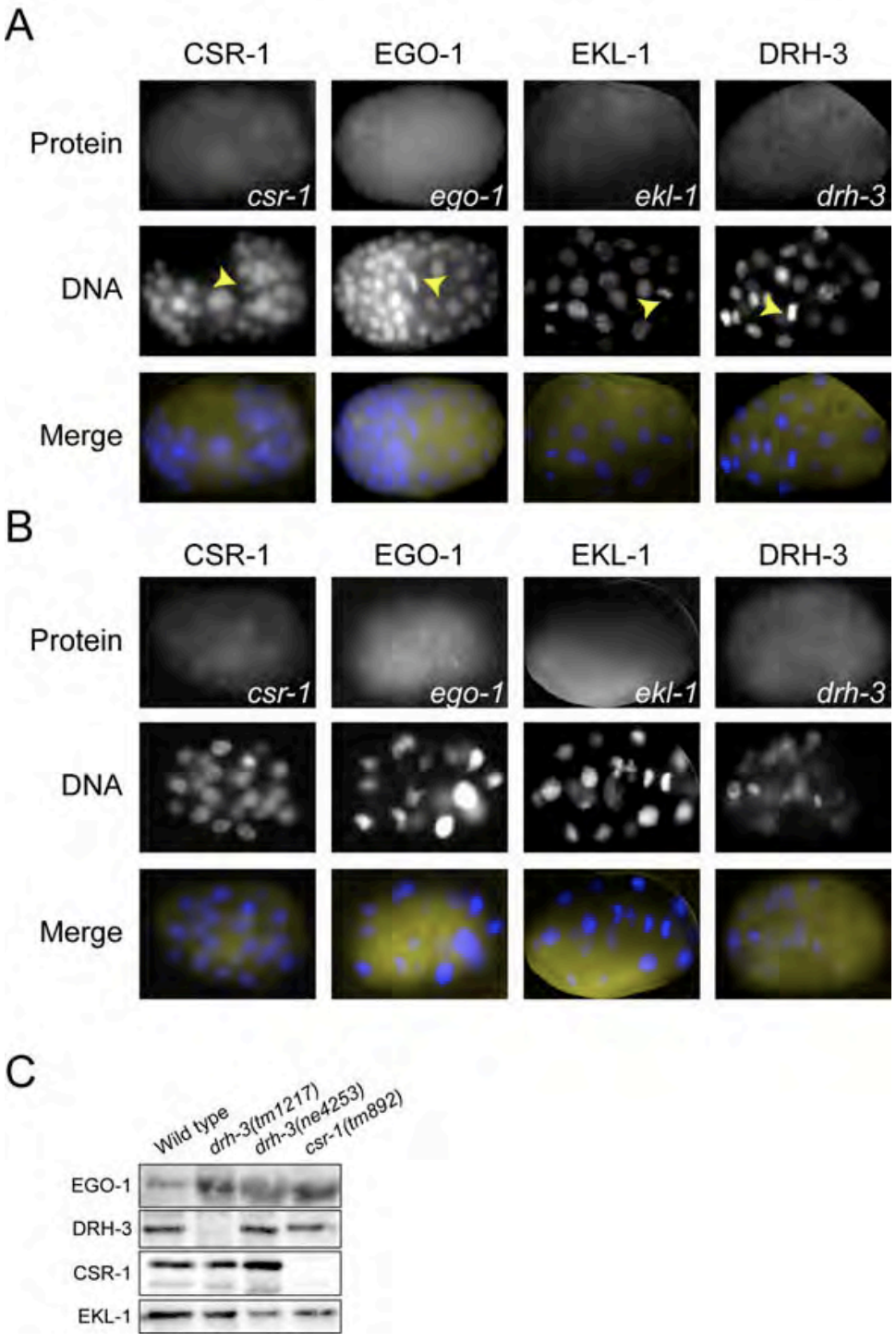
A
CSR-1
EGO-1
EKL-1
DRH-3
Protein
csr-1
ego-1
ekl-1
drh-3
DNA
Merge
B
CSR-1
EGO-1
EKL-1
DRH-3
Protein
csr-1
ego-1
ekl-1
drh-3
DNA
Merge
C
Wild type
drh-3(tm1217)
drh-3(ne4253)
csr-1(tm892)
EGO-1
DRH-3
CSR-1
EKL-1

Figure II-S5. Localization of CSR-1, EGO-1, EKL-1 and DRH-3 is ablated in respective mutant or RNAi-depleted embryos

(A) Embryos depleted by RNAi for each of *csr-1*, *ego-1*, *ekl-1*, and *drh-3* (as marked) were stained for CSR-1, EGO-1, EKL-1, or DRH-3, respectively (as marked, top, yellow in merge). DNA was stained with DAPI (center, blue in merge). At least one metaphase is observable for each embryo in (A) (yellow arrows).

(B) Additional embryos, as in (A), but without any metaphase nuclei.

(C) Western blot analysis of EGO-1, DRH-3, CSR-1, and EKL-1 in wild type, *csr-1(tm892)* mutant, or *drh-3* mutant adults (*drh-3(tm1217)* is a null allele; *drh- 3(ne4253)* is a hypomorphic allele). 50µg of protein lysate was loaded per lane.

Figure II-S6

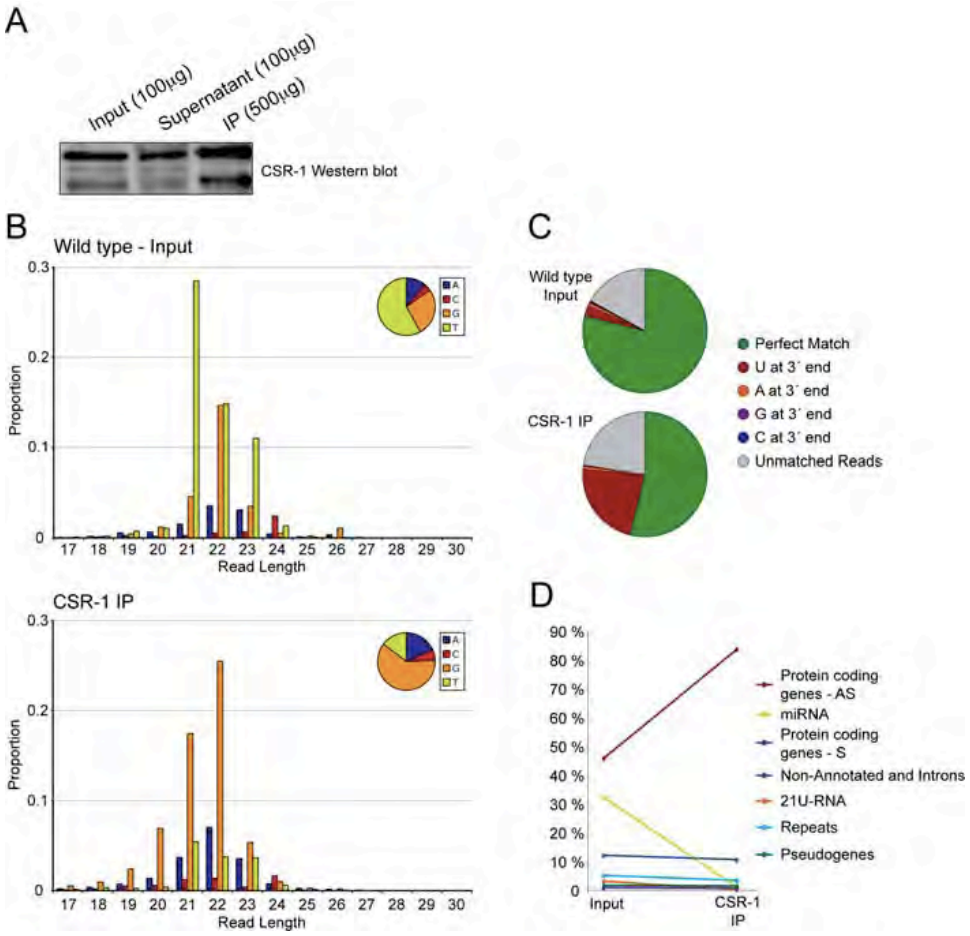
A
Input (100µg)
Supernatant (100µg)
IP (500µg)
CSR-1 Western blot
B
Wild type - Input
Proportion
Read Length
CSR-1 IP
C
Wild type Input
CSR-1 IP
Perfect Match
U at 3´ end
A at 3´ end
G at 3´ end
C at 3´ end
Unmatched Reads
D
Protein coding genes - AS
miRNA
Protein coding genes - S
Non-Annotated and Introns
21U-RNA
Repeats
Pseudogenes
Input
CSR-1 IP

Figure II-S6. Addition of untemplated uridine to the 3’ ends of CSR-1 22G-RNAs

(A) IP/Western blot analysis of CSR-1. 100µg of protein lysate was loaded for Input and Supernatant lanes. 10% of a CSR-1 IP from 5mg of protein lysate was loaded in the IP lane. Blots were probed with anti-CSR-1

(B) Size and first nucleotide distribution of small RNAs cloned in libraries from wild type (N2) Input and CSR-1 IP. Length of read is on the X axis, proportion of reads is on the Y axis and colors indicate the 5´ nucleotide as shown. Inset pie charts indicated the overall proportion of small RNA with each nucleotide at the 5’ position.

(C) Pie charts indicating the relative proportions of small RNAs with perfect matches to the genome (green) vs. having additional nucleotides added at their 3’ end for wild type (N2) Input and CSR-1 IP libraries using the TAP cloning method. A proportion of small RNA reads still did not match the genome after the removal of the additional 3’ nucleotides (gray).

(D) Line plot comparing the relative proportions of small RNA classes for the uridylated reads (reads that match the *C. elegans* genome after removal of uridine(s) from the 3’ end, as in (B.)), between wild-type (N2) Input (left) and CSR-1 IP (right) samples. (AS=antisense, S=sense)

Figure II-S7

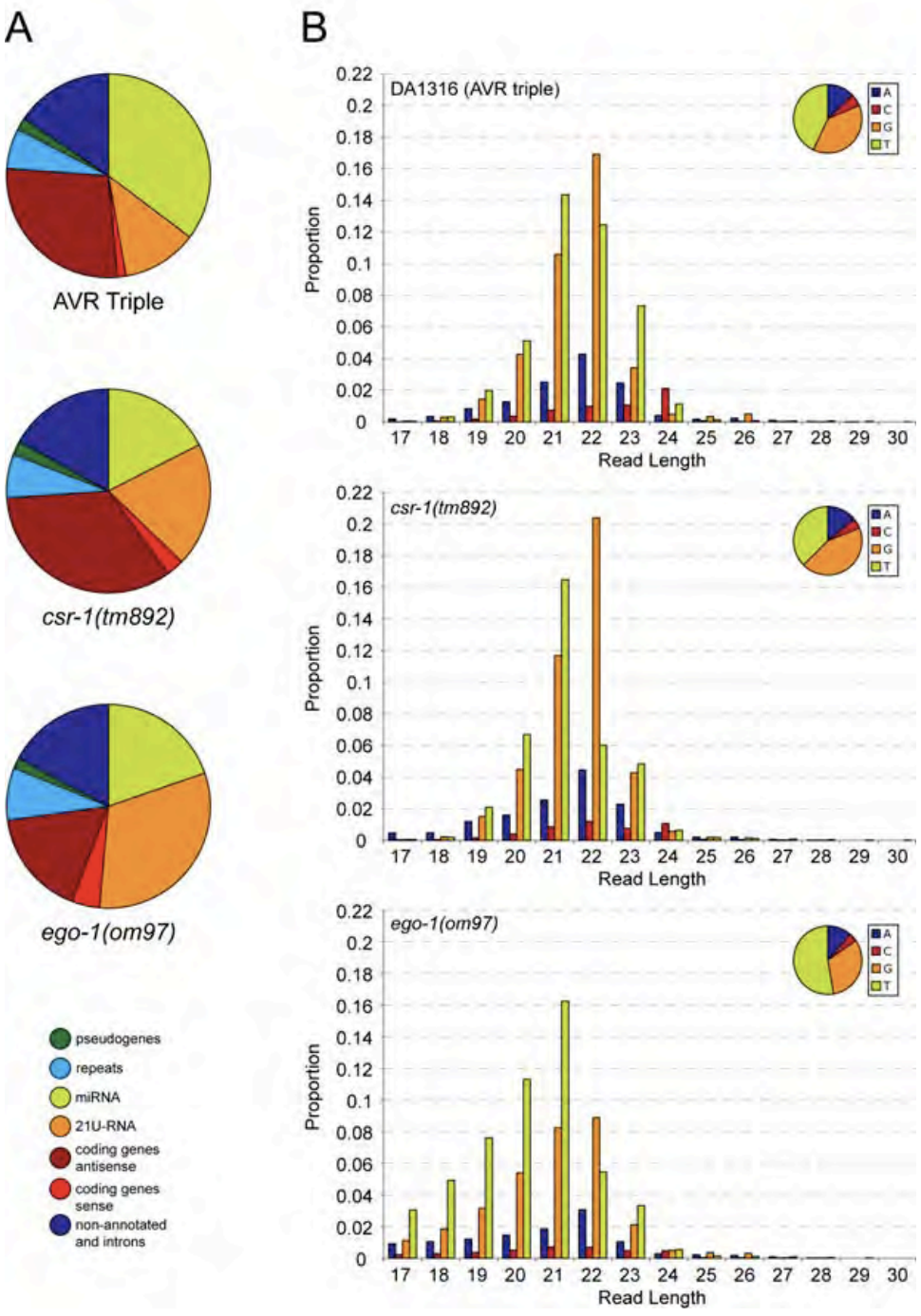
A
AVR Triple
csr-1(tm892)
ego-1(om97)
pseudogenes
repeats
miRNA
21U-RNA
coding genes antisense
coding genes sense
non-annotated and introns
B
DA1316 (AVR triple)
csr-1(tm892)
ego-1(om97)
Proportion
Read Length
A
C
G
T

Figure II-S7. Analysis of *csr-1(tm892)*, *ego-1(om97)*, and DA1316 small RNA libraries

(A) Pie charts indicating the relative proportions of various classes of small RNAs in *csr-1(tm892)* and *ego-1(om97)* mutants and a congenic wild-type strain (DA1316). DA1316 possesses three mutations that render it resistant to the drug ivermectin, which is used for selection of uniform populations of homozygous mutant adult worms (Duchaine et al., 2006).

(B) Size and first nucleotide distribution of small RNAs cloned in libraries from *csr-1(tm892)*, *ego-1(om97)* mutants and DA1316. Length of read is on the X axis, proportion of reads is on the Y axis and colors indicate the 5´ nucleotide as shown. Inset pie charts indicate the overall proportion of small RNA with each nucleotide at the 5´ position. *drh-3(ne4253)* and *ekl-1(tm1599)* small RNA compositions are described in (Gu et al., 2009), cosubmitted).

Figure II-S8

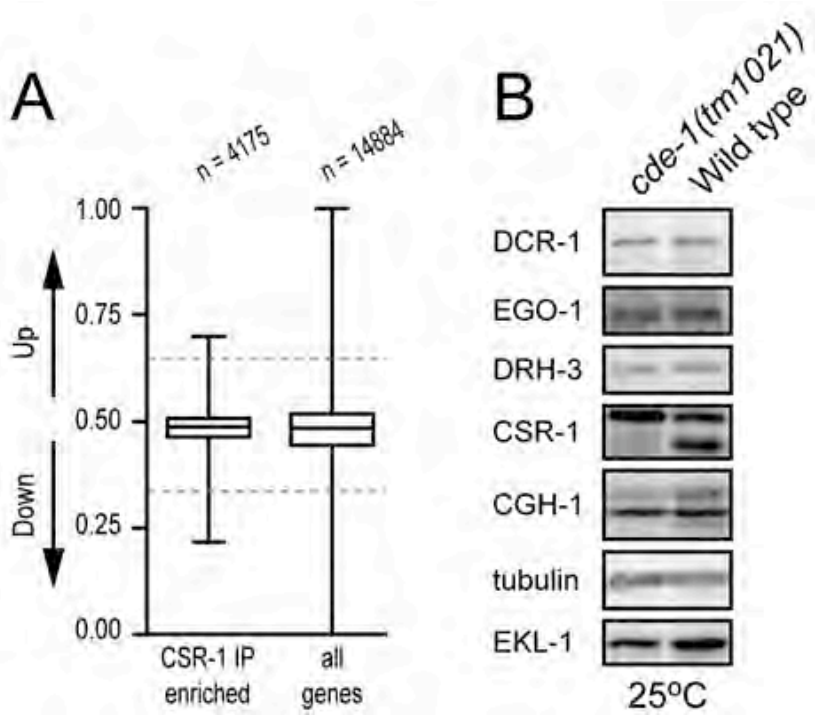
A
n = 4175
n = 14884
1.00
0.75
0.50
0.25
0.00
Up
Down
CSR-1 IP enriched
all genes
B
cde-1(tm1021)
Wild type
DCR-1
EGO-1
DRH-3
CSR-1
CGH-1
tubulin
EKL-1
25°C

Figure II-S8 CSR-1 22G-RNA target mRNA and protein levels are not changed in *drh-3(ne4253)* or *cde-1(tm1021)* mutants

(A) Box and whisker plot of CSR-1 target mRNA levels in the *drh-3(ne4253)* mutant as compared to wild type for CSR-1 22G-RNA targets (left) and for all genes (right).

(B) Western blot analysis of wild type and *cde-1(tm1021)* protein lysates, generated from animals grown at 25º C and probed for CSR-1 22G-RNA target proteins. EKL-1 is not a CSR-1 target and is shown as a loading control. [Note that only one isoform of CSR-1 seems to be expressed in *cde-1(tm1021)* lysates]

Figure II-S9

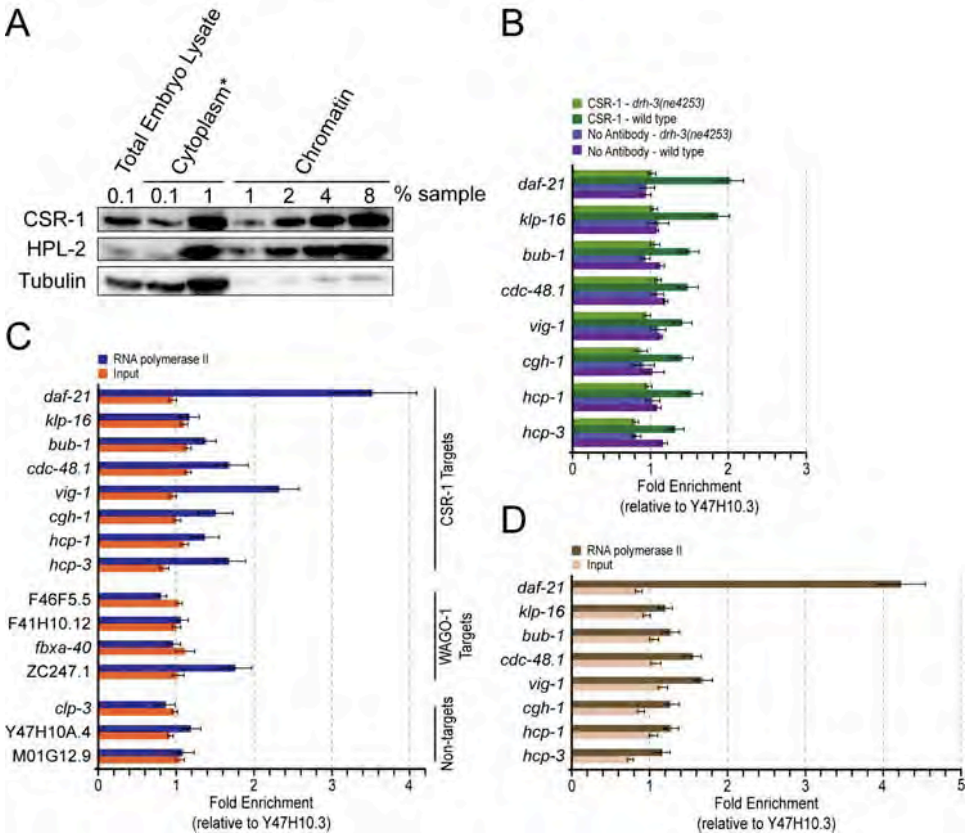
A
Total Embryo Lysate
Cytoplasm*
Chromatin
0.1 0.1 1 1 2 4 8 % sample
CSR-1
HPL-2
Tubulin
B
CSR-1 - drh-3(ne4253)
CSR-1 - wild type
No Antibody - drh-3(ne4253)
No Antibody - wild type
daf-21
klp-16
bub-1
cdc-48.1
vig-1
cgh-1
hcp-1
hcp-3
Fold Enrichment
(relative to Y47H10.3)
C
RNA polymerase II
Input
daf-21
klp-16
bub-1
cdc-48.1
vig-1
cgh-1
hcp-1
hcp-3
F46F5.5
F41H10.12
fbxa-40
ZC247.1
clp-3
Y47H10A.4
M01G12.9
CSR-1 Targets
WAGO-1 Targets
Non-targets
Fold Enrichment
(relative to Y47H10.3)
D
RNA polymerase II
Input
daf-21
klp-16
bub-1
cdc-48.1
vig-1
cgh-1
hcp-1
hcp-3
Fold Enrichment
(relative to Y47H10.3)

Figure II-S9. CSR-1 association with chromatin is 22G-RNA dependent

(A) Western blots of various amounts (as indicated) of total protein lysate, generated from embryos, cytoplasmic protein lysate and isolated chromatin (prepared according to (Chu et al., 2006) were probed for CSR-1, HPL-2 (one of two HP1 chromatin-binding proteins in *C. elegans*) (Couteau et al., 2002), and Tubulin. * The cytoplasmic fraction is invariably contaminated with nuclear content using this preparation method.

(B) Embryo ChIP/Quantitative Real-time PCR analysis of CSR-1 and agarose beads alone/no antibody control enrichment at CSR-1 22G-RNA target loci in wild type and *drh-3(ne4253)* embryos. These samples were prepared without the addition of DTBP, and are from a single, representative experiment. Fold enrichment is relative to Y47H10A.3. Error bars are standard deviation from the mean of three replicates of a single ChIP sample.

(C) Embryo ChIP/Quantitative Real-time PCR analysis of RNA Polymerase II enrichment and Input control enrichment at 22G-RNA target loci and non-targeted control loci, as in Figure 6. These samples were prepared from the same extracts as those in Figure 6 and are from a single, representative experiment. Fold enrichment is relative to Y47H10A.3. Error bars are standard deviation from the mean of three replicates of a single ChIP sample. Many CSR-1 22G-RNA target loci as well as the WAGO-1 22G-RNA target, ZC247.1, are enriched for RNA Polymerase II binding.

Table II-S1

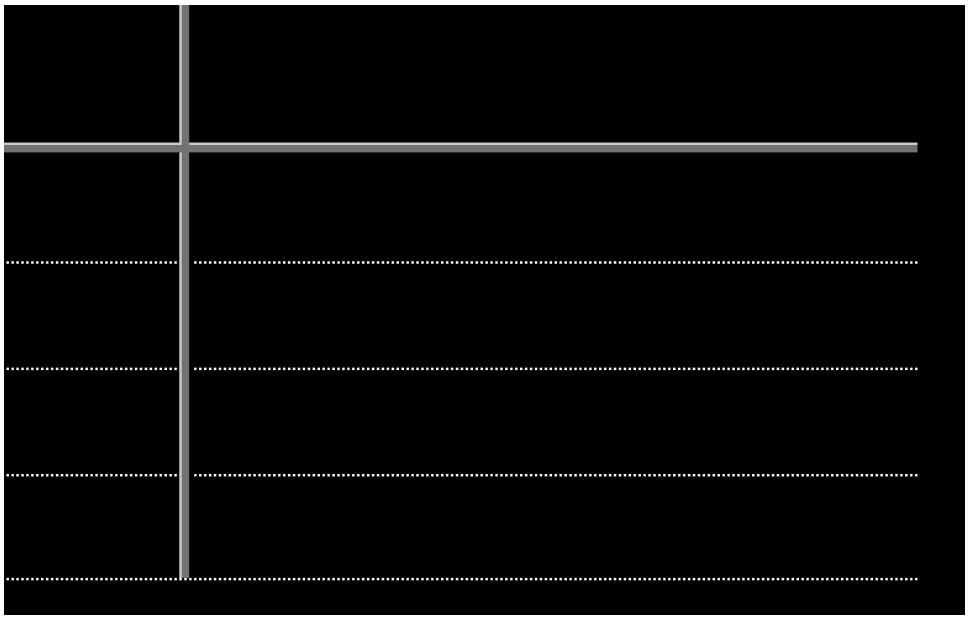

Table II-S1. Localization of RNAi Factors in RNAi depleted embryos

All determinations made on examination of 10-20 early embryos generated by RNAi to each of the factors. ---, lack of detectable staining in >95% of embryos. +, less than 25% of embryos counted had localization to disrupted metaphase plate (and embryo P granules, for DRH-3 and CSR-1). ++, up to 75% of embryos counted had localization to disrupted metaphase plate (and embryo P granules, for DRH-3 and CSR-1). +++, >95% of embryos counted had localization to disrupted metaphase plate (and embryo P granules, for DRH-3 and CSR-1).

* Residual levels in P granules; all staining lost as animals age

CHAPTER III

Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the *C. elegans* Germline

Author Contributions

Chapter III contributions. J.Vasale contributed the following to this section: She generated balanced RdRP double mutant (*rrf-1 ego-1)* strain with appropriate genetic markers to allow large-scale selection of RdRP double homozygous mutant. Also she performed small RNA extraction, purification, and preparation of cDNA libraries from this strain and other RdRP single mutants (*ego-1* and *rrf-1*) in parallel with the congenic wild-type control strain (DA1316). She performed small RNA Northern blot experiments on RdRP mutant small RNA populations and contributed to analyses and interpretation of deep-sequencing results. These data are presented in Figure III-3A and B, Figure III-5B, Figure IiI-6A, and Figure III-S7. This chapter has been published as "Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the C. elegans Germline" *Molecular Cell* 2009, Oct 23;36(2):231-44; *****Gu W, *****Shirayama M, *****Conte D, Jr., Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, Chen CC, Chaves DA, Duan S, Kasschau KD, Falgren N, Yates JR, 3rd, Mitani S, Carrington JC, Mello CC. License number 2374831415649

SUMMARY

Endogenous small RNAs (endo-siRNAs) interact with Argonaute proteins to mediate sequence specific regulation of diverse biological processes. Here, we combine deep-sequencing and genetic approaches to explore the biogenesis and function of endo-siRNAs in *C. elegans*. We describe conditional alleles of the Dicer-related helicase, *drh-3*, that abrogate both RNA interference and the biogenesis of endo-siRNAs, called 22G-RNAs. DRH-3 is a core component of RNA-dependent RNA polymerase (RdRP) complexes essential for several distinct 22G-RNA systems. We show that, in the germline, one system is dependent on worm-specific Argonautes, including WAGO-1, which localizes to germline nuage structures called P granules. WAGO-1 silences certain genes, transposons, pseudogenes, and cryptic loci. Finally, we demonstrate that components of the nonsense-mediated decay pathway function in at least one WAGO-mediated surveillance pathway. These findings broaden our understanding of the biogenesis and diversity of 22G-RNAs and suggest additional regulatory functions for small RNAs.

INTRODUCTION

Regulatory pathways related to RNA interference (RNAi) utilize small RNAs to guide the sequence-specific modulation of gene expression, chromatin structure, and innate immune function (Ding and Voinnet, 2007; Moazed, 2009). Small RNA classes can be distinguished based on a number of factors, including mechanism of biogenesis, mode of regulation or function, and the Argonaute proteins with which they interact (Ghildiyal and Zamore, 2009). Argonaute family members are structurally related to ribonuclease (RNase) H and bind to the ends of single stranded small RNAs, presenting the central residues for base pairing interactions with target nucleic acids (Parker et al., 2005; Song et al., 2004).

Diverse pathways have been implicated in the biogenesis and loading of various small RNA species onto their respective Argonaute proteins (Siomi and Siomi, 2009). Double-stranded RNA (dsRNA), including the stem-loop precursors of miRNAs, are processed into mature 5′ monophosphorylated short RNAs by Dicer, an RNase IIII-like enzyme. The biogenesis of PIWI-interacting RNAs (piRNAs) is less well understood and appears to be independent of Dicer. In flies, PIWI-mediated cleavage events appear to define the 5′ ends of new piRNAs (Brennecke et al., 2007; Gunawardane et al., 2007), whereas 3′ end maturation occurs through an undefined mechanism that ultimately results in the 2′-O-methylation of the 3′ residue (Klattenhoff and Theurkauf, 2008). In plants, fungi, and nematodes, silencing signals are amplified from target RNA by RdRPs. In some cases, RdRPs and Dicer function in a concerted manner to synthesize and

process dsRNA, producing siRNAs with 5′ monophosphate residues (Colmenares et al., 2007; Lee and Collins, 2007). In nematodes, RdRPs also catalyze the unprimed, de novo synthesis of 5′-triphosphorylated RNAs that appear to be loaded directly, without Dicer processing, onto members of an expanded clade of worm specific Argonautes (WAGOs) (Aoki et al., 2007; Pak and Fire, 2007; Sijen et al., 2007; Yigit et al., 2006).

In plants, flies, and mammals, several classes of small RNA species are derived from transposable elements and repeat sequences as well as a subset of nonrepetitive protein-coding sequences and pseudogenes (Czech et al., 2008; Ghildiyal et al., 2008; Kasschau et al., 2007; Okamura et al., 2008; Tam et al., 2008; Watanabe et al., 2008). In many cases, these small RNAs are derived from loci capable of dsRNA formation, including transposable elements, inverted repeats, and bidirectionally transcribed regions. Naturally occurring small RNAs are often coincident with pericentric heterochromatin and have been implicated in the establishment and/or maintenance of heterochromatin and in centromere function (Hall et al., 2003; Verdel et al., 2004). In *C. elegans*, endogenous small RNAs (endo-siRNAs) have been reported to target several hundred loci, including protein-coding as well as noncoding loci (Ambros et al., 2003; Lim et al., 2003b; Ruby et al., 2006).

Here, we demonstrate that the Dicer-related helicase DRH-3 is essential for the biogenesis of RdRP-derived small RNAs in *C. elegans*. We have named these small RNAs 22G-RNAs based on their strong propensity for having a 5′

guanosine (5´G) residue and a length of 22 nt. 22G-RNAs are abundantly expressed in the germline and are maternally deposited in oocytes. Surprisingly, the majority of 22G-RNAs target unique genome sequences, including ~50% of the annotated coding genes in *C. elegans*. In addition to DRH-3, we show that the RdRPs, RRF-1 and EGO-1, and the tudor-domain protein EKL-1 are required for the biogenesis of 22G-RNAs. 22G-RNAs can be divided into two major systems based on the associated Argonautes and their cofactors. One of these systems is dependent on RDE-3, MUT-7, and members of the WAGOs, including WAGO-1, while the second is dependent on the Argonaute CSR-1 and the nucleotidyl transferase CDE-1. The WAGO 22G-RNA system silences transposons,pseudogenes, and cryptic loci as well as certain genes, while the CSR-1 system functions to promote chromosome segregation (Claycomb et al., 2009). Finally, we demonstrate a role for components of the nonsense-mediated mRNA decay (NMD) pathway in 22G-RNA biogenesis, although our results do not necessarily implicate NMD per se in 22G-RNA biogenesis. Our findings uncover a surprisingly rich maternal inheritance of small RNAs and raise many questions about the potential significance of these RNA species in the transmission of epigenetic information and genome surveillance.

RESULTS

Novel Alleles of *drh-3* Disrupt RNAi

A previous study identified DRH-3 as a Dicer-interacting factor required for germline RNAi and for viability (Duchaine et al., 2006). Animals homozygous for the *drh-3(tm1217)* deletion (a putative null allele) are infertile, and RNAi targeting drh-3 results in a penetrant embryonic lethal phenotype with defects in chromosome segregation and in the production of both Dicer-dependent and Dicer-independent small RNA populations (Duchaine et al., 2006; Nakamura et al., 2007).

Our screens for RNAi-deficient (Rde) strains identified three additional alleles of *drh-3*. The homozygous mutants bearing these non-null alleles are viable at 20ºC but are infertile at 25ºC. Each allele alters a distinct amino acid within the HELICc domain of the putative helicase (Figure III-1A), and, based on genetic tests, each behaves like a partial loss-of-function mutation. Consistent with previous work demonstrating that DRH-3 is required for germline RNAi (Duchaine et al., 2006), these *drh-3* point mutants were defective for RNAi targeting the maternal gene *pos-1* (Figure III-1B) and exhibited varying degrees of somatic RNAi deficiency. For example, the *drh-3(ne3197)* mutant was strongly resistant to RNAi targeting the somatically expressed basement membrane collagen *let-2*, while the *ne4253* and *ne4254* alleles are partially and fully sensitive, respectively (Figure III-1B). Although the *drh-3* point mutants exhibit a range of phenotypes that increase in penetrance at 25ºC, they do not appear to

Figure III-1

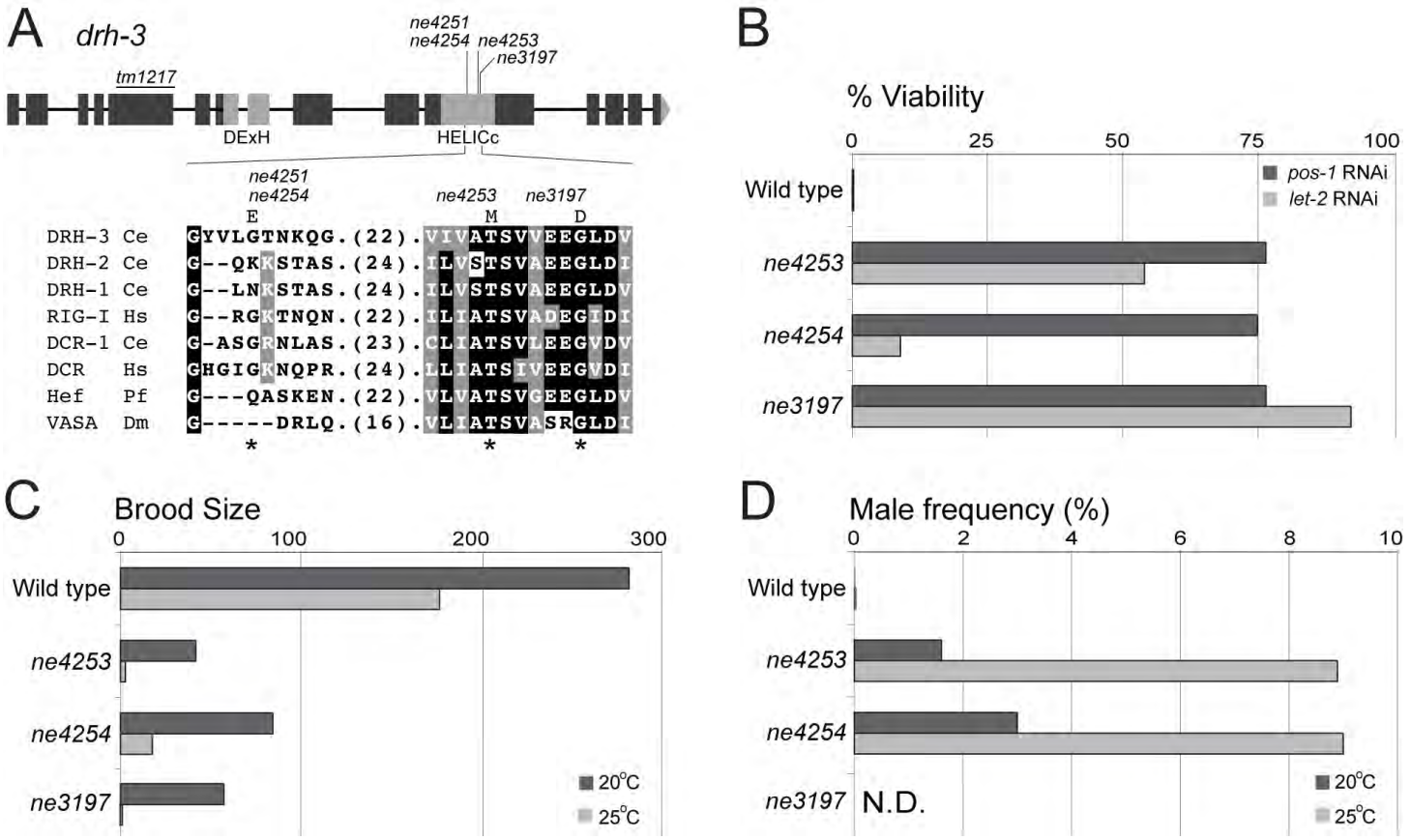
A
drh-3
tm1217
ne4251
ne4254
ne4253
ne3197
DExH
HELICc
ne4251
ne4254
ne4253
ne3197
E
M
D
DRH-3 Ce GYVLGTNKQG.(22).VIVATSVVEEGLDV
DRH-2 Ce G--QKKSTAS.(24).ILVSTSVAEEGLDI
DRH-1 Ce G--LNKSTAS.(24).ILVSTSVAEEGLDV
RIG-I Hs G--RGKTNQN.(22).ILIATSVADEGIDI
DCR-1 Ce G-ASGRNLAS.(23).CLIATSVLEEGVDV
DCR Hs GHGIGKNQPR.(24).LLIATSIVEEGVDI
Hef Pf G---QASKEN.(22).VLVATSVGEEGLDV
VASA Dm G-----DRLQ.(16).VLIATSVASRGLDI
*
*
*
B
% Viability
0
25
50
75
100
Wild type
ne4253
ne4254
ne3197
pos-1 RNAi
let-2 RNAi
C
Brood Size
0
100
200
300
Wild type
ne4253
ne4254
ne3197
20°C
25°C
D
Male frequency (%)
0
2
4
6
8
10
Wild type
ne4253
ne4254
ne3197
N.D.
20°C
25°C

Figure III-1. Hypomorphic alleles of *drh-3* are RNAi deficient and temperature sensitive

A) Schematic of the *drh-3* gene structure. Top panel: the conserved DExH and HELICc domains and the *drh-3* lesions; bottom panel: the four missense alleles, as indicated, map the HELICc domain.

B) RNAi deficient phenotypes of *drh-3* mutants. The fraction of viable embryos produced by animals exposed to *pos-1* RNAi food or the viability of animals reared on *let-2* RNAi food at 20°C.

(C and D) Brood size and Him phenotypes of *drh-3* mutants at 20°C (dark bars) and 25°C (light bars). Mean brood size (top panel) of at least 10 hermaphrodites was determined by counting the number of embryos produced. The frequency of males among viable offspring was determined (bottom panel).

be classic temperature-sensitive mutants. Null alleles of many germline factors, including several RNAi-pathway genes, cause similar conditional-sterile phenotypes that likely reflect an underlying temperature-dependent process in the germline that is uncovered in these mutant backgrounds. The spectrum of phenotypes observed in the *drh-3* point-mutant strains, including sterility, embryonic lethality, and high incidence of males (*him*) (Figures III-1C and III-1D), are similar to those observed for null alleles of several Mutator (Mut) class, Rde strains and, as observed with Mut strains, the RNAi defect associated with the *drh-3* alleles was not affected by temperature: each allele was fully resistant to *pos-1* RNAi at the permissive temperature of 20"C (Figure III-1A).

To determine whether the *drh-3* point mutants showed small RNA defects similar to the deletion mutant (Duchaine et al., 2006), small RNA populations were isolated and analyzed by northern blot to detect previously characterized endo-siRNAs (Figure III-S1). All of the previously examined DRH-3-dependent endo-siRNAs were dramatically reduced or undetected in the *drh-3* point-mutant samples, while miRNA biogenesis was unaffected. These results are consistent with the notion that these alleles of *drh-3* cause a partial loss of *drh-3* function.

DRH-3 Is Essential for the Biogenesis of 22G-RNAs

In the course of analyzing the small RNA phenotypes associated with the drh-3 mutants, we observed that a prominent small RNA species of ~22 nt was virtually absent in small RNA samples prepared from each of the *drh-3* mutants (Figure

III-2A). A second prominent small RNA species of ~21 nt appeared to be unaffected in *drh-3*. Neither small RNA species was altered in samples prepared from a *dcr-1(ok247)* deletion mutant (Figure III-2A). Thin-layer chromatography experiments indicated that the 22 nt small RNAs have a 5´G and are resistant to terminator exonuclease, suggesting the presence of a 5´ cap or polyphosphate (Figure III-2B). The 21 nt small RNAs are comprised of both 5´G and 5´ uracil (5´U) species, the latter of which we recently identified as the 21U-RNAs, piRNAs, associated with PRG-1 (Batista et al., 2008). In contrast to 21U-RNAs, the majority of the 22 nt small RNAs were sensitive to periodate, indicating that the 3´ end is not modified (Ruby et al., 2006). Both 22 nt and 21 nt 5´G small RNAs were dramatically reduced in the *drh-3* mutants, whereas 5´U small RNAs were unaffected. These data are consistent with the notion that the 22 nt small RNAs represent an abundant pool of 5´-triphosphorylated products of endogenous RdRP (Aoki et al., 2007; Pak and Fire, 2007; Sijen et al., 2007).

To identify and characterize the DRH-3-dependent small RNAs on a genome-wide level, small RNAs between 18 and 26 nt were cloned from wild-type and *drh-3* mutant animals using a protocol compatible with cloning small RNAs bearing 5´ triphosphates. Illumina sequencing of wild-type and *drh-3* libraries yielded 2.34 million and 4.33 million reads that perfectly match the *C. elegans* genome, respectively. Consistent with our biochemical analyses, both the size distribution and first nucleotide composition of small RNAs were

Figure III-2

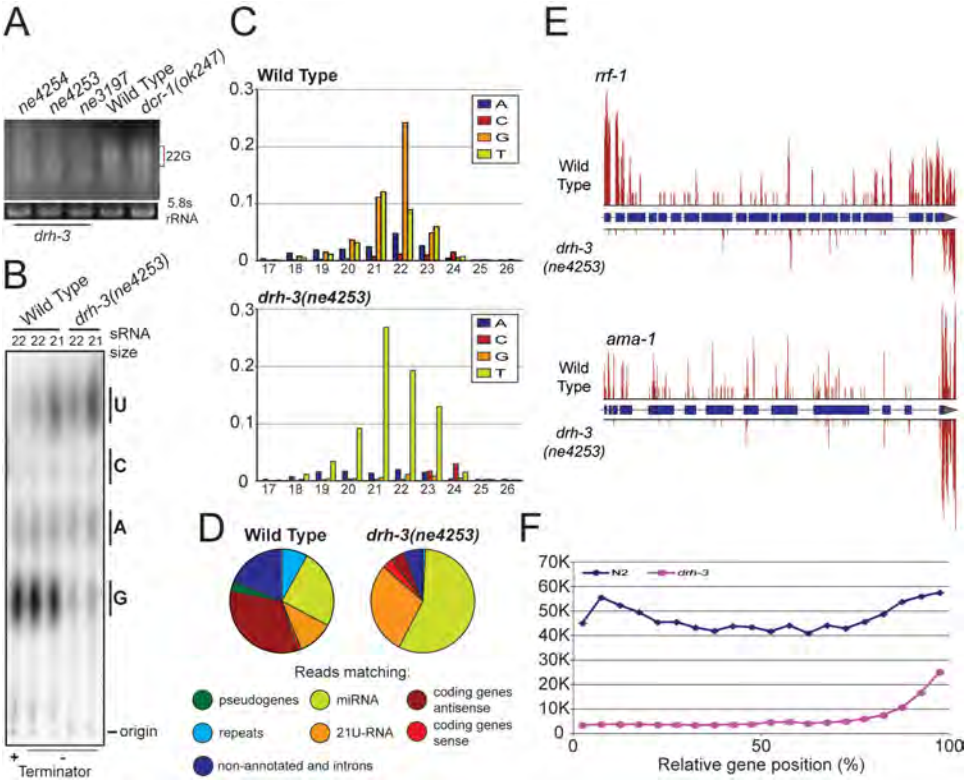
A
ne4254
ne4253
ne3197
Wild Type
dcr-1(ok247)
22G
5.8s rRNA
drh-3
B
Wild Type
drh-3(ne4253)
22 22 21 22 21
sRNA size
U
C
A
G
origin
+
-
Terminator
C
Wild Type
drh-3(ne4253)
A
C
G
T
0.3
0.2
0.1
0
17 18 19 20 21 22 23 24 25 26
D
Wild Type
drh-3(ne4253)
Reads matching:
pseudogenes
miRNA
coding genes antisense
repeats
21U-RNA
coding genes sense
non-annotated and introns
E
rrf-1
Wild Type
drh-3 (ne4253)
ama-1
Wild Type
drh-3 (ne4253)
F
N2
drh-3
70K
60K
50K
40K
30K
20K
10K
0
0
50
100
Relative gene position (%)

Figure III-2. DRH-3 is required for the biogenesis of 22G-RNAs

(A) Ethidium bromide staining of small RNAs. The 22nt position is denoted by bracket. Note persistence of a ~21nt band in all samples. 5.8S rRNA was used as loading control.

(B) Thin-layer chromatography analysis of the first nucleotide of gel-purified 22 and 21nt RNA, as indicated, from wild type and *drh-3* mutant samples. Bars (right), the position of each nucleotide; “+”, treated with Terminator exonuclease; “-”, untreated.

(C) Length and first nucleotide distribution of genome matching reads from wild type and *drh-3* mutant small RNA libraries. Sense structural RNAs were excluded in the analysis.

(D) Distribution of reads that match the indicated genome annotations sequenced in wild type and *drh-3* mutant small RNA libraries.

(E) Examples of small RNA distribution within *rrf-1* and *ama-1*. The density of antisense reads is indicated by vertical bars above (wild-type) or below (*drh-3*) the gene structure.

(F) Genome-wide analysis of small RNA distribution within genes. Total antisense reads (y-axis) plotted according to the relative position (%) within all genes (x-axis; from 5´ to 3´).

dramatically altered in the *drh-3* sample (Figure III-2C). First, wild-type reads peaked sharply at 21–22 nt, comprising 25% and 36% of the total reads, respectively. Whereas 21 nt reads had similar levels of 5´U and 5´G residues (11% and 10% of total reads, respectively), ~60% of 22 nt reads started with 5´G (~21% of total reads). Strikingly, reads with 5´G were strongly depleted from the *drh-3* mutant sample, resulting in a dramatic enrichment of reads with 5´U.

After removing structural RNA degradation products from the data set, about one-third of the wild-type reads matched to miRNAs (24.7%) and 21U-RNAs (11%) (Figure III-2D). The most abundant class of small RNA reads from wild-type samples (34%) were antisense to protein-coding genes. The remaining small RNA reads were derived from transposons and other repetitive loci (~10%) as well as nonannotated loci (~20%). In contrast, the *drh-3* sample was strongly depleted of endo-siRNAs targeting protein-coding genes, pseudogenes, repeats, and nonannotated loci but enriched proportionately for reads matching miRNAs and 21U-RNAs. Together, our biochemical and deep-sequencing data indicated that DRH-3 is essential for the biogenesis of an abundant class of endo-siRNAs expressed in *C. elegans*. Based on the propensity for 22 nt length and 5´G residue, we refer to these small RNAs as 22G-RNAs, which include what were previously identified as endogenous siRNAs (Ambros et al., 2003; Ruby et al., 2006).

22G-RNAs Are Enriched at Transcript Termini

While examining the distribution of 22G-RNAs targeting protein coding loci, we observed that 22G-RNAs were most abundant toward the 3´ end of many protein-coding transcripts. For example, 22G-RNAs were clearly enriched at the 3´ ends of both *rrf-1* and *ama-1* transcripts in the wild-type sample and tapered toward the 5´ end (Figure III-2E). Interestingly, 22G-RNAs were also enriched at the 5´ end of *rrf-1*. Remarkably, the remnant of *rrf-1* and *ama-1* 22G-RNAs that were cloned from drh-3 mutants mapped almost exclusively to the 3´ end of these loci. This pattern might be expected if RdRP initiates 22G-RNA biogenesis at or near the 3´ end of a transcript.

To examine this on a larger scale, each transcript for which 22G-RNA reads occur in both wild-type and *drh-3* mutant samples was divided into 20 consecutive intervals of equal size. The total number of 22G-RNAs that map to each interval was plotted for both wild-type and *drh-3*. In wild-type, 22G-RNAs were more abundant toward both the 5´ and 3´ termini of predicted transcripts (Figure III-2F). This terminal distribution of 22G-RNAs represented a general trend for most target genes and was not caused by a few genes with a high number of small RNAs at either end (Figure III-S2).

Markedly different results were obtained when this analysis was applied to the *drh-3* mutant data. In *drh-3* mutants, while 22G-RNAs were greatly depleted, they were not depleted uniformly across their targets (Figures III-2F and III-S2). The levels of 22G-RNAs within the first 15 bins (representing 75% transcript

length) were disproportionately depleted, resulting in a marked exponential trend in 22G RNA levels over the remaining 5 bins, peaking at the 3´ end. This 3´ end enrichment was evident for about 86% of the 22G-RNA targets. Together, these results raise the possibility that 22G-RNA biogenesis is initiated at the 3´ end of most targets, and the wild-type activity of *drh-3* promotes the propagation of 22G-RNA biogenesis by RdRP along the template RNA.

Distinct Genetic Requirements for 22G-RNA

To characterize the genetic pathways required for the biogenesis of 22G-RNAs, we examined a panel of RNAi-related mutants by northern blot analysis to detect small RNAs derived from representative abundant 22G-RNA loci. In addition, we performed small RNA cloning experiments to identify 22G-RNA populations that are enriched in the germline, soma, or oocyte (Figure III-S3, Tables III-S1 and III-S2). This exercise revealed that most 22G-RNAs are germline expressed and/or maternal (Figure III-S3). As expected from our deep-sequencing experiments, we found that 22G-RNAs enriched in both the soma (*Y47H10A.5*) and germline (*F37D6.3* and Tc1) were dependent on DRH-3. Likewise, RDE-3 and MUT-7 were required for 22G-RNAs targeting all three loci (Figure III-3A), indicating that at least one aspect of the mechanism of 22G-RNA expression is shared by these somatic and germline loci.

Previous work has shown that some somatic 22G-RNA loci (e.g., *K02E2.6* and X-cluster) are dependent on the ERI endo- RNAi complex as well as multiple

Figure III-3

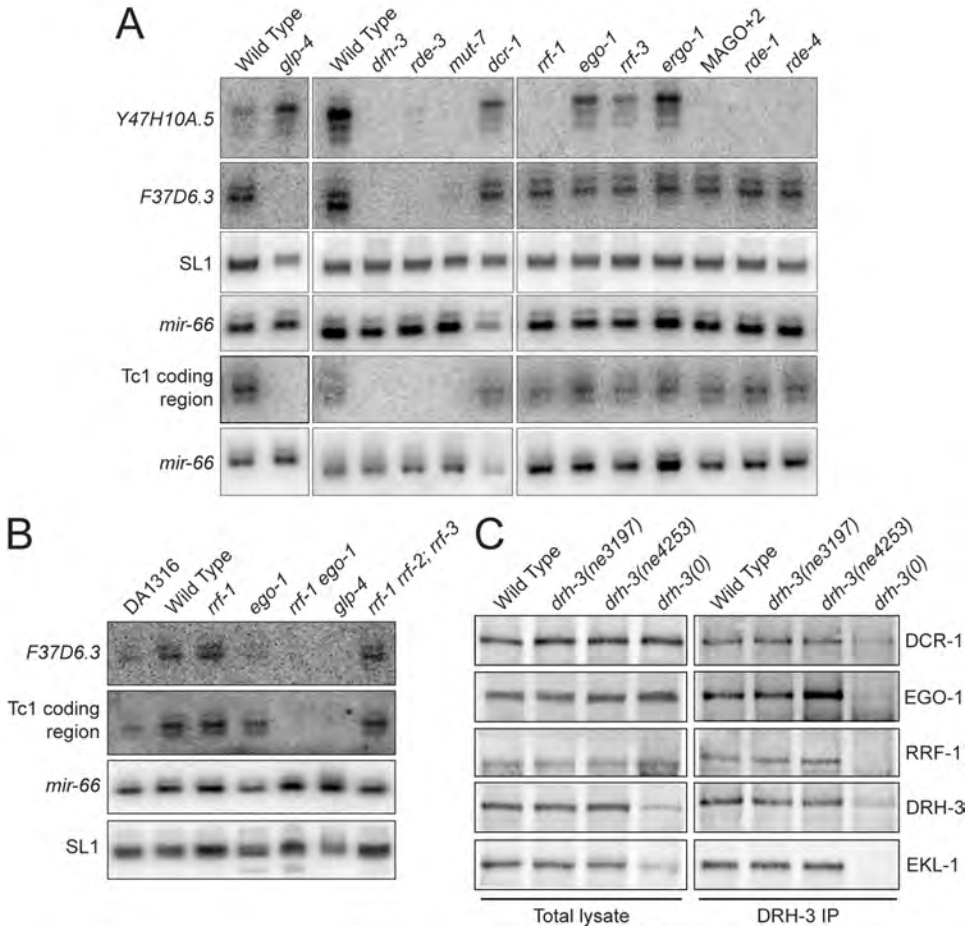
A
Wild Type
glp-4
Wild Type
drh-3
rde-3
mut-7
dcr-1
rrf-1
ego-1
rrf-3
ergo-1
MAGO+2
rde-1
rde-4
Y47H10A.5
F37D6.3
SL1
mir-66
Tc1 coding region
mir-66
B
DA1316
Wild Type
rrf-1
ego-1
rrf-1 ego-1
glp-4
rrf-1 rrf-2; rrf-3
F37D6.3
Tc1 coding region
mir-66
SL1
C
Wild Type
drh-3(ne3197)
drh-3(ne4253)
drh-3(0)
Wild Type
drh-3(ne3197)
drh-3(ne4253)
drh-3(0)
DCR-1
EGO-1
RRF-1
DRH-3
EKL-1
Total lysate
DRH-3 IP

Figure III-3. Distinct requirements and genetic redundancy in 22G-RNA pathways.

(A) Northern blots of 22G-RNA expression in somatic gene *Y47H10A.5*, and germline targets *F37D6.3* and Tc1 transposon in RNAi mutants. Loading controls: SL1 precursor and *mir-66*.

B) Northern blots of germline 22G-RNAs in RdRP mutants. DA1316 is a congenic wild type control for the *ego-1* and *rrf-1 ego-1* mutants. Loading controls: SL1 precursor and *mir-66*.

C) Coimmunoprecipitation analyses of DRH-3 from wild type and *drh-3* lysates. Total lysate (left panels) and DRH-3 immunoprecipitates (right panels) were analyzed by Western blot for candidate interacting proteins (indicated at right). The *drh-3(0)*: the deletion allele (*tm1217*) control. Some DRH-3 protein detected in the *drh-3(0)* sample likely represents a persistence of maternal product.

Argonaute proteins that interact with secondary siRNAs produced by RdRP (Duchaine et al., 2006; Yigit et al., 2006). However, somatic 22G-RNAs derived from *Y47H10A.5* (Figure III-3A) were independent of the ERI pathway genes *rrf-3* and *ergo-1*. Instead, *Y47H10A.5* 22GRNAs fail to accumulate in exogenous (exo-) RNAi pathway mutants, including *rde-4, rde-1,* and *rrf-1* as well as a derivative of the multiple Argonaute mutant MAGO (Yigit et al., 2006) with two additional Argonaute mutations (MAGO+2), suggesting that *Y47H10A.5* 22G-RNAs could be triggered by dsRNA.

Germline 22G-RNAs appear to be independent of the exo- RNAi and ERI pathways. For example, mutations in *dcr-1*, *rde-4*, or *ergo-1* caused no visible depletion in 22G-RNA populations based on ethidium bromide staining of small RNA (Figure III-2A) and were not required for the production of *F37D6.3* or Tc1 22G-RNAs (Figures III-3A and III-S4). These findings suggest that 22G-RNA biogenesis is not triggered by dsRNA at these and many other targets. Despite the biochemical evidence indicating that germline 22G-RNAs are the products of RdRP, we observed near-wild-type levels of 22G-RNAs in each of the individual RdRP mutants as well as in the MAGO+2 mutant (Figures III-3A and III-3B), suggesting additional redundancy within the respective RdRP and Argonaute families of proteins.

DRH-3, the Tudor-Domain Protein EKL-1, and Two Partially Redundant RdRPs Form a Core Complex Essential for 22G-RNA Biogenesis

RRF-1 is required for RNAi in somatic tissues (Sijen et al., 2007), while EGO-1 is required for fertility and has been implicated in RNAi targeting some but not all germline-expressed genes (Smardon et al., 2000). The latter finding could be explained if RRF-1 is functionally redundant with EGO-1 and is expressed within an overlapping domain in the germline. This might also account for the persistence of germline 22G-RNA expression in the RdRP single mutants analyzed (Figure III-3). A noncomplementation screen to generate the *rrf-1 ego-1* double mutant yielded a rearrangement, *neC1*, that disrupts *rrf-1* and results in a putative null allele of *rrf-1* linked to the *ego-1(om97)* nonsense allele (Figure III-S5). Northern blot analyses of small RNAs prepared from the *rrf-1(neC1) ego-1(om97)* double mutant revealed that germline 22G-RNAs fail to accumulate in animals null for both *rrf-1* and *ego-1* (Figure III-3B), demonstrating that RRF-1 and EGO-1 function redundantly in the germline to produce 22G-RNAs.

Consistent with the overlapping functions of RRF-1 and EGO-1 for germline 22G-RNA biogenesis, both RRF-1 and EGO-1 interacted with DRH-3 in immunoprecipitation (IP) experiments (Figure III-3C). A recent study demonstrated that DRH-3 interacts with RRF-1 and is required for RdRP activity in vitro (Aoki et al., 2007). We previously identified DRH-3 as a component of the ERI complex, which includes the tudor-domain protein ERI-5 (Duchaine et al., 2006). EKL-1 is a close homolog of ERI-5 and is required for fertility, RNAi, and chromosome segregation (Claycomb et al., 2009). Consistent with the

phenotypic similarities between *drh-3, ego-1, and ekl-1* mutants (Duchaine et al., 2006; Rocheleau et al., 2008; Vought et al., 2005), EKL-1 also interacted with DRH-3 (Figure III-3C). In addition, both EGO-1 and EKL-1 were among the most enriched proteins in DRH-3 IPs as assessed by multidimensional protein identification technology (MudPIT) (Figure III-S6). Although DCR-1 was detected in the DRH-3 IP by western blot, DCR-1 peptides were not identified in DRH-3 IP-MudPIT experiments. Combined with our deep-sequencing data (below), these data suggest that DRH-3, EKL-1, and RdRP form a core RdRP complex that is essential for the biogenesis of 22G-RNAs in *C. elegans*.

WAGO Family of Argonautes Function Redundantly and Interact with Germline 22G-RNAs

Previously, we demonstrated that the accumulation of secondary siRNAs generated by RdRP is dependent upon multiple redundant WAGO proteins. The previously described MAGO mutant was strongly defective for RNAi and failed to accumulate certain endo-siRNAs (Yigit et al., 2006). However, the MAGO+2 mutant, which includes *ppw-2(tm1120)* and *C04F12.1(tm1637)* as well as the MAGO mutations (Yigit et al., 2006), continued to express normal levels of germline 22G-RNAs targeting *F37D6.3* and Tc1 (Figure III-3A), indicating that additional Argonautes interact with germline 22G-RNAs.

Therefore, we generated additional combinations of mutants within the WAGO clade (Figure III-4A). This analysis identified other WAGO mutant

combinations with germline RNAi defects. For example, *ppw-2(tm1120); f58g1.1(tm1019)* double mutants were resistant to *pos-1*(RNAi), whereas the individual alleles are sensitive to *pos-1*(RNAi) (Yigit et al., 2006). A mutant lacking four of the branch III WAGOs, including *ppw-2(tm1120)*, *F55A12.1(tm2686), F58G1.1(tm1019),* and *ZK1248.7(tm1113),* was resistant to germline RNAi (data not shown) but still produced normal levels of *F37D6.3* germline 22G-RNAs (Figure III-4B, Quadruple). Deletion of the branch III WAGO, *wago-1(tm1414),* resulted in a Quintuple Argonaute mutant with dramatically reduced germline 22G-RNAs (Figures III-4B and III-4C). Furthermore, the *wago-1(tm1414)* mutant alone showed a reduction in *F37D6.3* germline 22G-RNAs that was comparable to the Quintuple Argonaute mutant (Figure II-4C), demonstrating that WAGO-1 plays a key role in germline 22G-RNA function. Transgenic lines expressing a GFP::WAGO-1 fusion under the control of the *wago-1* promoter revealed that WAGO-1 is expressed in the germline and localizes to perinuclear foci that resemble P granules (Figures III-4D and III-4E).

Finally, we generated a strain lacking all 12 of the WAGO genes (not including predicted pseudogenes). This duodecuple mutant (MAGO12) is viable and resistant to RNAi and exhibits a high frequency of spontaneous males and temperature-dependent sterility at 25°C. Germline 22G-RNAs were undetectable by northern blot analysis in the MAGO12 strain (Figure III-4C), demonstrating a clear dependence of the 22G-RNAs on WAGOs.

Figure III-4

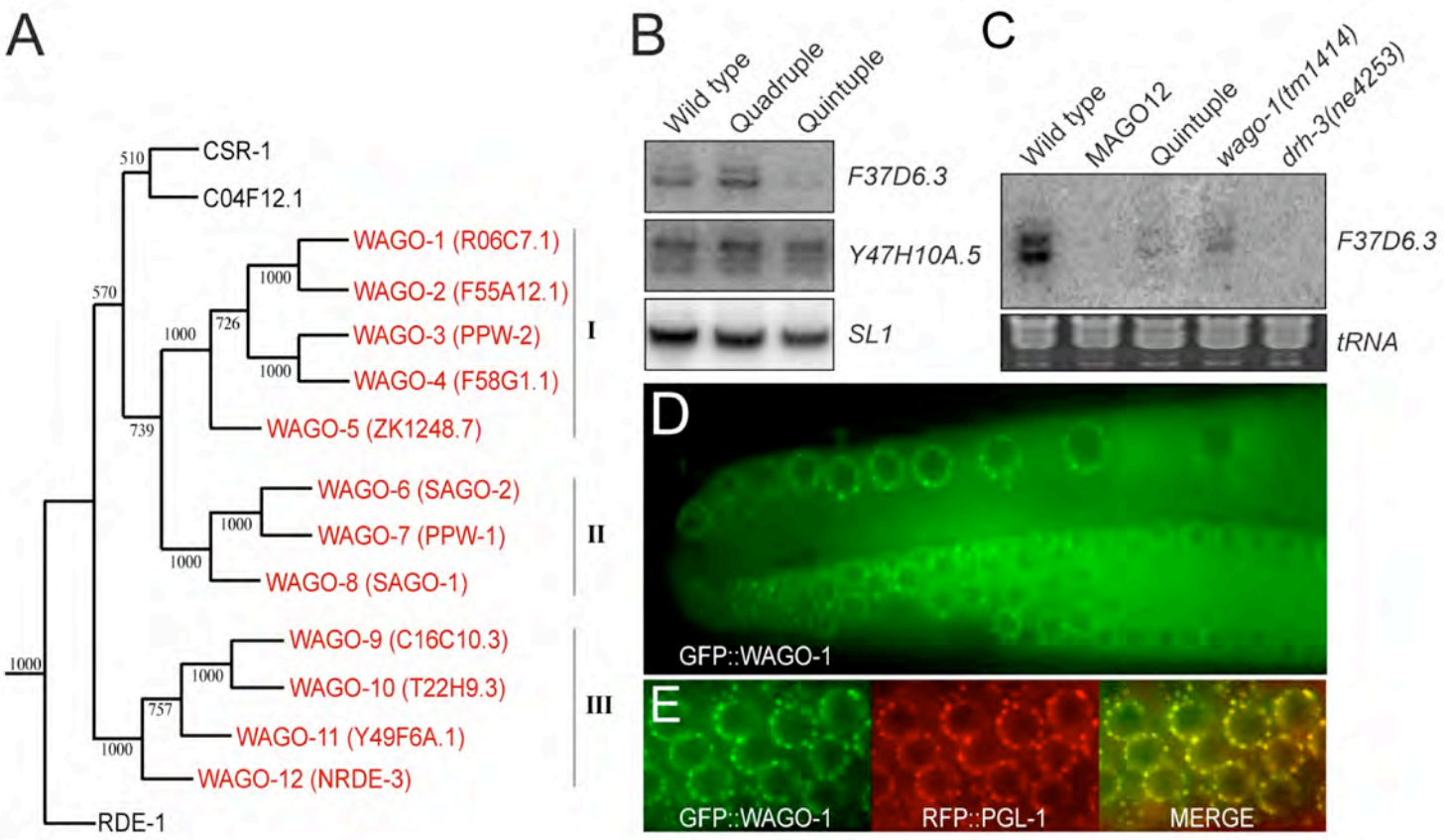
A
510
CSR-1
C04F12.1
WAGO-1 (R06C7.1)
1000
WAGO-2 (F55A12.1)
570
1000
726
WAGO-3 (PPW-2)
I
1000
WAGO-4 (F58G1.1)
739
WAGO-5 (ZK1248.7)
WAGO-6 (SAGO-2)
1000
WAGO-7 (PPW-1)
II
1000
WAGO-8 (SAGO-1)
1000
WAGO-9 (C16C10.3)
1000
WAGO-10 (T22H9.3)
757
WAGO-11 (Y49F6A.1)
III
1000
WAGO-12 (NRDE-3)
RDE-1
B
Wild type
Quadruple
Quintuple
F37D6.3
Y47H10A.5
SL1
C
Wild type
MAGO12
Quintuple
wago-1(tm1414)
drh-3(ne4253)
F37D6.3
tRNA
D
GFP::WAGO-1
E
GFP::WAGO-1
RFP::PGL-1
MERGE

Figure III-4. WAGO-1 and highly redundant WAGOs required for germline 22G-RNA biogenesis.

(A) Phylogenetic representation of the WAGO proteins (see Supplementary Methods). WAGOs deleted in MAGO12 are indicated in red.

(B), (C) Northern blots of *F37D6.3* 22G-RNAs in multiple WAGO mutants. Loading control: SL1 precursor (B) and tRNA staining (C). D), E) Fluorescence microscopy of GFP::WAGO-1 and RFP::PGL-1, which colocalize to P granules in the germline.

WAGOs, RDE-3, and MUT-7 Are Required for Germline 22G-RNA Silencing Pathways

To gain insight into the function of germline 22G-RNAs, we performed deep sequencing of small RNA populations from mutants with germline 22G-RNA defects. In addition, we generated transgenic animals that express a 3X FLAG::WAGO-1 fusion protein and deep sequenced the small RNAs that coprecipitate with WAGO-1. For each mutant, the fraction of reads matching coding genes, nonannotated loci, and repeat elements was reduced, with concomitant increases in the fraction of miRNA and 21U-RNA reads (Figure III-S7). Conversely, the fraction of reads matching to a particular set of coding genes, nonannotated loci, and repeat elements were enriched in the small RNA library prepared from the WAGO-1 immunoprecipitate, while miRNAs and 21U-RNAs were severely depleted (Figure III-5A).

We next asked whether the reduction of 22G-RNAs in each mutant occurred globally or at particular loci. 22G-RNAs targeting protein-coding loci were nearly completely eliminated in the *drh-3* and *ekl-1* single mutants and in the *rrf-1 ego-1* double mutant. Gene-targeted 22G-RNAs were far less likely to be depleted in the *rde-3*, *mut-7*, and MAGO12 mutant samples (Figure III-5B), with the notable exception that 22G-RNA species targeting a subset of genes with normally very high 22G-RNA levels were strongly depleted in each of these

Figure III-5

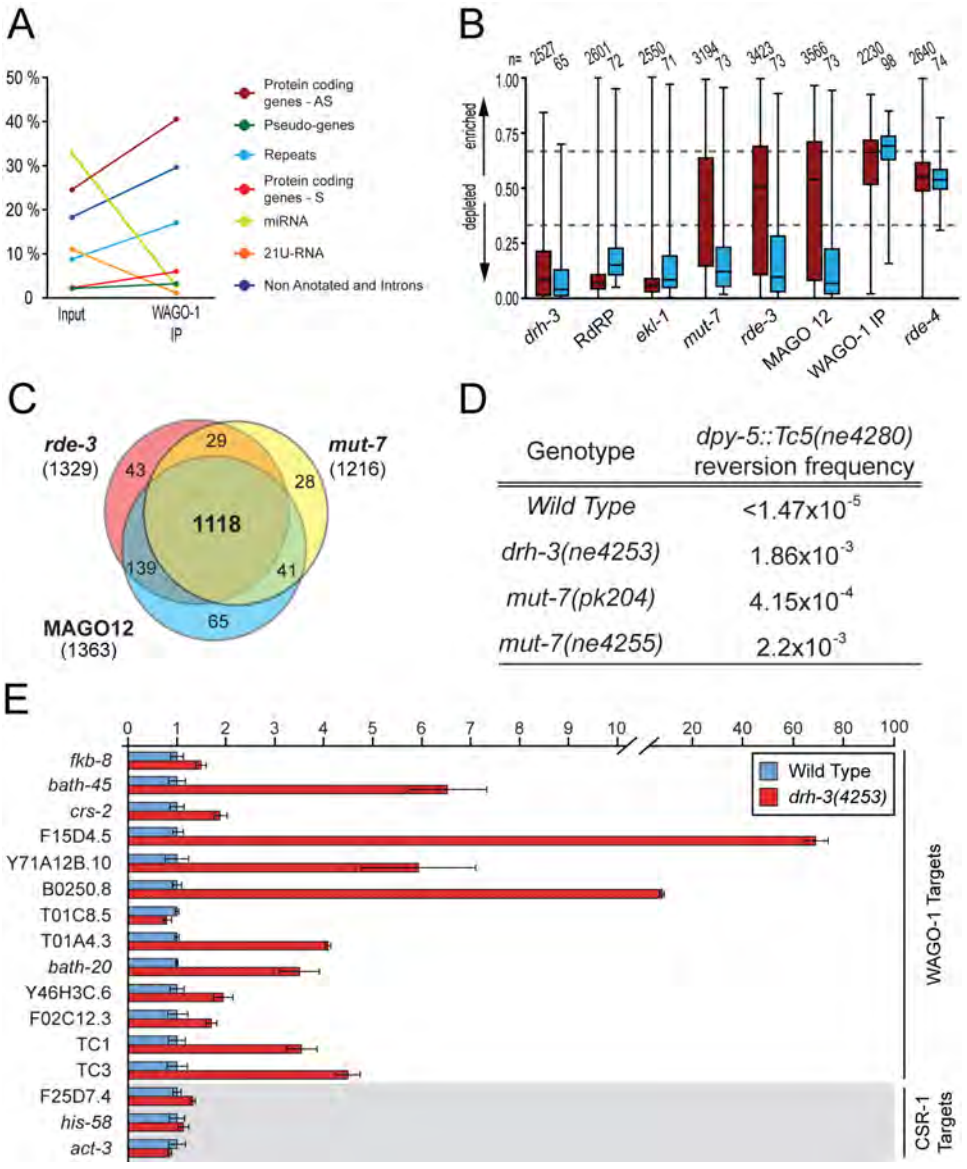
A
50 %
40 %
30 %
20 %
10 %
0
Input
WAGO-1 IP
Protein coding genes - AS
Pseudo-genes
Repeats
Protein coding genes - S
miRNA
21U-RNA
Non Anotated and Introns
B
n= 2527 65 2601 72 2550 71 3194 73 3423 73 3566 73 2230 98 2640 74
1.00
0.75
0.50
0.25
0.00
enriched
depleted
drh-3
RdRP
ekl-1
mut-7
rde-3
MAGO 12
WAGO-1 IP
rde-4
C
rde-3
(1329)
mut-7
(1216)
MAGO12
(1363)
29
43
28
1118
139
41
65
D
Genotype
dpy-5::Tc5(ne4280) reversion frequency
Wild Type
<1.47x10-5
drh-3(ne4253)
1.86x10-3
mut-7(pk204)
4.15x10-4
mut-7(ne4255)
2.2x10-3
E
0 1 2 3 4 5 6 7 8 9 10 20 40 60 80 100
Wild Type
drh-3(4253)
fkb-8
bath-45
crs-2
F15D4.5
Y71A12B.10
B0250.8
T01C8.5
T01A4.3
bath-20
Y46H3C.6
F02C12.3
TC1
TC3
F25D7.4
his-58
act-3
WAGO-1 Targets
CSR-1 Targets

Figure III-5. Deep-sequence analyses identified at least two distinct 22G-RNA pathways in the germline.

A) Graph depicting change in reads matching the indicated genome annotations between input and WAGO-1 IP samples.

B) Change of 22G-RNAs derived from genes (red) or transposons (blue) in each mutant. Relative enrichment calculated as the ratio of mutant / (mutant + wild type) for 'n' genes or transposons. WAGO-1 IP enrichment calculated as WAGO-1IP / (WAGO-1IP + IP input).

C) Venn comparison of genes depleted of 22G-RNAs (≥ 2-fold) in indicated mutants (loci below the lower dashed line in (B).

D) The frequency of reversion of *dpy-5::Tc5* as indicated (Ketting et al., 1999).

E) Quantitative RT-PCR analysis of 22G-RNA target expression in *drh-3* mutant (red) relative to wild-type (blue). WAGO and CSR-1 (gray) targets are indicated

mutants (Figure III-S8). In contrast, 22G-RNAs were largely unaffected in an *rde-4* mutant, which is required for ERI-class small RNAs (Duchaine et al., 2006); (Lee et al., 2006); J.V., W.G., and C.C.M., unpublished data). The 22G-RNAs depleted in the *rde-3*, *mut-7*, and MAGO12 mutants were almost completely overlapping (Figure III-5C and Table III-S3). Despite an overall reduction of 22GRNA reads, a subset of 22G-RNA species was not depleted in *rde-3*, *mut-7*, and MAGO12 mutants (Figures III-5B and III-S8); in fact, some were increased in proportion. These WAGO-independent 22G-RNA populations are associated with and dependent upon another germline-expressed Argonaute, CSR-1 (Figure III-S9) (Claycomb et al., 2009). The bimodal distribution of 22G-RNA loci indicates that at least two qualitatively distinct 22G-RNA pathways exist in the germline that depend on a core set of factors (DRH-3, EKL-1, and RdRP) whose small-RNA products interact with distinct Argonautes.

Consistent with the requirement for WAGO-pathway components in exo-RNAi, the WAGO-associated 22G-RNAs appear to be involved in silencing their respective targets. Loci with the highest levels of 22G-RNAs in wild-type were consistently derepressed in the *drh-3* mutant, as assessed by semiquantitative PCR with reverse transcription (qRT-PCR) (Figure III-5E) and Affymetrix tiling arrays (Figure III-S10). In contrast, CSR-1-associated 22G-RNAs do not appear to silence their targets (Claycomb et al., 2009), consistent with the biological distinction between these pathways.

Previous work has shown that Tc elements are silenced in the germline by

an RNAi mechanism (Ketting et al., 1999; Sijen and Plasterk, 2003; Tabara et al., 1999). Individual Repbase annotations, which include all major classes of transposons in *C. elegans*, were uniformly depleted of 22G-RNAs in *drh-3, ekl-1,* RdRP, *rde-3, mut-7*, and MAGO12 mutant samples, and most transposon 22G-RNAs were enriched in the WAGO-1 IP sample (Figure III-5B). Transposon loci showed normal levels of 22G-RNAs in an *rde-4* mutant sample (Figure III-5B) (Tabara et al., 1999). Thus, the transposon-silencing pathway in *C. elegans* consists of DRH-3, EKL-1, RdRPs, RDE-3, MUT-7, and multiple WAGOs, including WAGO-1.

The *drh-3* alleles described here display the hallmarks of Mut class, Rde mutants (Figure III-1). Indeed, spontaneous mutants with phenotypes that revert at high frequency were cloned from the *drh-3* mutants, including a dpy-5::Tc5 insertion. The frequency of reversion from Dumpy to wild-type, upon excision of Tc5 from *dpy-5* in *drh-3(ne4253),* was similar to an allele of *mut-7(ne4255)* that was isolated in the same screen and almost 5-fold higher than the nonsense allele *mut-7(pk204)* (Figure III-5D). Similar results were obtained with an unc-22::Tc1 insertion. Furthermore, Tc1 and Tc3 transcripts were derepressed in the *drh-3* mutant (Figure III-5E), demonstrating that DRH-3 is required for transposon silencing.

22G-RNAs and Surveillance

Approximately 15% of 22G-RNA reads were derived from nonannotated loci and

were dependent on RDE-3, MUT-7, and MAGO12. These loci primarily correspond to unique intergenic sequences and could represent pseudogenes or cryptic loci that lack open reading frames and are unrecognizable by current bioinformatic approaches. In some cases, we could predict potential splicing patterns based on antisense reads spanning these nonannotated regions (Figure III-S11A). Consistent with this notion, 22G-RNAs derived from many loci annotated as pseudogenes were also depleted in *rde-3*, *mut-7*, and MAGO12 mutants (Figure III-6A). As with annotated genes targeted by WAGO-associated 22G-RNAs, qRT-PCR and microarray analysis demonstrated that both pseudogenes and cryptic loci targeted by 22G-RNAs were desilenced in the *drh-3* mutant (Figures III-6B and III-S12).

Upon closer inspection of protein-coding loci targeted by the WAGO pathway, we noted that the 22G-RNA profile often did not correspond to the annotated gene prediction (Figure III-S11B and III-Table S4). In many cases, 22G-RNAs mapped within predicted introns, suggesting that the corresponding introns were not spliced in the target RNA. In other cases, 22G-RNAs started or ended abruptly in the middle of the annotation and extended well upstream or downstream of the gene prediction, suggesting that the annotation is incomplete or incorrect. Lastly, we noticed a number of WAGO target genes with intron

Figure III-6

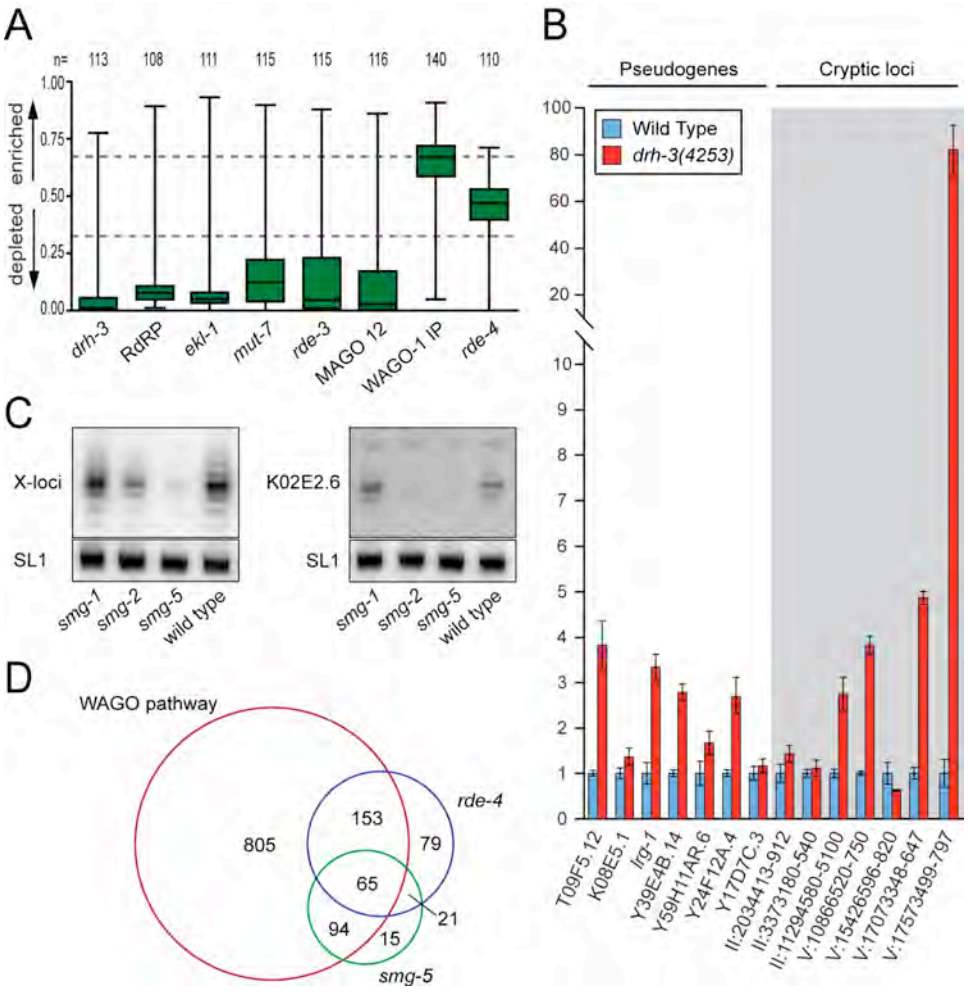
A
n= 113 108 111 115 115 116 140 110
enriched
depleted
drh-3
RdRP
eki-1
mut-7
rde-3
MAGO 12
WAGO-1 IP
rde-4
B
Pseudogenes
Cryptic loci
Wild Type
drh-3(4253)
T09F5.12
K08E5.1
lrg-1
Y39E4B.14
Y59H11AR.6
Y24F12A.4
Y17D7C.3
II:2034413-912
II:3373180-540
II:11294580-5100
V:10866520-750
V:15426596-820
V:17073348-647
V:17573499-797
C
X-loci
SL1
K02E2.6
smg-1
smg-2
smg-5
wild type
D
WAGO pathway
rde-4
smg-5
805
153
79
65
21
94
15

Figure III-6. WAGO-associated 22G-RNAs define a surveillance system.

A) Enrichment or depletion of 22G-RNAs derived from ‘n’ annotated pseudogenes.

B) Quantitative RT-PCR analysis of pseudogenes and cryptic loci targeted by WAGO pathway in *drh-3* mutant (red) relative to wild-type (blue).

C) Northern blots of 22G-RNAs in *smg* mutants grown at 20°C. Similar results were obtained with mutants grown at 25°C (not shown). Loading control: SL1 precursor.

D) Venn comparison of genes depleted of 22G-RNAs (≥2-fold) in mutants.

annotations in 3´ UTRs. Because pseudogenes and genes with 3´ UTR introns are expected to be targets of the NMD pathway, we asked whether 22G-RNA biogenesis was dependent on the PIN domain protein SMG-5, the Upf1 helicase SMG-2, and the phosphatidylinositolkinase SMG-1 (Anders et al., 2003; Glavan et al., 2006; Grimson et al., 2004; Page et al., 1999). 22G-RNAs derived from X-cluster and *K02E2.6*, which has a 3´ UTR intron, were reduced in the null mutant *smg-5(r860)* and to a lesser extent in the non-null *smg-2(r863)* (Figure III-6C). *K02E2.6* and X-cluster 22G-RNAs were unchanged in the temperature-sensitive mutant *smg-1(cc546)* at both permissive and nonpermissive temperatures (Figure III-6C). These data suggest a role for SMG-2 and SMG-5 in 22G-RNA biogenesis that is distinct from their recognized role in NMD and that NMD per se is not required for 22G-RNA biogenesis.

Deep-sequence analysis of *smg-5* mutant small RNAs revealed that SMG-5 is required for the biogenesis of 22G-RNAs targeting 15% of WAGO-dependent 22G-RNA target genes (Figure III-6C). Interestingly, SMG-5 was not required for most pseudogene-derived 22G-RNAs. Furthermore, the few published endogenous targets of NMD do not appear to be 22G-RNA targets. Roughly half of the SMG- 5-dependent 22G-RNA loci overlap with RDE-4-dependent 22G-RNA loci (Figure III-6D), which includes both ERI-dependent and ERI-independent 22G-RNA loci (J.V., W.G., and C.C.M., unpublished data). These findings indicate that multiple WAGO dependent 22G-RNA pathways exist, which together define a general surveillance system that silences transposons and

aberrant transcripts.

DISCUSSION

In this study, we have combined deep sequencing with the powerful genetics of *C. elegans* to identify and characterize an abundant class of endo-siRNAs that we call 22G-RNAs. In adult animals, 22G-RNAs are primarily germline expressed and are derived from unique sequences in the genome, including coding genes, transposons, pseudogenes, and nonannotated loci. Combining data from three small RNA libraries, including an Argonaute IP sample, we have identified 22G-RNAs antisense to over 50% of the annotated protein-coding genes.

DRH-3, EKL-1, and the partially redundant RdRPs, RRF-1 and EGO-1, form a core RdRP complex that functions in multiple 22G-RNA pathways (Figure III-7). RDE-3, MUT-7, and members of the WAGO clade, in particular WAGO-1, define a general 22G-RNA surveillance system that silences transposable elements and aberrant transcripts. A second pathway is dependent on CSR-1, which promotes kinetochore structure and chromosome segregation (Figure III-7) (Claycomb et al., 2009). Taken together, 22G-RNAs appear to engage targets that derive from both the actively expressed regions (CSR-1 associated), as well as the ‘‘silent’’ regions of the genome (WAGO-1 associated). These findings support a model in which the 22G-RNA pathways exert genome-scale surveillance important for maintenance of the germline.

Figure III-7

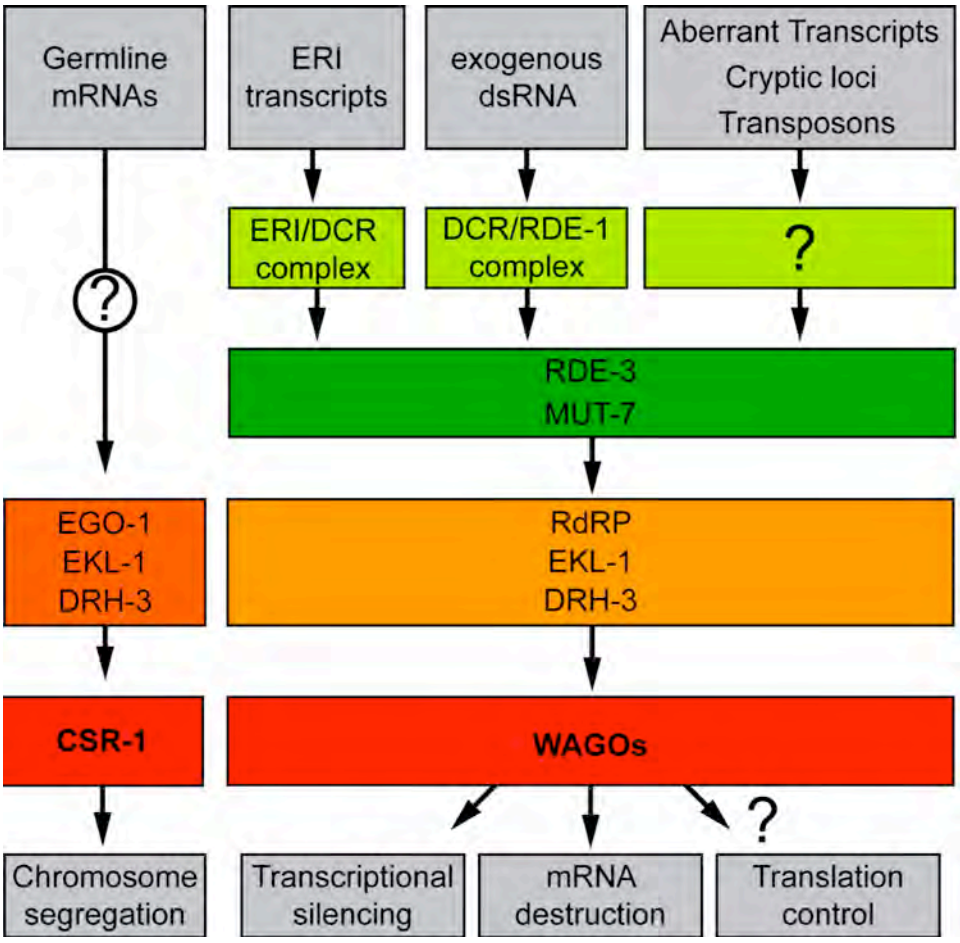

Figure III-7. Model of Germlline 22G-RNA pathways required for genome surveillance in *C. elegans*

Dicer-Dependent versus Independent 22G-RNA Biogenesis

It is formally possible that maternal DCR-1 is sufficient to generate primary siRNAs and that subsequent recruitment of the secondary RNAi machinery, dependent on RdRPs and WAGOs, results in a self-sustaining amplification cycle to produce 22G-RNAs in *dcr-1* zygotic mutants. However, extensive northern blot experiments demonstrate that ERI-dependent 22G-RNAs and miRNAs are depleted in the *dcr-1* mutant despite the maternal contribution of DCR-1, while germline 22G-RNAs are present at normal or, in some cases, elevated levels (Figure III-S4). Aside from miRNA and exo- RNAi complexes, the ERI complex appears to be the primary DCR-1 complex in *C. elegans* (Duchaine et al., 2006). We have shown that 22G-RNAs are largely independent of RDE-4, a DCR-1 cofactor in both exo-RNAi and ERI pathways (Duchaine et al., 2006; Lee et al., 2006; Tabara et al., 2002). Thus, we favor the model that the major 22G-RNA pathways are initiated in a DCR-1-independent fashion. New alleles of *dcr-1* will be important in order to resolve this issue in the future.

The Role of DRH-3 in 22G-RNA Biogenesis

How might Dicer-independent 22G-RNA biogenesis be triggered, and what role does DRH-3 play in this process? A recent report has shown that *drh-3* null mutant extracts are deficient in the in vitro synthesis of antisense small RNA by RRF-1 (Aoki et al., 2007). Our analysis of 22G-RNA levels in hypomorphic *drh-3* point-mutant alleles is consistent with a role for the conserved DRH-3 helicase

domain in 22G-RNA biogenesis. Two of the three *drh-3* missense alleles alter highly conserved residues within the HELICc domain. *The drh-3(ne4253)* lesion (T834M) alters a residue that contacts RNA in the Vasa crystal structure; the *drh-3(ne3197)* lesion (G840D) alters a residue that maps to the interface between the HELICc and DExH domains and coordinates the water molecule that is thought to be required for ATP hydrolysis (Sengoku et al., 2006). Both lesions are likely to abrogate ATPase and/or unwinding activity based on structural and biochemical studies with related proteins (Liang et al., 1994; Sengoku et al., 2006).

The enrichment of 22G-RNAs at the 3´ end of transcripts suggests that 22G-RNA biogenesis begins at the 3´ end of target RNAs, followed by cycles of 22G-RNA synthesis by RdRP and proceeding along the template toward the 5´ end. Interestingly, the 3´ localized 22G-RNAs are least diminished in the *drh-3* mutant, suggesting that these lesions do not prevent the initial loading of RdRP onto the template but rather interfere with the processivity of RdRP. DRH-3 could remove secondary structure from the template or facilitate transfer of the 22G-RNA to downstream WAGOs, allowing RdRP to initiate a second round of synthesis at the next available C residue in the template RNA. Whatever the mechanism by which DRH-3 promotes 22G-RNA biogenesis, our data support the idea that RdRP is recruited to the 3´ end of target transcripts. Templates lacking a poly(A) tail are better substrates for RdRP in vitro (Aoki et al., 2007) suggesting that defective 3´ end formation may be one trigger for 22G-RNA

biogenesis. In *Arabidopsis*, decapped, misspliced, and misterminated transcripts are recognized by, and activate, the RNA-silencing machinery (Gazzani et al., 2004; Herr et al., 2006). In fission yeast, two ß-nucleotidyl transferases, Cid12 and Cid14, determine whether transcripts are recognized by the RdRP complex (Cid12) or the TRAMP/exosome complex (Cid14) (Buhler et al., 2007; Buhler et al., 2008; Motamedi et al., 2004). In *cid14* mutants, transcripts that are normally turned over by the TRAMP surveillance pathway become substrates for the RdRP complex (Buhler et al., 2008), indicating that these pathways recognize a common feature. Perhaps the ß-nucleotidyl transferase RDE-3 and the 3´-to-5´ exonuclease MUT-7 function in an exosome-like pathway that recognizes and processes the 3´ end of aberrant transcripts, providing a signal that recruits the RdRP complex (Figure III-7) (Chen et al., 2005; Ketting et al., 1999; Lee et al., 2006; Tabara et al., 1999).

Multiple Distinct 22G-RNA Pathways

Based on studies in other model systems, we expected a significant fraction of germline 22G-RNAs to be derived from transposons. Indeed, transposable elements are targets of 22G-RNAs. However, 22G-RNAs derived from unique sequences, both genic and intergenic, comprise a major fraction of the 22G-RNAs that interact with WAGO-1. Furthermore, loci that produce the highest levels of 22G-RNAs appear to interact with WAGO-1 and are the most desilenced in the *drh-3* mutant. Remarkably, CSR-1 interacts with a

nonoverlapping population of 22G-RNAs derived almost exclusively from protein-coding loci (Figure III-S9) (Claycomb et al., 2009). Loci targeted by CSR-1 produce fewer 22G-RNAs than WAGO-1 loci and are not desilenced in either *csr-1* or *drh-3* mutants (Claycomb et al., 2009). These findings indicate that these pathways are mechanistically or functionally distinct (Figure III-7), a conclusion that is consistent with the genetically defined functions of CSR-1 and WAGO-1. Although both pathways seem to be important for an efficient response to foreign dsRNA (Claycomb et al., 2009; Yigit et al., 2006), it seems likely that the WAGO surveillance system is primarily involved, as WAGOs were shown to be limiting for RNAi and to interact directly with the secondary 22G-RNAs in the amplification cycle (Yigit et al., 2006). *csr-1* mutants disrupt the perinuclear localization of the germline nuage (Claycomb et al., 2009) wherein WAGO-1 resides and hence could indirectly affect WAGO-1 function by interfering with its proper localization within these germline structures. Further genetic and biochemical studies will be necessary to dissect the relative contributions of the WAGO-1 and CSR-1 pathways to RNAi.

At least some of the specificity of CSR-1 and WAGO 22G-RNA pathways can be attributed to the involvement of distinct RdRP complexes. We have uncovered a role for RRF-1 in the germline, where it is redundant with EGO-1 in a surveillance pathway that regulates transposons, pseudogenes, and cryptic loci. However, EGO-1 alone is required for the 22G-RNAs that associate with CSR-1 (Claycomb et al., 2009). In addition, van Wolfswinkel et al. (van

Wolfswinkel et al., 2009) implicate the ß-nucleotidyl transferase CDE-1 as a specificity factor for EGO-1 in the chromosome segregation pathway, as EGO-1 can also function with RDE-3 (a CDE-1 homolog) for the biogenesis of WAGO-associated 22G-RNAs (Figure III-7).

Despite clear genetic redundancy among the WAGOs, we expect that individual WAGOs normally function in distinct pathways (Figure III-7). Consistent with this idea, RDE-4 and SMG-5 are required for the biogenesis of distinct and overlapping subsets of WAGO-dependent 22G-RNAs. Domeier et al. (Domeier et al., 2000) showed that the exo-RNAi response is short-lived in *smg-2, -5, and -6* mutants. Our findings extend their work and provide molecular insight into the link between the NMD pathway and RNAi. Both studies connect 22G-RNA biogenesis to the translation apparatus and suggest that an alternative branch of the NMD pathway exists (Behm-Ansmant et al., 2007). Perhaps SMG proteins recognize a particular characteristic of 22G-RNA target transcripts and recruit the RdRP machinery. Our findings indicate that the signal is unlikely to be premature termination codons. Alternatively, NMD components could function as WAGO cofactors. WAGOs lack the catalytic residues important for Slicer activity and are not expected to cleave a target in vivo, suggesting that alternative turnover mechanisms are involved in silencing. It is interesting to note that the PIN domains of both SMG-5 and SMG-6 are structurally related to RNase H (Glavan et al., 2006), but only SMG-6 retains the catalytic residues important for the endonucleolytic cleavage that initiates NMD (Eberle et al., 2009; Huntzinger et

al., 2008). If WAGOs do indeed lack catalytic activity, perhaps SMG-6 could provide the endonuclease activity that reinforces the 22G-RNA amplification cycle for this set of targets. Additional complexity and bifurcation/convergence of WAGO-dependent 22G-RNA pathways is likely to emerge as we identify new factors required for 22G-RNA biogenesis.

The Mutator phenotypes of *mut-7, rde-3, and drh-3* could, in part, result from defects in WAGO-dependent chromatin silencing (III-7Figure 7). WAGO-12/NRDE-3 is a nuclear WAGO required for cotranscriptional silencing (Guang et al., 2008) but not for the accumulation of 22G-RNAs. However, mutants that block the biogenesis of 22G-RNAs prevent nuclear localization of NRDE-3 and exacerbate derepression of NRDE-3 targets, suggesting additional WAGOs are involved in a parallel, posttranscriptional silencing pathway. It will be of interest in the future to dissect the potential role of 22G-RNAs in different chromatin mediated silencing pathways.

Maternal Small RNAs

In several respects, including transposon control and DCR-1- independent small RNA biogenesis, the WAGO 22G-RNA system is analogous to the Drosophila and vertebrate piRNA pathways. Furthermore, we have shown that the WAGO and CSR-1 22G-RNA systems are maternal and that factors involved in these pathways localize to germline P granules (Claycomb et al., 2009), which are thought to function in the repression and storage of maternal mRNAs (Rajyaguru

and Parker, 2009). This appears to be a common feature of germline small RNA pathways in animals, as PIWI family members also localize to P granules or nuage (Batista et al., 2008; Li et al., 2009; Malone et al., 2009). In each case, the localization of Argonaute proteins to P granules appears to be dependent on small RNA biogenesis. For the CSR-1 pathway, the P granule structure itself seems to be dependent on small RNA biogenesis (Claycomb et al., 2009). The close association between cytoplasmic P granules and nuclear pores (Pitt et al., 2000) would allow Argonaute systems to survey the entire transcriptome as RNAs exit the nucleus and enter the P granule, reinforcing both the biogenesis and regulatory functions of small RNAs. Maternal small RNAs function in a number of epigenetic programs, from transposon silencing (Brennecke et al., 2007; Tam et al., 2008; Watanabe et al., 2008) to imprinting (Davis et al., 2005) to the maternal-zygotic transition (Giraldez et al., 2006; Lykke-Andersen et al., 2008). Our findings suggest that 22G-RNAs in *C. elegans* mirror the expression of many germline-expressed RNAs, including those destined for expression as well as silencing. Thus, 22G-RNAs and their Argonaute partners provide versatile regulators of both physical and epigenetic inheritance.

MATERIALS AND METHODS

Worm Strains

C. elegans culture and genetics were essentially as described (Brenner, 1974).

Strains used in this study:

The Bristol N2 strain was used as the wild-type strain. Alleles used in this study, grouped by chromosome:

LGI: *sago-2(tm894), smg-2(r863), ppw-1(tm914), ppw-2(tm1120), F55A12.1(tm2686), avr-14(ad1302), rde-3(ne3364), smg-1(cc546), smg-5(r860), ekl-1(tm1599), wago-1(tm1414), rrf-1(pk1417,neC1), ego-1(om97), drh-3(tm1217, ne3197, ne4253, ne4254), hT2[qIs48](I;III), glp-4(bn2);*

LGII: *Y49F6A.1*(*tm1127*), *ZK1248.7*(*tm1113*), *rrf-3(pk1426)*, *C06A1.4(tm887*), *F58G1.1*(*tm1019*);

LGIII: *C16C10.3*(*tm1200*), *eft-3(q145)*, *dcr-1(ok247)*, *qC1[neIs(myo-2::avr-15, rol-6(su1006), unc-22(RNAi))]*, *mut-7* (*pk204*, *ne4255*), *rde-4(ne337)*;

LGIV: *fem-1*(*hc17*), *M03D4.6(tm1144*);

LGV: *T22H9.3*(*tm1186*), *ergo-1(tm1860)*, *sago-1*(*tm1195*), *rde-1(ne300)*, *avr-15*(*ad1051*), *glc-1*(*pk54*);

X: *R04A9.2*(*tm1116*).

Strain List:

Strain name	Text name	Genotype
WM187	RdRP	*avr-14(ad1302) rrf-1(neC1) ego-1(om97)/hT2[qIs48]* I*; +/hT2 III; avr-15(ad1051) glc-1(pk54)* V
WM188	MAGO+2	*sago-2(tm0894) ppw-1(tm0914) ppw-2(tm1120) C04F12.1(tm1637)* I*; C06A1.4(tm0887) F58G1.1(tm1059)* II*; M03D4.6(tm1144)* IV*; sago-1(tm1195)* V
WM189	Quadruple	*ppw-2(tm1120) F55A12.1(tm2686) I; ZK1248.7(tm1113) F58G1.1(tm1019)* II
WM190	Quintuple	*ppw-2(tm1120) F55A12.1(tm2686) R06C7.1(tm1414)* I*; ZK1248.7(tm1113) F58G1.1(tm1019)* II
WM191	MAGO12	*sago-2(tm894) ppw-1(tm914) ppw-2(tm1120) F55A12.1(tm2686) R06C7.1(tm1414)* I*; Y49F6.1(tm1127) ZK1248.7(tm1113) F58G1.1(tm1019)* II*; C16C10.3(tm1200) K12B6.1(tm1195)* III*; T22H9.3(tm1186)* V*; R04A9.2(tm1116)* X
WM192	Flag::WAGO-1	*unc-119(ed3)* III; neIs21[*unc-119(+); 3xflag::wago-1*]
WM193	GFP::WAGO-1	*unc-119(ed3)* III; neIs22[*unc-119(+); gfp::wago-1*]
DA1316		*avr-14(ad1302)* I*; avr-15(ad1051) glc-1(pk54)* V

R06C7.1 translational fusions were created as follows: A 4.7kb NheI-SpeI DNA fragment released from cosmid R06C7 was inserted into the SpeI site of pBS-KS+ (Stratagene). A BamHI site was inserted immediately after the ATG of R06C7.1 by site-directed mutagenesis. A fragment containing the GFP ORF from pPD95.75 (gift of A. Fire) or 3xFLAG was inserted into the BamHI site. Finally, a 5.7 Kb EagI fragment containing the wild-type *unc-119* gene was inserted into the NotI site of each plasmid. Microparticle bombardment (Praitis et al., 2001) was used to generate integrated lines expressing GFP::R06C7.1 or 3xFLAG::R06C7.1.

Worm Culture

For large cultures, synchronous animals were grown at 20°C for ~65-70hr and harvested as gravid adults. *FLAG::WAGO-1* animals were grown at 24°C and harvested as young adults. Ooctyes were harvested from *fem-1* (*hc17*) animals (propagated at 25°C for 50hr) as described (Aroian et al., 1997). *glp-4*(*bn2*) were grown at 25^{o}C for 50hr. *drh-3*(*tm1217*)/*hT2*[*qIs48*](I:III) and WM187 were grown on NGM plates with 25 µg/L ivermectin at 20°C for 65-70hr. Animals were washed three times with M9 buffer (22mM KH_2PO_4, 42mM Na_2HPO_4, 85mM NaCl, 1mM $MgSO_4$), precipitated by gravity during each wash to remove *E. coli*, free embryos, and hatched L1 animals, if any, and then incubated in M9 buffer for ~30 minutes to clear bacteria in the gut. Finally, animals were tightly pelleted using cold M9 buffer and either quickly frozen in a dry ice/ethanol bath and stored at -80°C or processed immediately for protein or RNA.

Protein Immunoprecipitation

Proteins were extracted in a stainless steel dounce using lysis buffer containing 25 mM Tris-Cl (pH 7.5), 0.15 M NaCl, 2.5 mM MgCl2, 0.05% Igepal, 1 mM DTT, 1% SUPERase•In (Ambion), 0.4% phosphatase cocktail I (Sigma), 0.4% phosphatase cocktail II (Sigma), and Mini Protease Inhibitor Cocktail (4 Tablets/25 ml buffer, Roche), and cleared at 20,000 g for 10 min at 4°C. Immunoprecipitation of DRH-3 was performed as follows: ~20 mg of protein extract was precleared with 40 µl of Protein A/G PLUS-Agarose beads (Santa Cruz Biotechnologies) for 1hr at 4°C. The cleared extract was then incubated

with 10µl each of N-Ab and C-Ab for 1h at 4°C. Immune complexes were incubated with 50µl of Protein A/G PLUS-Agarose beads (Santa Cruz) at 4°C for 1hr. Precipitated immune complexes were washed three times with ice-cold lysis buffer, and eluted with 2X sample buffer. A 10% sample was used for Western blot analysis.

Western blot analysis

Proteins were resolved by SDS-PAGE on 4-15% acrylamide gradient gels (Biorad) and transferred to Hybond-C Extra membrane (GE Healthcare). Membranes were blocked with 5% nonfat dried milk in PBST (137 mM NaCl, 2.7 mM KCl, 10 mM Phosphate pH 7.4, 0.1% Tween-20) and incubated in the same buffer with primary antibodies at 4°C overnight. Membranes were washed 3 times in PBST, incubated with HRP-conjugated secondary antibody in PBST (1:5,000) at room temperature for 1hr and again washed with PBST. Chemiluminescence was performed using the Western Lightening ECL Kit (Perkin Elmer) and visualized using a CCD camera and LAS-3000 Intelligent Dark-Box (Fujifilm).

MudPIT Analysis of DRH-3 Immunoprecipitations

The proteins eluted from DRH-3 immunoprecipitates were precipitated with 5 volumes of acetone. Protein pellets were washed with 5 volumes acetone and resusupended in digestion buffer (0.1% (w/v) Rapigest (Waters Corporation) in 50 mM Ammonium Bicarbonate). Samples were then reduced with TCEP

(Tris(2-Carboxyethyl)-Phosphine Hydrochloride), alklyated with iodoacetamide and digested with trypsin. Digested peptides were analyzed by LC/LC/MS/MS using an LTQ-Orbitrap mass spectrometer. Multidimensional chromotography was performed online with salt steps of 20%, 50% and 100% (MacCoss et al., 2002). Tandem mass spectra were collected in a data-dependent manner with up to 5 ms2 scans performed for each initial scan (m/z range 400-1800).

The search program Prolucid was used to match data to a database obtained from Wormbase (WP180). Peptide identifications were filtered using DTASelect program (Tabb et al., 2002). The AC fold feature of the program Pattern Lab (Carvalho et al., 2008) was used to compare DRH-3 IPs from wild-type N2 and two *drh-3* missense mutants (*ne4253* and *ne3197*) to a DRH-3 IP from the deletion allele (*tm1217*). Proteins more than 5-fold enrichment in the non-deletion strain and a minimum p-value of 0.0001 were considered significant.

Antibody generation and purification

Antibodies used in this study include: (1) affinity-purified, anti-DRH-3 polyclonal (see Supplementary materials); (2) anti-DCR-1 polyclonal (Duchaine et al., 2006); (3) affinity-purified, anti-EKL-1 ((Claycomb et al., 2009) cosubmitted); (4) affinity-purified, anti-RRF-1 polyclonal (see Supplementary materials); (5) affinity-purified, anti-EGO-1 polyclonal (see Supplementary materials); (6) anti-FLAG M2 monoclonal (Sigma); (7) HRP-conjugated, anti-rabbit IgG secondary antibodies (Jackson Immunoresearch).

Polyclonal sera were raised in rabbits against a mixture of two N-terminal fragments (residues 1-222 and 125-360) and a C-terminal fragment (residues 889-1119) to generate the sera DRH3-N-Ab and DRH3-C-Ab, respectively (Capralogics Inc). The recombinant proteins, each bearing the N-terminal peptide MGSSHHHHHHSSGLVPRGSH, were expressed in *E. coli* and purified on Ni-NTA resin (QIAGEN). Antibodies were affinity-purified using the corresponding recombinant proteins and Affi-gel 10 (BioRad, 153-1000), as described (Duchaine et al., 2006). Polyclonal sera were raised against peptides for RRF-1 (YIDNDKSSFHKPFYERQK) and EGO-1 (KSVRSSDDVQKINMRLLV) by Alpha Diagnostics Intl. (San Antonio, TX). Control Westerns showing specificity are provided Figure S6.

Quantitative Real-time PCR

Quantitative Real-time PCR (qRT-PCR) is as described (Batista et al., 2008; Ruby et al., 2006). 100 µg RNA was digested with 20 U of DNase I (Ambion) at 37°C for 1 hr followed by phenol-extraction and ethanol-precipitation. cDNA was generated from 5µg of total RNA, using random hexamers with Superscript III Reverse Transcriptase (Invitrogen). qRT-PCR was performed on the ABI Prism 7500 Sequence Detection System using Applied Biosystems SYBR Green PCR Master mix. Thermocycling was done for 40 cycles in a 15 µl reaction containing 7.5ul SYBR master mix, 0.3µl of 10µM primers, 2µl of cDNA, and 5.2ul dH2O. Primer sequences are provided in Table III-S5. Each

experiment was composed of triplicate wild type and drh-3 mutant samples respectively, with three technical replicates in the PCR reactions. The expression level of each target RNA is first normalized to 18S RNA. The average ratio of target mRNA/18S RNA from 3 wild type samples was designated as 1 and the standard deviation was proportionally adjusted. To quantify the change of RNA expression in the *drh-3* mutants, the average ratio of the 3 mutant samples was further normalized to that of wild type samples with standard deviation proportionally adjusted.

Microarray Analysis

Probe signals were calculated using Affymetrix Tiling Analysis Software 1.1.2 (bandwidth: 30; intensities: PM/MM) with three *drh-3* (*ne4253*) replicates as the experimental dataset and three wild-type replicates as the control. Affymetrix probe coordinates (release WS170) were converted to release WS192 coordinates using a Perl script. Gene expression values were defined as the geometric mean of all probe signals within a gene that had a P-value of <0.1. Actin was used to normalize expression values prior to comparison.

Northern Blot Analysis

Total RNA was extracted in a stainless steel dounce with TRI Reagent (MRC, Inc.), according to the manufacturer's instructions, and separated using a Phase-Lock gel column (Eppendorf). Small RNAs less than ~200nt were

prepared from total RNA using *mir*Vana™ miRNA Isolation Kit or reagents (Ambion, AM1560).

For small RNA analysis, ~8µg of *mir*Vana-enriched RNA was resolved on a 15% acrylamide/7M urea gel and transferred by electrophoresis to Hybond N+ (GE Healthcare, RPN303B) using a Trans-Blot SD (Biorad). For mRNA analysis, ~15µg total RNA denatured with glyoxal loading buffer (Ambion Cat. No. 8551) at 50°C for 1hr was resolved on a 1% agarose gel and transferred to Hybond N+ (GE Healthcare, RPN303B) membrane by capillary transfer with mild hydrolysis (Sambrook et al.). For small RNA Northerns, membranes were probed with ^{32}P-labeled StarFire oligos (IDT) in Ultrahyb Oligo buffer (Ambion) at ~30°C. Northerns to detect the K02E2.6 mRNA (Duchaine et al., 2006) were probed with a T7-transcribed riboprobe in Ultrahyb buffer (Ambion) at 65°C. Starfire probe sequences are provided in Table III-S5.

Biochemical analysis of small RNAs

To analyze the 5´ nucleotide composition of small RNAs, gel-purified 21nt and 22nt RNAs were dephosphorylated using CIP (NEB), rephosphorylated with PNK (NEB) in the presence of γ-^{32}P-ATP. 32P-labeled small RNAs were digested with RNase P1 (US Biological) and nucleotides were resolved on TLC plates (PEI-cellulose; Sigma) using 1 M LiCl, as described (Gu et al., 2009). The 3´ end status of gel-purified small RNA was determined by β-elimination analysis (Alefelder et al., 1998)

Small RNA cloning

18-26 nt endogenous RNAs were purified from ~50 µg *MirVana*-enriched small RNAs using 15% PAGE/7 M urea, followed by overnight elution with buffer containing 10 mM Tris-Cl (pH7.5), 1 mM EDTA, and 0.3 M NaCl, and precipitation with 20 µg glycogen and 1 volume of isopropanol. Ethidium Bromide staining was used to visualize the RNA standards (18mer and 26mer RNAs) and the endogenous RNAs. A 5´ ligation-independent method was used to clone small RNAs from *wild-type* and *glp-4* (*bn2*) animals. A 5´ ligation-dependent method was used to clone small RNAs from wild-type, *drh-3(ne4253)*, *mut-7(pk204)*, *rde-3(ne3364)*, MAGO12, *ekl-1(tm1599)*, *rrf-1(neC1) ego-1(om97)*, *rde-4(ne337)*, and *smg-5(r860)*. A TAP cloning method was used to clone small RNAs from *fem-1(hc17)* ooctyes.

In the 5´ ligation independent cloning procedure, to enrich for RNAs with 5´ triphosphate, the purified small RNAs were first incubated with 0.5 Unit/µl Polynucleotide Kinase (NEB, M0201L) in 40 µl reaction buffer containing 0.4 mM ATP and 0.5 Unit/µl SUPERase•In (Ambion) at 37°C for 1 hr, followed by phenol extraction and ethanol precipitation, and by incubation with 0.08 U/µl of Terminator 5´-Phosphate-Dependent Exonuclease (Epicenter Biotechnologies, TER51020) in 50 µl reaction buffer at 30°C for 45 min. The reaction was stopped by phenol extraction and the small RNAs were precipitated with 4 volume of ethanol. Small RNAs treated with Terminator exonuclease were cloned using a

5′ ligation independent cloning protocol essentially as described (Ambros et al., 2003). In this method, small RNAs and 1μM of each standard (18mer GGC GTG TAG GGA TCC AAA and 24 mer GGC CAA CGU UCU CAA CAA TAG TG synthetic RNAs) were incubated with 10 μM of pre-adenylated DNA oligo (AppCTGTAGGCACCATCAAT/ddC/), 1 Unit/μl of SUPERase•In, 10% DMSO and 2 U/μl T4 RNA ligase (Takara Bio Inc., 2050A) in 10μl ligation buffer (50mM Tris-Cl pH7.5, 10mM $MgCl_2$, 0.06μg/μl BSA, 10mM DTT). The 3′ ligated products were gel-purified as described above, and reverse transcribed in a standard 20μl reaction using SMART Technology (Clontech): 1) RNAs were annealed with 6 pmole DNA RT oligo ATT GAT GGT GCC TAC AG in 9.6 μl H_2O at 65°C for 5 min and incubated on ice for 2 min; 2) the hybridized RNAs were incubated with 1 μl PowerScript Reverse Transcriptase (Clontech Cat. PT3396-2) and 6 pmol of 5′ template DNA oligo TCT ACA GTC CGA CGA TCG GG in 20 μl buffer containing 5mM $MgCl_2$, 10mM DTT, 0.5 mM dNTP and 1 Unit/μl SUPERase•In at 42°C for 1 hr; 3) the reaction was heat-inactivated at 85°C for 5min and small RNAs were digested with 0.1 U/μl RNase H at 37°C for 20 min. The cDNA was PCR-amplified using RT oligo and 5′ template oligo for 10~15 cylces with *Ex Taq* (Takara, RR001B) and further extended for 5~10 cycles to add Solexa linkers.

In the 5′ ligation dependent procedure, small RNAs were incubated with 1 U/μl Calf Intestinal Phosphatase (NEB) and 1 U/μl SUPERase•In, in buffer 3 (NEB) at 37°C for 1hr, phenol-extracted twice, ethanol-precipitated, and then 3′ ligated as described above. The 3′-ligated-RNAs were phosphorylated with 1

U/µl Polynucleotide Kinase in buffer (NEB) containing 1 U/µl SUPERase•In and 2 mM ATP at 37°C for 1 hr, phenol-extracted once, ethanol-precipitated, and 5´-ligated with 2 Unit/µl T4 RNA ligase and 30 µM 5´ linker GTT CTA CAG TCC GAC GAT C in 10 µl 1X buffer containing 1 U/µl SUPERase•In, 0.1 µg/µl BSA and 10% DMSO. cDNA was synthesized using SuperScript III with the same 3´ RT oligo but without the 5´ template oligo. The same PCR strategy was used to obtain small RNA amplicons containing Solexa linkers.

A “TAP” cloning method was used to minimize the cloning of rRNA and mRNA fragments that degrade during the purification of oocytes. Small RNAs purified from oocytes (see below) were incubated with 0.05 Unit/µl Tobacco Acid Pyrophosphatase (Epicenter Biotechnologies T19050) in 10 µl reaction buffer containing 1 Unit/µl SUPERase•In at 37°C for 1hr, followed by the phenol extraction and ethanol precipitation. The RNAs were then ligated and cDNA was amplified, as described above in the 5´ ligation dependent procedure, but without CIP and PNK treatment. Finally, cDNA was purified via a 10% native acrylamide gel, and sequenced using Illumina Genome Analyzer at UMASS CFAR and at Oregon State University Center for Genome Research and Biocomputing (Seo et al., 2004).

To clone the oocyte small RNAs, functional ooctyes were obtained from *fem-1* (*hc17*) animals as described (Aroian et al., 1997) which does not produce sperms and embryos at 25 ^{o}C, while accumulating lots of unfertilized oocytes. Therefore, the sample RNA should be free of sperm or embryonic RNAs. The

worms were precipitated in M9 buffer several times by gravity to remove any *E. coli* contamination. To obtain a high quality of oocytes, we only used *Serotonin* and *Levamisole* to induce oocyte-laying followed by purification via a 20 µm mesh. The oocytes were washed twice to remove cell debris and then were visually checked under microscope. During the whole process, we kept the worms intact and healthy at room temperature with gentle stirring, thus avoiding contamination from broken germline and somatic tissue. As a result, small RNAs derived from somatic target genes, such as *Y47H10A.5* and *C44B11.6*, were depleted at least 20 fold in the oocyte library than in the wild type worm library. Furthermore, mir-35-42 family was enriched 7 fold in the oocyte than in the gravid adult sample, reaching ~ 70% of total miRNAs. In the *mir-35-42* family, *mir-38* was highly expressed, covering more than 1/3 of total miRNA reads in the oocytes. Although this profile is similar to that of miRNA in the embryos, there were some apparent discrepancies between the oocyte sample and embryonic sample because *mir-41* was ~24 higher in the embryos than in the oocytes. Therefore, we believe we purified a highly pure oocyte RNA sample and free of contamination from germline and somatic tissue, sperm and embryos.

FLAG::R06C7.1 IP-cloning of small RNA

Approximately two million *FLAG::R06C7.1* young adult worms grown at 24°C were washed successively in M9, ice-cold water and ice-cold 20 mM Hepes-KOH (pH 7.3) plus 1mM DTT, 180 microgram/ml PMSF, 4 microgram/ml

Pepstatin A. The worm pellet (~1.3 ml) was suspended with the same volume of lysis buffer (20 mM Hepes-KOH [pH 7.3], 110 mM K-acetate, 2 mM Mg-Acetate, 0.1% Tween 20, 0.5 % Triton X100, 1 mM DTT, 180 µ g/ml PMSF, 4 µg/ml Pepstatin A, complete protease inhibitor cocktail (Roche), 1/5000 of Antifoam B Emulsion (Sigma)) and homogenized with a dounce homogenizer. The lysate was cleared twice at 16,000xg for 15 minutes at 4°C. ~12mg of lysate was precleared by adding 250µl of Protein G magnetic beads (Roche) and rocking at 4°C for 1hr. Precleared lysate was incubated with 10µg of anti-FLAG antibodies M2 (Sigma) at 4°C for 1hr. Immune complexes were capture by adding 50µl Protein G beads and gently rocking for 1hr at 4°C. Beads were washed with lysis buffer 3x15min at 4°C and once with wash buffer (100 mM Ammonium-acetate, 0.1 mM MgCl2, 0.2 % Tween 20, pH 7.5) for 5 minutes at room temperature. Beads were resuspended in 50µl wash buffer and extracted with 200 µl Trizol. Small RNAs were precipitated with 1/10th 3M NaOAc and 3 volumes of ethanol in the presence of 20µg of glycogen, washed with 80% of ethanol, and resuspended in 20µl of water for RNA cloning.

Small RNA Analyses

All small RNA analyses were performed using custom Perl (5.8.6) scripts and Wormbase release WS192. The following modifications or additions were made to WS192: 22 mitochondrial tRNAs and 4 non-coding RNAs (ncRNAs) were manually repaired to complete the missing chromosome strand or

sequence information; mir-1829a was reannotated to the Crick strand at the same WS192 locus; 79 transcripts annotated as both protein coding genes and ncRNAs were considered as protein coding genes to avoid redundancy; the mature *lin-4* sequence (TCCCTGAGACCTCAAGTGTGA) was added; 21U-RNAs were as defined (Batista et al., 2008; Ruby et al., 2006). Finally, cDNA loci included in our analysis were 20186 genes, 1467 pseudogenes, 156 miRNAs, 15915 21U-RNAs, 159 ncRNAs, 630 tRNAs, 138 snoRNAs, 21 rRNAs, 94 snRNAs, 1 scRNA, complex repeats annotated by HMMFS and RepeatMasker, simple repeats including tandem and inverted repeats, and consensus repeat sequences annotated by RepBase 13.07.

In the 5´-ligation-independent procedure, Solexa reads containing the 5´ linker sequences (GGG) at position 1-3 and a perfect match to the first 6nt of the 3´ linker (CTGTAG) are used to extract the inserted sequences. Inserts 17-27 nt long were used to identify perfect matches to the *C. elegans* genome, cDNAs, introns, repeats, and transposons. A single match returned a unique genomic locus, which was defined by a chromosome number, the start and the end of the matched sequence on the chromosome, and its orientation. The genome analysis returned all the matches but those to exon-exon junctions, which were obtained through cDNA analysis using WS192 cDNA annotations. To calculate small RNA reads derived from a single genomic locus, the reads of a unique small RNA were normalized using the total genomic loci it matched (repeat-

normalized reads), assuming that each of these loci contributes the same amount of this RNA.

For 5´-ligation-dependent libraries, the same procedure was used except that: 1) no 5´ linker was used to extract the inserts and 2) 28-30 nt small RNAs were also included in the analysis.

To compare 22G-RNAs derived from a gene, transposon, or pseudogene between two samples, each sample was normalized using the total number of reads less structural RNAs, i.e. sense small RNA reads likely derived from degraded ncRNAs, tRNAs, snoRNAs, rRNAs, snRNAs, and scRNAs. Degradation products of structural RNAs map to the sense strand, with a poorly defined size profile and 1^{st} nucleotide distribution. At least twenty-five 22G-RNA reads per million, non-structural reads in one of the two samples was arbitrarily chosen as a cutoff for comparison analyses. A change of 2-fold or more between samples was chosen as an enrichment threshold. Because some 21U-RNAs or miRNAs overlap with protein coding genes, reads derived from miRNAs loci within a window of ±4nt and all the known 21U-RNAs were filtered out prior to comparison analysis.

Relative distribution of small RNAs along protein coding loci

To analyze the relative distribution of 22G-RNAs along all protein coding genes, each gene was divided into 20 equal bins, named bin 1 to 20 according to their relative distance from the 5´ end of the gene and each bin representing 5% of the coding potential. For gene with multiple splice forms, all the coding exons

of this gene were collapsed into one spliced genomic locus according to the genomic coordinates, and then the spliced locus was divided into 20 bins. Repeat-normalized reads of 22G-RNAs of each gene were then sample normalized using the total non-structural reads, and then were mapped to each bin. To simplify the analysis, especially for small RNAs spanning exon-exon junctions, 50% reads of each 22G-RNA were assigned to the start bin where the 5′ end of the 22G-RNA was located and the rest 50% were assigned to the end bin where the 3′ end was located. In Wormbase WS192, 9367 genes don't have any annotated 3′UTRs, while 10819 genes have at least one splicing forms with annotated 3′ UTRs, among which 8113 genes have annotated 3′UTRs for all the splicing forms. Only considering genes with at least one 22G-RNAs in both *drh-3* mutant and wild-type animals, we analyzed the small RNA distribution both on 6103 genes with 3′UTRs from 10819 3′UTR-containing genes, and on 1916 non-3′UTR genes. The total number of reads in each bins was calculated by summing the corresponding bin of each gene. This was done for all the annotated protein-coding genes as well as those with annotated 3′UTRs and those without an annotated 3′UTR. To exclude the possibility that an individual gene could affect the overall distribution significantly, we performed a different but similar analysis, in which the contribution of a given gene to each of the 20 bins was represented by the 22G-RNA percentage in each of the 20 bins along the gene.

Phylogenetic analysis

T-Coffee was used for protein alignment (Poirot et al., 2003). Bootstrapping and neighbor joining methods using the Phylip software package were used to create the cladogram (Felsenstein, 2005).

Supplemental Information

Figure III-S1

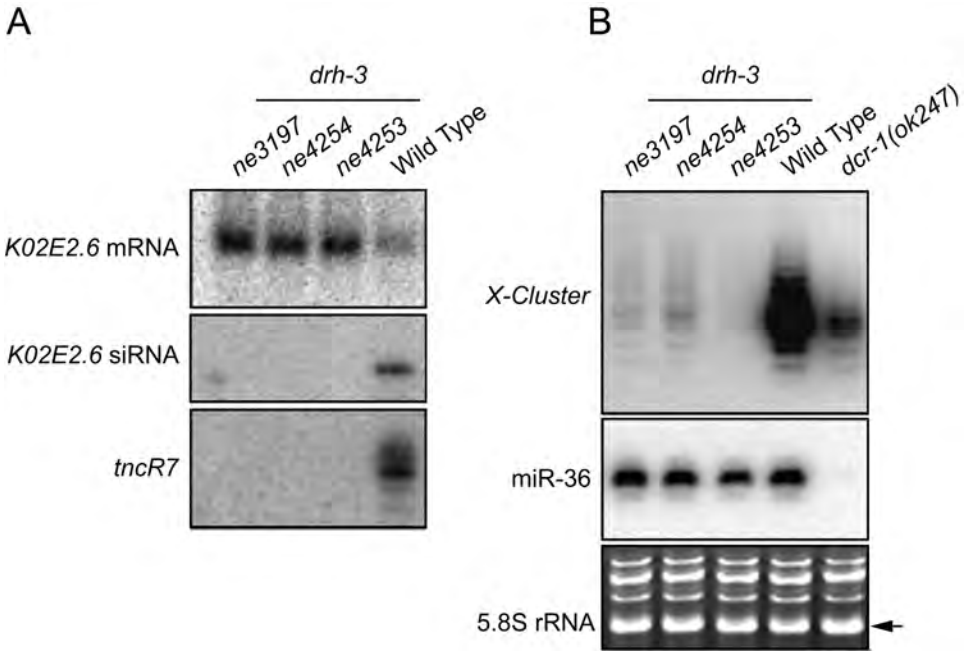

Figure III-S1. *drh-3* Mutants Behave Like Loss-of-Function for Endogenous Small RNA Biogenesis

(A) Northern blot analyses to detect *K02E2.6* mRNA and endo-siRNAs and the tiny noncoding RNA *tncR7* in wild-type and *drh-3* mutant animals.

(B) Northern blot analysis of X-loci small RNAs and *mir-36* miRNA in wild-type, *drh-3*, and *dcr- 1* mutant animals. 5S rRNA is shown as a loading control.

Figure III-S2

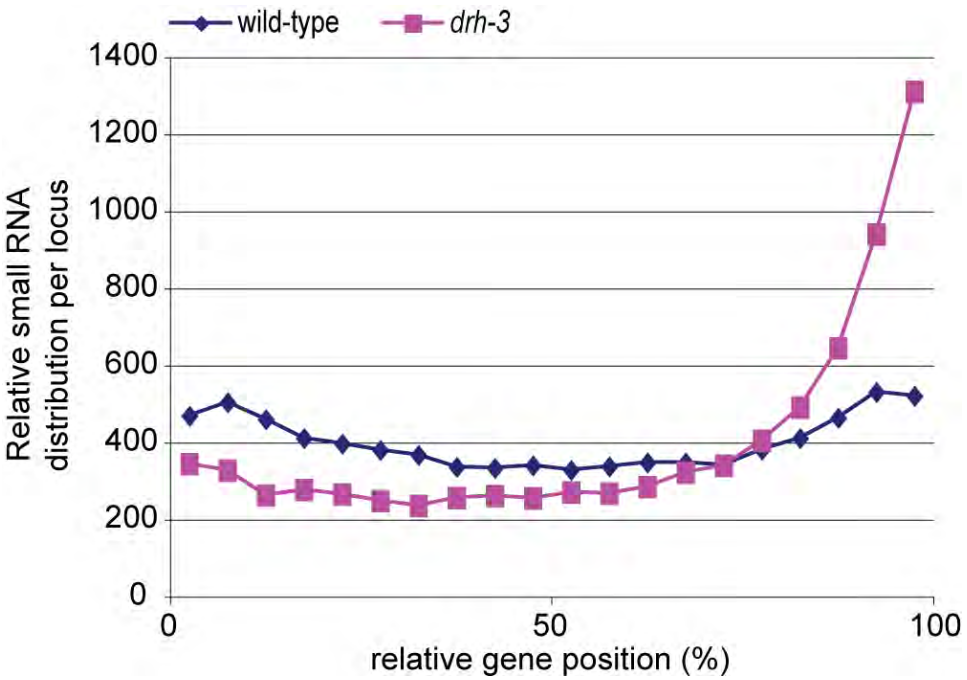

Figure III-S2. The Relative Distribution of Small RNAs Along Protein-Coding Genes

The relative fraction of antisense reads was determined along the length of each transcript, such that the total contribution equals 1.0. The graph represents the relative contribution of small RNAs for all transcripts (y axis) in both wild-type (diamonds) and *drh-3* mutant (squares) samples at the relative position along all transcripts (x axis; 5´ to 3´).

Figure III-S3

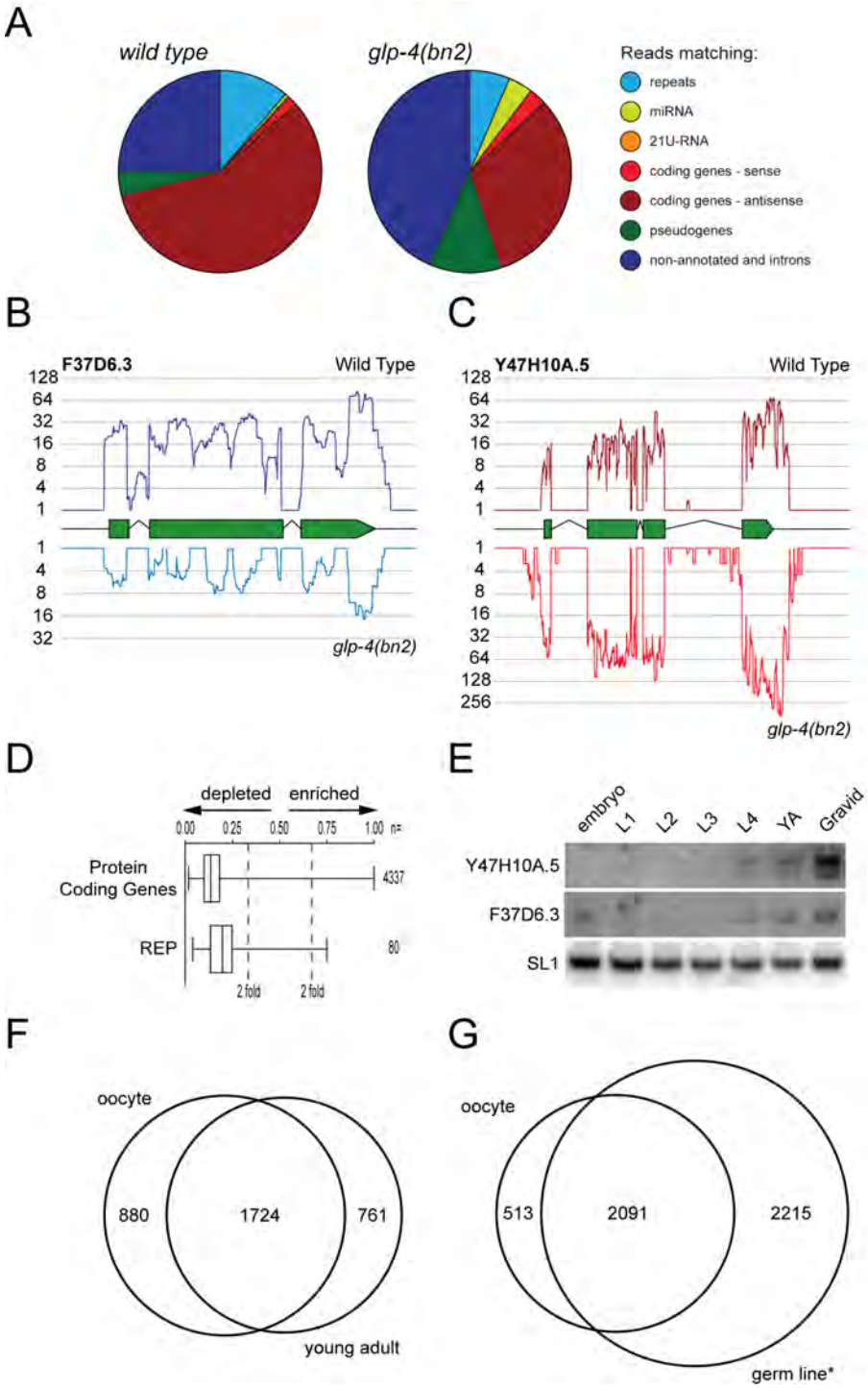

Figure III-S4

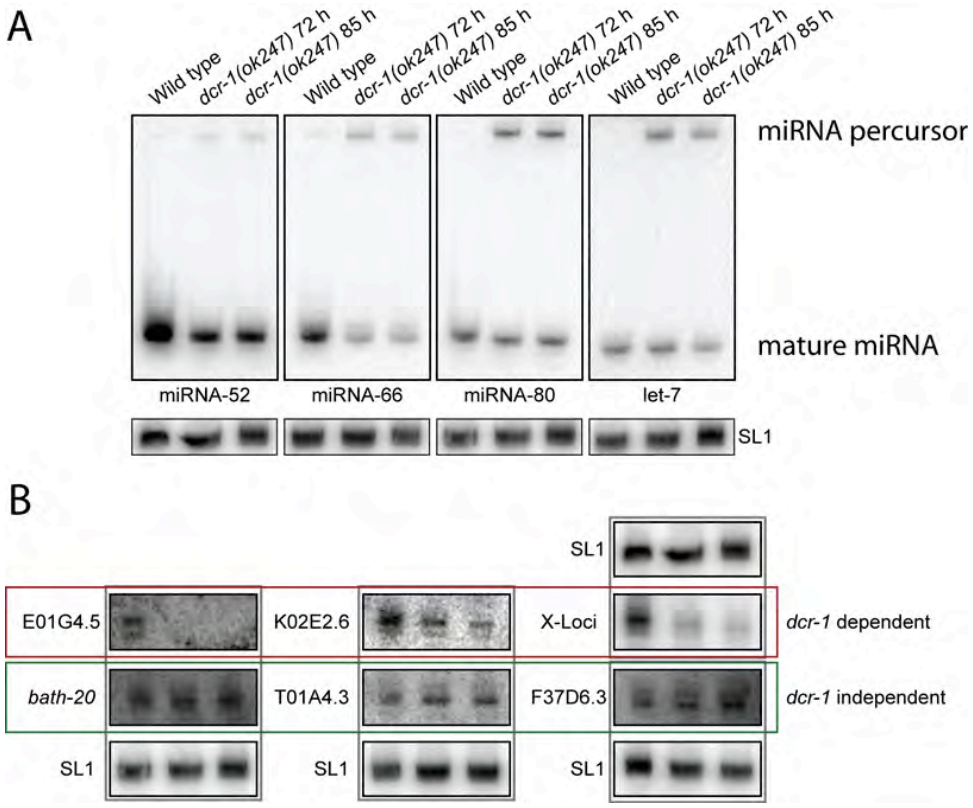

Figure III-S4. Northern Analysis of Dicer-Dependent or Independent Small RNAs

(A) Northern analysis of miRNAs, as indicated. Wild-type worms were grown for 72 hr at 20°C, and *dcr-1(ok247)* mutants were grown at 20°C for 72 hr and 85 hr to deplete the maternal Dicer in the adult worms. SL1 precursor was used as loading control for all the northern blots below. Mature miRNAs and accumulated miRNA precursors in *dcr-1(ok247)* mutant were labeled on the right.

(B) Northern analysis of Dicer-dependent 22G-RNAs (ERI target genes) and Dicer-independent 22G-RNAs, as indicated.

Figure III-S5

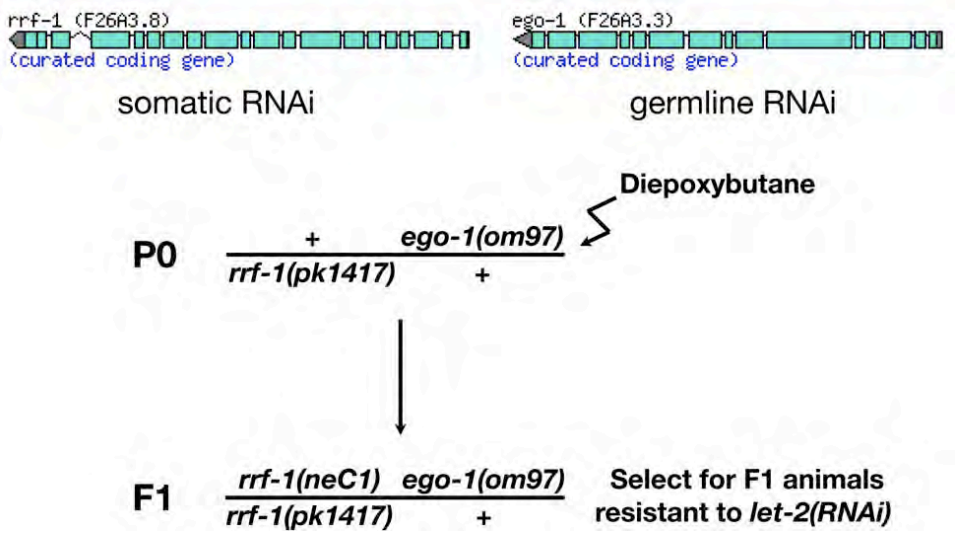

Figure III-S5. Noncomplementation Screen to Generate *rrf-1 ego-1* Double Mutants

The *rrf-1* and *ego-1* genes reside as tandem loci on chromosome I, separated by only ~650 bp of intergenic sequence, making it difficult to generate double mutants by recombination. To circumvent this complication, we performed an *rrf-1* noncomplementation screen to generate *rrf-1 ego-1* double mutants. P0 animals of the indicated genotype were mutated using diepoxybutane. F1 progeny were selected for viability on bacteria expressing *let-2* dsRNA to isolate candidate *rrf-1 ego-1* double mutant chromosomes balanced by an *rrf-1* mutant chromosome. This screen yielded a rearrangement, *neC1*, that disrupts *rrf-1*, resulting in a putative null allele of *rrf-1* linked to the *ego-1(om97)* nonsense allele. The *rrf-1(neC1) ego-1(om97) / rrf-1(0)* animals are resistant to somatic RNAi but sensitive to germline RNAi, and segregate sterile *rrf-1(neC1) ego-1(om97)* self progeny.

Figure III-S6

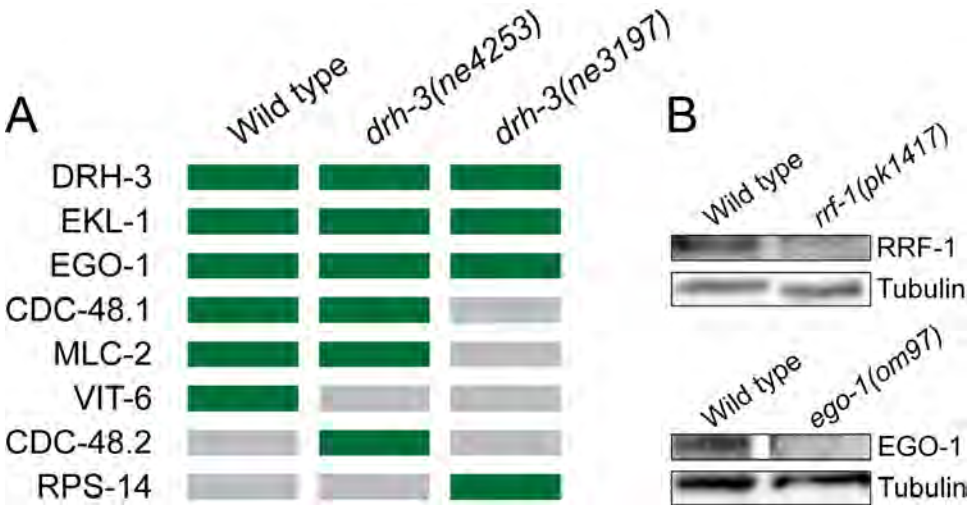

Figure III-S6.
Identification of DRH-3-Interacting Proteins Using MudPIT Analysis

(A) Proteins co-immunoprecipitated with DRH-3 in wild-type and *drh-3* mutant samples. Green boxes: protein enriched more than 5-fold in wild-type or *drh-3* mutants over *drh-3* deletion strain with a P-value of 0.0001; gray boxes: not enriched.

(B) Western blot analysis to detect RRF-1 and EGO-1 in wild-type samples and deletion mutants as indicated. α-tubulin was as used as loading control.

Figure III-S7

Figure III-S7. Distribution of Small RNA Reads in Various Mutant Backgrounds

Distribution of reads that match the indicated genome annotations sequenced in wild-type and mutant small RNA libraries. DA1316 is the congenic, wild-type control for *rrf-1 ego-1* and *ekl-1* mutant libraries.

Figure III-S8

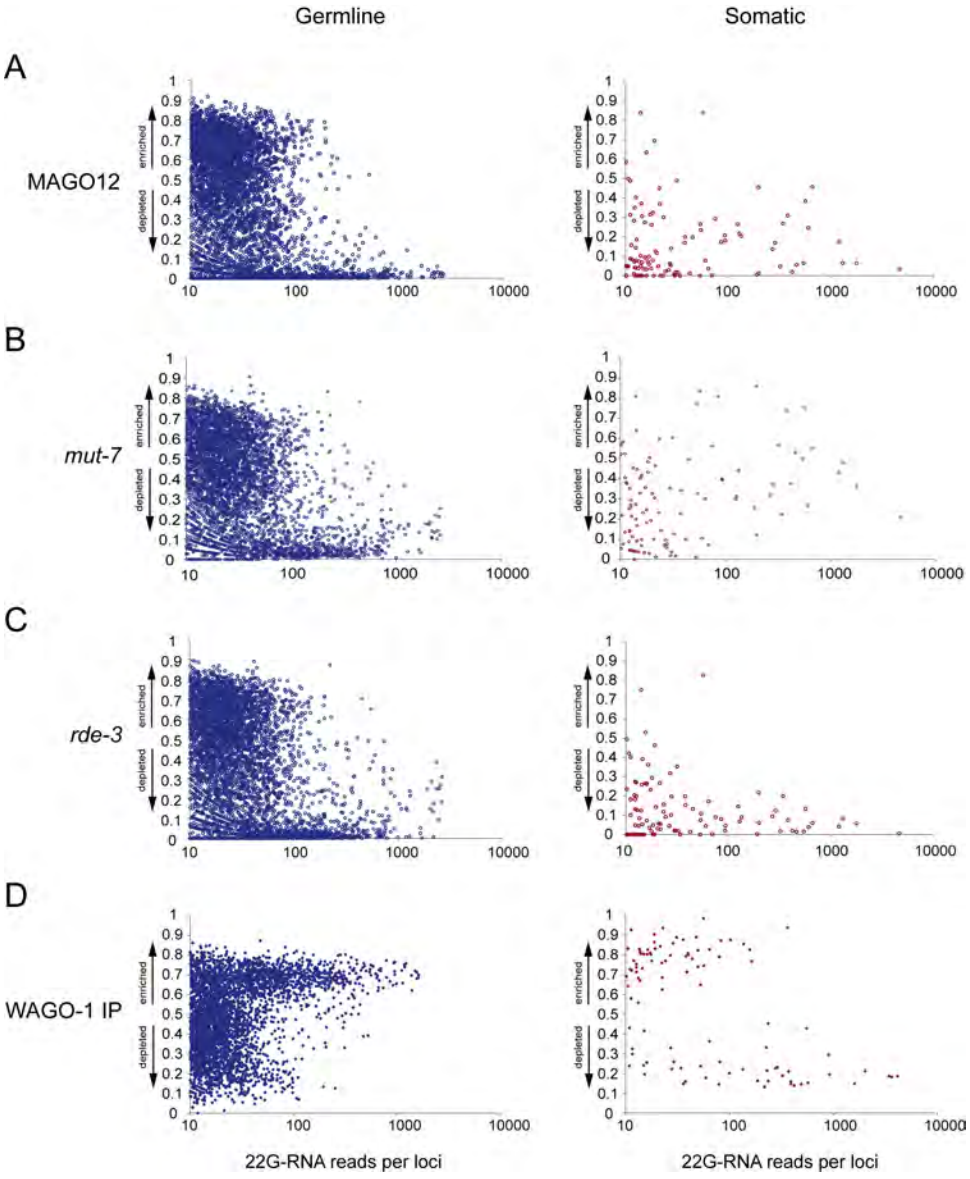

Figure III-S8. Normally Abundant 22G-RNA Loci Are Depleted in *rde-3*, *mut-7*, and MAGO12 Mutants and Enriched in WAGO-1 IP Samples

(A-C) Enrichment or depletion of 22G-RNAs derived from annotated protein-coding loci in the *rde-3* (A), *mut-7* (B) and MAGO12 (C) mutant samples (relative to wild-type) was plotted with respect to the number of small RNAs derived from each gene in the wild-type sample. Loci with at least 10 reads per 10^6 total (not including structural) reads in the wild-type sample were analyzed. Relative enrichment was calculated as the ratio of mutant / (mutant + wild-type) for each feature. Values approaching 0 indicate depletion of 22G-RNAs, while values approaching 1 indicate enrichment. Dotted-lines denote 2-fold enrichment (upper) and 2-fold depletion (lower). Germline 22G-RNA loci (left) and Somatic 22G-RNA loci (right) represent loci that were at least 2-fold reduced or increased, respectively, in the *glp-4(bn2)* mutant.

(D) Enrichment or depletion of 22G-RNAs derived from annotated protein-coding loci in the FLAG::WAGO-1 IP samples (relative to wild-type) was plotted with respect to the number of small RNAs derived from each gene in the FLAG::WAGO-1 input sample. Relative enrichment was calculated as the ratio of IP / (IP + input) for each feature.

Figure III-S9

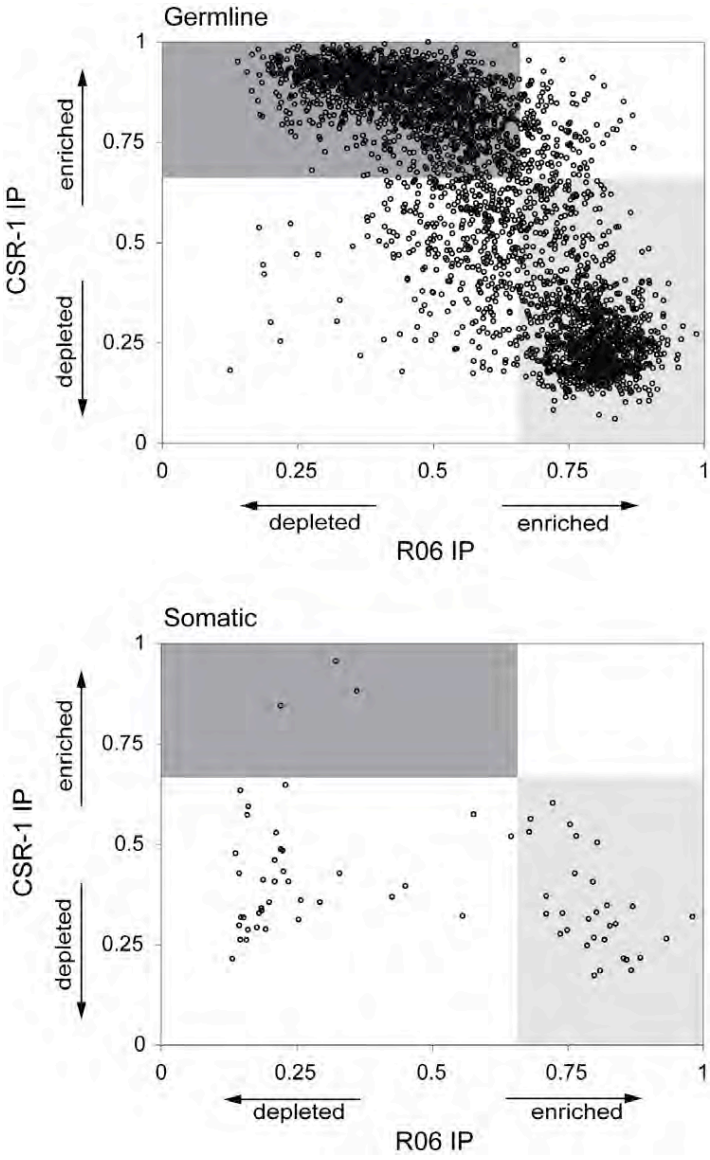

Figure III-S9. WAGO-1 and CSR-1 Bind Distinct 22G-RNAs

The enrichment or depletion of 22G-RNA protein-coding loci in the CSR-1 IP (Claycomb et al., 2009) was compared to the WAGO-1 IP (Figure S5). Loci increased 2-fold or more only in the CSR-1 IP are in the upper left quadrant (dark gray). Loci increased 2-fold or more in the WAGO-1 IP are in the lower right quadrant (light gray). Loci increased 2-fold or more in both IPs are in theupper right quadrant (white). Upper panel, germline loci; lower panel, somatic loci.

Figure III-S10

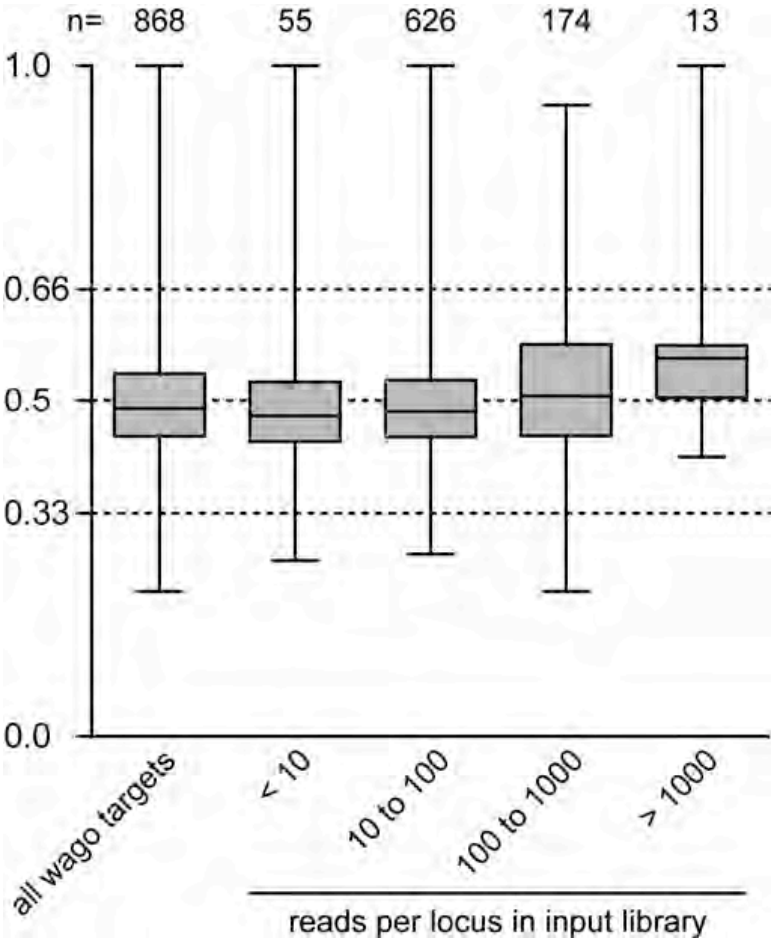

Figure III-S10. The Expression Change of WAGO Targets in the *drh-3* Mutant Relative to Wild- Type as a Function of Small RNA Number

Expression ratio expressed as *drh-3* / (*drh-3* + wild-type) was determined for each WAGO target, as well as for each target within the indicated range of small reads in the wild-type sample. “n” loci with a significant microarray signal were analyzed. Top and bottom solid lines in each category represent the highest and lowest values (i.e. most upregulated and most downregulated) respectively. The top and bottom ends of each box represent the 75th and 25th percentile, respectively; and the solid line within the box represents the median value. Dotted-lines denote 2- fold upregulation (upper) and 2-fold downregulation (lower).

Figure III-S11

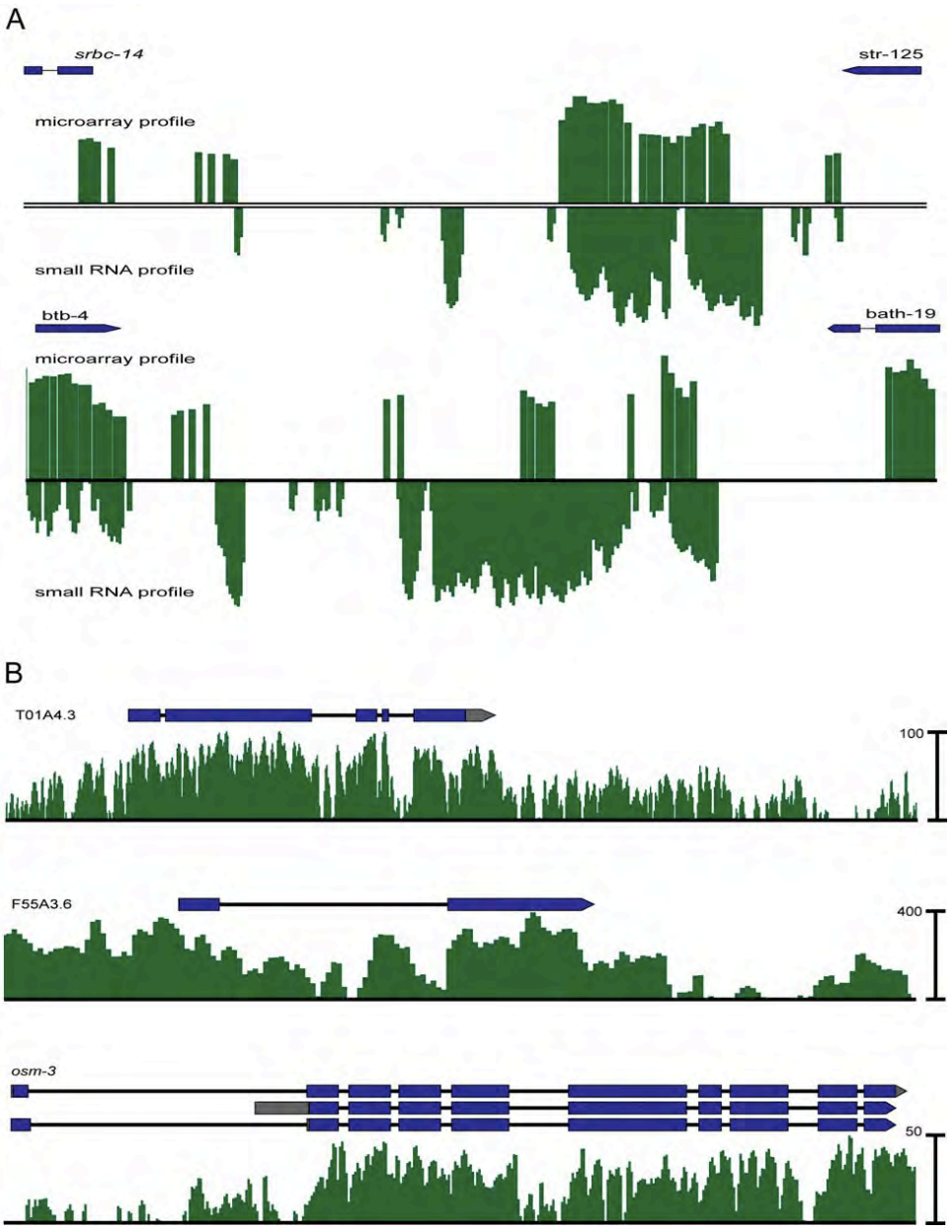

Figure III-S11. Examples of Nonannotated Genomic Loci of 22G-RNAs and R06 IP Target Loci

(A) Examples of nonannotated genomic loci of 22G-RNAs between *srbc-14* and *str-125*, and between *btb-4* and *bath-19*. The microarray signals (upward) and small RNA profile (downward) are shown. For the small RNA profile, each bar represents the number of small RNAs at each nucleotide averaged across a 10 bp window.

(B) Examples of 22G-RNA profiles from annotated protein-coding loci *T01A4.3*, *F55A3.6*, and *osm-3*, which are enriched in the WAGO-1 IP.

CHAPTER IV

Sequential Rounds of RNA-dependent RNA Transcription Drive Endogenous Small-RNA Biogenesis in the ERGO-1/Argonaute Pathway

Author Contributions

Chapter IV contributions. J.Vasale contributed to this section via the following: Small RNA isolation and purification of small RNAs in RdRP mutant (single and multiple mutant combinations of mutants: *ego-1; rrf-1, rrf-2 and rrf-3*) strains as well as in *ergo-1* mutant strain in addition to preparation of small RNA cDNA libraries and cloning. Performed ERGO-1 IP/Northern, Western, and IP/ small RNA cloning experiments, and contributed to analyses and interpretation of deep-sequencing results. These data are presented in Figure IV-1, Figure IV-2, Figure IV-3, Figure IV-4, Figure IV-5, Figure IV-S1, Figure IV-S2, Figure IV-S4, Figure IV-S5, Table IV-S1, Table IV-S3, and Table IV-S4. Chapter IV of this thesis was published as "Sequential rounds of RNA-dependent RNA Transcription Drive Endogenous Small-RNA Biogenesis in the ERGO-1/Argonaute Pathway", *****Vasale JJ, *****Gu W, Thivierge C, Batista PJ, Claycomb JM, Youngman EM, Duchaine TF, Mello CC, Conte D, Jr. (2010); Feb 23;107 (8):3582-7 *Proceedings of the National Academy of Sciences of the United States of America*. Authors need not obtain permission for use as part of their dissertations

SUMMARY

Argonaute proteins interact with distinct classes of small RNAs to direct multiple regulatory outcomes. In many organisms, including plants, fungi, and nematodes, cellular RdRPs use Argonaute targets as templates for amplification of silencing signals. Here, we show that distinct RdRPs function sequentially to produce small RNAs that target endogenous loci in *Caenorhabditis elegans*. We show that DCR-1, the RdRP RRF-3, and the dsRNA-binding protein RDE-4 are required for the biogenesis of 26-nt small RNAs with a 5′ guanine (26G-RNAs) and that 26G-RNAs engage the Piwi-clade Argonaute, ERGO-1. Our findings support a model in which targeting by ERGO-1 recruits a second RdRP (RRF-1 or EGO-1), which in turn transcribes 22G-RNAs that interact with worm-specific Argonautes (WAGOs) to direct gene silencing. ERGO-1 targets exhibit a nonrandom distribution in the genome and appear to include many gene duplications, suggesting that this pathway may control over expression resulting from gene expansion.

INTRODUCTION

RNA interference (RNAi) is a mechanism of gene regulation directed by Argonaute proteins in conjunction with their sequence-specific small RNA cofactors. A multitude of distinct Argonaute-mediated regulatory modules have been identified in plants, fungi, and animals (Chapman and Carrington, 2007). In all of these pathways, base pair interactions between the small RNA and a target

molecule provide specificity, whereas the Argonaute protein, which contains a conserved nuclease domain, can direct silencing through target cleavage or through recruitment of transcriptional or posttranscriptional regulators.

Argonaute pathways can be triggered by natural or exogenous sources of double-stranded (ds)RNA. The Dicer family of RNase III related enzymes processes dsRNA into small RNAs of 20–25 nucleotides (nt) in length (Ghildiyal and Zamore, 2009). For example, micro (mi)RNAs are processed from genomically encoded hairpins and mediate Argonaute-dependent silencing at the posttranscriptional level (Grishok et al., 2001; Hutvagner et al., 2001; Lee et al., 1993). Short interfering (si) RNAs are processed from both endogenous and exogenous sources of dsRNA and direct Argonaute-dependent cleavage of target mRNAs (Ghildiyal and Zamore, 2009). siRNAs processed from primary dsRNA sources (for example, hairpins or convergent transcripts) are referred to as primary siRNAs. In fungi, plants, and nematodes, RdRPs are required for the amplification of silencing signals. siRNAs that are processed from dsRNA generated by RdRP are referred to as “secondary siRNAs.” In *C. elegans*, secondary siRNAs appear to be directly synthesized by RdRP, independently of DCR-1, and are loaded onto Argonautes (Aoki et al., 2007; Pak and Fire, 2007; Sijen et al., 2007).

In *C. elegans*, two RdRPs, RRF-1 and EGO-1, are required for the biogenesis of an abundant class of endogenous small RNAs called 22G-RNAs (Gu et al., 2009), which are predominantly 22 nt in length and contain a

triphosphorylated 5′ guanine. Interestingly, 22G-RNAs are antisense to more than 50% of annotated genes (Gu et al., 2009). Two major 22G-RNA systems exist in *C. elegans:* those that interact with the Argonaute CSR-1 and those that interact with the expanded family of worm-specific Argonaute (WAGO) proteins (Claycomb et al., 2009; Gu et al., 2009). The CSR-1/22G-RNA system is required for the proper organization of holocentric chromosomes and is essential for chromosome segregation. The WAGO/22G-RNA system provides surveillance against transposable elements and aberrant endogenous transcripts (the endo- RNAi pathway) and is also required for the response to foreign dsRNA (the exo-RNAi pathway).

The ERI endo-RNAi pathway was defined by mutations that result in an enhanced exo-RNAi response (Duchaine et al., 2006; Kennedy et al., 2004). The Eri phenotype appears to reflect relaxed competition for limiting RNAi factors that are also required for the response to exogenous dsRNA (Duchaine et al., 2006; Nakamura et al., 2007; Yigit et al., 2006). The RdRP RRF-3 was identified as an Eri mutant (Simmer et al., 2002) and as a physical interactor with Dicer and other proteins defined genetically as Eri factors (Duchaine et al., 2006). On the basis of these and other studies, it was proposed that RRF-3 produces endogenous dsRNA that is processed by Dicer and loaded onto the Argonaute, ERGO-1, for which loss-of-function also results in an Eri phenotype (Duchaine et al., 2006; Yigit et al., 2006). WAGOs were identified as possible secondary Argonautes in the ERI pathway (Yigit et al., 2006).

Here, we show that the ERI pathway is indeed a two-step Argonaute pathway and that the pathway also involves two separate rounds of RdRP-mediated small RNA biogenesis. We show that, in embryos, ERGO-1 interacts with the previously described 26-nt small RNAs with a 5′ guanine (26G-RNAs) (Ruby et al., 2006). Furthermore, components of the ERI complex, including RRF-3, DCR-1, and the dsRNA-binding protein RDE-4, are required for the biogenesis of both 26G- and 22G-RNAs on Eri targets. In contrast, RRF-1 and WAGOs are required for the accumulation of 22G- but not 26G-RNAs. Hence, we propose that 26G-RNAs are the primary small RNAs in the ERI pathway that drive the downstream production of 22G-RNAs. Many ERGO-1 26G-RNA targets appear to be ancient duplications, suggesting that the function of this pathway may be to buffer the expression of rapidly expanding gene families.

RESULTS

ERGO-1 Interacts with 26G-RNAs

Previous work identified ERGO-1 as an Argonaute that functions in the ERI endo-siRNA pathway (Yigit et al., 2006). To identify small RNAs that interact with ERGO-1, we deep sequenced small RNAs prepared from ERGO-1 immunoprecipitation (IP) and input samples. Developmental expression studies indicated that ERGO-1 was primarily expressed in embryos and was virtually

absent from L3 and L4 larvae and young adults (lacking embryos) (Figure IV-1A and Figure IV-S1). Therefore, ERGO-1 IP experiments were performed using embryo lysates. Analyses of both size and first-nucleotide distribution of reads revealed that 26G-RNAs were enriched ≈2.2-fold in the ERGO- 1 IP sample over the input library (Figure IV-1B). Although 21- to 23- nt small RNAs with a 5′ uracil (5′U), including miRNAs and 21U-RNAs, were cloned at high levels, they were not enriched in the ERGO-1 IP sample (Figure IV-1B). The modest enrichment of 26G-RNAs and the high background of 21U-RNAs and miRNAs is consistent with the low efficiency of ERGO-1 IP (Figure IV-S1). However, the interaction between ERGO-1 and a representative 26G-RNA appears to be specific, because both the expression of the 26G-RNA and its interaction with ERGO-1 were abrogated in an *ergo-1* null mutant (Figure IV-S1).

Figure IV-1

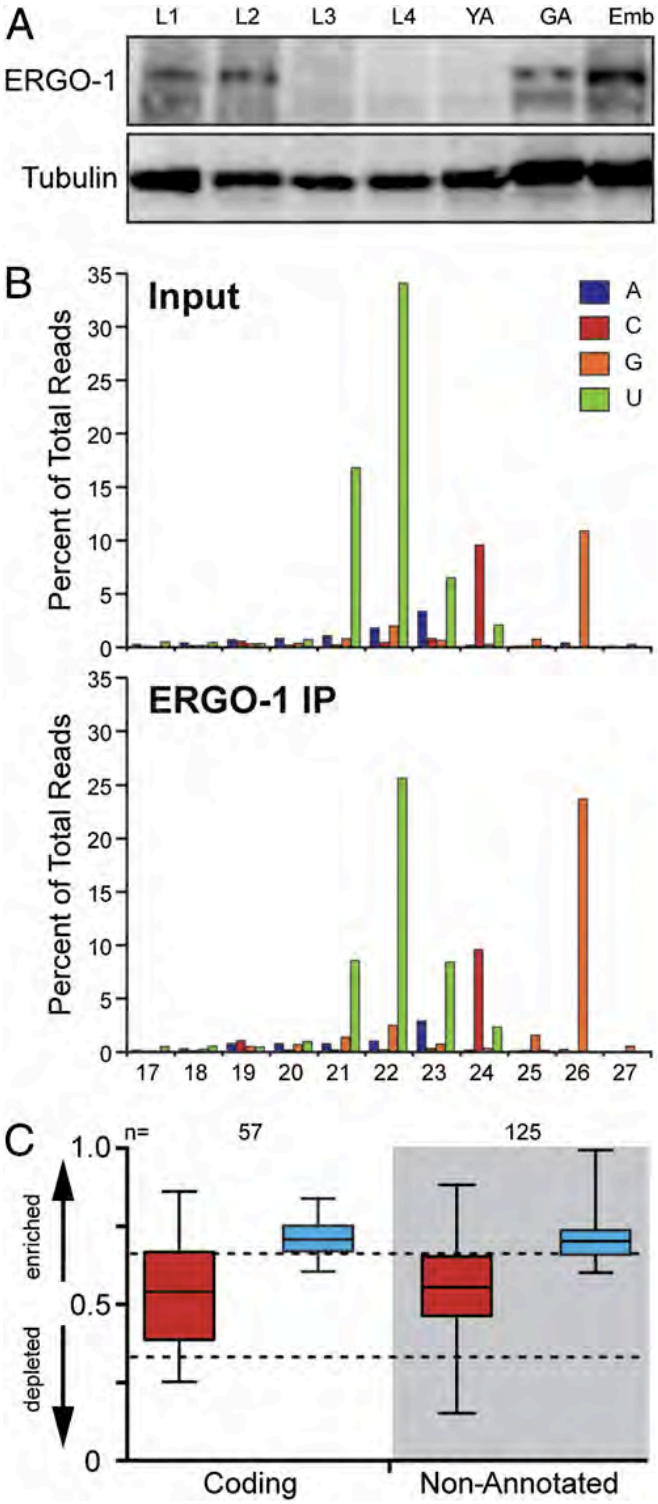
A
L1
L2
L3
L4
YA
GA
Emb
ERGO-1
Tubulin
B
Input
A
C
G
U
Percent of Total Reads
ERGO-1 IP
17 18 19 20 21 22 23 24 25 26 27
C
n=
57
125
1.0
0.5
0
enriched
depleted
Coding
Non-Annotated

Figure IV-1. ERGO-1 Interacts with 26G-RNAs in Embryos.

(A) Expression profile of ERGO-1 protein.

(B) Length and first nucleotide distribution of genome matching reads in Input and ERGO-1 IP small RNA libraries.

(C) Enrichment or depletion of small RNAs derived from 26G-RNA genes (white area) or nonannotated clusters (gray area) in the ERGO-1 IP. Small RNAs were separated by read length into 26 nt (blue) and <26 nt (red). Values approaching 1 indicate enrichment of small RNA; values approaching 0 indicate depletion. Relative enrichment was calculated as ratio of IP/(IP + wild-type). “n” loci with at least 10 reads per million total reads (not including structural) in either the wild-type or the mutant sample were analyzed. The top and bottom of each box represent the 75th and 25th percentiles, respectively. The horizontal line within each box represents the median value. Dotted lines denote 2-fold enrichment (Upper) and twofold depletion (Lower).

ERGO-1-associated 26G-RNAs mapped to genes (23%), pseudogenes (8%), and nonannotated loci (64%) in proportions similar to those observed for the 26G-RNA species present in our non-IP dataset (Figure IV-S1). Comparing these two datasets, 26GRNAs targeting a set of 57 genes were enriched by ≥1.5-fold in the ERGO-1 IP dataset (Figure IV-1C and Table IV-S1). When we examined each 5′G size population independently, we observed varying degrees of enrichment in the ERGO-1 IP. However, of the reads that matched 26G-RNA targets, only 24G–28G reads were enriched twofold (Figure IV-S2). The frequency distribution of 24G–28G reads was consistent with these populations being derived from 26G-RNAs (Figure IV-S2). Greater than 90% of the enriched 24G and 25G reads represented 3′ truncations of 26G-RNAs, whereas the 27G and 28G reads appeared to represent terminal transferase products (Claycomb et al., 2009). Small RNA species, ranging from 17 to 23 nt, including 22G-RNAs (Gu et al., 2009), were also present, and in some cases enriched, in the IP dataset. The significance of this enrichment is not clear. However, because such reads were present at very low levels, were derived from loci that are not targeted by 26G-RNAs, and were not dependent on ERGO-1 (Figure IV-1 and Figure IV-S2). Therefore, although we cannot rule out specific interactions with other small RNA species, the above data strongly support the direct association between ERGO-1 and 26G-RNAs that are abundant during embryogenesis.

26G-RNAs were previously shown to be 5′ monophosphorylated small RNAs with a 3′ modification that is resistant to oxidation by periodate (β-

elimination) (Ruby et al., 2006). To determine whether the ERGO-1 interacting 26G-RNAs correspond to those identified by Ruby et al. (Ruby et al., 2006), small RNAs purified from adult animals containing embryos were oxidized with periodate (β-eliminated) before cloning and deep sequencing. A second library was prepared in parallel from untreated small RNAs as a control and both libraries were prepared using a method compatible with cloning mono- or triphosphorylated small RNAs, i.e., 22G-RNAs (Gu et al., 2009).

Of 2.77 million genome-matching reads in the untreated sample, 0.5% corresponded to potential 26G-RNAs and 2.9% to 21U-RNAs, whereas the majority of reads represented 22G-RNAs (Figure IV-2A). In the β -eliminated sample, the 26G and 21U species were enriched 8.6- and 12.3-fold, constituting 4.3 and 35.8% of the 5.68 million genome matching reads, respectively. In both samples, ~40% of 26G-RNA reads mapped antisense to coding genes (30%) or pseudogenes (10%). More than half (~56%) of the reads mapped to genomic loci lacking any annotation (Figure IV-2B), similar to the assignment of ERGO- 1 interacting 26G-RNAs. We identified 49 genes with antisense 26G-RNAs that were enriched in the β-eliminated sample over the untreated control (Tabel IV-S1), 48 of which were enriched at least 1.5- fold in the ERGO-1 IP (Figure IV-S2). Analyzing deep-sequencing data across developmental time points (Batista et al., 2008), we observed that 26G-RNAs derived from these 48 genes were most abundant during embryogenesis and decrease dramatically during larval development (Figure IV-2C).

Figure IV-2

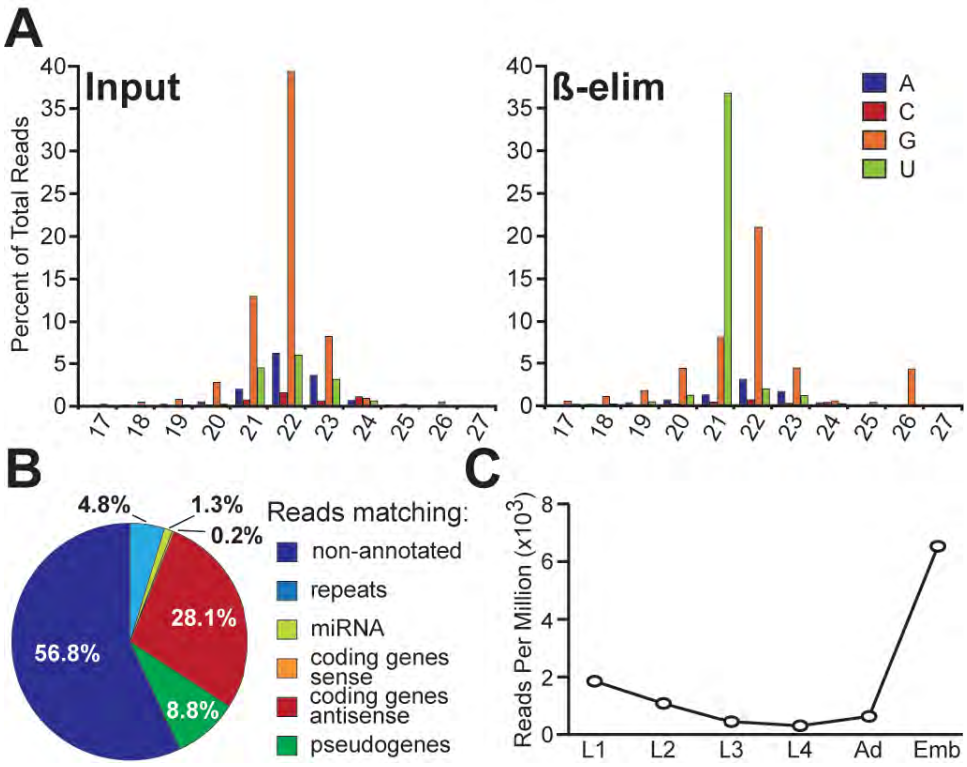

Figure IV-1. 26G-RNAs Cloned After β-elimination

(A) Plot of the first nucleotide composition and length of small RNA reads that were sequenced in Input and β-eliminated (β-elim) samples. (B) Pie chart indicating the assignment of genome-matching 26-nt reads according to genome annotation. (C) Expression profile of ERGO-1–dependent 26G-RNAs during development. Twenty six- nucleotide reads targeting 48 ERGO-1–dependent loci were extracted from deep-sequencing data generated by Batista et al. (16). Plot shows reads per million in each developmental stage.

Identification of Nonannotated 26G-RNA Loci.

The majority of 26G-RNAreads were derived from unique nonannotated genomic sequences. Despite the lack of annotation, these 26G-RNAs were in clusters and oriented on one strand as though antisense to an expressed transcript (Figure IV-3A). To further characterize these nonannotated genomic loci, 26G-RNA reads matching genome annotations were removed from the data and each chromosome strand was scanned using a 500-nt window to build and annotate 26G-RNA clusters (See Methods). This analysis defined 147 genomic loci with a 26G-RNA density of at least 10 reads per million (rpm) in our dataset (Table IV-S2). These clusters are much more extensive than recently reported clusters (Stoeckius et al., 2009) and appear to encompass complete transcription units that have not been annotated. Analysis of the ERGO-1 IP data revealed that 126 of these loci were enriched above a threshold of 1.5-fold in the IP relative to input (Figure IV-1C). By visual inspection, ~17 nonannotated 26G-RNA clusters appeared to extend from, or were very close to, annotated 26G-RNA genes, raising the possibility that these clusters may target incompletely annotated transcripts (Gu et al., 2009).

In total, 26G-RNAs targeting annotated and nonannotated loci defined a set of ≈180 26G-target loci with read densities >10 rpm. The genomic location of 26G-RNA targets appeared to be nonrandom. The most abundant 26G-RNA loci tended to map within 5 Mb of the chromosome ends (Figure IV-3A and Figure IV-

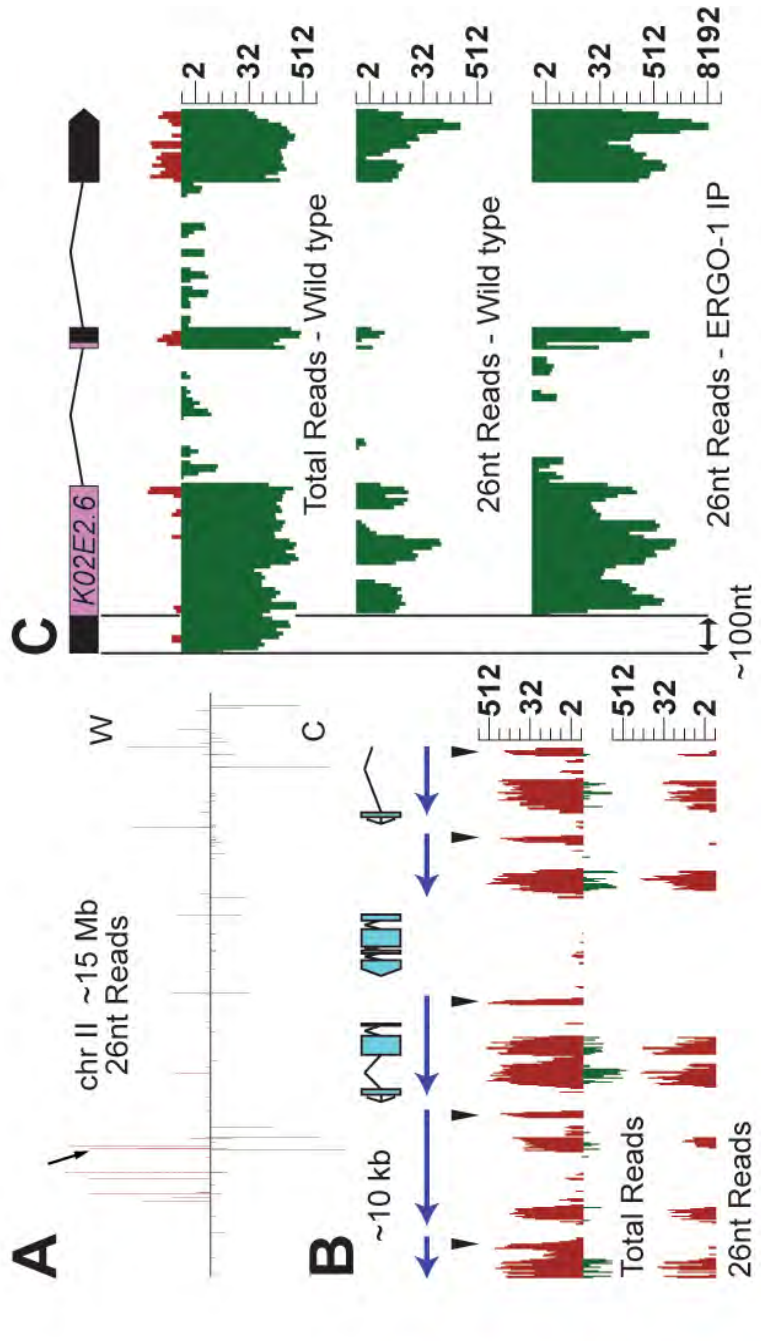
A
chr II ~15 Mb
26nt Reads
W
C
B
~10 kb
Total Reads
26nt Reads
512
32
2
C
K02E2.6
Total Reads - Wild type
26nt Reads - Wild type
26nt Reads - ERGO-1 IP
2
32
512
8192
~100nt

Figure IV-3

Figure IV-3. 26G-RNA Clusters are Targeted by 22G-RNAs

(A) Density profile of 26- nt reads along chromosome (chr) II, which is ~15 Mb. C, Crick; W, Watson. Arrow indicates location of 26G-RNA cluster shown in B.

(B) Density profile of small RNAs targeting an ~10-kb cluster in adults with embryos (gravid). Twenty-six-nucleotide read density shown in the Lower graph represents 26GRNAs. “Total reads” shown in Upper graph include both 26G- and 22G-RNA reads. Several peaks within this cluster lack 26G-RNA reads (arrowheads). The majority of reads from this cluster are on the Watson strand (red). Reads that map to the Crick strand (green) in this cluster are associated with 26G-RNA reads only. The blue arrows above density profiles predict transcription units based on the observed 26G-/22G-RNA patterns at ERGO-1 targets. The annotated gene predictions within this ~10-kb interval are illustrated above the density plots. Log2 scales are shown Right).

(C) Density profiles of small RNAs targeting K02E2.6 in wild-type adult and ERGO-1 IP datasets. Gene structure is shown at the top. Reads matching the Watson strand (red) are sense reads. Reads matching the Crick strand (green) are antisense. “Total Reads” include 26G- (“26-nt Reads”) and 22G-RNAs. 26G-RNAs are excluded from ~100 nt of the 5′-UTR of K02E2.6. A log2 scale is shown (Right).

S3). This pattern was strikingly different from the entire set of 22G-RNA loci, which were more evenly distributed throughout each chromosome (Claycomb et al., 2009; Gu et al., 2009). Interestingly, we observed groups of 26G-RNA targets that were oriented in tandem in the genome. For example, a ~10-kb region of chromosome II exhibited 5 different 26G-RNAloci (Figure IV-3B) that appear to be homologous, tandemly repeated units, suggesting that they are duplications that have diverged in sequence.

Essentially all 26G-RNA targets were also targeted by 22G-RNAs (Figure IV-3 and Tables IV-S3 and IV-S4) (Gu et al., 2009). In fact, 22G-RNAs were several orders of magnitude more abundant than 26G-RNAs at many loci in gravid adult samples. Interestingly, 26G-RNAs but not 22G-RNAs were excluded from the first ~100 nt at the 5′ end of target genes (Figure IV-3B). For example, 26G-RNAs mapped with similar density to both the exons and the 3′-UTR of K02E2.6, but were absent from the 5′-UTR (Figure IIV-3B). In contrast, the entire K02E2.6 transcript was targeted by 22G-RNAs, including the 5′-UTR. Sense reads were almost exclusively derived from regions targeted by 26G-RNAs and rarely derived from the regions targeted only by 22G-RNAs. This differential small RNA pattern was also observed at virtually all 26G-RNA target loci and helped us to define 26G-RNA loci (See Methods). Surprisingly, 22G-RNAs corresponding to more than half of the 26G-RNA loci were enriched in the somatic tissues of adult animals (Table IV-S3), where 26G-RNAs were relatively depleted (Figure IV-2C). Taken together, these observations are consistent with

the idea that targeting by 26G-RNAs drives the secondary biogenesis of 22G-RNAs both in the embryo and in subsequent developmental stages.

26G-RNA Biogenesis Is Dependent on Components of the ERI Pathway.

Having demonstrated that ERGO-1 interacts with 26G-RNAs, we next examined what other factors influence the expression of small RNAs targeting the ERGO-1 26G-RNA loci. Consistent with our deep-sequencing data, Northern blot analysis using a probe for siR26-1 (Ruby et al., 2006), targeting C40A11.10, revealed association of this 26G-RNA with the ERGO-1 IP complex in embryo lysates (Figure IV-4A). We found that siR26-1 expression was abrogated by mutations in the ERI pathway, including *ergo-1*, *rrf-3*, *rde-4*, and a viable, Eri allele of *dcr-1(mg375Eri)* (Pavelec et al., 2009) (Figure IV-4 A and B). The requirement for *rde-4* in the biogenesis of 26G-RNAs was independent of its role in the exo-RNAi pathway, as *rde-1* was not required for 26G-RNA biogenesis (Figure IV-4B).

Among the four *C. elegans* RdRP genes, only *rrf-3* was required for the expression of 26G-RNAs as determined by Northern blot analysis (Figure IV-4 A–C and Figure IV-S4). 26G-RNA expression was unaltered in both *rrf-1* and *rrf-2* mutants (Figure IV-4 A–C and Figure IV-S4). Although most 26G-RNAs were not detected in *ego-1* mutants (Figure IV-S4), this could reflect the fact that *ego-1* mutants are sterile and thus lack embryos, which is the stage when ERGO-1-dependent 26G-RNAs are most abundantly expressed (Figure IV-2C). As expected, the expression of siR26-1 was unaffected in mutants lacking WAGO-1

Figure IV-4

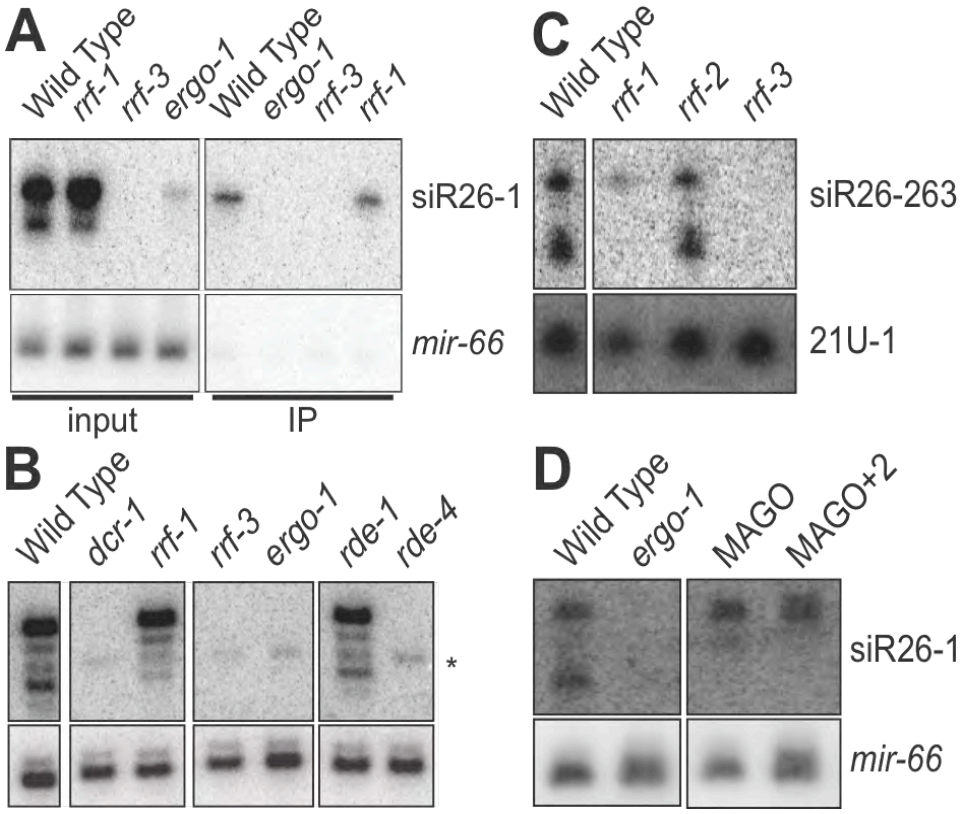
A
Wild Type
rrf-1
rrf-3
ergo-1
Wild Type
ergo-1
rrf-3
rrf-1
siR26-1
mir-66
input
IP
C
Wild Type
rrf-1
rrf-2
rrf-3
siR26-263
21U-1
B
Wild Type
dcr-1
rrf-1
rrf-3
ergo-1
rde-1
rde-4
*
D
Wild Type
ergo-1
MAGO
MAGO+2
siR26-1
mir-66

Figure IV-4. Genetic Requirements for 26G-RNA Expression

(A) Northern blot of siR26-1 in wild-type and mutant embryos. Input (Left) and ERGO-1 IP (Right) samples are indicated. Loading control: mir-66.

(B) Northern blot of siR26-1 in RNAi mutant (adult). The membrane was first hybridized to a mir-66 probe as a loading control, which could not be removed completely, and is indicated by the asterisk (*). All lanes shown were from the same membrane and exposure. Only relevant lanes are shown.

(C) Northern blot of siR26-263 in the indicated RdRP mutants (adult). Loading control: 21U-1. All lanes were from the same membrane and exposure. Only the relevant lanes are shown. (D) Northern blot of siR26-1 in mutants with multiple WAGO deletions (MAGO and MAGO+2). Loading control: mir-66. Adults were used. All lanes were from the same membrane and exposure. Only the relevant lanes are shown.

and other WAGO-class Argonautes that are required for, and interact with, 22G-RNAs (Figure IV-4D). Taken together, these findings indicate that 26G-RNA accumulation is dependent on components of the ERI pathway, but independent of several components of the exo-RNAi and 22G-RNA pathways.

ERI Pathway Stimulates 22G-RNA Accumulation.

As described above, ERGO-1-dependent 26G-RNA loci were also targeted by 22G-RNAs. Some of our probes (siR26-1 and siR26-263) detected both 26G-RNAs and a 22-nt RNA species, whereas the K02E2.6 probe detected only 22G-RNAs (Figure IV-4 and Figure IV-S4) (Gu et al., 2009). The ERI-pathway genes *rrf-3*, *ergo-1*, *rde-4*, and *dcr-1* were all required for expression of both the 26G- and 22G-RNA species at these target loci (Figure IV-4B and Figure IV-S4). In contrast, *rrf-1* and several *wago* Argonautes assayed were not required for 26G-RNA expression at these targets, but were required for 22G-RNA expression (Figure IV-4 C and D) (Gu et al., 2009). ERGO-1 still interacted with 26G-RNAs in *rrf-1* mutant embryos (Figure IV-4A). Together, these data suggest that the expression of ERGO-1 26G-RNAs is required for the RRF-1-dependent biogenesis of 22G-RNAs at these loci.

To examine the requirements of RRF-3, ERGO-1, and RRF-1 for 22G-RNA biogenesis on a genome scale, small RNAs were cloned and deep sequenced from *rrf-3*, *ergo-1*, and *rrf-1* mutants. Compared to our wild-type dataset (Gu et al., 2009), the overall distribution of small RNA classes was

Figure IV-5

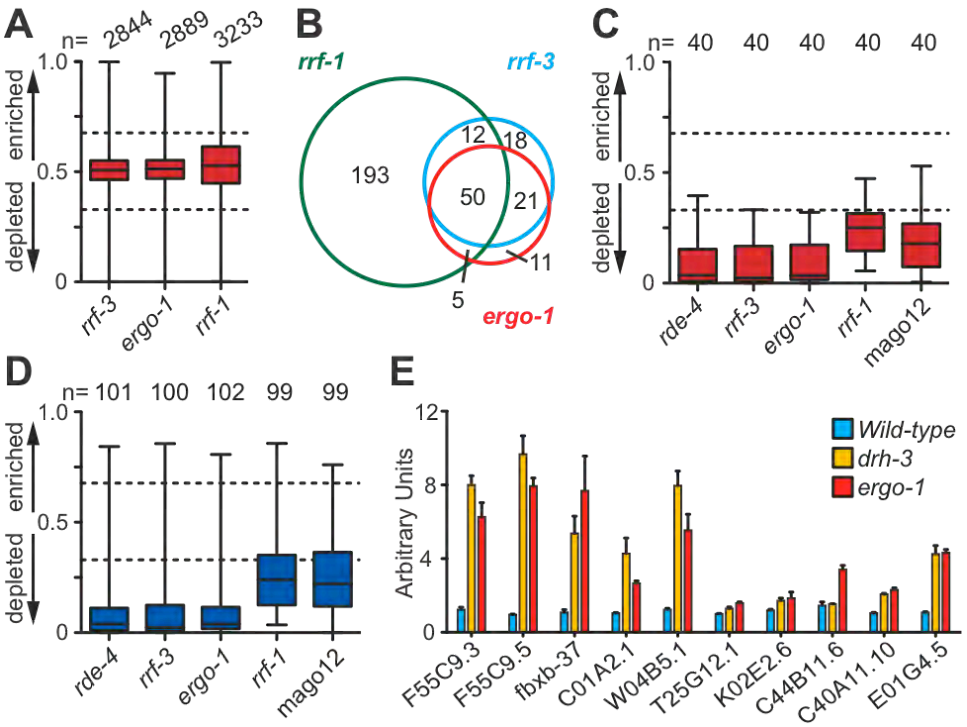
A
n= 2844 2889 3233
1.0
0.5
0
enriched
depleted
rrf-3
ergo-1
rrf-1
B
rrf-1
rrf-3
ergo-1
193
12
18
50
21
11
5
C
n= 40 40 40 40 40
1.0
0.5
0
enriched
depleted
rde-4
rrf-3
ergo-1
rrf-1
mago12
D
n= 101 100 102 99 99
1.0
0.5
0
enriched
depleted
rde-4
rrf-3
ergo-1
rrf-1
mago12
E
Arbitrary Units
12
8
4
0
Wild-type
drh-3
ergo-1
F55C9.3
F55C9.5
fbxb-37
C01A2.1
W04B5.1
T25G12.1
K02E2.6
C44B11.6
C40A11.10
E01G4.5

Figure IV-5. The Biogenesis of 22G-RNAs Targeting 26G-RNA Loci is Dependent on the ERI pathway.

(A) Overall enrichment or depletion of 22G-RNAs targeting n genes in each mutant, as described in Fig. 1C.

(B) Venn comparison of genes depleted of 22G-RNAs (≥2-fold) in indicated mutants (loci below the lower dashed line in A).

(C) Enrichment or depletion of 22G-RNAs derived from n 26G target genes, as described in Fig. 1C.

(D) Enrichment or depletion of 22G-RNAs derived from n 26G nonannotated clusters, as described in Fig. 1C.

(E) qRT-PCR analysis of 26G target genes expressionin *drh-3* (yellow) and *ergo-1* (red) mutants and wild type (blue). Error bars represent the standard deviation of the mean.

largely unaffected in each mutant (Figure IV-S5) and 22G-RNAs targeting annotated genes were unaffected as a whole (Figure IV-5A). However, *ergo-1* and *rrf-3* mutants were depleted of 22G-RNAs targeting 87 and 101 genes, respectively, and were highly overlapping with 71 genes in common (Figure IV-5B and Table IV-S4).

Within the set of 48 annotated genes targeted by 26G-RNAs, only 40 exhibited 22G-RNA levels that satisfied the rigorous criteria of 25 rpm in wild-type or mutant samples. All 40 of these genes were depleted of 22G-RNAs in both *rrf-3* and *ergo-1* mutants (Figure IV-5C). Of the ~100 nonannotated clusters that satisfy the 25-rpm cutoff, 90 were depleted of 22G-RNAs in both *ergo-1* and *rrf-3* mutants (Figure IV-5D). Consistent with our Northern data demonstrating a role for RDE-4 in 26G-RNA biogenesis, virtually all of the 26G-RNA loci (38 annotated genes and 92 nonannotated clusters) were depleted of 22G-RNAs in an *rde-4* mutant RNA sample (Figure IV-5 C and D) (Gu et al., 2009). Together, these data demonstrate that the ERI endo-siRNA pathway is required for the expression of both 26G-RNAs and 22G-RNAs at these ERGO-1 target loci.

RRF-1 and WAGOs Are Required for 22G-RNAs in the ERI Pathway.

In the *rrf-1* mutant, ~260 genes were depleted of 22G-RNAs (Table IV-S4), suggesting that RRF-1 is more broadly required for 22G-RNA biogenesis. Of the 71 loci depleted of 22G-RNAs in both *rrf-3* and *ergo-1* mutants, 62% were also RRF-1 dependent (Figure IV-5B). More significantly, 31 (78%) 26G-RNA target

genes were depleted of 22G-RNAs in the rrf-1 mutant dataset, all of which were depleted of 22G-RNAs in *rrf-3* and *ergo-1* mutants (Figure IV-5C). Of the genes that were not depleted of 22G-RNAs in the *rrf-1* mutant and that satisfy our 22G-RNA criteria for analysis, all but one was dependent on the RdRP EGO-1 (Claycomb et al., 2009). The remaining target was not depleted of 22G-RNAs in either *rrf-1* or *ego-1*, suggesting that RRF-1 and EGO-1 are redundant for 22G-RNAs targeting this gene. In addition, 76 nonannotated clusters were depleted of 22G-RNAs in the *rrf-1* mutant (Figure IV-5D), of which 74 were also dependent on *rrf-3*, *ergo-1*, and *rde-4*. Finally, we examined deep-sequencing data from a mutant bearing deletions in all 12 *wago* genes, MAGO12 (Gu et al., 2009). 22G-RNAs targeting 35 (88%) 26G-RNA target genes and 71 nonannotated clusters were depleted in the MAGO12 mutant (Figure IV-5 C and D). Thus, the biogenesis of 22G-RNAs at ERGO-1 target loci is dependent on *rrf-1* and WAGOs.

WAGO/22G-RNA pathways have been shown to silence their targets (Gu et al., 2009) and previous reports have indicated that the ERI pathway is an endogenous-silencing pathway (Duchaine et al., 2006; Pavelec et al., 2009). To look for silencing of the ERI targets, we used quantitative PCR after reverse transcription (qRT-PCR). Indeed, ERGO-1 26G-RNA targets were up-regulated in *drh-3* and *ergo-1* mutants (Figure IV-5E), both of which are required for the expression of 22G-RNAs targeting these loci (Figures. IV-4 and IV-5) (Gu et al., 2009).

DISCUSSION

Here, we have analyzed the genetic and small-RNA components of the ERI endogenous-RNAi pathway. We have identified Dicer-dependent 26-nt RNA species (26G-RNAs) as cofactors of the ERI pathway Argonaute ERGO-1. Components of the ERI complex, including DCR-1, the RdRP RRF-3, and the dsRNA-binding protein RDE-4, are required for the biogenesis of both 26G-RNAs and 22G-RNAs on ERI targets. A second RdRP, RRF-1, and additional WAGO Argonautes are required for the accumulation of 22G-RNAs (but not 26G-RNAs). These findings support a two-step model in which 26G-RNAs function upstream in the ERI pathway and drive the downstream production of 22G-RNAs (Figure IV-6).

During embryogenesis, 26G-RNAs drive the biogenesis of 22G-RNAs that persist into later developmental stages, when ERGO-1/26G-RNAs are present at very low levels (Figure IV-6). Thus, 22G-RNA expression may function to maintain silencing through a self-sustaining amplification loop in the absence of further 26G-RNA expression. This possibility is also consistent with the long-lasting silencing observed in response to exogenous dsRNA (Grishok et al., 2000), which involves a distinct upstream Argonaute but shares with the ERI pathway the RRF-1-dependent secondary 22G-RNA/ WAGO pathway (Gu et al., 2009; Yigit et al., 2006).

Figure IV-6

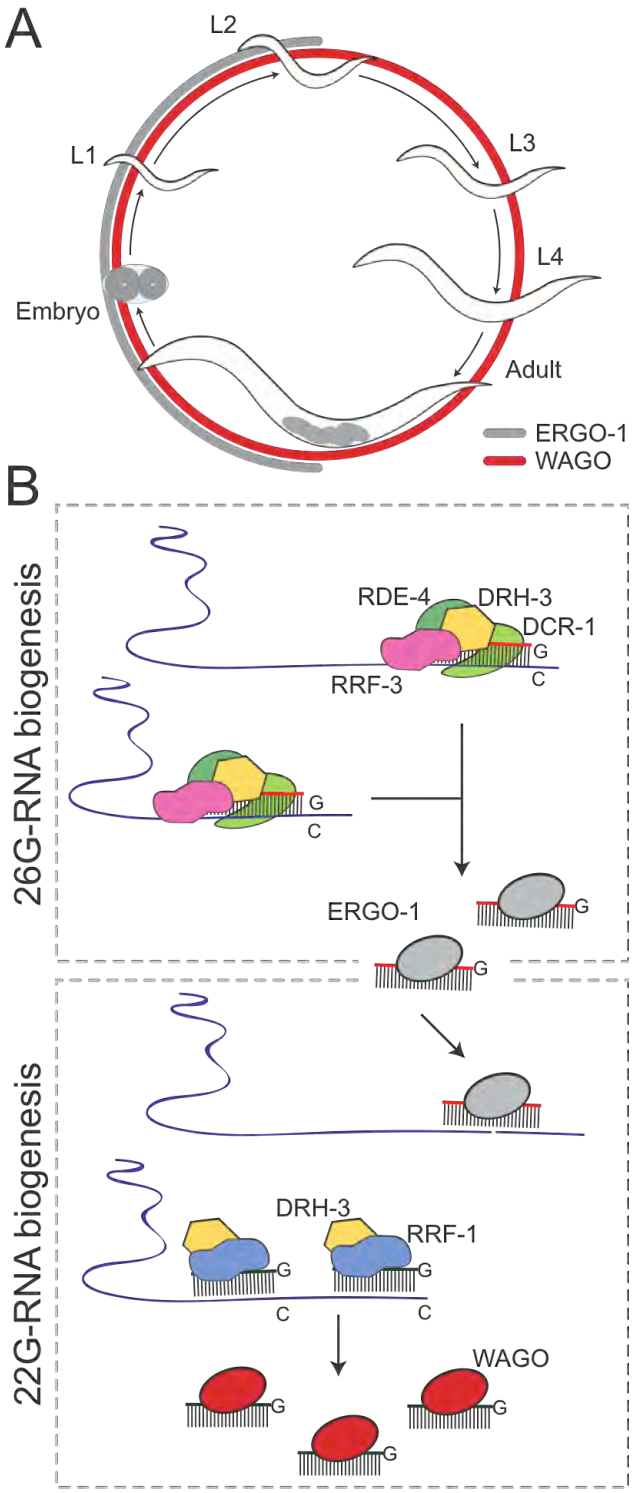
A
L2
L1
L3
L4
Embryo
Adult
ERGO-1
WAGO
B
26G-RNA biogenesis
RDE-4
DRH-3
DCR-1
RRF-3
G
C
ERGO-1
22G-RNA biogenesis
DRH-3
RRF-1
WAGO

Figure IV-6 Model for 26G-/22G-RNA biogenesis

(A) ERGO-1/26G-RNA (gray) and WAGO/22G-RNA (red) expression during the *C. elegans* life cycle: embryo, L1–L4 larval stages, and adult with embryos.

(B) Processive 26G-RNA biogenesis by the ERI complex and ERGO-1 loading. Targeting by ERGO-1 results in recruitment of the 22G-RNA RdRP machinery and WAGO loading.

In some cases, ERI pathway-dependent 22G-RNAs were derived from loci that were not significantly targeted by 26GRNAs. For example, the well-characterized X-cluster is an ERI dependent 22G-RNA locus (Ambros et al., 2003; Duchaine et al., 2006; Gu et al., 2009; Lee et al., 2006) but is not targeted by 26G-RNAs above the 10-rpm cutoff. However, the X-cluster shares a region of significant nucleotide identity with a 26GRNA- producing locus on chromosome 1 [linkage group (LG)I]. The region of identity in the LGI locus is targeted almost exclusively by 22G-RNAs. These and other similar findings are consistent with a transitive biogenesis of 22G-RNAs downstream of 26G-RNA targeting (Sijen et al., 2001).

The dependence of 26G-RNAs on components of the DCR-1/ ERI complex is consistent with a concerted mechanism of biogenesis (Colmenares et al., 2007; Duchaine et al., 2006; Lee and Collins, 2007). RNA duplexes generated by RRF-3 could be processed by DCR-1 to generate duplex siRNAs that are loaded into ERGO-1. However, several observations are not in agreement with 26G-RNA biogenesis via a direct, DCR-1-mediated cleavage of RRF-3-generated RNA duplexes. As noted by Ruby et al. (Ruby et al., 2006), the 5′G bias of 26G-RNAs is more consistent with direct synthesis by RdRP and does not fit well with the thermodynamic rules thought to govern the loading of Dicer products into Argonaute complexes (Khvorova et al., 2003; Schwarz et al., 2003). In addition, although sense small RNAs were cloned from 26G-RNA loci,

their length and position with respect to the corresponding 26G-RNAs are not consistent with Dicer processing. Although some phasing of 26G-RNAs was apparent (when the most abundant species were considered), for most loci we observed a highly overlapping distribution of 26G-RNAs. Rather than the phasing one might expect from Dicer-mediated processing, these findings are most consistent with cycles of 26-mer synthesis by RdRP initiating at multiple sites along the target mRNA. Further biochemical and genetic analyses will be necessary to understand the role of Dicer in the ERI pathway.

ergo-1 mutants are viable and exhibit no overt phenotypes other than an enhanced sensitivity to exogenous RNAi. Indeed, *ergo-1* mutants do not display the Him or male-specific, temperature sensitive sterile phenotypes associated with other Eri mutants (Pavelec et al., 2009). Instead, these male-specific functions depend on spermatogenesis expressed 26G-RNAs that engage two partially redundant Argonautes, ALG-3 and ALG-4 (Conine et al., 2010) *ergo-1* may function to regulate the exo-RNAi pathway in somatic tissues and/or may have other as yet undetected biological functions. Most ERGO-1 targets are not recognizable as genes and often reside in clusters of what appear to represent ancient duplications. Therefore, it is conceivable that the ERGO-1 pathway may function to buffer against deleterious affects arising from expression of these duplicated noncoding sequences. Whereas the specific biological function of the ERGO-1 pathway remains unclear, it provides a striking example of

interdependence and competition between Argonaute systems and points to the complexity and rapidly evolving nature of Argonaute /small-RNA networks.

MATERIALS AND METHODS

Worm Culture and Strains. *C. elegans* culture and genetics were essentially as described ((Brenner, 1974)). The Bristol N2 strain was used as the wild-type control. Alleles used are listed by chromosome: LGI, *sago-2 (tm894)*, *ppw-1 (tm914)*, *ppw-2 (tm1120)*, *avr-14 (ad1302)*, *rrf-1 (pk1417)*, *ego-1 (om97)*, *hT2[qIs48] (I;III)*, *C04F12.1 (tm1637)*, *rrf-2 (pk2040)*; LGII, *rrf-3 (pk1426)*, *C06A1.4 (tm887)*, *F58G1.1 (tm1019)*; LGIII, *dcr-1 (mg375Eri)*, *rde-4 (ne337)*; LGIV, *M03D4.6 (tm1144)*; and LGV, *ergo-1 (tm1860)*, *sago-1 (tm1195)*, *rde-1 (ne300)*, *avr-15 (ad1051)*, *glc-1(pk54)*.

Generation of ERGO-1 Antibodies. A C-terminal ERGO-1-specific peptide (CEVNKDMNVNEKLEGMTFV) was coupled to KLH and used to immunize four rabbits (Capralogics).

ERGO-1 Immunoprecipitation. Using a stainless steel dounce, proteins were extracted from embryos in cold lysis buffer [25mMHepes-KOH(pH7.4), 10 mM KOAc, 2 mM Mg(OAc)2, 100 mM KCl, 1% Triton X-100, 1 mM DTT] containing 1% SUPERase•In (Ambion), 1% Phosphatase Inhibitor Mixture 1 (Sigma), 1% Phosphatase Inhibitor Mixture 2 (Sigma), and Mini Protease Inhibitor Mixture (four tablets/ 25 mL buffer, Roche). Lysates were cleared at 20,000 X g for 10 min at 4°C and protein concentration was determined using the D_C Protein Assay (Bio-Rad). Immunoprecipitation of ERGO-1 was performed by incubating 45.6

mg of total protein extract, at a final concentration of 3.8 mg/mL, with 10 µL of ERGO-1 antiserum per mg of protein for 1 hour at 4 °C. A total of 20 µL of a 50% slurry of Protein A Sepharose beads (GE Healthcare) was added per mg of protein extract, and the mixture was incubated with rocking for 1 hour at 4°C. Immune complexes were washed three times with cold lysis buffer followed by three final washes with lysis buffer lacking Triton X-100 and DTT. Ten percent of the beads were set aside for Western blot analysis. RNA was extracted from the immune complexes (~100 µL packed protein A pellet) using 500 µL of TRI Reagent (MRC Reagents), according to the manufacturer's instructions. An independent replicate of ERGO-I immunoprecipitation from 9.6 mg of protein was repeated as above, at a final concentration of 4.6 mg/mL, using 0.05% Nonidet P-40 and 4µl of ERGO-1 antiserum per mg of protein.

Northern Blot Analysis. Small RNA Northern blots were performed as described (Chapter II Methods). Starfire probe sequences are provided in Table IV-S5.

Northern Blot Analysis following ERGO-1 Immunoprecipitation. Equal amounts of embryo protein lysate (9.6mg) were used to extract small RNAs from input and ERGO-1 immunoprecipiation as described (Chapter II Methods).

Radioactive Labeling of RNA Oligos. Individual 18-, 22-, and 26-nt RNA oligos

were radioactively labeled with T4 Polynucleotide Kinase (New England Biolabs), using 30 µCi of [γ-32P] ATP (Perkin- Elmer BLU502Z250UC). Excess free nucleotide was removed using Micro Bio-Spin 6 Columns (Bio-Rad no. 732-6221).

Western Blot Analysis. Proteins resolved by SDS/PAGE on 4–15% acrylamide gradient gels (Bio-Rad) were transferred to Hybond-C Extra membrane (GE Healthcare). Membranes were blocked with 5% nonfat dried milk in PBST [137mMNaCl, 2.7mMKCl, 10mM Phosphate (pH 7.4), 0.1% Tween-20] and incubated at 4 °C overnight with primary antibodies diluted in the same buffer. Rabbit anti-ERGO-1 was diluted 1:3,000 and mouse monoclonal antitubulin (clone DM1A, no. T6199, Sigma) was diluted 1:5,000. Membranes were washed in PBST, incubated with HRP-conjugated secondary antibody in PBST (1:5,000) at room temperature for 1 hour, and again washed with PBST. Chemiluminescence was performed using the Western Lightening ECL Kit (Perkin-Elmer) and visualized using a CCD camera and LAS-3000 Intelligent Dark-Box (Fujifilm).

Small RNA Purification and Cloning. Extraction of total RNA and enrichment for small RNA < ~200 nt were as described (Chapter II Methods).

Oxidation of Small RNA. Small RNA (~20 µg) was oxidized using 0.2 M NaIO4

(periodate) in 60 µL of 0.3 M borate buffer (pH 8.6) for 10 min at room temperature. Excess NaIO4 was destroyed by adding 2 µL of glycerol and incubating for 10 min at room temperature. Oxidized RNA was desalted using a Bio-spin 6 column (Bio-Rad) and precipitated with 4 vol of ethanol. Small RNA Cloning and Sequencing. RNAs (18 to 30 nt) were gel purified and cloned using a 5′ ligation-dependent protocol as described (Chapter II Methods). Some small RNA samples were pretreated with calf-intestine phosphatase and polynucleotide kinase (*rrf-3*, *ergo-1*, and *rrf-1*) or tobacco acid pyrophosphatase (β-elimination and input) to make the 5′ ends of 22G-RNAs available for ligation (Gu et al., 2009). cDNA libraries were sequenced by the University of Massachusetts (Worcester, MA) Deep Sequencing Core, using an Illumina Genome Analyzer II.

Real-Time PCR. qRT-PCR was performed as described (Chapter II Methods). cDNA was generated using 5 µg of total RNA, random hexamers, and SuperScript III Reverse Transcriptase (Invitrogen). The expression level of each target RNA was normalized to *gpd-2* or *act-3*. Primer sequences are provided in Table IV-S5.

Data Analysis. Small RNA sequences were processed and mapped to the *C. elegans* genome (Wormbase release WS192) as well as Repbase (13.07), using custom Perl scripts (Perl 5.8.6) as described (Chapter III Methods). Clusters were generated from 26G-RNAs that matched nonannotated genome

sequences, using the ERGO-1 IP dataset.

26G-RNA Clusters. Clusters were generated from 26G-RNAs that matched nonannotated genome sequences using the ERGO-1 IP dataset, using a custom Perl script as follows: A 500-bp window was used to scan the Watson or Crick strand of each chromosome. A cluster was initiated when a window included at least one nonannotated 26G-RNA and was extended as long as the next window had at least one 26G-RNA. The 3′ end of each cluster was further extended 1 kb to include a potential 22G-RNA locus. Each cluster was then refined visually using a generic genome browser. Adjacent 26G-RNAclusters were combined until a 22G-RNA-only locus was encountered or a cluster was split if a 22G-RNA-only locus was included. We analyzed 147 clusters, each with ≥10 26G-RNAs.

Supplemental Information

Figure IV-S1

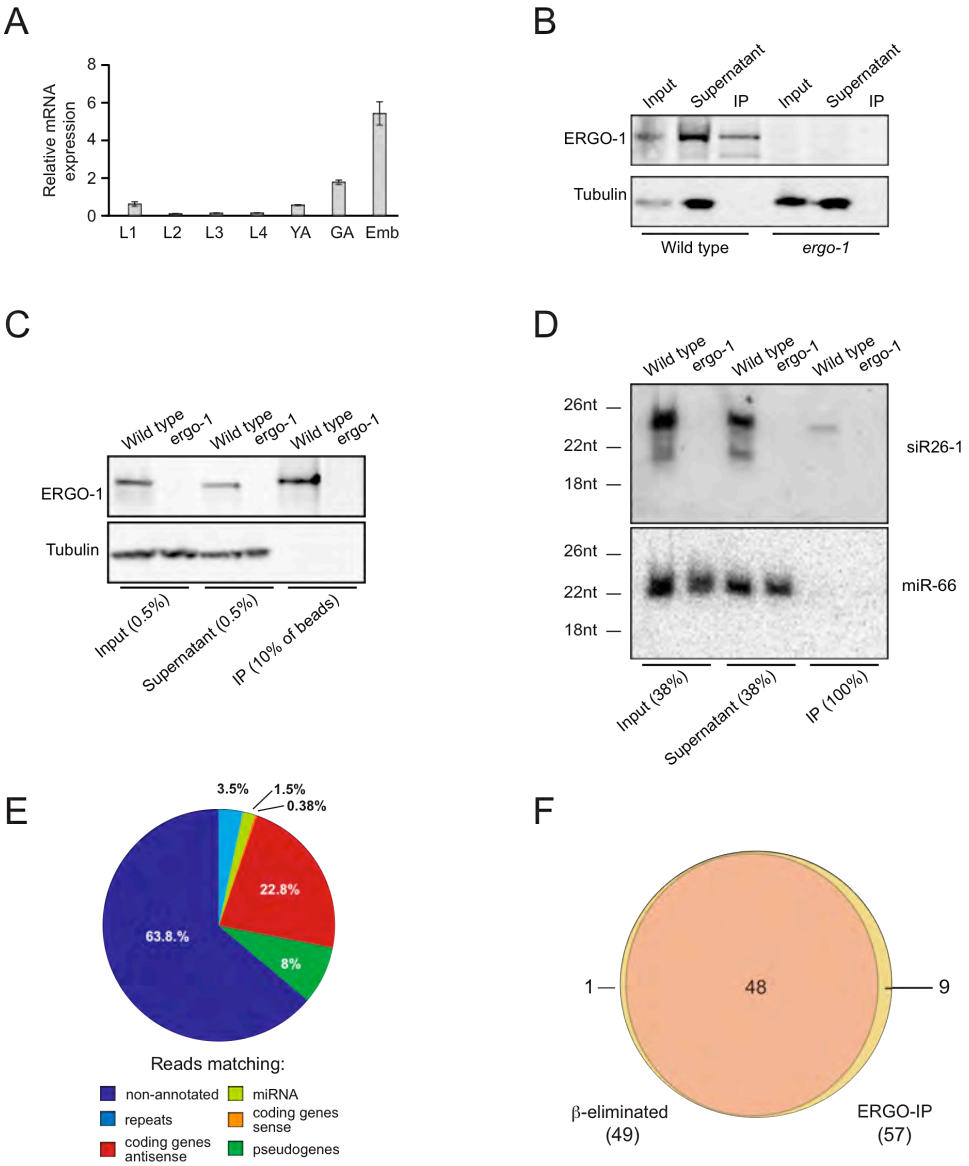
A
Relative mRNA expression
8
6
4
2
0
L1
L2
L3
L4
YA
GA
Emb
B
Input
Supernatant
IP
Input
Supernatant
IP
ERGO-1
Tubulin
Wild type
ergo-1
C
Wild type
ergo-1
Wild type
ergo-1
Wild type
ergo-1
ERGO-1
Tubulin
Input (0.5%)
Supernatant (0.5%)
IP (10% of beads)
D
Wild type
ergo-1
Wild type
ergo-1
Wild type
ergo-1
26nt
22nt
18nt
siR26-1
26nt
22nt
18nt
miR-66
Input (38%)
Supernatant (38%)
IP (100%)
E
3.5%
1.5%
0.38%
22.8%
63.8.%
8%
Reads matching:
non-annotated
miRNA
repeats
coding genes sense
coding genes antisense
pseudogenes
F
1
48
9
β-eliminated (49)
ERGO-IP (57)

Figure IV-S1. ERGO-1 expression and 26G-RNA analysis.

(A) Expression profile of *ergo-1* mRNA by qRT-PCR analysis. Error bars represent the standard deviation of the mean.

(B) Western blot of ERGO-1 after IP from wild-type or ergo-1 mutant embryonic lysates. This ERGO-1 IP from wild type was used for the IP cloning of ERGO-1-associated small RNAs

(C) Western blot of ERGO-1 IP. Independent IP of wild-type and ergo-1 embryos is shown, showing percentage of protein loaded in input, unbound supernatant, and IP fractions.

(D) Northern blot of small RNA extracted from ERGO-1 IP (C), showing percentage of small RNA loaded and probed with siR26-1 and control miR-66. Radiolabeled RNA oligos were used as size markers.

(E) Pie chart indicating the assignment of genome-matching 26-nt reads according to genome annotation.

(F) Venn diagram of genes with 26G-RNA reads enriched in a β-eliminated sample (49 genes) compared to those enriched in the ERGO-1 IP (57 genes). Only genes with 26G-RNAs ≥10 rpm and enriched ≥1.5-fold over untreated or input were used.

Fig IV-S2

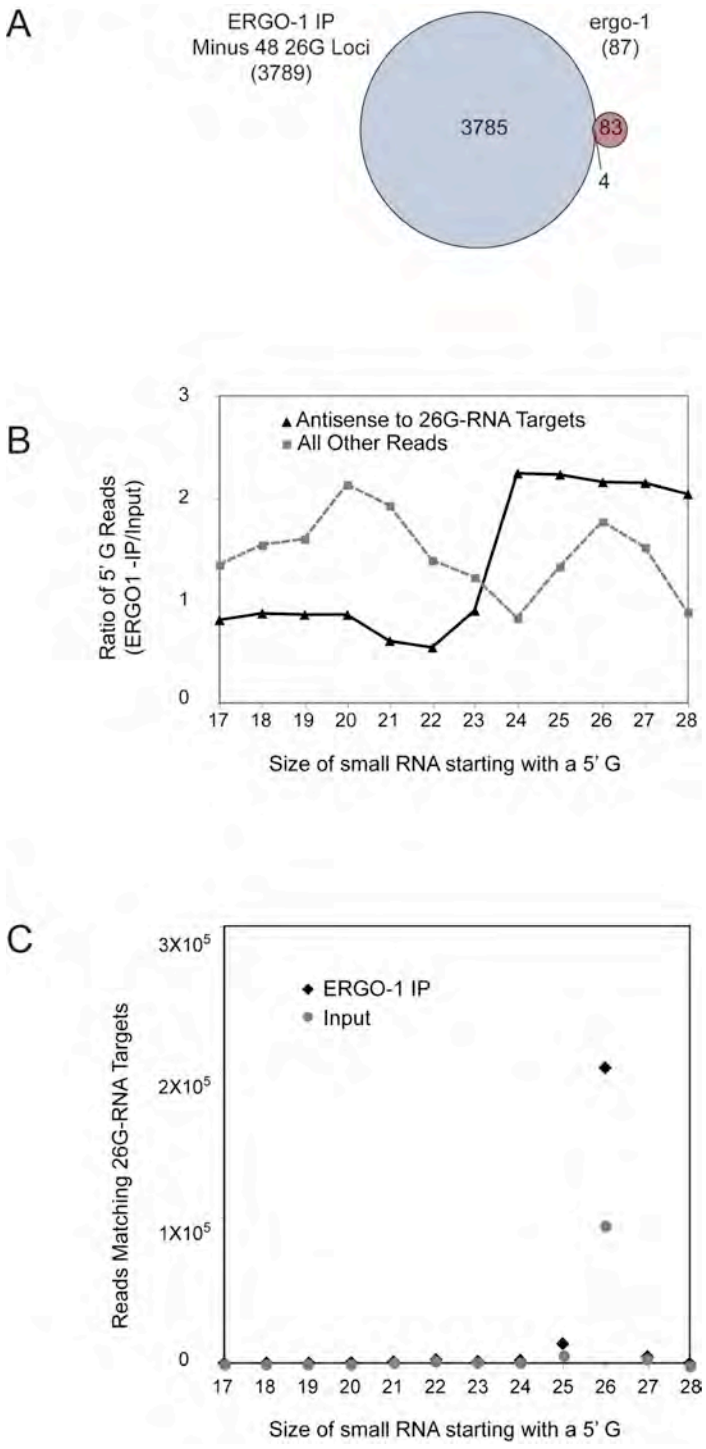
A
ERGO-1 IP
Minus 48 26G Loci
(3789)
ergo-1
(87)
3785
83
4
B
Antisense to 26G-RNA Targets
All Other Reads
Ratio of 5' G Reads
(ERGO1 -IP/Input)
0
1
2
3
17 18 19 20 21 22 23 24 25 26 27 28
Size of small RNA starting with a 5' G
C
ERGO-1 IP
Input
Reads Matching 26G-RNA Targets
0
1X10^5
2X10^5
3X10^5
17 18 19 20 21 22 23 24 25 26 27 28
Size of small RNA starting with a 5' G

Figure IV-S2. Analysis of 5′G Reads in ERGO-1 IP and Input Samples.

(A) Venn diagram of non-26G-RNA target genes with small RNAs enriched ≥2-fold in the ERGO-1 IP compared to the list of ERGO-1-dependent small RNA target genes from Table S4 (includes 26G-RNA targets). A minimum read cutoff was not used to generate the ERGO-1 IP enriched list.

(B) Enrichment profile of 5′G reads according to size. The enrichment of each size 5′G population is shown for reads derived from 26G-RNA target loci (solid black line with triangles) compared to all remaining reads (dashed gray line with squares). Reads were normalized to the sample size before determining the enrichment factor.

(C) Distribution of 5′G reads derived from 26G-RNA target loci. Read numbers were normalized to sample size.

Figure IV-S3

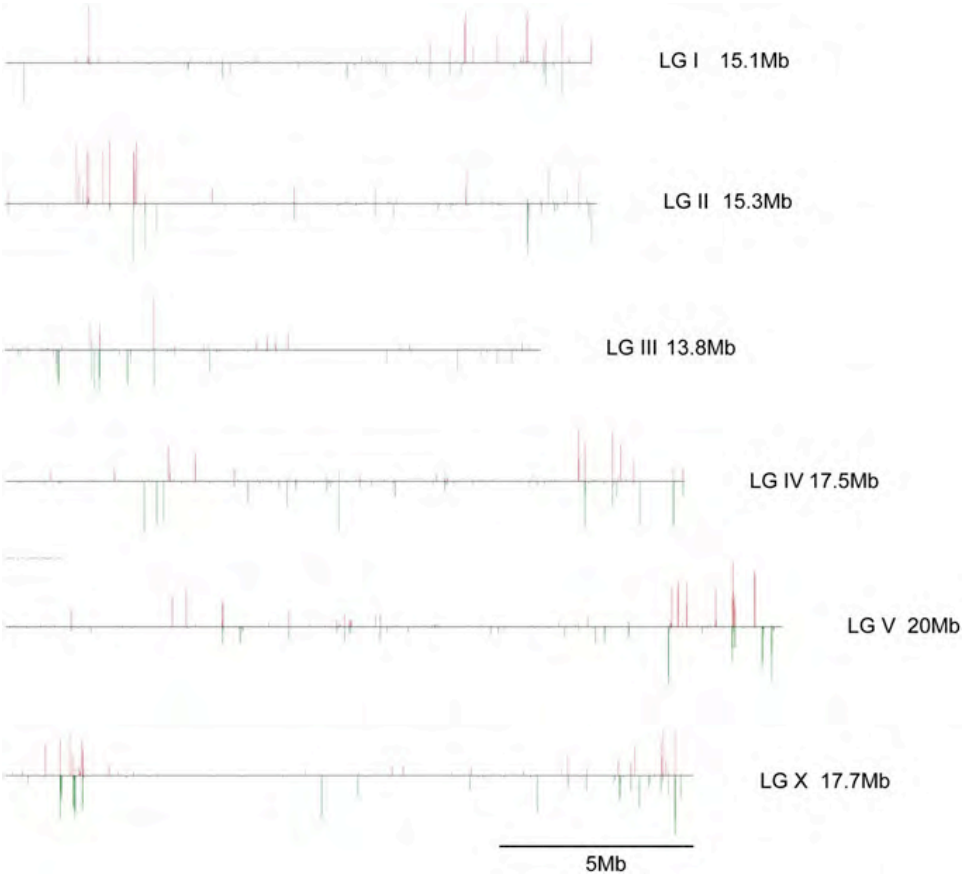

Figure IV-S3. Density Profile of 26-nt Reads Along Each *C. elegans* Linkage Group (LG)

Reads that map to the Watson strand are indicated in red and reads that map to the Crick strand are indicated in green. (Scale bar, 5 Mb.)

Figure IV-S4

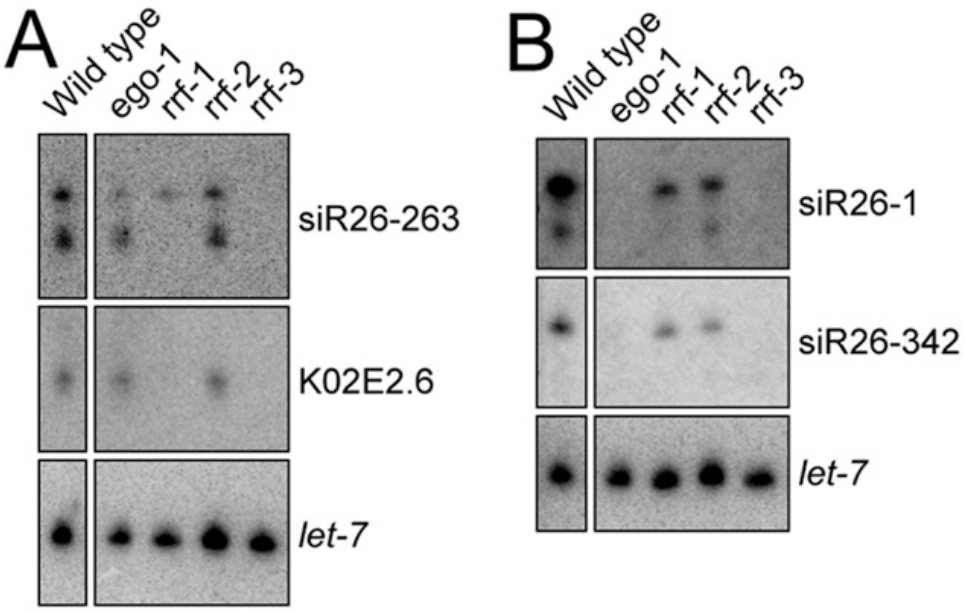

Figure IV-S4. 26G-/22G-RNA Analysis in RdRP Mutants

(A) Northern blot analysis of siR-263 (26G and 22G) and K02E2.6 (22G only) in wild-type and RdRP mutant animals. Loading control: let-7.

(B) Northern blot analysis of siR26-1 (26G and 22G) and siR26-342 (26G only). Loading control: let-7.

Figure IV-S5

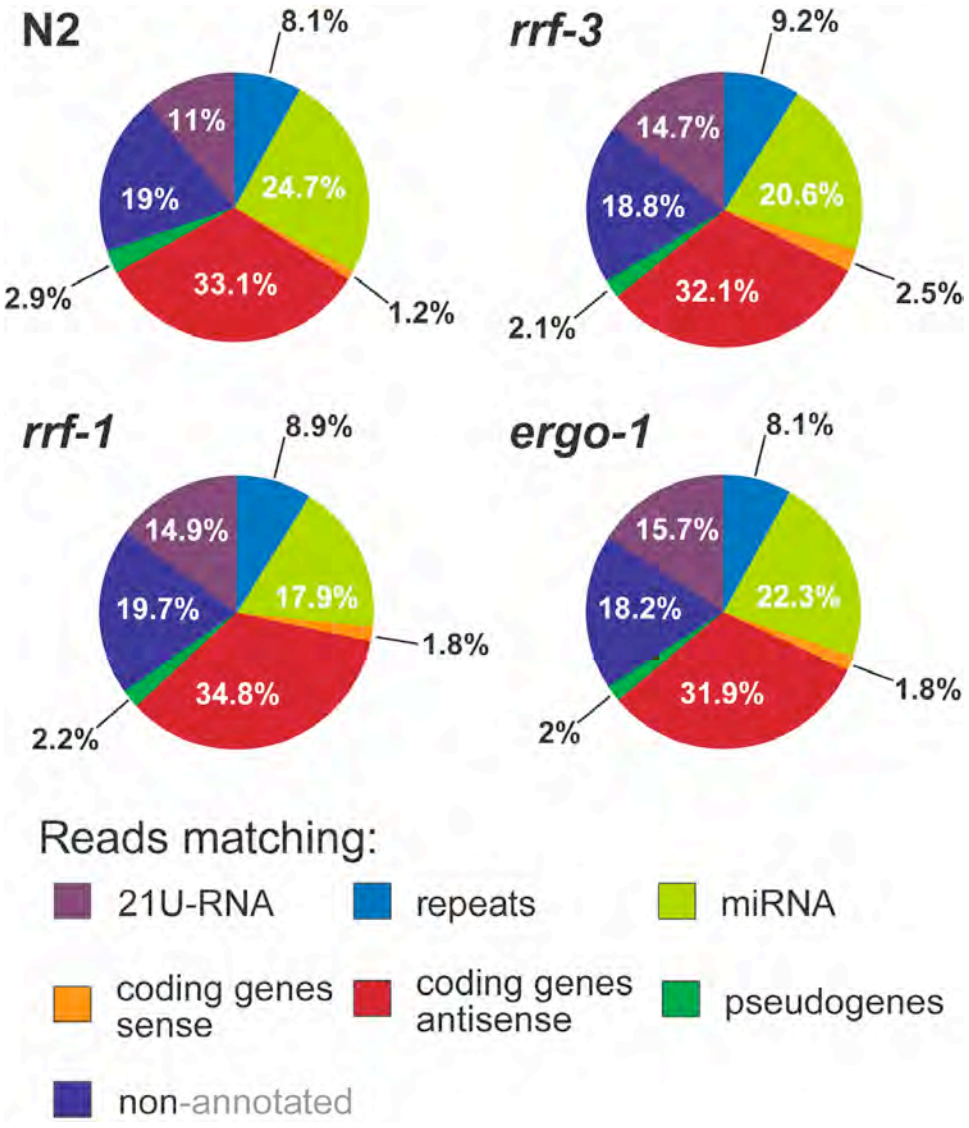

Figure IV-S5 Distribution of Reads that Match Indicated Genome Annotations

Sequenced in *rrf-1, rrf-3,* and *ergo-1* mutants compared to N2 (wild-type) control.

Table IV-S1

ß-elimination (49 genes)				ERGO-1 IP (57 genes)					
Gene Name	ß-eliminated Normalized (rpm)	Untreated Normalized (rpm)	Fold Enrichement	Gene Name	IP Normalized (rpm)	Input Normalized (rpm)	Fold Enrichement	48 genes in common between ß-elim and ERGO-1 IP	Annotation as predicted in WormBase Release WS213
F39F10.4	101.1762	5.0549	20.0156	Y57G11C.51	2762.151	536.4946	5.1485	C01A2.1	N/A
H09G03.1	18.2997	1.0832	16.8944	F49B2.3	56.9522	12.3256	4.6206	C02B8.2	F-box C domain
ZK380.5	68.2133	4.8743	13.9945	C39B5.2	23.9187	5.8335	4.1002	C11G6.2	DNA helicase PIF1/RRM3
Y82E9BR.6	14.6925	1.0832	13.5642	Y7A5A.8	29.492	7.3389	4.0186	C35D6.3	N/A
ZK402.2	22.4787	1.715	13.1071	Y82E9BR.20	313.7497	78.2819	4.0079	C36A4.11	N/A
C11G6.2	59.562	4.8743	12.2195	Y41C4A.17	13.9333	3.7636	3.7022	C40A11.10	N/A
W04B5.1	136.5732	11.3132	12.072	F52D2.5	26.0087	7.1508	3.6372	CD4.8	N/A
F31A9.2	56.3067	4.8141	11.6961	F58F9.1	18.5777	5.269	3.5259	E01G4.5	N/A
W05H12.2	305.4641	26.538	11.5104	fbxb-37	1467.6358	429.986	3.4132	E01G4.7	N/A
ZC132.4	27.0096	2.3469	11.5087	C35D6.3	246.1541	75.8356	3.2459	F31A9.2	N/A
fbxa-46	58.1543	5.0549	11.5046	Y43F8B.9	271.6984	83.9273	3.2373	F31A9.4	F-box C domain
Y57G11C.51	343.999	31.2318	11.0144	E01G4.5	1105.9519	343.7066	3.2177	F39E9.7	N/A
F52D2.6	323.5878	29.9681	10.7977	Y116F11A.1	2127.6075	663.0441	3.2088	F39F10.4	N/A
CD4.8	22.2147	2.0761	10.7002	CD4.8	130.1212	43.1397	3.0163	F52D2.6	N/A
H16D19.4	34.4879	3.2496	10.6131	F52D2.6	1217.8823	405.617	3.0025	F55A4.4	N/A
Y105C5A.14	383.4137	36.2867	10.5662	Y17D7B.4	262.6418	91.8307	2.8601	F55C9.3	N/A
F31A9.4	203.4081	19.4973	10.4326	ZK402.2	159.1487	56.4062	2.8215	F55C9.5	N/A
Y116F11A.1	336.1689	32.315	10.4029	ZK380.5	485.9996	176.4166	2.7548	H09G03.1	N/A
Y37E11B.2	709.8169	68.722	10.3288	Y43F8B.6	29.0276	10.538	2.7546	H16D19.4	DNA helicase PIF1/RRM3
Y116A8B.1	509.0482	50.368	10.1066	T05E12.8	1222.1784	445.793	2.7416	K02E2.6	Resembles retroviral aspartyl proteases
ZK402.3	40.837	4.1221	9.9069	F55C9.5	3934.2858	1442.382	2.7276	K06B9.6	N/A
C02B8.2	43.8137	4.5133	9.7078	Y71A12B.2	40.987	15.6187	2.6242	T05E12.8	N/A
K02E2.6	1177.8667	122.0387	9.6516	ZK402.5	190.1502	73.844	2.575	T08B6.2	N/A
E01G4.5	184.4046	19.1362	9.6364	T08B6.2	68.7374	27.4739	2.5019	T25G12.11	N/A
F55C9.5	1067.0128	114.8175	9.2931	Y71A12B.3	42.4964	17.0301	2.4954	W04B5.1	N/A
C40A11.10	1457.113	157.0616	9.2773	W05H12.2	1425.8361	576.3883	2.4737	W04B5.2	N/A
C01A2.1	134.4324	14.8035	9.0811	C01A2.1	548.2735	223.7433	2.4505	W05H12.2	N/A
F39E9.7	695.3003	78.3503	8.8743	W04B5.1	579.1202	236.508	2.4486	Y105C5A.14	N/A
ZK402.5	32.7136	3.7008	8.8395	F31A9.2	256.604	106.5713	2.4078	Y116A8B.1	N/A
C44B11.6	238.9518	27.4407	8.7079	F55A4.4	93.4689	38.9528	2.3995	Y116F11A.1	DNA helicase PIF1/RRM3
Y43F8B.9	43.2858	5.0549	8.5632	H16D19.4	76.6329	32.7429	2.3404	Y17D7B.4	N/A
T08B6.2	52.4357	6.138	8.5427	ZC132.4	189.26	80.9164	2.339	Y17D7C.1	N/A
C35D6.3	21.291	2.5274	8.424	F39F10.4	388.9699	166.7254	2.333	Y37E11B.2	N/A
Y17D7B.4	84.4601	10.4708	8.0663	K02E2.6	5391.704	2348.834	2.2955	Y41C4A.17	N/A
T05E12.8	128.4498	16.2478	7.9057	Y37E11B.2	3337.5169	1491.6532	2.2375	Y43F8B.9	N/A
Y71A12B.3	11.2613	1.4442	7.7974	Y7A5A.9	21.3643	9.5971	2.2261	Y57G11C.51	N/A
F55C9.3	1687.4428	218.4421	7.7249	ZK402.3	267.4411	124.307	2.1515	Y71A12B.2	N/A
F55A4.4	44.2536	5.777	7.6603	C02B8.2	431.5825	200.7856	2.1495	Y71A12B.3	N/A
Y17D7C.1	43.9896	5.777	7.6146	Y82E9BR.6	78.9551	36.8828	2.1407	Y82E9BR.20	N/A
K06B9.6	21.4669	2.8885	7.4319	C11G6.2	161.6257	75.7415	2.1339	Y82E9BR.6	N/A
Y71A12B.2	10.7335	1.4442	7.4319	F39E9.7	2699.2966	1276.033	2.1154	ZC132.4	iposase-encoded protein linked to pepti
E01G4.7	694.6845	93.5149	7.4286	F31A9.4	1255.7343	604.0504	2.0789	ZK380.5	N/A
W04B5.2	205.8715	28.5238	7.2175	T02G6.4	22.9899	11.2907	2.0362	ZK402.2	N/A
T25G12.11	117.8922	16.9699	6.9471	Y116A8B.1	2125.3434	1052.4779	2.0194	ZK402.3	N/A
Y82E9BR.20	37.4792	5.4159	6.9202	C40A11.10	5549.6142	2789.7344	1.9893	ZK402.5	N/A
C36A4.11	16.5401	2.5274	6.5442	F55C9.3	6370.7471	3342.9768	1.9057	fbxa-46	Cyclin-like F-box
fbxb-37	136.8958	20.9416	6.537	W04B5.2	702.9326	371.8391	1.8904	fbxa-52	DUF38
fbxa-52	22.0828	4.1522	5.3183	fbxa-52	42.3803	23.1459	1.831	fbxb-37	Cyclin-like F-box
Y41C4A.17	10.3816	2.5274	4.1076	F52G3.4	218.8682	122.5037	1.7866		
				Y105C5A.14	1584.9074	888.481	1.7838		
				E01G4.7	2877.4487	1625.1026	1.7706		
				fbxa-46	204.8188	120.6219	1.698		
				H09G03.1	19.2743	11.3848	1.693		
				T25G12.11	805.8064	481.5467	1.6734		
				Y17D7C.1	120.6387	73.8597	1.6333		
				C36A4.11	37.9681	23.7104	1.6013		
				K06B9.6	40.7548	26.7212	1.5252		

Table IV-S1
Genes Targeted by 26G-RNAs Enriched in ERGO-IP and ß-elimination

Chapter V

General Discussion

Discussion and Future Directions

In this thesis, I have described the identification and characterization of two classes of endogenous small RNAs that are 22nt and 26nt in length that start with a 5´ G, called 22G-RNAs and 26G-RNAs, respectively. The 22G-RNAs and 26G-RNAs are distinct from other previously identified endogenous small RNAs, such as miRNAs and piRNAs. Unlike miRNAs and piRNAs, the 22G- and 26G-RNAs are produced by the activity of distinct RdRPs and they target a diverse set of endogenous loci, including annotated protein coding genes, transposable elements, pseudogenes, and nonannotated (or cryptic) loci. The 22G-RNAs rely on the activity of the partially redundant RdRPs EGO-1 and RRF-1, are DICER independent, and possess a 5´ triphosphate. The 26G-RNAs depend on the RdRP RRF-3 and DICER for their production.

Although it is unclear how exactly these small RNAs are processed into 26-nt long species, it is clear that DICER is required for their biogenesis. Complete loss or mutation of the helicase domain of DICER leads to a loss of the 26G-RNAs (Welker et al., 2010). Additionally, the 26G-RNAs possess a hallmark of DICER processing, 5’ monophosphate (Ruby et al., 2006), despite their initial production by RRF-3. These observations support a model whereby RRF-3 synthesizes dsRNA longer than 26nt, using target transcripts as a template, and that this dsRNA is recognized by the ERI complex and processed into 26G-RNAs by DICER. The observation of low abundance small RNA species

complementary to 26G-RNAs (i.e. sense) that possess a 3’ 3nt overlap is consistent with this model.

Using a combination of deep sequencing, molecular, genetic, and cell biology studies, at least two populations of 22G-RNAs were identified that interact with distinct Argonautes. First, the WAGO-1 associated 22G-RNAs target protein coding genes, repetitive elements, and cryptic loci in the germline to regulate target RNA expression. Second, CSR-1 interacts with a set of 22G-RNAs that target germline-expressed protein coding genes. However, the 22G-RNAs that function in the CSR-1 pathway do not regulate target transcript levels, and instead may modulate chromosomal domains that are essential for chromosome segregation.

This work also demonstrated that the helicase, DRH-3 and the tudor-domain protein, EKL-1, along with functionally redundant RdRPs, RRF-1 and EGO-1, form a core complex that is required for the biogenesis of nearly all 22G-RNAs. Surprisingly, EGO-1 and RRF-1 are redundant in the germline for the biogenesis of WAGO-associated 22G-RNAs, which is consistent with the observation that *ego-1* mutants were sensitive to RNAi targeting some germline genes. However, EGO-1 alone is required for the biogenesis of CSR-1 22G-RNAs, and loss of *ego-1* leads to germline RNAi deficiency, sterility and embryonic lethality.

The work presented here suggests the existence of at least three distinct 22G-RNA pathways: the CSR-1 pathway and two WAGO pathways. The WAGO

22G-RNAs can be distinguished by those that are dependent on DICER and those are independent of DICER. The DICER-dependent 22G-RNAs are also dependent on the ERI pathway, including the RdRP RRF-3. Interestingly, the ERI loci are not only targets of 22G-RNAs, but are additionally targeted by 26G-RNAs that interact with the Argonaute ERGO-1. The production of ERI-dependent 22G-RNAs is dependent on the production of the 26G-RNAs that target these same loci.

These data indicate that the ERI pathway is a two-step RdRP-dependent pathway that requires distinct RdRPs and Argonaute proteins at each step. The biogenesis of 26G-RNAs is dependent on RRF-3 and DICER and is proposed to be the primary step in this pathway. The 26G-RNAs interact with the Argonaute ERGO-1 and are thought to guide ERGO-1 to the complementary mRNA. ERGO-1 possesses the DDH catalytic residues required for Slicer activity and is predicted to mediate mRNA cleavage. This cleavage event may facilitate the recruitment of an RdRP complex, consisting of RRF-1 or EGO-1, which produce 22G-RNAs. The 22G-RNAs interact with the functionally redundant set of WAGO Argonautes in this downstream, secondary step. Remarkably, the Eri-dependent 22G-RNAs are maintained throughout development in somatic tissues, whereas ERGO-1 and its 26G-RNA cofactors are robustly expressed during embryogenesis. Therefore, after initiation of the ERGO-1-dependent primary phase, the 22G-RNA/WAGO step would provide maintenance of silencing throughout development.

Loss of genes required specifically in the Eri endogenous silencing pathway (ERI-1, ERI-3, DCR-1, ERI-5, RRF-3) leads to temperature-sensitive sterility, which results from defective sperm (Conine et al., 2010; Duchaine et al., 2006; Han et al., 2009; Kim et al., 2009; Pavelec et al., 2009). The sterility of *eri* mutants can be rescued by mating *eri* mutant hermaphrodites with wild-type males. In addition to temperature-sensitive sterility, *eri* mutants exhibit a high incidence of males (Him) phenotype and an enhanced sensitivity to exogenous RNAi. Interestingly, some Eri mutants do not exhibit the Him or sterile phenotypes (Pavelec et al., 2009; Sijen et al., 2001). One example of an *eri* gene that does not display explicit developmental defects is *ergo-1*. Several alleles of *ergo-1* have been isolated and none display any obvious growth or fertility defects, even at increased temperatures, nor do they display a Him phenotype (Pavelec et al., 2009). The developmental expression pattern of ERGO-1 is consistent with a lack of germline defects in *ergo-1* mutants, as ERGO-1 is most strongly expressed in embryos.

In contrast to *ergo-1* mutants, loss of the redundant Argonautes *alg-3* and *alg-4* leads to defects in both spermatocyte cell division and activation of mature spermatids, but does not result in an *eri* phenotype (Conine et al., 2010; Gent et al., 2009; Pavelec et al., 2009). Consistent with a role in spermatogenesis, ALG-3 and ALG-4 are strongly expressed in the germline when sperm are produced (Conine et al., 2010). Although ERGO-1, ALG-3 and ALG-4 exhibit different aspects of the *eri* pathway phenotype, both rely on the activities of the shared

core ERI complex consisting of RRF-3, DICER, ERI-1, ERI-3, and ERI-5, mutants of which all display both enhanced RNAi and fertility defects.

Recent studies have described two distinct subclasses (Class I and Class II) of germline-enriched 26G-RNAs in *C. elegans* with differential temporal and spatial expression profiles (Han et al., 2009). Class I specific 26G-RNAs are enriched in sperm and regulate genes required for spermatogenesis. Class II specific 26G-RNAs are enriched during oogenesis and embryogenesis and may regulate maternally-deposited mRNAs during development (Han et al., 2009, Stoeckius et al., 2009). Consistent with our findings that ERGO-1 is required for 26G-RNAs and is expressed during embryogenesis, *ergo-1* mutants display defects in expression of these Class II but not Class I 26G-RNAs targets (Conine et al., 2010; Han et al., 2009; Pavelec et al., 2009). The expression of Class I 26G-RNAs has been shown to be dependent on the Argonautes ALG-3 and ALG-4 (Conine et al., 2010; Han et al., 2009). RRF-3 and ERI-1 are required for the expression of both Class I and II 26G-RNAs, again providing support that the ERI complex is required for upstream events in these pathways (Duchaine et al., 2006; Yigit et al., 2006).

This work raises a number of questions about the mechanism of 26G-RNA biogenesis and the function of the ERGO-1 26G-RNA pathway. 1.) What are the molecular signatures that define a 26G-RNA target? 2.) How does the ERI complex find a target? 3.) What is the mechanism of 26G-RNA biogenesis? 4.) What is the biological role of the ERGO-1 26G-RNA pathway? In the following

sections I will discuss several hypotheses and experiments to address some of these questions and gain insight into biological functions of the ERGO-1 26G-RNA pathway.

Function of ERGO-1 Pathway

It is intriguing that no overt phenotype is observed in *ergo-1* mutants or in the absence of ERGO-1 associated 26G-RNAs despite evidence that this pathway is involved in down-regulating target gene expression. ERGO-1 orthologs are present in several related nematode species (*C. briggsae, C. brenneri, C. japonica* and *P. pacificus*) and 26 nt small RNAs are expressed in C. briggsae embryos (J. Claycomb, personal communication). However, the 26G-RNA target loci appear to be species-specific (J. Claycomb, personal communication). These preliminary observations suggest that the ERGO-1 pathway may be functionally conserved but target different loci in each nematode species. Although it is possible that another Argonaute may function redundantly with ERGO-1, several observations argue against this possibility. For instance, both 26G-RNAs and associated 22G-RNAs are severely depleted in the *ergo-1* mutants, suggested that no other Argonaute can compensate for ERGO-1 function. A second possibility is that there is an entire endogenous silencing pathway (and not just an Argonaute) that functions redundantly with respect to regulating 26G-RNA Class II embryonic pathway targets. Additionally, the ERGO-1 pathway could be involved in a genetic buffering mechanism that is required for an environmental

stress response, which would not be revealed or observed under normal laboratory conditions.

To identify pathways that function redundantly with ERGO-1, forward genetic screening could be performed to identify mutants that are inviable or sick in the absence of ERGO-1 (a “synthetic lethal” screen). An alternative approach would be to perform a genome wide RNAi screen, in the same genetic background, to look for synthetic phenotypes using the publicly available RNAi feeding library (Fraser et al., 2000; Kamath and Ahringer, 2003). There are several advantages of this method, including that it would enable rapid, systematic identification of genes that are synthetic lethal with *ergo-1*, thus eliminating the need for mapping.

A Role of ERGO-1 26G-RNA Pathway in Regulating Chromosome Structure and Integrity?

The majority of ERGO-1 26G-RNAs derive from nonannotated intergenic regions that map to within 5 Megbases (Mb) of the ends of each chromosome arm. Recently, these chromosomal regions have also been shown to correlate with enriched levels of trimethylation at histone 3 lysine 9 (H3K9) (Gu and Fire, 2010), a histone modification that is generally associated with inactive chromatin (Kouzarides, 2007).

In many organisms, Argonaute proteins have been shown to play a role in chromatin remodeling events, often leading to the accumulation of H3K9

methylation (Djupedal and Ekwall, 2009; Verdel et al., 2009). Could the role of the ERGO-1 26G-RNA pathway be to regulate chromatin domains required for the establishment (or remodeling) of heterochromatin during embryonic mitosis?

Evidence for a role of the Eri pathway in the nucleus comes from studies of the WAGO Argonaute NRDE-3 (Guang et al., 2008). Over 70% of the 22G-RNAs that associate with NRDE-3 are also directed against ERGO-1 26G-RNAs targets, and mutations in *nrde-3* result in the increased expression of ERGO-1 26G-RNA target RNAs. For example, both the pre-mRNA and the mRNA levels of the ERGO-1/26G-RNA target, *E01G4.5*, are upregulated in *nrde-3* mutants (Guang et al., 2008). Localization studies of NRDE-3 demonstrated that NRDE-3 translocates from the cytoplasm to the nucleus, and is dependent on binding to 22G-endo-siRNAs (Guang et al., 2008). Evidence that NRDE-3 acts at the level of, or downstream of, endo-siRNA production is based on the observation that NRDE-3 nuclear localization fails to occur in *eri* mutants such as *eri-1* and *ergo-1* (Guang et al., 2008). These observations raise the possibility that the ERGO-1/NRDE-3 branch of the ERI pathway could be involved in regulating nuclear organization and even chromatin modification, such as H3K9 methylation, in embryos.

Perhaps the ERGO-1 branch of the Eri pathway is required to regulate and stabilize chromatin domains near the ends of chromosome arms, which may be inherently less stable due to their repetitive nature. Somatic tissue in *C. elegans* may be refractory to such inherent chromosome end instability due to

the redundancy of DNA repair pathways, including homologous recombination (HR), non-homologous end joining (NHEJ), single-strand annealing (SSA), and a recently describes alternative ENJ (alt-ENJ) pathway (Pontier and Tijsterman, 2009). Remarkably, loss of HR, NHEJ, and SSA pathways together leads to no apparent developmental defect (Pontier and Tijsterman, 2009). Likewise, if the ERGO-1 pathway also played a role in chromosome end stability, either as a part of one of these pathways or in an independent pathway, it may not come as a surprise that loss of *ergo-1* does not result in an overt developmental defects. It would be relatively simple to test the involvement of ERGO-1 components in DNA repair, by using existing reporter assays to directly quantify repair in *ergo-1* pathway mutants or by creating genetic mutants between the *ergo-1* pathway and the HR, NHEJ, and SSA pathway triple mutant.

Although it has been shown that NRDE-3 localizes to the nucleus, and we have determined that the developmental expression timing of NRDE-3 and ERGO-1 is identical (J. Claycomb, personal communication), the localization pattern of ERGO-1 has not yet been determined, despite our best efforts. Immunolocalization experiments and studies of *ergo-1::gfp* localization are necessary to determine the subcellular site of ERGO-1 accumulation. Furthermore, performing ChIP for NRDE-3 and ERGO-1 will be useful in determining whether these Eri Argonautes directly associate with chromatin at their target 26G loci. In addition, we can examine histone modifications present at 26G-RNA target loci using a modified histone candidate ChIP approach or by

using bioinformatics to correlate existing modified histone ChIP data from the ModENCODE project (Celniker et al., 2009) with 26G-RNA loci throughout the genome. After identifying which histone modifications are present at the 26G-RNA target loci, we can determine the contribution of the ERGO-1 pathway to the deposition of these modifications by performing ChIP experiments in *ergo-1* pathway mutants.

What are the molecular signatures that define a 26G-RNA target?

Many of the 26G-loci appear to be either mis-annotated genes or loci that altogether lack an annotation. Interestingly, when we examine the small RNA density profiles, we observe that there is an approximately 100nt region at the 5´ end of nearly all ERGO-1 target loci which is exclusively targeted by 22G-RNAs and devoid of 26G-RNAs. This profile is in contrast to the remainder of the length of target transcripts, which are targeted by both 22G-RNAs and 26G-RNAs. This observation begs the question as to why 26G-RNAs appear to be excluded from the 5´ ends of these loci? Perhaps the simplest explanation is that the 26G-RNA loci are preferentially expressed as different mRNA isoforms that vary at their 5´ region or may have distinct transcription start sites. For example, a shorter form expressed during embryogenesis might be used as a template for 26G-RNA biogenesis, while a second, longer form expressed during later stages of development might serve as a template for 22G-RNA biogenesis. This possibility could be tested by simply performing 5´ RACE (rapid amplification of cDNA ends)

on total RNA isolated from developmentally staged samples or by performing RNA in situ hybridizations to determine specific RNA expression patterns.

Recently, we have found that the 100 nt region targeted by 22G-RNAs, but not 26G-RNAs, appears to be conserved between 26G-RNA loci as three motifs with a loose consensus (Figure V-1). One interpretation is that the sequence motifs define a molecular signature that is necessary for identifying these specific loci as templates for the production of 26G-RNAs. Alternatively, the motif could function as a cis-regulatory element or binding site for an unidentified RNA-binding complex that prohibits 26G-RNA but not 22G-RNA synthesis.

To determine whether these motifs are required for 26G-RNA biogenesis, genomic deletions of the motifs at specific 26G-RNA loci and then mutants can be assayed for the production of the 26G-RNAs from that locus. In addition, reporter constructs driving the expression of visual maker, such as GFP, could be used to test if these motifs are necessary and sufficient to trigger 26G-RNA and 22G-RNA production. To identify proteins that binds to the motif, we could perform RNA pull-down experiments or RNA-based Affinity Chromatography followed by mass spectrometry (Tian, 2002).

Final Conclusion

These studies set the stage for future experiments aimed at answering many unresolved questions about 26G-RNA function and biogenesis. Learning more

Figure V-1.

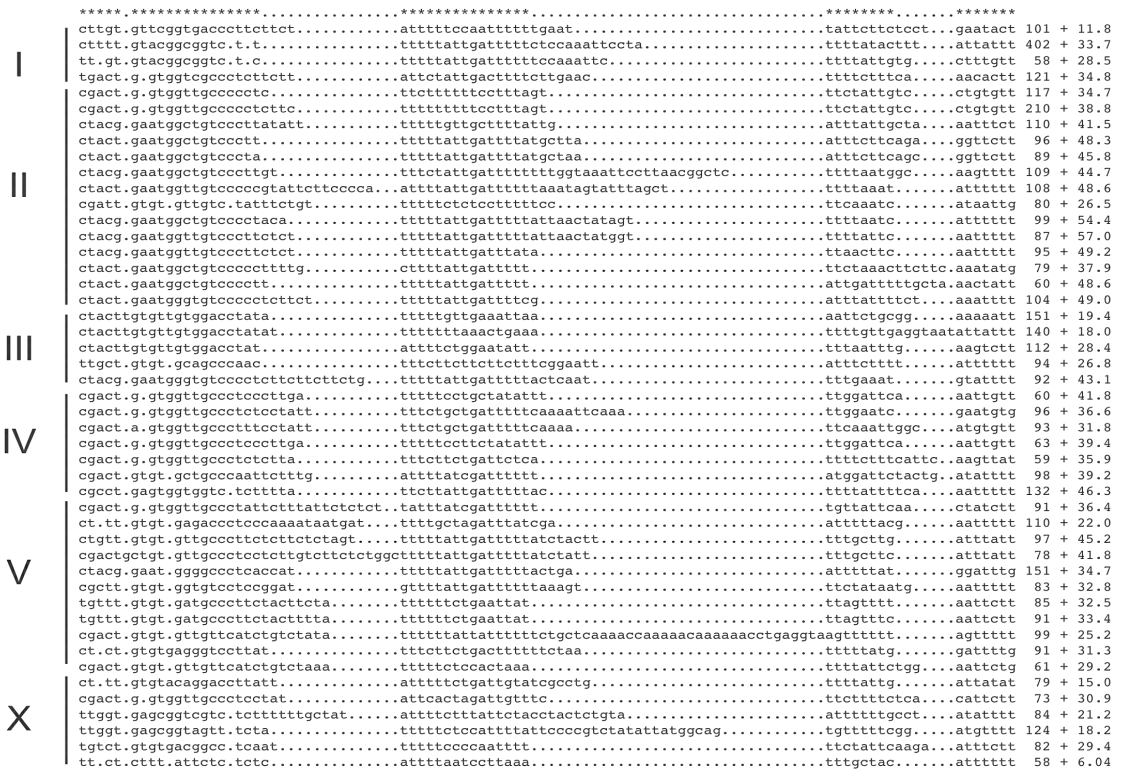

Figure V-1 Sequence alignment of 5’ Regions of Top ERGO-1 Associated 26G-RNA Target Loci According to Chromosome Distribution

about the mechanism of how the RNAi machinery, specifically the ERI complex, is recruited to the target RNAs will be extremely informative in understanding the mechanism of 26G-RNA biogenesis. Furthermore, determining how the 26G-RNA targets are distinguished from all of the other cellular endogenous transcripts that are also regulated by the endogenous small RNAs will also be a key to understanding the biological function(s) of this pathway.

Since the discovery of RNAi, astounding progress has been made regarding the study of non-coding small RNA biogenesis and function. By establishing a conserved mechanism of gene regulation, from repressing gene expression in a temporal and spatial manner, to directing epigenetic regulation that persists in a heritable manner, small RNAs are unmistakably seen as fundamental regulatory molecules that have supplanted the canonical role of RNA in the Central Dogma of molecular biology. It is undoubtedly the case that future biomedical research will reveal new and exciting insights into the diversity of mechanisms by which small RNAs regulate gene expression and the genome.

Bibliography

Albertson, D.G., and Thomson, J.N. (1982). The kinetochores of Caenorhabditis elegans. Chromosoma *86*, 409-428.

Alefelder, S., Patel, B.K., and Eckstein, F. (1998). Incorporation of terminal phosphorothioates into oligonucleotides. Nucleic Acids Res *26*, 4983-4988.

Almeida, R., and Allshire, R.C. (2005). RNA silencing and genome regulation. Trends Cell Biol *15*, 251-258.

Ambros, V., and Lee, R.C. (2004). Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. Methods Mol Biol *265*, 131-158.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in C. elegans. Curr Biol *13*, 807-818.

Anders, K.R., Grimson, A., and Anderson, P. (2003). SMG-5, required for C.elegans nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. Embo J *22*, 641-650.

Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K., and Tabara, H. (2007). In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. Embo J *26*, 5007-5019.

Aravin, A.A., and Hannon, G.J. (2008). Small RNA silencing pathways in germ and stem cells. Cold Spring Harb Symp Quant Biol *73*, 283-290.

Aroian, R.V., Field, C., Pruliere, G., Kenyon, C., and Alberts, B.M. (1997). Isolation of actin-associated proteins from Caenorhabditis elegans oocytes and their localization in the early embryo. Embo J *16*, 1541-1549.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

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

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S.*, et al.* (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol Cell *31*, 67-78.

Baulcombe, D. (1999). Viruses and gene silencing in plants. Arch Virol Suppl *15*, 189-201.

Beanan, M.J., and Strome, S. (1992). Characterization of a germ-line proliferation mutation in C. elegans. Development *116*, 755-766.

Behm-Ansmant, I., Kashima, I., Rehwinkel, J., Sauliere, J., Wittkopp, N., and Izaurralde, E. (2007). mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. FEBS Lett *581*, 2845-2853.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. Embo J *17*, 170-180.

Boisvert, M.E., and Simard, M.J. (2008). RNAi pathway in C. elegans: the argonautes and collaborators. Curr Top Microbiol Immunol *320*, 21-36.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

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

Buchwitz, B.J., Ahmad, K., Moore, L.L., Roth, M.B., and Henikoff, S. (1999). A histone-H3-like protein in C. elegans. Nature *401*, 547-548.

Buhler, M., Haas, W., Gygi, S.P., and Moazed, D. (2007). RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. Cell *129*, 707-721.

Buhler, M., and Moazed, D. (2007). Transcription and RNAi in heterochromatic gene silencing. Nat Struct Mol Biol *14*, 1041-1048.

Buhler, M., Spies, N., Bartel, D.P., and Moazed, D. (2008). TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway. Nat Struct Mol Biol *15*, 1015-1023.

Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873-886.

Caplen, N.J., Parrish, S., Imani, F., Fire, A., and Morgan, R.A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. Proc Natl Acad Sci U S A *98*, 9742-9747.

Carroll, C.W., and Straight, A.F. (2006). Centromere formation: from epigenetics to self-assembly. Trends Cell Biol *16*, 70-78.

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

Carvalho, P.C., Fischer, J.S., Chen, E.I., Yates, J.R., 3rd, and Barbosa, V.C. (2008). PatternLab for proteomics: a tool for differential shotgun proteomics. BMC Bioinformatics *9*, 316.

Celniker, S.E., Dillon, L.A., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M.*, et al.* (2009). Unlocking the secrets of the genome. Nature *459*, 927-930.

Chapman, E.J., and Carrington, J.C. (2007). Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet *8*, 884-896.

Cheeseman, I.M., and Desai, A. (2008). Molecular architecture of the kinetochore-microtubule interface. Nat Rev Mol Cell Biol *9*, 33-46.

Chekulaeva, M., and Filipowicz, W. (2009). Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. Curr Opin Cell Biol *21*, 452-460.

Chen, C.C., Simard, M.J., Tabara, H., Brownell, D.R., McCollough, J.A., and Mello, C.C. (2005). A member of the polymerase beta nucleotidyltransferase superfamily is required for RNA interference in C. elegans. Curr Biol *15*, 378-383.

Chu, D.S., Liu, H., Nix, P., Wu, T.F., Ralston, E.J., Yates, J.R., 3rd, and Meyer, B.J. (2006). Sperm chromatin proteomics identifies evolutionarily conserved fertility factors. Nature *443*, 101-105.

Chuang, C.F., and Meyerowitz, E.M. (2000). Specific and heritable genetic interference by double-stranded RNA in Arabidopsis thaliana. Proc Natl Acad Sci U S A *97*, 4985-4990.

Claycomb, J.M., Batista, P.J., Pang, K.M., Gu, W., Vasale, J.J., van Wolfswinkel, J.C., Chaves, D.A., Shirayama, M., Mitani, S., Ketting, R.F.*, et al.* (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell *139*, 123-134.

Claycomb, J.M., MacAlpine, D.M., Evans, J.G., Bell, S.P., and Orr-Weaver, T.L. (2002). Visualization of replication initiation and elongation in Drosophila. J Cell Biol *159*, 225-236.

Cogoni, C. (2001). Homology-dependent gene silencing mechanisms in fungi. Annu Rev Microbiol *55*, 381-406.

Collins, R.E., and Cheng, X. (2006). Structural and biochemical advances in mammalian RNAi. J Cell Biochem *99*, 1251-1266.

Colmenares, S.U., Buker, S.M., Buhler, M., Dlakic, M., and Moazed, D. (2007). Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. Mol Cell *27*, 449-461.

Conine, C.C., Batista, P.J., Gu, W., Claycomb, J.M., Chaves, D.A., Shirayama, M., and Mello, C.C. (2010). 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.

Couteau, F., Guerry, F., Muller, F., and Palladino, F. (2002). A heterochromatin protein 1 homologue in Caenorhabditis elegans acts in germline and vulval development. EMBO Rep *3*, 235-241.

Csankovszki, G., Collette, K., Spahl, K., Carey, J., Snyder, M., Petty, E., Patel, U., Tabuchi, T., Liu, H., McLeod, I.*, et al.* (2009). Three distinct condensin complexes control C. elegans chromosome dynamics. Curr Biol *19*, 9-19.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R.*, et al.* (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

D'Agostino, I., Merritt, C., Chen, P.L., Seydoux, G., and Subramaniam, K. (2006). Translational repression restricts expression of the C. elegans Nanos homolog NOS-2 to the embryonic germline. Dev Biol *292*, 244-252.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. (2000). An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell *101*, 543-553.

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R.*, et al.* (2008). 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.

Davis, E., Caiment, F., Tordoir, X., Cavaille, J., Ferguson-Smith, A., Cockett, N., Georges, M., and Charlier, C. (2005). RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. Curr Biol *15*, 743-749.

de Lange, P., van Blokland, R., Kooter, J.M., and Mol, J.N. (1995). Suppression of flavonoid flower pigmentation genes in Petunia hybrida by the introduction of antisense and sense genes. Curr Top Microbiol Immunol *197*, 57-75.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231-235.

Dernburg, A.F. (2001). Here, there, and everywhere: kinetochore function on holocentric chromosomes. J Cell Biol *153*, F33-38.

Ding, S.W., and Voinnet, O. (2007). Antiviral immunity directed by small RNAs. Cell *130*, 413-426.

Djupedal, I., and Ekwall, K. (2009). Epigenetics: heterochromatin meets RNAi. Cell Res *19*, 282-295.

Domeier, M.E., Morse, D.P., Knight, S.W., Portereiko, M., Bass, B.L., and Mango, S.E. (2000). A link between RNA interference and nonsense-mediated decay in Caenorhabditis elegans. Science *289*, 1928-1931.

Duchaine, T.F., Wohlschlegel, J.A., Kennedy, S., Bei, Y., Conte, D., Jr., Pang, K., Brownell, D.R., Harding, S., Mitani, S., Ruvkun, G*., et al.* (2006). Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways. Cell *124*, 343-354.

Duxbury, M.S., Ashley, S.W., and Whang, E.E. (2005). RNA interference: a mammalian SID-1 homologue enhances siRNA uptake and gene silencing efficacy in human cells. Biochem Biophys Res Commun *331*, 459-463.

Eberle, A.B., Lykke-Andersen, S., Muhlemann, O., and Jensen, T.H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat Struct Mol Biol *16*, 49-55.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res *30*, 207-210.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001a). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001b). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev *15*, 188-200.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001c). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. Embo J *20*, 6877-6888.

Fagard, M., Boutet, S., Morel, J.B., Bellini, C., and Vaucheret, H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc Natl Acad Sci U S A *97*, 11650-11654.

Felsenstein, J. (2005). Using the quantitative genetic threshold model for inferences between and within species. Philos Trans R Soc Lond B Biol Sci *360*, 1427-1434.

Fire, A., Albertson, D., Harrison, S.W., and Moerman, D.G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. Development *113*, 503-514.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Fjose, A., Ellingsen, S., Wargelius, A., and Seo, H.C. (2001). RNA interference: mechanisms and applications. Biotechnol Annu Rev *7*, 31-57.

Fox, S., Filichkin, S., and Mockler, T.C. (2009). Applications of ultra-high-throughput sequencing. Methods Mol Biol *553*, 79-108.

Frasch, M. (2008). A matter of timing: microRNA-controlled temporal identities in worms and flies. Genes Dev *22*, 1572-1576.

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

Gan, J., Tropea, J.E., Austin, B.P., Court, D.L., Waugh, D.S., and Ji, X. (2006). Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. Cell *124*, 355-366.

Gazzani, S., Lawrenson, T., Woodward, C., Headon, D., and Sablowski, R. (2004). A link between mRNA turnover and RNA interference in Arabidopsis. Science *306*, 1046-1048.

Gent, J.I., Schvarzstein, M., Villeneuve, A.M., Gu, S.G., Jantsch, V., Fire, A.Z., and Baudrimont, A. (2009). A Caenorhabditis elegans RNA-directed RNA polymerase in sperm development and endogenous RNA interference. Genetics *183*, 1297-1314.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z.*, et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat Rev Genet *10*, 94-108.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75-79.

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

Glavan, F., Behm-Ansmant, I., Izaurralde, E., and Conti, E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. Embo J *25*, 5117-5125.

Goday, C., Gonzalez-Garcia, J.M., Esteban, M.R., Giovinazzo, G., and Pimpinelli, S. (1992). Kinetochores and chromatin diminution in early embryos of Parascaris univalens. J Cell Biol *118*, 23-32.

Golden, D.E., Gerbasi, V.R., and Sontheimer, E.J. (2008). An inside job for siRNAs. Mol Cell *31*, 309-312.

Grimson, A., O'Connor, S., Newman, C.L., and Anderson, P. (2004). SMG-1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in Caenorhabditis elegans. Mol Cell Biol *24*, 7483-7490.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell *106*, 23-34.

Grishok, A., Tabara, H., and Mello, C.C. (2000). Genetic requirements for inheritance of RNAi in C. elegans. Science *287*, 2494-2497.

Gu, S.G., and Fire, A. (2010). Partitioning the C. elegans genome by nucleosome modification, occupancy, and positioning. Chromosoma *119*, 73-87.

Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J.*, et al.* (2009). Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the C. elegans Germline. Mol Cell.

Guang, S., Bochner, A.F., Pavelec, D.M., Burkhart, K.B., Harding, S., Lachowiec, J., and Kennedy, S. (2008). An Argonaute transports siRNAs from the cytoplasm to the nucleus. Science *321*, 537-541.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.

Hall, I.M., Noma, K., and Grewal, S.I. (2003). RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. Proc Natl Acad Sci U S A *100*, 193-198.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev *18*, 3016-3027.

Han, T., Manoharan, A.P., Harkins, T.T., Bouffard, P., Fitzpatrick, C., Chu, D.S., Thierry-Mieg, D., Thierry-Mieg, J., and Kim, J.K. (2009). 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans. Proc Natl Acad Sci U S A *106*, 18674-18679.

Hartig, J.V., Tomari, Y., and Forstemann, K. (2007). piRNAs--the ancient hunters of genome invaders. Genes Dev *21*, 1707-1713.

He, S., Zhang, D., Cheng, F., Gong, F., and Guo, Y. (2009a). Applications of RNA interference in cancer therapeutics as a powerful tool for suppressing gene expression. Mol Biol Rep *36*, 2153-2163.

He, Z., Kokkinaki, M., Pant, D., Gallicano, G.I., and Dym, M. (2009b). Small RNA molecules in the regulation of spermatogenesis. Reproduction *137*, 901-911.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118-120.

Herr, A.J., Molnar, A., Jones, A., and Baulcombe, D.C. (2006). Defective RNA processing enhances RNA silencing and influences flowering of Arabidopsis. Proc Natl Acad Sci U S A *103*, 14994-15001.

Hock, J., and Meister, G. (2008). The Argonaute protein family. Genome Biol *9*, 210.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol *17*, 1265-1272.

Hunter, C.P., Winston, W.M., Molodowitch, C., Feinberg, E.H., Shih, J., Sutherlin, M., Wright, A.J., and Fitzgerald, M.C. (2006). Systemic RNAi in Caenorhabditis elegans. Cold Spring Harb Symp Quant Biol *71*, 95-100.

Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. Rna *14*, 2609-2617.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol *9*, 22-32.

Jannot, G., Boisvert, M.E., Banville, I.H., and Simard, M.J. (2008). Two molecular features contribute to the Argonaute specificity for the microRNA and RNAi pathways in C. elegans. RNA *14*, 829-835.

Ji, X. (2008). The mechanism of RNase III action: how dicer dices. Curr Top Microbiol Immunol *320*, 99-116.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J. (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. *110*, 462-467

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

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol *5*, e57.

Kawasaki, I., Shim, Y.H., Kirchner, J., Kaminker, J., Wood, W.B., and Strome, S. (1998). PGL-1, a predicted RNA-binding component of germ granules, is essential for fertility in C. elegans. Cell *94*, 635-645.

Kennedy, S., Wang, D., and Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. Nature *427*, 645-649.

Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell *95*, 1017-1026.

Kennerdell, J.R., and Carthew, R.W. (2000). Heritable gene silencing in Drosophila using double-stranded RNA. Nat Biotechnol *18*, 896-898.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev *15*, 2654-2659.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133-141.

Ketting, R.F., and Plasterk, R.H. (2000). A genetic link between co-suppression and RNA interference in C. elegans. Nature *404*, 296-298.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

Kim, J.K., Gabel, H.W., Kamath, R.S., Tewari, M., Pasquinelli, A., Rual, J.F., Kennedy, S., Dybbs, M., Bertin, N., Kaplan, J.M.*, et al.* (2005). Functional genomic analysis of RNA interference in C. elegans. Science *308*, 1164-1167.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139.

Klahre, U., Crete, P., Leuenberger, S.A., Iglesias, V.A., and Meins, F., Jr. (2002). High molecular weight RNAs and small interfering RNAs induce systemic posttranscriptional gene silencing in plants. Proc Natl Acad Sci U S A *99*, 11981-11986.

Klattenhoff, C., and Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. Development *135*, 3-9.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kumar, A. (2008). RNA interference: a multifaceted innate antiviral defense. Retrovirology *5*, 17.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. Science *313*, 363-367.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

Lee, R.C., Hammell, C.M., and Ambros, V. (2006). Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans. Rna *12*, 589-597.

Lee, S.R., and Collins, K. (2007). Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. Nat Struct Mol Biol *14*, 604-610.

Lewis, D.L., Hagstrom, J.E., Loomis, A.G., Wolff, J.A., and Herweijer, H. (2002). Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. Nat Genet *32*, 107-108.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M.*, et al.* (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell *137*, 509-521.

Li, Y.X., Farrell, M.J., Liu, R., Mohanty, N., and Kirby, M.L. (2000). Double-stranded RNA injection produces null phenotypes in zebrafish. Dev Biol *217*, 394-405.

Liang, L., Diehl-Jones, W., and Lasko, P. (1994). Localization of vasa protein to the Drosophila pole plasm is independent of its RNA-binding and helicase activities. Development *120*, 1201-1211.

Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. (2003a). Vertebrate microRNA genes. Science *299*, 1540.

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

Lipardi, C., and Paterson, B.M. (2009). Identification of an RNA-dependent RNA polymerase in Drosophila involved in RNAi and transposon suppression. Proc Natl Acad Sci U S A *106*, 15645-15650.

Liu, Q., Feng, Y., and Zhu, Z. (2009). Dicer-like (DCL) proteins in plants. Funct Integr Genomics *9*, 277-286.

Liu, Q., and Paroo, Z. (2010). Biochemical Principles of Small RNA Pathways. Annu Rev Biochem.

Lohmann, J.U., Endl, I., and Bosch, T.C. (1999). Silencing of developmental genes in Hydra. Dev Biol *214*, 211-214.

Lykke-Andersen, K., Gilchrist, M.J., Grabarek, J.B., Das, P., Miska, E., and Zernicka-Goetz, M. (2008). Maternal Argonaute 2 is essential for early mouse development at the maternal-zygotic transition. Mol Biol Cell *19*, 4383-4392.

MacCoss, M.J., McDonald, W.H., Saraf, A., Sadygov, R., Clark, J.M., Tasto, J.J., Gould, K.L., Wolters, D., Washburn, M., Weiss, A*., et al.* (2002). Shotgun identification of protein modifications from protein complexes and lens tissue. Proc Natl Acad Sci U S A *99*, 7900-7905.

MacRae, I.J., and Doudna, J.A. (2007). Ribonuclease revisited: structural insights into ribonuclease III family enzymes. Curr Opin Struct Biol *17*, 138-145.

Maddox, P.S., Oegema, K., Desai, A., and Cheeseman, I.M. (2004). "Holo"er than thou: chromosome segregation and kinetochore function in C. elegans. Chromosome Res *12*, 641-653.

Maida, Y., Yasukawa, M., Furuuchi, M., Lassmann, T., Possemato, R., Okamoto, N., Kasim, V., Hayashizaki, Y., Hahn, W.C., and Masutomi, K. (2009). An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. Nature *461*, 230-235.

Maine, E.M., Hauth, J., Ratliff, T., Vought, V.E., She, X., and Kelly, W.G. (2005). EGO-1, a putative RNA-dependent RNA polymerase, is required for

heterochromatin assembly on unpaired dna during C. elegans meiosis. Curr Biol *15*, 1972-1978.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell *137*, 522-535.

Martens, H., Novotny, J., Oberstrass, J., Steck, T.L., Postlethwait, P., and Nellen, W. (2002). RNAi in Dictyostelium: the role of RNA-directed RNA polymerases and double-stranded RNase. Mol Biol Cell *13*, 445-453.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell *15*, 185-197.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature *431*, 343-349.

Meneely, P.M., Farago, A.F., and Kauffman, T.M. (2002). Crossover distribution and high interference for both the X chromosome and an autosome during oogenesis and spermatogenesis in Caenorhabditis elegans. Genetics *162*, 1169-1177.

Milhavet, O., Gary, D.S., and Mattson, M.P. (2003). RNA interference in biology and medicine. Pharmacol Rev *55*, 629-648.

Mito, Y., Sugimoto, A., and Yamamoto, M. (2003). Distinct developmental function of two Caenorhabditis elegans homologs of the cohesin subunit Scc1/Rad21. Mol Biol Cell *14*, 2399-2409.

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

Mochizuki, K., and Gorovsky, M.A. (2004a). Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement. Genes Dev *18*, 2068-2073.

Mochizuki, K., and Gorovsky, M.A. (2004b). Small RNAs in genome rearrangement in Tetrahymena. Curr Opin Genet Dev *14*, 181-187.

Morozova, O., Hirst, M., and Marra, M.A. (2009). Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genomics Hum Genet *10*, 135-151.

Morris, K.V. (2008). RNA-mediated transcriptional gene silencing in human cells. Curr Top Microbiol Immunol *320*, 211-224.

Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789-802.

Nagaki, K., Kashihara, K., and Murata, M. (2005). Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant Luzula nivea. Plant Cell *17*, 1886-1893.

Nakamura, M., Ando, R., Nakazawa, T., Yudazono, T., Tsutsumi, N., Hatanaka, N., Ohgake, T., Hanaoka, F., and Eki, T. (2007). Dicer-related drh-3 gene functions in germ-line development by maintenance of chromosomal integrity in Caenorhabditis elegans. Genes Cells *12*, 997-1010.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279-289.

Newmark, P.A., Reddien, P.W., Cebria, F., and Sanchez Alvarado, A. (2003). Ingestion of bacterially expressed double-stranded RNA inhibits gene expression in planarians. Proc Natl Acad Sci U S A *100 Suppl 1*, 11861-11865.

Ngo, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. Proc Natl Acad Sci U S A *95*, 14687-14692.

Nishida, K.M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S., Siomi, H., and Siomi, M.C. (2007). Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. Rna *13*, 1911-1922.

Nykanen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell *107*, 309-321.

Oegema, K., Desai, A., Rybina, S., Kirkham, M., and Hyman, A.A. (2001). Functional analysis of kinetochore assembly in Caenorhabditis elegans. J Cell Biol *153*, 1209-1226.

Okamura, K., Balla, S., Martin, R., Liu, N., and Lai, E.C. (2008). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. Nat Struct Mol Biol *15*, 581-590.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell *130*, 89-100.

Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev *18*, 1655-1666.

Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Dev Biol *216*, 671-680.

Page, M.F., Carr, B., Anders, K.R., Grimson, A., and Anderson, P. (1999). SMG-2 is a phosphorylated protein required for mRNA surveillance in Caenorhabditis elegans and related to Upf1p of yeast. Mol Cell Biol *19*, 5943-5951.

Pak, J., and Fire, A. (2007). Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science *315*, 241-244.

Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. (2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. Mol Cell *9*, 315-327.

Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669-672.

Palauqui, J.C., Elmayan, T., Pollien, J.M., and Vaucheret, H. (1997). Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. Embo J *16*, 4738-4745.

Parker, G.S., Eckert, D.M., and Bass, B.L. (2006). RDE-4 preferentially binds long dsRNA and its dimerization is necessary for cleavage of dsRNA to siRNA. Rna *12*, 807-818.

Parker, J.S., Roe, S.M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663-666.

Parrish, S., Fleenor, J., Xu, S., Mello, C., and Fire, A. (2000). Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. Mol Cell *6*, 1077-1087.

Pavelec, D.M., Lachowiec, J., Duchaine, T.F., Smith, H.E., and Kennedy, S. (2009). Requirement for ERI/DICER Complex in Endogenous RNAi and Sperm Development in Caenorhabditis elegans. Genetics.

Pimpinelli, S., and Goday, C. (1989). Unusual kinetochores and chromatin diminution in Parascaris. Trends Genet *5*, 310-315.

Pitt, J.N., Schisa, J.A., and Priess, J.R. (2000). P granules in the germ cells of Caenorhabditis elegans adults are associated with clusters of nuclear pores and contain RNA. Dev Biol *219*, 315-333.

Poirot, O., O'Toole, E., and Notredame, C. (2003). Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. Nucleic Acids Res *31*, 3503-3506.

Pontier, D.B., and Tijsterman, M. (2009). A robust network of double-strand break repair pathways governs genome integrity during C. elegans development. Curr Biol *19*, 1384-1388.

Praitis, V., Casey, E., Collar, D., and Austin, J. (2001). Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. Genetics *157*, 1217-1226.

Qiao, L., Lissemore, J.L., Shu, P., Smardon, A., Gelber, M.B., and Maine, E.M. (1995). Enhancers of glp-1, a gene required for cell-signaling in Caenorhabditis elegans, define a set of genes required for germline development. Genetics *141*, 551-569.

Rajyaguru, P., and Parker, R. (2009). CGH-1 and the control of maternal mRNAs. Trends Cell Biol *19*, 24-28.

Ratcliff, F., Harrison, B.D., and Baulcombe, D.C. (1997). A similarity between viral defense and gene silencing in plants. Science *276*, 1558-1560.

Ratcliff, F.G., MacFarlane, S.A., and Baulcombe, D.C. (1999). Gene silencing without DNA. rna-mediated cross-protection between viruses. Plant Cell *11*, 1207-1216.

Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

Reinke, V., Smith, H.E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S.J., Davis, E.B., Scherer, S., Ward, S*., et al.* (2000). A global profile of germline gene expression in C. elegans. Mol Cell *6*, 605-616.

Resnick, T.D., McCulloch, K.A., and Rougvie, A.E. (2010). miRNAs give worms the time of their lives: small RNAs and temporal control in Caenorhabditis elegans. Dev Dyn *239*, 1477-1489.

Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol *12*, 340-349.

Robert, V.J., Sijen, T., van Wolfswinkel, J., and Plasterk, R.H. (2005). Chromatin and RNAi factors protect the C. elegans germline against repetitive sequences. Genes Dev *19*, 782-787.

Rocheleau, C.E., Cullison, K., Huang, K., Bernstein, Y., Spilker, A.C., and Sundaram, M.V. (2008). The Caenorhabditis elegans ekl (enhancer of ksr-1 lethality) genes include putative components of a germline small RNA pathway. Genetics *178*, 1431-1443.

Rocheleau, C.E., Downs, W.D., Lin, R., Wittmann, C., Bei, Y., Cha, Y.H., Ali, M., Priess, J.R., and Mello, C.C. (1997). Wnt signaling and an APC-related gene specify endoderm in early C. elegans embryos. Cell *90*, 707-716.

Rocheleau, C.E., Yasuda, J., Shin, T.H., Lin, R., Sawa, H., Okano, H., Priess, J.R., Davis, R.J., and Mello, C.C. (1999). WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in C. elegans. Cell *97*, 717-726.

Roignant, J.Y., Carre, C., Mugat, B., Szymczak, D., Lepesant, J.A., and Antoniewski, C. (2003). Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in Drosophila. Rna *9*, 299-308.

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol Microbiol *6*, 3343-3353.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193-1207.

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83-86.

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev *21*, 1603-1608.

Sanchez Alvarado, A., and Newmark, P.A. (1999). Double-stranded RNA specifically disrupts gene expression during planarian regeneration. Proc Natl Acad Sci U S A *96*, 5049-5054.

Satzinger, H. (2008). Theodor and Marcella Boveri: chromosomes and cytoplasm in heredity and development. Nat Rev Genet *9*, 231-238.

Schiebel, W., Haas, B., Marinkovic, S., Klanner, A., and Sanger, H.L. (1993a). RNA-directed RNA polymerase from tomato leaves. I. Purification and physical properties. J Biol Chem *268*, 11851-11857.

Schiebel, W., Haas, B., Marinkovic, S., Klanner, A., and Sanger, H.L. (1993b). RNA-directed RNA polymerase from tomato leaves. II. Catalytic in vitro properties. J Biol Chem *268*, 11858-11867.

Schisa, J.A., Pitt, J.N., and Priess, J.R. (2001). Analysis of RNA associated with P granules in germ cells of C. elegans adults. Development *128*, 1287-1298.

Schutz, S., and Sarnow, P. (2006). Interaction of viruses with the mammalian RNA interference pathway. Virology *344*, 151-157.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.

Schwarz, D.S., Tomari, Y., and Zamore, P.D. (2004). The RNA-induced silencing complex is a Mg2+-dependent endonuclease. Curr Biol *14*, 787-791.

Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. Cell *125*, 287-300.

Seo, T.S., Bai, X., Ruparel, H., Li, Z., Turro, N.J., and Ju, J. (2004). Photocleavable fluorescent nucleotides for DNA sequencing on a chip constructed by site-specific coupling chemistry. Proc Natl Acad Sci U S A *101*, 5488-5493.

She, X., Xu, X., Fedotov, A., Kelly, W.G., and Maine, E.M. (2009). Regulation of heterochromatin assembly on unpaired chromosomes during Caenorhabditis elegans meiosis by components of a small RNA-mediated pathway. PLoS Genet *5*, e1000624.

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

Shrey, K., Suchit, A., Nishant, M., and Vibha, R. (2009). RNA interference: emerging diagnostics and therapeutics tool. Biochem Biophys Res Commun *386*, 273-277.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. Cell *107*, 465-476.

Sijen, T., and Plasterk, R.H. (2003). Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. Nature *426*, 310-314.

Sijen, T., Steiner, F.A., Thijssen, K.L., and Plasterk, R.H. (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science *315*, 244-247.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S.P., Nonet, M.L., Fire, A., Ahringer, J., and Plasterk, R.H. (2002). Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi. Curr Biol *12*, 1317-1319.

Siomi, H., and Siomi, M.C. (2009). On the road to reading the RNA-interference code. Nature *457*, 396-404.

Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine, E.M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. Curr Biol *10*, 169-178.

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

Stoeckius, M., Maaskola, J., Colombo, T., Rahn, H.P., Friedlander, M.R., Li, N., Chen, W., Piano, F., and Rajewsky, N. (2009). Large-scale sorting of C. elegans embryos reveals the dynamics of small RNA expression. Nat Methods *6*, 745-751.

Tabara, H., Grishok, A., and Mello, C.C. (1998). RNAi in C. elegans: soaking in the genome sequence. Science *282*, 430-431.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell *99*, 123-132.

Tabara, H., Yigit, E., Siomi, H., and Mello, C.C. (2002). 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.

Tabb, D.L., McDonald, W.H., and Yates, J.R., 3rd (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res *1*, 21-26.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M.*, et al.* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534-538.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol *14*, 1025-1040.

Thomson, T., and Lin, H. (2009). The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. Annu Rev Cell Dev Biol *25*, 355-376.

Tian, H. (2002). Combinatorial selection of RNA ligands for complex cellular targets : the RNA liagands-based proteomics. Mol Cell Proteomics *1*, 99-103.

Tijsterman, M., Ketting, R.F., Okihara, K.L., Sijen, T., and Plasterk, R.H. (2002). RNA helicase MUT-14-dependent gene silencing triggered in C. elegans by short antisense RNAs. Science *295*, 694-697.

Tijsterman, M., May, R.C., Simmer, F., Okihara, K.L., and Plasterk, R.H. (2004). Genes required for systemic RNA interference in Caenorhabditis elegans. Curr Biol *14*, 111-116.

Timmons, L., Court, D.L., and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. Gene *263*, 103-112.

Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature *395*, 854.

Timmons, L., Tabara, H., Mello, C.C., and Fire, A.Z. (2003). Inducible systemic RNA silencing in Caenorhabditis elegans. Mol Biol Cell *14*, 2972-2983.

Tops, B.B., Plasterk, R.H., and Ketting, R.F. (2006). The Caenorhabditis elegans Argonautes ALG-1 and ALG-2: almost identical yet different. Cold Spring Harb Symp Quant Biol *71*, 189-194.

Tops, B.B., Tabara, H., Sijen, T., Simmer, F., Mello, C.C., Plasterk, R.H., and Ketting, R.F. (2005). RDE-2 interacts with MUT-7 to mediate RNA interference in Caenorhabditis elegans. Nucleic Acids Res *33*, 347-355.

Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P., and Sharp, P.A. (1999). Targeted mRNA degradation by double-stranded RNA in vitro. Genes Dev *13*, 3191-3197.

Updike, D.L., and Strome, S. (2009). a genomewide rnai screen for genes that affect the stability, distribution and function of P granules in Caenorhabditis elegans. Genetics *183*, 1397-1419.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.

van der Krol, A.R., Mur, L.A., de Lange, P., Mol, J.N., and Stuitje, A.R. (1990). Inhibition of flower pigmentation by antisense CHS genes: promoter and minimal sequence requirements for the antisense effect. Plant Mol Biol *14*, 457-466.

van Wolfswinkel, J.C., Claycomb, J.M., Batista, P.J., Mello, C.C., Berezikov, E., and Ketting, R.F. (2009). CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. Cell *139*, 135-148.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell *16*, 69-79.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Verdel, A., Vavasseur, A., Le Gorrec, M., and Touat-Todeschini, L. (2009). Common themes in siRNA-mediated epigenetic silencing pathways. Int J Dev Biol *53*, 245-257.

Voinnet, O. (2005). Non-cell autonomous RNA silencing. FEBS Lett *579*, 5858-5871.

Voinnet, O., Pinto, Y.M., and Baulcombe, D.C. (1999). Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. Proc Natl Acad Sci U S A *96*, 14147-14152.

Voinnet, O., Vain, P., Angell, S., and Baulcombe, D.C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. Cell *95*, 177-187.

Volpe, T., Schramke, V., Hamilton, G.L., White, S.A., Teng, G., Martienssen, R.A., and Allshire, R.C. (2003). RNA interference is required for normal centromere function in fission yeast. Chromosome Res *11*, 137-146.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Vos, L.J., Famulski, J.K., and Chan, G.K. (2006). How to build a centromere: from centromeric and pericentromeric chromatin to kinetochore assembly. Biochem Cell Biol *84*, 619-639.

Vought, V.E., Ohmachi, M., Lee, M.H., and Maine, E.M. (2005). EGO-1, a putative RNA-directed RNA polymerase, promotes germline proliferation in parallel with GLP-1/notch signaling and regulates the spatial organization of nuclear pore complexes and germline P granules in Caenorhabditis elegans. Genetics *170*, 1121-1132.

Wang, G., and Reinke, V. (2008). A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. Curr Biol *18*, 861-867.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T.*, et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Welburn, J.P., and Cheeseman, I.M. (2008). Toward a molecular structure of the eukaryotic kinetochore. Dev Cell *15*, 645-655.

Welker, N.C., Pavelec, D.M., Nix, D.A., Duchaine, T.F., Kennedy, S., and Bass, B.L. (2010). Dicer's helicase domain is required for accumulation of some, but not all, C. elegans endogenous siRNAs. Rna *16*, 893-903.

Wianny, F., and Zernicka-Goetz, M. (2000). Specific interference with gene function by double-stranded RNA in early mouse development. Nat Cell Biol *2*, 70-75.

Winston, W.M., Sutherlin, M., Wright, A.J., Feinberg, E.H., and Hunter, C.P. (2007). Caenorhabditis elegans SID-2 is required for environmental RNA interference. Proc Natl Acad Sci U S A *104*, 10565-10570.

Wu-Scharf, D., Jeong, B., Zhang, C., and Cerutti, H. (2000). Transgene and transposon silencing in Chlamydomonas reinhardtii by a DEAH-box RNA helicase. Science *290*, 1159-1162.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol *2*, E104.

Yao, M.C., and Chao, J.L. (2005). RNA-guided DNA deletion in Tetrahymena: an RNAi-based mechanism for programmed genome rearrangements. Annu Rev Genet *39*, 537-559.

Yao, M.C., Fuller, P., and Xi, X. (2003). Programmed DNA deletion as an RNA-guided system of genome defense. Science *300*, 1581-1584.

Yeom, K.H., Lee, Y., Han, J., Suh, M.R., and Kim, V.N. (2006). Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. Nucleic Acids Res *34*, 4622-4629.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev *17*, 3011-3016.

Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.C., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., and Mello, C.C. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell *127*, 747-757.

Zamore, P.D. (2001). RNA interference: listening to the sound of silence. Nat Struct Biol *8*, 746-750.

Zamore, P.D. (2006). RNA interference: big applause for silencing in Stockholm. Cell *127*, 1083-1086.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25-33.

Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. Cell *118*, 57-68.